Analog Subspace Coding: A New Approach to Coding for Non-Coherent Wireless Networks

Mahdi Soleymani and Hessam Mahdavifar, *Member, IEEE*

***Abstract*—We provide a novel framework to study subspace codes for non-coherent communications in wireless networks. To this end, an *analog operator channel* is defined with inputs and outputs being subspaces of $\mathbb{C}^n$. Then a certain distance is defined to capture the performance of subspace codes in terms of their capability to recover from interference and rank-deficiency of the network. We also study the robustness of the proposed model with respect to an additive noise. Furthermore, we propose a new approach to construct subspace codes in the analog domain, also regarded as Grassmann codes, by leveraging polynomial evaluations over finite fields together with characters associated to finite fields that map their elements to the unit circle in the complex plane. The constructed codes, referred to as character-polynomial (CP) codes, are shown to perform better comparing to other existing constructions of Grassmann codes in terms of the trade-off between the rate and the normalized minimum distance, for a wide range of values for n.**

I. Introduction

Wireless networks are rapidly growing in size, are becoming more hierarchical, and are becoming increasingly distributed. In the next generation of wireless cellular networks, namely 5G, tens of small cells, hundreds of mobile users demanding ultra-high data rates, and thousands of Internet-of-Things (IoT) devices will be all operating within the coverage of one single cell [1]. While the efforts for 5G standardization are still ongoing, several new features have been introduced in the recent releases of the Long-Term Evolution (LTE) standard to start supporting the diverse requirements of the wide range of use cases in 5G. Started with Release 10 the deployment of small cells in LTE is becoming increasingly popular to deliver enhanced spectral capacity and extended network coverage [2], which is also fundamental to enhanced mobile broadband (eMBB) and massive machine type communications (mMTC) scenarios in 5G. Moreover, features such as coordinated multipoint (CoMP) transmission and reception [3] together with enhanced intercell interference coordination (eICIC) [4] have been introduced and used since Release 10 and evolved since then.

The aforementioned techniques are, however, difficult to scale as the number of small cells, that can be also regarded as relays, keeps increasing and as more layers are added in the hierarchical network. More specifically, conventional methods including channel estimation of point-to-point wireless links, link-level block coding, and successive interference cancellation do not properly scale with the size of such massive networks. Motivated by the emergence of such massive networks we study coding for wireless networks consisting of many relays operating in a non-coherent fashion, where the network nodes are oblivious to the channel gains of the point-to-point wireless links as well as the structure of the network. In a sense, this resembles a random linear network coding scenario, though completely in the physical layer, where physical-layer transport blocks are linearly combined in the relay nodes as they receive the spatial sum of blocks sent by the neighboring nodes. This holds assuming omni-directional radio frequency (RF) transmitter and receiver antennas are deployed at the network nodes. Also, in the considered setup, the relay nodes, such as small cells, do not attempt to decode messages and only amplify and forward the received physical-layer blocks.

In this paper, we define a new framework for reliable communications over wireless networks in a non-coherent fashion, as discussed above, using *analog subspace codes*. Let $\mathcal{W}$ denote an ambient vector space of dimension n over a field $\mathbb{L}$, i.e., $\mathcal{W} = \mathbb{L}^n$. A subspace code in $\mathcal{W}$ is a non-empty subset of the set of all the subspaces of $\mathcal{W}$. We observe that subspace codes in the analog domain, where the underlying field $\mathbb{L}$ is $\mathbb{R}$ or $\mathbb{C}$, become relevant for conveying information across networks in such a scenario.

This work is mainly inspired by the seminal work by Koetter and Kschischang [5], who defined a new framework for correcting errors and erasures in a randomized network coding scenario [6]. They defined an *operator channel* to capture the effect of errors and erasures in such a scenario and showed that subspace codes over finite fields are instrumental to provide reliability for communications over operator channels. In a sense, we develop a counterpart for Koetter-Kschischang's operator channel in the analog domain, referred to as *analog operator channel*. More specifically, the analog operator channel models the *rank-deficiency* of the network, caused by relay failures or lacking a sufficient number of active relays, as subspace erasures. Also, it models the interference from neighboring cells/small cells as subspace errors. We further discuss various methods for constructing subspace codes for the analog operator channel. In particular, we propose a novel construction method by leveraging characters associated to Abelian groups and finite fields, and mapping them to the unit circle in the complex plane.

It is worth noting that the setup considered in this paper fundamentally differs from Koetter-Kschischang's setup in two main aspects. First, due to the fundamental differences between the structure of finite fields and the analog fields of $\mathbb{R}$ or $\mathbb{C}$ constructing codes for analog operator channels requires entirely different approaches comparing to subspace codes constructed over finite fields in [5]. Second, the effect of physical layer is abstracted out in the setup considered in [5] as it is often the case in the network coding literature. However, in this work, we arrive at the notion of analog operator channels of subspace codes with an innovative perspective, namely, physical layer

The material in this paper was presented in part at the IEEE International Symposium on Information Theory in June 2020.

This work was supported by the National Science Foundation under grants CCF–1763348, CCF–1909771, and CCF–1941633.

M. Soleymani and H. Mahdavifar are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48104 (email: mahdy@umich.edu and hessam@umich.edu).

communications over wireless networks. Hence, we consider the additive noise that is always present in the physical layer, in addition to subspace errors and erasures discussed above, and characterize the robustness of the analog operator channel model with respect to the additive noise.

Analog subspace codes can be also viewed as codes in Grassmann space, also referred to as Grassmann codes, provided that the dimensions of all the subspace codewords are equal. There is a long history on studying bounds [7]–[11], using packing and covering arguments, and capacity analysis in Grassmann space, mostly motivated by space-time coding for multiple-input multiple-output (MIMO) wireless systems [12]–[14]. In such systems, a separate block code is needed to guarantee the reliability regardless, and the space-time code can be interpreted as the means of improving the reliability by exploiting the diversity the MIMO channel offers. However, we arrive at the problem of constructing subspace codes from the analog operator channel. In other words, subspace codes are used for reliable communications over analog operator channels the same way block codes are conventionally used for reliable communications over point-to-point links. A more detailed overview of prior works on Grassmann codes and their relations to our approach is provided later in Section II-C.

The rest of this paper is organized as follows. In Section II the analog operator channel is defined and an overview of the related prior work on Grassmann codes is provided. In Section III a new notion of subspace distance is defined and its relation with correcting subspace errors and erasures is discussed. The robustness of the analog operator channel with respect to the additive noise is analyzed in Section IV. Also, new constructions of analog subspace codes are discussed in Section V. Finally, the paper is concluded in Section VI.

II. Preliminaries

A. Notation Convention

Let $[n]$ denote the set of positive integers less than or equal to n, i.e., $[n] = \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$. Also, for $x \in \mathbb{R}$, $(x)_+ \stackrel{\text{def}}{=} \max(0, x)$.

In this paper, matrices are represented by bold capital letters. The row space of a matrix $\boldsymbol{X}$ is denoted by $\langle \boldsymbol{X} \rangle$. Also, for a square matrix $\boldsymbol{X}$, the trace of $\boldsymbol{X}$, denoted by $\mathrm{tr}(\boldsymbol{X})$, is defined to be the sum of elements of $\boldsymbol{X}$ on the main diagonal.

The ambient vector space is denoted by W. The parameter n is reserved for the dimension of W throughout the paper. Also, we have $W = \mathbb{L}^n$, where $\mathbb{L}$ can be either $\mathbb{R}$ or $\mathbb{C}$. In order to state results for $\mathbb{L}$, which could be either $\mathbb{R}$ or $\mathbb{C}$, a parameter β is defined, where $\beta = 1$ for $\mathbb{L} = \mathbb{R}$ and $\beta = 2$ for $\mathbb{L} = \mathbb{C}$. Let $\mathcal{P}(W)$ denote the set of all subspaces of W. For a subspace $V \in \mathcal{P}(W)$, the dimension of V is denoted by $\dim(V)$. The sum of two subspaces $U, V \in \mathcal{P}(W)$ is defined as

$$U + V \stackrel{\text{def}}{=} \{u + v : u \in U, v \in V\}. \tag{1}$$

Note that if U and V intersect trivially, i.e., $U \cap V = \{\boldsymbol{0}\}$, where $\boldsymbol{0}$ is the all-zero vector, then $U + V$ is a direct sum and is denoted by $U \oplus V$.

The set of all m-dimensional subspaces of $\mathbb{L}^n$ is denoted by $G_{m,n}(\mathbb{L})$, which is referred to as Grassmann space or Grassmannian in the literature. Given $\mathbb{L} = \mathbb{C}$, $G_{m,n}(\mathbb{C})$ can be also described as follows:

$$G_{m,n}(\mathbb{C}) \stackrel{\text{def}}{=} \{\langle \boldsymbol{Z} \rangle : \boldsymbol{Z} \in \mathbb{C}^{m \times n}, \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{H}} = \boldsymbol{I}_m\}, \tag{2}$$

where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix. The elements of $G_{m,n}(\mathbb{L})$ are also referred to as m-planes.

The Frobenius norm of a matrix $\boldsymbol{A}$ is defined as

$$\|\boldsymbol{A}\| \stackrel{\text{def}}{=} \sqrt{\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})} = \sqrt{\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})}. \tag{3}$$

By fixing a basis for W, any vector in W is represented by n-tuples of coordinates with respect to the chosen basis. The inner product between $\boldsymbol{u}, \boldsymbol{v} \in W$ is then defined as: $\boldsymbol{u}.\boldsymbol{v} \stackrel{\text{def}}{=} \sum_{i=1}^{n} u_i v_i$. Then the orthogonal subspace of $U \in \mathcal{P}(W)$ is defined as

$$U^{\perp} \stackrel{\text{def}}{=} \{\boldsymbol{v} \in W : \boldsymbol{u}.\boldsymbol{v} = 0, \forall \boldsymbol{u} \in U\}. \tag{4}$$

For a set M, a σ-quasimetric on M is a function $d : M \times M \to \mathbb{R}$ that satisfies all the conditions of a metric except the triangle inequality being relaxed to

$$\forall x, y, z \in M, \quad d(x, z) < \sigma\big(d(x, y) + d(y, z)\big), \tag{5}$$

for a constant $\sigma > 1$. This inequality is referred to as *σ-relaxed triangle inequality*.

B. Analog operator channel

This model is motivated by non-coherent communications over wireless networks, as discussed in Section I. Hence, each piece of the model is followed by a brief explanation from this perspective. Let $\boldsymbol{x}_i \in \mathbb{C}^n$, for $i \in [m]$, denote the input vectors. The input vectors, as physical layer transport blocks, can be sent by several antennas of a transmitter, e.g., a cellular base station, at different time frames. By discarding the interference and the additive noise, the output of the channel is a set of vectors $\boldsymbol{y}_j = \sum_{i=1}^{m} h_{j,i}\boldsymbol{x}_i$, where $j \in [l]$. Each vector $\boldsymbol{y}_j$ is the received transport block by an antenna of the receiver at a certain time frame. Note that a time-frame-level synchronization is assumed across the wireless links, e.g., by employing specific patterns in a designated subset of orthogonal frequency-division multiplexing (OFDM) symbols in each time frame as in LTE networks [15]. Also, the relays in the network, e.g., small cells, are assumed to be amplify-and-forward relays. They can forward a transport block, received during a certain time frame, in a subsequent time frame. This is because the communication is assumed to be done in the unit of time frame, i.e., the relay has to wait for the current time frame to end before it can begin forwarding what it received. Then, due to the different delays, in the unit of time frames, that the transport blocks may encounter as they are propagated through the network, the received $\boldsymbol{y}_j$'s can be the combination of transmitted $\boldsymbol{x}_i$'s across different antennas and time frames. Under a non-coherent scenario, both the transmitter and the receiver are oblivious to $h_{j,i}$'s, the topology of the network, and the link-level channel gains. It is possible that several interference blocks, e.g., up to t of them, from neighboring cells/small cells are also received by the receiver. Hence, we have

$$\boldsymbol{Y}_{l \times n} = \boldsymbol{H}_{l \times m}\boldsymbol{X}_{m \times n} + \boldsymbol{G}_{l \times t}\boldsymbol{E}_{t \times n}, \tag{6}$$

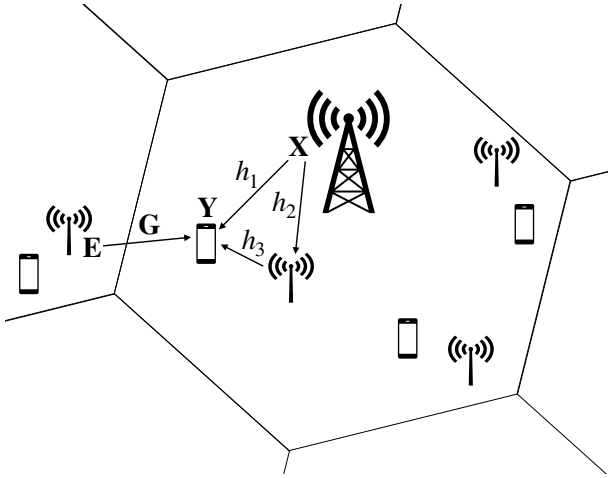

Fig. 1: An example of a non-coherent wireless network with the input-output relation specified in (6).

where $\boldsymbol{X}$'s rows are the transmitted blocks $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$, $\boldsymbol{E}$'s rows are the interference blocks $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_t$, $\boldsymbol{Y}$'s rows are the received blocks $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_l$, and $\boldsymbol{H} = [h_{j,i}]_{l\times m}$ and $\boldsymbol{G} = [g_{j,i}]_{l\times t}$ are assumed to be unknown to the transmitter and the receiver. Note that both $\boldsymbol{H}$ and $\boldsymbol{G}$ depend on the network topology as well as the link-level channel gains, however, $\boldsymbol{G}$ also depends on the specific nodes where the interference blocks have entered the network.

An example of the communication scenario, described by (6), is illustrated in Figure 1. Here, all the considered nodes have one transmit and one receive antennas and the communication is done in two time frames. We have $\boldsymbol{X} = [\boldsymbol{x}]_{1\times n}$, $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}_{2\times n}$, $\boldsymbol{E} = [\mathbf{e}]_{1\times n}$, $\boldsymbol{G} = \begin{bmatrix} g \\ 0 \end{bmatrix}_{2\times 1}$, and $\boldsymbol{H} = \begin{bmatrix} h_1 \\ h_2 h_3 \end{bmatrix}_{2\times 1}$. In other words, the receiver receives $h_1\boldsymbol{x} + g\mathbf{e}$ in the first time frame and receives $h_2 h_3 \boldsymbol{x}$ in the second time frame.

In the scenario described by (6), even in the absence of the interference blocks $\boldsymbol{E}$, the only way to convey information to the receiver is through the subspace spanned by the rows of $\boldsymbol{X}$. This is mainly due to the underlying assumption on non-coherent communications, where $\boldsymbol{H}$ is assumed to be completely unknown to both the transmitter and the receiver. Furthermore, $\boldsymbol{H}$ may not be full column rank, e.g., when $l < n$, which implies that $\langle \boldsymbol{X} \rangle$ can not be fully recovered. In order to capture the rank-deficiency of $\boldsymbol{H}$, a stochastic erasure operator is defined as follows. For some $k \geqslant 0$, $\mathcal{H}_k(U)$ returns a random k-dimensional subspace of U, if $\dim(U) > k$, and returns U otherwise. Then the analog operator channel is defined as follows:

Definition 1: An analog operator channel associated with W is a channel with input $U \in \mathcal{P}(W)$ and output $V \in \mathcal{P}(W)$ together with the following input-output relation:

$$V = \mathcal{H}_k(U) \oplus E, \tag{7}$$

where E is the interference subspace, also referred to as the error subspace, with $E \cap U = \{\mathbf{0}\}$. Then $\rho = \dim(U) - k$ is referred to as the dimension of erasures and $t = \dim(E)$ is referred to as the dimension of errors.

Remark 1. In the communication scenario described by (6), the additive noise of the physical layer, often modeled as additive white Gaussian noise (AWGN), is discarded. Note that the intermediate relay nodes in the wireless network, such as small cells, are not often limited by power constraints as the end mobile users are. Hence, it is natural to assume that the relay nodes operate at high signal-to-noise ratio (SNR). Nevertheless, it is essential to investigate the effect of additive noise as a perturbation of the transformation described by (6). In other words, instead of $\boldsymbol{Y}$, $\boldsymbol{Y} + \boldsymbol{N}$ is received by the receiver, where $\boldsymbol{N}$ is a matrix of i.i.d. Gaussian random variables. This, in turn, results in a perturbation in the analog operator channel, defined in Definition 1. We will discuss the robustness of the considered channel model with respect to the additive noise in Section IV.

C. Related prior work

The Grassmann space can be turned into a metric space using *chordal* distance. Roughly speaking, *chordal* distance generalizes the notion of angle between two lines to subspaces of equal dimension, which will be defined more precisely in the next section. Let δ_c denote the *chordal* distance normalized by n. Also, let the rate R of a code $\mathcal{C}$ be defined as $\ln|\mathcal{C}|/n$.

The problem of deriving bounds on the minimum distance of subspace codes of fixed dimension, i.e., packing subspaces in $G_{m,n}(\mathbb{R})$, was first studied by Shannon for the special case of $m = 1$ [7]. The packing problem in $G_{1,n}(\mathbb{R})$ is also related to designing spherical codes, i.e., packing points on a hyper sphere in the Euclidean space with a given angular separation. A lower bound on the best rate R of $\mathcal{C}$ with the minimum angle θ, i.e., $\delta_c = \sin(\theta)$, is derived by Shannon [7] as $R > -\ln(\sin(\theta))$, assuming $n \to \infty$. Lower and upper bounds on the largest achievable rate R, given a fixed δ_c and m while $n \to \infty$, were derived in [8]. The upper bound was later improved in [9]–[11].

In terms of lower bounds for the packing problem in the Grassmann space, an achievability bound for the minimum distance of the codes for finite values of n was derived in [16]. The construction of subspace codes in $G_{m,n}(\mathbb{R})$ and analysis of their minimum *chordal* distance was also studied in [17], and some of these constructions were observed to be optimal. However, the suggested construction methods in [17] are numerical, making them computationally infeasible for general parameters. Another numerical method for constructing codes based on alternative projection is proposed in [18]. In another line of work, motivated by quantum error-correcting codes, constructions based on group structures are suggested [19]–[23] (see, e.g., [24] for a brief survey). A connection between these codes and massive multiple access scenarios is observed in [25], where low-complexity decoders are also proposed. Another related line of work is the *frame* design problem in a Hilbert space, where a frame is a set of overcomplete unit norm vectors having small mutual inner product. Such a design has applications to a broad range of problems, including problems in signal processing, distributed sensing, parallel processing, etc [26]–[30].

The most notable line of work on Grassmann codes is motivated by space-time coding for MIMO channels. In other

words, the problem of constellation design for communications over a non-coherent MIMO is observed to be closely related to the packing problem in the complex Grassmann space. In this line of work, initiated by Marzetta and Hochwald [12], it is assumed that neither the transmitter nor the receiver knows the fading coefficients of the channel which are assumed to be fixed in the coherence time of the channel. They also proposed a constellation scheme in the Grassmann space called unitary space-time modulation [13]. Zheng and Tse [14] derived the capacity of non-coherent MIMO channel in high SNR in terms of the channel coherence time and the number of receive and transmit antennas. They further give a geometric interpretation of the capacity expression as sphere packing in the Grassmann manifold [14]. Furthermore, it is observed by Han and Rosenthal [31] that maximizing the chordal distance is the appropriate design criterion for the design of unitary space-time constellation at the low-SNR regime in non-coherent wireless communication systems. In another related work by Hochwald *et al.* , a lower bound on pairwise error probability between two subspaces of equal dimension is derived in [32], which is then used to derive a certain criterion for constellation design by maximizing the *chordal* distance between subspaces in the Grassmann space. They also provide a Fourier-based construction method and its equivalent algebraic construction employing linear codes to design constellations for non-coherent MIMO channels. Their approach involves random search procedure and, hence, is not scalable with n. Another numerical optimization method is described in [33]. Other constructions of Grassmann packings based on Nordstrom-Robinson Codes [34] and Reed-Muller codes [35] were also shown to closely approximate the channel capacity of non-coherent MIMO channels.

Another major related line of work includes a wide range of signal processing tasks where the pairwise geometry of subspaces plays an important role in characterizing the performance of the system, e.g., the misclassification probability for the optimal maximum-a-posteriori (MAP) classifier in a Gaussian mixture model (GMM). In particular, a duality between the problems of classification of k-dimensional subspaces from noisy features and the communication over non-coherent MIMO channel are observed in [36] and [37]. Also, the results on the capacity of non-coherent MIMO channels are used to provide necessary conditions for successful classification in [38]. The probability of misclassification is further analyzed in [39] and is characterized in terms of the principal angles.

Note that non-coherent MIMO communication can be considered as a special case of the non-coherent wireless networking scenario, described by (6), where there is no interference term while discarding additive noise. Also, there is no relay node and the number of transmitter and receiver antennas are known. In other words, the structure of the network and consequently, the rank of transform matrix $\boldsymbol{H}$ is known. Accordingly, prior works on non-coherent MIMO do not deal with subspace errors and erasures, and the underlying communication channel model is totally different from the analog operator channel considered in this paper, as defined in Definition 1.

Note also that the aforementioned prior works on code constructions in the Grassmann space do not often provide *good* solutions, in terms of the trade-off between R and the normalized minimum distance, for general m and n. However, a comparison, in terms of the trade-off between R and the normalized minimum distance, between a new construction of analog subspace codes proposed in this paper and the best existing constructions in the literature is done in Section V-C.

III. Analog Metric Space, Subspace Codes, and Error Correction

In this section we provide a precise description of the *chordal* distance defined for Grassmann space. Then we extend and modify the *chordal* distance to arrive at a new notion of distance, defined for the set of all subspaces of the ambient space, i.e., $\mathcal{P}(W)$, and show that it conveniently captures the error-correction capability of subspace codes when used over analog operator channels.

Given the structure of analog operator channels and their input and output alphabets being $\mathcal{P}(W)$, as defined in Definition 1, it is natural to design codes over $\mathcal{P}(W)$ in order to correct errors and erasures associated with such channels. To this end, the first step is to define a distance function that *properly* captures the effect of errors and erasures imposed by the analog operator channel. Before that, we discuss the *chordal* distance between two m-planes, which makes $G_{m,n}(\mathbb{L})$ a metric space.

The chordal distance $d_c : G_{m,n}(\mathbb{L}) \times G_{m,n}(\mathbb{L}) \rightarrow \mathbb{R}$ was first introduced for $\mathbb{L} = \mathbb{R}$ in [17] and was extended to $\mathbb{L} = \mathbb{C}$ in [8]. Consider two m-planes U and V. Let $\boldsymbol{u}_i \in U$ and $\boldsymbol{v}_i \in V$ be row vectors having unit length such that $|\boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}}|$ is maximal, subject to the conditions $\boldsymbol{u}_i \boldsymbol{u}_j^{\mathrm{H}} = 0$ and $\boldsymbol{v}_i \boldsymbol{v}_j^{\mathrm{H}} = 0$ for all i, j with $i > j \geqslant 1$. Then the *principal* angle θ_i, for $i \in [m]$, between U and V is defined as $\theta_i = \arccos |\boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}}|$, see, e.g., [8], [40]. Then the chordal distance between U and V is defined as follows:

$$d_c(U, V) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{m} \sin^2(\theta_i)}. \tag{8}$$

Note that this is not the only possible definition for distance between subspaces (see, e.g., [8], [41] for other similar notions). However, we focus on extending a certain variation of this notion of distance in this paper, to be discussed next. We will show later that it can be made suitable for capturing the error correcting capabilities of subspace codes for the analog operator channel.

Let $\boldsymbol{Z}$ denote an orthonormal matrix spanning $V \in G_{m,n}(\mathbb{L})$, i.e. ,

$$V = \langle \boldsymbol{Z} \rangle, \quad \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{H}} = \boldsymbol{I}_m.$$

Then, the matrix $\boldsymbol{P}_V = \boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z}$ is an orthogonal projection operator from $\mathbb{L}^n$ on V. It is shown in [17] that $G_{m,n}(\mathbb{R})$ with chordal distance can be isometrically embedded into a sphere in the Euclidean space $\mathbb{R}^D$, where $D = \binom{n+1}{2} - 1$, using the projection matrices associated with subspaces. More specifically,

$$\left\| \boldsymbol{P}_V - \frac{m}{n} \boldsymbol{I}_n \right\|^2 = \sqrt{\frac{m(n-m)}{n}}, \tag{9}$$

for all $V \in G_{m,n}(\mathbb{R})$. It is also shown that the chordal distance between two m-planes is equal to

$$d_c(U,V) = \frac{1}{\sqrt{2}} \left\| \boldsymbol{P}_U - \boldsymbol{P}_V \right\|. \tag{10}$$

Moreover, $G_{0,n}(\mathbb{R}), G_{1,n}(\mathbb{R}), \ldots, G_{n,n}(\mathbb{R})$ can also be embedded into a larger sphere in $\mathbb{R}^{D+1}$, i.e., for all subspaces V of the ambient space $\mathbb{R}^n$ we have

$$\left\| \boldsymbol{P}_V - \frac{1}{2}\boldsymbol{I}_n \right\|^2 = \frac{1}{4}n. \tag{11}$$

Note that $G_{m,n}(\mathbb{R})$ is the intersection of this sphere with the plane described by $\mathrm{tr}(\boldsymbol{P}_V) = m$, which is characterized by (9) (see [17, Figure 10]). Since the Frobenius norm induces a metric on the set of all $n \times n$ matrices, regardless of whether they are projection matrices or not, one can use (10) to generalize the notion of chordal distance to subspaces of different dimensions. This generalized distance is the Euclidean norm of the *chord* connecting the points associated with U and V on the sphere characterized by (11) normalized by $\sqrt{2}$. Note that, this definition coincides with (8) if U and V have equal dimensions. The proposed generalization of the chordal distance definition that also includes subspaces with different dimensions is similar to the one considered in [35, Definition 1]. The only minor difference is that the one considered in [35, Definition 1] has an extra multiplicative factor of $\sqrt{2}$.

Note also that principal angles, used in the definition of chordal distance, do not depend on the choice of basis for the ambient space. In other words, roughly speaking, the chordal distance is invariant under rotation of subspaces. This is shown more formally in the following lemma.

Lemma 1: Let $U, V \in \mathcal{P}(W)$. Given any two orthonormal bases for W, namely $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ and $\{\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n\}$, referred to as basis 1 and basis 2, respectively, we have

$$\left\| \boldsymbol{P}_U - \boldsymbol{P}_V \right\| = \left\| \boldsymbol{P}'_U - \boldsymbol{P}'_V \right\|,$$

where $\boldsymbol{P}_V$ and $\boldsymbol{P}'_V$ are matrix representations for the orthogonal projection operator on V in basis 1 and basis 2, respectively.

Proof: Let $\boldsymbol{Z}$ and $\boldsymbol{T}$ be matrices with orthonormal rows, represented in basis 1, which span U and V, respectively. Then, they are represented in basis 2 as follows:

$$\boldsymbol{Z}' = \boldsymbol{Z}\boldsymbol{Q}, \quad \boldsymbol{T}' = \boldsymbol{T}\boldsymbol{Q},$$

for a unitary matrix $\boldsymbol{Q}$. Then we have the following series of equalities by noting that $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$ for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that both $\boldsymbol{B}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{B}$ are well-defined and that the projection matrices are Hermitian:

$$\begin{aligned}
&\left\| \boldsymbol{P}'_U - \boldsymbol{P}'_V \right\| = \left\| \boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{P}_V - \boldsymbol{P}_U)\boldsymbol{Q} \right\| \\
&= \sqrt{\mathrm{tr}(\boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{P}_V - \boldsymbol{P}_U)\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{P}_V - \boldsymbol{P}_U)\boldsymbol{Q})} \\
&= \sqrt{\mathrm{tr}(\boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{P}_V - \boldsymbol{P}_U)^2\boldsymbol{Q})} = \sqrt{\mathrm{tr}(\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}}(\boldsymbol{P}_V - \boldsymbol{P}_U)^2)} \\
&= \sqrt{\mathrm{tr}((\boldsymbol{P}_V - \boldsymbol{P}_U)^2)} = \left\| \boldsymbol{P}_U - \boldsymbol{P}_V \right\|,
\end{aligned}$$

which complete the proof. ∎

The generalized chordal distance, discussed above, is further modified to arrive at a new notion of *distance* over $\mathcal{P}(W)$, defined as follows.

Definition 2: The *distance* $d : \mathcal{P}(W) \times \mathcal{P}(W) \to \mathbb{R}$ is defined as

$$d(U,V) \stackrel{\mathrm{def}}{=} \left\| \boldsymbol{P}_U - \boldsymbol{P}_V \right\|^2 = \mathrm{tr}\big((\boldsymbol{P}_U - \boldsymbol{P}_V)^2\big), \tag{12}$$

where $U, V \in \mathcal{P}(W)$ and $\boldsymbol{P}_U, \boldsymbol{P}_V$ are the projection matrices associated to U, V, respectively.

Note that Lemma 1 implies that $d(.,.)$ is well-defined. Note also that $d(.,.) = 2d_c(.,.)^2$ by (10) and (12) and, equivalently, is equal to the square of the metric considered in [35, Definition 1]. It is shown in Lemma 15 in the appendix that the square of a metric is a 2-quasimetric, where a quasimetric is defined in Section II-A. Hence, $d(.,.)$ is a 2-quasimetric. It is further shown in Lemma 16 in the appendix that for $U, V, T \in \mathcal{P}(W)$, $d(.,.)$ satisfies the triangle inequality, i.e., $\sigma = 1$ in (5), as long as P_U and P_V are simultaneously diagonalizable, i.e., one can find a basis in which both P_U and P_V are diagonal matrices. This property is later utilized to characterize the error-and-erasure correction capability of codes used over analog operator channels in terms of their *minimum distance* the same way it is done given an underlying metric. Hence, we refer to $d(.,.)$ as a distance through the rest of this paper keeping in mind that it is indeed a 2-quasimetric.

An equivalent expression for the distance $d(.,.)$ is derived in the following lemma.

Lemma 2: For any $\langle \boldsymbol{Z} \rangle, \langle \boldsymbol{T} \rangle \in G_{m,n}(\mathbb{L})$, where the rows of $\boldsymbol{Z}$ and $\boldsymbol{T}$ are orthonormal, we have

$$d(\langle \boldsymbol{Z} \rangle, \langle \boldsymbol{T} \rangle) = 2(m - \left\| \boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}} \right\|^2).$$

Proof: By noting that $\mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{H}}) = \mathrm{tr}(\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}) = m$ and by using (12) one can write

$$\begin{aligned}
&d(\langle \boldsymbol{Z} \rangle, \langle \boldsymbol{T} \rangle) = \mathrm{tr}((\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z} - \boldsymbol{T}^{\mathrm{H}}\boldsymbol{T})^2) = 2(m - \mathrm{tr}(\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T})) \\
&= 2(m - \mathrm{tr}(\boldsymbol{T}\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}})) = 2(m - \mathrm{tr}((\boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}})^{\mathrm{H}}(\boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}}))) \\
&= 2(m - \left\| \boldsymbol{Z}\boldsymbol{T}^{\mathrm{H}} \right\|^2),
\end{aligned}$$

which completes the proof. ∎

Definition 3: An analog subspace code $\mathcal{C}$ is a subset of $\mathcal{P}(W)$. The size of $\mathcal{C}$ is denoted by $|\mathcal{C}|$. The minimum distance of $\mathcal{C}$ is defined as

$$d_{\min}(\mathcal{C}) \stackrel{\mathrm{def}}{=} \min_{U,V \in \mathcal{C}, U \neq V} d(U,V),$$

where $d(.,.)$ is defined in Definition 2. The maximum dimension of the codewords of $\mathcal{C}$ is denoted by

$$l(\mathcal{C}) \stackrel{\mathrm{def}}{=} \max_{U \in \mathcal{C}} \dim(U).$$

The code $\mathcal{C}$ is then referred to as an $[n, l(\mathcal{C}), |\mathcal{C}|, d_{\min}(\mathcal{C})]$ subspace code, where n is the dimension of the ambient space W.

If the dimension of all codewords in $\mathcal{C}$ are equal, then the code is referred to as a *constant-dimension* code, which is also called a code on Grassmannian or a Grassmann code in the literature.

The *dual* subspace code associated with subspace code $\mathcal{C}$ is the code $\mathcal{C}^{\perp} \stackrel{\mathrm{def}}{=} \{U^{\perp} : U \in \mathcal{C}\}$. Lemma 17 implies that $d_{\min}(\mathcal{C}^{\perp}) = d_{\min}(\mathcal{C})$. Note that if $\mathcal{C}$ is a constant-dimension

code of type $[n, l, M, d_{\min}]$, then $\mathcal{C}^{\perp}$ is a constant-dimension code of type $[n, n-l, M, d_{\min}]$.

Definition 4: Let $\mathcal{C}$ be an $[n, l, M, d_{\min}(\mathcal{C})]$ subspace code. The normalized weight λ, the rate R, and the normalized minimum distance δ of $\mathcal{C}$ are defined as follows:

$$\lambda \stackrel{\text{def}}{=} \frac{l}{n}, \quad R \stackrel{\text{def}}{=} \frac{\ln M}{n}, \quad \delta \stackrel{\text{def}}{=} \frac{d_{\min}(\mathcal{C})}{2l}.$$

Note that the normalized weight λ and the normalized minimum distance δ are always between 0 and 1. However, while designing constant-dimension codes one can limit the attention to $\lambda \in [0, \frac{1}{2}]$. This is because for any code $\mathcal{C}$ with $l > \frac{n}{2}$, there exists a dual code $\mathcal{C}^{\perp}$ with $l < \frac{n}{2}$ and having the same distance properties.

As in conventional block codes, one can associate a minimum distance decoder to a subspace code $\mathcal{C}$, e.g., when used for communication over an analog operator channel, in order to recover from subspace errors and erasures. Such a decoder returns the nearest codeword $V \in \mathcal{C}$ given $U \in \mathcal{P}(W)$ as its input, i.e., for any $V' \in \mathcal{C}$, $d(U, V) \leqslant d(U, V')$. The following lemma plays a key rule in relating the minimum distance of $\mathcal{C}$ to its error-and-erasure correction capability under minimum distance decoding.

Lemma 3: Let $U, V \in \mathcal{P}(W)$ denote the input and the output of an analog operator channel, respectively, with the relation specified in (7). Then for any $T \in \mathcal{P}(W)$ we have

$$d(U, T) \leqslant \rho + t + d(V, T), \tag{13}$$

where ρ and t denote the dimension of erasures and errors, respectively, as specified in Definition 1.

Proof: Let $U' = \mathcal{H}_k(U)$. Then $d(U, U') = \rho$ by Lemma 18. Also, as shown in the proof of Lemma 18, P_U and $P_{U'}$ are simultaneously diagonalizable. Hence, by Lemma 16 we have

$$d(U, T) \leqslant \rho + d(U', T), \tag{14}$$

for any $T \in \mathcal{P}(W)$. Moreover, $V = U' \oplus E$, where E denotes the error space with $\dim(E) = t$. Using the same argument $P_{U'}$ and P_V are also simultaneously diagonalizable. Similarly, by Lemma 16 we have

$$d(U', T) \leqslant t + d(V, T), \tag{15}$$

for any $T \in \mathcal{P}(W)$. Combining (14) with (15) completes the proof. ∎

Theorem 4: Consider a subspace code $\mathcal{C}$ used for communication over an analog operator channel, as defined in Definition 1, i.e., the input to the channel is $U \in \mathcal{C}$. Let t and ρ denote the dimension of errors and erasures, respectively. Then the minimum distance decoder successfully recovers the transmitted codeword $U \in \mathcal{C}$ from the received subspace V if

$$2(\rho + t) < d_{\min}(\mathcal{C}), \tag{16}$$

where $d_{\min}(\mathcal{C})$ is the minimum distance of $\mathcal{C}$ defined in Definition 3.

Proof: By Lemma 3 we have

$$d(U, T) \leqslant \rho + t + d(V, T), \tag{17}$$

for any $T \in \mathcal{P}(W)$. In particular,

$$d(U, V) \leqslant \rho + t, \tag{18}$$

by letting $T = V$. Now, let $T \in \mathcal{C}$ be a codeword other than U. By (16) and the definition of minimum distance we have

$$d(U, T) \geqslant d_{\min}(\mathcal{C}) > 2(\rho + t). \tag{19}$$

This together with (17) and (18) yields

$$d(V, T) > \rho + t \geqslant d(U, V). \tag{20}$$

Hence, the minimum-distance decoder returns U. ∎

Theorem 4 implies that erasures and errors have equal costs in the subspace domain as far as the minimum-distance decoder is concerned. In other words, the minimum-distance decoder for a code $\mathcal{C}$ can correct up to $\left\lfloor \frac{d_{\min}(\mathcal{C})-1}{2} \right\rfloor$ errors and erasures.

Remark 2. If one uses the chordal distance, instead of the distance $d(.,.)$ defined in Definition 2, and follows the similar arguments as we did in this section, a result similar to Theorem 4 can be obtained while the condition in (16) is replaced by $\sqrt{2}(\sqrt{\rho} + \sqrt{t})$ being strictly less than the minimum chordal distance of the code. Since $d(.,.) = 2d_c(.,.)^2$, where $d_c(.,.)$ is the generalized chordal distance, this condition can be expressed in terms of $d_{\min}(\mathcal{C})$ as follows:

$$4(\sqrt{\rho} + \sqrt{t})^2 < d_{\min}(\mathcal{C}). \tag{21}$$

Note that the left hand side of (21) is greater than that of (16) by a multiplicative factor that is between 2 and 4. This shows the clear advantage in using the new distance $d(.,.)$ instead of the chordal distance in characterizing the error-and-erasure correction capability of analog subspace codes. The advantage is due to the fact that although $d(.,.)$ does not always satisfy the triangle inequality, it exhibits properties of a metric when dealing with inputs and outputs of analog operator channels.

IV. Robustness Against Additive Noise

In this section, we analyze the robustness of the analog operator channel in the presence of an additive noise.

The additive noise is denoted by $\boldsymbol{N} \in \mathbb{C}^{l \times n}$, referred to as the *noise matrix*. In the presence of the additive noise, the transform equation described in (6) is extended as follows:

$$\boldsymbol{Y}_{l \times n} = \boldsymbol{H}_{l \times m}\boldsymbol{X}_{m \times n} + \boldsymbol{G}_{l \times t}\boldsymbol{E}_{t \times n} + \boldsymbol{N}_{l \times n}. \tag{22}$$

More specifically, the effect of all the noise terms added to the blocks across the wireless network is included in the noise matrix $\boldsymbol{N}$. For ease of notation, let $\boldsymbol{A}$ denote the term $\boldsymbol{HX} + \boldsymbol{GE}$, consisting of terms associated to the transmitted blocks as well as the interference blocks, referred to as the *signal matrix*.

In the rest of this section, we aim at characterizing the perturbation imposed by the additive noise in terms of the subspace distance. In other words, we derive bounds on subspace distance between the signal matrix and the signal matrix perturbed by noise, i.e., $d(\langle \boldsymbol{A} \rangle, \langle \boldsymbol{A} + \boldsymbol{N} \rangle)$, in terms of various characteristics of both the noise and the signal matrix.

First, we consider the case where $\boldsymbol{A}$ is full row rank. In a sense, this implies that the receiver does not *oversample* from the network. In this case, we show that the subspace distance caused by the noise is bounded in terms of two parameters: (1)

the ratio of the maximum singular value, also referred to as the spectral norm, of the signal matrix to the Frobenius norm of the noise matrix, i.e., $\frac{\|A\|_2}{\|N\|}$; and (2) the spectral norm condition number of the signal matrix, defined as

$$\kappa(\boldsymbol{A}) \stackrel{\text{def}}{=} \|\boldsymbol{A}\|_2 \|\boldsymbol{A}^+\|_2, \tag{23}$$

where $\boldsymbol{A}^+$ is the pseudo-inverse of $\boldsymbol{A}$, defined as $\boldsymbol{A}^+ \stackrel{\text{def}}{=} \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}$. It is well-known that $\kappa(\boldsymbol{A}) = \frac{\sigma_{\max}}{\sigma_{min}} \geqslant 1$, where $\sigma_{\max} = \|\boldsymbol{A}\|_2$ and $\sigma_{\min}$ are the largest and the smallest singular values of $\boldsymbol{A}$, respectively.

Our analysis is based on the analysis of perturbation of RQ-factorization, see., e.g., [42]. Recall, from the linear algebra literature, that the RQ-factorization of a given matrix $\boldsymbol{A} \in \mathbb{C}^{l\times n}$ decomposes it as $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{Q}$, where $\boldsymbol{Q} \in \mathbb{C}^{l\times n}$ is an orthonormal matrix with $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}} = \boldsymbol{I}_l$ and $\boldsymbol{R} \in \mathbb{C}^{l\times l}$ is an upper triangular matrix with positive diagonal elements. It is well known that the RQ-factorization is unique [43]. The following result relates the perturbation of $\boldsymbol{A}$ to the perturbation of $\boldsymbol{Q}$ in its RQ-factorization.

Theorem 5: [42, Theorem 1.6] (rephrased) Let $\boldsymbol{A} \in \mathbb{C}^{l\times n}$ be a full row rank matrix with RQ-factorization $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{Q}$. Then for any $\boldsymbol{E} \in \mathbb{C}^{l\times n}$ that satisfies $\left\|\boldsymbol{A}^+\right\|_2 \|\boldsymbol{E}\|_2 < 1$, we have the following RQ-factorization for $\boldsymbol{A}+\boldsymbol{E}$:

$$\boldsymbol{A}+\boldsymbol{E} = (\boldsymbol{R}+\boldsymbol{R}_\Delta)(\boldsymbol{Q}+\boldsymbol{Q}_\Delta),$$

with $(\boldsymbol{Q}+\boldsymbol{Q}_\Delta)(\boldsymbol{Q}+\boldsymbol{Q}_\Delta)^{\mathrm{H}} = \boldsymbol{I}_l$ such that

$$\|\boldsymbol{Q}_\Delta\| \leqslant \frac{(1+\sqrt{2})\kappa(\boldsymbol{A})}{1-\|\boldsymbol{A}^+\|_2\|\boldsymbol{E}\|_2}.\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|_2}. \tag{24}$$

An upper bound on the subspace distance caused by the noise is derived in the following theorem.

Theorem 6: Let $\boldsymbol{A} \in \mathbb{C}^{l\times n}$ be a full row rank matrix. Then for any $\boldsymbol{E} \in \mathbb{C}^{l\times n}$ that satisfies $\left\|\boldsymbol{A}^+\right\|_2 \|\boldsymbol{E}\|_2 < 1$, we have

$$d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{E}\rangle) \leqslant 2\epsilon+\epsilon^2, \tag{25}$$

where

$$\epsilon \stackrel{\text{def}}{=} \left(\frac{(1+\sqrt{2})\kappa(\boldsymbol{A})}{1-\|\boldsymbol{A}^+\|_2\|\boldsymbol{E}\|_2}.\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|_2}\right)^2. \tag{26}$$

Proof: Suppose that $\boldsymbol{Q}$ and $\boldsymbol{Q}_\Delta$ are derived according to Theorem 5. Then we have

$$d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{E}\rangle) = \left\|(\boldsymbol{Q}+\boldsymbol{Q}_\Delta)^{\mathrm{H}}(\boldsymbol{Q}+\boldsymbol{Q}_\Delta) - \boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}\right\|^2 \tag{27}$$

$$\leqslant \left\|\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\|^2 + \left\|\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}\right\|^2 + \left\|\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\|^2 \tag{28}$$

$$= 2\left\|\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\|^2 + \left\|\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\|^2, \tag{29}$$

where (27) is by (12), (28) follows from triangle inequality and (29) is by observing that $\|\boldsymbol{A}\| = \left\|\boldsymbol{A}^{\mathrm{H}}\right\|$ for any matrix $\boldsymbol{A}$. Moreover, for the first term in (29) we have

$$\left\|\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\|^2 = \mathrm{tr}(\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}_\Delta) = \mathrm{tr}(\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}_\Delta) = \|\boldsymbol{Q}_\Delta\|^2, \tag{30}$$

since $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{H}} = \boldsymbol{I}_l$. Applying Lemma 19, in the appendix, to the second term in (29) results in

$$\left\|\boldsymbol{Q}_\Delta^{\mathrm{H}}\boldsymbol{Q}_\Delta\right\| \leqslant \|\boldsymbol{Q}_\Delta\|^2. \tag{31}$$

Then (29), (30), and (31) together with the result of Theorem 5 yield (25). ∎

The result of Theorem 6 can be applied to upper bound the subspace distance $d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle)$ caused by the additive noise $\boldsymbol{N}$ as long as the signal matrix A is full row rank. If A is not full row rank, which means that the receiver is somewhat *oversampling* from the network, the addition of $\boldsymbol{N}$, even with very small norm, may result in a large $d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle)$. This is because $\boldsymbol{A}+\boldsymbol{N}$ is full row rank with probability 1 if entries of N are Gaussian random variables. Consequently, by Lemma 18,

$$d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle) \geqslant l - \mathrm{rank}(\boldsymbol{A}), \tag{32}$$

regardless of how small $\|\boldsymbol{N}\|$ is. The aim here is to characterize the error correction capability of subspace codes in the presence of additive noise, i.e., to extend the result of Theorem 4 to take into account the effect of the additive noise as well as subspace errors and erasures. In order to do so for the general case, where the signal matrix $\boldsymbol{A}$ is not necessarily full row rank, one can model the effect of the noise partially as an interference of dimension $l - \mathrm{rank}(\boldsymbol{A})$ and partially as an addition of noise to a full row rank sub-matrix of the signal matrix, whose effect can be bounded using Theorem 6. This is elaborated through the rest of this section.

Let $r = \mathrm{rank}(\boldsymbol{A})$ and r_d denote the rank-deficiency of $\boldsymbol{A}$, i.e.,

$$r_d \stackrel{\text{def}}{=} l - r. \tag{33}$$

Also, let $\boldsymbol{A}_1$ be an $r \times n$ full row rank sub-matrix of $\boldsymbol{A}$ and $\boldsymbol{A}_2$ denote the sub-matrix of $\boldsymbol{A}$ consisting of its remaining rows. Let $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ be sub-matrices of $\boldsymbol{N}$ with row indices associated with row indices of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, respectively. Without loss of generality one can write

$$A = \left[\frac{\boldsymbol{A}_1}{\boldsymbol{A}_2}\right], \quad N = \left[\frac{\boldsymbol{N}_1}{\boldsymbol{N}_2}\right], \tag{34}$$

where both $\boldsymbol{A}_1$ and $\boldsymbol{N}_1$ are $r \times n$ matrices and both $\boldsymbol{A}_2$ and $\boldsymbol{N}_2$ are $r_d \times n$ matrices. Then we have the following theorem.

Theorem 7: Let $\boldsymbol{A} \in \mathbb{C}^{l\times n}$ with $\mathrm{rank}(\boldsymbol{A}) = r$. Let $\boldsymbol{A}_1 \in \mathbb{C}^{r\times n}$ denote a foll row rank sub-matrix of $\boldsymbol{A}$. Then for any $\boldsymbol{N} \in \mathbb{C}^{l\times n}$ that satisfies $\left\|\boldsymbol{A}_1^+\right\|_2 \|\boldsymbol{N}\|_2 < 1$ we have

$$d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle) \leqslant (\sqrt{r_d}+\sqrt{\Delta})^2, \tag{35}$$

where r_d is the rank-deficiency of $\boldsymbol{A}$, as defined in (33). Also,

$$\Delta \stackrel{\text{def}}{=} 2\epsilon+\epsilon^2, \tag{36}$$

where

$$\epsilon \stackrel{\text{def}}{=} \left(\frac{(1+\sqrt{2})\kappa(\boldsymbol{A}_1)}{1-\left\|\boldsymbol{A}_1^+\right\|_2\|\boldsymbol{N}\|_2}.\frac{\|\boldsymbol{N}\|}{\|\boldsymbol{A}_1\|_2}\right)^2. \tag{37}$$

Proof: Let $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{N}_1$, and $\boldsymbol{N}_2$ be as specified in (34). Then we have the following:

$$d(\langle\boldsymbol{A}\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle)^{1/2} \tag{38}$$

$$= d(\langle\boldsymbol{A}_1\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle)^{1/2} \tag{39}$$

$$\leqslant d(\langle\boldsymbol{A}_1\rangle, \langle\boldsymbol{A}_1+\boldsymbol{N}_1\rangle)^{1/2} + d(\langle\boldsymbol{A}_1+\boldsymbol{N}_1\rangle, \langle\boldsymbol{A}+\boldsymbol{N}\rangle)^{1/2}, \tag{40}$$

where (39) holds because $\langle\boldsymbol{A}\rangle = \langle\boldsymbol{A}_1\rangle$ and (40) is by triangle inequality for $d_c(.,.)$ and noting that $d(U,V) = 2d_c(U,V)^2$. We will bound the two terms in (40) separately.

To bound the first term in (40), first note that $\|\boldsymbol{N}_1\| \leqslant \|\boldsymbol{N}\|$. Also, for the spectral norm, we have $\|\boldsymbol{N}_1\|_2 \leqslant \|\boldsymbol{N}\|_2$ since adding a row to a rectangular matrix does not reduce its maximum singular value, see, e.g., [44]. This together with Theorem 6 yield the following upper bound on the first term in (40):

$$d(\langle \boldsymbol{A}_1 \rangle, \langle \boldsymbol{A}_1 + \boldsymbol{N}_1 \rangle) \leqslant \Delta = 2\epsilon + \epsilon^2 \tag{41}$$

where ϵ is specified in (37).

For the second term in (40) we have

$$\begin{aligned} &d(\langle \boldsymbol{A}_1 + \boldsymbol{N}_1 \rangle, \langle \boldsymbol{A} + \boldsymbol{N} \rangle) \\ &= d(\langle \boldsymbol{A}_1 + \boldsymbol{N}_1 \rangle, \langle \boldsymbol{A}_1 + \boldsymbol{N}_1 \rangle + \langle \boldsymbol{A}_2 + \boldsymbol{N}_2 \rangle) \qquad (42) \\ &\leqslant \operatorname{rank}(\langle \boldsymbol{A}_2 + \boldsymbol{N}_2 \rangle) \leqslant r_d, \qquad (43) \end{aligned}$$

where (42) holds by the definition in (1), and (43) follows from Lemma 18 and by noting that for any $U, V \in \mathcal{P}(W)$ one can always write $U + V = U \oplus V'$ for some $V' \in \mathcal{P}(W)$ with $\dim(V') \leqslant \dim(V)$. The proof is complete by combining (41) and (43) together with (40). ∎

Theorem 7 shows that the additive noise affects the output of the analog operator channel in two ways. It, in a sense, *rotates* the output subspace by a value upper bounded by Δ and also, implicitly, induces an extra interference term of dimension upper bounded by r_d (For simplicity, we consider the worst case scenario where this dimension is r_d). This motivates us to define the *noisy* analog operator channel as follows. First, we define a stochastic operator $\mathcal{R}_\Delta$, referred to as the *rotation operator*, which takes a subspace $U \in \mathcal{P}(W)$ as the input and returns a random subspace $V \in \mathcal{P}(W)$ with $\dim(V) = \dim(U)$ as the output in such a way that

$$d(U, V) \leqslant \Delta.$$

Definition 5: A noisy analog operator channel associated with the analog ambient space W is a channel with input U and output V, where $U, V \in \mathcal{P}(W)$, with the following input-output relation:

$$V = \mathcal{R}_\Delta\big(\mathcal{H}_k(U) \oplus E\big) \oplus F, \tag{44}$$

where $\mathcal{H}_k(U)$ and E induce subspace erasures and errors, respectively, as in the analog operator channel model, defined in Definition 1, $\mathcal{R}_\Delta$ is the rotation operation defined above, and F is the implicit interference caused by the additive noise.

The following theorem extends the result of Theorem 4 to take into account the effect of the additive noise as well as the subspace errors and erasures.

Theorem 8: Consider a subspace code $\mathcal{C}$ used for communication over a noisy analog operator channel, as defined in Definition 5, i.e., the input to the channel is $U \in \mathcal{C}$. Let t, ρ, and r_d denote the dimension of errors, erasures, and the implicit noise interference F, respectively. Then the minimum distance decoder successfully recovers the transmitted codeword $U \in \mathcal{C}$ from the received subspace V if

$$\rho + t + (\sqrt{\rho + t + \Delta} + \sqrt{\Delta} + 2\sqrt{r_d})^2 < d_{\min}(\mathcal{C}). \tag{45}$$

Proof: Let $V_1 = \mathcal{H}_k(U) \oplus E$ and $V_2 = \mathcal{R}_\Delta(V_1)$. Note that we have $V = V_2 \oplus F$. Then by applying Lemma 3 to the analog operator channel with input U and output V_1 we have

$$d(U, V_2) \leqslant \rho + t + d(V_1, V_2) = \rho + t + \Delta. \tag{46}$$

Also, by using the triangle inequality for the chordal distance $d_c(.,.)$ and noting that $d(.,.) = 2d_c(.,.)^2$ we have

$$d(U, V) \leqslant \big(\sqrt{d(U, V_2)} + \sqrt{d(V_2, V)}\big)^2. \tag{47}$$

By Lemma 18 we have $d(V_2, V) = r_d$. This together with (46) and (47) yields

$$d(U, V) \leqslant (\sqrt{\rho + t + \Delta} + \sqrt{r_d})^2. \tag{48}$$

Now consider a codeword $T \in \mathcal{C}$ other than U. Again, by applying Lemma 3 to the analog operator channel with the input U and the output V_1 and by rearranging the terms we have

$$d(T, V_1) \geqslant d(T, U) - \rho - t \geqslant d_{\min}(\mathcal{C}) - \rho - t, \tag{49}$$

where the last inequality is by the definition of minimum distance. Also, by using the triangle inequality for the chordal distance we have

$$d(T, V) \geqslant \big(\sqrt{d(T, V_1)} - \sqrt{d(V_1, V_2)} - \sqrt{d(V_2, V)}\big)^2. \tag{50}$$

Again by noting that $d(V_1, V_2) \leqslant \Delta$, $d(V_2, V) = r_d$, and by combining (50) with (49) we have

$$d(T, V) \geqslant \big(\sqrt{d_{\min}(\mathcal{C}) - \rho - t} - \sqrt{\Delta} - \sqrt{r_d}\big)^2. \tag{51}$$

It can be observed that the condition given in (45) implies that the right hand side of (51) is strictly greater than that of (48). In other words, (45) guarantees that

$$d(T, V) > d(U, V),$$

for any codeword $T \in \mathcal{C}$ other than U. Hence, the minimum distance decoder returns U which completes the proof. ∎

Remark 3. Note that Theorem 8 reduces to Theorem 4 by setting $r_d = \Delta = 0$. In other words, Theorem 8 *properly* extends the result of Theorem 4, on relating the minimum distance of analog subspace codes to their error-and-erasure correction capability, to the noisy analog operator channel scenario. In practice, the implicit noise interference term F and, consequently, the term r_d in (45) can be potentially removed by simply discarding a certain number of received blocks at the receiver. However, this requires knowing the rank of the received signal by the receiver which may not be readily available due to the assumptions on non-coherent communications. This can be further explored when considering a practical wireless networking scenario to see whether such information, i.e., the rank of the received signal, can be obtained or well-approximated, e.g., using principal component analysis (PCA) methods, by the receiver. Also, as shown in Theorem 7, the other term, besides r_d, resulting from the additive noise that affects the output subspace is Δ. Note that for a fixed signal matrix $\boldsymbol{A}$, as $\|\boldsymbol{N}\| \to 0$, we have $\epsilon \to 0$ as well as $\Delta \to 0$, where ϵ and Δ are specified in (37) and (36), respectively. This together with a procedure to remove the r_d term, as discussed above, show that the analog operator channel can be made *robust* with respect to the additive noise, i.e., the subspace distance perturbation in the received signal matrix, caused by the additive noise, goes to

zero as $\|\boldsymbol{N}\| \to 0$.

V. Constructions of Analog Subspace Codes

In this section, we explore three approaches for constructing analog subspace codes. In particular, the novel approach based on character sums results in explicit constructions with better rate-minimum distance trade-off compared to the previously known constructions for a wide range of parameters.

More specifically, we are concerned with the following equivalent questions. For a given dimension of the ambient space n and the size of the subspace code $|\mathcal{C}|$, or equivalently the rate of $\mathcal{C}$, construct $\mathcal{C}$ with the maximum possible $d_{\min}$. Alternatively, given n and a desired $d_{\min}$, construct the subspace code $\mathcal{C}$ with the maximum size/rate. Although the exact answer to these equivalent questions are not known in general, one can derive sphere-packing-type upper bounds and Gilbert-Varshamov-type lower bounds for codes in the subspace domain.

By precisely characterizing the volume of balls in the Grassmann space Barg and Nogin derived lower and upper bounds for R as $n \to \infty$ while m and δ are fixed [8]. More specifically, they analyzed the asymptotic behavior of the volume of a sphere with a fixed radius on $G_{m,n}(\mathbb{L})$ that is then used in a packing-type and a covering-type argument. Their result is recapped in the following theorem. Note that we use the new notion of distance, as defined in Definition 2, to state the results.

Theorem 9: [8, Theorem 2] There exists a sequence of codes in $G_{m,n}(\mathbb{L})$ with fixed m and δ, while $n \to \infty$, and asymptotic rate

$$R > -\frac{1}{2}\beta m \ln(\delta). \tag{52}$$

Moreover, for any such sequence of codes

$$R < -\beta m \ln(\sqrt{1-\sqrt{1-\frac{\delta}{2}}}), \tag{53}$$

where $\beta = 1, 2$ for $\mathbb{L} = \mathbb{R}, \mathbb{C}$, respectively, as discussed in Section II-A.

Note that the lower bound in Theorem 9 is based on existence-type arguments. However, a somewhat *stronger* result can be established to conclude that such codes, perhaps by sacrificing in the rate, can be found with probability arbitrarily close to 1 in a random ensemble. This is the focus of the next subsection. Also, it is discussed how such a result can be used in constructing explicit codes.

A. Constructions based on random ensembles

For large values of n, the volume of a ball with a certain radius r in $G_{m,n}(\mathbb{C})$ is characterized in [8]. This is done by utilizing the relation between the principal angles of two uniformly distributed subspaces in $G_{m,n}(\mathbb{C})$ and the singular values of Wishart matrices, the matrices of the form $\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}}$, where the elements of $\boldsymbol{N} \in \mathbb{C}^{m\times n}$ are i.i.d standard normal random variables. Note that the subspace spanned by the rows of such a random matrix $\boldsymbol{N}$ is uniformly distributed on the corresponding sphere in the Grassmann space $G_{m,n}(\mathbb{C})$.

Let $\mathcal{R}$ denote a random ensemble of subspace codes with code size $M = \exp(nR)$, wherein each codeword is the row space of a randomly generated $m \times n$ matrix with i.i.d entries from the $\mathcal{N}(0,1)$ distribution. In the next theorem, it is shown that a random subspace code *almost* achieves half of the Gilbert-Varshamov lower bound, stated in (52), as $n \to \infty$, with a probability that approaches 1 exponentially fast in n.

Theorem 10: Consider a random ensemble $\mathcal{R}$ of subspace codes with the rate

$$R = -\frac{1}{4}\beta m \ln(\delta) - \epsilon,$$

for some $\epsilon > 0$. Then the normalized minimum distance of a code $\mathcal{C}$ randomly picked from $\mathcal{R}$ is at least δ with probability at least $1 - \exp(-2n\epsilon + o(n))$.

Proof: Let $\mathcal{C} = \{\Phi_i : 1 \leqslant i \leqslant M\}$, where $M = \exp(nR)$ and $\mathcal{C}$ is randomly picked from $\mathcal{R}$. Then the probability p that two arbitrary codewords Φ_i and Φ_j have distance at most r is equal to the volume of a ball with radius r, which can be characterized as follows [11, Theorem 1]:

$$p = (\frac{r}{2m})^{\frac{\beta nm}{2}+o(n)}. \tag{54}$$

Let $X_{i,j}$ be an indicator random variable which is 1 if $d(\Phi_i, \Phi_j) \leqslant r$; otherwise, $X_{i,j} = 0$. Let

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{M}\sum_{j=i+1}^{M} X_{i,j}.$$

Using Markov inequality together with (54) and the linearity of expectation we have:

$$\Pr[X \geqslant 1] < E[X] = \binom{M}{2}\Pr[X_{i,j} = 1] \tag{55}$$

$$= \binom{M}{2} p < \exp\Big(2nR + (\frac{\beta mn}{2} + o(n))\ln(\frac{r}{2m})\Big). \tag{56}$$

Note that if the random variable X, associated to $\mathcal{C}$, is zero, then it implies that $d_{\min}(\mathcal{C}) \geqslant r$, i.e.,

$$\Pr[d_{\min}(\mathcal{C}) \geqslant r] = \Pr[X = 0] = 1 - \Pr[X \geqslant 1]. \tag{57}$$

This together with (56) complete the proof. ∎

Remark 4. Note that random ensembles, in general, do not lead to explicit constructions of subspace codes that can be constructed with complexity that is polynomial in n. However, a potential approach to utilize such ensembles is to pick another parameter n' that is much smaller than n, e.g., $n' = O(\log n)$, and consider a random ensemble of subspace codes with the dimension of ambient space equal to n'. Then a brute-force search within the ensemble is feasible as its complexity is exponential in n' and, hence, it is polynomial in n. Also, a minimum distance decoding is feasible for such a code. Then a *proper* concatenation scheme can be potentially used, by concatenating the random inner code with some explicit construction of an outer code in order to construct explicit codes with non-vanishing rates given a fixed δ as $n \to \infty$. The details are left for future work.

B. Packing lines using binary codes

A special and yet practically interesting case of the analog operator channel is when $m = 1$. For instance, from the non-coherent wireless communications perspective, elaborated in

Section II-B, this can be a reasonable scenario in the uplink transmission of a wireless node with one antenna transmitting the node's data in one time frame. In such cases, it is natural to assume that there is no rank deficiency; otherwise, reliable communication is not possible. A possible approach to construct codes in $G_{1,n}(\mathbb{L})$ is to use well-known constructions of binary linear codes and map the binary coded data into real/complex symbols resembling a joint coded modulation design.

The idea of constructing codes in $G_{1,n}(\mathbb{R})$ using binary linear codes was first suggested in [17]. Consider a binary linear code of length n that forms a closed set under the completion, i.e., it contains the all-one codeword, with normalized minimum Hamming distance γ. Then a possible mapping to real-valued symbols is to map zeros to 1's and ones to -1's. This results in a code in $G_{1,n}(\mathbb{R})$ with the normalized minimum distance $\delta = 32\gamma^2(1-\gamma)^2$. The same result also holds in the complex domain, i.e., packing lines in $G_{1,n}(\mathbb{C})$, where one can map ones to $(1+i)$'s and zeros to $-(1+i)$'s. Hence, given a family of binary linear codes with fixed Hamming distance and asymptotically non-vanishing rate in terms of n, one can construct a code in $G_{1,n}(\mathbb{L})$ with a fixed minimum distance and non-vanishing rate as well. To this end, in the rest of this subsection, we briefly overview various families of binary linear codes, from this perspective, in the coding theory literature.

There are several well-known constructions of binary linear codes, mainly based on code concatenations, with asymptotically good minimum distance. The idea of code concatenation was first introduced by Elias in the form of product codes [45] and was later developed by Forney [46]. Also, the well-known Zyablov lower bound for the normalized minimum distance of a concatenated code with rate R is characterized as follows [47]:

$$\delta_Z(R) = \max_{R \leqslant x \leqslant 1} \delta_{GV}(x)(1-\frac{R}{x}), \tag{58}$$

where $\delta_{GV}(x)$ is the Gilbert-Varshamov normalized distance at rate x for the binary codes. Furthermore, by using multilevel concatenation, codes with minimum distance even larger than Zyablov bound (58) can be constructed. More specifically, a generalization to (58) by letting the number of concatenation levels going to infinity was given later, known as Blokh-Zyablov bound [48], stated as follows:

$$R_{BZ} = 1 - h(\delta) - \delta \int_0^{1-h(\delta)} \frac{dx}{\delta_{GV}(x)}. \tag{59}$$

Another line of work on combining codes is due to Tanner [49] with the general theme of using one or more shorter codes, referred to as subcodes, in combination with a certain bipartite graph. In particular, a certain family of such graph-based codes is referred to as expander codes, which are proved to have asymptotically good minimum distance. For instance, Barg and Zemor [50] proposed a family of expander codes meeting the Zyablov bound, specified in (58), with the construction complexity at most $O(n^2)$ and a decoder, with complexity that is linear in n, that corrects up to half of the minimum distance. They further improved this result by introducing another family of codes that exceeds Zyablov bound with construction complexity not more than $O(n \log n)$.

There are also other families of concatenated codes, based on algebraic-geometry (AG) codes as their outer codes, which can somewhat provide better minimum distance comparing to graph-based codes. More specifically, a certain concatenated code family with a short binary inner code and an AG outer code can surpass the Blokh-Zyablov bound [51], characterized in (59). These codes have construction complexity of $O(n^3 \log^3 n)$ [52] and are currently known to have the largest asymptotic rate, given a certain fixed normalized minimum, while having polynomial construction complexity. Also, the decoding complexity of such codes is $O(n^3)$.

A main drawback of aforementioned construction methods based on concatenation is that they often require n to be very large in order to meet the promised performance. In other words, they exhibit excellent asymptotic performance, however, they often fall short for finite values of n that are of practical interest. Hence, it is desirable to focus on explicit constructions of subspace codes that can be constructed for a wide range of n, regardless of how small or large n is, while providing reasonable performance. In a sense, we aim at constructing subspace codes that resemble well-known families of block codes such as Reed-Solomon codes in the subspace domain, and that can be constructed for a broad range of parameters. This is the focus of the next subsection.

C. A new family of analog subspace codes: Character-polynomial codes

In this section, we propose a new family of subspace codes in $G_{1,n}(\mathbb{C})$ by leveraging polynomial evaluations over finite fields and mapping the finite field symbols to the roots of unity. Then results on character sums from analytic number theory [53], discussed next, are used to bound the minimum distance of the constructed codes. The resulting codes are referred to as character-polynomial (CP) codes.

Consider a cyclic group G of order $|G|$. A character χ associated to G is a homomorphism from G to the unit circle in the complex plane with the regular multiplication of complex numbers, i.e.,

$$\chi(g_1 g_2) = \chi(g_1)\chi(g_2), \tag{60}$$

for all $g_1, g_2 \in G$. It can be observed that

$$\chi_g^* = \chi(g^{-1}), \tag{61}$$

where g^{-1} is the inverse of g in G and x^* is the complex conjugate of x for $x \in \mathbb{C}$. Given a certain and finite number of characters $\chi_1, \ldots, \chi_l$ one can define the product character $\chi_1\chi_2\cdots\chi_l$ by setting

$$(\chi_1\chi_2\cdots\chi_l)(g) = \chi_1(g)\chi_2(g)\cdots\chi_l(g),$$

for all $g \in G$. The set of all characters associated to G together with this product form an Abelian group of order $|G|$, where the elements χ_j, for $j = 1, 2, \ldots, |G|$, are described as follows [53]:

$$\chi_j(g'^k) = \mathrm{e}(\frac{jk}{|G|}), \tag{62}$$

for $k = 0, 1, \ldots, |G|-1$, where g' is a generator of G and $\mathrm{e}(x) \stackrel{\text{def}}{=} \exp(2\pi i x)$. Note that $\chi_0(g) = 1$ for $g \in G$ and hence, it is referred to as the *trivial* character.

A finite field $\mathbb{F}_q$ is naturally equipped with two finite Abelian groups, i.e., the additive and the multiplicative group. Then, the *additive* and *multiplicative* characters of $\mathbb{F}_q$ are defined as the characters associated with the additive and the multiplicative group in $\mathbb{F}_q$, respectively. Using (62), the additive characters, denoted by χ_j, for $j = 0, 1, \ldots, q-1$, are described as follows [53]:

$$\chi_j(\alpha) = \mathrm{e}(\frac{\mathrm{tr_a}(j\alpha)}{p}) \tag{63}$$

for $j \in \mathbb{F}_q$, where p is the characteristic of $\mathbb{F}_q$, and

$$\mathrm{tr_a}(\gamma) \stackrel{\mathrm{def}}{=} \gamma + \gamma^p + \cdots + \gamma^{p^{m-1}}$$

is the *absolute* trace function from $\mathbb{F}_q$ to $\mathbb{F}_p$, where $q = p^m$. Note that (63) implies that $\chi_j(\alpha) = \chi_1(j\alpha)$ and the trivial additive character is $\chi_0(\alpha) = 1$ for $\alpha \in \mathbb{F}_q$.

The following result, due to Weil [54], on the summations over characters, which are commonly known as exponential sums or character sums in the literature, will be utilized in bounding the minimum distance of CP codes, to be discussed next.

Theorem 11: [53, Theorem 5.35] Consider a polynomial $f \in \mathbb{F}_q[x]$ of degree $d \geqslant 1$ with $\gcd(d, q) = 1$. Let χ be a nontrivial additive character of $\mathbb{F}_q$. Then

$$|\sum_{\alpha \in \mathbb{F}_q} \chi(f(\alpha))| \leqslant (d-1)\sqrt{q}. \tag{64}$$

Next, for some $k < q$, let

$$\mathcal{F} \stackrel{\mathrm{def}}{=} \{f \in \mathbb{F}_q[x] : f(x) = \sum_{i \in [k], i \bmod p \neq 0} f_i x^i\}. \tag{65}$$

Note that $|\mathcal{F}| = q^{\lceil k(p-1)/p \rceil}$. We fix $n = q-1$ in our construction.

Definition 6: The code $\mathcal{C}(\mathcal{F}) \subseteq G_{1,n}(\mathbb{C})$, referred to as a character-polynomial (CP) code, is defined as follows:

$$\mathcal{C}(\mathcal{F}) \stackrel{\mathrm{def}}{=} \{\langle (c_1, c_2, \ldots, c_n)\rangle : c_i = \chi\big(f(\alpha_i)\big), \quad \forall f \in \mathcal{F}, \alpha_i \in \mathbb{F}_q \backslash \{0\}\}, \tag{66}$$

where χ is a fixed nontrivial additive character of $\mathbb{F}_q$, and α_i's are distinct non-zero elements of $\mathbb{F}_q$.

The following theorem provides a lower bound on the normalized minimum distance of $\mathcal{C}(\mathcal{F})$ in terms of q and d.

Theorem 12: The code $\mathcal{C}(\mathcal{F})$ has size $q^{\lceil k(p-1)/p \rceil}$ and

$$\delta \geqslant 1 - \frac{\big((k-1)\sqrt{q}+1\big)^2}{n^2}, \tag{67}$$

where $\delta = d_{\min}/2m$ (here $m = 1$) is the normalized minimum distance of the code.

Proof: Consider distinct codewords $\langle \boldsymbol{C}_1\rangle, \langle \boldsymbol{C}_2\rangle \in \mathcal{C}(\mathcal{F})$ with corresponding distinct polynomials $f_1, f_2 \in \mathcal{F}_1$. Let $f = f_2 - f_1$. Then we have

$$\begin{aligned} n\left\|\boldsymbol{C}_1\boldsymbol{C}_2^{\mathrm{H}}\right\| &= |\sum_{\alpha \in \mathbb{F}_q \backslash \{0\}} \chi^*\big(f_1(\alpha)\big)\chi\big(f_2(\alpha)\big)| \\ &\stackrel{(a)}{=} |\sum_{\alpha \in \mathbb{F}_q} \chi\big(f(\alpha)\big) - 1| \\ &\stackrel{(b)}{\leqslant} |\sum_{\alpha \in \mathbb{F}_q} \chi\big(f(\alpha)\big)| + 1 \\ &\stackrel{(c)}{\leqslant} (k-1)\sqrt{q} + 1, \end{aligned}$$

where (a) follows by (61) and (60) and noting that $\chi\big(f(0)\big) = 1$, (b) is by the triangle inequality, and (c) is by Theorem 11 applied to $f = f_1 - f_2$. Note that $f \in \mathcal{F}$. This implies that $\deg(f) \geqslant 1$ and $\gcd(\deg(f), q) = 1$ since polynomials in $\mathcal{F}$, as defined in (65), do not have a monomial of degree divisible by p. Hence, $f = f_1 - f_2$ satisfies the conditions in Theorem 11.

Using *Lemma* 2, we have

$$\delta = \frac{d(\langle \boldsymbol{C}_1\rangle, \langle \boldsymbol{C}_2\rangle)}{2} = 1 - \left\|\boldsymbol{C}_1\boldsymbol{C}_2^{\mathrm{H}}\right\|^2 > 1 - \frac{\big((k-1)\sqrt{q}+1\big)^2}{n^2},$$

which completes the proof. ∎

Note that the right hand side of (67) can be approximated in terms of the rate of the code. This results in a bound on the trade-off between the normalized minimum distance and the rate of the code. We plot this bound and compare it with other bounds/constructions next. In particular, in order to have a simplified numerical analysis, we limit our attention to the case where q is a prime, i.e., $q = p$, and $k < p$. In this case, we have $|\mathcal{C}(\mathcal{F})| = q^k$ and $R = \frac{k \ln q}{n}$. Also, the bound in (67) is simplified as follows:

$$\delta \geqslant 1 - \frac{\big((k-1)\sqrt{q}+1\big)^2}{n^2} > 1 - \frac{qR^2}{(\ln q)^2}. \tag{68}$$

It is worth mentioning that all nontrivial choices for the

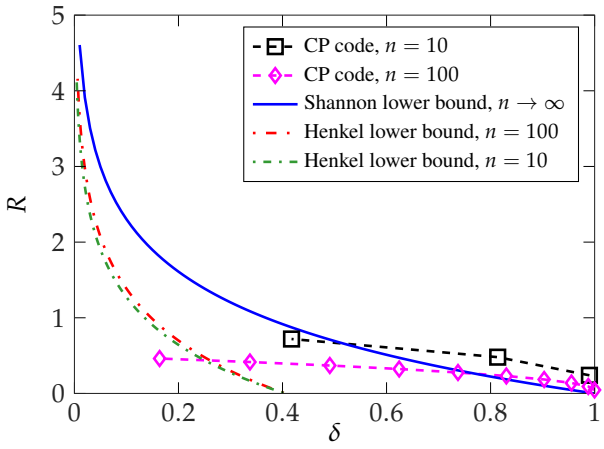

Fig. 2: Comparison of CP codes with lower-bounds in terms of the trade-off between the rate R and the normalized minimum distance δ for $m = 1$.

character χ result in the same codebook. To observe this, let χ_a and χ_b denote two distinct nontrivial characters for some $a, b \in \mathbb{F}_q$, and $\mathcal{C}_a(\mathcal{F})$ and $\mathcal{C}_b(\mathcal{F})$ denote the corresponding codebooks described in Definition 6, respectively. Since both a and b are non-zero, $c = ba^{-1}$ is a well-defined non-zero element of $\mathbb{F}_q$. Then, for any $f(x) \in \mathcal{F}$, one can write

$$\begin{aligned}\chi_b(f(x)) =& \mathrm{e}\left(\frac{\mathrm{tr_a}(bf(x))}{p}\right) = \mathrm{e}\left(\frac{\mathrm{tr_a}(acf(x))}{p}\right)\\ =& \mathrm{e}\left(\frac{\mathrm{tr_a}(af'(x))}{p}\right) = \chi_a(f'(x)).\end{aligned} \tag{69}$$

Note that $f'(x) \overset{\text{def}}{=} cf(x)$ is also in $\mathcal{F}$. This together with (69) imply that $\mathcal{C}_a(\mathcal{F}) = \mathcal{C}_b(\mathcal{F})$.

In Figure 2, we compare the trade-off between the rate R and the normalized minimum distance δ that the CP codes offer at different values of n with Shannon's lower bound [7], for $n \to \infty$, and lower bounds derived by Henkel [16, Theorem 4.2] for finite values of n. Note that these lower bounds are of the same type as Gilbert-Varshamov bound and do not yield explicit constructions. Nevertheless, it can be observed that CP codes outperform these lower bounds at low rates, thereby improving these bounds while providing explicit constructions. Note also that the trade-off between R and δ shown in Figure 2 for CP codes is derived from the bound established in Theorem 12. In other words, the actual value of δ for the given values of R can be larger than what is shown in Figure 2.

Remark 5. Given a subspace code in $\mathcal{C} \subseteq G_{m,n}(\mathbb{C})$ one can construct a code in $G_{2m,2n}(\mathbb{R})$ by mapping $\boldsymbol{C}_i \in \mathcal{C}$ to

$$\begin{bmatrix} \Re(\boldsymbol{C}_i) & \Im(\boldsymbol{C}_i) \\ -\Im(\boldsymbol{C}_i) & \Re(\boldsymbol{C}_i) \end{bmatrix}, \tag{70}$$

where $\Re$ and $\Im$ represent the real part and the imaginary part of their input, respectively. It can be observed that this mapping preserves the normalized distance between the codewords. Hence, the normalized minimum distance of $\mathcal{C}$ is also preserved. This mapping enables us to construct codes in $G_{2,n}(\mathbb{R})$ using the proposed CP codes, while keeping the normalized minimum distance and the size of the code the same, in order to have fair comparisons with existing code constructions in the real Grassmann space.

In Figure 3, we compare CP codes with two existing constructions of Grassmann codes, that are constructed explicitly for a wide range of n, in the literature. In [21], Calderbank *et al.* introduce a group-theoretic framework for packing in $G_{2^i,2^k}(\mathbb{R})$ for any pair of integers (i,k) with $i \leqslant k$. In another prior work, Ashikhmin *et al.* [35] provide a code construction method in $G_{2^i,2^k}(\mathbb{C})$ based on binary Reed-Muller (RM) codes. It is worth noting that the subspaces in this construction correspond to certain projection operators determined by Pauli matrices appearing in quantum computing, where the main objective is to enable error correction in quantum computing. Therefore, one should not expect such constructions to maximize the rate. However, they are competitive in small dimensions as illustrated in Figure 3. By utilizing the mapping specified in (70) for both the CP codes and the codes constructed in [35] with $i = 0$, we compare the log-size of the codes obtained in $G_{2,n}(\mathbb{R})$ with that of codes in $G_{2,n}(\mathbb{R})$ from [21], while fixing

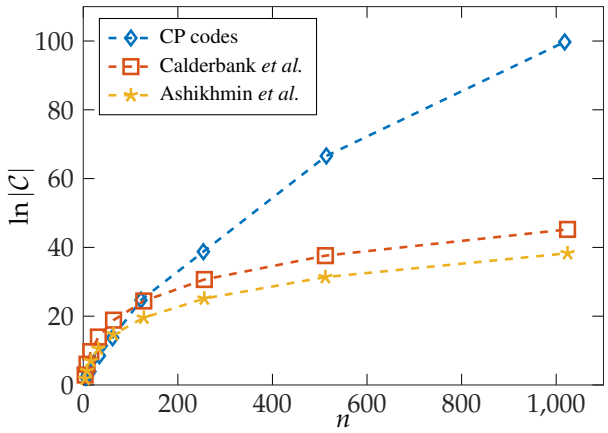

Fig. 3: Comparison of the codes in $G_{2,n}(\mathbb{R})$ obtained from our proposed CP codes in $G_{1,\frac{n}{2}}(\mathbb{C})$ with the codes constructed by Calderbank *et al.* [21] and Ashikhmin *et al.* [35]. The codes from [21] [35] have $\delta = \frac{1}{2}$, and for CP codes we have $\delta \geqslant \frac{1}{2}$, for all considered values of n.

$\delta = \frac{1}{2}$ for the codes from [21] [35] and having $\delta \geqslant \frac{1}{2}$ for the CP codes, for all the considered values of n. The results are shown in Figure 3. Note that n is equal to 2^k for the constructions in [21] and [35], while for CP codes we pick $n = 2p$, where p is the largest p with $p < 2^{k-1}$, for $k \in \{3, \ldots, 10\}$. It can be observed that CP codes offer significantly larger code size and, consequently, rate comparing to the other explicit constructions, as n grows large.

Remark 6. Note that the lower bound of Theorem 12 on $d_{\min}$ of CP codes is at most two. Even with $d_{\min}$ slightly greater than two Theorem 4 implies that the correction of only one error can be guaranteed using a minimum distance decoder. A possible solution to this issue is to consider list decoders in order to guarantee error recovery beyond $d_{\min}/2$. In the finite field domain, several prior works have studied list decoding for algebraic subspace codes, see, e.g., [55]–[59]. In particular, it is observed in [57] that an unbounded number of errors can be corrected by increasing the list size, while the dimension of subspace codewords is one. However, obtaining similar results in the analog domain necessitates further investigation and is left for future work.

Remark 7. The Weil bound, recapped in Theorem 11, has been utilized in various coding theoretic contexts , e.g., to provide bounds on the minimum distance of the duals of BCH codes [60] and to estimate the covering radius of BCH codes with large block lengths [61], [62]. Also, it has inspired the design of certain families of sequences over finite fields of prime size with low auto-correlation and cross-correlation in [63] in a similar fashion. Notably, Kumar *et al.* derive a counterpart of the Weil bound over Galios rings [64]. They further design families of phase-shift-keying sequences with low correlation where this bound is then utilized to provide a guarantee on the correlation level of the construction. The main difference between the design criterion considered in the constructions in [63] and [64] and that of our approach is that these prior works require the correlation between the circular shifts of any two codewords to be bounded, in addition to the auto-correlation, while we only

require the correlation between two distinct codewords to be small. Furthermore, as it is shown in the next section, a certain property of the construction provided in this work is leveraged to construct subspace codes in higher dimensions ($m > 1$).

D. Higher Dimensional Character-Polynomial Codes

The *character-polynomial* (CP) codes, demonstrated in Section V-C, provide a family of one-dimensional subspaces in a complex Grassmann space, i.e., a packing of lines in $G_{1,n}(\mathbb{C})$. Next, we provide a slightly different version of one-dimensional CP codes that enables us to provide a new construction in $G_{m,n}(\mathbb{C})$ for $m > 1$. We fix $n = q$ in this section.

Definition 7: The code $\mathcal{C}'(\mathcal{F}) \subseteq G_{1,n}(\mathbb{C})$ is defined as follows:

$$\mathcal{C}'(\mathcal{F}) \stackrel{\text{def}}{=} \{\langle (c_1, c_2, \ldots, c_n)\rangle : c_i = \chi\big(f(\alpha_i)\big), \forall f \in \mathcal{F}, \alpha_i \in \mathbb{F}_q\}, \tag{71}$$

where χ is a fixed nontrivial additive character of $\mathbb{F}_q$, $\mathcal{F}$ is defined in (65), and α_i's are distinct elements of $\mathbb{F}_q$.

The definition of CP codes provided in (66) excludes the zero element of $\mathbb{F}_q$ from the set of evaluation points. Including the zero element in the alternative version, specified in (71), leads to a certain property that is discussed in the following lemma. Before that, we define the following. For any two sets of orthonormal bases B_1 and B_2 for W, the *cross-correlation* between B_1 and B_2 is defined as

$$\Delta_{B_1,B_2} \stackrel{\text{def}}{=} \max_{v_1 \in B_1, v_2 \in B_2} |v_1 \cdot v_2|, \tag{72}$$

where the operation $\cdot$ denotes the inner product.

Lemma 13: The set of normal vectors representing revised one-dimensional CP codewords, defined in Definition 7, can be split into $q^{\lceil k(p-1)/p\rceil - 1}$, denoted by b, collections B_i's, for $i \in [b]$, where each B_i is an orthonormal basis for W. Furthermore, the cross-correlation between B_i's is upper bounded as follows:

$$\max_{i,j \in [b], i \neq j} \Delta_{B_i,B_j} \leqslant \frac{(k-1)}{\sqrt{n}}. \tag{73}$$

Proof: The set of polynomials $\mathcal{F}$, defined in (65), can be split into disjoint subsets such that the polynomials belonging to the same subset differ only in the coefficient of the degree-one monomial, i.e., the coefficient of x. Note that the constant coefficient of all the polynomials in $\mathcal{F}$ is equal to zero according to (65). Then, two distinct polynomials f and f' in $\mathcal{F}$ belong to the same subset if and only if $\deg(f - f') = 1$. Consequently, it can be observed that $\mathcal{F}$ is partitioned into $q^{\lceil k(p-1)/p\rceil - 1}$ of such subsets each of size q. Let $\boldsymbol{c} = \frac{1}{\sqrt{n}}(c_1, \cdots, c_n)$ and $\boldsymbol{c}' = \frac{1}{\sqrt{n}}(c'_1, \cdots, c'_n)$, where $c_i = \chi\big(f(\alpha_i)\big)$ and $c'_i = \chi\big(f'(\alpha_i)\big)$ for $i \in [n]$, and f and f' belong to the subset of $\mathcal{F}$ as discussed above. Then,

$$|\boldsymbol{c} \cdot \boldsymbol{c}'| = \frac{1}{n}\Big|\sum_{\alpha \in \mathbb{F}_q} \chi^*\big(f(\alpha)\big)\chi\big(f'(\alpha)\big)\Big| \stackrel{(a)}{=} \frac{1}{n}\Big|\sum_{\alpha \in \mathbb{F}_q} \chi\big((f' - f)(\alpha)\big)\Big| \stackrel{(b)}{\leqslant} 0 \tag{74}$$

where (a) follows by (63) and (b) is by the Weil bound, specified in (64), together with noting that $\deg(f - f') = 1$. Therefore, (74) implies that we must have $\boldsymbol{c} \cdot \boldsymbol{c}' = 0$, i.e., $\boldsymbol{c}$ and $\boldsymbol{c}'$ must be orthogonal. Hence, each of the subsets of $\mathcal{F}$, as discussed above, consists of q mutually orthogonal lines in W. Consequently, the set of unit-norm vectors representing these lines is an orthonormal basis for W. The upper bound in (73) can be derived again using the Weil bound and by noting that $\deg(f - f') \leqslant k$ for any f and f' in $\mathcal{F}$, i.e., for any two normalized distinct $\boldsymbol{c}$ and $\boldsymbol{c}'$ in the CP code, the upper bound in (73) holds. ∎

Inspired by Lemma 13, we provide a construction for packing m-planes in $G_{m,n}(\mathbb{C})$ for $m > 1$. Let $\boldsymbol{v}_1^{(i)}, \cdots, \boldsymbol{v}_q^{(i)}$ denote the orthonormal basis vectors in B_i, for $i \in [b]$, where b is defined in Lemma 13. Also, let

$$\Phi_{ij} = \begin{bmatrix} \boldsymbol{v}_{(j-1)m+1}^{(i)} \\ \vdots \\ \boldsymbol{v}_{jm}^{(i)} \end{bmatrix}, \tag{75}$$

for all $i \in [b]$ and $j \in [\lfloor \frac{q}{m} \rfloor]$. Then,

$$\mathcal{C} \stackrel{\text{def}}{=} \{\langle \Phi_{ij} \rangle : \forall i \in [b], \forall j \in [\Big\lfloor \frac{q}{m} \Big\rfloor]\} \tag{76}$$

is a subspace code in $G_{m,n}(\mathbb{C})$. Note that we have

$$|\mathcal{C}| = q^{\lceil k(p-1)/p\rceil - 1} \Big\lfloor \frac{q}{m} \Big\rfloor.$$

The normalized minimum distance of $\mathcal{C}$ is characterized in the next theorem.

Theorem 14: The normalized minimum distance δ of the code $\mathcal{C}$, defined in (76), is lower bounded as

$$\delta \geqslant 1 - \frac{m(k-1)^2}{n}. \tag{77}$$

Proof: Consider two distinct codewords $C_1 = \langle \Phi_{i,j} \rangle$ and $C_2 = \langle \Phi_{i',j'} \rangle \in \mathcal{C}$. Note that the rows in $\Phi_{i,j}$ and $\Phi_{i',j'}$ are orthonormal. Note also that for $i = i'$, $\Phi_{i,j}\Phi_{i',j'}^{\mathrm{H}} = \mathbf{0}$, since all the rows of both matrices belong to the same orthonormal basis which implies that $\delta = 1$ in this case. Otherwise, i.e., when $i \neq i'$, we have

$$\Big\|\Phi_{i,j}\Phi_{i',j'}^{\mathrm{H}}\Big\|^2 \leqslant \frac{m^2(k-1)^2}{n}, \tag{78}$$

since $\Phi_{i,j}\Phi_{i',j'}^{\mathrm{H}}$ is an $m \times m$ matrix whose elements' absolute value is upperbounded by (73). Then, one can write

$$\delta = \frac{d(C_1, C_2)}{2m} \tag{79}$$

$$= \frac{2(m - \Big\|\Phi_{i,j}\Phi_{i',j'}^{\mathrm{H}}\Big\|^2)}{2m} \tag{80}$$

$$\geqslant 1 - \frac{m(k-1)^2}{n}, \tag{81}$$

where (79) is by the definition of the normalized distance provided in Definition 4, (80) follows by utilizing the alternative characterization provided in Lemma 2 for the distance function defined in (12), and (81) results from plugging in (78) into (80). ∎

In Table I, we compare the parameters of our proposed codes with those of the closest explicit construction of Grassmann

	Our construction			Calderbank *et al.* [21]		
m	n	$\ln(\lvert\mathcal{C}\rvert)$	$d_{\min}$	n	$\ln(\lvert\mathcal{C}\rvert)$	$d_{\min}$
4	254	28.36	2.43	256	32.62	2
4	502	43.51	2.44	512	40.25	2
4	1018	74.09	2.10	1024	48.57	2
4	2042	110.2	2.37	2048	57.59	2
8	1018	48.47	4.92	1024	49.87	4
8	2024	81.76	4.3	2048	59.57	4
8	4078	120.5	4.47	4096	69.97	4
8	8186	189.9	4.22	8192	81.07	4
16	2042	53.34	9.86	2048	59.52	8
16	4078	89.36	8.40	4096	70.61	8
32	4078	58.19	19.67	4096	69.19	16
32	8186	97.03	16.86	8192	81.67	16

TABLE I: Parameters of the new construction in Grassmannian space provided in this paper to those of the proposed scheme by Calderbank *et al.* [21, Theorem 1].

codes proposed in [21]. By utilizing the mapping specified in (70) for our codes in $G_{\frac{m}{2},\frac{n}{2}}(\mathbb{C})$, we compare the blocklength, logarithm of the code size, and the minimum distance of the codes obtained in $G_{m,n}(\mathbb{R})$ with those of the codes in $G_{m,n}(\mathbb{R})$ from [21]. In all the instances of n and $m = 4, 8, 16, 32$ in Table I, the normalized minimum distance is equal to $\frac{1}{2}$ for codes from [21] while it is at least $\frac{1}{2}$ for our codes. Note that n is equal to 2^k for the construction in [21], while for our codes we pick $n = 2p$, where p is the largest prime number with $p < 2^{k-1}$, for various choices of k. It can be observed that our proposed construction offer significantly larger code size and, consequently, rate comparing to the explicit construction of [21], as n grows large. Note also that even for small n in a few rows of Table I where the log-size of the proposed CP codes is not larger than that of the codes constructed in [21] yet, the minimum distance offered by CP codes is still larger while having a smaller blocklength. The other advantage of our construction is that it does not have any constraint on the codeword dimension m but m has to be a power of two in [21].

VI. Conclusion and Future Work

In this paper, motivated by the emergence of massive wireless networks, we provided a new coding framework for noncoherent communications across such networks in order to mitigate the network deficiency and interference, e.g., from neighbouring cells in a cellular network. To this end, the concept of analog operator channel was introduced that captures the effect of network deficiency and interference as subspace erasure and errors, respectively. Also, a new distance is defined and relations between the error-and-erasure correction capability of a subspace code in the analog domain and its minimum distance is established. This leads to a code design criteria to correct errors/erasures over the analog operator channel. Furthermore, we extended the framework to the case with additive noise, that naturally exists in physical layer links, and showed that the analog operator channel is robust with respect to the additive noise. As a consequence, the effect of noise is shown to be negligible in high signal-to-noise ratio regimes. Finally, we proposed a novel algebraic construction for subspace codes in the complex domain that outperforms the existing constructions in the literature, in terms of the rate-minimum distance tradeoff, for a wide range of code blocklength.

There are several directions for future research. Extending the results derived for minimum distance decoder by exploring list decoding algorithms is a natural direction for future research. In the case of noisy operator channel, obtaining tighter bounds on the subspace perturbation imposed by the additive noise is another research direction.

References

[1] "5G; Study on new radio (NR) access technology physical layer aspects," 3GPP TR 38.802, March 2017.

[2] T. Nakamura, S. Nagata, A. Benjebbour, Y. Kishiyama, T. Hai, S. Xiaodong, Y. Ning, and L. Nan, "Trends in small cell enhancements in LTE advanced," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 98–105, 2013.

[3] S. Sun, Q. Gao, Y. Peng, Y. Wang, and L. Song, "Interference management through CoMP in 3GPP LTE-advanced networks," *IEEE Wireless Communications*, vol. 20, no. 1, pp. 59–66, 2013.

[4] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Transactions on Networking (ToN)*, vol. 22, no. 1, pp. 137–150, 2014.

[5] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3579–3591, 2008.

[6] T. Ho, R. Koetter, M. Medard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," *Proceedings of IEEE International Symposium on Information Theory*, 2003.

[7] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.

[8] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2450–2454, 2002.

[9] C. Bachoc, "Linear programming bounds for codes in Grassmannian spaces," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2111–2125, 2006.

[10] C. Bachoc, Y. Ben-Haim, and S. Litsyn, "Bounds for codes in the Grassmann manifold," in *2006 IEEE 24th Convention of Electrical & Electronics Engineers in Israel*. IEEE, 2006, pp. 25–29.

[11] A. Barg and D. Nogin, "A bound on Grassmannian codes," *Journal of Combinatorial Theory, Series A*, vol. 113, no. 8, pp. 1629–1635, 2006.

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

[13] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[14] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[15] C. Cox, *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. John Wiley & Sons, 2012.

[16] O. Henkel, "Sphere-packing bounds in the Grassmann and Stiefel manifolds," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3445–3456, 2005.

[17] J. H. Conway, R. H. Hardin, and N. J. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[18] I. S. Dhillon, J. R. Heath, T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Experimental mathematics*, vol. 17, no. 1, pp. 9–35, 2008.

[19] G. Nebe, E. M. Rains, and N. J. Sloane, "The invariants of the Clifford groups," *Designs, Codes and Cryptography*, vol. 24, no. 1, pp. 99–122, 2001.

[20] P. Shor and N. J. A. Sloane, "A family of optimal packings in Grassmannian manifolds," *Journal of Algebraic Combinatorics*, vol. 7, no. 2, pp. 157–163, 1998.

[21] A. Calderbank, R. Hardin, E. Rains, P. Shor, and N. J. A. Sloane, "A group-theoretic framework for the construction of packings in Grassmannian spaces," *Journal of Algebraic Combinatorics*, vol. 9, no. 2, pp. 129–140, 1999.

[22] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. Sloane, "Quantum error correction and orthogonal geometry," *Physical Review Letters*, vol. 78, no. 3, p. 405, 1997.

[23] A. R. Calderbank, E. M. Rains, P. Shor, and N. J. Sloane, "Quantum error correction via codes over GF(4)," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1369–1387, 1998.

[24] N. J. A. Sloane, "Packing planes in four dimensions and other mysteries," *arXiv preprint math/0208017*, 2002.

[25] O. Tirkkonen1 and R. Calderbank, "Codebooks of complex lines based on binary subspace chirps," in *Proceedings of Information Theory Workshop (ITW)*, 2019.

[26] T. Strohmer and R. W. Heath Jr, "Grassmannian frames with applications to coding and communication," *Applied and computational harmonic analysis*, vol. 14, no. 3, pp. 257–275, 2003.

[27] P. Xia, S. Zhou, and G. B. Giannakis, "Achieving the Welch bound with difference sets," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1900–1907, 2005.

[28] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Transactions on information theory*, vol. 51, no. 1, pp. 188–209, 2005.

[29] G. Kutyniok, A. Pezeshki, R. Calderbank, and T. Liu, "Robust dimension reduction, fusion frames, and Grassmannian packings," *Applied and Computational Harmonic Analysis*, vol. 26, no. 1, pp. 64–76, 2009.

[30] C. Ding and T. Feng, "A generic construction of complex codebooks meeting the Welch bound," *IEEE Transactions on information theory*, vol. 53, no. 11, pp. 4245–4250, 2007.

[31] G. Han and J. Rosenthal, "Geometrical and numerical design of structured unitary space–time constellations," *IEEE transactions on information theory*, vol. 52, no. 8, pp. 3722–3735, 2006.

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

[33] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2618–2626, 2001.

[34] V. Aggarwal, A. Ashikhmin, and A. R. Calderbank, "A Grassmannian packing based on the Nordstrom-Robinson code," in *Proceedings of IEEE Information Theory Workshop*. IEEE, 2006, pp. 1–5.

[35] A. Ashikhmin and A. R. Calderbank, "Grassmannian packings from operator Reed–Muller codes," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5689–5714, 2010.

[36] H. Reboredo, F. Renna, R. Calderbank, and M. R. Rodrigues, "Compressive classification of a mixture of gaussians: Analysis, designs and geometrical interpretation," *arXiv preprint arXiv:1401.6962*, 2014.

[37] M. Nokleby, M. Rodrigues, and R. Calderbank, "Discrimination on the Grassmann manifold: Fundamental limits of subspace classifiers," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2133–2147, 2015.

[38] M. Nokleby, R. Calderbank, and M. R. Rodrigues, "Information-theoretic limits on the classification of Gaussian mixtures: Classification on the grassmann manifold," in *2013 IEEE Information Theory Workshop (ITW)*. IEEE, 2013, pp. 1–5.

[39] J. Huang, Q. Qiu, and R. Calderbank, "The role of principal angles in subspace classification," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1933–1945, 2015.

[40] S. Roy, "A note on critical angles between two flats in hyperspace with certain statistical applications," *Sankhyā: The Indian Journal of Statistics*, pp. 177–194, 1947.

[41] K. Ye and L.-H. Lim, "Distance between subspaces of different dimensions," *arXiv preprint arXiv:1407.0900*, vol. 4, 2014.

[42] J.-G. Sun, "Perturbation bounds for the Cholesky and QR factorizations," *BIT Numerical Mathematics*, vol. 31, no. 2, pp. 341–352, 1991.

[43] H. Golub and C. F. Van Loan, "Matrix computations," *Press, London*, 1996.

[44] S.-G. Hwang, "Cauchy's interlace theorem for eigenvalues of Hermitian matrices," *The American Mathematical Monthly*, vol. 111, no. 2, pp. 157–159, 2004.

[45] P. Elias, "Error-free coding," *IRE Trans. on Inform. Theory,*, pp. 29–37, 1954.

[46] G. D. Forney, "Concatenated codes." 1965.

[47] V. V. Zyablov, "An estimate of the complexity of constructing binary linear cascade codes," *Problemy Peredachi Informatsii*, vol. 7, no. 1, pp. 5–13, 1971.

[48] E. Blokh and V. V. Zyablov, "Linear concatenated codes," *Moscow, USSR: Nauka*, 1982.

[49] R. Tanner, "A recursive approach to low complexity codes," *IEEE Transactions on information theory*, vol. 27, no. 5, pp. 533–547, 1981.

[50] A. Barg and G. Zemor, "Distance properties of expander codes," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 78–90, 2006.

[51] G. Katsman, M. Tsfasman, and S. Vladut, "Modular curves and codes with a polynomial construction," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 353–355, 1984.

[52] K. W. Shum, I. Aleshnikov, P. V. Kumar, H. Stichtenoth, and V. Deolalikar, "A low-complexity algorithm for the construction of algebraic-geometric codes better than the Gilbert-Varshamov bound," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2225–2241, 2001.

[53] R. Lidl and H. Niederreiter, *Finite fields*. Cambridge university press, 1997, vol. 20.

[54] A. Weil, "On some exponential sums," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 34, no. 5, p. 204, 1948.

[55] H. Mahdavifar and A. Vardy, "Algebraic list-decoding on the operator channel," *Proceedings of IEEE International Symposium on Information Theory*, pp. 1193–1197, 2010.

[56] V. Guruswami, S. Narayanan, and C. Wang, "List decoding subspace codes from insertions and deletions," *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pp. 183–189, 2012.

[57] H. Mahdavifar and A. Vardy, "Algebraic list-decoding of subspace codes," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7814–7828, 2013.

[58] V. Guruswami, C. Wang, and C. Xing, "Explicit list-decodable rank-metric and subspace codes via subspace designs," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2707–2718, May 2016.

[59] H. Mahdavifar and A. Vardy, "Algebraic list-decoding in projective space: Decoding with multiplicities and rank-metric codes," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 1085–1100, 2019.

[60] D. R. Anderson, "A new class of cyclic codes," *SIAM Journal on Applied Mathematics*, vol. 16, no. 1, pp. 181–197, 1968. [Online]. Available: http://www.jstor.org/stable/2099415

[61] T. Helleseth, "On the covering radius of cyclic linear codes and arithmetic codes," *Discrete Applied Mathematics*, vol. 11, no. 2, pp. 157–173, 1985.

[62] A. Tietäinen, "On the covering radius of long binary BCH codes," *Discrete Applied Mathematics*, vol. 16, no. 1, pp. 75–77, 1987.

[63] I. Blake and J. Mark, "A note on complex sequences with low correlations (corresp.)," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 814–816, 1982.

[64] P. V. Kumar, T. Helleseth, and A. R. Calderbank, "An upper bound for Weil exponential sums over Galois rings and applications," *IEEE transactions on Information Theory*, vol. 41, no. 2, pp. 456–468, 1995.

APPENDIX

Lemma 15: Given a metric $d_0 : M \times M \to \mathbb{R}$ on a set M, the function $d(x,y) = d_0(x,y)^2 : M \times M \to \mathbb{R}$ is a 2-quasimetric on M.

Proof: We only need to show that (5) holds for $d(.,.)$ with $\sigma = 2$. The proof is by noting that for any $x, y, z \in M$ we have

$$d(x,z) = d_0(x,z)^2 \leqslant (d_0(x,y) + d_0(y,z))^2 \tag{82}$$
$$\leqslant 2(d_0(x,y)^2 + d_0(y,z)^2) \tag{83}$$
$$= 2(d(x,y) + d(y,z)), \tag{84}$$

where (82) follows from the triangle inequality and (83) is by Cauchy-Schwarz inequality. The remaining properties of a quasimetric follow in a straightforward way given that d_0 is a metric. ∎

Lemma 16: Suppose that the projection matrices of two subspaces $U, V \in \mathcal{P}(W)$ are simultaneously diagonalizable. Then the distance $d(.,.)$, as defined in (2), satisfies the triangle inequality, i.e.,

$$d(U,V) + d(V,T) \geqslant d(T,U),$$

for any $T \in \mathcal{P}(W)$.

Proof: By definition of simultaneously diagonalizable matrices, there exists an orthonormal basis for the ambient space W in which both $\boldsymbol{P}_U$ and $\boldsymbol{P}_V$ are diagonal. Let $u_{i,j}$,

$v_{i,j}$ and $t_{i,j}$, for $i,j \in [n]$, denote the entries of the projection matrices $\boldsymbol{P}_U$, $\boldsymbol{P}_V$, and $\boldsymbol{P}_T$ in this basis, respectively. Note that $u_{i,j} = v_{i,j} = 0$ for $i \neq j$. Also, the diagonal entries of $\boldsymbol{P}_U$ and $\boldsymbol{P}_V$ are either 0 or 1, since $\boldsymbol{P}_U$ and $\boldsymbol{P}_V$ are projection matrices. Let

$$I_U \stackrel{\text{def}}{=} \{i : u_{i,i} = 1, i \in [n]\},$$

and

$$I_V \stackrel{\text{def}}{=} \{i : v_{i,i} = 1, i \in [n]\}.$$

Also, we have $0 \leqslant t_{i,i} \leqslant 1$, for $i \in [n]$, a property that holds for diagonal entries of any projection matrix $\boldsymbol{P}_T$. By (12) we have the following relations for the pairwise distances between U, V, and T:

$$d(U,V) = |I_U \cap I_V^c| + |I_U^c \cap I_V|, \tag{85}$$

$$d(T,U) = \sum_{i,j\in[n]} |t_{i,j}|^2 + \sum_{i\in I_U} (1 - 2t_{i,i}), \tag{86}$$

$$d(V,T) = \sum_{i,j\in[n]} |t_{i,j}|^2 + \sum_{i\in I_V} (1 - 2t_{i,i}), \tag{87}$$

where the complements in (85) are taken with respect to $[n]$. Subtracting (87) from (86) yields

$$d(T,U) - d(V,T) = \sum_{i\in I_U\cap I_V^c} (1 - 2t_{i,i}) - \sum_{i\in I_U^c\cap I_V} (1 - 2t_{i,i}) \tag{88}$$

$$\leqslant |I_U \cap I_V^c| + |I_U^c \cap I_V|, \tag{89}$$

where (89) is by noting that $-1 \leqslant 1 - 2t_{i,i} \leqslant 1$ for $i \in [n]$. This together with (85) complete the proof. ∎

Lemma 17: For any $U, V \in \mathcal{P}(W)$ we have

$$d(U^\perp, V^\perp) = d(U,V), \tag{90}$$

where $d(.,.)$ is defined in Definition 2.

Proof: It is well known that $\boldsymbol{P}_{U^\perp} = \boldsymbol{I} - \boldsymbol{P}_U$. Hence,

$$d(U^\perp, V^\perp) = \mathrm{tr}((\boldsymbol{P}_U^\perp - \boldsymbol{P}_V^\perp)^2) = \mathrm{tr}((\boldsymbol{P}_V - \boldsymbol{P}_U)^2) = d(U,V).$$

∎

Lemma 18: Suppose that $U, T \in \mathcal{P}(W)$ and let $V = U \oplus T$. Then we have

$$d(U,V) = \dim(T), \tag{91}$$

where $d(.,.)$ is defined in Definition 2.

Proof: Let $t = \dim T$ and $u = \dim U$. Then $\dim V = t + u$. One can always find a basis for W, namely $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$, such that $U = \langle \boldsymbol{I}_u \rangle$ and $V = \langle \boldsymbol{I}_v \rangle$ where $\boldsymbol{I}_u$ and $\boldsymbol{I}_v$ are identity matrices of dimensions u and v, respectively. Consequently, the orthogonal projection matrices associated with these subspaces are $\boldsymbol{P}_U = \left[\begin{array}{c|c} \boldsymbol{I}_u & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{0} \end{array}\right]$ and $\boldsymbol{P}_V = \left[\begin{array}{c|c} \boldsymbol{I}_v & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{0} \end{array}\right]$. Then the lemma follows immediately by using (12) and noting that the distance is rotation invariant by Lemma 1. ∎

Lemma 19: Let $\boldsymbol{B} \in \mathbb{C}^{l\times n}$. Then $\left\|\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right\| \leqslant \|\boldsymbol{B}\|^2$.

Proof: Let $\sigma_i \in \mathbb{R}$, for $i \in [l]$, denote the singular values of $\boldsymbol{B}$. Then the singular values of $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ are λ_i^2's. Hence, by using (3) we have

$$\left\|\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\right\| = \sqrt{\sum_{i=1}^{l} \sigma_i^4} \leqslant \sum_{i=1}^{l} \sigma_i^2 = \|\boldsymbol{B}\|^2.$$

∎

Mahdi Soleymani (Student Member, IEEE) received his B.S. and M.S. degrees in Electrical Engineering at Sharif University of Technology, Tehran, Iran, in 2014 and 2016, respectively. He is currently pursuing his Ph.D. degree in Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. He received the Honourable Mention Award at the International Physics Olympiad (IPhO) in 2010 and a Gold Medal at Iran National Physics Olympiad in 2009. His research interests lie in the area of coding and information theory with applications to distributed storage and computing systems, wireless networks, and machine learning.

Hessam Mahdavifar (Member, IEEE) is an Assistant Professor in the Department of Electrical Engineering and Computer Science at the University of Michigan Ann Arbor. He received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2007, and the M.Sc. and the Ph.D. degrees from the University of California San Diego (UCSD), La Jolla, in 2009, and 2012, respectively, all in electrical engineering. He was with the Samsung US R&D between 2012 and 2016, in San Diego, US, as a staff research engineer.

He received the NSF career award in 2020. He also received Best Paper Award in 2015 IEEE International Conference on RFID, and the 2013 Samsung Best Paper Award. He also received two Silver Medals at International Mathematical Olympiad in 2002 and 2003, and two Gold Medals at Iran National Mathematical Olympiad in 2001 and 2002. His main area of research is coding and information theory with applications to wireless communications, security, privacy, and machine learning.