Codes with Local Regeneration

Govinda M. Kamath, N. Prakash, V. Lalitha and P. Vijay Kumar
Department of ECE, Indian Institute of Science, Bangalore, India
Email: {govinda, prakashn, lalitha, vijay}@ece.iisc.ernet.in.

***Abstract*—Regenerating codes and codes with locality are two schemes that have recently been proposed to ensure data collection and reliability in a distributed storage network. In a situation where one is attempting to repair a failed node, regenerating codes seek to minimize the amount of data downloaded for node repair, while codes with locality attempt to minimize the number of helper nodes accessed. In this paper, we provide several constructions for a class of vector codes with locality in which the local codes are regenerating codes, that enjoy both advantages. We derive an upper bound on the minimum distance of this class of codes and show that the proposed constructions achieve this bound. The constructions include both the cases where the local regenerating codes correspond to the MSR as well as the MBR point on the storage-repair-bandwidth tradeoff curve of regenerating codes.**

I. Introduction

Apart from ensuring reliability of stored data, the principal goals in a distributed storage network relate to efficient data collection and node repair. Our interest is in coding schemes which store the data across n nodes in such a way that a data collector can recover the data by connecting to a small number k of nodes in the network. Node repair is to be accomplished by connecting to a subset of nodes and downloading a uniform amount of data from each node. The number of nodes contacted for repair is termed the repair degree while the total amount of data downloaded for repair is called the repair bandwidth. It is of interest to minimize both repair degree as well as repair bandwidth. It is also desirable to have multiple options for both data collection and node repair.

Distributed storage systems found in practice, include Windows Azure Storage [1] and the Hadoop-based systems [2] used in Facebook and Yahoo. Maximum-distance separable (MDS) codes are commonly used in distributed storage systems, for example in HDFS RAID [3]. MDS coding schemes, while optimal in terms of storage overhead, are however inefficient in terms of node repair, as the repair degree as well as repair bandwidth are both large. Two alternative approaches to coding have recently been advocated to enable more efficient node repair, namely, regenerating codes [4] and codes with locality [5].

A. Regenerating Codes

In the regenerating-code framework [4], a file of B symbols over some finite field $\mathbb{F}_q$ is encoded to $n\alpha$ symbols over $\mathbb{F}_q$ and stored across n nodes in the network, each node storing α code symbols. The codes are structured in such a way that a data collector can download the data by connecting to any k nodes and node repair can be accomplished by connecting to any d nodes while downloading $\beta \leq \alpha$ symbols from each node. The quantity $d\beta$ is termed the repair bandwidth. A cut-set bound from network coding, tells us that given code parameters $((n,k,d),(\alpha,\beta),B)$ the size of a data file is upper bounded [4] by

$$B \quad \leq \quad \sum_{i=0}^{k-1} \min\{\alpha,(d-i)\beta\}. \qquad (1)$$

For fixed values of parameters $\{B,k,d\}$, there are multiple pairs (α,β) that satisfy (1), which results in a storage-repair-bandwidth trade-off. At the Minimum Storage Regeneration (MSR) point, the total storage $n\alpha$ is as small as possible while at the Minimum Bandwidth Regeneration (MBR) point $d\beta$ is minimized. The regenerating codes in this paper carry out exact repair and thus the contents of the failed and replacement nodes are identical. Explicit constructions of MSR codes for some parameters are presented in [6], [7], [8], [9]. Existence of MSR codes for all $(n,k,d),\ n > d \geq k$, is shown in [10]. Explicit MBR codes for all $(n,k,d),\ n > d \geq k$ are presented in [7]. In [11], a class of MBR codes with $d=(n-1)$ are presented, which are termed as repair-by-transfer MBR codes as they enable node repair without need for any operation other than simple data transfer.

B. Codes with Locality

Let $\mathcal{C}$ be an $[n,\kappa,d_{\text{min}}]$ linear code[1] over $\mathbb{F}_q$, having code symbols $\{c_i\}_{i=1}^{n}$. The j^{th} code symbol c_j is said to have (r,δ)-*locality* if there exists a subset of code symbols that includes c_j and that forms a "local" code with parameters $[r+\delta-1, \leq r, \delta]$. The code is said to have information locality if the code has (r,δ) locality for a collection S of code symbols from which the message symbols can be recovered. The code has all-symbol locality if all code symbols have (r,δ) locality. The notion of locality was introduced in [5] by Gopalan et. al. . The upper bound below

$$d_{\text{min}} \quad \leq \quad (n-\kappa+1) - \left(\left\lceil \frac{\kappa}{r} \right\rceil - 1\right)(\delta-1), \qquad (2)$$

(with $\delta = 2$) was derived and two constructions provided. The first construction arises from an earlier construction of class of codes termed as pyramid codes [15]. The second construction (of an all-symbol locality code) is existential in nature and based on a counting argument. A class of code closely related

[1] We use κ in place of k to avoid a clash with regenerating code notation. Also, the bounds derived in [5], [12], [13], [14] apply to local codes with parameters $[\leq r+\delta-1, \leq r, \geq \delta]$. We adopt the slightly more restrictive definition of locality here for simplicity in presentation.

to the pyramid code has been employed in Windows Azure Storage [1]. The authors in [16] implement a class of codes with locality (called locally repairable codes) in HDFS and compare the performance with Reed Solomon codes.

The results in [5] were subsequently extended in [12] and [13] to scalar codes with arbitrary δ, and vector codes with $\delta = 2$ respectively. Upper bounds and constructions are provided in both papers for their respective settings. A general construction of explicit and optimal codes with all-symbol locality for $\delta = 2$ is provided in [17].

C. Overview of Results

In this paper, we construct codes (over a vector alphabet) with locality in which the local codes are themselves, regenerating codes. This makes the codes efficient both in terms of download bandwidth as well as repair degree. We term such codes as codes with local regeneration or equivalently, locally regenerating codes. We present a bound on both the minimum distance and code size for this class of codes and the constructions provided are optimum with respect to this bound. In an independent and parallel work[2], the authors of [18] also consider codes with all-symbol locality where the local codes are regenerating codes. Bounds on minimum distance are provided and a construction for optimal codes with MSR all-symbol locality based on rank-distance codes is presented.

Section II introduces vector codes and a class of vector codes called uniform rank accumulation(URA) codes. Section III provides bounds on minimum distance of codes with locality where the local codes have the URA property. Optimal constructions of codes with locality, where the local codes are MSR and MBR codes are presented in Sections IV, V respectively. A performance comparison between locally regenerating codes, regenerating codes and scalar codes with locality is presented in Section VI. Most proofs are omitted for lack of space and and can be found in [14].

II. Vector Codes

By vector codes, we mean codes over a vector alphabet of the form $\mathbb{F}_q^\alpha$ for some $\alpha \geq 1$, that are linear over $\mathbb{F}_q$, i.e., $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ and $a, b \in \mathbb{F}_q \Rightarrow a\mathbf{c} + b\mathbf{c}' \in \mathcal{C}$. As a vector space over $\mathbb{F}_q$, $\mathcal{C}$ has dimension K, termed the scalar dimension (or file size) of the code and as a code over the alphabet $\mathbb{F}_q^\alpha$, the code has minimum distance $d_{\min}$. Associated with the vector code $\mathcal{C}$ is an $\mathbb{F}_q$-linear scalar code $\mathcal{C}^{(s)}$ of length $N = n\alpha$, where $\mathcal{C}^{(s)}$ is obtained by expanding each vector symbol within a codeword into α scalar symbols (in some prescribed order). Given a generator matrix G for the scalar code $\mathcal{C}^{(s)}$, the first code symbol in the vector code is naturally associated with the first α columns of G etc. We will refer to the collection of α columns of G associated with the i^{th} code symbol $\mathbf{c}_i$ as the i^{th} thick column. To avoid having to deal with degeneracy, we will assume that all the α columns comprising a thick column in the generator matrix of a vector code are linearly independent. We will also refer to a vector code of block length n, scalar dimension K, minimum distance $d_{\min}$, and (vector-size parameter) α as an $[n, K, d_{\min}, \alpha]$ code. We note that vector codes have been extensively studied in literature as array codes [19].

A. Uniform Rank Accumulation Codes

Let $\mathcal{C}$ be an $[n, K, d_{\min}, \alpha]$ vector code having generator matrix G and let $S_i, 1 \leq i \leq n$ be an arbitrary subset of i thick columns of G. The code $\mathcal{C}$ is said to be a Uniform Rank Accumulation (URA) code or a code possessing the URA property, if $\text{Rank}\,(G|_{S_i}) = \sum_{j=1}^{i} a_j$, for some set $\{a_1, a_2, \cdots, a_n\}$ of non-negative integers, that are independent of the specific set S_i of i thick columns chosen. Then $\{a_i, 1 \leq i \leq n\}$ will be called the rank accumulation profile of $\mathcal{C}$. We have that

$$\alpha = a_1 \geq a_2 \geq \cdots a_{n-2} \geq a_{n-1} \geq a_n \geq 0, \qquad (3)$$

with $\sum_{i=1}^{n} a_i = K$, where K is the scalar dimension of $\mathcal{C}$. Further, in terms of the minimum distance of the code:

$$a_{n-i} = 0, \; 0 \leq i \leq (d_{\min} - 2), \quad a_{n-d_{\min}+1} > 0. \qquad (4)$$

It can be shown that any $((n, k, d), (\alpha, \beta))$ MSR code is an URA code, having URA profile

$$a_i \;=\; \alpha, \; 1 \leq i \leq k, \qquad a_i \;=\; 0, \; k+1 \leq i \leq n. \quad (5)$$

Similarly, from the results in [11], it can be shown that MBR codes also belong to the URA family with URA profile:

$$a_i = \alpha - (i-1)\beta, \; 1 \leq i \leq k, \quad a_i = 0, \; k+1 \leq i \leq n. \quad (6)$$

III. Codes with URA Locality

In this section, we discuss locality in the context of vector codes and provide a bound on minimum distance under the assumption that the local codes have the URA property. With additional assumptions, we also present the structure of a code which achieves the bound on minimum distance with equality.

Analogous to the scalar case, the j^{th} code symbol $\mathbf{c}_j \in \mathbb{F}_q^\alpha$ of an $[n, K, d_{\min}, \alpha]$ vector code $\mathcal{C}$ has (r, δ)-*locality* if there exists a subset of code symbols that includes $\mathbf{c}_j$ and that forms a (vector) local code with parameters $[r + \delta - 1, \leq r\alpha, \delta, \alpha]$. The code has information locality if the code has (r, δ) locality for a collection S of code symbols from which the K message symbols can be recovered. The code has all-symbol locality if all code symbols have (r, δ) locality. We also let $\{\mathcal{C}_i\}_{i=1}^{m}$ to denote the collection of all local codes of $\mathcal{C}$ with parameters $[r + \delta - 1, \leq r\alpha, \delta]$. The case of locality in vector codes with $\delta = 2$ was previously considered in [13], where it was shown that under $(r, \delta = 2)$-all-symbol locality, the minimum distance $d_{\min}$ of $\mathcal{C}$ is upper bounded by

$$d_{\text{min}} \;\; \leq \;\; n - \left\lceil \frac{K}{\alpha} \right\rceil + 1 - \left(\left\lceil \frac{K}{r\alpha} \right\rceil - 1 \right). \qquad (7)$$

Let $\mathcal{U}$ (for URA) denote the class of $\mathbb{F}_q$-linear vector codes $\mathcal{C}$, where each code $\mathcal{C}$ is an $[n, K, d_{\text{min}}, \alpha]$ vector code

- possessing (r, δ) information locality with $\delta \geq 2$, and
- where all the local codes $\{\mathcal{C}_i\}_{i=1}^{m}$ of $\mathcal{C}$ are URA codes with rank accumulation profile $\{a_i, 1 \leq i \leq (r+\delta-1)\}$.

[2] See submissions [14] and [18] on arXiv.

We use n_L, K_L to denote the block length and scalar dimension of the local codes $\{\mathcal{C}_i\}_{i=1}^m$ respectively, i.e.,

$$n_L \triangleq r+\delta-1, \quad K_L \triangleq \sum_{i=1}^{n_L} a_i. \tag{8}$$

A. Sub-Additivity

We extend the finite length vector $(a_1, a_2, \cdots, a_{n_L})$ to a semi-infinite sequence $\{a_i\}_{i=1}^{\infty}$ of period n_L by defining

$$a_{i+jn_L} \;=\; a_i, \; 1 \le i \le n_L, \; j \ge 1. \tag{9}$$

We use $P(s), s \ge 0$, to denote the sequence of leading sums of this semi-infinite sequence, i.e.,

$$P(s) \;=\; \sum_{i=1}^{s} a_i, \quad s \ge 1, \quad P(0)=0. \tag{10}$$

It follows from the periodicity of $\{a_i\}$ that

$$P(u_1 n_L + u_0) = u_1 K_L + P(u_0) \;; u_1 \ge 0, \; 1 \le u_0 \le n_L.$$

Additionally, it can be verified that $P(\cdot)$ is sub-additive, i.e.,

$$P(s+s') \;\le\; P(s)+P(s'), \; \forall \; s, s' \ge 0, \tag{11}$$

We next define the function $P^{(\text{inv})}$ by setting $P^{(\text{inv})}(\nu)$, for $\nu \ge 1$, to be the smallest integer s such that $P(s) \ge \nu$. It can be verified that for all $v_1 \ge 0, \quad 1 \le v_0 \le K_L$,

$$P^{(\text{inv})}(v_1 K_L + v_0) \;=\; v_1 n_L + P^{(\text{inv})}(v_0).$$

B. Upper Bound on Minimum Distance

Theorem 3.1: Let $\mathcal{C}$ belong to Class $\mathcal{U}$. Then the minimum distance of $\mathcal{C}$ is upper bounded by

$$d_{\min} \;\le\; n+1-P^{(\text{inv})}(K). \tag{12}$$

When $K_L \mid K$, the bound takes on the form

$$d_{\min} \;\le\; n+1-\left(\frac{K}{K_L}\right)r - \left(\frac{K}{K_L}-1\right)(\delta-1). \tag{13}$$

Sketch of Proof: We use the fact that given any set $T \subseteq [n]$ such that rank $(G|_T) < K$, we have $d_{\min} \le n-|T|$. Thus it suffices to construct a set $T \subseteq [n]$ such that $|T| \ge P^{(\text{inv})}(K)-1$ and Rank$(G|_T) < K$. Such a T is easily constructed if there are enough local codes having disjoint support that also contribute K_L to the scalar dimension of $\mathcal{C}$. In this case, T is simply the union of supports of $\lfloor \frac{K}{K_L} \rfloor$ disjoint local codes, to which we also add partial support ($< r$) of a further disjoint local code. In the general case, where the local codes do not have disjoint support, the set T is obtained using an algorithm similar to the one used in [5] (for scalar locality). The analysis of the algorithm in this case, makes use of the sub-additive properties of $P(\cdot)$, see [14]. ■

Corollary 3.2: Let $\mathcal{C}$ belong to Class $\mathcal{U}$. Then given n and $d_{\min}$, the scalar dimension of $\mathcal{C}$ is upper bounded by

$$K \;\le\; P(n-d_{\min}+1).$$

We say that $\mathcal{C}$ is distance-optimal if $d_{\min} = n+1-P^{(\text{inv})}(K)$ and rate-optimal if $K = P(n-d_{\min}+1)$.

C. Structure of Optimal Codes

We say that the function P is strictly sub-additive if for any $s \ge 1, s' \ge 1$ such that $s+s' \le n_L$, we have $P(s+s') < P(s)+P(s')$. A necessary and sufficient condition for this is that $a_2 < a_1$.

Theorem 3.3: Let $\mathcal{C} \in \mathcal{U}$ be both distance and rank optimal. If either,

- $K_L \mid K$, or
- P is strictly sub-additive,

the local codes $\{\mathcal{C}\}_{i=1}^m$ must all have disjoint supports.

Note that for an MBR code, P is strictly sub-additive and hence Theorem 3.3 readily applies.

IV. MSR-Local Codes

Four constructions of distance and rate optimal MSR-local codes with $\delta \ge 3$ are presented here of which the first two are explicit. The third construction will prove the existence, for sufficiently large field size, of MSR-local codes for a wider range of code parameters. The fourth construction will establish the existence of all-symbol MSR-local codes whenever $n_L \mid n$. The bound on $d_{\min}$ in (12) when specialized to the case of codes with (r,δ) MSR locality, yields

$$\begin{aligned} d_{\min} \;&\le\; n+1-P^{(\text{inv})}(K) \\ &= \left(n - \left\lceil \frac{K}{\alpha} \right\rceil + 1\right) - \left(\left\lceil \frac{K}{r\alpha} \right\rceil - 1\right)(\delta-1). \end{aligned} \tag{14}$$

A. Sum-Parity MSR-Local Codes

Theorem 4.1: Let $\mathcal{C}_0$ be an $((n_L+\Delta, r, d), (\alpha,\beta), K_L = r\alpha)$ MSR code with $n_L = (r+\delta-1)$, $d \le n_L - 1$ and $\Delta \le \delta$. Let $G_0 = [G_L \mid Q_\Delta]$ be a generator matrix of $\mathcal{C}_0$, where G_L and Q_Δ are submatrices having $n_L\alpha$ and $\Delta\alpha$ columns respectively. Let the code $\mathcal{C}$ have generator matrix

$$G = \left[\begin{array}{ccc|c} G_L & & & Q_\Delta \\ & \ddots & & \vdots \\ & & G_L & Q_\Delta \end{array}\right], \tag{15}$$

in which both matrices G_L and Q_Δ appear $m \ge 1$ times. Then $\mathcal{C}$ is an optimal MSR-local code with (r,δ) information locality, having parameters $K = mr\alpha$, $n = mn_L + \Delta$, $\alpha = (d-r+1)\beta$ and d_{min} given by equality in (14).

Proof: Since $d \le n_L - 1$, the matrix G_L generates an $((n_L, r, d), (\alpha, \beta), K_L)$ MSR code and hence C is an MSR-local code with (r,δ) information locality. To calculate the minimum distance, note that if $\mathbf{c}$ is any non-zero codeword and has non-zero components belonging to two or more local codes, then Hamming weight$(\mathbf{c}) \ge 2\delta \ge \delta + \Delta$. On the other hand, if the non-zero components of $\mathbf{c}$ are restricted to one of the local codes and the global parities, then it follows from the minimum distance of $\mathcal{C}_0$ that its weight is $\ge \delta+\Delta$ and thus $d_{\min} \ge \delta+\Delta$. Finally, note that the bound in (14), for the given code parameters, reduces to $d_{\min} \le \delta+\Delta$. ■

B. Pyramid-Like MSR-Local Codes

The construction below mimics the construction of pyramid codes in [15], with the difference that we are now dealing with vector symbols in place of scalars and local MSR codes in place of local MDS codes.

Theorem 4.2: Let $\mathcal{C}'$ be an $((n' = mr + \delta - 1 + \Delta, k' = mr, d), (\alpha, \beta), K = mr\alpha)$ exact repair MSR code such that $d \leq n' - \Delta - 1 = mr + \delta - 2$. Let the (systematic) generator matrix G' of $\mathcal{C}'$ be given by $G' = \left[\begin{array}{c|c|c} I_{mr\alpha} & Q & Q' \end{array}\right]$, where $I_{mr\alpha}$ denotes an identity matrix of size $mr\alpha$ and where the matrices Q, Q' have $(\delta - 1)\alpha$ and $\Delta\alpha$ columns respectively. The "punctured" generator matrix $G'' \triangleq [I_{mr\alpha} \mid Q]$ generates an $((n' - \Delta, k', d), (\alpha, \beta))$ MSR code, say $\mathcal{C}''$. Then the generator matrix G of the desired code $\mathcal{C}$ is obtained by splitting and rearranging the columns of Q, as shown below

$$G = \left[\begin{array}{cccccc|c} I_{r\alpha} & Q_1 & & & & & \\ & & \ddots & \ddots & & & Q' \\ & & & & I_{r\alpha} & Q_m & \end{array}\right], \quad (16)$$

where the $\{Q_i\}$ are matrices of size $(r\alpha \times (\delta - 1)\alpha)$. Then $\mathcal{C}$ is an MSR-local code with (r, δ) information locality, having parameters $K = mr\alpha$, $n = m(r+\delta-1)+\Delta$, $\alpha = (d-r+1)\beta$ and $d_{\min}$ given by equality in (14), i.e., $d_{\min} = \Delta + \delta$.

Proof: The code $\mathcal{C}$ clearly has (r, δ) information locality, where the local codes are generated by $[I_{r\alpha} \mid Q_i]$, $i \in [m]$. Also, it is easily seen that all the local codes are shortened codes of $\mathcal{C}''$. It is shown in Theorem 6 of [7] that shortening an MSR code results in a second MSR code, from which we conclude that $\mathcal{C}$ is an MSR-local code with (r, δ) information locality. The fact that the code has the required minimum distance follows by observing that $d_{\min}(\mathcal{C}) \geq d_{\min}(\mathcal{C}')$. ■

Remark 1: The existence of MSR codes for all possible $[n, k, d]$ has been shown in [10] and these codes could be used as the codes $\mathcal{C}_0, \mathcal{C}'$ in the two constructions above. In terms of known, explicit constructions, the code $\mathcal{C}_0$ can be picked from the product-matrix class [7] of MSR codes. The product-matrix construction requires $d \geq 2r - 2$, which combined with $d \leq r + \delta - 2$, leads to the constraint $r \leq \delta$ on the applicability of this construction in Theorem 4.1. When combined with the requirement $d \leq mr + \delta - 2$, it leads to the constraint $mr \leq \delta$ on the applicability of this construction in Theorem 4.2.

C. Existence of MSR-Local Codes

Theorem 4.3: Given the existence of an $((n_L, r, d), (\alpha, \beta), K_L = r\alpha)$ exact-repair MSR code over $\mathbb{F}_q$ with $n_L = (r + \delta - 1)$, there exists a distance optimal $[n, K, d_{\min}, \alpha]$ MSR-local code with

(a) (r, δ) information locality, whenever $K = mK_L, m \geq 2$ and field size $q > \binom{n}{mr}$

(b) (r, δ) all symbol locality, whenever $n = mn_L, m \geq 2$, $K = \ell\alpha, r \leq \ell \leq mr$ and field size $q > \binom{n}{\ell}$.

For the case of information locality, unlike in the case of Theorems 4.1 and 4.2, there is no constraint here on the repair degree d involving r and δ and thus Theorem 4.3 is applicable for a wider range of parameters.

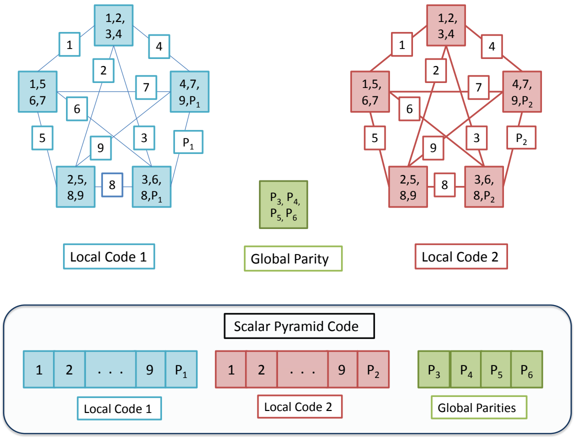

Fig. 1. Repair-by-Transfer MBR-Local code shown on top, with parameters $n = 11, K = 18, r = 3, \delta = 3$ and local code parameters $((n_L = 5, k_L = 3, d = 4), (\alpha = 4, \beta = 1), K_L = 9)$. Code symbols are drawn from the underlying scalar pyramid code, shown in bottom.

V. MBR-Local codes

Two constructions of distance-optimal MBR-local codes will be presented here, the first is an explicit construction with information locality and the second is an existential proof of MBR-local codes with all-symbol locality. The local MBR codes appearing in both constructions are the repair-by-transfer MBR codes contained in [11].

A. MBR-Local Codes with (r, δ) Information Locality

Construction 5.1: The construction proceeds in 3 stages:

Stage 1: Let $N_L = \binom{n_L}{2}$ and $\Delta_L = N_L - K_L + 1$. A pyramid code $\mathcal{A}$ with (K_L, Δ_L) information locality and having parameters $[mN_L + \Delta\alpha, mK_L, \Delta_L + \Delta\alpha]$ is first constructed that is composed of $\Delta\alpha$ global parity symbols and m support-disjoint local $[N_L, K_L, \Delta_L]$ MDS codes $\{\mathcal{A}_i\}_{i=1}^m$.

Stage 2: The N_L MDS-coded symbols corresponding to the i^{th} local code $\mathcal{A}_i$ are then used to construct a repair-by-transfer $(n_L, r, d, (\alpha, \beta = 1), K_L)$ MBR code where $n_L = r + \delta - 1$, $\alpha = d = n_L - 1$, $K_L = r\alpha - \binom{r}{2}$.

Stage 3: Finally, the $\Delta\alpha$ global parities are distributed amongst Δ code symbols, each code symbol corresponding to α symbols over $\mathbb{F}_q$.

An example construction is presented in Fig. 1.

Theorem 5.2: The code in Construction 5.1 has length $n = mn_L + \Delta$, scalar dimension $K = mK_L$ and is distance-optimal.

Sketch of Proof: An upper bound on $d_{\min}$ of the code $\mathcal{C}$ obtained via Construction 5.1 is given by (13) (since $K_L | K$) and for the parameters of the code, this simplifies to $d_{\min} \leq \delta + \Delta$. We then show that any pattern of $\delta + \Delta - 1$ erasures can be corrected by the code. Since pyramid codes are optimal scalar codes with information locality, we obtain using (2) that the scalar code $\mathcal{A}$ employed in Construction 5.1 has minimum distance given by $D_{\min} = \Delta_L + \Delta\alpha = \Delta\alpha + \binom{\delta-1}{2} + 1$. We note that at least $\delta - 1$ code symbols out of any pattern of $\delta + \Delta - 1$ erased code symbols, arise from the union of the local codes. It can be argued, using the sub-additivity of the P function for MBR codes, that the maximum number of scalar symbols (of code $\mathcal{A}$) lost upon erasing $\delta - 1$ code symbols from the union of the local codes is $\binom{\delta-1}{2}$. Hence it will follow that the number of scalar symbols of $\mathcal{A}$ that are erased by $\delta + \Delta - 1$ erasures in $\mathcal{C}$, is at most $D_{\min} - 1$. Hence the pyramid code, $\mathcal{A}$, can recover from this many erasures and hence, so can the vector code $\mathcal{C}$. ■

B. Existence of MBR-Local Codes with All-Symbol Locality

Theorem 5.3: Consider an optimal $[mN_L, \ell K_L, d_{\min}], \ell \leq m$ scalar code $\mathcal{A}$ with (K_L, Δ_L) all-symbol locality, where $\Delta_L = N_L - K_L + 1$. The existence of such a code having support disjoint local MDS codes is known for the parameters assumed here (see Theorems 5, 9 in [12]). Using $\mathcal{A}$ in the Stage 1 and 2 of Construction 5.1 results in an (r, δ) all-symbol MBR locality code. The code has $n = mn_L$ and scalar dimension $K = \ell K_L$ and is distance-optimal.

VI. Performance Comparison

In Fig. 2, the performance of a representative set of codes having common length $n = 60$ and common minimum distance $d_{\min} = 8$ of various classes of codes (obtained via both explicit constructions and existential arguments) discussed above are plotted, for the case of a single node failure. In the plots, the X-axis denotes the storage overhead, given by $\frac{n\alpha}{K}$, where K is the file size. In the first plot, the Y-axis denotes the normalized repair bandwidth, calculated as $\frac{n\omega}{K}$, where ω denotes the average repair bandwidth for repairing a single node, assuming i.i.d node failures. When n is large, the number of failures in the system is proportional to n assuming a Poisson-failure model and thus the repair-bandwidth cost per unit time per unit file size is proportional to $\frac{n\omega}{K}$. The Y-axis in the second plot denotes the average number of nodes accessed during repair. Note that for similar values of storage overhead, codes with local regeneration have moderate values of both repair bandwidth and access, while regenerating codes have high access and low repair bandwidth, and scalar codes with locality have high repair bandwidth and low access.

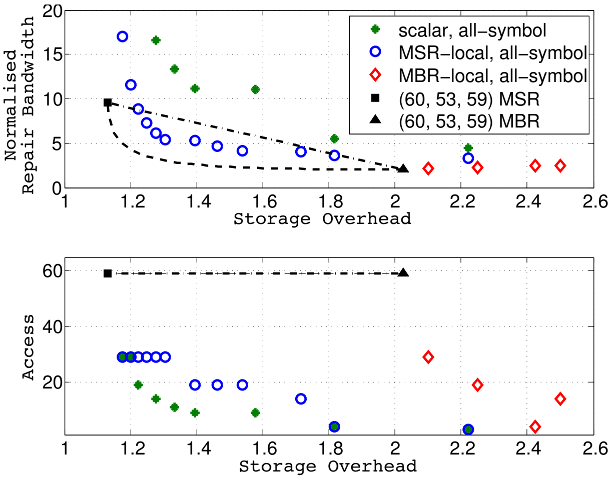

Fig. 2. The performance of various code constructions presented in this paper as well as that of regenerating codes, all having common length 60 and minimum distance 8 are plotted. Taken together, the two plots permit a comparison of the various codes in terms of normalized repair bandwidth, storage overhead and access (i.e., repair degree).

Acknowledgements

This research is supported in part by the National Science Foundation under Grant 0964507 and in part by the NetApp Faculty Fellowship program. The work of V. Lalitha is supported by a TCS Research Scholarship.

References

[1] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in windows azure storage," in *Proc. USENIX Annual Technical Conference (ATC)*, Boston, MA, 2012, pp. 15–26.

[2] "Hadoop." [Online]. Available: http://hadoop.apache.org

[3] D. Borthakur, R. Schmit, R. Vadali, S. Chen, and P. Kling, "HDFS RAID," *Tech talk. Yahoo Developer Network*, 2010.

[4] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[5] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the Locality of Codeword Symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, Nov. 2012.

[6] C. Suh and K. Ramchandran, "Exact-repair MDS code construction using interference alignment," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1425–1442, Mar. 2011.

[7] K. V. Rashmi, 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," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5227–5239, Aug. 2011.

[8] D. S. Papailiopoulos, A. G. Dimakis, and V. R. Cadambe, "Repair Optimal Erasure Codes through Hadamard Designs," 2011. [Online]. Available: arXiv:1106.1634

[9] I. Tamo, Z. Wang, and J. Bruck, "Zigzag Codes: MDS Array Codes with Optimal Rebuilding," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1597–1616, Mar. 2013.

[10] V. R. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and Changho Suh, "Asymptotic Interference Alignment for Optimal Repair of MDS codes in Distributed Data Storage," 2012. [Online]. Available: http://www.mit.edu/ viveck/resources/Research/asymptotic_storage.pdf

[11] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Distributed Storage Codes With Repair-by-Transfer and Nonachievability of Interior Points on the Storage-Bandwidth Tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.

[12] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, Jul. 2012, pp. 2776–2780.

[13] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, Jul. 2012, pp. 2771–2775.

[14] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with Local Regeneration," 2012. [Online]. Available: arXiv:1211.1932

[15] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Proc. Int. Symp. Netw. Comp. and Applications (NCA)*, 2007, pp. 79–86.

[16] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: Novel erasure codes for big data," 2013. [Online]. Available: arXiv:1301.3791

[17] N. Silberstein, A. S. Rawat, and S. Vishwanath, "Error resilience in distributed storage via rank-metric codes," in *Proc. 50th Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, Urbana-Champaign, IL, Oct. 2012, pp. 1150 –1157.

[18] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," Oct. 2012. [Online]. Available: arXiv:1210.6954

[19] Mario Blaum, Patrick G. Farrell, and Henk C.A. van Tilborg, "Array Codes," *Handbook of Coding Theory*, vol. 2, pp. 1855–1909, 1998.