Error Correction Based on Partial Information

Itzhak Tamo, *Member, IEEE*, Min Ye, and Alexander Barg, *Fellow, IEEE*

***Abstract*— We consider the decoding of linear and array codes from errors when we are only allowed to download a part of the codeword. More specifically, suppose that we have encoded k data symbols using an (n, k) code with code length n and dimension k. During storage, some of the codeword coordinates might be corrupted by errors. We aim to recover the original data by reading the corrupted codeword with a limit on the transmission bandwidth, namely, we can only download an α proportion of the corrupted codeword. For a given α, our objective is to design a code and a decoding scheme such that we can recover the original data from the largest possible number of errors. A naive scheme is to read αn coordinates of the codeword. This method used in conjunction with MDS codes guarantees recovery from any $\lfloor (\alpha n - k)/2 \rfloor$ errors. In this paper we show that we can instead download an α proportion from each of the codeword's coordinates. For a well-designed MDS code, this method can guarantee recovery from $\lfloor (n - k/\alpha)/2 \rfloor$ errors, which is $1/\alpha$ times more than the naive method, and is also the maximum number of errors that an (n, k) code can correct by downloading only an α proportion of the codeword. We present two families of such optimal constructions and decoding schemes of which one is based on Interleaved Reed-Solomon codes and the other on Folded Reed-Solomon codes. We further show that both code constructions attain asymptotically optimal list decoding radius when downloading only a part of the corrupted codeword. We also construct an ensemble of random codes that with high probability approaches the upper bound on the number of correctable errors when the decoder downloads an α proportion of the corrupted codeword.**

***Index Terms*— Distributed storage, α-decoding radius, MDS codes, random coding, Reed-Solomon codes.**

I. Introduction

RECOVERY of information under limitations on the repair bandwidth has received signification attention in information theory literature. In particular, a well-known approach to enhance resilience of distributed storage systems against failures of storage disks relies on Maximum Distance Separable (MDS) codes which are optimal in terms of the redundancy-reliability tradeoff. More specifically, an MDS code with r parity symbols can recover the original data from any r erasures of the codeword coordinates. In practice, single disk failure is the most common scenario. Upon observing this, Dimakis *et al.* [2] introduced the concept of repair bandwidth, which is the minimum possible amount of data one needs to download in order to recover any single node failure. An MDS code with optimal (minimum) repair bandwidth is called Minimum Storage Regenerating (MSR) code. In the low rate regime, Rashmi et al. gave an explicit construction of MSR codes [3]. Constructions of optimal-repair regenerating codes with no limitations on the code rate were given in several works of the authors [4]–[7]. Guruswami and Wootters studied the repair bandwidth of Reed-Solomon (RS) codes [8]. Constructions of RS codes with optimal repair bandwidth were given in [9]–[11].

In this paper we consider the problem of decoding linear and array codes from errors when we are allowed to rely only on a part of the corrupted codeword. Before proceeding to a more detailed description of this problem, we first provide one possible application to motivate it. Suppose that the code is deployed in a wireless system, for instance, in low-power wide-area networks (LP-WAN) or narrow-band Internet of Things, wherein the links between the nodes are prone to errors that arise because of physical separation and energy constraints. An associated and natural constraint in such systems is the requirement to transmit as little information as possible to the data collector (whose goal is to decode the codeword). The information received by the collector may therefore be corrupted by errors either because of unreliable storage devices or of the noisy links. Moreover, bandwidth considerations may require to limit communication to only a part of the codeword that encodes the data. This motivates the problem of error correction under communication constraints called here fractional decoding.

More precisely, if we encode the original data using an (n, k) MDS code with code length n and dimension k, it is well known that we can recover the original data from any $\lfloor (n - k)/2 \rfloor$ errors when we receive the whole codeword. In a distributed system, reading the whole codeword requires certain amount of disk I/Os and transmission bandwidth. Now suppose that we have a limit on the bandwidth and we can only download an $\alpha < 1$ proportion of the whole codeword; a natural question then is how many errors we can guarantee to correct in this setup. Rephrasing this question, we are interested how much of the error correcting capability is sacrificed by reducing the transmission bandwidth.

Similarly to the study of MSR codes, we also resort to *array codes* [12]. An (n, k, l) array code $\mathcal{C}$ over a finite field F

Manuscript received September 9, 2018; revised May 30, 2019; accepted November 13, 2019. Date of publication November 19, 2019; date of current version February 14, 2020. The work of I. Tamo was supported in part by the ISF under Grant 1030/15 and in part by the NSF-BSF under Grant 2015814. The work of M. Ye was supported by the NSF under Grant CCF1422955. The work of A. Barg was supported by the NSF under Grant CCF1814487, Grant CCF1618603, and Grant CCF1422955. This article was presented at the 2017 IEEE International Symposium on Information Theory.

I. Tamo is with the Department of EE-Systems, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: zactamo@gmail.com).

M. Ye is with the Data Science and Information Technology Research Center, Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518055, China (e-mail: yeemmi@gmail.com).

A. Barg is with the Department of ECE and ISR, University of Maryland, College Park, MD 20742 USA, and also with the IITP, Russian Academy of Sciences, 127051 Moscow, Russia (e-mail: abarg@umd.edu).

Communicated by A. Rudra, Associate Editor for Complexity.

Digital Object Identifier 10.1109/TIT.2019.2954409

maps an $l \times k$ data matrix $(D_1, \ldots, D_k) \in (F^l)^k$ to an $l \times n$ codeword matrix $(C_1, \ldots, C_n) \in (F^l)^n$. Each column C_i of the matrix is a codeword coordinate. The distance between two codewords is measured by the Hamming metric in a usual way: $d(C^{(1)}, C^{(2)}) = |\{i : C_i^{(1)} \neq C_i^{(2)}\}|$. Motivated by the works in distributed storage, we assume that each coordinate is stored on a separate node in the system, and we call the parameter l that determines the dimension of the column vector C_i *sub-packetization*. Note that a scalar code can also be viewed as an array code with $l = 1$.

Definition I.1 (Fractional decoding and α-decoding radius)**.** *Consider an (n, k, l) array code $\mathcal{C} = \{(C_1, \ldots, C_n)\}$ over F, where $C_i \in F^l, i = 1, \ldots, n$.*

(i) We say that $\mathcal{C}$ can correct up to t errors by downloading an α proportion of the codeword if there exist $n+1$ functions $f_i : F^l \to F^{\alpha_i l}, i = 1, 2, \ldots, n$ with $\sum_{i=1}^n \alpha_i \leq n\alpha$ and $g : F^{(\sum_{i=1}^n \alpha_i)l} \to F^{nl}$ such that

$$\begin{aligned} &g(f_1(C_1 + E_1), f_2(C_2 + E_2), \ldots, f_n(C_n + E_n)) \\ &= (C_1, C_2, \ldots, C_n) \end{aligned} \tag{1}$$

for any codeword $(C_1, \ldots, C_n) \in \mathcal{C}$ and any error vector $(E_1, E_2, \ldots, E_n)$ of Hamming weight $|\{i : E_i \neq 0\}| \leq t$.

(ii) For $\alpha \geq k/n$, define the α-decoding radius $r_\alpha(\mathcal{C})$ as the maximum number of errors that the code $\mathcal{C}$ can correct by downloading an α-proportion of the codeword.

(iii) For $\alpha \geq k/n$, we further define the α-decoding radius of (n, k) codes as

$$r_\alpha(n, k) = \max_{\mathcal{C} \in \mathcal{M}_{n,k}} r_\alpha(\mathcal{C}),$$

where $\mathcal{M}_{n,k}$ is the set of all (n, k) codes.

Remark I.2. *Since the information contents of the codeword C is kl symbols of the field F, the inequality $\alpha \geq k/n$ forms a trivial necessary condition for decoding even without errors. This condition will be assumed throughout the paper.*

It is well known that for any (n, k) code $\mathcal{C}$, we have $r_1(\mathcal{C}) \leq \lfloor (n-k)/2 \rfloor$, and the equality holds for MDS codes. Thus $r_1(n, k) = \lfloor (n-k)/2 \rfloor$. Moreover, we have an obvious lower bound for an MDS code $\mathcal{C}$:

$$r_\alpha(\mathcal{C}) \geq \lfloor (\alpha n - k)/2 \rfloor. \tag{2}$$

To see this, we can simply read any αn coordinates of the codeword. Since a punctured MDS code is still an MDS code with the same dimension, we obtain the lower bound (2).

In this paper, we show that

$$r_\alpha(n, k) = \lfloor (n - k/\alpha)/2 \rfloor \tag{3}$$

for any n, k and α, and we give two families of explicit constructions of MDS codes together with the decoding schemes which achieve the optimal α-decoding radius in (3). The optimal α-decoding radius in (3) improves upon the lower bound (2) obtained from the naive decoding strategy by a factor of $1/\alpha$.

The underlying idea of the two optimal code constructions and decoding schemes is to download from each of the codeword coordinates a number of field symbols that forms an α proportion of the coordinate's size, and to ensure that the downloaded symbols constitute a codeword in an $(n, k/\alpha, \alpha l)$ MDS code, which can be used to recover the original data. One of our constructions is based on Interleaved Reed-Solomon codes [13], [14]. In particular, we propose an optimal fractional decoding scheme for Interleaved RS codes. The other construction is based on Folded Reed-Solomon (FRS) codes of Guruswami and Rudra [15]. While FRS codes solve the problem somewhat trivially, our solution based on Interleaved RS code has the advantage of smaller sub-packetization as well as smaller encoding/decoding/fractional decoding complexity.

Furthermore, we show that random codes with high probability asymptotically achieve the bound (3) on the α-decoding radius. The ensemble of random codes that we consider is based on randomly chosen "contracting" linear maps of the coordinates of an MDS code. Finally, we take up the question of constructing MDS codes with optimal repair bandwidth (also called MSR codes) which at the same time have the optimal α-decoding radius. A construction of codes with both these properties is obtained by using an idea in a recent paper [5] by the authors.

The paper is organized as follows. In Section II we prove an upper bound on the α-decoding radius, which we show to be attainable in several ways. Specifically, in Section III we show that random linear mappings are asymptotically optimal for fractional decoding. Subsequently, in Sections IV and V we present the two families of code constructions achieving the upper bound for finite code length. Then in Section V we introduce the notion of α-list decoding capacity, and show that both code constructions achieve it. Finally, in Section VI-B, we present the MSR code construction with optimal α-decoding radius.

II. Upper Bound on the α-Decoding Radius

Theorem II.1. *Let $n \geq \alpha n \geq k$. Then*

$$r_\alpha(n, k) \leq \lfloor (n - k/\alpha)/2 \rfloor. \tag{4}$$

Proof. Let $\mathcal{C}$ be an (n, k, l) code, and let $f_i, i = 1, \ldots, n$ be the mappings that satisfy the conditions in Definition I.1. Consider the "projected" code C^α obtained by applying the functions $f_i, i = 1, \ldots, n$ to the coordinates of the codewords of $\mathcal{C}$:

$$\mathcal{C}^\alpha = \{(f_1(C_1), \ldots, f_n(C_n)) : (C_1, \ldots, C_n) \in \mathcal{C}\}$$

We will argue that the minimum Hamming distance of the code $\mathcal{C}^\alpha$ (as defined above) is at most $n - \lceil \frac{k}{\alpha} \rceil + 1$, implying (4). Suppose otherwise, then the code $\mathcal{C}^\alpha$ corrects any $n - \lceil \frac{k}{\alpha} \rceil + 1$ erasures, i.e., it is possible to recover the codeword from any given subset of $s := \lceil \frac{k}{\alpha} \rceil - 1$ of its coordinates.

Assume w.l.o.g. that $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$. By the assumption, it is possible to recover the codeword from the first s coordinates, i.e., the projection mapping on the first s coordinates is injective, or, rephrasing again, $\sum_{i=1}^s \alpha_i \geq k$. This implies that $\alpha_{s+1} \geq \alpha_s \geq k/s$. With this we obtain

$$\sum_{i=1}^n \alpha_i = \sum_{i=1}^s \alpha_i + \sum_{i=s+1}^n \alpha_i \geq k + (n - s)\frac{k}{s} = \frac{nk}{s} > \alpha n$$

since $s < \frac{k}{\alpha}$. At the same time, by Def. I.1, the sum $\sum_{i=1}^{n} \alpha_i \le \alpha n$, which contradicts the assumption. The proof is complete. □

III. Random Coding Bounds

Here we examine another view of the codes defined above with the aim of estimating the parameters of codes $\mathcal{C}^\alpha$ obtained from MDS codes under a random contracting mapping. To put the arguments in context, recall the construction of concatenated codes which combine two codes, say an $[n,k]$ MDS code $\mathcal{C}_1$ over the finite field $F = \mathbb{F}_{q^l}$ and an $[m,l]$ code $\mathcal{C}_2$ over the field $\mathbb{F}_q$, into a code of length nm over $\mathbb{F}_q$. To transform a codeword $C = (C_1, \dots, C_n)$ of $\mathcal{C}_1$ to the codeword of the concatenated code, each symbol C_i is replaced with a codeword of the code $\mathcal{C}_2$ using some injective map from $\mathbb{F}_{q^l}$ to $\mathcal{C}_2$. Thereby, the number of coordinates in the q-ary representation of C_i is increased from l to m. In our current situation, we are interested in the code obtained by mapping the coordinate C_i to an element in the field $\mathbb{F}_{q^{\alpha l}}$, where $\alpha < 1$ (Definition I.1 considers a slightly more general case wherein α depends on i, while the constructions in the next sections assume equal α_i's). Thus, codes for fractional decoding may be viewed as "inverse concatenation codes" which shrink the dimension of each coordinate of the original codes instead of expanding it.

This point of view suggests an approach to random coding bounds similar to the earlier results on concatenated codes e.g., [16]. Namely, we start with an $[n,k]$ MDS code $\mathcal{C}_1$ over the field $\mathbb{F}_{q^l}$ and map each coordinate to an element in $\mathbb{F}_{q^{\alpha l}}$ using a uniformly random linear mapping. Specifically, suppose that $A = (A_1, \dots, A_n)$ is an n-tuple of linear maps $\mathbb{F}_{q^l} \to \mathbb{F}_{q^{\alpha l}}$ and let

$$\begin{aligned}\mathcal{C}^\alpha &:= A(\mathcal{C}_1)\\ &= \{(A_1(C_1), \dots, A_n(C_n)) : (C_1, \dots, C_n) \in \mathcal{C}_1\}\end{aligned}$$

be the resulting linear code. In this section we compute the typical parameters of the code $\mathcal{C}^\alpha$, which will be shown to meet or approach the bound (3) with high probability. We consider two different asymptotic regimes, of fixed n and $l \to \infty$, and of $n = q^l \to \infty$, with the above conclusion applying to both of them.

We will call the mapping A *optimal* for the fractional decoding of $\mathcal{C}_1$ if for every subset $\mathcal{I} \subset [n]$ of size $L = k/\alpha + 1$, the restriction of A to $\mathcal{I}$ defined as

$$\begin{aligned}A_{\mathcal{I}} : \qquad & \mathcal{C}_1 \to \left(\mathbb{F}_{q^{\alpha l}}\right)^L\\ & (C_1, \dots, C_n) \mapsto (A_i(C_i), i \in \mathcal{I})\end{aligned} \tag{5}$$

is injective. Recalling Definition I.1 and the bound (3), if A is optimal, then the code $\mathcal{C}^\alpha$ corrects $n - L$ erasures, and so its distance equals $n - k/\alpha$. Suppose that $A = (A_1, \dots, A_n)$ is realized by random $l \times \alpha l$ matrices A_i whose elements are chosen from $\mathbb{F}_q$ independently and with uniform distribution.

Before proceeding, recall the following classic fact about the weight distribution of an $[n,k]$ MDS code $\mathcal{C}_1$ over $\mathbb{F}_{q^l}$:

$$|\{C \in \mathcal{C}_1 : \mathrm{wt}(C) = i\}| \le \binom{n}{i} q^{l(i-n+k)}, \quad i \ge n-k+1.$$

Indeed, the restriction of $\mathcal{C}_1$ to any k coordinates is injective. Once we fix $n-i$ coordinates to 0 in any of the possible $\binom{n}{i}$ ways, there are $(q^l-1)^{k-(n-i)}$ possible choices of nonzero coordinates before the codeword is identified uniquely. This gives the claimed upper bound.

Proposition III.1. *Let $\mathcal{C}_1$ be an $[n,k]$ MDS code over the field $\mathbb{F}_{q^l}$. Let $\alpha > k/n$ and let $A : \mathcal{C}_1 \to \mathcal{C}^\alpha$ be the random linear mapping defined above. Suppose that n, k are fixed and $l \to \infty$, then A is an optimal mapping for the fractional decoding of $\mathcal{C}_1$ with probability $1 - o(1)$.*

Proof. Let $C = (C_1, \dots, C_n), C \neq 0$ be a codeword of $\mathcal{C}_1$ and suppose its Hamming weight is $\mathrm{wt}(C) = w$. Since $A = (A_1, \dots, A_n)$ is linear, $A_i(C_i) = 0$ if $C_i = 0$ and $\Pr(A_i(C_i) = 0) = q^{-\alpha l}$ if $C_i \neq 0$. Therefore

$$\Pr(A(C) = 0) = q^{-\alpha w l}.$$

Observe that for any subset $\mathcal{I} \subset [n]$ of size $L > k$, the code $\mathcal{C}_1$ restricted to the coordinates in $\mathcal{I}$ is an $[L,k]$ MDS code. Now let us fix a subset $\mathcal{I} \subseteq [n]$ of size $L > k/\alpha$ and show that the mapping $A : \mathcal{C}_1 \to \mathcal{C}^\alpha$ with high probability has a trivial kernel. We have

$$\begin{aligned}\Pr(\ker(A_{\mathcal{I}}) \neq 0\}) &\le \sum_{C \in \mathcal{C}_1, C \neq 0} \Pr(A_{\mathcal{I}}(C) = 0)\\ &= \sum_{w=L-k+1}^{L} \sum_{\mathrm{wt}(C)=w} \Pr(A_{\mathcal{I}}(C) = 0)\\ &\le \sum_{w=L-k+1}^{L} \binom{L}{w} q^{l(w-L+k)} q^{-\alpha w l}\\ &= \sum_{w=L-k+1}^{L} \binom{L}{w} q^{l(w-\alpha w-L+k)}. \end{aligned} \tag{6}$$

The exponent in the last expression, given by $w - \alpha w - L + k$, is an increasing function of w, so $w - \alpha w - L + k \le k - \alpha L < 0$ for all $w \le L$. Therefore $q^{l(w-\alpha w-L+k)} \to 0$ for all $w \le L$ when $l \to \infty$, and thus $\Pr(\ker(A_{\mathcal{I}}) \neq 0\}) \to 0$ for every subset $\mathcal{I} \subseteq [n]$ of size $L > k/\alpha$. Since there are only finitely many such subsets, we conclude that with probability approaching one, the mapping $A_{\mathcal{I}}$ is injective for every choice of $\mathcal{I}$. This completes the proof of the proposition. □

Now let us analyze the case when the code length $n = q^l \to \infty$. In this case it is more convenient to consider asymptotic optimality of the mapping A. Given an $[n, k = Rn]$ MDS code $\mathcal{C}_1$ over the field $\mathbb{F}_{q^l}$ and a linear mapping $A : \mathcal{C}_1 \mapsto \mathcal{C}^\alpha$, we call A *asymptotically optimal* for the fractional decoding of $\mathcal{C}_1$ if the following two conditions are satisfied:

1) A is injective;
2) the distance of the code $\mathcal{C}^\alpha$ satisfies $d(\mathcal{C}^\alpha) \ge n(1 - R/\alpha - o(1))$.

In other words, the mapping A is asymptotically optimal if the cardinality of the code C^α is unchanged from that of $\mathcal{C}_1$, and its relative distance asymptotically satisfies the bound (3).

Proposition III.2. *Let $\mathcal{C}_1$ be an $[n,k]$ MDS code over $\mathbb{F}_{q^l}$, where $n = q^l$ and $k = Rn$. Let $A = (A_1, \dots, A_n)$ be the random linear mapping $\mathcal{C}_1 \to \mathcal{C}^\alpha$ defined above, where*

$\alpha > R$. Suppose that R is fixed and $n \to \infty$,[1] then A is an asymptotically optimal mapping for the fractional decoding of $\mathcal{C}_1$ with probability $1 - o(1)$.

Proof. Let us prove the injectivity condition. Proceeding as in (6), we have

$$\begin{aligned}\Pr(\ker(A) \neq 0\}) &\leq \sum_{C \in \mathcal{C}_1, C \neq 0} \Pr(A(C) = 0) \\ &= \sum_{w=n-k+1}^{n} \sum_{\mathrm{wt}(C)=w} \Pr(A(C) = 0) \\ &\leq \sum_{w=n-k+1}^{n} \binom{n}{w} q^{l(w-n+k)} q^{-\alpha w l} \\ &= q^{-nl(1-R)} \sum_{w=n-k+1}^{n} \binom{n}{w} q^{wl(1-\alpha)} \\ &\leq q^{-nl(1-R)} \sum_{w=0}^{n} \binom{n}{w} q^{wl(1-\alpha)} \\ &= q^{-nl(1-R)} (1 + q^{l(1-\alpha)})^n \\ &= (q^{-l(1-R)} + q^{-l(\alpha-R)})^n \to 0.\end{aligned}$$

This shows that the mapping A is injective with probability $1 - o(1)$.

Next we prove that with probability $1 - o(1)$ the distance $d(\mathcal{C}^\alpha)$ satisfies

$$d(\mathcal{C}^\alpha) \geq n - \frac{k}{\alpha} - \frac{2n}{\alpha \log_4 n} = n\Big(1 - \frac{R}{\alpha} - o(1)\Big). \quad (7)$$

Starting with a nonzero codeword $C \in \mathcal{C}_1$ of weight $\mathrm{wt}(C) = w$, let us estimate the probability that it maps on a codeword of $\mathcal{C}^\alpha$ of weight no larger than i for some $i \leq w$:

$$\Pr(\mathrm{wt}(A(C)) \leq i) \leq \binom{w}{i} q^{-\alpha l(w-i)}.$$

By the union bound,

$$\begin{aligned}\Pr(d(C^\alpha) \leq i) &\leq \Pr(\{\exists C \in \mathcal{C}_1 : 1 \leq \mathrm{wt}(A(C)) \leq i\}) \\ &\leq \sum_{C \in \mathcal{C}_1, C \neq 0} \Pr(\mathrm{wt}(A(C)) \leq i) \\ &= \sum_{w=n-k+1}^{n} \sum_{\mathrm{wt}(C)=w} \Pr(\mathrm{wt}(A(C)) \leq i) \\ &\leq \sum_{w=n-k+1}^{n} \binom{n}{w} q^{l(w-n+k)} \binom{w}{i} q^{-\alpha l(w-i)} \\ &= \sum_{w=n-k+1}^{n} \binom{n}{w} \binom{w}{i} n^{w-\alpha w-n+k+\alpha i} \\ &\overset{(a)}{\leq} \sum_{w=n-k+1}^{n} 4^n n^{-\alpha n+k+\alpha i} \\ &\leq k 4^{n+(k-\alpha n+\alpha i)\log_4 n},\end{aligned}$$

where inequality (a) follows from the facts that $\binom{n}{w} \leq 2^n$, $\binom{w}{i} \leq 2^n$, and $w - \alpha w - n + k + \alpha i < -\alpha n + k + \alpha i$ for all $w \leq n$. Thus if $i = n - \frac{k}{\alpha} - \frac{2n}{\alpha \log_4 n}$, then

$$\Pr(d(C^\alpha) \leq i) \leq 4^{-n} k \to 0$$

when $n \to \infty$. This implies (7) and concludes the proof. □

Concluding this section, we note a difference between the results for classic binary concatenated codes [16] and the results above. In the former case, symbols of the MDS code are mapped on random binary codewords, and the resulting code with high probability approaches the Gilbert-Varshamov bound, matching the best known parameters for the binary case (under some additional assumption on the component codes, derived in [16].) In our case, the alphabet size of the resulting code $\mathcal{C}^\alpha$ is allowed to grow, and the rate and distance of $\mathcal{C}^\alpha$ are as good as those obtained from MDS codes in a deterministic way in the next two sections.

IV. Optimal Fractional Decoding Scheme for Interleaved RS Codes

In this section we propose an optimal fractional decoding scheme for Interleaved RS codes. This code family as well as the other constructions in this paper derive from the standard RS codes, defined as follows.

Definition IV.1. *A* Reed-Solomon code $\mathrm{RS}_G(n, k, \Omega) \subseteq G^n$ *of dimension k over a field G with evaluation points $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \subseteq G$ is the set of vectors*

$$\{(h(\omega_1), \ldots, h(\omega_n)) \in G^n : h \in G[x], \deg h \leq k-1\}.$$

The idea behind the concept of Interleaved RS codes is as follows. Suppose that the points ω_i in the above definition are elements of a subfield F of G such that $[G : F] = l$. By expanding the coefficients of the polynomial h into vectors over F, we can view the codevector $(h(\omega_1), \ldots, h(\omega_n))$ as l vectors of an RS code over F. This motivates the following definition.

Definition IV.2. *An* Interleaved Reed-Solomon code $\mathrm{IRS}_F(n, k, l, \Omega) \subseteq (F^l)^n$ *is an (n, k, l) array code consisting of l independent RS codes with the same set of evaluation points $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \subseteq F$; see* (8) *at the bottom of the next page.*

Note that a codeword $C \in \mathrm{IRS}_F(n, k, l, \Omega)$ is completely determined by the choices of the polynomials $h_1, \ldots, h_l$. Therefore, we write a codeword in an IRS code as $C(h_1, \ldots, h_l)$. We assume throughout that α is rational, noting that this constraint does not incur any loss of generality in terms of the code parameters.

Proposition IV.3. *Let $\alpha = m/l < 1$, where m and l are positive integers. Given n and k satisfying that $n \geq kl/m$ and $m|k$, and given a finite field F with $|F| \geq n$, the Interleaved RS code $\mathrm{IRS}_F(n, k, l, \Omega)$ with the evaluation points $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \subseteq F$ has the optimal α-decoding radius.*

The proof is given in the remainder of this section. Let $A_1, A_2, \ldots, A_m \subseteq F$ be m pairwise disjoint subsets of the

[1] It doesn't matter how q and l scale as long as $n = q^l \to \infty$. In the proof, we only use the condition $n = q^l \to \infty$, and we do not use any property of q and l themselves.

field F, each of size k/m. For $j = 1, 2, \ldots, m$, define the annihilator polynomial of the set A_j to be

$$p_j(x) = \prod_{\omega \in A_j} (x - \omega).$$

We further define m polynomials $g_1, \ldots, g_m$:

$$g_j(x) = h_{l-m+j}(x)(p_j(x))^{l-m} + \sum_{u=1}^{l-m} h_u(x)(p_j(x))^{u-1}, \quad j = 1, 2, \ldots, m. \tag{10}$$

It is clear that $(g_1(\omega_i), g_2(\omega_i), \ldots, g_m(\omega_i))$ can be calculated from $(h_1(\omega_i), h_2(\omega_i), \ldots, h_l(\omega_i))$. Our strategy is to download the m-dimensional vector $(g_1(\omega_i), g_2(\omega_i), \ldots, g_m(\omega_i))$ from the ith codeword coordinate, which is exactly an m/l proportion of the codeword. In other words, for a codeword $C = C(h_1, \ldots, h_l) \in \text{IRS}_F(n, k, l, \Omega)$, we download the $m \times n$ matrix in (9) at the bottom of this page.

Since $\deg(p_j) = k/m$, we have $\deg(g_j) < kl/m$. Thus $g(C) \in \text{IRS}_F(n, kl/m, m, \Omega)$. As a result, we can recover all the coefficients of polynomials $\{g_j(x)\}_{j=1}^m$ as long as there are no more than $\lfloor (n - kl/m)/2 \rfloor$ errors in the received vector. Now we only need to show that given polynomials $\{g_j(x)\}_{j=1}^m$, we can recover the polynomials $\{h_j(x)\}_{j=1}^l$. To see this, we notice that for $j = 1, 2 \ldots, m$,

$$g_j(\omega) = h_1(\omega) \text{ for all } \omega \in A_j.$$

Consequently, we know the evaluations of $h_1(x)$ at all the points in $\cup_{j=1}^m A_j$. There are k distinct points in the set $\cup_{j=1}^m A_j$ and the degree of $h_1(x)$ is less than k, so we can recover $h_1(x)$. From $h_1(x)$ and $\{g_j(x)\}_{j=1}^m$, we can calculate the polynomials

$$g_j^{(1)}(x) = \frac{g_j(x) - h_1(x)}{p_j(x)} = h_{l-m+j}(x)(p_j(x))^{l-m-1} + \sum_{u=2}^{l-m} h_u(x)(p_j(x))^{u-2}$$

for $j = 1, 2, \ldots, m$. Since

$$g_j^{(1)}(\omega) = h_2(\omega) \text{ for all } \omega \in A_j,$$

we know the evaluations of $h_2(x)$ at all the points in $\cup_{j=1}^m A_j$, so we can also recover $h_2(x)$. From $h_1(x), h_2(x)$ and $\{g_j(x)\}_{j=1}^m$, we can calculate the polynomials

$$g_j^{(2)}(x) = \frac{g_j^{(1)}(x) - h_2(x)}{p_j(x)} = h_{l-m+j}(x)(p_j(x))^{l-m-2} + \sum_{u=3}^{l-m} h_u(x)(p_j(x))^{u-3}$$

for $j = 1, 2, \ldots, m$. Since

$$g_j^{(2)}(\omega) = h_3(\omega) \text{ for all } \omega \in A_j,$$

we know the evaluations of $h_3(x)$ at all the points in $\cup_{j=1}^m A_j$, so we can also recover $h_3(x)$. It is clear that we can repeat this procedure until we recover $\{h_j(x)\}_{j=1}^{l-m}$. Then the polynomials $\{h_{l-m+j}(x)\}_{j=1}^m$ can be easily recovered by

$$h_{l-m+j}(x) = \frac{g_j(x) - \sum_{u=1}^{l-m} h_u(x)(p_j(x))^{u-1}}{(p_j(x))^{l-m}}, \quad j = 1, \ldots, m.$$

This shows that we can recover the polynomials $\{h_j(x)\}_{j=1}^l$ from the polynomials $\{g_j(x)\}_{j=1}^m$, and consequently recover the original codeword.

A. Advantages of Interleaved RS Codes Over Naive MDS Code Constructions

There is a much simpler way to give MDS code constructions achieving the optimal α-decoding radius. Indeed, we can take any (nl, kl) scalar MDS code over a finite field F with size $|F| \geq nl$ and group together blocks of l coordinates of it into a vector in F^l. It is clear that in this way we obtain an (n, k, l) MDS array code $\tilde{\mathcal{C}}$. Moreover, by reading αl symbols of F from each of the coordinates of $\tilde{\mathcal{C}}$ we obtain an $(n, k/\alpha, \alpha l)$ MDS array code $\mathcal{C}^\alpha$ which can correct up to $\lfloor (n - k/\alpha)/2 \rfloor$ errors, and thus $\mathcal{C}$ forms an optimal code for fractional decoding.

The main advantage of Interleaved RS codes over this naive MDS code construction (as well as the Folded Reed-Solomon (FRS) code construction considered in Section V) is that Interleaved RS codes require a smaller field size ($|F| \geq n$ for IRS compared to $|F| \geq nl$ for the other two options). Note that the encoding and decoding procedures, as well as the fractional decoding procedure, are all performed over the underlying finite field F. Therefore, the smaller field size leads to lower complexity of all these operations.

$$\text{IRS}_F(n, k, l, \Omega) := \left\{ \left(\begin{pmatrix} h_1(\omega_1) \\ h_2(\omega_1) \\ \vdots \\ h_l(\omega_1) \end{pmatrix}, \begin{pmatrix} h_1(\omega_2) \\ h_2(\omega_2) \\ \vdots \\ h_l(\omega_2) \end{pmatrix}, \ldots, \begin{pmatrix} h_1(\omega_n) \\ h_2(\omega_n) \\ \vdots \\ h_l(\omega_n) \end{pmatrix} \right) : h_i \in F[x], \deg h_i < k \ \forall 1 \leq i \leq l \right\} \tag{8}$$

$$g(C) := \left\{ \left(\begin{pmatrix} g_1(\omega_1) \\ g_2(\omega_1) \\ \vdots \\ g_m(\omega_1) \end{pmatrix}, \begin{pmatrix} g_1(\omega_2) \\ g_2(\omega_2) \\ \vdots \\ g_m(\omega_2) \end{pmatrix}, \ldots, \begin{pmatrix} g_1(\omega_n) \\ g_2(\omega_n) \\ \vdots \\ g_m(\omega_n) \end{pmatrix} \right) : \deg g_i < kl/m \ \forall 1 \leq i \leq m \right\} \tag{9}$$

Another advantage is that IRS code allows for very efficient decoding from random errors. In Section 2.3.2 of [17], Devet et al. described a "linear" variant of the Cohn-Heninger algorithm [18] that can uniquely decode $\mathrm{IRS}_F(n,k,l,\Omega)$ from random errors with high probability when the number of errors does not exceed $\frac{l}{l+1}(n-k)$ (this error correction radius was also attained earlier in [13], albeit with inferior running time). This algorithm runs extremely fast in practice. Notice that in the fractional decoding procedure of IRS codes, we need to decode another IRS code $g(C)$ in (9). Therefore, for both standard decoding and fractional decoding, we can use this fast decoding algorithm for IRS codes if the errors are random, which is the case for most applications in practice. Note that a probabilistic fractional decoding procedure of Interleaved Reed-Solomon codes, following in the footsteps of [14], was recently proposed in [19].

V. Folded Reed-Solomon Codes

Folded RS (FRS) codes were introduced by Guruswami and Rudra [15] for the problem of optimal list decoding. In this section we show that FRS codes are optimal for the fractional decoding in a rather straightforward way.

Let us recall the definition of FRS codes.

Definition V.1. *Let F be a finite field with cardinality $|F| > nl$. Let γ be a primitive element of F. A Folded Reed-Solomon code* $\mathrm{FRS}(n,k,l) \subseteq (F^l)^n$ *is an MDS array code with each codeword coordinate being a vector in F^l defined as follows:*

$$\big\{(C_1, C_2, \ldots, C_n) : C_i = (h(\gamma^{(i-1)l}), h(\gamma^{(i-1)l+1}), \ldots, h(\gamma^{(i-1)l+l-1})) \in F^l \text{ for } 1 \le i \le n, h \in F[x], \deg h \le kl-1\big\}.$$

We limit ourselves to those values of sub-packetization l for which αl is an integer.

Proposition V.2. *The α-decoding radius of FRS codes satisfies*

$$r_\alpha(\mathrm{FRS}(n,k,l)) = \lfloor (n - k/\alpha)/2 \rfloor.$$

Proof. We will construct $n+1$ functions $f_i : F^l \to F^{\alpha_i l}, i = 1, 2, \ldots, n$ and $g : F^{(\sum_{i=1}^n \alpha_i)l} \to F^{nl}$ that (I.1). The functions $f_i : F^l \to F^{\alpha l}$ will simply project a symbol on its first αl coordinates, i.e., $f_i = f$, where for $(d_1, d_2, \ldots, d_l) \in F^l$,

$$f((d_1, d_2, \ldots, d_l)) = (d_1, d_2, \ldots, d_{\alpha l}). \tag{11}$$

Thus, the code $\mathcal{C}^\alpha$ is a projection of the code $\mathcal{C}$,

$$\mathcal{C}^\alpha = \big\{(C_1^\alpha, C_2^\alpha, \ldots, C_n^\alpha) = (f(C_1), f(C_2), \ldots, f(C_n)) : (C_1, C_2, \ldots, C_n) \in \mathrm{FRS}(n,k,l)\big\} \tag{12}$$

Equivalently, we can write C^α as

$$C^\alpha = \big\{(C_1^\alpha, C_2^\alpha, \ldots, C_n^\alpha) : C_i^\alpha = (h(\gamma^{(i-1)l}), h(\gamma^{(i-1)l+1}), \ldots, h(\gamma^{(i-1)l+\alpha l-1}) \in F^l \text{ for } 1 \le i \le n, h \in F[x], \deg h \le kl-1\big\}. \tag{13}$$

Since any k/α coordinates of $\mathcal{C}^\alpha$ contain $(k/\alpha)(\alpha l)$ evaluations of the encoding polynomial h with degree less than kl, we can recover h and thus the whole codeword from any k/α coordinates of $\mathcal{C}^\alpha$. We thus conclude that $\mathcal{C}^\alpha$ is an $(n, k/\alpha, \alpha l)$ MDS array code, so it can correct up to $\lfloor (n-k/\alpha)/2 \rfloor$ errors.

If E_i is the error in the ith coordinate of the codeword, we can write $f(C_i + E_i) = f(C_i) + f(E_i)$ for $i = 1, 2, \ldots, n$. Suppose that $(C_1, C_2, \ldots, C_n) \in \mathrm{FRS}(n,k,l)$ and $|\{i : E_i \neq 0\}| \le \lfloor (n - k/\alpha)/2 \rfloor$, then $(f(C_1), f(C_2), \ldots, f(C_n)) \in \mathcal{C}^\alpha$ and $|\{i : f(E_i) \neq 0\}| \le \lfloor (n - k/\alpha)/2 \rfloor$. As a result, we can recover the codeword $(f(C_1), f(C_2), \ldots, f(C_n)) \in \mathcal{C}^\alpha$ and thus recover the encoding polynomial h and finally the codeword $(C_1, C_2, \ldots, C_n) \in \mathrm{FRS}(n,k,l)$ from $(f(C_1 + E_1), f(C_2 + E_2), \ldots, f(C_n + E_n))$. By our definition in (I.1), this shows that $r_\alpha(\mathrm{FRS}(n,k,l)) \ge \lfloor (n - k/\alpha)/2 \rfloor$, and proof is concluded with a reference to the upper bound (4). □

Remark V.3. *Given multiple values $\alpha_1, \alpha_2, \ldots, \alpha_m$, if we choose l in such a way that $\alpha_1 l, \alpha_2 l, \ldots, \alpha_m l$ are all integers, then* $\mathrm{FRS}(n,k,l)$ *achieves the optimal α_i-decoding radius for $1 \le i \le m$ simultaneously.*

VI. Further Observations

A. α-List Decoding Capacity

In this section we extend our study of fractional decoding to the list decoding problem. Under unique decoding, the decoder outputs the correct codeword as long as the received vector is within a certain distance r_u from it. Under list decoding, the decoder finds a list of all codewords that are within a certain distance r_l from the received vector. Denote the size of this list by L. We say that a code corrects r_l errors under list-of-L decoding if sphere of radius r_l centered at any received vector contains at most L codewords.

Complexity considerations suggest that L is a slowly growing function of the code length n (or even a constant). In this paper, following a long line of work in algebraic list decoding, we assume that L is a polynomial function of n. The main result of [15] amounts to stating that (n,k,l) FRS codes of rate $R := k/n$ correct the asymptotically maximum number of errors $r_l = n(1 - R - o(1))$ under lists of polynomial size. It turns out that FRS codes are also optimal under fractional list decoding.

Let us define formally the fractional decoding problem.

Definition VI.1 ((α, L) list decoding radius)**.** *Consider an (n,k,l) array code $\mathcal{C} = \{(C_1, \ldots, C_n)\}$ over F, where $C_i \in F^l, i = 1, \ldots, n$.*

(i) We say that $\mathcal{C}$ corrects up to t errors under list-of-L decoding by downloading an α proportion of the codeword if there exist $n+1$ functions $f_i : F^l \to F^{\alpha_i l}, i = 1, 2, \ldots, n$, $\sum_{i=1}^n \alpha_i \le n\alpha$ and $g : F^{(\sum_{i=1}^n \alpha_i)l} \to (F^{nl})^L$ such that for any codeword $C = (C_1, \ldots, C_n) \in \mathcal{C}$ and any error vector $E = (E_1, E_2, \ldots, E_n)$ of Hamming weight $\le t$, we have

$$g(f_1(C_1 + E_1),\ f_2(C_2 + E_2), \ldots, f_n(C_n + E_n)) = \{C^{(i)}, i = 1, \ldots, L\},\ \text{and } C \in \{C^{(i)}, i = 1, \ldots, L\}. \tag{14}$$

(ii) For $\alpha \ge k/n$, define the (α, L)-list decoding radius $r_{\alpha,L}(\mathcal{C})$ to be the maximum number of errors that the code $\mathcal{C}$

can correct under decoding into a list of size L by downloading an α proportion of the codeword.

(iii) For $\alpha \geq R$, we further define the (normalized) α-list decoding capacity of codes of rate at least R as

$$\rho_\alpha(R) = \sup\Big\{ \frac{r_{\alpha,L}(\mathcal{C})}{n} : \text{rate}(\mathcal{C}) \geq R \text{ and } L \text{ is polynomial in } n \Big\},$$

where $n(\mathcal{C})$ is the code length of $\mathcal{C}$. More formally,

$$\rho_\alpha(R) = \sup_{m\in\mathbb{N}} \limsup_{n\to\infty} \frac{r_{\alpha,n^m}(n, Rn)}{n}$$

where $r_{\alpha,n^m}(n, Rn)$ is the maximum of $r_{\alpha,L}(\mathcal{C})$ over all codes of length n and rate R.

Repeating the proof of Theorem II.1, we can easily show that $\rho_\alpha(R) \leq 1 - R/\alpha$. At the same time, we can show that the two families of RS-type codes shown above to be optimal for α-decoding are also optimal for the fractional list decoding problem in the sense of achieving the α-list decoding capacity.

1) α-List Decoding of Interleaved RS Codes in Sect. IV: We recall that the IRS codes in Sect. IV can also be viewed as RS codes with evaluation points in a subfield. Such codes have appeared in several previous works on array codes; in particular, in [20], Guruswami and Xing presented a list decoding algorithm for them. This algorithm can be easily modified for the problem of α-list decoding IRS codes.

Theorem VI.2 ([20])**.** *Let $F = \mathbb{F}_q, E = \mathbb{F}_{q^l}$ and let $\mathcal{C}$ be the code $RS_E(n, k, \Omega)$, where $\Omega = F$. For every $R = \frac{k}{n} \in (0, 1)$, and $\epsilon, \gamma > 0$, there exists a sufficiently large positive integer l such that the code can be list decoded from a fraction of $1 - R - \epsilon$ of errors in $|\mathcal{C}|^\gamma$ time, outputting a list of size at most $|\mathcal{C}|^\gamma$.*

This result can be modified for the α-list decoding problem, where as before $\alpha = m/l$. This is simply because in (9), we have $g(C) \in \text{IRS}_F(n, kl/m, m, \Omega)$, and we can view it as an RS code with evaluation points in a subfield.

This concludes the description, justifying the optimality claim for α-list decoding of the codes considered here.

2) α-List Decoding of FRS Codes: It is also possible to show that there exists a family of FRS codes of growing length n and sub-packetization l that can be list-decoded from an $1 - R/\alpha$ fraction of errors by downloading an α proportion of the codeword. To justify this claim, we again need to construct $n + 1$ functions $f_i : F^l \to F^{\alpha_i l}, i = 1, 2, \ldots, n$ and $g : F^{(\sum_{i=1}^n \alpha_i)l} \to F^{nl}$ that satisfy (14). It turns out that the projection functions suffice, and we take $f_1 = f_2 = \cdots = f_n = f$, where f is defined in (11). Downloading an α proportion from each of the codeword coordinates, we obtain the code $\mathcal{C}^\alpha$ defined in (13) whose rate is R/α. When the code length n and sub-packetization l of the FRS code become large enough, we can use the list decoding algorithm introduced in [15] to decode $\mathcal{C}^\alpha$ up to a fraction arbitrarily close to $1 - R/\alpha$ of errors.

Thus we conclude that

$$\rho_\alpha(R) = 1 - R/\alpha,$$

and FRS codes achieve the α-list decoding capacity.

Remark VI.3. *The code $\mathcal{C}^\alpha$ differs from an FRS code in the sense that the evaluation points in two consecutive coordinates are not consecutive powers of the primitive element. However, the list decoding algorithm introduced in [15] only requires that within each codeword coordinate, the evaluation points are consecutive powers of the primitive element. The code $\mathcal{C}^\alpha$ satisfies this constraint, so it is possible to rely on this algorithm in our arguments.*

Note that when the code length n and the sub-packetization l of FRS codes become large enough, they achieve the α-list decoding capacity uniformly for all values of α.

B. Minimum Storage Regenerating Codes With Optimal α-Decoding Radius

In this section we give an explicit construction of MDS codes with optimal bandwidth for repairing single erasure and optimal α-decoding radius simultaneously. The construction is a simple extension of the MSR code construction in [5].

We first recall the repair bandwidth and the cut-set bound. Given an (n, k, l) MDS array code $\mathcal{C}$ over a finite field F, a failed node C_i and a set of $d \geq k$ helper nodes $\{C_j, j \in \mathcal{R}\}$, define $N(\mathcal{C}, i, \mathcal{R})$ as the smallest number of symbols of F one needs to download in order to recover the failed node C_i from the helper nodes $\{C_j, j \in \mathcal{R}\}$. The repair bandwidth of the code is defined as follows.

Definition VI.4 (Repair bandwidth)**.** *Let $\mathcal{C}$ be an (n, k, l) MDS array code over a finite field F. Let $d \geq k$ be the number of helper nodes. The* d-repair bandwidth *of the code $\mathcal{C}$ is given by*

$$\beta(d) := \max_{i\in[n], |\mathcal{R}|=d, i\notin\mathcal{R}} N(\mathcal{C}, i, \mathcal{R}). \tag{15}$$

According to the *cut-set bound* derived in [2],

$$N(\mathcal{C}, i, \mathcal{R}) \geq \frac{dl}{d - k + 1}$$

for all $\mathcal{R} \subseteq ([n] \setminus \{i\})$ with cardinality d. If the d-repair bandwidth meets the cut-set bound with equality, i.e.,

$$\beta(d) = \frac{dl}{d - k + 1},$$

we say that the code $\mathcal{C}$ has the *d-optimal repair property*, and $\mathcal{C}$ is referred to as MSR code in the literature.

Let $\alpha = m/s < 1$, where m and s are positive integers. In this section we present an $(n, k, l = s(d-k+1)^n)$ MDS array code $\mathcal{C}$ over a finite field F with d-optimal repair property and optimal α-decoding radius simultaneously, where the field size $|F| \geq s(d - k + 1)n$. We write a codeword of $\mathcal{C}$ as $(C_1, C_2, \ldots, C_n)$ and write each coordinate as $C_i = (c_{i,j,\underline{a}} : j \in [s], \underline{a} \in \{0, 1, \ldots, d - k\}^n)$, i.e., the coordinates of C_i is indexed by a scalar $j \in [s]$ and a vector $\underline{a} = (a_1, a_2, \ldots, a_n) \in \{0, 1, \ldots, d - k\}^n$, so each C_i indeed has $l = s(d - k + 1)^n$ coordinates. Let $\{\lambda_{i,j,t} : i \in [n], j \in [s], t \in \{0, 1, \ldots, d-k\}\}$ be $s(d-k+1)n$ distinct elements of F. The code $\mathcal{C}$ is defined by the following set of parity check equations:

$$\sum_{i=1}^{n}\sum_{j=1}^{s} \lambda_{i,j,a_i}^t c_{i,j,\underline{a}} = 0, \tag{16}$$
$$t = 0, 1, \ldots, (n-k)s - 1, \quad \underline{a} \in \{0, 1, \ldots, d - k\}^n.$$

We can see that for each fixed $\underline{a} \in \{0,1,\ldots,d-k\}^n$, the vector $(c_{i,j,\underline{a}} : i \in [n], j \in [s])$ forms a Generalized Reed-Solomon (GRS) code with length sn and dimension sk, so $\mathcal{C}$ is indeed an $(n, k, l = s(d-k+1)^n)$ MDS array code.

Proposition VI.5. *The code $\mathcal{C}$ has optimal α-decoding radius.*

Proof. From each C_i we download $f(C_i) := (c_{i,j,\underline{a}} : j \in [m], \underline{a} \in \{0,1,\ldots,d-k\}^n) \in F^{m(d-k+1)^n}$, which contains a $m/s = \alpha$ proportion of coordinates in C_i. Since $(c_{i,j,\underline{a}} : i \in [n], j \in [s])$ forms an (sn, sk) MDS code for every $\underline{a} \in \{0,1,\ldots,d-k\}^n$, we can calculate $(c_{i,j,\underline{a}} : i \in [n], j \in [s])$ from $\{f(C_i) : i \in \mathcal{I}\}$ for every $\underline{a} \in \{0,1,\ldots,d-k\}^n$ and every subset $\mathcal{I} \subseteq [n]$ with cardinality $|\mathcal{I}| \geq sk/m = k/\alpha$. In other words, we can recover the original codeword $(C_1, C_2, \ldots, C_n)$ from $\{f(C_i) : i \in \mathcal{I}\}$ from every subset $\mathcal{I} \subseteq [n]$ with cardinality $|\mathcal{I}| \geq k/\alpha$. We thus conclude that we can do fractional decoding up to $\lfloor (n - k/\alpha)/2 \rfloor$ errors. □

Proposition VI.6. *The code $\mathcal{C}$ has the d-optimal repair property.*

Proof. Without loss of generality suppose that we want to repair C_1. For $u \in \{0,1,\ldots,d-k\}$, we write $\underline{a}(1,u) := (u, a_2, a_3, \ldots, a_n)$, namely we replace a_1 with u in vector $\underline{a}$ to obtain $\underline{a}(1,u)$. Replacing $\underline{a}$ with $\underline{a}(1,u)$ in (16), we obtain that for every $u \in \{0,1,\ldots,d-k\}$,

$$\sum_{j=1}^{s} \lambda_{1,j,u}^t c_{1,j,\underline{a}(1,u)} + \sum_{i=2}^{n}\sum_{j=1}^{s} \lambda_{i,j,a_i}^t c_{i,j,\underline{a}(1,u)} = 0,$$
$$t = 0,1,\ldots,(n-k)s-1, \quad \underline{a} \in \{0,1,\ldots,d-k\}^n.$$

Summing these equations over $u \in \{0,1,\ldots,d-k\}$, we have

$$\sum_{u=0}^{d-k}\sum_{j=1}^{s} \lambda_{1,j,u}^t c_{1,j,\underline{a}(1,u)} + \sum_{i=2}^{n}\sum_{j=1}^{s} \lambda_{i,j,a_i}^t \Big(\sum_{u=0}^{d-k} c_{i,j,\underline{a}(1,u)}\Big) = 0,$$
$$t = 0,1,\ldots,(n-k)s-1, \quad \underline{a} \in \{0,1,\ldots,d-k\}^n.$$

Since all the λ's in the equation above are distinct, we conclude that for every fixed $\underline{a} \in \{0,1,\ldots,d-k\}^n$, the vector

$$\Big(\{c_{1,j,\underline{a}(1,u)} : u \in \{0,1,\ldots,d-k\}, j \in [s]\}, \Big\{\sum_{u=0}^{d-k} c_{i,j,\underline{a}(1,u)} : i \in \{2,3,\ldots,n\}, j \in [s]\Big\}\Big) \quad (17)$$

forms a GRS code with length $s(d-k+1) + s(n-1) = s(d-k+n)$ and dimension $s(d-k+n) - s(n-k) = sd$. As an immediate consequence, we can calculate the vector in (17) from

$$\Big\{\sum_{u=0}^{d-k} c_{i,j,\underline{a}(1,u)} : i \in \mathcal{R}, j \in [s]\Big\}$$

for any subset $\mathcal{R} \subseteq [n]$ with cardinality $|\mathcal{R}| = d$. Therefore we can download the following $\frac{dl}{d-k+1}$ symbols in F

$$\Big\{\sum_{u=0}^{d-k} c_{i,j,\underline{a}(1,u)} : i \in \mathcal{R}, j \in [s], \underline{a} \in \{0,1,\ldots,d-k\}^n, a_1 = 0\Big\}$$

from the d helper nodes $\{C_i : i \in \mathcal{R}\}$, and we will be able to calculate

$$\Big\{c_{1,j,\underline{a}(1,u)} : u \in \{0,1,\ldots,d-k\}, j \in [s], \underline{a} \in \{0,1,\ldots,d-k\}^n, a_1 = 0\Big\} = \{c_{1,j,\underline{a}}, j \in [s], \underline{a} \in \{0,1,\ldots,d-k\}^n\},$$

which is the set of all the coordinates of C_1. This completes the proof of the d-optimal repair property. □

REFERENCES

[1] I. Tamo, M. Ye, and A. Barg, "Fractional decoding: Error correction from partial information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 998–1002.

[2] 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.

[3] 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.

[4] 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.

[5] M. Ye and A. Barg, "Explicit constructions of high-rate MDS array codes with optimal repair bandwidth," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2001–2014, Apr. 2017.

[6] M. Ye and A. Barg, "Explicit constructions of optimal-access MDS codes with nearly optimal sub-packetization," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6307–6317, Oct. 2017.

[7] M. Ye and A. Barg, "Cooperative repair: Constructions of optimal mds codes for all admissible parameters," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1639–1656, Mar. 2018.

[8] V. Guruswami and M. Wootters, "Repairing Reed–Solomon codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5684–5698, Sep. 2017.

[9] I. Tamo, M. Ye, and A. Barg, "Optimal repair of Reed–Solomon codes: Achieving the cut-set bound," in *Proc. IEEE 58th Annu. Symp. Found. Comput. Sci. (FOCS)*, Berkeley, CA, USA, Oct. 2017, pp. 216–227.

[10] M. Ye and A. Barg, "Explicit constructions of MDS array codes and RS codes with optimal repair bandwidth," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1202–1206.

[11] I. Tamo, M. Ye, and A. Barg, "The repair problem for Reed–Solomon codes: Optimal repair of single and multiple erasures with almost optimal node size," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2673–2695, May 2019.

[12] M. Blaum, P. G. Farell, and H. van Tilborg, "Array codes," in *Handbook Coding Theory*, vol. 2, V. Pless and W. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier, 1998, ch. 22, pp. 1855–1909.

[13] D. Bleichenbacher, A. Kiayias, and M. Yung, "Decoding of interleaved Reed Solomon codes over noisy data," in *Automata, Languages and Programming* (Lecture Notes in Computer Science), vol. 2719. Berlin, Germany: Springer, 2003, pp. 97–108.

[14] G. Schmidt, V. R. Sidorenko, and M. Bossert, "Collaborative decoding of interleaved Reed–Solomon codes and concatenated code designs," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2991–3012, Jul. 2009.

[15] V. Guruswami and A. Rudra, "Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 135–150, Jan. 2008.

[16] C. Thommesen, "The existence of binary linear concatenated codes with Reed–Solomon outer codes which asymptotically meet the Gilbert-Varshamov bound," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 850–853, Nov. 1983.

[17] C. Devet, I. Goldberg, and N. Heninger, "Optimally robust private information retrieval," in *Proc. 21st USENIX Security Symp.*, 2012, pp. 269–283.

[18] H. Cohn and N. Heninger, "Approximate common divisors via lattices," *Open Book Series*, vol. 1, pp. 271–293, Nov. 2013.

[19] W. Santos, "On fractional decoding of Reed–Solomon codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1552–1556.

[20] V. Guruswami and C. Xing, "List decoding Reed–Solomon, algebraic-geometric, and Gabidulin subcodes up to the Singleton bound," in *Proc. 45th Annu. ACM Symp. Theory Comput. (STOC)*. New York, NY, USA, Jun. 2013, pp. 843–852.

Itzhak Tamo (M'10) was born in Israel in 1981. He received the B.A. degree in mathematics, the B.Sc. degree in electrical engineering, and the Ph.D. degree in electrical engineering from Ben-Gurion University, Israel, in 2008 and 2012, respectively. From 2012 to 2014, he was a Post-Doctoral Researcher with the Institute for Systems Research, University of Maryland, College Park. Since 2015, he has been a Senior Lecturer with the Electrical Engineering Department, Tel Aviv University, Israel. His research interests include storage systems and devices, coding, information theory, and combinatorics. He was a co-recipient (with Z. Wang and J. Bruck) of the IEEE Communication Society Data Storage Technical Committee 2013 Best Paper Award. Together with A. Barg, he received the 2015 IEEE Information Theory Society Paper Award.

Min Ye received the B.S. degree in electrical engineering from Peking University, Beijing, China, in 2012, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, in 2017. He then spent two years as a Post-Doctoral Researcher at Princeton University. Since 2019, he has been an Assistant Professor with the Data Science and Information Technology Research Center, Tsinghua-Berkeley Shenzhen Institute, Shenzhen, China. His research interests include coding theory, information theory, differential privacy, and machine learning. He received the 2017 IEEE Data Storage Best Paper Award.

Alexander Barg (M'00–SM'01–F'08) is currently a Professor with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD, USA. He is broadly interested in information and coding theory, applied probability, and algebraic combinatorics. He has published about 100 research articles. He received the 2015 Information Theory Society Paper Award (joint with I. Tamo) and the 2017 IEEE Data Storage Award (joint with M. Ye). He was a plenary Speaker at the 2016 IEEE International Symposium on Information Theory. He currently serves as the Editor-in-Chief of TRANSACTIONS on Information Theory.