Coding for Errors and Erasures in Random Network Coding

Ralf Kötter, *Senior Member, IEEE*, and Frank R. Kschischang, *Fellow, IEEE*

Abstract—The problem of error-control in random linear network coding is considered. A "noncoherent" or "channel oblivious" model is assumed where neither transmitter nor receiver is assumed to have knowledge of the channel transfer characteristic. Motivated by the property that linear network coding is vector-space preserving, information transmission is modeled as the injection into the network of a basis for a vector space V and the collection by the receiver of a basis for a vector space U. A metric on the projective geometry associated with the packet space is introduced, and it is shown that a minimum-distance decoder for this metric achieves correct decoding if the dimension of the space $V \cap U$ is sufficiently large. If the dimension of each codeword is restricted to a fixed integer, the code forms a subset of a finite-field Grassmannian, or, equivalently, a subset of the vertices of the corresponding Grassmann graph. Sphere-packing and sphere-covering bounds as well as a generalization of the Singleton bound are provided for such codes. Finally, a Reed–Solomon-like code construction, related to Gabidulin's construction of maximum rank-distance codes, is described and a Sudan-style "list-1" minimum-distance decoding algorithm is provided.

Index Terms—Network coding, network error correction, subspace metric.

I. INTRODUCTION

RANDOM network coding [1]–[3] is a powerful tool for disseminating information in networks, yet it is susceptible to packet transmission errors caused by noise or intentional jamming. Indeed, in the most naive implementations, a single error in one received packet would typically render the entire transmission useless when the erroneous packet is combined with other received packets to deduce the transmitted message. It might also happen that insufficiently many packets from one generation reach the intended receivers, so that the problem of deducing the information cannot be completed.

In this paper, we formulate a coding theory in the context of a "noncoherent" or "channel oblivious" transmission model for random linear network coding that captures the effects both of errors, i.e., erroneously received packets, and of erasures, i.e., insufficiently many received packets. We are partly motivated by the close analogy between the $\mathbb{F}_q$-linear channel produced in random linear network coding and the $\mathbb{C}$-linear channel

produced in noncoherent multiple-antenna channels [4], where neither the transmitter nor the receiver is assumed to have knowledge of the channel transfer characteristic. In contrast with previous approaches to error control in random linear network coding, e.g., [5]–[8], the noncoherent transmission strategy taken in this paper is oblivious to the underlying network topology and to the particular linear network coding operations performed at the various network nodes. Here, information is encoded in the choice at the transmitter of a *vector space* (not a vector), and the choice of vector space is conveyed via transmission of a generating set for the space.

Just as codes defined on the complex Grassmann manifold play an important role in noncoherent multiple-antenna channels [4], we find that codes defined in an appropriate Grassmannian associated with a vector space over a finite field play an important role here, but with a different metric associated with the structure of the corresponding Grassman graph.

The standard, widely advocated approach to random linear network coding (see, e.g., [2]) involves transmission of packet "headers" that are used to record the particular linear combination of the components of the message present in each received packet. As we will show, this "uncoded" transmission may be viewed as a particular code of subspaces, but a "suboptimal" one, in the sense that the Grassmannian contains more spaces of a particular dimension than those obtained by prepending a header to the transmitted packets. Indeed, the very notion of a header or local and global encoding vectors, crucial in [2], [3], [8], is moot in our context.

A somewhat more closely related approach is that of [9], which deals with reliable communication in networks with so-called "Byzantine adversaries," who are assumed to have some ability to inject packets into the network and sometimes also to eavesdrop (i.e., read packets transmitted in the network) [10]. It is shown that an optimal communication rate (which depends on the adversary's eavesdropping capability) is achievable with high probability with codes of sufficiently long block length. The work of this paper, in contrast, concentrates more on the possibility of code constructions with a prescribed deterministic correction capability, which, however, asymptotically can achieve the same rates as would be achieved in the so-called "omniscient adversary model" of [9].

The remainder of this paper is organized as follows.

In Section II, we introduce the "operator channel" as a concise and convenient abstraction of the channel encountered in random linear network coding, when neither transmitter nor receiver has knowledge of the channel transfer characteristics. The input and output alphabet for an operator channel is the projective geometry (the set of all subspaces) associated with a given

R. Kötter is with the Institute for Communications Engineering, Technical University of Munich, D-80333 Munich, Germany (e-mail: ralf.koetter@tum.de).

F. R. Kschischang is with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: frank@comm.utoronto.ca).

vector space over a finite field $\mathbb{F}_q$. In Section III, we define a metric on this set that is natural and suitable in the context of random linear network coding. The transmitter selects a space V for transmission, indicating this choice by injection into the network of a set of packets that generate V. The receiver collects packets that span some received space U. We show that correct decoding is possible with a minimum-distance decoder if the dimension of the space $V \cap U$ is sufficiently large, just as correct decoding in the conventional Hamming metric is possible if the received vector u agrees with the transmitted vector v in sufficiently many coordinates.

We will usually confine our attention to constant-dimension codes, i.e., codes in which all codewords have the same dimension. In this case, the code is a subset of the corresponding Grassmannian, or, equivalently, a subset of the vertices of the corresponding Grassmann graph. Coding in the Grassmann graph has been an active area of research in combinatorics [11]–[15], where the problem has been studied as a packing problem that arises naturally. In Section IV, we derive elementary coding bounds, analogous to the sphere-packing (Hamming) upper bounds and the sphere-covering (Gilbert–Varshamov) lower bounds for such codes. By defining an appropriate notion of puncturing, we also derive a Singleton bound. Asymptotic versions of these bounds are also given. Some Johnson-type bounds on constant dimension codes can be found in [16].

A notable application for constant-dimension codes is their use as so-called linear authentication codes introduced by Wang, Xing, and Safavi-Naini [17]. There, the properties of codes in the Grassmann graphs are used to detect tampering with an authenticated message. The authors of [17] describe a construction by which constant-dimension codes that are suitable for linear authentication can be obtained from rank-distance codes, in particular from the maximum rank-distance codes of Gabidulin [18]. In Section V, we revisit this construction in the context of the coding metric defined in this paper, and we show that these codes achieve the Singleton bound asymptotically. The main result of Section V is the development of an efficient Sudan-style "list-1" minimum-distance decoding algorithm for these codes. A related polynomial-reconstruction-based decoder for Gabidulin codes has been described by Loidreau [19]. The connection between rank-metric codes and a generalized rank-metric decoding problem induced by random linear network coding is explored in [20].

II. OPERATOR CHANNELS

We begin by formulating our problem for the case of a single unicast, i.e., communication between a single transmitter and a single receiver. Generalization to multicasting is straightforward.

To capture the essence of random linear network coding, recall [2], [3] that communication between transmitter and receiver occurs in a series of rounds or "generations;" during each generation, the transmitter injects a number of fixed-length packets into the network, each of which may be regarded as a row vector of length N over a finite field $\mathbb{F}_q$. These packets propagate through the network, possibly passing through a

number of intermediate nodes between transmitter and receiver. Whenever an intermediate node has an opportunity to send a packet, it creates a random $\mathbb{F}_q$-linear combination of the packets it has available and transmits this random combination. Finally, the receiver collects such randomly generated packets and tries to infer the set of packets injected into the network. There is *no* assumption here that the network operates synchronously or without delay or that the network is acyclic.

The set of successful packet transmissions in a generation induces a directed multigraph with the same vertex set as the network, in which edges denote successful packet transmissions. The rate of information transmission (packets per generation) between the transmitter and the receiver is upper-bounded by the min-cut between these nodes, i.e., by the minimum number of edge deletions in the graph that would cause the separation of the transmitter and the receiver. It is known that random linear network coding in $\mathbb{F}_q$ is able to achieve a transmission rate that achieves the min-cut rate with probability approaching one as $q \to \infty$ [3].

Let $\{p_1, p_2, \ldots, p_M\}$, $p_i \in \mathbb{F}_q^N$, denote the set of injected vectors. In the error-free case, the receiver collects packets $y_j, j = 1, 2, \ldots, L$ where each y_j is formed as $y_j = \sum_{i=1}^{M} h_{j,i} p_i$ with unknown, randomly chosen coefficients $h_{j,i} \in \mathbb{F}_q$.

We note that *a priori* L is not fixed and the receiver would normally collect as many packets as possible. However, as noted above, properties of the network such as the min-cut between the transmitter and the receiver may influence the joint distribution of the $h_{j,i}$ and, at some point, there will be no benefit from collecting further redundant information.

If we choose to consider the injection of T erroneous packets, this model is enlarged to include error packets $e_t, t = 1, \ldots, T$ to give

$$y_j = \sum_{i=1}^{M} h_{j,i} p_i + \sum_{t=1}^{T} g_{j,t} e_t$$

where again $g_{j,t} \in \mathbb{F}_q$ are unknown random coefficients. Note that since these erroneous packets may be injected anywhere within the network, they may cause widespread error propagation. In particular, if $g_{j,1} \neq 0$ for all j, even a single error packet e_1 has the potential to corrupt each and every received packet.

In matrix form, the transmission model may be written as

$$y = Hp + Ge \tag{1}$$

where H and G are random $L \times M$ and $L \times T$ matrices, respectively, p is the $M \times N$ matrix whose rows are the transmitted vectors, y is the $L \times N$ matrix whose rows are the received vectors, and e is the $T \times N$ matrix whose rows are the error vectors.

The network topology will certainly impose some structure on the matrices H and G. For example, H may be rank-deficient if the min-cut between transmitter and receiver is not large enough to support the transmission of M independent packets during the lifetime of one generation.[1] While the possibility may exist to exploit the structure of the network, in the strategy

---

[1]This statement can be made precise once the precise protocol for transmission of a generation has been fixed. However, for the purpose of this paper it is sufficient to summarily model "rank deficiency" as one potential cause of errors.

adopted in this paper we do not take any possibly finer structure of the matrix H into account. Indeed, any such fine structure can be effectively obliterated by randomization at the source, i.e., if, rather than injecting packets p_i into the network, the transmitter were instead to inject random linear combinations of the p_i.

At this point, since H is random, we may ask what property of the injected sequence of packets remains invariant in the channel described by (1), even in the absence of noise ($e = 0$)? Since H is a random matrix, all that is fixed by the product Hp is the row space of p. Indeed, as far as the receiver is concerned, any of the possible generating sets for this space are equivalent. We are led, therefore, to consider information transmission not via the choice of p, but rather by the choice of the vector space spanned by the rows of p. This simple observation is at the heart of the channel models and transmission strategies considered in this paper. Indeed, with regard to the vector space selected by the transmitter, the only deleterious effect that a multiplication with H may have is that Hp may have smaller rank than p, due to, e.g., an insufficient min-cut or packet erasures, in which case Hp generates a subspace of the row space of p.

Let W be a fixed N-dimensional vector space over $\mathbb{F}_q$. All transmitted and received packets will be vectors of W; however, we will describe a transmission model in terms of subspaces of W spanned by these packets. Let $\mathcal{P}(W)$ denote the set of all subspaces of W, an object often called the projective geometry of W. The dimension of an element $V \in \mathcal{P}(W)$ is denoted as $\dim(V)$. The sum of two subspaces U and V of W is $U + V = \{u + v : u \in U, v \in V\}$. Equivalently, $U + V$ is the smallest subspace of W containing both U and V. If $U \cap V = \{0\}$, i.e., if U and V have trivial intersection, then the sum $U + V$ is a direct sum, denoted as $U \oplus V$. Clearly, $\dim U \oplus V = \dim(U) + \dim(V)$. For any subspaces U and V we have $V = (U \cap V) \oplus V'$ for some subspace V' isomorphic to the quotient space $V/(U \cap V)$. In this case, $U + V = U + ((U \cap V) \oplus V') = U \oplus V'$.

For integer $k \geq 0$, we define a stochastic operator $\mathcal{H}_k$, called an "erasure operator," that operates on the subspaces of W. If $\dim(V) > k$, then $\mathcal{H}_k(V)$ returns a randomly chosen k-dimensional subspace of V; otherwise, $\mathcal{H}_k(V)$ returns V. For the purposes of this paper, the distribution of $\mathcal{H}_k(V)$ is unimportant; for example, it could be chosen to be uniform. Given two subspaces U and V of W, it is always possible to realize U as $U = \mathcal{H}_k(V) \oplus E$ for some subspace E of W, assuming that $k = \dim(U \cap V)$ and that $\mathcal{H}_k(U)$ realizes $U \cap V$.

We define the following "operator channel" as a concise transmission model for network coding.

Definition 1: An operator channel C associated with the ambient space W is a channel with input and output alphabet $\mathcal{P}(W)$. As described above, the channel input V and channel output U can always be related as

$$U = \mathcal{H}_k(V) \oplus E \qquad (2)$$

where $k = \dim(U \cap V)$ and E is an error space. In transforming V to U, we say that the operator channel commits $\rho = \dim(V) - k$ *erasures* and $t = \dim(E)$ *errors*.

Note that we have chosen to model the error space E as intersecting trivially with the transmitted subspace V, and thus

the choice of E is not independent of V. However, if we were to model the received space as $U = \mathcal{H}_k(V) + E$ for an arbitrary error space E, then, since E always decomposes for some space E' as $E = (E \cap V) \oplus E'$, we would get $U = \mathcal{H}_k(V) + (E \cap V) \oplus E' = \mathcal{H}_{k'}(V) \oplus E'$ for some $k' \geq k$. In other words, components of an error space E that intersect with the transmitted space V would only be helpful, possibly decreasing the number of erasures seen by the receiver.

In summary, an operator channel takes in a vector space and puts out another vector space, possibly with erasures (deletion of vectors from the transmitted space) or errors (addition of vectors to the transmitted space).

This definition of an operator channel makes a very clear connection between network coding and classical information theory. Indeed, an operator channel can be seen as a standard discrete memoryless channel with input and output alphabet $\mathcal{P}(W)$. By imposing a channel law, i.e., transition probabilities between spaces, it would (at least conceptually) be straightforward to compute capacity, error exponents, etc. Indeed, only slight extensions would be necessary concerning the ergodic behavior of the channel. For the present paper we constrain our attention to the question of constructing good codes in $\mathcal{P}(W)$, which is an essentially combinatorial problem. The codes we construct may be regarded as "one-shot" codes, i.e., codes of length one, since the transmission of a codeword will induce exactly one use of the operator channel.

III. CODING FOR OPERATOR CHANNELS

Definition 1 concisely captures the effect of random linear network coding in the presence of networks with erasures, varying min-cuts and/or erroneous packets. Indeed, we will show how to construct codes for this channel that correct combinations of errors and erasures. Before we give such a construction we need to define a suitable metric.

A. A Metric on $\mathcal{P}(W)$

Let $\mathbb{Z}_+$ denote the set of nonnegative integers. We define a function $d : \mathcal{P}(W) \times \mathcal{P}(W) \to \mathbb{Z}_+$ by

$$d(A, B) := \dim(A + B) - \dim(A \cap B). \qquad (3)$$

Since $\dim(A + B) = \dim(A) + \dim(B) - \dim(A \cap B)$, we may also write

$$d(A, B) = \dim(A) + \dim(B) - 2\dim(A \cap B)$$
$$= 2\dim(A + B) - \dim(A) - \dim(B).$$

The following lemma is a cornerstone for code design for the operator channel of Definition 1.

Lemma 1: The function

$$d(A, B) := \dim(A + B) - \dim(A \cap B)$$

is a metric for the space $\mathcal{P}(W)$.

Proof: We need to check that for all subspaces $A, B, X \in \mathcal{P}(W)$ we have: i) $d(A, B) \geq 0$ with equality if and only if $A = B$, ii) $d(A, B) = d(B, A)$, and iii) $d(A, B) \leq d(A, X) +$

$d(X, B)$. The first two conditions are clearly true and so we focus on the third condition, the triangle inequality. We have

$$
\begin{aligned}
\frac{1}{2}&(d(A,B) - d(A,X) - d(X,B)) \\
&= \dim(A \cap X) + \dim(B \cap X) - \dim(X) - \dim(A \cap B) \\
&= \underbrace{\dim(A \cap X + B \cap X) - \dim(X)}_{\leq 0} \\
&\quad + \underbrace{\dim(A \cap B \cap X) - \dim(A \cap B)}_{\leq 0} \\
&\leq 0
\end{aligned}
$$

where the first inequality follows from the property that $A \cap X + B \cap X \subseteq X$ and the second inequality follows from the property that $A \cap B \cap X \subseteq A \cap B$. $\qquad \square$

Remark 1: The fact that $d(A, B)$ is a metric also follows from the fact that this quantity represents the distance of a geodesic between A and B in the undirected Hasse graph representing the lattice of subspaces of W [11] partially ordered by inclusion, i.e., where $X \preceq Y$ if and only if X is a subspace of Y. In this graph, the vertices correspond to the elements of $\mathcal{P}(W)$ and an edge joins a subspace X with a subspace Y if and only if $|\dim(X) - \dim(Y)| = 1$ and either $X \subset Y$ or $Y \subset X$. Just as the hypercube provides the appropriate setting for coding in the Hamming metric, the undirected Hasse graph represents the appropriate setting for coding in the context considered here. $\square$

If a basis for W is fixed, then since W is an N-dimensional vector space over $\mathbb{F}_q$, the elements of W may be represented by N-tuples of $\mathbb{F}_q$-valued coordinates with respect to this basis. We take the usual inner product between vectors $u = (u_1, \ldots, u_N)$ and $v = (v_1, \ldots, v_N)$ as $(u, v) = \sum_{i=1}^{N} u_i v_i$. If U is a k-dimensional subspace of W, then the orthogonal subspace

$$
U^{\perp} = \{v \in W : (u, v) = 0 \text{ for all } u \in U\}
$$

is a space of dimension $N - k$. It is well known that for any subspaces U and V of W that $(U^{\perp})^{\perp} = U$

$$
(U + V)^{\perp} = U^{\perp} \cap V^{\perp} \quad \text{and} \quad (U \cap V)^{\perp} = U^{\perp} + V^{\perp}.
$$

It follows that

$$
\begin{aligned}
d(U^{\perp}, V^{\perp}) &= \dim(U^{\perp} + V^{\perp}) - \dim(U^{\perp} \cap V^{\perp}) \\
&= \dim((U \cap V)^{\perp}) - \dim((U + V)^{\perp}) \\
&= (N - \dim(U \cap V)) - (N - \dim(U + V)) \\
&= \dim(U + V) - \dim(U \cap V) \\
&= d(U, V). \qquad (4)
\end{aligned}
$$

Thus, the distance between subspaces U and V is perfectly mirrored by the distance between the orthogonal subspaces $U^{\perp}$ and $V^{\perp}$.

B. Codes

Let W be an N-dimensional vector space over $\mathbb{F}_q$. A *code* for an operator channel with ambient space W is simply a nonempty subset of $\mathcal{P}(W)$, i.e., a nonempty collection of subspaces of W.

The size of a code $\mathcal{C}$ is denoted by $|\mathcal{C}|$. The minimum distance of $\mathcal{C}$ is denoted by

$$
D(\mathcal{C}) := \min_{X, Y \in \mathcal{C}: X \neq Y} d(X, Y).
$$

The maximum dimension of the codewords of $\mathcal{C}$ is denoted by

$$
\ell(\mathcal{C}) := \max_{X \in \mathcal{C}} \dim(X).
$$

If the dimension of each codeword of $\mathcal{C}$ is the same, then $\mathcal{C}$ is said to be a *constant-dimension* code.

In analogy with the $[n, k, d]$ triple that describes the parameters of a classical linear error correcting code of length n, dimension k, and minimum Hamming distance d, a code $\mathcal{C}$ for an operator channel with an N-dimensional ambient space over $\mathbb{F}_q$ is said to be of type $[N, \ell(\mathcal{C}), \log_q |\mathcal{C}|, D(\mathcal{C})]$.

The *complementary code* corresponding to a code $\mathcal{C}$ is the code $\mathcal{C}^{\perp} = \{U^{\perp} : U \in \mathcal{C}\}$ obtained from the orthogonal subspaces of the codewords of $\mathcal{C}$. In view of (4), we have $D(\mathcal{C}^{\perp}) = D(\mathcal{C})$. If $\mathcal{C}$ is a constant-dimension code of type $[N, \ell, M, D]$, then $\mathcal{C}^{\perp}$ is a constant-dimension code of type $[N, N - \ell, M, D]$.

Before we study bounds and constructions of codes in $\mathcal{P}(W)$, we need a proper definition of rate. Let $\mathcal{C} \subset \mathcal{P}(W)$ be a code of type $[N, \ell, \log_q |\mathcal{C}|, D]$. To transmit a space $V \in \mathcal{C}$ would require the transmitter to inject up to $\ell(\mathcal{C})$ (basis) vectors from V into the network, corresponding to the transmission of $N\ell$ q-ary symbols. This motivates the following definition.

Definition 2: Let $\mathcal{C}$ be a code of type $[N, \ell, \log_q(|\mathcal{C}|), D]$. The normalized weight λ, the rate R, and the normalized minimum distance δ of $\mathcal{C}$ are defined as

$$
\lambda = \frac{\ell}{N}, \quad R = \frac{\log_q(|\mathcal{C}|)}{N\ell}, \quad \text{and} \quad \delta = \frac{D}{2\ell}. \qquad \square
$$

The parameters λ, R, and δ are quite natural. The normalized weight λ takes the role of the energy of a spherical code in Euclidean space, or the equivalent weight parameter for constant-weight codes. As such, λ is naturally limited to the range $[0, 1]$. For constant-dimension codes, just as in the case of constant-weight codes, the interesting range can actually be limited to $[0, \frac{1}{2}]$ as any code $\mathcal{C}$ with $\ell > N/2$ corresponds to the complementary code $\mathcal{C}^{\perp}$ with $\ell < N/2$ and having the identical distance properties. The definition of δ gives a natural range of $[0, 1]$. Indeed, a normalized distance of 1 could only be obtained by spaces having trivial intersection. The rate R of a code is restricted to the range $[0, 1]$, with a rate of 1 only being approachable for $\lambda \to 0$.

The fundamental code construction problem for the operator channel of Definition 1 thus becomes the determination of achievable tuples $[\lambda, R, \delta]$ as the dimension of ambient space N becomes arbitrarily large. We note that this setup may lack physical reality since it assumes that the network can operate with arbitrarily long packets; thus, we will try to express our results for finite length N whenever possible. Furthermore, as noted above, the codes we consider here are "one-shot" codes that induce just a single use of the operator channel. In situations where the channel characteristics (such as the network min-cut) are time-varying, it may be interesting and useful to define codes that induce n uses of the operator channel, consid-

ering what performance is attainable as $n \to \infty$. We will not pursue this direction in this paper.

C. Error and Erasure Correction

A minimum-distance decoder for a code $\mathcal{C}$ is one that takes the output U of an operator channel and returns a nearest codeword $V \in \mathcal{C}$, i.e., a codeword $V \in \mathcal{C}$ satisfying, for all $V' \in \mathcal{C}, d(U,V) \le d(U,V')$.

The importance of the minimum distance $D(\mathcal{C})$ for a code $\mathcal{C} \subset \mathcal{P}(W)$ is given in the following theorem, which provides the combined error-and-erasure-correction capability of $\mathcal{C}$ under minimum-distance decoding. Define $(x)_+$ as $(x)_+ := \max\{0,x\}$.

Theorem 2: Assume we use a code $\mathcal{C}$ for transmission over an operator channel. Let $V \in \mathcal{C}$ be transmitted, and let

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

be received, where $\dim(E) = t$. Let $\rho = (\ell(\mathcal{C}) - k)_+$ denote the maximum number of erasures induced by the channel. If

$$2(t + \rho) < D(\mathcal{C}) \qquad (5)$$

then a minimum distance decoder for $\mathcal{C}$ will produce the transmitted space V from the received space U.

Proof: Let $V' = \mathcal{H}_k(V)$. From the triangle inequality we have

$$d(V,U) \le d(V,V') + d(V',U) \le \rho + t.$$

If $T \ne V$ is any other codeword in $\mathcal{C}$, then

$$D(\mathcal{C}) \le d(V,T) \le d(V,U) + d(U,T)$$

from which it follows that

$$d(U,T) \ge D(\mathcal{C}) - d(V,U) \ge D(\mathcal{C}) - (\rho + t).$$

Provided that the inequality (5) holds, then $d(U,T) > d(U,V)$ and hence a minimum-distance decoder must produce V. $\square$

Not surprisingly, given the symmetry in this setup between erasures (deletion of dimensions due to, e.g., an insufficient min-cut in the network or an unfortunate choice of coefficients in the random linear network code) and errors (insertion of dimensions due to errors or deliberate malfeasance), erasures and errors are equally costly to the decoder. This stands in apparent contrast with traditional error correction (where erasures cost less than errors); however, this difference is merely an accident of terminology. A perhaps more closely related classical concept would be that of "insertions" and "deletions."

If we can be sure that the projection operation is moot (expressed by choosing operator $\mathcal{H}_{\dim(W)}$ which operates as an identity on each subspace of W) or that the network produces no errors (expressed by choosing the error space $E = \{0\}$), we get the following corollary.

Corollary 3: Assume we use a code $\mathcal{C}$ for transmission over an operator channel where $V \in \mathcal{C}$ is transmitted. If

$$U = \mathcal{H}_{\dim(W)}(V) \oplus E = V \oplus E$$

is received, and if $2t < D(\mathcal{C})$ where $\dim(E) = t$, then a minimum distance decoder for $\mathcal{C}$ will produce V. Symmetrically, if

$$U = \mathcal{H}_k(V) \oplus \{0\} = \mathcal{H}_k(V)$$

is received, and if $2\rho < D(\mathcal{C})$ where $\rho = (\ell(\mathcal{C}) - k)_+$, then a minimum-distance decoder for $\mathcal{C}$ will produce V.

In other words, the first part of the corollary states that in the absence of erasures a minimum-distance decoder uniquely corrects errors up to dimension

$$t \le \left\lfloor \frac{D(\mathcal{C}) - 1)}{2} \right\rfloor,$$

precisely in parallel to the standard error correction situation.

D. Constant-Dimension Codes

In the context of network coding, it is natural to consider codes in which each codeword has the same dimension, as knowledge of the codeword dimension can be exploited by the decoder to initiate decoding. Constant-dimension codes are analogous to constant-weight codes in Hamming space (in which every codeword has constant Hamming weight) or to spherical codes in Euclidean space (in which every codeword has constant energy).

As noted above, when considering constant-dimension codes we may restrict ourselves to codes of type $[N, \ell, M, D]$ with $\ell \le N - \ell$, since a code of type $[N, \ell, M, D]$ with $\ell > N - \ell$ may be replaced with its complementary code $C^\perp$ while maintaining all distance properties (therefore maintaining all error- and erasure-correcting capability).

Constant-dimension codes are naturally described as particular vertices of a so-called *Grassmann graph*, also called a q-Johnson scheme, where the latter name emphasizes that these objects constitute association schemes. A formal definition is given as follows.

Definition 3: Denote by $\mathcal{P}(W, \ell)$ the set of all subspaces of W of dimension ℓ. This object is known as a Grassmannian. The Grassmann graph $G_{W,\ell}$ has vertex set $\mathcal{P}(W, \ell)$ with an edge joining vertices U and V if and only if $d(U,V) = 2$. $\square$

Remark: It is well known that $G_{W,\ell}$ is distance regular [21] and an association scheme with relations given by the distance between spaces. As such, practically all techniques for bounds in the Hamming association scheme apply. In particular, sphere-packing and sphere-covering concepts have a natural equivalent formulation. We explore these directions in Section IV. We also note that the distance between two spaces U, V in $\mathcal{P}(W, \ell)$ introduced in (3) is, like in the case of constant-weight codes in the Hamming metric, an even number equal to twice the graph distance in the Grassmann graph.[2] As noted Section I, codes in the Grassmann graph have been considered previously in [11]–[15], [17]. $\square$

When modeling the operation of random linear network coding by the operator channel of Definition 1, there is no further need to specify the precise network protocol. In particular,

---

[2]Defining a distance as half of $d(U,V)$ would give noninteger values for packings in $\mathcal{P}(W)$.

we assume the receiver knows that a codeword V from $\mathcal{P}(W, \ell)$ was transmitted. In this situation, a receiver could choose to collect packets until the collected packets, interpreted as vectors, span an ℓ-dimensional space. This situation would correspond to an operator channel of type $U = \mathcal{H}_{\ell - t(E)}(V) \oplus E$, corresponding to $t(E)$ erasures and $t(E)$ errors. According to Theorem 2, we can thus correct up to an error dimension $\lfloor \frac{D(\mathcal{C}) - 1}{4} \rfloor$. To some extent, this additional factor of two reflects the choice of the distance measure as being twice the graph distance in $G_{W, \ell}$. Note that this situation also would arise if the errors originated in network-coded transmissions through the min-cut edges in the graph. If the errors do not affect the min-cut but may have arisen anywhere else in the network, a receiver can choose to collect packets until an $\ell + t(E)$-dimensional space $V \oplus E$ has been recovered.[3] In this case, the error correction capability would increase to an error dimension $\lfloor \frac{D(\mathcal{C}) - 1}{2} \rfloor$. We do not study the implications of this observation further in this paper, since the coding-theoretic goal of constructing good codes in $\mathcal{P}(W)$ is not affected by this. Nevertheless, we point out that a properly designed protocol can (and should) take advantage of these differences.

E. Examples of Codes

We conclude this section with two examples of codes in $\mathcal{P}(W, \ell)$.

Example 1: Let W be the vector space of N-tuples over $\mathbb{F}_q$. Consider the set $\mathcal{C} \subset \mathcal{P}(W, \ell)$ of spaces $U_i, i = 1, 2, \ldots, |\mathcal{C}|$ with generator matrices $G(U_i) = (I | A_i)$ where I is an $\ell \times \ell$ identity matrix and the A_i are all possible $\ell \times (N - \ell)$ matrices over $\mathbb{F}_q$. It is easy to see that all $G(U_i)$ generate different spaces, intersecting in subspaces of dimension at most $\ell - 1$ and that, hence, the minimum distance of the code is $2\ell - 2(\ell - 1) = 2$. The code is a constant-dimension code of type $[N, \ell, \ell(N - \ell), 2]$ with normalized weight $\lambda = \ell/N$, rate $R = 1 - \lambda$, and normalized distance $\delta = \frac{1}{\lambda N}$. $\qquad\square$

The first example corresponds to a trivial code that offers no error protection at all. While this code has been advocated widely for random linear network coding it is by no means the optimal code for a given distance $D = 2$, as can be seen in the following example.

Example 2: Again let W be the space of vectors of length N. We now choose the code $\mathcal{C}' = \mathcal{P}(W, \ell)$, which yields a constant-dimension code of type $[N, \ell, \log_q |\mathcal{P}(W, \ell)|, 2]$ which is clearly larger than the code $\mathcal{C}$ of Example 1. As explained in Section IV, $|\mathcal{P}(W, \ell)|$ is equal to the Gaussian coefficient $\begin{bmatrix} N \\ \ell \end{bmatrix}_q$.

We note that $\mathcal{C}''$, defined as $\mathcal{C}'' = \bigcup_{i=1}^{\ell} \mathcal{P}(W, i)$ (no longer a constant-dimension code) is obviously an even bigger code (albeit with minimum distance $D = 1$) that can be used for random linear network coding while not using more network resources than $\mathcal{C}$ and $\mathcal{C}'$. However, in contrast to $\mathcal{C}'$, the receiver must be able to determine when the transmission of the code space is complete. This information is implicit in $\mathcal{C}'$ and $\mathcal{C}$ since the dimension of the transmitted space is fixed beforehand. $\qquad\square$

---

[3]Not knowing the effective dimension of E, i.e., the dimension of $E/(E \cap V)$, in practice the receiver would just collect as many packets as possible and attempt to reconstruct the corresponding space.

In the next section, we provide a few standard bounds for codes in our setup.

IV. BOUNDS ON CODES

A. Preliminaries

We will be interested in constructing constant-dimension codes. We start this section by introducing some notation that will be relevant for packings in $\mathcal{P}(W, \ell)$ where W is an N-dimensional vector space over $\mathbb{F}_q$.

The q-ary *Gaussian coefficient*, the q-analogue of the binomial coefficient, is defined, for nonnegative integers ℓ and n with $\ell \leq n$, by

$$\begin{bmatrix} n \\ \ell \end{bmatrix}_q := \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-\ell+1} - 1)}{(q^\ell - 1)(q^{\ell-1} - 1) \cdots (q - 1)}$$
$$= \prod_{i=0}^{\ell-1} \frac{q^{n-i} - 1}{q^{\ell-i} - 1}$$

where the empty product obtained when $\ell = 0$ is interpreted as 1.

As is well known (see, e.g., [22, Ch. 24]), the Gaussian coefficient $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ gives the number of distinct ℓ-dimensional subspaces of an n-dimensional vector space over $\mathbb{F}_q$.

For $q > 1$, the asymptotic behavior of $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ is given by the following lemma.

Lemma 4: The Gaussian coefficient $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ satisfies

$$1 < q^{-\ell(n-\ell)} \begin{bmatrix} n \\ \ell \end{bmatrix}_q < 4$$

for $0 < \ell < n$, so that we may write $\begin{bmatrix} n \\ \lambda n \end{bmatrix}_q = \Theta(q^{n^2 \lambda (1-\lambda)}), 0 < \lambda < 1$.

Proof: The quantity $q^{\ell(n-\ell)}$ may be interpreted as the number of ℓ-dimensional subspaces of $\mathbb{F}_q^n$ that occur as the row space of a matrix of the form $[I | A]$, where I is an $\ell \times \ell$ identity matrix and A is an arbitrary $\ell \times (n - \ell)$ matrix over $\mathbb{F}_q$. (This is the number of codewords in the code $\mathcal{C}$ of Example 1 with $N = n$.) Since $\ell > 0$, this set does not contain all ℓ-dimensional subspaces of $\mathbb{F}_q^n$ and hence the left-hand inequality results. For the right-hand inequality we observe that $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ may be written as

$$\begin{bmatrix} n \\ \ell \end{bmatrix}_q = q^{\ell(n-\ell)} \frac{(1 - q^{-n})(1 - q^{-n+1}) \ldots (1 - q^{-n+\ell-1})}{(1 - q^{-\ell})(1 - q^{-\ell+1}) \ldots (1 - q^{-1})}$$
$$< q^{\ell(n-\ell)} \frac{1}{(1 - q^{-\ell})(1 - q^{-\ell+1}) \ldots (1 - q^{-1})}$$
$$< q^{\ell(n-\ell)} \prod_{j=1}^{\infty} \frac{1}{(1 - q^{-j})}.$$

The function $f(x) = \prod_{j=1}^{\infty} \frac{1}{(1 - x^j)}$ is the generating function of integer partitions [22, Ch. 15] which is increasing in x. As we are interested in $f(1/q)$ for $q \geq 2$, we find that

$$\prod_{j=1}^{\infty} \frac{1}{1 - q^{-j}} \leq \prod_{j=1}^{\infty} \frac{1}{1 - 2^{-j}} = 1/Q_0 < 4$$

where $Q_0 \approx 0.288788095$ is a probabilistic combinatorial constant (see, e.g., [23]) that gives the probability that a large, randomly chosen square binary matrix over $\mathbb{F}_2$ is nonsingular. $\qquad\square$

We remarked earlier that the Grassmann graph constitutes an association scheme, which lets us use simple geometric arguments to give the standard sphere-packing upper bounds and sphere-covering lower bounds. In order to establish the bounds we need the notion of a sphere.

Definition 4: Let W be an N–dimensional vector space and let $\mathcal{P}(W, \ell)$ be the set of ℓ dimensional subspaces of W. The sphere $S(V, \ell, t)$ of radius t centered at a space V in $\mathcal{P}(W, \ell)$ is defined as the set of all subspaces U that satisfy $d(U, V) \leq 2t$

$$S(V, \ell, t) = \{U \in \mathcal{P}(W, \ell) : d(U, V) \leq 2t\}. \qquad \square$$

Note that we prefer to define the radius in terms of the graph distance in the Grassmann graph. The radius can therefore take on any nonnegative integer value.

Theorem 5: The number of spaces in $S(V, \ell, t)$ is independent of V and equals

$$|S(V, \ell, t)| = \sum_{i=0}^{t} q^{i^2} \begin{bmatrix} \ell \\ i \end{bmatrix} \begin{bmatrix} N - \ell \\ i \end{bmatrix}$$

for $t \leq \ell$.

Proof: The claim that $S(V, \ell, t)$ is independent of V follows from the fact that $\mathcal{P}(W, \ell)$ constitutes a distance regular graph [21]. We give an expression for the number of spaces U that intersect V in an $\ell - i$-dimensional subspace. We can choose the $\ell - i$-dimensional subspace of intersection in $\begin{bmatrix} \ell \\ \ell - i \end{bmatrix} = \begin{bmatrix} \ell \\ i \end{bmatrix}$ ways. Once this is done we can complete the subspace in

$$\frac{(q^N - q^\ell)(q^N - q^{\ell+1})\dots(q^N - q^{\ell+i-1})}{(q^\ell - q^{\ell-i})(q^\ell - q^{\ell-i+1})\dots(q^\ell - q^{\ell-1})} = q^{i^2} \begin{bmatrix} N - \ell \\ i \end{bmatrix}$$

ways. Thus, the cardinality of a shell of spaces at distance $2i$ around V equals $q^{i^2} \begin{bmatrix} N - \ell \\ i \end{bmatrix} \begin{bmatrix} \ell \\ i \end{bmatrix}$. Summing the cardinality of the shells gives the theorem. $\qquad \square$

Note that $|S(V, \ell, t)| = |S(V, N - \ell, t)|$, as expected from (4).

B. Sphere-Packing and Sphere-Covering Bounds

We now can simply state the sphere-packing and sphere-covering bounds as follows.

Theorem 6: Let $\mathcal{C}$ be a collection of spaces in $\mathcal{P}(W, \ell)$ such that $D(\mathcal{C}) \geq 2t$, and let $s = \lfloor \frac{t-1}{2} \rfloor$. The size of $\mathcal{C}$ must satisfy

$$|\mathcal{C}| \leq \frac{|\mathcal{P}(W, \ell)|}{|S(V, \ell, s)|} = \frac{\begin{bmatrix} N \\ \ell \end{bmatrix}}{|S(V, \ell, s)|} < \frac{\begin{bmatrix} N \\ \ell \end{bmatrix}}{q^{s^2} \begin{bmatrix} \ell \\ s \end{bmatrix} \begin{bmatrix} N - \ell \\ s \end{bmatrix}}$$

$$< 4q^{(\ell-s)(N-s-\ell)}.$$

Conversely, there exists a code $\mathcal{C}'$ with distance $D(\mathcal{C}') \geq 2t$ such that $|\mathcal{C}'|$ is lower-bounded by

$$|\mathcal{C}| \geq \frac{|\mathcal{P}(W, \ell)|}{|S(V, \ell, t-1)|} = \frac{\begin{bmatrix} N \\ \ell \end{bmatrix}}{|S(V, \ell, t-1)|}$$

$$> \frac{\begin{bmatrix} N \\ \ell \end{bmatrix}}{(t-1)q^{(t-1)^2} \begin{bmatrix} \ell \\ t \end{bmatrix} \begin{bmatrix} N - \ell \\ t - 1 \end{bmatrix}}$$

$$> \frac{1}{16t} q^{(\ell-t+1)(N-t-\ell+1)}.$$

Proof: Given the expression for the size of a sphere in $\mathcal{P}(W, \ell)$, the upper and lower bounds are just the familiar packing and covering bounds for codes in distance regular graphs. $\qquad \square$

Again we note, as a consequence of (4), that these bounds are symmetric in ℓ and $N - \ell$.

We can express the bounds of Theorem 6 in terms of normalized parameters.

Corollary 7: Let $\mathcal{C}$ be a collection of spaces in $\mathcal{P}(W, \ell)$ with normalized minimum distance $\delta = \frac{D(\mathcal{C})}{2\ell}$. The rate of $\mathcal{C}$ is bounded from above by

$$R \leq (1 - \delta/2)(1 - \lambda(1 + \delta/2)) + o(1)$$

where $o(1)$ approaches zero as N grows. Conversely, there exists a code $\mathcal{C}'$ with normalized distance δ such that the rate of $\mathcal{C}'$ is lower-bounded as

$$R \geq (1 - \delta)(1 - \lambda(\delta + 1)) + o(1)$$

where again $o(1)$ approaches zero as N grows. $\qquad \square$

As in the case of the Hamming scheme, the upper bound is not very good, especially since it easily can be seen that δ cannot be larger than one. We next derive a Singleton-type bound for packings in the Grassmann graph.

C. Singleton Bound

We begin by defining a suitable puncturing operation on codes. Suppose $\mathcal{C}$ is a collection of spaces in $\mathcal{P}(W, \ell)$, where W has dimension N. Let W' be any subspace of W of dimension $N - 1$. A punctured code $\mathcal{C}'$ is obtained from $\mathcal{C}$ by replacing each space $V \in \mathcal{C}$ by $V' = \mathcal{H}_{\ell-1}(V \cap W')$ where $\mathcal{H}_{\ell-1}$ denotes the erasure operator defined earlier. In other words, V is replaced by $V \cap W'$ if $V \cap W'$ has dimension $\ell - 1$; otherwise, V is replaced by some $(\ell - 1)$-dimensional subspace of V. Although this puncturing operation does not in general result in a unique code, we denote any such punctured code as $\mathcal{C}|_{W'}$.

We have the following theorem.

Theorem 8: If $\mathcal{C} \subseteq \mathcal{P}(W, \ell)$ is a code of type $[N, \ell, \log_q |\mathcal{C}|, D]$ with $D > 2$ and W' is an $(N - 1)$-dimensional subspace of W, then $\mathcal{C}' = \mathcal{C}|_{W'}$ is a code of type $[N - 1, \ell - 1, \log_q |\mathcal{C}|, D']$ with $D' \geq D - 2$.

Proof: Only the cardinality and the minimum distance of $\mathcal{C}'$ are in question. We first verify that $D' \geq D - 2$. Let U and V be two codewords of $\mathcal{C}$, and suppose that $U' = \mathcal{H}_{\ell-1}(U \cap W')$ and $V' = \mathcal{H}_{\ell-1}(V \cap W')$ are the corresponding codewords in $\mathcal{C}'$. Since $U' \subseteq U$ and $V' \subseteq V$, we have $U' \cap V' \subseteq U \cap V$, so that

$$2 \dim(U' \cap V') \leq 2 \dim(U \cap V) \leq 2\ell - D$$

where the latter inequality follows from the property that $d(U, V) = 2\ell - 2 \dim(U \cap V) \geq D$. Now in $\mathcal{C}'$ we have

$$\begin{aligned} d(U', V') &= \dim(U') + \dim(V') - 2 \dim(U' \cap V') \\ &= 2(\ell - 1) - 2 \dim(U' \cap V') \\ &\geq 2\ell - 2 - (2\ell - D) \\ &= D - 2. \end{aligned}$$
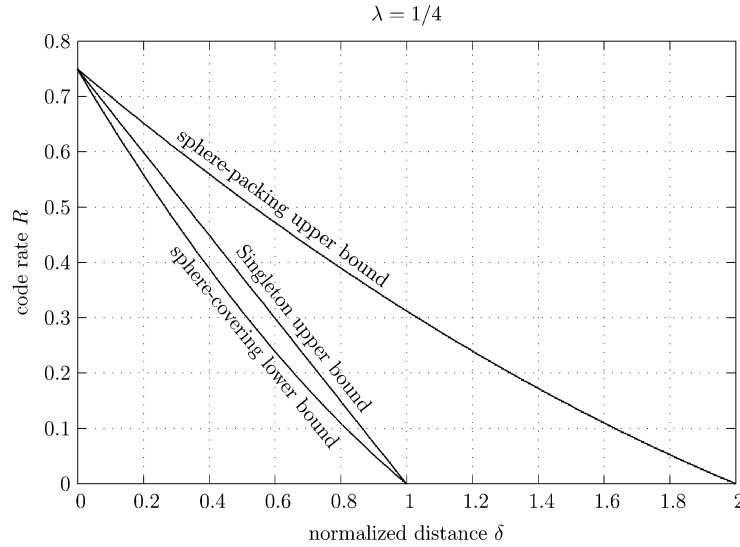
Fig. 1. Upper and lower asymptotic bounds on the largest rate of a code in the Grassmann graph $G_{W,\ell}$ where the dimension N of ambient vector space is asymptotically large and $\lambda = \frac{\ell}{N}$ is chosen as $1/4$.

Since $D > 2, d(U', V') > 0$, so U' and V' are distinct, which shows that C' has as many codewords as C. □

We may now state the Singleton bound.

Theorem 9: A q-ary code of $\mathcal{C} \subseteq \mathcal{P}(W, \ell)$ of type $[N, \ell, \log_q |\mathcal{C}|, D]$ must satisfy

$$|\mathcal{C}| \leq \begin{bmatrix} N - (D-2)/2 \\ \max\{\ell, N - \ell\} \end{bmatrix}_q.$$

Proof: If $\mathcal{C}$ is punctured a total of $(D-2)/2$ times, a code $\mathcal{C}'$ of type $[N - (D-2)/2, \ell - (D-2)/2, \log_q |C|, D']$ is obtained, with every codeword having dimension $\ell - (D-2)/2$ and with $D' \geq 2$. Such a code cannot have more codewords than the corresponding Grassmannian, which contains

$$a = \begin{bmatrix} N - (D-2)/2 \\ \ell - (D-2)/2 \end{bmatrix}_q = \begin{bmatrix} N - (D-2)/2 \\ N - \ell \end{bmatrix}_q$$

points. Applying the same argument to $\mathcal{C}^{\perp}$ yields the upper bound $b = \begin{bmatrix} N-(D-2)/2 \\ \ell \end{bmatrix}_q$. Now $a < b$ if and only if $\ell < N - \ell$, from which the bound follows. □

This bound is easily expressed in terms of normalized parameters. We consider only the case where $\ell \leq N - \ell$, i.e., $\lambda \leq 1/2$.

Corollary 10: Let $\mathcal{C}$ be a collection of spaces in $\mathcal{P}(W, \ell)$, with $\ell \leq \dim(W)/2$ and with normalized minimum distance $\delta = \frac{D(\bar{C})}{2\ell}$. The rate of $\mathcal{C}$ is bounded from above by

$$R \leq (1 - \delta)(1 - \lambda) + \frac{1}{\lambda N}(1 - \lambda + o(1)).$$

The three bounds are depicted in Fig. 1, for $\lambda = 1/4$ and in the limit as $N \to \infty$.

V. A REED–SOLOMON-LIKE CODE CONSTRUCTION AND DECODING ALGORITHM

We now turn to the problem of constructing a code capable of correcting errors and erasures at the output of the operator channels defined in Section II. The code construction is equivalent to

that given by Wang, Xing, and Safavi-Naini [17] in the context of linear authentication codes, which in turn can be regarded as an application of the maximum rank-distance construction of Gabidulin [18]. (The connection between constant-dimension codes and rank-metric codes and the generalized rank-metric decoding problem induced by the operator channel is studied in detail in [20].) The main contribution of this section is a Sudan-style "list-1" minimum-distance decoding algorithm, given in Section V-C.

A. Linearized Polynomials

Let $\mathbb{F}_q$ be a finite field and let $\mathbb{F} = \mathbb{F}_{q^m}$ be an extension field. Recall from [24, Ch. 11], [25, Sec. 3.4] that a polynomial $L(x)$ is called a *linearized polynomial* over $\mathbb{F}$ if it takes the form

$$L(x) = \sum_{i=0}^{d} a_i x^{q^i} \tag{6}$$

with coefficients $a_i \in \mathbb{F}, i = 0, \dots, d$. If all coefficients are zero, so that $L(x)$ is the zero polynomial, we will write $L(x) \equiv 0$; more generally, we will write $L_1(x) \equiv L_2(x)$ if $L_1(x) - L_2(x) \equiv 0$. When q is fixed under discussion, we will let $x^{[i]}$ denote x^{q^i}. In this notation, a linearized polynomial over $\mathbb{F}$ may be written as

$$L(x) = \sum_{i=0}^{d} a_i x^{[i]}.$$

If $L_1(x)$ and $L_2(x)$ are linearized polynomials over $\mathbb{F}$, then so is any $\mathbb{F}$-linear combination $\alpha_1 L_1(x) + \alpha_2 L_2(x), \alpha_1, \alpha_2 \in F$. The ordinary product $L_1(x)L_2(x)$ is not necessarily a linearized polynomial. However, the composition $L_1(L_2(x))$, often written as $L_1(x) \otimes L_2(x)$, of two linearized polynomials over $\mathbb{F}$ is again a linearized polynomial over $\mathbb{F}$. Note that this operation is not commutative, i.e., $L_1(x) \otimes L_2(x)$ need not be equal to $L_2(x) \otimes L_1(x)$.

The product $L_1(x) \otimes L_2(x)$ of linearized polynomials is computed explicitly as follows. If $L_1(x) = \sum_{i \geq 0} a_i x^{[i]}$ and $L_2(x) = \sum_{j \geq 0} b_j x^{[j]}$, then

$$L_1(x) \otimes L_2(x) = L_1(L_2(x)) = \sum_{i \geq 0} a_i (L_2(x))^{[i]}$$

$$= \sum_{i \geq 0} a_i \left(\sum_{j \geq 0} b_j x^{[j]} \right)^{[i]}$$

$$= \sum_{i \geq 0} \sum_{j \geq 0} a_i b_j^{[i]} x^{[i+j]} = \sum_{k \geq 0} c_k x^{[k]}$$

where

$$c_k = \sum_{i=0}^{k} a_i b_{k-i}^{[i]}.$$

Thus, the coefficients of $L_1(x) \otimes L_2(x)$ are obtained from those of $L_1(x)$ and $L_2(x)$ via a modified convolution operation. If $L_1(x)$ has degree q^{d_1} and $L_2(x)$ has degree q^{d_2}, then both $L_1(x) \otimes L_2(x)$ and $L_2(x) \otimes L_1(x)$ have degree $q^{d_1+d_2}$.

Under addition $+$ and composition $\otimes$, the set of linearized polynomials over $\mathbb{F}$ forms a noncommutative ring with identity. Although noncommutative, this ring has many of the properties of a Euclidean domain including, for example, an absence of zero divisors. The degree of a nonzero element forms a natural norm. There are two division algorithms: a left division and a right division, i.e., given any two linearized polynomials $a(x)$ and $b(x)$, it is easy to prove by induction that there exist unique linearized polynomials $q_L(x), q_R(x), r_L(x)$ and $r_R(x)$ such that

$$a(x) = q_L(x) \otimes b(x) + r_L(x) = b(x) \otimes q_R(x) + r_R(x),$$

where $r_L(x) \equiv 0$ or $\deg(r_L(x)) < \deg(b(x))$ and, similarly, where $r_R(x) \equiv 0$ or $\deg(r_R(x)) < \deg(b(x))$.

The polynomials $q_R(x)$ and $r_R(x)$ are easily determined by the following straightforward variation of ordinary polynomial long division. Let $\mathrm{lc}(a(x))$ denote the leading coefficient of $a(x)$, so that if $a(x)$ has degree q^d, i.e., $a(x) = a_d x^{[d]} + a_{d-1} x^{[d-1]} + \cdots + a_0 x^{[0]}$ with $a_d \neq 0$, then $\mathrm{lc}(a(x)) = a_d$.

procedure RDiv$(a(x), b(x))$

    input: a pair $a(x), b(x)$ of linearized polynomials over $\mathbb{F} = \mathbb{F}_{q^m}$, with $b(x) \not\equiv 0$.

    output: a pair $q(x), r(x)$ of linearized polynomials over $\mathbb{F}_{q^m}$

  begin

    if $\deg(a(x)) < \deg(b(x))$ **then**

        return $(0, a(x))$

    else

        $d := \deg(a(x)), e := \deg(b(x)), a_d := \mathrm{lc}(a(x))$, $b_e := \mathrm{lc}(b(x))$

        $t(x) := (a_d/b_e)^{[m-e]} x^{[d-e]}$ $(*)$

        return $(t(x), 0) + \mathbf{RDiv}\,(a(x) - b(x) \otimes t(x), b(x))$
        $(**)$

    endif

  end

Note that the parameter m in step $(*)$ is equal to the dimension of $\mathbb{F}_{q^m}$ as a vector space over $\mathbb{F}_q$. This algorithm terminates when it produces polynomials $q(x)$ and $r(x)$ with the property that $a(x) = b(x) \otimes q(x) + r(x)$ and either $r(x) \equiv 0$ or $\deg r(x) < \deg b(x)$.

The left-division procedure is essentially the same; "**RDiv**" is replaced with "**LDiv**" and $(*)$ and $(**)$ are replaced with the following:

    $t(x) := (a_d/(b_e^{[d-e]})) x^{[d-e]}$

    return $(t(x), 0) + \mathbf{LDiv}(a(x) - t(x) \otimes b(x), b(x))$

With this change, the algorithm terminates when it produces polynomials $q(x)$ and $r(x)$ with the property that $a(x) = q(x) \otimes b(x) + r(x)$.

Linearized polynomials receive their name from the following property. Let $L(x)$ be a linearized polynomial over $\mathbb{F}$, and let K be an arbitrary extension field of $\mathbb{F}$. Then K may be regarded as a vector space over $\mathbb{F}_q$. The map taking $\beta \in K$ to $L(\beta) \in K$ is *linear* with respect to $\mathbb{F}_q$, i.e., for all $\beta_1, \beta_2 \in K$ and all $\lambda_1, \lambda_2 \in \mathbb{F}_q$

$$L(\lambda_1 \beta_1 + \lambda_2 \beta_2) = \lambda_1 L(\beta_1) + \lambda_2 L(\beta_2).$$

Suppose that K is chosen to be large enough to include all the zeros of $L(x)$. The zeros of $L(x)$ then correspond to the kernel of $L(x)$ regarded as a linear map, so they form a vector space over $\mathbb{F}_q$. If $L(x)$ has degree q^d, this vector space has dimension at most d, but the dimension could possibly be smaller if $L(x)$ has repeated roots (which occurs if and only if $a_0 = 0$ in (6)).

On the other hand, if V is an n-dimensional subspace of K, then

$$L(x) = \prod_{\beta \in V} (x - \beta)$$

is a monic linearized polynomial over K (though not necessarily over $\mathbb{F}$). See [25, Lemma 21] or [26, Theorem 3.52].

The following lemma shows that if two linearized polynomials of degree at most q^{d-1} agree on at least d linearly independent points, then the two polynomials coincide.

Lemma 11: Let d be a positive integer and let $f(x)$ and $g(x)$ be two linearized polynomials over $\mathbb{F}$ of degree less than q^d. If $\alpha_1, \alpha_2, \ldots, \alpha_d$ are linearly independent elements of K such that we have $f(\alpha_i) = g(\alpha_i)$ for $i = 1, \ldots, d$, then $f(x) \equiv g(x)$.

    Proof: Observe that $h(x) = f(x) - g(x)$ has $\alpha_1, \ldots, \alpha_d$ as zeros, and hence also has all q^d linear combinations of these elements as zeros. Thus, $h(x)$ has at least q^d distinct zeros. However, since the actual degree of $h(x)$ is strictly smaller than q^d, this is only possible if $h(x) \equiv 0$.   □

B. Code Construction

Just as traditional Reed–Solomon codeword components may be obtained via the evaluation of an *ordinary* message polynomial, we obtain here a basis for the transmitted vector space via the evaluation of a *linearized* message polynomial.

Let $\mathbb{F}_q$ be a finite field, and let $\mathbb{F} = \mathbb{F}_{q^m}$ be a (finite) extension field of $\mathbb{F}_q$. As in the previous subsection, we may regard $\mathbb{F}$ as a vector space of dimension m over $\mathbb{F}_q$. Let $A = \{\alpha_1, \ldots, \alpha_\ell\} \subset \mathbb{F}$ be a set of linearly independent elements in this vector space.

These elements span an ℓ-dimensional vector space $\langle A \rangle \subseteq \mathbb{F}$ over $\mathbb{F}_q$. Clearly, $\ell \leq m$. We will take as ambient space the direct sum $W = \langle A \rangle \oplus \mathbb{F} = \{(\alpha, \beta) : \alpha \in \langle A \rangle, \beta \in \mathbb{F}\}$, a vector space of dimension $\ell + m$ over $\mathbb{F}_q$.

Let $u = (u_0, u_1, \ldots, u_{k-1}) \in \mathbb{F}^k$ denote a block of message symbols, consisting of k symbols over $\mathbb{F}$ or, equivalently, mk symbols over $\mathbb{F}_q$. Let $\mathbb{F}^k[x]$ denote the set of linearized polynomials over $\mathbb{F}$ of degree at most q^{k-1}. Let $f(x) \in \mathbb{F}^k[x]$, defined as

$$f(x) = \sum_{i=0}^{k-1} u_i x^{[i]},$$

be the linearized polynomial with coefficients corresponding to u. Finally, let $\beta_i = f(\alpha_i)$. Each pair $(\alpha_i, \beta_i), i = 1, \ldots, \ell$, may be regarded as a vector in W. Since $\{\alpha_1, \ldots, \alpha_\ell\}$ is a linearly independent set, so is $\{(\alpha_1, \beta_1), \ldots, (\alpha_\ell, \beta_\ell)\}$; hence, this set spans an ℓ-dimensional subspace V of W. We denote the map that takes the message polynomial $f(x) \in \mathbb{F}^k[x]$ to the linear space $V \in \mathcal{P}(W, |A|)$ as ev_A.

Lemma 12: if $|A| \geq k$ then the map $\mathrm{ev}_A : \mathbb{F}^k[x] \rightarrow \mathcal{P}(W, |A|)$ is injective.

Proof: Suppose $|A| \geq k$ and $\mathrm{ev}_A(f) = \mathrm{ev}_A(g)$ for some $f(x), g(x) \in \mathbb{F}^k[x]$. Let $h(x) = f(x) - g(x)$. Clearly, $h(\alpha_i) = 0$ for $i = 1, \ldots, \ell$. Since $h(x)$ is a linearized polynomial, it follows that $h(x) = 0$ for all $x \in \langle A \rangle$. Thus, $h(x)$ has at least $q^{|A|} \geq q^k$ zeros, which is only possible (since $h(x)$ has degree at most q^{k-1}) if $h(x) \equiv 0$, so that $f(x) \equiv g(x)$. □

Henceforth, we will assume that $\ell \geq k$. Lemma 12 implies that, provided this condition is satisfied, the image of $\mathbb{F}^k[x]$ is a code $\mathcal{C} \subseteq \mathcal{P}(W, \ell)$ with q^{mk} codewords. The minimum distance of $\mathcal{C}$ is given by the following theorem; however, first we need the following lemma.

Lemma 13: If $\{(\alpha_1, \beta_1), \ldots, (\alpha_r, \beta_r)\} \subseteq W$ is a collection of r linearly independent elements satisfying $\beta_i = f(\alpha_i)$ for some linearized polynomial f over $\mathbb{F}$, then $\{\alpha_1, \ldots, \alpha_r\}$ is a linearly independent set.

Proof: Suppose that for some $\gamma_1, \ldots, \gamma_r \in \mathbb{F}_q$ we have $\sum_{i=1}^r \gamma_i \alpha_i = 0$. Then, in W, we would have

$$\sum_{i=1}^r \gamma_i(\alpha_i, \beta_i) = \left(\sum_{i=1}^r \gamma_i \alpha_i, \sum_{i=1}^r \gamma_i \beta_i\right) = \left(0, \sum_{i=1}^r \gamma_i f(\alpha_i)\right)$$
$$= \left(0, f\left(\sum_{i=1}^r \gamma_i \alpha_i\right)\right) = (0, f(0)) = (0, 0)$$

which is possible (since the (α_i, β_i) pairs are linearly independent) only if $\gamma_1, \ldots, \gamma_r = 0$. □

Theorem 14: Let $\mathcal{C}$ be the image under ev_A of $\mathbb{F}^k[x]$, with $\ell = |A| \geq k$. Then $\mathcal{C}$ is a code of type $[\ell+m, \ell, mk, 2(\ell-k+1)]$.

Proof: Only the minimum distance is in question. Let $f(x)$ and $g(x)$ be distinct elements of $\mathbb{F}^k[x]$, and let $U = \mathrm{ev}_A(f)$ and $V = \mathrm{ev}_A(g)$. Suppose that $U \cap V$ has dimension r. This means it is possible to find r linearly independent elements $(\alpha_1', \beta_1'), \ldots, (\alpha_r', \beta_r')$ such that $f(\alpha_i') = g(\alpha_i') = \beta_i', i = 1, \ldots, r$. By Lemma 13, $\alpha_1', \ldots, \alpha_r'$ are linearly independent and hence they span an r-dimensional space B with the property that $f(b) - g(b) = 0$ for all $b \in B$. If $r \geq k$, then $f(x)$

and $g(x)$ would be two linearized polynomials of degree less than q^k that agree on at least k linearly independent points, and hence by Lemma 11, we would have $f(x) \equiv g(x)$. Since this is not the case, we must have $r \leq k - 1$. Thus

$$d(U, V) = \dim(U) + \dim(V) - 2\dim(U \cap V)$$
$$= 2(\ell - r) \geq 2(\ell - k + 1).$$

It is easy to exhibit two codewords U and V that satisfy this bound with equality. □

The Singleton bound, evaluated for the code parameters of Theorem 14, states that

$$|\mathcal{C}| \leq \begin{bmatrix} N - (D-2)/2 \\ \ell - (D-2)/2 \end{bmatrix}_q = \begin{bmatrix} m + k \\ k \end{bmatrix}_q < 4q^{mk}.$$

This implies that a true Singleton-bound-achieving code could have no more than four times as many codewords as $\mathcal{C}$. When N is large enough, the difference in rate between $\mathcal{C}$ and a Singleton-bound-achieving code becomes negligible. Indeed, in terms of normalized parameters, we have

$$R = (1 - \lambda)\left(1 - \delta + \frac{1}{\lambda N}\right)$$

which certainly has the same asymptotic behavior as the Singleton bound in the limit as $N \rightarrow \infty$. We claim, therefore, that these Reed–Solomon-like codes are nearly Singleton-bound-achieving.

We note also that the traditional network code $\mathcal{C}$ of Example 1, a code of type $[m + \ell, \ell, m\ell, 2]$, is obtained as a special case of these codes by setting $k = \ell$.

This code construction involving the evaluation of linearized polynomials is clearly closely related to the rank-metric code construction of Gabidulin [18]. However, in our setup, the codewords are not arrays, but rather the vector spaces spanned by the rows of the array, and the relevant decoding metric is not the rank metric, but rather the distance measure defined in (3). The connection between subspace codes and rank-metric codes is explored further in [20].

C. Decoding

Suppose now that $V \in \mathcal{C}$ is transmitted over the operator channel described in Section II and that an $(\ell - \rho + t)$-dimensional subspace U of W is received, where $\dim(U \cap V) = \ell - \rho$. In this situation, we have ρ erasures and an error norm of t, and $d(U, V) = \rho + t$. We expect to be able to recover V from U provided that $\rho + t < D/2 = \ell - k + 1$, and we will describe a Sudan-style "list-1" minimum-distance decoding algorithm to do so (see, e.g., [27, Sec. 9.3]). Note that, even if $t = 0$, we require $\rho < \ell - k + 1$, or $\ell - \rho \geq k$, i.e., not surprisingly (given that we are attempting to recover mk information symbols), the receiver must collect enough vectors to span a space of dimension at least k.

Let $r = \ell - \rho + t$ denote the dimension of the received space U, and let $(x_i, y_i), i = 1, \ldots, r$ be a basis for U. At the decoder, we suppose that it is possible to construct a nonzero bivariate polynomial $Q(x, y)$ of the form

$$Q(x, y) = Q_x(x) + Q_y(y),$$

$$\text{such that } Q(x_i, y_i) = 0 \text{ for } i = 1, \ldots, r \quad (7)$$

where $Q_x(x)$ is a linearized polynomial over $\mathbb{F}_{q^m}$ of degree at most $q^{\tau-1}$ and $Q_y(y)$ is a linearized polynomial over $\mathbb{F}_{q^m}$ of degree at most $q^{\tau-k}$. Although $Q(x, y)$ is chosen to interpolate only a basis for U, since $Q(x, y)$ is a linearized polynomial, it follows that in fact $Q(x, y) = 0$ for all $(x, y) \in U$.

We note that (7) defines a homogeneous system of r equations in $2\tau - k + 1$ unknown coefficients. This system has a nonzero solution when it is under-determined, i.e., when

$$r = \ell - \rho + t < 2\tau - k + 1. \tag{8}$$

Since $f(x)$ is a linearized polynomial over $\mathbb{F}_{q^m}$, so is $Q(x, f(x))$, given by

$$Q(x, f(x)) = Q_x(x) + Q_y(f(x)) = Q_x(x) + Q_y(x) \otimes f(x).$$

Since the degree of $f(x)$ is at most q^{k-1}, the degree of $Q(x, f(x))$ is at most $q^{\tau-1}$.

Now let $\{(a_1, b_1), \dots, (a_{\ell-\rho}, b_{\ell-\rho})\}$ be a basis for $U \cap V$. Since all vectors of U are zeros of $Q(x, y)$, we have $Q(a_i, b_i) = 0$ for $i = 1, \dots, \ell - \rho$. However, since $(a_i, b_i) \in V$ we also have $b_i = f(a_i)$ for $i = 1, \dots, \ell - \rho$. In particular

$$Q(a_i, b_i) = Q(a_i, f(a_i)) = 0, \quad i = 1, \dots, \ell - \rho$$

thus, $Q(x, f(x))$ is a linearized polynomial having $a_1, \dots, a_{\ell-\rho}$ as roots. By Lemma 13, these roots are linearly independent. Thus, $Q(x, f(x))$ is a linearized polynomial of degree at most $q^{\tau-1}$ that evaluates to zero on a space of dimension $\ell - \rho$. If the condition

$$\ell - \rho \geq \tau \tag{9}$$

holds, then $Q(x, f(x))$ has more zeros than its degree, which is only possible if $Q(x, f(x)) \equiv 0$. Since in general

$$Q(x, y) = Q_y(y - f(x)) + Q(x, f(x))$$

we have, when $Q(x, f(x)) \equiv 0$,

$$Q(x, y) = Q_y(y - f(x))$$

and so we may hope to extract $y - f(x)$ from $Q(x, y)$. Equivalently, we may hope to find $f(x)$ from the equation

$$Q_y(x) \otimes f(x) + Q_x(x) \equiv 0. \tag{10}$$

However, this equation is easily solved using the **RDiv** procedure described in Section V-A, with $a(x) = -Q_x(x)$ and $b(x) = Q_y(x)$. Alternatively, we can expand (10) into a system of equations involving the unknown coefficients of $f(x)$; this system is readily solved recursively (i.e., via back-substitution).

In summary, to find nonzero $Q(x, y)$ we must satisfy (8) and to ensure that $Q(x, f(x)) \equiv 0$ we must satisfy (9). When both (8) and (9) hold for some τ, we say that the received space U is *decodable*.

Suppose that the received space U is decodable. Substituting (9) into (8), we obtain the condition $\ell - \rho + t < 2(\ell - \rho) - k + 1$ or equivalently

$$\rho + t < \ell - k + 1 \tag{11}$$

i.e., not surprisingly decodability implies (11).

Conversely, suppose (11) is satisfied. From (11) we get $\ell - \rho \geq t + k$, or

$$\ell - \rho + t + k = r + k \leq 2(\ell - \rho). \tag{12}$$

By selecting

$$\tau = \left\lceil \frac{r + k}{2} \right\rceil$$

(which is possible to do since the receiver knows both r and k), we satisfy (8). With this choice of τ, and applying condition (12), we see that

$$\tau \leq \ell - \rho + 1/2;$$

however, since ℓ, ρ and τ are integers, we see that (9) is also satisfied. In other words, condition (11) implies decodability, which is precisely what we would have hoped for.

The interpolation polynomial $Q(x, y)$ can be obtained from the r-basis vectors $(x_1, y_1), (x_2, y_2), \dots, (x_r, y_r)$ for U via any method that provides a nonzero solution to the homogeneous system (7). We next describe an efficient algorithm to accomplish this task. This algorithm is closely related to the work of Loidreau [19], who provides a polynomial-reconstruction-based procedure for rank-metric decoding of Gabidulin codes.

Let $f(x, y) = f_x(x) + f_y(y)$ be a bivariate linearized polynomial, which means that both $f_x(x)$ and $f_y(y)$ are linearized polynomials. Let the degree of $f_x(x)$ and $f_y(y)$ be $q^{d_x(f)}$ and $q^{d_y(f)}$, respectively. The $(1, k-1)$-weighted degree of $f(x, y)$ is defined as

$$\deg_{1,k-1}(f(x, y)) := \max\{d_x(f), k - 1 + d_y(f)\}$$

Note that this definition is different from the weighted degree definitions for usual bivariate polynomials. However, it should become more natural by observing that we may write $f(x, y)$ as $f(x, y) = f_x(x) + f_y(x) \otimes y$.

The following adaptation of an algorithm for the interpolation problem in Sudan-type decoding algorithms (see, e.g., [28], [29]) provides an efficient way to find the required bivariate linearized polynomial $Q(x, y)$. Let a vector space U be spanned by r linearly independent points $(x_i, y_i) \in W$.

procedure Interpolate(U)

    input: a basis $(x_i, y_i) \in W, i = 1, \dots, r$, for U

    output: a linearized bivariate polynomial
$$Q(x, y) = Q_x(x) + Q_y(y)$$

    initialization : $f_0(x, y) = x, f_1(x, y) = y$

begin

    for $i = 1$ **to** r **do**

        $\Delta_0 := f_0(x_i, y_i); \Delta_1 := f_1(x_i, y_i)$

        if $\Delta_0 = 0$ **then**

            $f_1(x, y) := f_1^q(x, y) - \Delta_1^{q-1} f_1(x, y)$

        elseif $\Delta_1 = 0$ **then**

            $f_0(x, y) := f_0^q(x, y) - \Delta_0^{q-1} f_0(x, y)$

        else

            if $\deg_{1,k-1}(f_0) \leq \deg_{1,k-1}(f_1)$ **then**

                $f_1(x, y) := \Delta_1 f_0(x, y) - \Delta_0 f_1(x, y)$

                $f_0(x, y) := f_0^q(x, y) - \Delta_0^{q-1} f_0(x, y)$

else

$$f_0(x,y) := \Delta_1 f_0(x,y) - \Delta_0 f_1(x,y)$$

$$f_1(x,y) := f_1^q(x,y) - \Delta_1^{q-1} f_1(x,y)$$

   endif

   endfor

   if $\deg_{1,k-1}(f_1) < \deg_{1,k-1}(f_0)$ **then**

      return $f_1(x,y)$

   else

      return $f_0(x,y)$

   endif

  end

For completeness, we provide a proof of correctness of this algorithm, which mimics the proof in the case of standard bivariate interpolation, finding the ideal of polynomials that vanishes at a given set of points [28], [29].

Define an order $\prec$ on bivariate linearized polynomials as follows: write $f(x,y) \prec g(x,y)$ if

$$\deg_{1,k-1}(f(x,y)) < \deg_{1,k-1}(g(x,y)).$$

In case

$$\deg_{1,k-1}(f(x,y)) = \deg_{1,k-1}(g(x,y))$$

write $f(x,y) \prec g(x,y)$ if

$$d_y(f) + k - 1 < \deg_{1,k-1}(f(x,y))$$

and

$$d_y(g) + k - 1 = \deg_{1,k-1}(g(x,y)).$$

If none of these conditions is true, we say that $f(x,y)$ and $g(x,y)$ are not comparable. While $\prec$ is clearly not a total order on polynomials, it is granular enough for the proof of correctness of procedure **Interpolate**. In particular, $\prec$ gives a total order on monomials and we can, hence, define a leading term $\mathrm{lt}_\prec(f)$ as the maximal monomial (without its coefficient) in f under the order $\prec$.

Lemma 15: Assume that we have two bivariate linearized polynomials $f(x,y)$ and $g(x,y)$ which are not comparable under $\prec$. We can create a linear combination $h(x,y) = f(x,y) + \gamma g(x,y)$ which for a suitably chosen γ yields a polynomial $h(x,y)$ with $h(x,y) \prec f(x,y)$ and $h(x,y) \prec g(x,y)$.

Proof: If $f(x,y)$ and $g(x,y)$ are not comparable then we have $\mathrm{lt}_\prec(f) = \mathrm{lt}_\prec(g)$. Choosing γ as the quotient of the corresponding coefficients of $\mathrm{lt}_\prec(f)$ in f and $\mathrm{lt}_\prec(g)$ in g yields a polynomial $h(x,y)$ such that $\mathrm{lt}_\prec(h) \prec \mathrm{lt}_\prec(f) = \mathrm{lt}_\prec(g)$. $\square$

Let A be a set of r linearly independent points $(x_i, y_i) \in W$. We say that a nonzero polynomial $f(x,y)$ is x-minimal with respect to A if $f(x,y)$ is a minimal polynomial under $\prec$ such that $\mathrm{lt}_\prec(f) = x^{[d_x(f)]}$ and $f(x,y)$ vanishes at all points of A. Similarly, a nonzero polynomial $g(x,y)$ is said to be y-minimal with respect to A if $g(x,y)$ is a minimal polynomial under $\prec$

such that $\mathrm{lt}_\prec(g) = y^{[d_y(g)]}$ and $g(x,y)$ vanishes in all points of A.

Theorem 16: The polynomials $f_0(x,y)$ and $f_1(x,y)$ that are output by procedure **Interpolate** are x-minimal and y-minimal with respect to the given set of r linearly independent points $(x_i, y_i) \in W$.

Proof: First we note that x-minimal and y-minimal polynomials can always be compared under $\prec$ since they have different leading monomials. The proof proceeds by induction. We first verify that the polynomials x and y are x-minimal and y-minimal with respect to the empty set. We thus assume that after j iterations of the interpolation algorithm the polynomials $f_0(x,y)$ and $f_1(x,y)$ are x-minimal and y-minimal with respect to the points $(x_i, y_i), i = 1, 2, \ldots, j$. It is easy to check that the set of polynomials constructed in the next iteration also vanishes at the point (x_{j+1}, y_{j+1}) so this part of the definition of x- and y-minimality with respect to points $(x_i, y_i), i = 1, 2, \ldots, j+1$ will not be a problem.

Assume first the generic case that $\Delta_0 \neq 0$ and $\Delta_1 \neq 0$. Assume that $f_1(x,y) \prec f_0(x,y)$ holds. Let

$$f_0'(x,y) = \Delta_1 f_0(x,y) - \Delta_0 f_1(x,y).$$

In this case, $lt_\prec(f_0'(x,y)) = lt_\prec(f_0(x,y))$ and the x-minimality of $f_0'(x,y)$ follows from the x-minimality of $f_0(x,y)$. Let

$$f_1'(x,y) = f_1^q(x,y) - \Delta_1^{q-1} f_1(x,y).$$

We will show that $f_1'(x,y)$ is y-minimal with respect to points $(x_i, y_i), i = 1, 2, \ldots, j+1$. To this end and in order to arrive at a contradiction assume that $f_1'(x,y)$ is not y-minimal. This would imply that there exists a y-minimal polynomial $f_1''(x,y)$ with respect to points $(x_i, y_i), i = 1, 2, \ldots, j+1$ such that $f_1''(x,y) \neq f_1(x,y)$ which has the same leading term as $f_1(x,y)$. The two polynomials are clearly different since $f_1''(x,y)$ would vanish at (x_{j+1}, y_{j+1}) while $f_1(x,y)$ does not. But this would imply that we can find a polynomial $h(x,y)$ as linear combination of $f_1''(x,y)$ and $f_1(x,y)$ which would precede both $f_0(x,y)$ and $f_1(x,y)$ under the order $\prec$ and which would vanish at all points $(x_i, y_i), i = 1, 2, \ldots, j$, thus contradicting the minimality of $f_0(x,y)$ and $f_1(x,y)$. A virtually identical arguments holds if we have $f_0(x,y) \prec f_1(x,y)$.

Next we consider the case that Δ_0 equals 0 while we have $\Delta_1 \neq 0$. In this case, $f_0(x,y)$ is unchanged and, hence, inherits its x-minimality from the previous iteration. We only have to check that the newly constructed $f_1'(x,y) = f_1^q(x,y) - \Delta_1^{q-1} f_1(x,y)$ is y-minimal with respect to points $(x_i, y_i), i = 1, 2, \ldots, j+1$. Again, assuming the opposite would imply that there exists a polynomial $f_1''(x,y) \neq f_1(x,y)$ with the same leading term as $f_1(x,y)$. The two polynomials are again different since $f_1''(x,y)$ would vanish at (x_{j+1}, y_{j+1}) while $f_1(x,y)$ does not. Again, we form a suitable linear combination $h(x,y)$ of $f_1''(x,y)$ and $f_1(x,y)$ which would precede $f_1(x,y)$ under $\prec$. If the leading term of $h(x,y)$ is of type $y^{[d_y(h)]}$ or $h(x,y) \prec f_0(x,y)$, we have arrived at a contradiction negating the y-minimality of $f_1(x,y)$ or the x-minimality of $f_0(x,y)$. Otherwise, note that $h(x,y)$ does not vanish at (x_{j+1}, y_{j+1}) and hence is not a multiple (under $\otimes$ of $f_0(x,y)$.) Hence, for a suitably chosen t, we can

find a polynomial $h''(x, y)$ as a linear combination of $h(x, y)$ and $x^{[t]} \otimes f_0(x, y)$ which precedes $h(x, y)$. Repeating this procedure we arrive at a polynomial $\hat{h}(x, y)$ which either has a leading term of type $y^{[d_y(\hat{h})]}$ or which precedes $f_0(x, y)$ under $\prec$, either contradicting the y-minimality of $f_1(x, y)$ or the x-minimality of $f_0(x, y)$. Finally, we note that the case $\Delta_0 \neq 0$ and $\Delta_1 = 0$ follows from similar arguments. For the case $\Delta_0 = 0$ and $\Delta_1 = 0$ there is nothing to prove. $\qquad \square$

Based on Theorem 16 we can claim that the **Interpolate** procedure solves the problem—as required in (7)—of finding the bivariate linearized polynomial $Q(x, y)$ of minimal $(1, k-1)$ weighted degree $\tau - 1$, which is identified as the polynomial $f_0(x, y)$ or $f_1(x, y)$ of smaller $(1, k-1)$ weighted degree.

Let $V \in \mathcal{C}$ be transmitted over the operator channel described in Section II and assume that an $(\ell - \rho + t)$-dimensional subspace U of W is received. Decoding comprises the following steps:

1. Invoke **Interpolate**(U) to find a bivariate linearized polynomial $Q(x, y) = Q_x(x) + Q_y(y)$ of minimal $(1, k - 1)$ weighted degree that vanishes on the vector space U.
2. Invoke **RDiv**$(-Q_x(x), Q_y(x))$ to find a linearized polynomial $f(x)$ with the property that $-Q_x(x) \equiv Q_y(x) \otimes f(x)$. If no such polynomial can be found declare "failure."
3. Output $f(x)$ as the information polynomial corresponding the codeword $\hat{V} \in \mathcal{C}$ if $d(U, \hat{V}) < \ell - k + 1$.

The time-complexity of this procedure is dominated by the **Interpolate** step, which requires $O((\ell + m)^2)$ operations in $\mathbb{F}_{q^m}$.

VI. Conclusion

In this paper, we have defined a class of operator channels as the natural transmission models in "noncoherent" random linear network coding. The inputs and outputs of operator channels are subspaces of some given ambient vector space. We have defined a coding metric on these subspaces which gives rise to notions of erasures (dimension reduction) and errors (dimension enlargement). In defining codes, it is natural to consider constant-dimension codes; in this case, the code forms a subset of a finite-field Grassmannian. Sphere-packing and sphere-covering bounds as well as a Singleton-type bound are obtained in this context. Finally, a Reed–Solomon-like code construction (equivalent to the construction of linear authentication codes in [17]) is given, and a Sudan-style "list-1" unique decoding algorithm is described, resulting in codes that are capable of correcting various combinations of errors and erasures.

References

[1] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. 2003 IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 442.

[2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 2003 Allerton Conf. Communications, Control and Computing*, Monticello, IL, Oct. 2003.

[3] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

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

[5] N. Cai and R. W. Yeung, "Network coding and error correction," in *Proc. 2002 IEEE Information Theory Workshop*, Bangalore, India, Oct. 2002, pp. 119–122.

[6] R. W. Yeung and N. Cai, "Network error correction, Part I: Basic concepts and upper bounds," *Comm. Inf. and Systems*, vol. 6, no. 1, pp. 19–36, 2006.

[7] N. Cai and R. W. Yeung, "Network error correction, Part II: Lower bounds," *Commun. Inf. and Systems*, vol. 6, no. 1, pp. 37–54, 2006.

[8] Z. Zhang, "Linear network error correction codes in packet networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 209–219, Jan. 2008.

[9] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard, "Resilient network coding in the presence of Byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.

[10] S. Jaggi and M. Langberg, "Resilient network codes in the presence of eavesdropping Byzantine adversaries," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 541–545.

[11] P. V. Ceccherini, "A q-analogous of the characterization of hypercubes as graphs," *J. Geom.*, vol. 22, pp. 57–74, 1984.

[12] L. Chihara, "On the zeros of the Askey-Wilson polynomials, with applications to coding theory," *SIAM J. Math. Anal.*, vol. 18, no. 1, pp. 191–207, 1987.

[13] W. J. Martin and X. J. Zhu, "Anticodes for the Grassmann and bilinear forms graphs," *Des., Codes Cryptogr.*, vol. 6, pp. 73–79, 1995.

[14] R. Ahlswede, H. K. Aydinian, and L. H. Khachatrian, "On perfect codes and related concepts," *Des., Codes Cryptogr.*, vol. 22, pp. 221–237, 2001.

[15] M. Schwartz and T. Etzion, "Codes and anticodes in the Grassmann graph," in *J. Comb. Theory Ser. A*, 2002, vol. 97, pp. 27–42.

[16] S.-T. Xia and F.-W. Fu, "Johnson type bounds on constant dimension codes," *Des., Codes Cryptogr.*, Available online at arxiv.org/abs/1709.1074, submitted for publication.

[17] H. Wang, C. Xing, and R. Safavi-Naini, "Linear authentication codes: Bounds and constructions," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 866–872, Apr. 2003.

[18] E. M. Gabidulin, "Theory of codes with maximal rank distance," *Probl. Inf. Transm.*, vol. 21, pp. 1–12, Jul. 1985.

[19] P. Loidreau, "A Welch-Berlekamp like algorithm for decoding Gabidulin codes," in *Proc. Int. Workshop on Coding and Cryptography, WCC 2005 (Lecture Notes in Computer Science)*, Ø. Ytrehus, Ed. Berlin, Germany: Springer-Verlag, 2006, vol. 3969.

[20] D. Silva, F. R. Kschischang, and R. Koetter, "A rank-metric approach to error-control in random network coding," *IEEE Trans, Inf. Theory*. [Online]. Available: http://arxiv.org/abs/0711.0708, submitted for publication

[21] A. E. Brouwer, A. M. Cohen, and A. Neumaier, *Distance-Regular Graphs*. New York: Springer-Verlag, 1989.

[22] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[23] E. R. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE*, vol. 68, pp. 564–593, May 1980.

[24] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[25] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. New York: North-Holland, 1977.

[26] R. Lidl and H. Niederreiter, "Finite Fields," in *The Encyclopedia of Mathematics*, G.-C. Rota, Ed. Reading, MA: Addison-Wesley, 1983, vol. 20.

[27] R. M. Roth, *Introduction to Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[28] R. J. McEliece, "The Guruswami-Sudan decoding algorithm for Reed-Solomon codes," Jet Propulsion Lab., Pasadena, CA, Interplanetary Network Progr. Rep. 42-153, 2003. [Online]. Available: www.systems.caltech.edu/EE/Faculty/rjm/papers/RSD-JPL.pdf

[29] R. J. McEliece, B. Wang, and K. Watanabe, "Koetter interpolation over modules," in *Proc. 2005 Allerton Conf. Communications Control and Computing*, Monticello, IL, Oct. 2005.