Exact-Repair MDS Codes for Distributed Storage Using Interference Alignment

Changho Suh
Wireless Foundations
University of California at Berkeley
Email: chsuh@eecs.berkeley.edu

Kannan Ramchandran
Wireless Foundations
University of California at Berkeley
Email: kannanr@eecs.berkeley.edu

***Abstract*—The high repair cost of (n,k) Maximum Distance Separable (MDS) erasure codes has recently motivated a new class of codes, called Regenerating Codes, that optimally trade off storage cost for repair bandwidth. In this paper, we address bandwidth-optimal (n,k,d) Exact-Repair MDS codes, which allow for any failed node to be repaired exactly with access to arbitrary d survivor nodes, where $k \leq d \leq n-1$. Under *scalar-linear* codes which do not permit symbol-splitting, we construct Exact-Repair MDS codes that are optimal in repair bandwidth for the case of $k/n \leq 1/2$ and $d \geq 2k-1$. Our codes are deterministic and require a finite-field size of at most $2(n-k)$. Under *vector-linear* codes which allow for the break-up of stored symbols into arbitrarily small subsymbols, we show the existence of optimal Exact-Repair codes for the entire admissible range of possible (n,k,d), i.e., $k < n$ and $k \leq d \leq n-1$. That is, we establish the existence of vector-linear Exact-Repair MDS codes that match the fundamental cutset lower bound. Our approach for both the constructive scalar-linear code design and for the existence of vector-linear codes is based on interference alignment techniques.**

I. Introduction

In distributed storage systems, Maximum Distance Separable (MDS) erasure codes are well-known coding schemes that can offer maximum reliability for a given storage overhead. For an (n,k) MDS code for storage, a source file of size $\mathcal{M}$ bits is divided equally into k units (of size $\frac{\mathcal{M}}{k}$ bits each), and these k data units are expanded into n encoded units, and stored at n nodes. The code guarantees that a user or Data Collector (DC) can reconstruct the source file by connecting to any arbitrary k nodes. While MDS codes are optimal in terms of reliability versus storage overhead, they come with a significant maintenance overhead when it comes to repairing failed encoded nodes.

This challenge has motivated a new class of coding schemes, called Regenerating Codes [1], [2], which target the optimal tradeoff between storage cost and repair bandwidth. On one end of this spectrum of Regenerating Codes are Minimum Storage Regenerating codes that can match the minimum storage cost of MDS codes while also significantly reducing repair bandwidth. In [1], [2], the fundamental tradeoff between storage cost α and repair bandwidth γ was characterized by

$$(\alpha,\gamma) = \left(\frac{\mathcal{M}}{k}, \frac{\mathcal{M}}{k}\cdot\frac{d}{d-k+1}\right), \quad (1)$$

where d denotes the number of nodes that are connected to repair a failed node, simply called the degee d where $k \leq d \leq n-1$. Note that this code requires the same minimal storage cost as that of conventional MDS codes, while substantially reducing repair bandwidth by a factor of $\frac{k(d-k+1)}{d}$ (e.g., for $(n,k,d) = (31,6,30)$, there is a 5x bandwidth reduction). In this paper, we call this code Repair MDS code.

While Repair MDS codes enjoy substantial benefits over conventional MDS codes, they come with some limitations in construction. Specifically, the achievable schemes in [1], [2] restore failed nodes in a *functional* manner only, using a random-network-coding based framework. This means that the replacement nodes maintain the MDS-code property but do not *exactly* replicate the information content of the failed nodes. In many settings of interest, it is important to have exact repair of failed nodes. This forms the subject of this paper.

Specifically, we ask the following question: is there a price for attaining the optimal tradeoff of (1) with the extra constraint of exact repair? The work in [3] sheds some light on this question. Specifically, it was shown that under *scalar linear* codes[1], when $\frac{k}{n} > \frac{1}{2} + \frac{2}{n}$, there is a price for exact repair. For large n, this case boils down to $\frac{k}{n} > \frac{1}{2}$. Now what about for $\frac{k}{n} \leq \frac{1}{2}$?

The first contribution of this paper is to resolve this open problem by showing that scalar-linear Exact-Repair MDS codes come with no extra cost over the optimal tradeoff of (1) for the case of $\frac{k}{n} \leq \frac{1}{2}$ and $d \geq 2k-1$. Our codes are deterministic and require a field size of at most $2(n-k)$. Our result draws its inspiration from the work in [3], which guarantees exact repair of systematic node, while satisfying the MDS code property, but which does not provide exact repair of failed parity nodes. In providing a constructive solution for the exact repair of *all* nodes, we use geometric insights to propose a large family of repair codes. This both provides insights into the structure of codes for exact repair of all nodes, as well as opens up a rich design space for constructive solutions. This will be explained in Section III.

This research was funded in part by an AFOSR grant (FA9550-09-1-0120), a DTRA grant (HDTRA1-09-1-0032), and an NSF grant (CCF-0830788).

[1]In scalar linear codes, symbols are not allowed to be split into arbitrarily small subsymbols.

The second contribution is to establish the following fact. Under *vector linear* codes which allow for the break-up of stored symbols into arbitrarily small subsymbols, we show the existence of Exact-Repair MDS codes that achieve the optimal tradoff of (1) for the entire admissible spectrum of (n, k, d), i.e., $k < n$ and $k \leq d \leq n-1$.[2] That is we show that there is no theoretical gap between exact repair and functional repair codes for the entire range of (n, k, d). This will be explained in Section IV.

Our results for both constructive scalar-linear codes and vector-linear codes build on the concept of *interference alignment*, which was introduced in the context of wireless communication networks [5], [6].

Due to space constraints, we will provide only a terse description of our proposed constructions and refer the reader to [7] for details related to constructive scalar linear Exact-Repair codes for $\frac{k}{n} \leq \frac{1}{2}$ and $d \geq 2k-1$, and to [8] for details related to the existence of vector linear Exact-Repair codes for the entire admissible spectrum of (n, k, d).

II. Interference Alignment for Repair Codes

Our achievable scheme builds on on the concept of interference alignment. The idea of interference alignment is to align multiple interference signals in a signal subspace whose dimension is smaller than the number of interferers [5], [6], [9]. This concept relates intimately to our repair problem that involves recovery of a subset (related to the subspace spanned by a failed node) of the overall aggregate signal space (related to the entire user data dimension). This attribute was first observed in [10], where it was shown that the interference alignment concept could be exploited for Exact-Repair MDS codes having small k ($k = 2$). However, generalizing interference alignment to large values k (even $k = 3$) proves to be challenging, as we describe in the sequel. In order to appreciate this better, let us first review the scheme of [10] that was applied to the exact repair problem. We will then address the difficulty of extending interference alignment for larger systems and describes how to address this in Section III.

Review of $(4, 2)$ Exact-Repair MDS Codes [10]: We consider the degree $d = 3$. In this paper, we normalize the repair-bandwidth-per-link ($\frac{\gamma}{d}$) to be 1, making $\mathcal{M} = k(d-k+1) = 4$. One can partition a whole file into smaller chunks so that each has a size of $k(d-k+1)$. The optimal tradeoff point (1) then gives storage cost $\alpha = 2$.

Fig. 1 illustrates interference alignment for exact repair of failed node 1. We introduce matrix notation for illustrative purpose. Let $\mathbf{a} = (a_1, a_2)^t$ and $\mathbf{b} = (b_1, b_2)^t$ be 2-dimensional information-unit vectors, where $(\cdot)^t$ indicates a transpose. Let $\mathbf{A}_i$ and $\mathbf{B}_i$ be 2-by-2 encoding submatrices for parity node i ($i = 1, 2$). Let us explain the interference alignment scheme. First notice that since the repair-bandwidth-per-link is 1, each survivor node j uses a 2-dimensional projection

[2]Independently, Cadambe-Jafar-Maleki [4] have shown the existence of vector linear Exact-Repair MDS codes that attain the optimal tradeoff of (1) for (n, k, d) where $k < n$ and $d = n-1$.

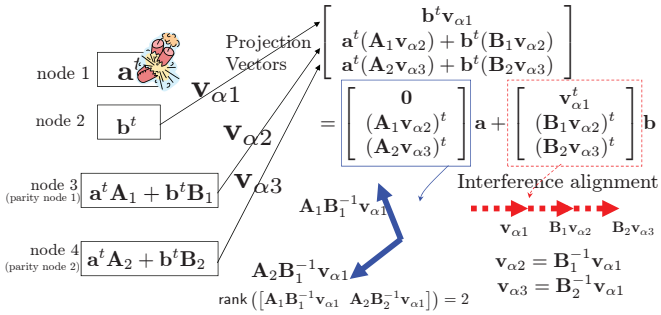


Fig. 1. Interference alignment for exact repair of failed node 1 for a $(4, 2)$ Exact-Repair MDS code.

vector $\mathbf{v}_{\alpha j}$ to project its data into a scalar. By connecting to three nodes, we get: $\mathbf{b}^t\mathbf{v}_{\alpha 1}$; $\mathbf{a}^t(\mathbf{A}_1\mathbf{v}_{\alpha 2}) + \mathbf{b}^t(\mathbf{B}_1\mathbf{v}_{\alpha 2})$; $\mathbf{a}^t(\mathbf{A}_2\mathbf{v}_{\alpha 3}) + \mathbf{b}^t(\mathbf{B}_2\mathbf{v}_{\alpha 3})$.

Here the goal is to decode 2 desired unknowns out of 3 equations including 4 unknowns. To achieve this goal, we need: (i) the matrix associated with the desired signal "$\mathbf{a}$" should have full rank; (ii) the matrix associated with interference "$\mathbf{b}$" should have rank 1. The second condition can be met by setting $\mathbf{v}_{\alpha 2} = \mathbf{B}_1^{-1}\mathbf{v}_{\alpha 1}$ and $\mathbf{v}_{\alpha 3} = \mathbf{B}_2^{-1}\mathbf{v}_{\alpha 1}$. This choice forces the interference space to be collapsed into a one-dimensional linear subspace, thereby achieving interference alignment. We can also satisfy the first condition by carefully choosing $\mathbf{A}_i$'s and $\mathbf{B}_i$'s.

III. A Constructive Framework for Scalar-Linear Codes

This idea cannot be generalized to arbitrary (n, k): it provides the optimal codes only for the case of $k = 2$. For $k \geq 3$ (more-than-two interfering information units), achieving interference alignment for exact repair turns out to be significantly more complex than the $k = 2$ case. We propose a *common-eigenvector* based conceptual framework to overcome the challenge. Specifically, our approach is based on a certain *elementary matrix* property [11].

Our framework consists of four components: (1) developing a family of codes for exact repair of systematic codes based on the common-eigenvector concept; (2) drawing a *dual* relationship between the systematic and parity node repair; (3) guaranteeing the MDS-code property; (4) constructing codes with finite-field alphabets. Step (2) of our framework is a significant distinction from that of [3] and is needed to tackle the full exact repair problem not addressed there. The framework covers the case of $n \geq 2k$ (and $d \geq 2k-1$). It turns out that the $(2k, k, 2k-1)$ code case contains the key design ingredients and the case of $n \geq 2k$ can be derived from this. Hence, we first focus on the simplest example: $(6, 3, 5)$ Exact-Repair MDS codes. Later in Section III-E, we will generalize this to arbitrary (n, k, d) repair codes in the class.

A. Systematic Node Repair

Fig. 2 illustrates the example of repairing systematic node 1 for a $(6, 3, 5)$ code. By the optimal tradeoff (1), the choice

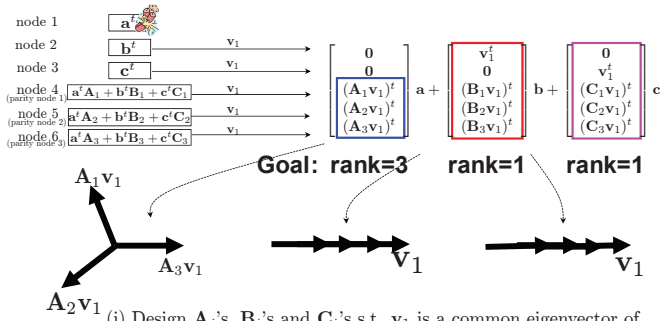


Fig. 2. Illustration of exact repair of systematic node 1 for $(6, 3, 5)$ Exact-Repair MDS codes.

of $\mathcal{M} = 9$ gives $\alpha = 3$ and $\frac{\gamma}{d} = 1$. Let $\mathbf{a} = (a_1, a_2, a_3)^t$, $\mathbf{b} = (b_1, b_2, b_3)^t$ and $\mathbf{c} = (c_1, c_2, c_3)^t$. We define 3-by-3 encoding submatrices of $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{C}_i$ (for $i = 1, 2, 3$); and 3-dimensional projection vectors $\mathbf{v}_{\alpha i}$'s.

By connecting to the five nodes, we get the five equations. In order to successfully recover the desired signal components of $\mathbf{a}$, the matrix associated with $\mathbf{a}$ should have full rank of 3, while the other matrices corresponding to $\mathbf{b}$ and $\mathbf{c}$ should have rank 1, respectively. In accordance with the $(4, 2)$ code example in Fig. 1, if one were to set $\mathbf{v}_{\alpha 3} = \mathbf{B}_1^{-1}\mathbf{v}_{\alpha 1}$, $\mathbf{v}_{\alpha 4} = \mathbf{B}_2^{-1}\mathbf{v}_{\alpha 1}$ and $\mathbf{v}_{\alpha 5} = \mathbf{B}_3^{-1}\mathbf{v}_{\alpha 1}$, then it is possible to achieve interference alignment with respect to $\mathbf{b}$. However, this choice also specifies the interference space of $\mathbf{c}$. If the $\mathbf{B}_i$'s and $\mathbf{C}_i$'s are not designed judiciously, interference alignment is not guaranteed for $\mathbf{c}$. Hence, it is not evident how to achieve interference alignment *at the same time*.

In order to address the challenge of simultaneous interference alignment, we invoke a common eigenvector concept. The idea consists of two parts: (i) designing the $(\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i)$'s such that $\mathbf{v}_1$ is a common eigenvector of the $\mathbf{B}_i$'s and $\mathbf{C}_i$'s, but not of $\mathbf{A}_i$'s[3]; (ii) repairing by having survivor nodes *project* their data onto a linear subspace spanned by this common eigenvector $\mathbf{v}_1$. We can then achieve interference alignment for $\mathbf{b}$ and $\mathbf{c}$ at the same time, by setting $\mathbf{v}_{\alpha i} = \mathbf{v}_1, \forall i$. As long as $[\mathbf{A}_1\mathbf{v}_1, \mathbf{A}_2\mathbf{v}_1, \mathbf{A}_3\mathbf{v}_1]$ is invertible, we can also guarantee the decodability of $\mathbf{a}$. See Fig. 2.

The challenge is now to design encoding submatrices to guarantee the existence of a common eigenvector while also satisfying the decodability of desired signals. The difficulty comes from the fact that in our $(6, 3, 5)$ repair code example, these constraints need to be satisfied for *all* six possible failure configurations. The structure of elementary matrices [11] (generalized matrices of Householder and Gauss matrices) gives insights into this. To see this, consider a 3-by-3 elementary matrix $\mathbf{A} = \mathbf{u}\mathbf{v}^t + \alpha\mathbf{I}$, where $\mathbf{u}$ and $\mathbf{v}$ are 3-dimensional vectors. Here is an observation that motivates our proposed structure: the dimension of the null space of $\mathbf{v}$ is 2 and the null vector $\mathbf{v}^\perp$ is an eigenvector of $\mathbf{A}$, i.e., $\mathbf{A}\mathbf{v}^\perp = \alpha\mathbf{v}^\perp$.

This motivates the following structure:

$$\begin{aligned}
&\mathbf{A}_1 = \mathbf{u}_1\mathbf{v}_1^t + \alpha_1\mathbf{I};\ \mathbf{B}_1 = \mathbf{u}_1\mathbf{v}_2^t + \beta_1\mathbf{I};\ \mathbf{C}_1 = \mathbf{u}_1\mathbf{v}_3^t + \gamma_1\mathbf{I}\\
&\mathbf{A}_2 = \mathbf{u}_2\mathbf{v}_1^t + \alpha_2\mathbf{I};\ \mathbf{B}_2 = \mathbf{u}_2\mathbf{v}_2^t + \beta_2\mathbf{I};\ \mathbf{C}_2 = \mathbf{u}_2\mathbf{v}_3^t + \gamma_2\mathbf{I}\\
&\mathbf{A}_3 = \mathbf{u}_3\mathbf{v}_1^t + \alpha_3\mathbf{I};\ \mathbf{B}_3 = \mathbf{u}_3\mathbf{v}_2^t + \beta_3\mathbf{I};\ \mathbf{C}_3 = \mathbf{u}_3\mathbf{v}_3^t + \gamma_3\mathbf{I},
\end{aligned} \tag{2}$$

where

$$(C1) \quad \begin{cases} \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \text{ is invertible;} \\ \mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] \text{ is invertible,} \end{cases} \tag{3}$$

and the values of the α_i's, β_i's and γ_i's can be arbitrary non-zero values. Only for this section, we consider the simple case of $\mathbf{V} = \mathbf{I}$, although these need not be orthogonal, but only invertible. We then see that: $\mathbf{A}_i\mathbf{v}_1 = \alpha_i\mathbf{v}_1 + \mathbf{u}_i; \mathbf{B}_i\mathbf{v}_1 = \beta_i\mathbf{v}_1; \mathbf{C}_i\mathbf{v}_1 = \gamma_i\mathbf{v}_1$, for $i = 1, 2, 3$. Importantly, notice that $\mathbf{v}_1$ is a common eigenvector of the $\mathbf{B}_i$'s and $\mathbf{C}_i$'s, while simultaneously ensuring that the vectors of $\mathbf{A}_i\mathbf{v}_1$ are linearly independent. Hence, setting $\mathbf{v}_{\alpha i} = \mathbf{v}_1$ for all i, it is possible to achieve simultaneous interference alignment while also guaranteeing the decodability of the desired signals. See Fig. 2. On the other hand, this structure also guarantees exact repair for $\mathbf{b}$ and $\mathbf{c}$. We can use $\mathbf{v}_2$ for exact repair of $\mathbf{b}$. Similarly, $\mathbf{v}_3$ is used for $\mathbf{c}$.

B. Dualization for Parity Node Repair

Parity nodes can be repaired by drawing a *dual* relationship with systematic nodes. The procedure has two steps. The first is to remap parity nodes with $\mathbf{a}'$, $\mathbf{b}'$, and $\mathbf{c}'$, respectively:

$$\begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \\ \mathbf{c}' \end{bmatrix} := \begin{bmatrix} \mathbf{A}_1^t & \mathbf{B}_1^t & \mathbf{C}_1^t \\ \mathbf{A}_2^t & \mathbf{B}_2^t & \mathbf{C}_2^t \\ \mathbf{A}_3^t & \mathbf{B}_3^t & \mathbf{C}_3^t \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix}.$$

Systematic nodes can then be rewritten in terms of the prime notations:

$$\begin{aligned}
\mathbf{a}^t &= \mathbf{a}'^t\mathbf{A}_1' + \mathbf{b}'^t\mathbf{B}_1' + \mathbf{c}'^t\mathbf{C}_1',\\
\mathbf{b}^t &= \mathbf{a}'^t\mathbf{A}_2' + \mathbf{b}'^t\mathbf{B}_2' + \mathbf{c}'^t\mathbf{C}_2',\\
\mathbf{c}^t &= \mathbf{a}'^t\mathbf{A}_3' + \mathbf{b}'^t\mathbf{B}_3' + \mathbf{c}'^t\mathbf{C}_3',
\end{aligned} \tag{4}$$

where the newly mapped encoding submatrices $(\mathbf{A}_i', \mathbf{B}_i', \mathbf{C}_i)$'s are defined as:

$$\begin{bmatrix} \mathbf{A}_1' & \mathbf{A}_2' & \mathbf{A}_3' \\ \mathbf{B}_1' & \mathbf{B}_2' & \mathbf{B}_3' \\ \mathbf{C}_1' & \mathbf{C}_2' & \mathbf{C}_3' \end{bmatrix} := \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \\ \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 \\ \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 \end{bmatrix}^{-1}. \tag{5}$$

With this remapping, one can dualize the relationship between systematic and parity node repair. Specifically, if all of the $\mathbf{A}_i'$'s, $\mathbf{B}_i'$'s, and $\mathbf{C}_i'$'s are *elementary matrices* and form a similar structure as in (2), exact repair of the parity nodes becomes transparent.

The challenge is now how to guarantee the dual structure. We show that a special relationship between $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ through $(\alpha_i, \beta_i, \gamma_i)$'s can guarantee this dual relationship.

[3]Of course, five additional constraints also need to be satisfied for the other five failure configurations for this $(6, 3, 5)$ code example.

Lemma 1: Suppose

$$\mathbf{P} := \left[\begin{array}{c|c|c} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{array}\right] \text{ is invertible.} \quad (6)$$

$$(C2) \quad \kappa\mathbf{U} = \mathbf{V}'\mathbf{P}. \quad (7)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$, $\mathbf{V}' = [\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3']$ is the dual basis matrix i.e., $\mathbf{v}_i'^t\mathbf{v}_j = \delta(i-j)$ and κ is an arbitrary non-zero value s.t. $1-\kappa^2 \neq 0$. Then, all of the $\mathbf{A}_i'$'s, $\mathbf{B}_i'$'s, and $\mathbf{C}_i'$'s are elementary matrices and form a similar structure as in (2), thus ensuring exact repair of parity nodes.

Proof: See [7] for the detailed proof. ■

C. The MDS-Code Property

The third part of the framework is to guarantee the MDS-code property, which allows us to identify specific constraints on the $(\alpha_i, \beta_i, \gamma_i)$'s and/or $(\mathbf{V}, \mathbf{U})$. Consider all four possibilities corresponding to the Data Collector (DC) (i) (1) 3 systematic nodes; (ii) 3 parity nodes; (iii) 1 systematic and 2 parity nodes; (iv) 1 systematic and 2 parity nodes. It turns out that the following condition ensures the MDS-code property:

$$(C3) \quad \textit{Any submatrix of } \mathbf{P} \text{ is invertible.} \quad (8)$$

We use the Gaussian elimination method to check invertibility of the composite matrix for each case. See [7] for the detailed verification.

D. Code Construction with Finite-Field Alphabets

The last part is to design invertible matrices ($\mathbf{P}$, $\mathbf{V}$, $\mathbf{U}$) such that the conditions $(C1), (C2), (C3)$ hold. In order to guarantee $(C3)$, we can use a Cauchy matrix as introduced for the code in [3].

Definition 1 (A Cauchy Matrix [12]): A Cauchy matrix $\mathbf{P}$ is an $m \times l$ matrix with entries p_{ij} in the form:

$$p_{ij} = \frac{1}{x_i - y_j}, \forall i = 1, \cdots, m, j = 1, \cdots, l, x_i \neq y_j,$$

where x_i and y_j are elements of a field and $\{x_i\}$ and $\{y_j\}$ are injective sequences, i.e., elements of the sequence are distinct.

The injective property of $\{x_i\}$ and $\{y_j\}$ requires a finite field size of $2m$ for an $m \times m$ Cauchy matrix. Therefore, in our $(6, 3, 5)$ repair code example, the finite field size of 6 suffices. The field size condition for guaranteeing invertibility of $\mathbf{V}$ is more relaxed.

Theorem 1 ($(6, 3, 5)$ Exact-Repair MDS Codes): Suppose $\mathbf{P}$ of (6) is a Cauchy matrix. Each element of $\mathbf{P}$ is in $\mathsf{GF}(q)$ and $q \geq 6$. Suppose $(\mathbf{V}, \mathbf{U})$ satisfy $(C1), (C2)$. Then, the code is the Exact-Repair MDS code that achieves the optimal tradeoff of (1).

Example: Fig. 3 shows a numerical example for exact repair of (a) systematic node 1 and (b) parity node 1 where $[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = [2, 2, 2; 2, 3, 1; 2, 1, 3]$. Notice that the projection vector solution for systematic node repair is simple: $\mathbf{v}_{\alpha i} = 2^{-1}\mathbf{v}_1 = (1, 1, 1)^t, \forall i$. Note that this choice enables simultaneous interference alignment, while guaranteeing the decodability of $\mathbf{a}$. Notice that (b_1, b_2, b_3) and (c_1, c_2, c_3) are aligned into $b_1 + b_2 + b_3$ and $c_1 + c_2 + c_3$, respectively, while three equations associated with $\mathbf{a}$ are linearly independent. The dual structure also guarantees exact repair of parity nodes. Importantly, we have chosen $(\mathbf{P}, \mathbf{V}, \mathbf{U})$ from a family of codes in [7] such that parity node repair is quite simple. As shown in Fig. 3 (b), downloading only the first equation from each survivor node ensures exact repair. Notice that the five downloaded equations contain only five unknown variables of $(a_1', a_2', a_3', b_1', c_1')$ and three equations associated with $\mathbf{a}'$ are linearly independent. Hence, we can successfully recover $\mathbf{a}'$.

E. Generalization

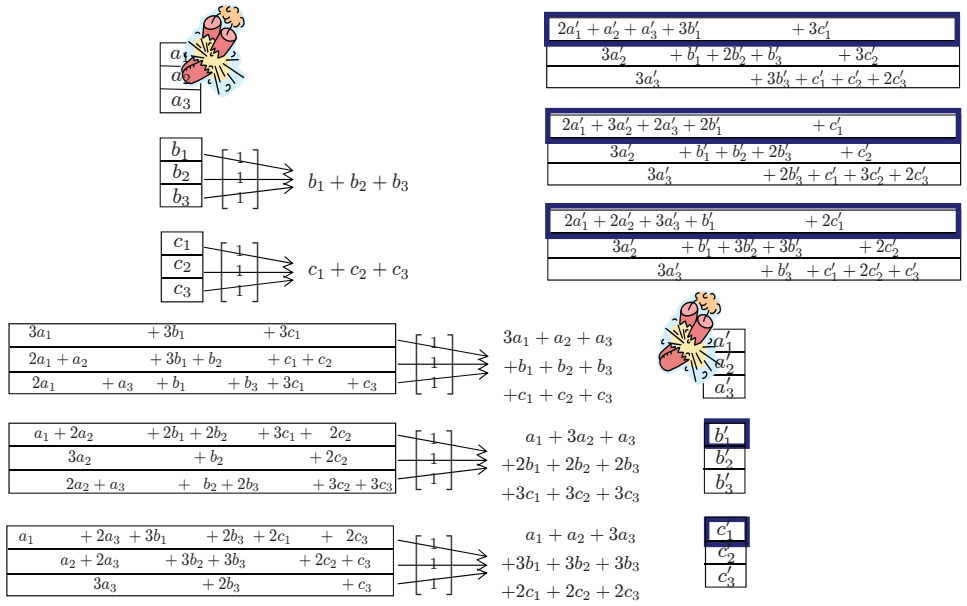
The proposed framework gives insights into generalization to $(2k, k, 2k-1)$ Exact-Repair MDS codes. The key observation is that when assuming $\mathcal{M} = k(d-k+1)$, storage cost is $\alpha = \mathcal{M}/k = d-k+1 = k$ and this number is equal to the number of systematic nodes and furthermore matches the number of parity nodes. Notice that the storage size matches the size of encoding submatrices, which determines the size of $\mathbf{V}$. Therefore, we can easily design $\mathbf{P}, \mathbf{V}, \mathbf{U} \in \mathbb{F}_q^{k \times k}$ such that $(C1), (C2), (C3)$ hold, as long as $q \geq 2k$. This immediately provides $(2k, k, 2k-1)$ Exact-Repair MDS codes.

Now what if k is less than the size $(= \alpha = d-k+1)$ of encoding submatrices, i.e., $d \geq 2k-1$? Note that this case automatically implies that $n \geq 2k$, since $n \geq d+1$. The key observation in this case is that the encoding submatrix size is bigger than k, and therefore we have more degrees of freedom than the number of constraints. Hence, exact repair of systematic nodes becomes transparent. This was observed in [3], where it was shown that for this regime, exact repair of systematic nodes only can be guaranteed by carefully manipulating $(2k, k, 2k-1)$ codes through a puncturing operation. We show that the puncturing technique in [3] can also carry over to ensure exact repair of *all* nodes for our family of codes. We state the theorem only. See [7] for technical details.

Theorem 2 ($\frac{k}{n} \leq \frac{1}{2}, d \geq 2k-1$): Suppose that $(k-1)$ systematic nodes are connected for exact repair. Then, under exact repair constraints of all nodes, the optimal tradeoff of (1) can be attained with a deterministic scheme requiring a field size of at most $2(n-k)$.

IV. Existence of Optimal Exact-Repair MDS Codes

The interference alignment scheme in [6] that permits an arbitrarily large number of symbol extensions (i.e., *vector* linear codes) forms the basis of our results here. The results in [3] say that under scalar linear codes, the case of either $\frac{k}{n} > \frac{1}{2} + \frac{2}{n}$ or $k+1 \leq d \leq \max(k+1, 2k-4)$ induces more constraints than the available number of design variables. This parallels the problem encountered by Cadambe and Jafar [6] in the conceptually similar but physically different context of wireless interference channels. Cadambe and Jafar resolve this issue using the idea of symbol-extension, which is analogous to the idea of vector linear codes for the distributed storage repair problem studied here. Building on the connection described in [7] between the wireless interference channel and the distributed storage repair problems, we leverage the



(a) Exact repair of systematic node 1 (b) Exact repair of parity node 1

Fig. 3. A $(6, 3, 5)$ Exact-Repair MDS code defined over $\mathsf{GF}(4)$ where a generator polynomial $g(x) = x^2 + x + 1$. Since we choose $\mathbf{U} = \mathbf{I}$, for parity node repair, the projection vector solution is much simpler. We download only the first equation from each survivor node; systematic node repair is a bit involved: setting all of the projection vectors as $2^{-1}\mathbf{v}_1 = (1, 1, 1)^t$.

scheme in [6] to the repair problem, showing the existence of Exact-Repair MDS codes that achieve minimum repair bandwidth (matching the cutset lower bound) for all admissible values of (n, k, d). Due to space constraints, we state the theorem only. See [8] for the proof and more technical details.

Theorem 3 ((n, k, d) Exact-Repair MDS Codes): There exist vector linear Exact-Repair MDS codes that achieve the minimum repair bandwidth corresponding to the cutset bound of (1), allowing for any failed node to be exactly repaired with access to any arbitrary d survivor nodes, where $k \leq d \leq n - 1$, provided storage symbols can be split into a sufficiently large number of subsymbols, and the field size can be made sufficiently large.

V. Conclusion

We have systematically developed interference alignment techniques for both scalar-linear and vector-linear Exact-Repair MDS codes. Under scalar-linear codes, we have constructed Exact-Repair MDS codes that achieve the cutset lower bound on repair band for the case of $\frac{k}{n} \leq \frac{1}{2}$; and $d \geq 2k - 1$. Our codes provide insights into a dual relationship between the systematic and parity node repair, as well as opens up a larger constructive design space of solutions. Furthermore, we have shown the existence of vector-linear Exact-Repair MDS codes that are optimal in repair bandwidth, for all admissible values of (n, k, d).

Acknowledgment

We gratefully acknowledge Prof. P. V. Kumar (of IISc) and his students, N. B. Shah and K. V. Rashmi, for insightful discussions and fruitful collaboration related to the structure of Exact-Repair codes.

References

[1] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE INFOCOM*, 2007.

[2] Y. Wu, A. G. Dimakis, and K. Ramchandran, "Deterministic regenerating codes for distributed storage," *Allerton Conference on Control, Computing and Communication*, Sep. 2007.

[3] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Explicit codes minimizing repair bandwidth for distributed storage," *IEEE ITW, online avaiable at arXiv:0908.2984v2*, Jan. 2010.

[4] V. R. Cadambe, S. A. Jafar, and H. Maleki, "Distributed data storage with minimum storage regenerating codes - exact and functional repair are asymptotically equally efficient," *arXiv:1004.4299v1*, Apr. 2010.

[5] M. A. Maddah-Ali, S. A. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, pp. 3457–3470, Aug. 2008.

[6] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degree of freedom for the K user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[7] C. Suh and K. Ramchandran, "Exact regeneration codes for distributed storage repair using interference alignment," *arXiv:1001.0107v2*, Apr. 2010.

[8] ——, "On the existence of optimal exact-repair mds codes for distributed storage," *arXiv:1004.4663v1*, Apr. 2010.

[9] C. Suh and D. Tse, "Interference alignment for cellular networks," *Allerton Conference on Control, Computing and Communication*, Sep. 2008.

[10] Y. Wu and A. G. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," *Proc. of IEEE ISIT*, 2009.

[11] A. S. Householder, *The Theory of Matrices in Numerical Analysis*. Dover, Toronto, Cananda, 1974.

[12] D. S. Bernstein, *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*. Princeton University Press, 2005.