Coding for Errors and Erasures in Random Network Coding

Ralf Koetter
Institute for Communications Engineering
TU Munich
D-80333 Munich
`ralf.koetter@tum.de`

Frank R. Kschischang
The Edward S. Rogers Sr. Department
of Electrical and Computer Engineering
University of Toronto
`frank@comm.utoronto.ca`

Submitted to *IEEE Transactions on Information Theory*
March 12, 2007

Abstract

The problem of error-control in random 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 random network coding is vector-space preserving, information transmission is modelled 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. We introduce a metric on the space of all subspaces of a fixed vector space, and show that a minimum distance decoder for this metric achieves correct decoding if the dimension of the space $V \cap U$ is large enough. If the dimension of each codeword is restricted to a fixed integer, the code forms a subset of a finite-field Grassmannian. Sphere-packing and sphere-covering bounds as well as 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 provided.

1. 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 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 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 network coding [5–9], 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 or its orthogonal complement.

Just as codes defined over complex Grassmannians 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. The standard, widely advocated approach to random 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 on the Grassmannian, 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 [5–9], is moot in our context. A somewhat more closely related approach is that of [10], which deals with reliable communication in networks with so-called “Byzantine adversaries,” who are are assumed to have some ability to inject packets into the network and also to eavesdrop (i.e., extract packets from the network). 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 [10].

The remainder of this paper is organized as follows.

In Section 2, we introduce the "operator channel" as a concise and convenient abstraction of the channel encountered in random 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 vector space over a finite field $\mathbb{F}_q$. In Section 3, we define a metric on this set that is natural and suitable in the context of random 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 (or its orthogonal complement). 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 codes having codewords all of the same dimension, in which case the code is a subset of the corresponding Grassmannian. In Section 4, 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.

In Section 5 we provide a Reed-Solomon-like code construction that achieves the Singleton bound asymptotically. This construction is closely related to the Gabidulin construction of maximum rank-distance codes [11]. The connection between (certain) codes defined over finite-field Grassmannians and rank-metric codes is explored more fully in [12].

2. 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. Generalizations to multicasts and sets of disjoint unicasts in a network are relatively straightforward and so we will not comment further on these.

To capture the essence of random network coding, recall [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 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.

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, properties of the network such as min-cut between the transmitter and the receiver may influence the joint distribution of the $h_{i,j}$ 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, 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 transmission of M independent packets during the lifetime of one generation[1]. While the possibility may exist to exploit this structure, in the strategy 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 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 particular vector space that is spanned by the rows 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 generated by p. This simple observation is at the heart of the channel models

[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.

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.

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 sub-spaces of W, an object often called the projective geometry of W. The dimension of an element $V \in \mathcal{P}(W)$ is denoted as $\dim(V)$.

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 $\mathcal{P}(W)$. The channel input V and channel output U are related as*

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

where $\mathcal{H}_k$ is an erasure operator, $E \in \mathcal{P}(W)$ is an arbitrary error space, and $\oplus$ denotes the direct sum. If $\dim(V) > k$, the erasure operator $\mathcal{H}_k$ acts to project V onto a randomly chosen k-dimensional subspace of V; otherwise, $\mathcal{H}_k$ leaves V unchanged. If the erasure operator $\mathcal{H}_k$ satisfies $\dim(V) - \dim(\mathcal{H}_k(V)) = \rho$ we say that $\mathcal{H}_k$ corresponds to ρ erasures. The dimension of E is called the error norm $t(E)$ of E.

The direct sum $\mathcal{H}_k(V) \oplus E$ is by definition the set $\{v + e : v \in \mathcal{H}_k(V), e \in E\}$, where we may always assume that E intersects trivially with V, and therefore also with $\mathcal{H}_k(V)$. Indeed, 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 $k' \geqslant k$. 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.

3. Coding for Operator Channels

Definition 1 concisely captures the effect of random 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.

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

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

$$d(A, B) := \dim(A + B) - \dim(A \cap B), \tag{2}$$

where $A + B = \{a + b : a \in A,\ b \in B\}$ denotes the sum of spaces A and B, i.e., the smallest space containing both A and B as subspaces. (We do *not* assume that A and B intersect trivially, hence $A + B$ is not in general equal to the direct sum $A \oplus B$.) Since $\dim(A + B) = \dim(A) + \dim(B) - \dim(A \cap B)$, we may also write

$$\begin{aligned} d(A, B) &= \dim(A) + \dim(B) - 2\dim(A \cap B) \\ &= 2\dim(A + B) - \dim(A) - \dim(B). \end{aligned}$$

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 P(W).

Proof. We need to check that for all subspaces $A, B, X \in \mathcal{P}(W)$ we have: i) $d(A, B) \geqslant 0$ with equality if and only if $A = B$, ii) $d(A, B) = d(B, A)$, and iii) $d(A, B) \leqslant d(A, X) + d(X, B)$. The first two conditions are clearly true and so we focus on the third condition, the triangle inequality. We have

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

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 $A \cap B \cap X \subseteq A \cap B$. ∎

Remark: 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 [13]. 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. ∎

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 elements of $\mathcal{C}$ is denoted by

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

3.2. 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 $V' \in \mathcal{C}$, $d(U, V) \leqslant d(U, V')$.

The importance of the minimum distance $D(\mathcal{C})$ for a code $\mathcal{C} \subset \mathcal{P}(\mathcal{W})$ 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

$$2(t + \rho) < D(\mathcal{C}), \tag{3}$$

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 $d(V, U) \leqslant d(V, V') + d(V', U) \leqslant \rho + t$. If $T \neq V$ is any other codeword in $\mathcal{C}$, then $D(\mathcal{C}) \leqslant d(V, T) \leqslant d(V, U) + d(U, T)$, from which it follows that $d(U, T) \geqslant D(\mathcal{C}) - d(V, U) \geqslant D(\mathcal{C}) - (\rho + t)$. Provided that the inequality (3) holds, then $d(U, T) > d(U, V)$ and hence a minimum distance decoder must produce V. ∎

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 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

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

precisely in parallel to the standard error-correction situation.

3.3. Codes on the Grassmannian

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).

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 2 *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 $d(U, V) = 2$.* ∎

Remark: It is well known that $G_{W,\ell}$ is distance regular [14] 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 4. We also note that the distance measure between two spaces U, V in $\mathcal{P}(W)$ introduced in (2) is equal to twice the graph distance in the Grassmann graph.[2] Codes (particularly the non-existence of perfect codes) in the Grassmann graph have been considered previously in [15–18]. ∎

When modelling the operation of random network coding by the operator channel of Definition 1, there is no further need to specify the precise protocol for random network coding. In particular, we assume the receiver knows that a codeword U 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.

3.4. Code Parameters

Before we study bounds and constructions of codes in $\mathcal{P}(W)$, we need a proper definition of rate. Let $C \subset \mathcal{P}(W)$ be a code. 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. This would correspond to $N\ell$ q-ary symbols if ambient space W is a vector space over $\mathbb{F}_q$. This motivates the following definition.

Definition 3 *Let W be a vector space of dimension N over $\mathbb{F}_q$. Let $\mathcal{C}$ be a subset of $\mathcal{P}(W)$ such that the dimension of any $V \in \mathcal{C}$ is at most ℓ and the minimum distance of $\mathcal{C}$ equals D. We say that $\mathcal{C}$ is a q-ary code of type $[N, \ell, \log_q(|\mathcal{C}|), D]$. The normalized weight λ, rate R and the normalized minimum distance δ of the code are defined as*

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

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

[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.

The parameters λ, R and δ are indeed 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 $[0, 1]$. Just as in constant-weight codes, the interesting range can actually be limited to $[0, \frac{1}{2}]$ as spaces with dimension larger than $\frac{N}{2}$ can be transmitted as the null space of a space of dimension less than $\frac{N}{2}$. 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. Still, 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.

3.5. Examples of Codes

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

Example 1 Let W be the space of vectors over $\mathbb{F}_q$ of length N. 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 of type $[N, \ell, \ell(N - \ell), 2]$ with normalized weight $\lambda = \ell/N$, rate $R = 1 - \lambda$ and normalized distance $\delta = \frac{1}{\lambda N}$. ∎

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

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 code of type $[N, \ell, \log_q |\mathcal{P}(W, \ell)|, 2]$ which is clearly larger than the code $\mathcal{C}$ of Example 1. We will give a precise expression for $|\mathcal{P}(W, \ell)|$ in Section 4. We note that $\mathcal{C}''$, defined as $\mathcal{C}'' = \bigcup_{i=1}^{\ell} \mathcal{P}(W, i)$ is obviously an even bigger code that can be used for random network coding while still using the same resources as $\mathcal{C}$ 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 ℓ is fixed beforehand. ∎

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

4. Bounds on Codes

4.1. Preliminaries

We start this section by introducing some notation that will be relevant for packings in $\mathcal{P}(W)$ where W is a vector space over $\mathbb{F}_q$.

For any non-negative integer i, define

$$[\![i]\!]_q := \begin{cases} 1 & \text{if } i = 0, \\ q^i - 1 & \text{if } i > 0. \end{cases}$$

We also write $[\![i]\!]$ if q is implied from the context. A function $[\![i]\!]!$, mimicking the usual factorial function, is defined as

$$[\![i]\!]! := \prod_{j=0}^{i} [\![j]\!].$$

With this notation we can define the Gaussian coefficient $\begin{bmatrix} n \\ m \end{bmatrix}_q$ as:

$$\begin{bmatrix} n \\ m \end{bmatrix}_q := \begin{cases} \frac{[\![n]\!]_q!}{[\![m]\!]_q! [\![n-m]\!]_q!} & 0 \leqslant m \leqslant n \\ 0 & \text{otherwise} \end{cases}$$

Again we write $\begin{bmatrix} n \\ m \end{bmatrix}$ if q is implied from the context. It is easily verified that $\lim_{q \to 1} \begin{bmatrix} n \\ m \end{bmatrix}_q$ equals the binomial coefficient $\binom{n}{m}$.

The importance of the Gaussian coefficients is reflected in the following proposition.

Theorem 4 *The number of ℓ-dimensional subspaces of an n dimensional vector space over $\mathbb{F}_q$ equals $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$.*

Proof. See, e.g., [19, Ch. 24]. ∎

For $q > 1$ the asymptotic behavior of $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ is given given by the following lemma. We use the notation $f(n) \doteq g(n)$ to mean that

$$\lim_{n \to \infty} \frac{1}{n} \log \left(\frac{f(n)}{g(n)} \right) = 0.$$

Lemma 5 *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 $\begin{bmatrix} n \\ \lambda n \end{bmatrix}_q \doteq q^{n^2 \lambda(1-\lambda)}$, $0 < \lambda < 1$.

Proof. *The left hand inequality is obvious, as $P(W,\ell)$ must contain at least as many subspaces as codewords in the code $\mathcal{C}$ of Example 1. For the right hand inequality we observe that $\begin{bmatrix} n \\ \ell \end{bmatrix}_q$ may be written as*

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

The function $f(x) = \prod_{j=1}^{\infty} \frac{1}{(1-x^j)}$ is the generating function of integer partitions [19, Ch. 15] which is increasing in x. As we are interested in $f(1/q)$ for $q \geqslant 2$, we find that

$$\prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \leqslant \prod_{j=1}^{\infty} \frac{1}{1-2^{-j}} < 4. \quad \blacksquare$$

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 $d(U,V) \leqslant 2t$,*

$$S(V,\ell,t) = \{U \in \mathcal{P}(W,\ell) : d(U,V) \leqslant 2t\}. \quad \blacksquare$$

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 non-negative integer value.

Theorem 6 *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 $t \leqslant \ell$.

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 [14]. 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 $\left[{\ell \atop \ell-i}\right] = \left[{\ell \atop i}\right]$ 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 $q^{i^2}\left[{N-\ell \atop i}\right]\left[{\ell \atop i}\right]$. Summing the cardinality of the shells gives the theorem. ∎

4.2. Sphere-Packing and Sphere-Covering Bounds

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

Theorem 7 *Let $\mathcal{C}$ be a collection of spaces in $\mathcal{P}(W,\ell)$ such that $D(\mathcal{C}) \geqslant 2t$, and let $s = \lfloor \frac{t-1}{2} \rfloor$. The size of $\mathcal{C}$ must satisfy*

$$|\mathcal{C}| \leqslant \frac{|\mathcal{P}(W,\ell)|}{|S(V,\ell,s)|} = \frac{\left[{N \atop \ell}\right]}{|S(V,\ell,s)|} < \frac{\left[{N \atop \ell}\right]}{q^{s^2}\left[{\ell \atop s}\right]\left[{N-\ell \atop s}\right]} < 4q^{(\ell-s)(N-s-\ell)}$$

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

$$|\mathcal{C}| \geqslant \frac{|\mathcal{P}(W,\ell)|}{|S(V,\ell,t-1)|} = \frac{\left[{N \atop \ell}\right]}{|S(V,\ell,t-1)|} > \frac{\left[{N \atop \ell}\right]}{(t-1)q^{(t-1)^2}\left[{\ell \atop t}\right]\left[{N-\ell \atop t-1}\right]} > \frac{1}{16t} q^{(\ell-t+1)(N-t-\ell+1)}$$

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. ∎

We can express the bound of Theorem 7 in terms of normalized parameters.

Corollary 8 *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*

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

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:

$$R \geqslant (1 - \delta)(1 - \lambda(\delta + 1)) + o(1),$$

where again $o(1)$ approaches zero as N grows. ∎

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.

4.3. 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 9 *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 $D' \geqslant D - 2$.*

Proof. Only the cardinality and the minimum distance of $\mathcal{C}'$ are in question. We first verify that $D' \geqslant D-2$. 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 $2\dim(U' \cap V') \leqslant 2\dim(U \cap V) \leqslant 2\ell - D$, where the latter inequality follows from the property that $d(U, V) = 2\ell - 2\dim(U \cap V) \geqslant D$. Now in $\mathcal{C}'$ we have

$$\begin{aligned} d(U', V') &= \dim(U') + \dim(V') - 2\dim(U' \cap V') \\ &= 2(\ell - 1) - 2\dim(U' \cap V') \\ &\geqslant 2\ell - 2 - (2\ell - D) \\ &= D - 2. \end{aligned}$$

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

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

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 $D' \geqslant 2$. Such a code cannot have more codewords than the corresponding Grassmannian. ∎

This bound is easily expressed in terms of normalized parameters.

Corollary 11 *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*

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

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

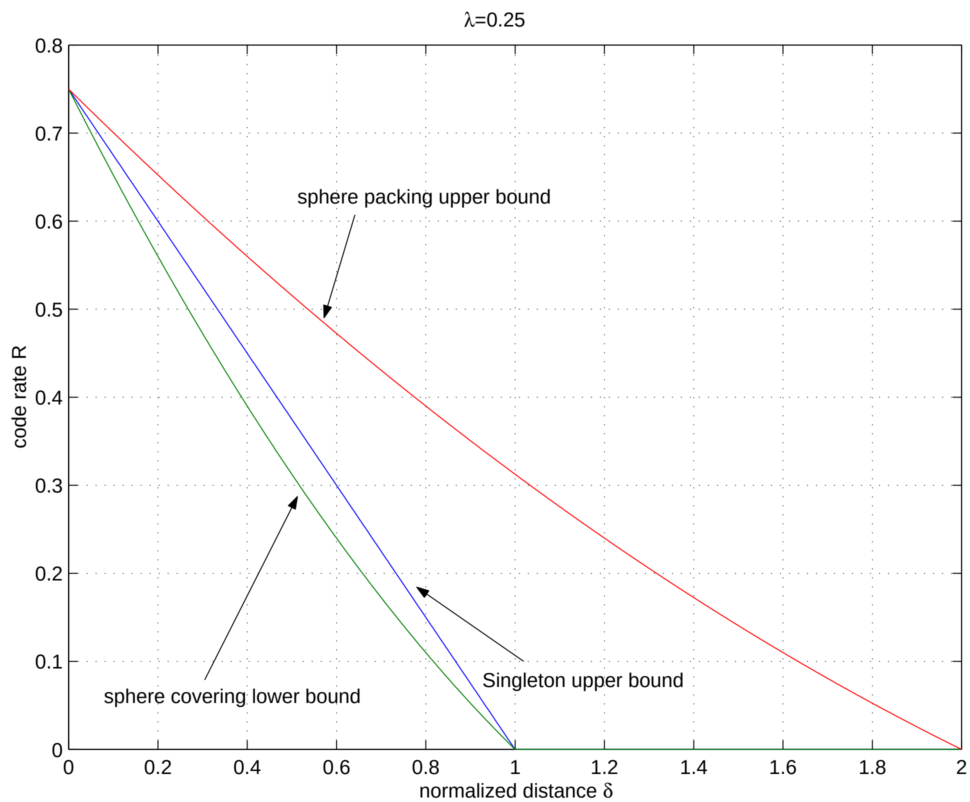

Figure 1: Upper and lower asymptotic bounds on the largest rate of a code in the Grassmann graph $G_{W,\ell}$ where the dimension N of ambient vector space is asymptotically large and $\lambda = \frac{\ell}{N}$ is chosen as $1/4$.

5. A Reed-Solomon-like Code Construction

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 2.

5.1. Linearized Polynomials

Let $\mathbb{F}_q$ be a finite field and let $\mathbb{F} = \mathbb{F}_{q^m}$ be an extension field. Recall from [20, Ch. 11], [21, Sec. 4.9], [22, Sec. 3.4] that a polynomial $L(x)$ is called a *linearized polynomial* over $\mathbb{F}$ if it takes the form

$$L(x) = \sum_{i=0}^{d} a_i x^{q^i}, \tag{4}$$

with coefficients $a_i \in \mathbb{F}, i = 0, \ldots, d$. 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

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

The linearized polynomial $L(x)$ in (4) has *conventional* q*-associate*

$$\ell(x) = \sum_{i=0}^{d} a_i x^{i}.$$

If $L(x)$ is a linearized polynomial, we will refer to the degree of its conventional q-associate as the *associate degree* of $L(x)$. Clearly a linearized polynomial of associate degree d has *actual* degree q^d.

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 $L_1(x) = \sum_{i \geqslant 0} a_i x^{[i]}$ and $L_2(x) = \sum_{j \geqslant 0} b_j x^{[j]}$, then

$$\begin{aligned} L_1(x) \otimes L_2(x) &= L_1(L_2(x) = \sum_{i \geqslant 0} a_i (L_2(x))^{[i]} \\ &= \sum_{i \geqslant 0} a_i \left(\sum_{j \geqslant 0} b_j x^{[j]} \right)^{[i]} \\ &= \sum_{i \geqslant 0} \sum_{j \geqslant 0} a_i b_j^{[i]} x^{[i+j]} = \sum_{k \geqslant 0} c_k x^{[k]} \end{aligned}$$

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 associate degree d_1 and $L_2(x)$ has associate degree d_2, then both $L_1(x) \otimes L_2(x)$ and $L_2(x) \otimes L_1(x)$ have associate degree $d_1 + d_2$.

Under addition $+$ and composition $\otimes$, the set of linearized polynomials over $\mathbb{F}$ forms a non-commutative ring with identity. Although non-commutative, this ring has many of the properties of a Euclidean domain including, for example, an absence of zero-divisors. The degree (or associate 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 associate degree 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 F = F_q^m, with b(x) ≢ 0.
    output: a pair q(x), r(x) of linearized polynomials over 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 := lc(a(x)), b_e := lc(b(x))
        t(x) := (a_d/b_e)^[m−e] x^[d−e]                          (*)
        return (t(x), 0) + RDiv(a(x) − b(x) ⊗ 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) + LDiv(a(x) − t(x) ⊗ 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 $\lambda_1, \lambda_2 \in F_q$,

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

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 $\mathbb{F}_q$. This vector space has dimension equal to at most the associate degree of $L(x)$, but the dimension could possibly be smaller if $L(x)$ has repeated roots (which occurs if and only if $a_0 = 0$ in (4)).

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 [21, Lemma 21] or [22, Theorem 3.52].

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

Lemma 12 *Let d be a positive integer and let $f(x)$ and $g(x)$ be two linearized polynomials over $\mathbb{F}$ of associate degree less than d. If $\alpha_1, \alpha_2, \ldots, \alpha_d$ are linearly independent elements of K such that 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 $h(x) \equiv 0$. ∎

5.2. 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 $\ell \leqslant m$. 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 associate degree at most $k - 1$. 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 (α_i, β_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 13 *If $|A| \geqslant k$ then the map $\mathsf{ev}_A : \mathbb{F}^k[x] \to \mathcal{P}(W, |A|)$ is injective.*

Proof. Suppose $|A| \geqslant k$ and $\mathsf{ev}_A(f) = \mathsf{ev}_A(g)$ 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 $q^{|A|} \geqslant q^k$ zeros, which is only possible (since $h(x)$ has associate degree at most $k - 1$) if $h(x) \equiv 0$, so that $f(x) \equiv g(x)$. ∎

Henceforth we will assume that $\ell \geqslant k$. Lemma 13 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 14 *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

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

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

Theorem 15 *Let $\mathcal{C}$ be the image under ev_A of $\mathbb{F}^k[x]$, with $\ell = |A| \geqslant k$. 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 $U = \mathsf{ev}_A(f)$ and $V = \mathsf{ev}_A(g)$. 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 14, $\alpha'_1, \ldots, \alpha'_r$ 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 $r \geqslant k$ then $f(x)$ and $g(x)$ would be two linearized polynomials of associate degree less than k that agree on at least k linearly independent points, and hence by Lemma 12, we would have $f(x) \equiv g(x)$. Since this is not the case, we must have $r \leqslant k - 1$. Thus

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

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 15, states that

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

This implies that a true Singleton-bound-achieving code could have no more than 4 times as many codewords as $\mathcal{C}$. When N is large enough, the difference in rate between $\mathcal{C}$ and a Singleton-bound-achieving becomes negligible. Indeed, in terms of normalized parameters, we have

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

which certainly has the same asymptotic behavior as the Singleton bound in the limit as $N \to \infty$. 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 [11]. 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 (2).

5.3. Decoding

Suppose now that $V \in \mathcal{C}$ is transmitted over the operator channel described in Section 2 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., [23, Sec. 9.3]). Note that, even if $t = 0$, we require $\rho < \ell - k + 1$, or $\ell - \rho \geqslant k$, 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

$$Q(x, y) = Q_x(x) + Q_y(y), \text{ such that } Q(x_i, y_i) = 0 \text{ for } i = 1, \ldots, r, \tag{5}$$

where $Q_x(x)$ is a linearized polynomial over $\mathbb{F}_{q^m}$ of associate degree at most $\tau - 1$ and $Q_y(y)$ is a linearized polynomial over $\mathbb{F}_{q^m}$ of associate degree at most $\tau - k$. 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 (5) 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

$$r = \ell - \rho + t < 2\tau - k + 1. \tag{6}$$

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 associate degree of $f(x)$ is at most $k - 1$, the associate degree of $Q(x, f(x))$ is at most $\tau - 1$.

Now let $\{(a_1, b_1), \ldots, (a_{\ell-\rho}, b_{\ell-\rho})\}$ 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 $i = 1, \ldots, \ell - \rho$. However, since $(a_i, b_i) \in V$ we also have $b_i = f(a_i)$ for $i = 1, \ldots, \ell - \rho$. In particular,

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

thus $Q(x, f(x))$ is a linearized polynomial having $a_1, \ldots, a_{\ell-\rho}$ as roots. By Lemma 14, these roots are linearly independent. Thus $Q(x, f(x))$ is a linearized polynomial of associate degree at most τ that evaluates to zero on a space of dimension $\ell - \rho$. If the condition

$$\ell - \rho \geqslant \tau \tag{7}$$

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

$$Q_y(x) \otimes f(x) + Q_x(x) \equiv 0. \tag{8}$$

However, this equation is easily solved using the `RDiv` procedure described in Section 5.1, with $a(x) = -Q_x(x)$ and $b(x) = Q_y(x)$. Alternatively, we can expand (8) 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 (6) and to ensure that $Q(x, f(x)) \equiv 0$ we must satisfy (7). When both (6) and (7) hold for some τ, we say that the received space U is *decodable*.

Suppose that the received space U is decodable. Substituting (7) into (6), we obtain the condition $\ell - \rho + t < 2(\ell - \rho) - k + 1$ or, equivalently,

$$\rho + t < \ell - k + 1, \tag{9}$$

i.e., not surprisingly decodability implies (9).

Conversely, suppose (9) is satisfied. From (9) we get $\ell - \rho \geqslant t + k$, or

$$\ell - \rho + t + k = r + k \leqslant 2(\ell - \rho). \tag{10}$$

By selecting

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

(which is possible to do since the receiver knows both r and k), we satisfy (6). With this choice of τ, and applying condition (10), we see that

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

however, since ℓ, ρ and τ are integers, we see that (7) is also satisfied. In other words, condition (9) 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), ..., (x_r, y_r) for U via any method that provides a nonzero solution to the homogeneous system (5). We next describe an efficient algorithm to accomplish this task.

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 associate degree of $f_x(x)$ and $f_y(y)$ be $d_x(f)$ and $d_y(f)$, 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. [24, 25]) 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) ∈ W, i = 1, ..., 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
        Δ_0 := f_0(x_i, y_i); Δ_1 := f_1(x_i, y_i)
        if Δ_0 = 0 then
            f_1(x,y) := f_1^q(x,y) − Δ_1^{q−1} f_1(x,y)
        elseif Δ_1 = 0 then
            f_0(x,y) := f_0^q(x,y) − Δ_0^{q−1} f_0(x,y)
        else
            if deg_{1,k−1}(f_0) ⩽ deg_{1,k−1}(f_1) then
                f_1(x,y) := Δ_1 f_0(x,y) − Δ_0 f_1(x,y)
                f_0(x,y) := f_0^q(x,y) − Δ_0^{q−1} f_0(x,y)
            else
                f_0(x,y) := Δ_1 f_0(x,y) − Δ_0 f_1(x,y)
                f_1(x,y) := f_1^q(x,y) − Δ_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_0(x,y)
    else
        return f_1(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 [24, 25].

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 16 *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 $\mathrm{lt}_\prec(h) \prec \mathrm{lt}_\prec(f) = \mathrm{lt}_\prec(g)$. ∎

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 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 17 *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 $f_1''(x,y)$ w.r.t. 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), $i = 1, 2, \ldots, j+1$, 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. ∎

Based on Theorem 17 we can claim that the `Interpolate` procedure solves the problem—as required in equation (5)—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 2 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 $V \in \mathcal{C}$ if $d(U,V) < \ell - k + 1$.

6. Conclusions

In this paper we have defined a class of operator channels as the natural transmission models in "noncoherent" random 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 restrict each codeword to have some fixed dimension; 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 is given, resulting in codes that are capable of correcting various combinations of errors and erasures.

Acknowledgment

The authors thank Danilo Silva for useful discussions.

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. on Inform. Theory*, (Yokohama), p. 442, June 29 – July 4, 2003.

[2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 2003 Allerton Conf. on Commun., 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. on Inform. Theory*, vol. 52, 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. on Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[5] N. Cai and R. W. Yeung, "Network coding and error correction," in *Proc. 2002 IEEE Inform. Theory Workshop*, pp. 119–122, Oct. 20–25, 2002.

[6] L. Song, R. W. Yeung, and N. Cai, "Zero-error network coding for acyclic networks," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 3129–3139, Dec. 2003.

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

[8] R. W. Yeung and N. Cai, "Network error correction, part II: Lower bounds," *Comm. in Inform. and Systems*, vol. 6, no. 1, pp. 37–54, 2006.

[9] Z. Zhang, "Network error correction coding in packetized networks," in *Proc. 2006 IEEE Inform. Theory Workshop*, Oct. 22–26, 2006.

[10] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard, "Resilient network coding in the presence of Byzantine adversaries," in *Proc. 26th Annual IEEE Conf. on Computer Commun., INFOCOM*, (Anchorage, AK), May 6–12, 2007. (To appear.).

[11] E. M. Gabidulin, "Theory of codes with maximal rank distance," *Problems of Information Transmission*, vol. 21, pp. 1–12, July 1985.

[12] D. Silva and F. R. Kschischang, "Using rank-metric codes for error correction in random network coding." (Submitted to the 2007 IEEE International Symposium on Information Theory.), 2007.

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

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

[15] 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.

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

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

[18] M. Schwartz and T. Etzion, "Codes and anticodes in the Grassmann graph," *J. of Combin. Theory Series A*, vol. 97, pp. 27–42, 2002.

[19] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*. Cambridge, UK: Cambridge University Press, second ed., 2001.

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

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

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

[23] R. M. Roth, *Introduction to Coding Theory*. Cambridge, UK: Cambridge University Press, 2006.

[24] R. J. McEliece, "The Guruswami-Sudan decoding algorithm for Reed-Solomon codes," tech. rep., JPL Interplanetary Network Progress Report 42-153, 2003. Available online at `www.systems.caltech.edu/EE/Faculty/rjm/papers/RSD-JPL.pdf`.

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