Asymptotic Interference Alignment for Exact Repair in Distributed Storage Systems

Viveck R. Cadambe, Syed A. Jafar, Hamed Maleki
Electrical Engineering and Computer Science
University of California Irvine,
Irvine, California, 92697, USA
Email: {vcadambe, syed, hmaleki}@uci.edu

***Abstract*—In this paper, we consider a distributed storage system where a file of size M is stored in n distributed storage nodes using an (n, k) systematic maximum distance separable (MDS) code. The (n, k) MDS code can protect the storage system from data loss in in case of failure (erasure) of storage nodes, as long as the number of failures is smaller than or equal to $(n - k)$, because of the MDS property of the code. The problem of interest of this paper is to repair failed nodes in the storage system, by replacing them by their replicas (exact repair), as efficiently as possible, i.e., by downloading the minimum possible amount of data from the surviving nodes. Recently, the problem, termed as the exact repair bandwidth problem, has been solved for the special case of $r = 1$ failure using the asymptotic interference alignment scheme developed by Cadambe and Jafar in the context of the wireless interference channel. In this paper, we extend this result to find the minimum repair bandwidth for the more general case of $r > 1$ failures, as long as the number of failures r is smaller than $(n - k)$ - the maximum number of failures that can be tolerated by the system.**

I. Introduction

Recently, there is an increased interest in applications of network codes for distributed storage systems. The motivation of application of network codes for distributed storage comes from the idea that they offer efficient repair strategies for failure of storage nodes in the system [1]. In this paper, we apply interference alignment based techniques to solve an open problem in this regard. Specifically, we apply the asymptotic interference alignment solution of Cadambe and Jafar [2] to construct a class of structured random network codes which efficiently repair failed nodes in distributed storage systems. Consider a set up where there is file of size M to be stored in n distributed storage nodes. The file is split into k equal parts of size M/k and stored in the first k storage nodes, also known as systematic nodes. The remaining $(n - k)$ nodes, known as parity nodes or non-systematic nodes, store data of the same size, i.e. M/k, for redundancy to protect from failure of storage nodes. The parity nodes are designed so that the original file can be completely recovered by a new node using any subset of k nodes of the original n nodes, i.e., so that a failure of up to $(n - k)$ storage nodes can be tolerated. Clearly, for this problem, storing the data using a (n, k) maximum distance separable (MDS) code suffices to achieve the required reconstruction criterion, since a MDS code protects the data from $(n-k)$ erasures. Thus, in general, a new node can download data of total size M by downloading the data stored in any of the k nodes to reconstruct the file and repair up to $n-k$ failed nodes. Now, consider the case where only $r < k$ nodes fail, and a repair center is to replace to the failed nodes. The amount of data to be downloaded by the repair center to repair $r < k$ failed nodes will be henceforth referred to as the *repair bandwidth*. Clearly, a repair bandwidth of M suffices to repair the failed nodes since the repair center can download data of size M from any k of the remaining $n-r$ *healthy* nodes to reconstruct the node exactly. However, note the inherent inefficiency in the solution - to replace data of size rM/k, the repair center downloads data of size M, i.e. k/r times the size of the data to be repaired. In the extreme case of $r = 1$, this factor if inefficiency is equal to k. A question of interest is whether this inefficiency is fundamental, or whether the node can be repaired with the new comer downloading data of size less than M. More specifically, the question of interest of this paper is *what is the minimum repair bandwidth required to repair $r < k$ failed nodes?*

This question of minimum repair bandwidth (See Figure 1) has been studied previously, for the special case of $r = 1$ failed node, from two perspectives [1], [3], [4], [5], [6]. The first is called functional regeneration [1], [3] and the second is called exact (or systematic) regeneration [4], [5], [6]. In functional regeneration, the requirement is to replace failed nodes by functions of the original data, so that the new repaired nodes combined with the surviving nodes form an MDS code. In other words, the repaired nodes are information equivalent to the failed nodes. In contrast, the exact repair problem requires the failed nodes to be replaced by replicas. In other words, the new nodes have to be identical, and not just information equivalent, to the original nodes. Recent results have shown that, for the case of $r = 1$ failed node, surprisingly, exact repair is asymptotically as efficient as functional repair in terms of repair bandwidth. Note that the equivalence is surprising because the constraints of exact regeneration are stricter than functional regeneration and one may expect the former to be more inefficient as compared to the latter owing to the extra constraints. The equivalence of exact and functional repair were shown for the special case of $k \leq n/2$ in references [6], [5] by drawing parallels between the repair problem and the wireless interference channel. The parallels enabled the authors to use interference alignment - an interference management tool - to construct random and even explicit

The ordering of the authors is alphabetical.

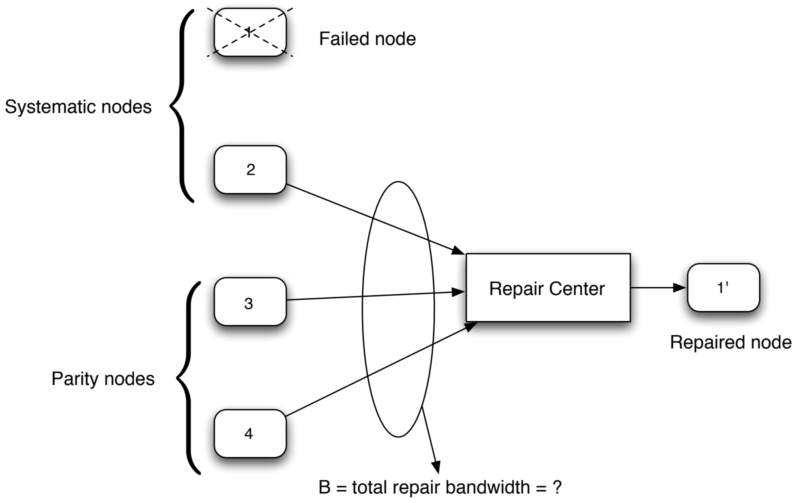

Fig. 1. Pictorial Representation of Problem Definition for $n = 4, k = 2, r = 1$

codes for the distributed storage problem for $k \leq n/2$ in [5], [6], [7], [8]. More recently, references [9], [10], [11] extended the equivalence, albeit in an asymptotic sense, for all (n, k), including the previously open case of $k > n/2$. Thus, the conclusion of [9], [10], [11] implied that, in the limit of large file sizes, exact regeneration is asymptotically as efficient as functional regeneration for the case of $r = 1$ failed node. The references showed this equivalence by adopting into the context of exact repair, the asymptotic alignment scheme developed for obtaining the degrees of freedom of the K-user interference channel in [2]. However, as far as we are aware, the repair efficiency problem for more than 1 failed node has not been studied in MDS code based distributed storage systems. In the main result of this paper, we show that the equivalence is maintained even for the case of $r > 1, r \leq n-k$ failed nodes. Our main result is stated as follows.

Theorem 1: Consider any tuple (n, k) such that $n > k$. For a file of size M stored in n distributed storage nodes as a part of a (n, k) systematic MDS code, the minimum repair bandwidth B for exact repair of any set of r failed nodes where $r < \min(k, n-k)$ satisfies

$$\lim_{M \to \infty} \frac{B}{rM/k} = \frac{n-r}{n-k}.$$

Equivalently, we can write

$$B = \frac{Mr(n-r)}{k(n-k)} + o(M)$$

Further, functional repair of r failed nodes in the system requires a minimum repair bandwidth B_f of

$$\frac{B_f}{rM/k} = \frac{n-r}{n-k}$$

In this paper, we focus on the achievability of the above bound for the exact repair problem. The lower bound on the repair bandwidth is obtained in a manner similar to the $r = 1$ case, and is described in the extended version of this paper [12].

Remark 1: Note that we need $r < (n-k)$ since the code can tolerate upto $n-k$ failures. If $r \geq k$, then the naive strategy of downloading the entire data stored in any k surviving nodes to reconstruct the original data, and then replace the failed nodes is trivially optimal. Therefore, ther regime of interest for the above result is $r < \min(k, n-k)$.

Remark 2: We note that the factor of inefficiency, i.e., the ratio of the repair bandwidth to the amount of data repaired is $\frac{B}{rM/k}$ is equal to $\frac{n-r}{n-k}$. Note that this ratio is always bigger than 1 and smaller than k/r for $r < \min(k, n-k)$. The fact that this factor is smaller than k/r implies that our solution is always more efficient than the trivial solution of downloading the entire data of size M from any k surviving nodes. The ratio being smaller than 1 implies that there is, always, a cost of inefficiency to be paid for using MDS codes during repair, i.e., the amount of data downloaded is always greater than the amount of data repaired. Finally, since $\frac{n-r}{n-k}$ approaches 1 as n becomes large for a fixed k, the inefficiency becomes vanishingly small as n becomes large.

II. Proof of Achievability of Theorem 1

We only provide an intuitive understanding of the achievable scheme here. The complete proof with all the technical details can be found in the extended version of the paper [12]. Consider a distributed storage system storing a total data of M using an (n, k) MDS code. The total data is represented by the $M/k \times k$ dimensional matrix $[\mathbf{x}_1 \ \ \mathbf{x}_2 \ \ \ldots \ \ \mathbf{x}_k]$, where $\mathbf{x}_i$ is an $M/k \times 1$ dimensional vector stored by systematic node $i \in \{1, 2, \ldots, k\}$. Node j, where $j \in \{k+1, k+2, \ldots, n\}$ being a parity node stores the $M/k \times 1$ vector $\mathbf{A}_{j,1}\mathbf{x}_1 + \mathbf{A}_{j,2}\mathbf{x}_2 + \ldots + \mathbf{A}_{j,k}\mathbf{x}_k$, where $\mathbf{A}_{j,i}$ is a $M/k \times M/k$ square matrix for $i \in \{1, 2, \ldots, k\}$. Henceforth, we assume that for $j \leq k$,

$$\mathbf{A}_{j,i} = \left\{ \begin{array}{ll} \mathbf{0} & j \neq i \\ \mathbf{I} & j = i \end{array} \right. , \forall i \in \{1, 2, \ldots, k\}. \qquad (1)$$

The above assumption implies that the data stored in node $j \in \{1, 2, \ldots, n\}$ is the $M/k \times 1$ vector

$$\mathbf{D}_j = \sum_{i=1}^{k} \mathbf{A}_{j,i}\mathbf{x}_i. \tag{2}$$

Note that $\mathbf{A}_{j,i}$ for $j = k+1, k+2, \ldots, n$ are a design choice that define the code; these matrices will henceforth be referred to as the *coding matrices*. We need to choose these matrices so that the code is an MDS code, i.e., using any subset of k nodes, the entire $M \times 1$ vector of data must be reconstructable. Thus, we need to ensure that

$$\text{rank}\left(\begin{bmatrix} \mathbf{A}_{j_1,1} & \mathbf{A}_{j_1,2} & \ldots & \mathbf{A}_{j_1,k} \\ \mathbf{A}_{j_2,1} & \mathbf{A}_{j_2,2} & \ldots & \mathbf{A}_{j_2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{j_k,1} & \mathbf{A}_{j_k,2} & \ldots & \mathbf{A}_{j_k,k} \end{bmatrix}\right) = M \tag{3}$$

for any distinct $j_1, j_2, \ldots, j_k \in \{1, 2, \ldots, n\}$.

Now, when $r, r \leq \min(k, n-k)$ nodes fail, the repair center collects a $\beta \times 1$ vector from each of the remaining $(n-r)$ healthy nodes where $\beta = \frac{B}{n-r}$, so that the total repair bandwidth is B. Our goal is to find the coding matrices $\mathbf{A}_{j,i}, (j,i) \in \{k+1, k+2, \ldots, n\} \times \{1, 2, \ldots, k\}$ and design the $\beta \times 1$ vector to be downloaded by the repair center so as to meet the required bound (presented in the statement of the theorem). We now describe our solution assuming that r systematic nodes fail. The extended version of the paper [12] describes how the solution can be adapted to repair failures of parity nodes. Without loss of generality, let us assume that the first r nodes fail. We provide a linear solution to this problem, so that the $\beta \times 1$ vector downloaded by the repair center from node $j > r$ to repair nodes $1, \ldots, r$ is $\mathbf{V}_j^T\mathbf{D}_j$, where $\mathbf{V}_j$ is a $M/k \times \beta$ matrix. The matrices $\mathbf{V}_j$ will be henceforth referred to as the repair vectors. The repair center now has to regenerate the r $M/k \times 1$ vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$, using $(n-r)$ vectors of the form $\mathbf{V}_j^T\mathbf{D}_j, j = r+1, r+2, \ldots, n$, each of dimension $\beta \times 1$. Notice that the $(k-r)$ vectors (of dimension $\beta \times 1$) downloaded using the $(k-r)$ systematic nodes do not contain any information about the desired vector $\mathbf{x}_1$ and can be interpreted as interference. Therefore, the repair center has, apart from the interference, $(n-k)$ vectors of dimension $\beta \times 1$ containing linear combinations of the desired data. Thus, the vectors available at the repair center can be described as follows.

- $(k-r)$ vectors of the form $\mathbf{V}_j^T\mathbf{x}_j, r < j \leq k$ - these vectors are downloaded from the $(k-r)$ healthy systematic nodes. They contain no information about the desired data, and will be used to cancel interference.
- $(n-k)$ vectors of the $\mathbf{V}_j^T \sum_{i=1}^{k} \mathbf{A}_{j,i}\mathbf{x}_i, k < j \leq n$ - these vectors contain both the desired signal and components of the interference.

The goal of our solution will be to completely cancel the interference from the latter $(n-k)$ vectors using the former $(k-r)$ vectors listed above, and then to regenerate $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ using the latter $(n-k)$ vectors. In order to completely cancel the interference related to $\mathbf{x}_i$ using $\mathbf{V}_i^T\mathbf{x}_i$ by linear techniques, we will need, $\forall j = k+1, k+2, \ldots, n$, $i = r+1, r+2, \ldots, k$, and for some $\beta \times \beta$ matrix Λ_j,

$$\begin{aligned} \mathbf{V}_j^T\mathbf{A}_{j,i}\mathbf{x}_i &= \Lambda_j\mathbf{V}_i^T\mathbf{x}_i & (4) \\ \Rightarrow \text{rowspan}(\mathbf{V}_j^T\mathbf{A}_{j,i}) &\subseteq \text{rowspan}(\mathbf{V}_i^T), & (5) \\ \Rightarrow \text{colspan}(\mathbf{A}_{j,i}^T\mathbf{V}_j) &\subseteq \text{colspan}(\mathbf{V}_i), & (6) \end{aligned}$$

where (5) follows from the fact that the matrices $\mathbf{A}_{j,i}$ and $\mathbf{V}_i$ are picked independent of the data $\mathbf{x}_i$, and therefore need to satisfy (4) for any data vector $\mathbf{x}_i$.

While the above condition ensures that the entire interference can be cancelled, we also need to ensure that, on interference cancellation, the $(n-k)$ vectors of dimension $\beta \times 1$ are sufficient to reconstruct $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$. Note that after interference cancellation, we get $(n-k)\beta$ equations in the rM/k variables formed by the components of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$. All we need to ensure is that these equations have a rank of rM/k, i.e.,

$$\text{colspan}\left(\begin{bmatrix} \mathbf{A}_{k+1,1}^T\mathbf{V}_{k+1} & \mathbf{A}_{k+2,1}^T\mathbf{V}_{k+2} & \ldots & \mathbf{A}_{n,1}^T\mathbf{V}_n \\ \mathbf{A}_{k+1,2}^T\mathbf{V}_{k+1} & \mathbf{A}_{k+2,2}^T\mathbf{V}_{k+2} & \ldots & \mathbf{A}_{n,2}^T\mathbf{V}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{k+1,r}^T\mathbf{V}_{k+1} & \mathbf{A}_{k+2,r}^T\mathbf{V}_{k+2} & \ldots & \mathbf{A}_{n,r}^T\mathbf{V}_n \end{bmatrix}\right)$$
$$= \frac{r.M}{k} \tag{7}$$

Therefore, our goal is to design $\mathbf{A}_{j,i}$ and $\mathbf{V}_l$ for $j \in \{k+1, k+2, \ldots, n\}, i \in \{1, 2, \ldots, k\}, l = r+1, r+2, \ldots, n$ so that

- The code is a (n,k) MDS code.
- The interference is aligned appropriately so that it can be completely canceled.
- The desired signals $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ can be regenerated at the repair center.

Thus, essentially we need to pick $\mathbf{A}_{j,i}$ and $\mathbf{V}_l$ for $j \in \{k+1, k+2, \ldots, n\}, i \in \{1, 2, \ldots, k\}, l \in \{r+1, r+2, \ldots, n\}$ so that (3), (6) and (7) are satisfied. Further, as noted in Remark 1, the field size q and M are also design choices (for large file sizes) that we can use to satisfy these conditions.

A. The solution : Choosing $\mathbf{A}_{j,i}, \mathbf{V}_l, M$ and q

For $k \leq \max(3, n/2), r = 1$, the solutions of [5], [6] design these matrices using Cauchy matrices to satisfy these conditions. Here, note that the conditions (6), (7) are similar to the interference alignment conditions in the interference channel [2]. Specifically, (6) is analogous to the condition that all the interference must align in the K-user interference channel, and (7) is similar to the condition that the desired signal must be linearly independent for linear decoding in the interference channel [2]. These parallels will enable us to build a solution based on the asymptotically perfect interference alignment scheme of the same reference.

On noting that there are $\Gamma = (n-k)(k-r)$ alignment equations in (6), like in [2], we choose $M = k(n-k)\Delta^\Gamma$ and $\beta = (\Delta+1)^\Gamma$, where $\Delta \geq 1$ can be any integer[1]. For any

[1] The intuition for these choices of M and β will hopefully become clear later in this section for a reader unfamiliar with [2].

value of Δ, we show the existence of a field size q, matrices $\mathbf{A}_{j,i}, i \in \{1,2,\ldots,k\}, j \in \{k+1,k+2,\ldots,n\}$ and $\mathbf{V}_l, l \in \{r+1,r+2,\ldots,n\}$ so that (3), (6), (7) are satisfied and the failed nodes can be repaired. Before we proceed to give a random coding based construction of the coding matrices and repair vectors, we will evaluate the repair bandwidth achieved by our scheme. Noting that our construction is applicable for any value of Δ, we can make Δ, a design parameter, arbitrarily large. As $\Delta \to \infty$, we have $M \to \infty$ and

$$\lim_{M\to\infty} \frac{B}{M} = \lim_{\Delta\to\infty} \frac{r(n-r)(\Delta+1)^{\Gamma}}{k(n-k)\Delta^{\Gamma}} = \frac{r(n-r)}{k(n-k)}$$

.

We now proceed to explain our construction of coding matrices and repair vectors satisfying the constraints of repair (3), (6), (7). Our solution, unlike those in references [5], [6], is a random coding solution. Specifically, we choose the coding matrices $\mathbf{A}_{j,i}, i \in \{1,2,\ldots,k\}, j \in \{k+1,k+2,\ldots,n\}$ randomly. We then provide an expression for $\mathbf{V}_l, l \in \{r+1,r+2,\ldots,n\}$ as a (random) function of $\mathbf{A}_{j,i}$ so that (6) is satisfied. Then we show for large field size q, that (3) and (7) are satisfied with a non-zero probability. This implies that there exists at least one choice of coding matrices $\mathbf{A}_{j,i}$ so that all the desired conditions, i.e., (3), (6), (7) are satisfied.

Design of Coding Matrices, $\mathbf{A}_{j,i}$*:* The alignment constraints, (6), are similar to the alignment constraints for the interference channel (See equation (50) in [2]). Note that the matrices $\mathbf{A}_{j,i}$ play a role analogous to channel matrices in wireless interference channels [2]. Drawing inspiration from [2], we choose the $M/k \times M/k$ dimensional matrices $\mathbf{A}_{j,i} \forall j = k+1,k+2,\ldots,n$ to be random *diagonal* matrices with each diagonal entry of each matrix chosen independently and uniformly distributed over the non-zero elements of the field $\mathbb{F}_q$. In other words, we choose

$$\mathbf{A}_{i,j} = \begin{bmatrix} a_{i,j}^{1} & 0 & \ldots & 0 \\ 0 & a_{i,j}^{2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{i,j}^{\frac{M}{k}} \end{bmatrix} \tag{8}$$

with all the diagonal entries chosen independent of each other and independent of all the diagonal entries of all other coding matrices, i.e., with $a_{i,j}^{m}$ chosen independent of $a_{\tilde{i},\tilde{j}}^{\tilde{m}}$ from the non-zero elements of the field, for all $i \neq \tilde{i}$ or $j \neq \tilde{j}$ or $m \neq \tilde{m}$, where $i, \tilde{i} \in \{1,2,\ldots,k\}$, $j, \tilde{j} \in \{k+1, k+2, \ldots, n\}$ and $m, \tilde{m} \in \{1,2,\ldots,\frac{M}{k}\}$. Note that all the coding matrices are full rank since all the diagonal elements are non-zero. In the extended version of this paper [12], we show that that this code is an MDS code with non-zero probability.

Design of Repair Vectors, $\mathbf{V}_l$*:* Here, we provide a set of repair vectors that satisfy (6). We first set the columns of vectors $\mathbf{V}_l$ (which are analogous to beamforming vectors in interference channels)

$$\mathbf{V}_{r+1} = \mathbf{V}_{r+2} = \ldots = \mathbf{V}_k = \mathbf{V}'$$

$$\mathbf{V}_{k+1} = \mathbf{V}_{k+2} = \ldots = \mathbf{V}_n = \mathbf{V}$$

where, $\mathbf{V}$ and $\mathbf{V}'$ are $M/k \times \beta$ dimensional matrices. Then the relations (6) can be re-written as

$$\text{colspan}(\mathbf{A}_{j,i}\mathbf{V}) \subseteq \text{colspan}(\mathbf{V}'), i = r+1, r+2, \ldots, k \tag{9}$$

for $j = k+1, k+2, \ldots, n$. Note that there are $(k-r)(n-k) = \Gamma$ conditions contained in (9). We wish to find $\mathbf{V}, \mathbf{V}'$ so that all these conditions are satisfied.

Intuitive understanding of asymptotic alignment: Before we provide precise expressions for $\mathbf{V}, \mathbf{V}'$, we will intuitively explain the extent of alignment required to to satisfy (6), (7). Since our bandwidth is restricted by β, we need $\text{rank}(\mathbf{V}) \leq \beta = (\Delta+1)^{\Gamma}$ and $\text{rank}(\mathbf{V}') \leq (\Delta+1)^{\Gamma}$. Further, noting that (7) implies $\sum_{j=k+1}^{n} \text{rank}(\mathbf{V}_j) \geq \frac{M}{k}$, we get $\text{rank}(\mathbf{V}) \geq \frac{M}{k(n-k)} = \Delta^{\Gamma}$. Therefore $\mathbf{V}$ must have at least Δ^{Γ} non-zero linearly independent columns. In order to satisfy (9), the span of the $\Gamma\Delta^{\Gamma}$ non-zero column vectors of the matrix

$$[\mathbf{A}_{k+1,1}\mathbf{V} \;\; \mathbf{A}_{k+2,1}\mathbf{V} \;\; \ldots \;\; \mathbf{A}_{n,1}\mathbf{V} \;\; \mathbf{A}_{k+1,2}\mathbf{V} \;\; \mathbf{A}_{k+2,2}\mathbf{V} \;\; \ldots \;\; \mathbf{A}_{n,k}\mathbf{V}]$$

should align in the space spanned by the $(\Delta+1)^{\Gamma}$ column vectors of $\mathbf{V}'$. For large values of Δ, since $\frac{\Delta^{\Gamma}}{(\Delta+1)^{\Gamma}} \to 1$, and all the coding matrices have a full rank of M/k, we have $\frac{\text{rank}(\mathbf{A}_{j,i}\mathbf{V})}{\text{rank}(\mathbf{V}')} \to 1$ for any $j \in \{k+1,\ldots,n\}, i \in \{1,2,\ldots,k\}$. From (9) this implies that $\text{colspan}(\mathbf{A}_{j,i}\mathbf{V}) \approx \text{colspan}(\mathbf{V}')$. In other words, the alignment between the Γ matrices on the left hand side of the Γ relations indicated by (9) is asymptotically perfect for large Δ. Next we return to the mathematical construction of the alignment scheme.

Following the arguments of [2], we choose the set of non-zero column vectors of $\mathbf{V}, \mathbf{V}'$ as shown at the top of the next page[2], where the entries of the $M/k \times 1$ column vector $\mathbf{w}$ are chosen uniformly over the non-zero elements of the field and independent of all the coding matrices.

Thus, the elements of $\mathbf{V}$ contain products of (diagonal) coding matrices corresponding to interference symbols contained in the parity nodes, with each matrix raised to an exponent that is allowed to take integer values from 0 up to $\Delta-1$. Since there are $\Gamma = (k-r)(n-k)$ coding matrices and Δ possible distinct values for the exponent of each matrix, the total number of elements, i.e. column vectors, in $\mathbf{V}$ is Δ^{Γ}. Similarly, the total number of column vectors in $\mathbf{V}'$ is $(\Delta+1)^{\Gamma}$. To understand the notation better, consider, e.g., the case where $\Delta = 1$. Then, $\mathbf{V} = \mathbf{w}$, i.e., just one column vector, and $\mathbf{V}'$ contains all the 2^{Γ} vectors of the form

$$\mathbf{A}_{k+1,r+1}^{\alpha_{k+1,r+1}} \mathbf{A}_{k+1,r+2}^{\alpha_{k+1,r+2}} \ldots \mathbf{A}_{k+1,n}^{\alpha_{k+1,n}} \mathbf{A}_{k+2,r+1}^{\alpha_{k+2,r+1}} \ldots \mathbf{A}_{n,k}^{\alpha_{n,k}} \mathbf{w}$$

where $\alpha_{j,i} \in \{0,1\}$. For any general value of Δ, the columns of $\mathbf{V}$ are of the form

$$\mathbf{A}_{k+1,r+1}^{\alpha_{k+1,r+1}} \mathbf{A}_{k+1,r+2}^{\alpha_{k+1,r+2}} \ldots \mathbf{A}_{k+1,n}^{\alpha_{k+1,n}} \mathbf{A}_{k+2,r+1}^{\alpha_{k+2,r+1}} \ldots \mathbf{A}_{n,k}^{\alpha_{n,k}} \mathbf{w}$$

[2] For convenience, we ignore the abuse in notation of these equations; the quantity on the left denotes the matrix, whereas the quantity on the right only denotes the set of non-zero columns of the matrix.

$$\mathbf{V} = \left\{ \left(\prod_{\substack{j=k+1,...,n \\ i=r+1,...,k}} \mathbf{A}_{j,i}^{\alpha_{j,i}} \right) \mathbf{w} : \alpha_{k+1,r+1}, ..., \alpha_{n,k} \in \{0, 1, ..., \Delta - 1\} \right\} \tag{10}$$

$$\mathbf{V}' = \left\{ \left(\prod_{\substack{j=k+1,...,n \\ i=r+1,...,k}} \mathbf{A}_{j,i}^{\alpha_{j,i}} \right) \mathbf{w} : \alpha_{k+1,r+1}, ..., \alpha_{n,k} \in \{0, 1, 2, ..., \Delta\} \right\} \tag{11}$$

where $\alpha_{j,i} \in \{0, 1, \ldots, \Delta - 1\}$ and $\mathbf{V}^{'}$ has columns of the form

$$\mathbf{A}_{k+1,r+1}^{\alpha_{k+1,r+1}} \mathbf{A}_{k+1,r+2}^{\alpha_{k+1,r+2}} \ldots \mathbf{A}_{k+1,n}^{\alpha_{k+1,n}} \mathbf{A}_{k+2,r+1}^{\alpha_{k+2,r+1}} \ldots \mathbf{A}_{n,k}^{\alpha_{n,k}} \mathbf{w}$$

where $\alpha_{j,i} \in \{0, 1, \ldots, \Delta\}$. Note that the ordering of the matrices $\mathbf{A}_{j,i}$ in the above notation is irrelevant, since the coding matrices, being diagonal, commute. This commuting property is the key to the alignment scheme. Because the ordering of matrices is irrelevant, it is readily verified that multiplying any column vector from $\mathbf{V}$ by any of the $\mathbf{A}_{j,i}$ involved, produces a column vector contained in $\mathbf{V}^{'}$. This is because multiplication by $\mathbf{A}_{j,i}$ simply raises the corresponding exponent of the element in $\mathbf{V}$ by one, but the elements of $\mathbf{V}^{'}$ already include all such terms. Since the set of columns of $\mathbf{A}_{j,i}\mathbf{V}$ is a sub-set of the columns of $\mathbf{V}^{'}$ for any $j \in \{k+1, k+2, \ldots, n\}, i \in \{r+1, r+2, \ldots, k\}$, it is evident that this choice of repair vectors satisfies (9), and equivalently, (6). Finally, we need to show that (7) is satisfied. Intuitively, we note that the repair vectors $\mathbf{V}, \mathbf{V}^{'}$ in (10),(11) are chosen independent of the coding matrices involved in (7). Thus, the (non-zero) columns of the left hand side of (7) are all linearly independent with high probability, if the field size is large enough, since each column entry is a function of a random entry which is independent of all other entries in the matrix. Since the number of non-zero columns of the matrix on the left hand side of (7) is equal to rM/k, the matrix has the desired rank of rM/k. A rigorous proof of (7) can be found in [12].

III. Conclusion

Recent literature in wireless networks has shown that, in a network with multiple sources, random network coding does not suffice, and we need structured codes in general to align interference. This insight extends to wired networks as well, where, while random network coding achieves capacity for a single-source network, we need a more structured approach to choosing network coding co-efficients in networks with multiple sources to align interference. Recently, this connection has been increasingly used to obtain new insights on multi-source multicast wired networks [13]. Since exact repair in distributed storage is related to the multi-source multicast problem, insights from interference alignment are useful in finding capacity of the associated class of wired networks. The results of this paper can thus be viewed as an expansion of the class of networks where asymptotic alignment is useful to achieve capacity (See extended version [12] for details). A natural direction of research in this regard is explore the extent to which the techniques of interference alignment could reveal insights into the capacity of multi-source multicast wired networks. From a practical perspective, while the exact regeneration problem is effectively solved from a perspective of existence of optimal codes, explicit construction of codes which achieve efficient repair is an open problem.

References

[1] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, pp. 4539 –4551, Sep. 2010.

[2] V. Cadambe and S. Jafar, "Interference alignment and the degrees of freedom of the k user interference channel," *IEEE Trans. on Information Theory*, vol. 54, pp. 3425–3441, Aug. 2008.

[3] Y. Wu, "Existence and construction of capacity-achieving network codes for distributed storage," in *IEEE International Symposium on Information Theory*, pp. 1150 –1154, 28 2009-july 3 2009.

[4] Y. Wu and A. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *IEEE International Symposium on Information Theory*, pp. 2276 –2280, 28 2009-july 3 2009.

[5] C. Suh and K. Ramchandran, "Exact regeneration codes for distributed storage repair using interference alignment," *CoRR*, vol. abs/1001.0107, 2010. http://arxiv.org/abs/1001.0107.

[6] N. B. Shah, R. K. V., P. V. Kumar, and K. Ramachandran, "Explicit codes minimizing repair bandwidth for distributed storage," *CoRR*, vol. abs/0908.2984, 2009. http://arxiv.org/abs/0908.2984.

[7] R. K. V., N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the msr and mbr points via a product-matrix construction," *CoRR*, vol. abs/1005.4178, 2010. http://arxiv.org/abs/1005.4178.

[8] B. Gaston and J. Pujol, "Double circulant minimum storage regenerating codes," *CoRR*, vol. abs/1007.2401, 2010. http://arxiv.org/abs/1007.2401.

[9] V. R. Cadambe, S. Jafar, and H. Maleki, "Distributed data storage with minimum storage regenerating codes - exact and functional repair are asymptotically equally efficient," *CoRR*, vol. abs/1004.4299, April 2010. http://arxiv.org/abs/1004.4299.

[10] C. Suh and K. Ramchandran, "On the existence of optimal exact-repair mds codes for distributed storage," *CoRR*, vol. abs/1004.4663, April 2010. http://arxiv.org/abs/1004.4663.

[11] V. R. Cadambe, S. Jafar, and H. Maleki, "Minimum repair bandwidth for exact regeneration in distributed storage," in *Proceedings of the IEEE Wireless Network Coding Conference (WiNC), Boston, MA, USA*, June 2010.

[12] V. R. Cadambe, S. Jafar, and H. Maleki, "Distributed data storage with minimum storage regenerating codes - exact and functional repair are asymptotically equally efficient." Preprint available on author's.

[13] A. Das, S. Vishwanath, S. Jafar, and A. Markopoulou, "Network coding for multiple unicasts: An interference alignment approach," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1878 –1882, june 2010.