
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Grezet, Matthias; Freij-Hollanti, Ragnar; Westerbäck, Thomas; Hollanti, Camilla
Alphabet-Dependent Bounds for Linear Locally Repairable Codes Based on Residual Codes

Published in:
IEEE Transactions on Information Theory

DOI:
[10.1109/TIT.2019.2911595](https://doi.org/10.1109/TIT.2019.2911595)

Published: 01/01/2019

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Grezet, M., Freij-Hollanti, R., Westerbäck, T., & Hollanti, C. (2019). Alphabet-Dependent Bounds for Linear Locally Repairable Codes Based on Residual Codes. *IEEE Transactions on Information Theory*, 65(10), 6089-6100. Article 8700214. <https://doi.org/10.1109/TIT.2019.2911595>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Alphabet-Dependent Bounds for Linear Locally Repairable Codes Based on Residual Codes

Matthias Grezet, Ragnar Freij-Hollanti, Thomas Westerbäck, and Camilla Hollanti

Abstract—Locally repairable codes (LRCs) have gained significant interest for the design of large distributed storage systems as they allow a small number of erased nodes to be recovered by accessing only a few others. Several works have thus been carried out to understand the optimal rate–distance tradeoff, but only recently the size of the alphabet has been taken into account. In this paper, a novel definition of locality is proposed to keep track of the precise number of nodes required for a local repair when the repair sets do not yield MDS codes. Then, a new alphabet-dependent bound is derived, which applies both to the new definition and the initial definition of locality. The new bound is based on consecutive residual codes and intrinsically uses the Griesmer bound. A special case of the bound yields both the extension of the Cadambe-Mazumdar bound and the Singleton-type bound for codes with locality (r, δ) , implying that the new bound is at least as good as these bounds. Furthermore, an upper bound on the asymptotic rate–distance tradeoff of LRCs is derived, and yields the tightest known upper bound for large relative minimum distances. Achievability results are also provided by deriving the locality of the family of Simplex codes together with a few examples of optimal codes.

I. INTRODUCTION

In recent times, many service providers allow users to access and store data remotely to avoid overwhelming the limited storage capacity of single users. This leads naturally to the design of large distributed storage systems that reliably store data while minimizing the redundancy necessary to deal with server failures.

The use of erasure-correcting codes together with network coding techniques for distributed storage systems, initiated in [1], has become popular since these so-called regenerating codes achieve the optimal tradeoff between the required repair bandwidth and storage overhead. For a standard erasure code of length n , dimension k and minimum distance d , any $d - 1$ failures can be repaired by contacting at most k other nodes.

The work of M. Grezet and C. Hollanti was supported in part by the Academy of Finland, under grants #276031, #282938, and #303819, and by the Technical University of Munich – Institute for Advanced Study, funded by the German Excellence Initiative and the EU 7th Framework Programme under grant agreement #291763, via a Hans Fischer Fellowship.

R. Freij-Hollanti was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under grant WA3907/1-1.

M. Grezet, R. Freij-Hollanti, and C. Hollanti are with the Department of Mathematics and Systems Analysis, Aalto University, FI-00076 Aalto, Finland (e-mail: matthias.grezet@aalto.fi; ragnar.freij@aalto.fi; camilla.hollanti@aalto.fi).

T. Westerbäck is with the Division of Applied Mathematics, UKK, Mälardalen University, Högscoleplan 1, Box 883, 721 23 Västerås, Sweden (e-mail: thomas.westerback@mdh.se).

Digital Object Identifier 10.1109/TIT.2019.2911595

Copyright © 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

In addition to this property, and at the cost of the failure tolerance, regenerating codes also enable efficient repair of failed nodes. This was long thought to be in contrast to the traditional maximum distance separable (MDS) codes that have to reconstruct the whole file in order to repair a single node. However, [2], [3] showed that this claim is not true in general, namely some MDS codes can also be efficiently repaired. Nevertheless, the number of nodes contacted for repair can be a bottleneck for the system efficiency. To reduce the repair network traffic, [4] and later [5] introduced the notion of locality r allowing the repair of a single failure to be done by contacting only r nodes with $r \ll k$. Erasure codes satisfying this requirement are called locally repairable (or recoverable) codes (LRCs).

A natural extension was presented in [6], [7] where the authors defined the locality $(r, \delta)_i$ for the information symbols to allow $\delta - 1$ failures to be still corrected locally. This requirement was extended in [6] to all symbols without differentiating between the information symbols and the parity symbols. In this paper, we focus only on all-symbol locality and therefore drop the specification. A *repair set* R of an LRC is a set of coordinates such that any $\delta - 1$ code symbols c_i with $i \in R$ can be obtained from the remaining code symbols with indices in R . Other extensions of the locality property include codes with availability [8], sequential repair of several erasures [9], cooperative repair [10], local repair on graphs [11] and many others.

Abundant literature has been devoted to understanding the best possible parameters of LRCs and providing optimal constructions. The authors of [4] gave the first tradeoff between the parameters n, k, d , and r by showing that the minimum distance d of an (n, k, d, r) -LRC with locality r is bounded as follows:

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \quad (1)$$

This bound was extended in [6] for any (n, k, d, r, δ) -LRC with locality (r, δ) :

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \quad (2)$$

The two bounds have been proven to be tight for large alphabet size with constructions provided in [5]–[7], [12]–[20]. Bounds for codes with availability were established in [8], [12], [21], [22]. For a summary on various bounds for LRCs, see [23].

The pioneering work done in [24] improves on the bound (1) by including a dependence on the alphabet size in the

bound, that is, for any (n, k, d, r) -LRC over the alphabet Q with $|Q| = q$, we have

$$k \leq \min_{1 \leq t \leq n/(r+1)} \left\{ tr + k_{\text{opt}}^{(q)}(n - t(r+1), d) \right\}, \quad (3)$$

where $k_{\text{opt}}^{(q)}(n, d)$ is the maximal dimension of a code over Q of length n and minimum distance d . This has led to further construction of optimal LRCs over small alphabets, for example in [25], [26].

Recently, the authors of [27] proposed the first alphabet-dependent bounds on (n, k, d, r, δ) -LRCs over Q using an upper bound $B_{l-c}(r + \delta - 1, \delta)$ on the cardinality of a code given its length $r + \delta - 1$ and minimum distance δ , with the extra requirement that the upper bound is a log-convex function on the length. The global bound is as follows:

$$k \leq \left(\left\lceil \frac{n-d+1}{r+\delta-1} \right\rceil + 1 \right) \log_q B_{l-c}(r + \delta - 1, \delta). \quad (4)$$

A linear-programming bound for LRCs with locality (r, δ) was also derived in [27] under the extra assumption that the repair sets are disjoint.

Finally, in [28], the authors presented a Singleton-type bound for binary linear LRCs. This bound uses the local dimension of a repair set instead of the parameter r and a more precise understanding of the intersection between two repair sets. As such, the work in this paper generalizes these two ideas.

While so far no distinction was made between non-linear and linear codes, the next results in this paragraph are only true for linear codes. In [29], Griesmer proved the existence of a residual code for any binary linear code (over \mathbb{F}_2), *i.e.*, a code obtained by a restriction with certain specific parameters. Griesmer then derived a lower bound on the length of the code given its dimension and minimum distance. The two results were later extended to an arbitrary field \mathbb{F}_q in [30]. We present here the most general form. For any $[n, k, d]$ linear code \mathcal{C} over \mathbb{F}_q , there exists \mathcal{C}' , a restriction of \mathcal{C} called the *residual code of \mathcal{C}* , such that \mathcal{C}' has parameters $[n-d, k-1, d' \geq \lceil d/q \rceil]$. By recursively taking residual codes, the authors of [30] obtained the following bound on the length n of a linear code, known as the *Griesmer bound*, and denoted here by $\mathcal{G}(k, d)$:

$$n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil =: \mathcal{G}(k, d). \quad (5)$$

A. Our contributions

In this paper, we first highlight the differences between the initial motivation for introducing the notion of locality in [4], [5] and the definition of locality given in [6], where the authors constrained the size of the repair sets. We show, through some examples, how the definition in [6] yields only a loose upper bound on the number of nodes contacted during the repair process when δ is larger than the alphabet size of the code. To remedy this, we introduce a new definition for locality called dimension-locality and compare it to the definition of locality in [6].

Then, we focus on linear LRCs and derive a new alphabet-dependent bound of the type of the bound (3) for linear codes

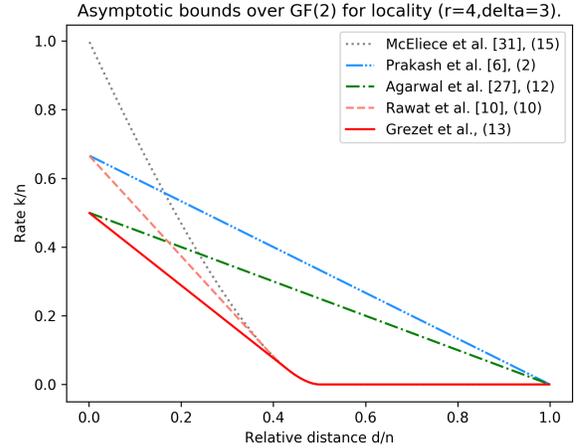


Fig. 1. Comparison of the upper bounds on the tradeoff between the rate k/n and the relative minimum distance d/n from [6], [27], [31], and the bounds (10) and (13) over the binary field for large values of n and fixed locality $(r = 4, \delta = 3)$.

with dimension-locality using the repair sets and chains of consecutive residual codes. Given the definition of dimension-locality, this bound also applies to linear LRCs with locality (r, δ) by using a weaker bound on the dimension of the local codes. As a corollary of our results, we also derive a new Singleton-type bound that reflects better the actual dimension of the local codes. Furthermore, the new bound can be used to obtain a straightforward extension of the bound (3) for locality (r, δ) and the bound (2), which shows that our bound is always at least as good as these bounds.

Next, we derive the asymptotic formulas of the new bound and the new Singleton-type bound when $n \rightarrow \infty$ to obtain the bounds on the tradeoff between the rate and the relative minimum distance. We also use these formulas to compare our bounds to the bounds in [27] (Equation (4) and (12) here). We show that there are cases where the new asymptotic Singleton-type bound is either better, equal, or worse than the asymptotic version of the bound (2). The comparison with our main bound (8) is more direct as we can prove that there is always an interval in the relative minimum distance where the new bound is strictly better than the bound (2). Moreover, the improvement is quite significant since our bound benefits from the locality-unaware bounds on the rate–distance tradeoff. As an example, Figure 1 displays the comparison between the known bounds and the new bound (13) for linear LRCs with locality $(4, 3)$ over the binary field, where we use the McEliece-Rodemich-Rumsey-Welch (MRRW) bound in [31] as the intrinsic bound on the rate. Finally, we prove the achievability of the new bounds by studying the locality of the Simplex codes and providing a few optimal examples.

The rest of the paper is organized as follows. In Section II, we discuss the relation between the initial motivation that led to the introduction of locality and the definition given in [6]. Then, we define the notion of dimension-locality and compare it to the original definition of locality in [6]. In Section III, we derive a new bound for linear LRCs with dimension-locality

and extend it to linear LRCs with locality (r, δ) . We derive also a new Singleton-type bound for these codes. In Section IV, we prove first that our bound is always as good as the straightforward extension of the bound (3) for locality (r, δ) and the bound (2). Then, we derive the asymptotic formulas for the new bounds and use them for the comparison to the bound (4). While the comparison with the new Singleton-type bound depends on the parameters of the codes, we prove that our bound always improves on the bound (4) for large relative minimum distances. Finally, in Section V, we provide a family of codes that achieves our bounds by studying the locality of the Simplex codes.

II. MATHEMATICAL PRELIMINARIES AND LOCALITY REVISITED

We denote the set $\{1, 2, \dots, n\}$ by $[n]$ and the set of all subsets of $[n]$ by $2^{[n]}$. The set of all positive integers including 0 is denoted by \mathbb{Z}_+ . For a length- n vector \mathbf{v} and a set $I \subseteq [n]$, the vector \mathbf{v}_I denotes the restriction of the vector \mathbf{v} to the coordinates in the set I . A linear code \mathcal{C} of length n , dimension k , and minimum distance d is denoted by $[n, k, d]$ and a generator matrix for \mathcal{C} is $G_{\mathcal{C}} = (\mathbf{g}_1, \dots, \mathbf{g}_n)$ where $\mathbf{g}_i \in \mathbb{F}_q^k$ is a column vector for $i \in [n]$. The number of codewords in \mathcal{C} is the cardinality of \mathcal{C} , $|\mathcal{C}| = q^k$. The *shortening* of a code \mathcal{C} to the set of coordinates $I \subseteq [n]$ is defined by

$$\mathcal{C}/I = \{\mathbf{c}_{[n] \setminus I} : \mathbf{c} \in \mathcal{C} \text{ such that } c_i = 0 \text{ for all } i \in I\},$$

and the *restriction* of code \mathcal{C} to I is defined by

$$\mathcal{C}|_I = \{\mathbf{c}_I : \mathbf{c} \in \mathcal{C}\}.$$

For convenience, we shortly call the codes obtained by a restriction *restricted codes*. Equivalently, the *puncturing* of \mathcal{C} on the set I is defined as the restriction of \mathcal{C} to the set $[n] - I$. For an $[n, k, d]$ linear code \mathcal{C} , if \mathcal{C} meets the Singleton bound, i.e., if $d = n - k + 1$, then \mathcal{C} is called a maximum distance separable (MDS) code.

To measure the dimension of restricted linear codes, or more generally, the amount of information contained in the restriction of arbitrary codes, we use the notion of an entropy function on the subsets $I \subseteq [n]$, where n is the length of the code. We state it here for quasi-uniform codes over the alphabet Q .

Quasi-uniform codes are a general class of error-correcting codes containing all the linear codes, group codes, and almost affine codes. More precisely, let \mathcal{C} be an arbitrary subset of Q^n . We can associate to \mathcal{C} a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with joint probability distribution

$$P(\mathbf{X} = \mathbf{x}) = \begin{cases} 1/|\mathcal{C}| & \text{if } \mathbf{x} \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, \mathcal{C} is a *quasi-uniform* code if for all $A \subseteq [n]$, the restricted random vector \mathbf{X}_A is uniformly distributed over its support $\lambda(\mathbf{X}_A) = \{\mathbf{x}_A : P(\mathbf{X}_A = \mathbf{x}_A) > 0\}$. We refer to [32] for more information about quasi-uniform codes and [33] for the entropy function on these codes.

Definition 1. Let \mathcal{C} be a quasi-uniform code of length n over the alphabet Q and $I \subseteq [n]$. The entropy associated to \mathcal{C} is the function $H_{\mathcal{C}} : [n] \rightarrow \mathbb{R}$ with

$$H_{\mathcal{C}}(I) = \frac{\log(|\mathcal{C}|_I)}{\log |Q|} = \frac{\log(|\{\mathbf{c}_I : \mathbf{c} \in \mathcal{C}\}|)}{\log |Q|}.$$

For ease of notation, if the underlying code of $H_{\mathcal{C}}$ is clear, we drop the specification to \mathcal{C} . For linear codes, this function measures exactly the dimension of the restricted codes and for a subset $I \subseteq [n]$, $H_{\mathcal{C}}(I)$ is equivalent to the rank of the submatrix formed by the columns \mathbf{g}_i with $i \in I$ or, equivalently, to the rank function of I in the associated matroid of \mathcal{C} . The entropy function has the following polymatroidal properties as shown in [34].

Proposition 1. Let \mathcal{C} be a quasi-uniform code of length n over the alphabet Q and H the entropy function associated to \mathcal{C} . For $I, J \subseteq [n]$, we have

- 1) $H(I) \leq |I|$,
- 2) If $I \subseteq J$ then $H(I) \leq H(J)$,
- 3) $H(I \cup J) + H(I \cap J) \leq H(I) + H(J)$.

The entropy function also behaves nicely for restricted codes. Let $I \subseteq [n]$ and $\mathcal{C}|_I$ be the restriction of \mathcal{C} to the set I . Then for $J \subseteq I$, we have $H_{\mathcal{C}|_I}(J) = H_{\mathcal{C}}(J)$. Finally, we define a closure operation on the subsets of $[n]$ for linear codes.

Definition 2. Let \mathcal{C} be an $[n, k, d]$ linear code and $I \subseteq [n]$. The closure operator $\text{cl} : 2^{[n]} \rightarrow 2^{[n]}$ is

$$\text{cl}(I) = \{e \in [n] : H(I \cup e) = H(I)\}.$$

One can think of the closure operator via the generator matrix $G_{\mathcal{C}}$ of \mathcal{C} , where $\text{cl}(I)$ is the set of all columns in $G_{\mathcal{C}}$ contained in the linear span of the columns indexed by I .

The following table summarizes the notation used throughout the paper. The formal definitions for some of them will only appear later in the paper.

$[n, k, d]$	Linear code of length n , dimension k and minimum distance d
(n, k, d, r, δ)	LRC with locality (r, δ) and δ the local minimum distance
$(n, k, d)(\kappa, \delta)$	LRC with dimension-locality (κ, δ)
$\mathcal{G}(k, d)$	Griesmer bound on the length n of a linear code
κ_B	Bound on the dimension of a code
κ_A	Log-convex bound on the dimension of a code
$B_{l-c}(n, d)$	Log-convex bound on the cardinality of a code
\mathcal{R}	Rate k/n
δ_n	Relative minimum distance d/n
<i>res-chain</i>	Chain of consecutive residual codes

A. Definition of locality and relation with the number of nodes contacted for repairing

In this part, we explain how the definition of locality (r, δ) in [6] diverges from the initial motivation of introducing the

notion of locality, and state a new definition of locality called dimension-locality. As mentioned in the introduction, [4] and then [5] introduced the notion of locality r to reduce the repair traffic by designing storage codes such that one failure can be repaired by contacting only a small number of nodes $r \ll k$ in the storage system. A natural extension of the above definition is to allow multiple erasures to be corrected locally by still accessing fewer nodes than k . For this, we need the local restrictions to have a minimum distance of δ so that up to $\delta - 1$ erasures can be repaired locally. The definition presented in [6] is the following.

Definition 3. An $[n, k, d]$ linear code \mathcal{C} has all-symbol locality (r, δ) if for all code symbol indices $i \in [n]$ there exists a set $R \subseteq [n]$, called a repair set, such that

- 1) $i \in R$,
- 2) $|R| \leq r + \delta - 1$,
- 3) The minimum distance of the restriction of \mathcal{C} to the set R is at least δ .

We refer to \mathcal{C} as an (n, k, d, r, δ) -LRC.

With a slight abuse of language, we say that a repair set R has dimension κ and minimum distance δ if the restricted code $\mathcal{C}|_R$ has dimension κ and minimum distance δ . We also say that a repair set R is MDS if $\mathcal{C}|_R$ is an MDS code.

In Definition 3, any $\delta - 1$ coordinates of R are determined by the values of the remaining $|R| - \delta + 1 \leq r$ coordinates, thus enabling local repair by contacting at most r other nodes. The problem with Definition 3 is that it implicitly requires the repair sets to be MDS in order for r to be the dimension of the local codes. In other words, if a repair set R is not MDS, then the number of nodes needed to repair any $\delta - 1$ failures in R is strictly less than r . Thus, Definition 3 diverges from the initial meaning of controlling precisely the number of nodes contacted during the repair process when considering non-MDS repair sets.

This observation is particularly relevant when the field size is fixed and δ is too large for MDS codes to exist. For example, if we consider binary codes and require that $\delta > 2$ in order to correct more than one erasure locally, then none of the repair sets can be MDS and r is no longer the local dimension. We illustrate this phenomenon by a concrete example.

Example 1. Let \mathcal{C} be the binary linear $[10, 4, 4]$ -code given by the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

We define the three repair sets by their corresponding columns in G : $R_1 = \{1, 2, 3, 5, 6, 8\}$, $R_2 = \{2, 3, 6, 7, 9, 10\}$, and $R_3 = \{1, 4, 6, 7, 8, 10\}$. Every repair set has size 6, minimum distance $d_{\mathcal{C}|_{R_i}} = 3$, and entropy $H(R_i) = 3$. Thus, according to Definition 3, we get $r = |R_i| - \delta + 1 = 4$ and hence \mathcal{C} is a binary linear $(10, 4, 4, 4, 3)$ -LRC. However, even though $r = k = 4$, we can repair up to two failures by contacting at most 3 nodes, as shown by the entropy. While this is not really in contrast with the above definition, it only

gives a loose bound $r = 4$ for the number of nodes we need to contact.

To be able to precisely keep track of the number of nodes contacted during the repair process, we propose a slightly different definition for locally repairable codes tolerating multiple local erasures. We do this by replacing the condition on the size of a repair set by a condition on its dimension.

Definition 4. An $[n, k, d]$ linear code \mathcal{C} has all-symbol dimension-locality (κ, δ) if for all code symbol indices $i \in [n]$, there exists a set $R \subseteq [n]$ such that

- 1) $i \in R$,
- 2) $H(R) = \log_q(|\{\mathbf{c}_R : \mathbf{c} \in \mathcal{C}\}|) \leq \kappa$,
- 3) The minimum distance of the restriction of \mathcal{C} to the set R is at least δ .

We refer to \mathcal{C} as an $(n, k, d)(\kappa, \delta)$ -LRC.

With this definition, we regain the fact that every $\delta - 1$ coordinates can be recovered by contacting at most κ other coordinates and κ can be made tight. When we do not restrict the field size, optimal repair sets are MDS with size equal to $\kappa + \delta - 1$. Thus, both definitions coincide for large alphabets. When the field size is fixed and we have non-MDS repair sets, the parameter κ in Definition 4 still measures the local dimension. On the contrary, the parameter r in Definition 3 measures the local dimension with the addition of a non-zero part of the local size that varies depending on the field size and the local minimum distance. Moreover, the new definition allows more flexibility on the size of the repair sets. Specifically, the size of a repair set can be smaller or bigger than $\kappa + \delta - 1$, since κ and δ are only an upper bound on the dimension of a code and a lower bound on its minimum distance, respectively.

Obviously, every code with locality (r, δ) is a code with dimension-locality $(\kappa = r, \delta)$. To obtain a closer relation between the two notions of locality, we can replace r by the value of an upper bound on the dimension of a code given its length $r + \delta - 1$ and minimum distance δ . Let κ_B be the maximal dimension obtained by such a bound. Then, every code with locality (r, δ) is a code with dimension-locality $(\kappa = \kappa_B, \delta)$. The problem is that κ_B might not be tight, *i.e.*, there is no repair set R such that $H(R) = \kappa_B$, which goes against the purpose of the new definition. This is illustrated in the following example.

Example 2. Let \mathcal{C}_1 be the $[7, 3, 4]$ binary linear code obtained by taking the dual of the $[7, 4, 3]$ Hamming code, and let \mathcal{C}_2 be the $[6, 3, 3]$ code obtained by puncturing \mathcal{C}_1 on the last coordinate. Consider the $[13, 6, 3]$ binary linear code $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ with generator matrix

$$G = \begin{pmatrix} G_1 & \mathbf{0} \\ \mathbf{0} & G_2 \end{pmatrix},$$

where G_1 and G_2 are the generator matrices of \mathcal{C}_1 and \mathcal{C}_2 , respectively.

By defining the repair sets to be $R_1 = \{1, \dots, 7\}$ and $R_2 = \{8, \dots, 13\}$, the code \mathcal{C} is a $(13, 6, 3, 5, 3)$ -LRC since $\delta = 3$ and $|R_1| = 7$. The maximal entropy of a repair set is 3, so \mathcal{C} is a $(13, 6, 3)(3, 3)$ -LRC. To get an upper bound on the

local dimension from the parameters (r, δ) , we can use an upper bound on the dimension of a binary code of length $r + \delta - 1 = 7$ and minimum distance $\delta = 3$. Since the Hamming code is a $[7, 4, 3]$ binary code, the best upper bound gives a local dimension of 4. However, this is strictly greater than the maximal dimension of a local code in this example, which is 3. Thus, it is impossible to obtain the exact maximal dimension of a local code from the parameters (r, δ) in Definition 3.

The previous example also illustrates how the flexibility of the sizes of the repair sets, obtained from Definition 4, can be used to achieve a code of length n that is not divisible by any of the sizes of the repair sets, while keeping the dimension-locality parameters (κ, δ) .

III. BOUNDS FOR DIMENSION-LOCALITY (κ, δ) AND LOCALITY (r, δ)

In this section, we study the structure of linear codes with dimension-locality (κ, δ) and derive a bound on their parameters. Following the general framework of [24], we construct a set with a large size and a small entropy. This is done by using a detailed analysis of the repair sets based on the work done in [29] and [30]. It yields a bound of the form of the bound (3) handling both MDS and non-MDS repair sets. Then, we extend our bound to linear codes with locality (r, δ) . Finally, using a weaker estimation of our results, we derive a new Singleton-type bound for (n, k, d, r, δ) -LRCs.

We start by presenting the new bound for linear codes with dimension-locality. Here, $k_{\text{opt}}^{(q)}$ is an upper bound on the dimension of a code and $\mathcal{G}(\kappa, \delta)$ is the Griesmer bound taken over \mathbb{F}_q .

Theorem 1. *Let \mathcal{C} be a linear $(n, k, d)(\kappa, \delta)$ -LRC over \mathbb{F}_q . Then we have*

$$k \leq \min_{\lambda \in \mathbb{Z}_+} \left\{ \lambda + k_{\text{opt}}^{(q)}(n - (a + 1) \cdot \mathcal{G}(\kappa, \delta) + \mathcal{G}(\kappa - b, \delta), d) \right\} \quad (6)$$

where $a, b \in \mathbb{Z}$ are such that $\lambda = a\kappa + b, 0 \leq b < \kappa$.

Proof. The proof is given in the appendix. \square

In order to prove this bound, we need a better understanding of the bound (3) and the implications of having a non-MDS repair set. The bound (3) relies mainly on two results. The first result is a construction of a set with an upper bound on its entropy and a lower bound on its size. The second result is a shortening argument that governs the part inside $k_{\text{opt}}^{(q)}$ in the bound (3). This is reproduced here with a slight rephrasing.

Lemma 1 ([24], Lemma 2). *Let \mathcal{C} be an $[n, k, d]$ linear code over \mathbb{F}_q and $I \subseteq [n]$ such that $H(I) < k$. Then the shortened code \mathcal{C}/I has parameters $[n - |I|, k - H(I), d' \geq d]$.*

Regarding the first result, the technique used to construct large sets relies on taking the union of repair sets. If two repair sets happen to intersect, which will reduce both the entropy and the size of their union, a correction is performed by adding arbitrary elements to their union. The main difficulty to extend this technique to non-MDS repair sets is to deal with the intersection of the repair sets and find the appropriate

correction. Specifically, the intersection of two repair sets can now have a size strictly larger than its entropy (take, for example, R_1 and R_2 in Example 1). Thus, it is not possible anymore to correct their union by an arbitrary set since the resultant set might exceed the upper bound on the entropy.

In order to correct the intersection, the main idea is to create a set using consecutive residual codes. As mentioned in the introduction, for any $[n, k, d]$ linear code \mathcal{C} over \mathbb{F}_q , there exists \mathcal{C}' , a restriction of \mathcal{C} called the residual code of \mathcal{C} , such that \mathcal{C}' has parameters $[n - d, k - 1, d' \geq \lceil d/q \rceil]$. We define the sequence of consecutive residual codes as a chain of subsets of $[n]$.

Definition 5. *Let \mathcal{C} be a $[n, k, d]$ linear code over \mathbb{F}_q . The res-chain of \mathcal{C} is a sequence of sets $(S_i)_{i=0}^k$ with $S_i \subseteq [n]$ constructed recursively by starting with $S_0 = [n]$ and S_i is such that $\mathcal{C}|_{S_i}$ is a residual code of $\mathcal{C}|_{S_{i-1}}$.*

This definition is well-defined since by the proof of [30, Theorem 1'], the residual code \mathcal{C}' of \mathcal{C} is constructed by restricting \mathcal{C} to a well-chosen set of coordinates. Therefore, we can interpret the recursive residual code chain as a sequence of sets in $[n]$. Furthermore, as the dimension of the residual code is one less than the dimension of the original code, the chain has length $k + 1$ and for all $0 \leq \alpha \leq k$, there is a set S in the res-chain of \mathcal{C} such that $H(S) = \alpha$. Finally, by a recursive argument, if S is a set in the res-chain of \mathcal{C} , then the minimum distance d_S of the restriction to S is bounded from below by

$$d_S \geq \left\lceil \frac{d}{q^{k-H(S)}} \right\rceil.$$

We now present two lemmas that are used to prove Theorem 1. The first lemma states how to increase the size of a set when the addition of an entire repair set will exceed the desired entropy.

Lemma 2. *Let \mathcal{C} be a linear $(n, k, d)(\kappa, \delta)$ -LRC over \mathbb{F}_q . Let $F \subseteq [n]$ be such that $\text{cl}(F) = F$ and α an integer with $1 \leq \alpha \leq \kappa$. If there exists a repair set R such that $H(R) - H(F \cap R) \geq \alpha$, then, there exists $F' \subseteq [n]$ with $\text{cl}(F') = F'$ such that*

- $H(F') \leq H(F) + \alpha$,
- $|F'| \geq |F| + \mathcal{G}\left(\alpha, \left\lceil \frac{\delta}{q^{\kappa-\alpha}} \right\rceil\right)$.

Proof. The proof is given in the appendix. \square

Using the above lemma, we can prove the following second lemma that represents the challenging part of proving the new bound.

Lemma 3. *Let \mathcal{C} be a linear $(n, k, d)(\kappa, \delta)$ -LRC over \mathbb{F}_q . Let $F \subseteq [n]$ be such that $\text{cl}(F) = F$ and $H(F) + \kappa \leq k$. Then, there exists $F_c \subseteq [n]$ with $\text{cl}(F_c) = F_c$ such that*

- $H(F_c) \leq H(F) + \kappa$,
- $|F_c| \geq |F| + \mathcal{G}(\kappa, \delta)$.

Proof. The proof is given in the appendix. \square

The intuition behind the proofs of Lemma 2 and 3 is the following. If all the repair sets are disjoint and have dimension κ , then Lemma 3 follows directly since no correction is

needed. If two repair sets intersect each other or have a dimension less than κ , both the entropy and the size of their union will be smaller than expected and a correction is required. In this case, we can use the chain of residual codes to get a set that, when added to their union, increases the entropy by exactly the amount left. The last trick is to evaluate both the size of the union and the set in the *res*-chain using the Griesmer bound. First, it is a bound on the length of a code where the minimum distance plays a more important role compared to the dimension. This fits the lower bound on the local minimum distance for codes with dimension-locality. Secondly, the Griesmer bound has the nice property that $\mathcal{G}(\kappa, \delta) = \mathcal{G}(\alpha, \delta) + \mathcal{G}\left(\kappa - \alpha, \left\lceil \frac{\delta}{q^\alpha} \right\rceil\right)$. The first term in the sum can be used to get a lower bound on the size of the repair set minus its intersection with a given set. The second term, under some conditions, gives a lower bound on the size of a particular set in the *res*-chain of a repair set. Thus, this relation is really useful when we add the extra set to correct the union of a repair set to a given set. Finally, the Griesmer bound is also consistent with our construction based on residual codes.

As a corollary of Theorem 1, we can force the parameter λ to only be a multiple of κ . This gives a bound resembling the original bound in [24].

Corollary 1. *Let \mathcal{C} be a linear $(n, k, d)(\kappa, \delta)$ -LRC over \mathbb{F}_q . Then we have*

$$k \leq \min_{t \in \mathbb{Z}_+} \left\{ t\kappa + k_{\text{opt}}^{(q)}(n - t \cdot \mathcal{G}(\kappa, \delta), d) \right\}. \quad (7)$$

Even if the wider range of the parameter λ makes the bound (6) theoretically better than the bound (7), the two bounds show similar experimental results. More precisely, we randomly generated some feasible parameters (n, k, d, r, δ) for LRCs over the binary field, which yielded results showing that the bound (7) is equal to the bound (6). One possible justification is that for two consecutive dimensions $\lambda - 1$ and λ inside the minimum in (6), the length in the second term decreases by the largest value when $\lambda = t\kappa$. Therefore, the optimal condition on the global dimension k would always happen when λ is a multiple of the local dimension κ . However, a formal proof is impossible due to the unknown intrinsic bound $k_{\text{opt}}^{(q)}$.

A. New bounds for locality (r, δ)

As already explained in Section II, to obtain a bound on LRCs with locality (r, δ) instead of dimension-locality (κ, δ) , we can estimate κ by an upper bound on the maximal dimension of a code given its length $r + \delta - 1$ and its minimum distance δ . Let us call κ_B the value of an arbitrary upper bound on the maximal dimension of a repair set. Since we never used that κ is actually tight, our previous results apply directly to codes with locality (r, δ) via the estimated dimension κ_B . Therefore, we get the following new bound.

Theorem 2. *Let \mathcal{C} be a linear (n, k, d, r, δ) -LRC over \mathbb{F}_q and κ_B the upper bound on the local dimension. Then*

$$k \leq \min_{\lambda \in \mathbb{Z}_+} \left\{ \lambda + k_{\text{opt}}^{(q)}(n - (a + 1)\mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B - b, \delta), d) \right\}, \quad (8)$$

where $a, b \in \mathbb{Z}$ are such that $\lambda = a\kappa_B + b, 0 \leq b < \kappa_B$.

It is crucial to estimate the length in the shortened part of the bound via the Griesmer bound instead of replacing it by $r + \delta - 1$. The reason is that $r + \delta - 1$ is an upper bound on the size of a repair set while we need something of the form of a lower bound. However, what we need is not exactly a lower bound since the dimension of a repair set can be lower than κ_B . We present a counter-example.

Example 3. *Let \mathcal{C} be the binary linear code given by the following generator matrix*

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

\mathcal{C} is a $(10, 3, 3, 2, 3)$ -LRC with obvious repair sets. Estimating κ_B using the Griesmer bound yields $\kappa_B = 1$. However, there is no sets with an entropy less than 2 and a size greater than $2 \cdot (r + \delta - 1) = 8$ since every set of size 8 has already an entropy equal to 3. The problem here is that the repair set of size 4, which gives the upper bound $r + \delta - 1$, has a minimum distance strictly greater than $\delta = 3$.

Using Lemma 3, we can derive a Singleton-type bound that take into consideration non-MDS repair sets.

Theorem 3. *Let \mathcal{C} be a linear (n, k, d, r, δ) -LRC over \mathbb{F}_q and κ_B the upper bound on the local dimension. Then*

$$d \leq n - \left\lceil \frac{k}{\kappa_B} \right\rceil \mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B - b, \delta) \quad (9)$$

where $b = k - 1 - \left(\left\lceil \frac{k}{\kappa_B} \right\rceil - 1 \right) \kappa_B$.

Proof. Let $a, b \in \mathbb{Z}$ be such that $k - 1 = a\kappa_B + b$ with $0 \leq b < \kappa_B$. By Lemma 3 and the proof of Theorem 1, there is a set $I \subseteq [n]$ such that $H(I) \leq k - 1$ and $|I| \geq (a + 1)\mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B - b, \delta)$. Then, the minimum distance d is bounded by

$$d \leq n - |I| \leq n - \left\lceil \frac{k}{\kappa_B} \right\rceil \mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B - b, \delta).$$

□

IV. ANALYSIS AND COMPARISONS

This section is devoted to the comparison between our bounds and the previously known bounds for LRCs. In the first part, we show that the bound (8) leads to the straightforward extension of the bound (3) for locality (r, δ) and the bound (9) leads to the Singleton-type bound (2) when the field size is sufficiently large. In the second part, we derive the asymptotic formulas of the bounds (8) and (9) to obtain the bounds on the tradeoff between the rate and the relative minimum distance of LRCs for fixed locality. This also enables a cleaner comparison between the new bounds and the bound (4) from [27]. Notice

that we do not compare our bounds to the linear programming bound derived in [27] since it is impossible to derive an asymptotic formula from it and we do not assume that the repair sets are disjoint.

Our results show that the comparison between the new asymptotic Singleton-type bound and the asymptotic version of bound (4) depends on the performance of the Griesmer bound compared to the log-convex bounds. Specifically, we give some examples where our bound is better, equal, or worse than the bound (4). On the other hand, we prove that the bound (8) is always better than the bound (4) for large relative minimum distances by using the Plotkin bound as the intrinsic bound in (8).

We start by showing that the bound (8) leads to the straightforward extension of the bound (3) for locality (r, δ) .

Corollary 2. *Let \mathcal{C} be a linear (n, k, d, r, δ) -LRC over \mathbb{F}_q . Then*

$$k \leq \min_{t \in \mathbb{Z}_+} \left\{ tr + k_{\text{opt}}^{(q)}(n - t(r + \delta - 1), d) \right\}. \quad (10)$$

This extension was presented in [10, Remark 3] for codes with (r', l) -cooperative locality. It can be adapted to LRCs by using the fact that a code with locality (r, δ) is a code with $(r' = r, l = \delta - 1)$ -cooperative locality.

Proof. Let κ_B be the upper bound on the local dimension. We want to show that for all $t \in \mathbb{Z}_+$ with $t \leq \frac{k}{r}$, there is a set I with $H(I) \leq tr$ and $|I| \geq t(r + \delta - 1)$. For t fixed, define $\lambda, a, b \in \mathbb{Z}_+$ such that $\lambda = tr = a\kappa_B + b$. By the same arguments as in the proof of Theorem 1, there exists a set I such that $H(I) \leq \lambda = tr$. It remains to show that $|I| \geq t(r + \delta - 1)$. First, we have $a \geq t$ since $a = \frac{tr-b}{\kappa_B} \geq \frac{t\kappa_B - b}{\kappa_B} = t - \frac{b}{\kappa_B}$. Now a is an integer, so $a \geq \left\lceil t - \frac{b}{\kappa_B} \right\rceil = t$. Using the fact that the Griesmer bound is greater than or equal to the Singleton bound, i.e., $\mathcal{G}(\kappa_B, \delta) \geq \kappa_B + \delta - 1$, we have

$$\begin{aligned} |I| &\geq a\mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B, \delta) - \mathcal{G}(\kappa_B - b, \delta) \\ &\geq a(\kappa_B + \delta - 1) + b \\ &= tr + a(\delta - 1) \\ &\geq t(r + \delta - 1). \end{aligned}$$

Hence, using Lemma 1 with this approximation on the size of I , we obtain the desired bound on k . \square

Now, we prove that the new Singleton-type bound can be used to obtain the bound (2).

Proposition 2. *For any linear (n, k, d, r, δ) -LRC, the bound (9) is at least as strong as the bound (2).*

Proof. We rewrite the bound of Theorem 3 to have something closer to the form of the bound (2). First, we rewrite the Griesmer bound as

$$\begin{aligned} \mathcal{G}(\kappa_B, \delta) &= \sum_{i=0}^{\kappa_B-1} \left\lceil \frac{\delta}{q^i} \right\rceil = \kappa_B + \sum_{i=0}^{\kappa_B-1} \left(\left\lceil \frac{\delta}{q^i} \right\rceil - 1 \right) \\ &= \kappa_B + \sum_{i=0}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor. \end{aligned}$$

Let $a, b \in \mathbb{Z}$ be such that $k - 1 = a\kappa_B + b, 0 \leq b < \kappa_B$. The bound of Theorem 3 can be transformed as follows:

$$\begin{aligned} d &\leq n - a\mathcal{G}(\kappa_B, \delta) - \mathcal{G}(\kappa_B, \delta) + \mathcal{G}(\kappa_B - b, \delta) \\ &= n - a \left(\kappa_B + \sum_{i=0}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor \right) - \left(b + \sum_{i=\kappa_B-b}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor \right) \\ &= n - k + 1 - \left(\left\lfloor \frac{k}{\kappa_B} \right\rfloor - 1 \right) \sum_{i=0}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor \\ &\quad - \sum_{i=\kappa_B-b}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor. \end{aligned}$$

By using the fact that $\kappa_B \leq r \Rightarrow \left\lfloor \frac{k}{\kappa_B} \right\rfloor \geq \left\lfloor \frac{k}{r} \right\rfloor$ and $\sum_{i=0}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor \geq \delta - 1$, we obtain

$$\begin{aligned} n - k + 1 - \left(\left\lfloor \frac{k}{\kappa_B} \right\rfloor - 1 \right) \sum_{i=0}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor - \sum_{i=\kappa_B-b}^{\kappa_B-1} \left\lfloor \frac{\delta - 1}{q^i} \right\rfloor \\ \leq n - k + 1 - \left(\left\lfloor \frac{k}{r} \right\rfloor - 1 \right) (\delta - 1). \end{aligned}$$

\square

This shows that the bounds (8) and (9) are at least as good as the bounds (10) and (2), respectively. Furthermore, we can see that the bounds (8) and (9) improve on the previous bounds when $\kappa_B < r$ or when $\delta > q$. The latter case is of particular interest for small alphabets. For example, when considering binary LRCs, the new bounds are already better than the bound (2) for all $\delta \geq 3$.

A. Asymptotic regime

For the rest of this section, we look at the asymptotic regime where $n \rightarrow \infty$. Let $\mathcal{R} = k/n$ be the rate of a code and $\delta_n = d/n$ its relative minimum distance. The goal is to obtain the bounds on the tradeoff between the rate and the relative minimum distance of LRCs when the locality (r, δ) is fixed and $n \rightarrow \infty$. This also makes the comparison to the bound (4) easier.

We start with the Singleton-type bound (9). By dividing the bound (9) by n and letting $n \rightarrow \infty$, its asymptotic formula is as follows :

$$\mathcal{R} \leq \frac{\kappa_B}{\mathcal{G}(\kappa_B, \delta)} (1 - \delta_n) + o(1). \quad (11)$$

Following the same method, we can derive the asymptotic version of the bound (4). For the ease of reading, we reproduce here the bound : For any (n, k, d, r, δ) -LRC over \mathbb{F}_q and $B_{l-c}(n, d)$ a bound on the cardinality of a code, which is log-convex in n and such that $B_{l-c}(0, d) = 1$, we have

$$k \leq \left(\left\lfloor \frac{n - d + 1}{r + \delta - 1} \right\rfloor + 1 \right) \log_q B_{l-c}(r + \delta - 1, \delta).$$

Its asymptotic version is therefore :

$$\mathcal{R} \leq \frac{\log_q B_{l-c}(r + \delta - 1, \delta)}{r + \delta - 1} (1 - \delta_n) + o(1). \quad (12)$$

The following table summarizes the asymptotic formulas for the Singleton-type bounds with different locality assumptions. Notice that the last three are truly comparable since they share the same locality assumptions. When looking at the table, we can see how the locality assumption reduces the rate by the fraction of the local dimension over the local size.

Singleton bound	$\mathcal{R} \leq 1 - \delta_n + o(1)$
Gopalan et al. [4]	$\mathcal{R} \leq \frac{r}{r+1}(1 - \delta_n) + o(1)$
Prakash et al. [6]	$\mathcal{R} \leq \frac{r}{r+\delta-1}(1 - \delta_n) + o(1)$
Agarwal et al. [27]	$\mathcal{R} \leq \frac{\log_q B_{l-c}(r+\delta-1, \delta)}{r+\delta-1}(1 - \delta_n) + o(1)$
Theorem 3	$\mathcal{R} \leq \frac{\kappa_B}{\mathcal{G}(\kappa_B, \delta)}(1 - \delta_n) + o(1)$

Following the method in [24], we can derive the asymptotic formula for the bound (8). Define $R_{\text{opt}}(\delta_n) = \lim_{n \rightarrow \infty} \frac{k_{\text{opt}}^{(q)}(n, \delta_n n)}{n}$. By dividing the bound (8) by n , we obtain its asymptotic version

$$\mathcal{R} \leq \min_{0 \leq x < \frac{1}{\nu}} x + (1 - x\nu) R_{\text{opt}} \left(\frac{\delta_n}{1 - x\nu} \right), \quad (13)$$

where $\nu = \mathcal{G}(\kappa_B, \delta) / \kappa_B$.

We can now compare the asymptotic formulas between (11), (13), and (12). Notice that for linear codes, $\log_q B_{l-c}(r + \delta - 1, \delta)$ is a bound on the dimension of a repair set. From now on, we denote by κ_A the bound $\kappa_A := \log_q B_{l-c}(r + \delta - 1, \delta)$. By definition, κ_A gives a valid upper bound on the local dimension in Theorem 2. However, the best upper bound on the dimension of a code is not necessarily log-convex and hence, $\kappa_B \leq \kappa_A$. In particular, the Griesmer bound on the cardinality of a code is not a log-convex function on the length n as demonstrated next.

Remember that a positive function $f(j)$ of the integer argument is called log-convex if $f(j_1)f(j_2) \leq f(j_1 - 1)f(j_2 + 1)$ for any $j_1 \leq j_2$ in the support of f . For any $[n, k, d]$ linear code \mathcal{C} over \mathbb{F}_q , the Griesmer bound on k given n and d , denoted by $\mathcal{G}_k(n, d)$, is obtained by taking the maximal $k' \in \mathbb{Z}_+$ such that $\mathcal{G}(k', d) \leq n$. Thus, it gives a bound on the cardinality, $|\mathcal{C}| \leq q^{k'}$. Let us consider the parameters $n_1 = n_2 = 8$, $d = 5$, and $q = 2$. Then, we obtain

$$\begin{aligned} \mathcal{G}_{k_1}(8, 5) &= 2, \\ \mathcal{G}_{k'_1}(7, 5) &= 1, \\ \mathcal{G}_{k'_2}(9, 5) &= 2. \end{aligned}$$

Hence, we have $2^{k_1}2^{k_2} = 2^4 > 2^{k'_1}2^{k'_2} = 2^3$ and the Griesmer bound on the cardinality of a code is not log-convex on n . Therefore, there is no obvious answer to the comparison between the bounds (11) and (12) since we need to compare $\frac{\kappa_B}{\mathcal{G}(\kappa_B, \delta)}$ and $\frac{\kappa_A}{r+\delta-1}$, and both the numerator and the denominator of the former are smaller than or equal to their respective correspondents in the latter.

To be more specific, the comparison between the two bounds (11) and (12) mainly depends on the performance of the Griesmer bound compared to the log-convex bounds. For example, if there exists a log-convex bound such that $\kappa_B = \kappa_A$

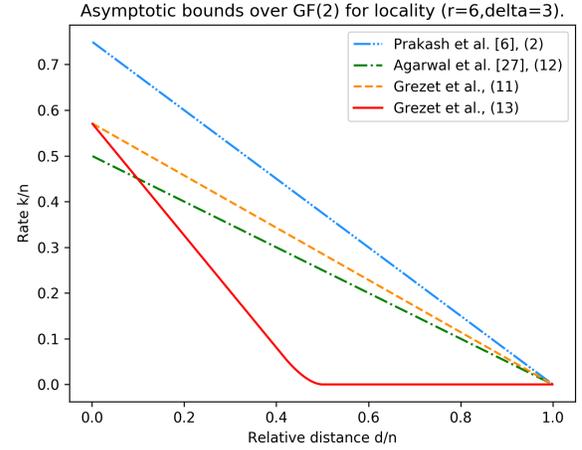


Fig. 2. Comparison of the asymptotic upper bounds on the rate k/n from [6], [27], and the bounds (11) and (13) over the binary field with fixed locality ($r = 6, \delta = 3$).

but $\mathcal{G}(\kappa_B, \delta) < r + \delta - 1$, then the bound (12) is strictly better than the new Singleton-type bound. This is illustrated in Figure 2, which displays the rate–distance tradeoff for binary codes with locality (6, 3). To evaluate the local dimension, we use the Hamming bound as a log-convex bound to get $\kappa_A = 4$, which is optimal. The Griesmer bound gives $\mathcal{G}(4, 3) = 7 < 8$. Hence the green line representing the bound (12) is better than the orange line displaying the bound (11).

On the other hand, if $\mathcal{G}_k(r + \delta - 1, \delta) < \kappa_A$ for all log-convex bounds then the bound (11) is strictly better than the bound (12) because we have $\mathcal{G}(\kappa_A, \delta) > r + \delta - 1$. Since it is impossible to give a proper example due to the fact that we would need to prove it for all log-convex bounds, we restrict here the comparison between the two bounds by considering the three bounds proven to be log-convex in [27], namely the Singleton, Hamming and Plotkin bounds. Let \mathcal{C} be a linear LRC with locality (12, 9). The Singleton bound gives an upper bound on the local dimension of 12 and the Hamming bound gives a bound of 7. The Plotkin bound is not applicable here since $\delta < (r + \delta - 1)/2$. Now, the Griesmer bound on the dimension of a code gives an upper bound of 5 and is therefore better than the Hamming bound. As we can see in Figure 3 displaying the asymptotic bounds for binary codes with locality (12, 9), the bound (11) in orange is always better than the bound (12) in green.

Finally, if $\kappa_A = \kappa_B$ and $\mathcal{G}(\kappa_A, \delta) = r + \delta - 1$, the two bounds become the same. This happens, for example, in Figure 1 where the locality is (4, 3) and both the Griesmer bound and the Plotkin bound give $\kappa_A = 3$ and $\mathcal{G}(3, 3) = 6 = r + \delta - 1$.

Nonetheless, these are just special cases of the comparison between the bounds (11) and (12) and the final comparison needs to be done case-specific. In particular, it depends on three parameters impossible to compute theoretically, which are the best upper bound on the dimension of a code κ_B , the best log-convex upper bound on the dimension of a code κ_A , and the performance of the Griesmer bound regarding both κ_A and κ_B .

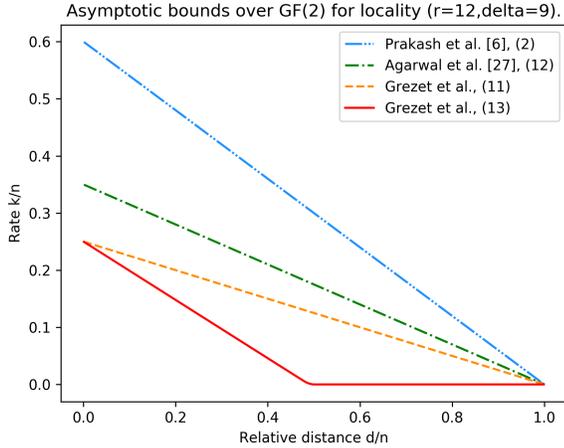


Fig. 3. Comparison of the asymptotic upper bounds on the rate k/n from [6], [27], and the bounds (11) and (13) over the binary field with fixed locality ($r = 12, \delta = 9$).

The comparison between the bound (12) and the bound (13) is more straightforward. Since the latter is at least as good as the bound (11), we automatically get that the bound (13) is stronger than the bound (12) when the new Singleton-type is stronger than or equal to the bound (12). Furthermore, we will see that the bound (13) is always stronger than the bound (12) for large relative minimum distances, *i.e.*, there is a threshold value δ_t such that for all relative minimum distances $\delta_n \in [\delta_t, 1]$, the bound (13) is better than the bound (12).

To prove this, we use the asymptotic Plotkin bound [35, Theorem 5.2.5] for R_{opt} given by

$$R_{\text{opt}}(\delta_n) \leq 1 - \frac{q}{q-1} \delta_n + o(1).$$

Combining it with the bound (13) and solving the optimization problem yields the following bound on the rate

$$\mathcal{R} \leq \frac{\kappa_B}{\mathcal{G}(\kappa_B, \delta)} \left(1 - \frac{\delta_n}{1 - 1/q} \right) + o(1). \quad (14)$$

We can now state formally our claim.

Proposition 3. *Let \mathcal{C} be a linear (n, k, d, r, δ) -LRC and κ_A be the bound on the local dimension given by the best log-convex bound $B_{l-c}(n, d)$. Assume that $\mathcal{G}(\kappa_A, \delta) < r + \delta - 1$ and let*

$$\delta_t := \frac{1}{1 + \frac{1}{q-1} \left(\frac{1}{1 - \frac{\mathcal{G}(\kappa_A, \delta)}{r + \delta - 1}} \right)}.$$

Then the bound (14) is stronger than the bound (12) for all relative minimum distances $\delta_n \in (\delta_t, 1]$.

Proof. The proof is given in the appendix. \square

The proof follows from the fact that the bound (14) using κ_A and the bound (12) are two lines with different slopes and that the bound (14) is equal to 0 when δ_n is larger than $\frac{q-1}{q}$. Thus, the two lines intersect exactly in δ_t and the bound (14) is better than the bound (12) for relative minimum distances strictly greater than δ_t .

Finally, any bound on the rate improving on the asymptotic Plotkin bound will thus increase the size of the interval where the bound (14) is better than the bound (12). In particular, this is true for the rate–distance bound given in [31], which is the best known bound for binary code. The MRRW bound is as follows:

$$\mathcal{R}(\delta_n) \leq h(1/2 - \sqrt{\delta_n(1 - \delta_n)}) + o(1), \quad (15)$$

where $h(x) := -x \log_2 x - (1 - x) \log_2(1 - x)$ is the binary entropy function.

We can thus replace the asymptotic Plotkin bound with the MRRW bound in (13). By numerically solving the optimization problem, we obtain the red curve in Figures 1, 2, and 3. We see that the bound (13) combining with the MRRW bound improves significantly on the bound (12) even when the maximal size of a repair set is larger than the Griesmer bound.

V. ACHIEVABILITY RESULTS

Several constructions of LRCs achieving the bound (2) already exist, for example in [6], [14], [15], [19], [20]. Many of these constructions require an alphabet size to be exponential in the code length. A construction of LRCs with locality (r, δ) achieving the bound (2) and with linear field size was presented in [17]. Since the bound (8) approaches the bound (2) for large alphabets, the bound (8) is tight. In this section, we show that the bound (8) is also tight for some parameter values for every fixed field size, in particular small ones, by considering the family of the Simplex codes.

Definition 6. *Let $n = \frac{q^m - 1}{q - 1}$ and $G_{m,q}(\mathcal{C})$ be an $m \times n$ matrix over \mathbb{F}_q with non-zero pairwise independent columns. The code \mathcal{C} generated by $G_{m,q}(\mathcal{C})$ is called a q -ary Simplex code $S(m, q)$ with parameters $[(q^m - 1)/(q - 1), m, q^{m-1}]$.*

Since the Simplex codes are known to achieve the Griesmer bound, they will achieve the bound (8) by taking $\lambda = 0$ and using the Griesmer bound for k_{opt} . Therefore, the locality parameters do not influence the optimality of the code. This is in fact true in general. If a code already achieves a bound on $[n, k, d]$ without locality constraints and has a certain locality, then it will be an optimal locally repairable code for these locality parameters by the bound (8).

The Simplex codes, as locally repairable codes with $\delta = 2$, were already considered in [24]. In [25], the authors used them to construct new LRCs. Here, we want to derive the locality for larger dimensions and $\delta > 2$.

The first thing to notice is that, for every coordinate $e \in [n]$, there exists a codeword c in $S(m, q)$, different from the zero codeword, such that $c_e = 0$. Indeed, it is enough to take two different codewords not a multiple of each other and subtract them in an appropriate manner to obtain the desired codeword. Since every codeword of $S(m, q)$ has the same weight, we can take the residual code associated to c , which is the Simplex code $S(m - 1, q)$. By recursion, for every $1 \leq \kappa \leq m$, the coordinate e is contained in a Simplex code $S(\kappa, q)$ obtained by a restriction of $S(m, q)$. Since the minimum distance of $S(\kappa, q)$ is $q^{\kappa-1}$, by letting $\kappa \geq 2$, we ensure that $\delta > 1$. Hence,

the Simplex code $S(m, q)$ has dimension-locality $(\kappa, q^{\kappa-1})$ for all $2 \leq \kappa \leq m$. Finally, $r = \frac{q^\kappa - 1}{q - 1} - q^{\kappa-1} + 1 = \frac{q^{\kappa-1} + q - 2}{q - 1}$.

To get examples that achieve the bounds (8) and (9) in a less obvious manner, we prove that all the examples presented in this paper are optimal. Let us start with Example 1, where the code \mathcal{C} is a binary $(10, 4, 4, 4, 3)$ -LRC with $\kappa = 3$. By using the Plotkin bound, we get $\kappa_B = \kappa = 3$ and thus $\mathcal{G}(3, 3) = 6$. We can now compute the bound (9):

$$d \leq 10 - \left\lceil \frac{4}{3} \right\rceil \cdot 6 + 6 = 4.$$

Hence, \mathcal{C} is optimal.

In Example 2, we presented a binary code with parameters $(13, 6, 3, 5, 3)$ and $\kappa = 3$. Since the purpose of this example is to illustrate the fact that κ_B might be not equal to κ , it is necessary here to use the bound (6) instead of the bound (8). Let $\lambda = 5 = \kappa + 2$. By using the Plotkin bound for $k_{\text{opt}}^{(2)}$, we have

$$k \leq 5 + k_{\text{opt}}^{(2)}(13 - 2 \cdot 6 + \mathcal{G}(1, 3), 3) = 5 + k_{\text{opt}}^{(2)}(4, 3) = 6.$$

Hence, this code reaches the bound (6).

Finally, the binary code in Example 3 has parameters $(10, 3, 3, 2, 3)$ and $\kappa = 1$. By using the Plotkin bound, we get $\kappa_B = \kappa = 1$ and $G(1, \delta) = 3$. Let $\lambda = 2 = 2\kappa$. We compute the bound (8) using the Plotkin bound for $k_{\text{opt}}^{(2)}$. We have

$$k \leq 2 + k_{\text{opt}}^{(2)}(10 - 2 \cdot 3, 3) = 2 + k_{\text{opt}}^{(2)}(4, 3) = 3.$$

Hence, the code is an optimal LRC.

Interestingly enough, to prove the optimality of both codes from Examples 2 and 3, we used a set of entropy $k - 1$ in the bounds (6) and (8). However, none of the codes reaches the Singleton-type bound (9) obtained via a set of the same entropy. This is because we have an extra dependency on the field size by the bound $k_{\text{opt}}^{(q)}$. Indeed there is no binary code with parameters $[4, 2, 3]$, while there is already an MDS code satisfying these parameters over \mathbb{F}_3 .

Finding a good family of LRCs achieving the bounds (6) or (8) is left for future work.

VI. CONCLUSION

In this paper, we highlighted the fact that the parameter r of an LRC with locality (r, δ) is only a loose upper bound on the number of nodes required for a local repair when the repair sets do not yield MDS codes. To remedy this, we proposed a new definition of locality called dimension-locality that restricts the dimension of the local codes instead of the size of the repair sets. We derived a new alphabet-dependent bound for LRCs of the same form as the bound in [24], which intrinsically uses the Griesmer bound. We showed that the bound is at least as good as the extension of the bound [24] for LRCs with locality (r, δ) and the Singleton-type bound. We derived an asymptotic upper bound on the rate–distance tradeoff of LRCs and proved that the bound is the tightest known upper bound for large relative minimum distances. Finally, we derived the locality parameters of the Simplex codes for every local dimension to show that our bound is tight for every field size.

VII. ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and the associate editor for a number of very helpful comments that improved the presentation and quality of this paper. The first author would like to thank the Department of Electrical and Computer Engineering at the Technical University of Munich for hosting him while part of this work was carried out.

REFERENCES

- [1] 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.
- [2] V. Guruswami and M. Wootters, "Repairing Reed-Solomon codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5684–5698, Sep. 2017.
- [3] A. S. Rawat, I. Tamo, V. Guruswami, and K. Efremenko, "MDS code constructions with small sub-packetization and near-optimal repair bandwidth," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6506–6525, Oct. 2018.
- [4] 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.
- [5] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, Jul. 2012, pp. 2771–2775.
- [6] 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*, Cambridge, MA, Jul. 2012, pp. 2776–2780.
- [7] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with local regeneration," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1606–1610.
- [8] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6979–6987, Nov. 2014.
- [9] N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with locality for two erasures," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, Jun. 2014, pp. 1962–1966.
- [10] A. S. Rawat, A. Mazumdar, and S. Vishwanath, "Cooperative local repair in distributed storage," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 107, Dec. 2015.
- [11] A. Mazumdar, "Storage capacity of repairable networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5810–5821, Nov. 2015.
- [12] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath, "Locality and availability in distributed storage," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4481–4493, Aug. 2016.
- [13] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Proc. 6th IEEE Int. Symp. Netw. Comput. Appl.*, Cambridge, MA, Jul. 2007, pp. 79–86.
- [14] G. M. Kamath, N. Silberstein, N. Prakash, A. S. Rawat, V. Lalitha, O. O. Koyluoglu, P. V. Kumar, and S. Vishwanath, "Explicit MBR all-symbol locality codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 504–508.
- [15] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 212–236, Jan. 2014.
- [16] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6661–6671, Dec. 2016.
- [17] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.
- [18] S. Goparaju and R. Calderbank, "Binary cyclic codes that are locally repairable," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, Jun. 2014, pp. 676–680.
- [19] T. Ernvall, T. Westerback, R. Freij-Hollanti, and C. Hollanti, "Constructions and properties of linear locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1129–1143, Mar. 2016.
- [20] T. Westerback, R. Freij-Hollanti, T. Ernvall, and C. Hollanti, "On the combinatorics of locally repairable codes via matroid theory," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5296–5315, Oct. 2016.
- [21] I. Tamo, A. Barg, and A. Frolov, "Bounds on the parameters of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3070–3083, Jun. 2016.

- [22] P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siegel, "Binary linear locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6268–6283, Nov. 2016.
- [23] R. Freij-Hollanti, C. Hollanti, and T. Westerbäck, "Matroid theory and storage codes: bounds and constructions," in *Network Coding and Subspace Designs*. Springer, Jan. 2018, pp. 385–425.
- [24] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5787–5794, Nov. 2015.
- [25] N. Silberstein and A. Zeh, "Anticode-based locally repairable codes with high availability," *Designs, Codes and Cryptography*, vol. 86, no. 2, pp. 419–445, Feb. 2018.
- [26] A. Zeh and E. Yaakobi, "Optimal linear and cyclic locally repairable codes over small fields," in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, Apr. 2015, pp. 1–5.
- [27] A. Agarwal, A. Barg, S. Hu, A. Mazumdar, and I. Tamo, "Combinatorial alphabet-dependent bounds for locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3481–3492, May 2018.
- [28] M. Grezet, R. Freij-Hollanti, T. Westerbäck, O. Olmez, and C. Hollanti, "Bounds on binary locally repairable codes tolerating multiple erasures," in *The International Zurich Seminar on Information and Communication (IZS)*, Proceedings, Zurich, Switzerland, Feb. 2018, pp. 103–107.
- [29] J. H. Griesmer, "A bound for error-correcting codes," *IBM Journal of Research and Development*, vol. 4, no. 5, pp. 532–542, 1960.
- [30] G. Solomon and J. J. Stiffler, "Algebraically punctured cyclic codes," *Information and Control*, vol. 8, no. 2, pp. 170–179, Apr. 1965.
- [31] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 157–166, Mar. 1977.
- [32] T. H. Chan, A. Grant, and T. Britz, "Properties of quasi-uniform codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, Jun. 2010, pp. 1153–1157.
- [33] T. Westerbäck, M. Grezet, R. Freij-Hollanti, and C. Hollanti, "On the polymatroidal structure of quasi-uniform codes with applications to heterogeneous distributed storage," in *International Symposium on Mathematical Theory of Networks and Systems*, Hong Kong, Hong Kong, Jul. 2018, pp. 641–647.
- [34] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Information and control*, vol. 39, no. 1, pp. 55–72, Oct. 1978.
- [35] J. H. van Lint, *Introduction to coding theory*, 3rd ed., ser. Graduate Texts in Mathematics. Berlin: Springer-Verlag, 1999, vol. 86.

APPENDIX

Proof of Lemma 2. Remember that $H(R) = H(\text{cl}(R))$ and $d_{C|\text{cl}(R)} \geq \delta$. Now there exists S in the *res*-chain of $\text{cl}(R)$ such that $H(S) - H(S \cap F) = \alpha$. Indeed if S_i and S_{i+1} are two consecutive sets in the *res*-chain of $\text{cl}(R)$, i.e., $S_{i+1} \subset S_i$ and $H(S_{i+1}) = H(S_i) - 1$, then we have

$$\begin{aligned} H(S_1) - H(F \cap S_1) &= H(S_2) + 1 - H(F \cap S_1) \\ &\leq H(S_2) + 1 - H(F \cap S_2). \end{aligned}$$

Therefore, the entropy difference between S_1 and its intersection with F increases by at most 1 compared to the entropy difference between S_2 and F . Since the last set in the *res*-chain of $\text{cl}(R)$ has entropy equal to 0, we reach all possible integers in between 0 and $H(\text{cl}(R)) - H(F \cap \text{cl}(R))$ by evaluate every set in the *res*-chain of $\text{cl}(R)$.

By construction of the *res*-chain of $\text{cl}(R)$, we have $d_S \geq \left\lfloor \frac{d_{\text{cl}(R)}}{q^{H(R)-H(S)}} \right\rfloor \geq \left\lfloor \frac{\delta}{q^{H(R)-H(S)}} \right\rfloor$. Thus, we can estimate the size of the added part via the Griesmer bound to have

$$\begin{aligned} |S| - |F \cap S| &\geq \mathcal{G}(H(S) - H(F \cap S), d_S) \\ &\geq \mathcal{G}\left(H(S) - H(F \cap S), \left\lfloor \frac{\delta}{q^{H(R)-H(S)}} \right\rfloor\right). \end{aligned}$$

Now since $H(S) \geq \alpha$, we have $H(R) - H(S) \leq \kappa - \alpha$. Hence we have

$$\begin{aligned} \mathcal{G}\left(H(S) - H(F \cap S), \left\lfloor \frac{\delta}{q^{H(R)-H(S)}} \right\rfloor\right) \\ \geq \mathcal{G}\left(\alpha, \left\lfloor \frac{\delta}{q^{\kappa-\alpha}} \right\rfloor\right). \end{aligned}$$

Thus, by defining $F' = \text{cl}(F \cup S)$, we have indeed

- $H(F') \leq H(F) + H(S) - H(F \cap S) = H(F) + \alpha$,
- $|F'| \geq |F| + |S| - |F \cap S| \geq |F| + \mathcal{G}\left(\alpha, \left\lfloor \frac{\delta}{q^{\kappa-\alpha}} \right\rfloor\right)$.

□

Proof of Lemma 3. The idea of the proof is the following. We first add repair sets to F as long as they do not increase the entropy too much. When there is one repair set that, when added, exceeds the bound on the entropy, we use Lemma 2 to add a smaller part and complete the set. We then prove that the corrected set has the desired size.

We first define recursively the set that contains F and some repair sets conditioned on the fact that there is no repair sets satisfying the hypotheses of Lemma 2. The construction is the following.

- 1) Define $F_0 = F$ and $\gamma_0 = 0$.
- 2) If for all repair sets R we have $H(R) - H(F_{i-1} \cap R) < \kappa - \gamma_{i-1}$ then choose R_i a repair set such that $\text{cl}(R_i) \not\subseteq F_{i-1}$ and define
 - $F_i = \text{cl}(F_{i-1} \cup R_i)$,
 - $\gamma_i = \gamma_{i-1} + H(R_i) - H(F_{i-1} \cap R_i)$.

The construction ends when there is at least one repair set with $H(R) - H(F_{i-1} \cap R) \geq \kappa - \gamma_{i-1}$. To see that the procedure always stops, assume for a contradiction that there is $j \in \mathbb{Z}_+$ such that $\{R : R \text{ repair set}\} \subseteq F_j$. Since every coordinate $e \in [n]$ is contained in a repair set, we have $F_j = [n]$. But by construction, we have $H(R_j) - H(F_{j-1} \cap R_j) < \kappa - \gamma_{j-1}$. This implies that

$$\begin{aligned} k = H(F_j) &\leq H(F_{j-1}) + H(R_j) - H(F_{j-1} \cap R_j) \\ &< H(F_{j-1}) + \kappa - \gamma_{j-1} \\ &\leq H(F) + \gamma_{j-1} + \kappa - \gamma_{j-1} \\ &= H(F) + \kappa, \end{aligned}$$

which contradict the assumption that $H(F) + \kappa \leq k$.

Let F_j be the set obtain by the previous algorithm and let R' the repair set such that $\text{cl}(R') \not\subseteq F_j$ and $H(R') - H(R' \cap F_j) \geq \kappa - \gamma_j$. We first give an upper bound on the entropy and a lower bound on the size of F_j . By construction, we have $H(F_j) \leq H(F) + \gamma_j$. For the size of F_j , we will use the following approximation of $\mathcal{G}(a, \delta) + \mathcal{G}(b, \delta)$ for $a, b \in \mathbb{Z}_+$. We have

$$\mathcal{G}(a, \delta) + \mathcal{G}(b, \delta) \geq \mathcal{G}(a, \delta) + \mathcal{G}\left(b, \left\lfloor \frac{\delta}{q^a} \right\rfloor\right) = \mathcal{G}(a + b, \delta).$$

We can now give an estimation of the size of F_j .

$$\begin{aligned}
|F_j| &\geq |F_{j-1}| + |R_j| - |F_{j-1} \cap R_j| \\
&\geq |F_{j-1}| + \mathcal{G}(\gamma_j - \gamma_{j-1}, \delta) \\
&\geq |F| + \sum_{i=1}^j \mathcal{G}(\gamma_i - \gamma_{i-1}, \delta) \\
&\geq |F| + \mathcal{G}\left(\sum_{i=1}^j (\gamma_i - \gamma_{i-1}), \delta\right) \\
&\geq |F| + \mathcal{G}(\gamma_j, \delta).
\end{aligned}$$

Now we can apply Lemma 2 to F_j by letting $\alpha = \kappa - \gamma_j$. Let F_c be the set obtained from Lemma 2. We check that F_c has indeed the desired entropy and size. For the entropy, we have

$$H(F_c) \leq H(F_j) + \kappa - \gamma_j \leq H(F) + \kappa.$$

For the size of F_c , we have

$$\begin{aligned}
|F_c| &\geq |F_j| + \mathcal{G}\left(\kappa - \gamma_j, \left\lceil \frac{\delta}{q^{\gamma_j}} \right\rceil\right) \\
&\geq |F| + \mathcal{G}(\gamma_j, \delta) + \mathcal{G}\left(\kappa - \gamma_j, \left\lceil \frac{\delta}{q^{\gamma_j}} \right\rceil\right) \\
&= |F| + \mathcal{G}(\kappa, \delta).
\end{aligned}$$

Hence F_c has both the required entropy and size and thus this concludes the proof. \square

Proof of Theorem 1. Let $\lambda, a, b \in \mathbb{Z}_+$ such that $0 \leq \lambda \leq k$ and $\lambda = a\kappa + b$ with $0 \leq b < \kappa$. The proof is divided into three parts. First we prove that there exists a starting set F_s with

- $H(F_s) \leq b$,
- $|F_s| \geq \mathcal{G}\left(b, \left\lceil \frac{\delta}{q^{\kappa-b}} \right\rceil\right)$.

To do this, we distinguish two cases. If there exists a repair set R such that $H(R) \geq b$, then we can use Lemma 2 with $F = \text{cl}(\emptyset)$ and $\alpha = b$ since $H(F) = 0$. Define F_s to be the set obtained from Lemma 2. F_s has indeed the required entropy and size.

If there is no such repair set, it means that b is a tighter upper bound on the entropy than κ . Therefore, \mathcal{C} is also an $(n, k, d)(b, \delta)$ -LRC. By Lemma 3, if $F = \text{cl}(\emptyset)$, then there exists F_s with $H(F_s) \leq b$ and $|F_s| \geq |\text{cl}(\emptyset)| + \mathcal{G}(b, \delta) \geq \mathcal{G}\left(b, \left\lceil \frac{\delta}{q^{\kappa-b}} \right\rceil\right)$.

Secondly, we iterate Lemma 3 a times. Let I be the final set. We have

- $H(I) \leq H(F_s) + a\kappa \leq a\kappa + b = \lambda$,
- $|I| \geq a\mathcal{G}(\kappa, \delta) + |F_s| \geq (a+1)\mathcal{G}(\kappa, \delta) - \mathcal{G}(\kappa - b, \delta)$.

Finally, now that we have the existence of this set, by the same argument as in [24] with Lemma 1, we can approximate the entropy of the shortened part by k_{opt} with n minus the bound on the size and the minimum distance d . Then, the best upper bound on k is the minimum bound obtained over all λ . This concludes the proof. \square

Proof of Proposition 3. We use κ_a as the upper bound on the dimension in (14) and let δ_n be such that $\delta_n \geq \delta_t$. We have

$$\begin{aligned}
\delta_n &\geq \frac{1}{1 + \frac{1}{q-1} \left(\frac{1}{1 - \frac{\mathcal{G}(\kappa_a, \delta)}{r+\delta-1}} \right)} && \iff \\
\delta_n &\geq \frac{1}{1 + \frac{r+\delta-1}{(q-1)(r+\delta-1-\mathcal{G}(\kappa_a, \delta))}} && \iff \\
\delta_n &\geq \frac{1}{\frac{q}{q-1}(r+\delta-1) - \mathcal{G}(\kappa_a, \delta)} && \iff \\
\delta_n \kappa_a &\geq \kappa_a \frac{r + \delta - 1 - \mathcal{G}(\kappa_a, \delta)}{\frac{q}{q-1}(r + \delta - 1) - \mathcal{G}(\kappa_a, \delta)}
\end{aligned}$$

Then, we have

$$\begin{aligned}
\delta_n \kappa_a &\left(\mathcal{G}(\kappa_a, \delta) - \frac{q}{q-1}(r + \delta - 1) \right) \\
&\leq \mathcal{G}(\kappa_a, \delta) \kappa_a - (r + \delta - 1) \kappa_a && \iff \\
\delta_n \kappa_a &\frac{\mathcal{G}(\kappa_a, \delta) - \frac{q}{q-1}(r + \delta - 1)}{\mathcal{G}(\kappa_a, \delta)(r + \delta - 1)} \\
&\leq \frac{\mathcal{G}(\kappa_a, \delta) \kappa_a - (r + \delta - 1) \kappa_a}{\mathcal{G}(\kappa_a, \delta)(r + \delta - 1)} && \iff \\
\delta_n &\left(\frac{\kappa_a}{r + \delta - 1} - \frac{\kappa_a}{\mathcal{G}(\kappa_a, \delta)} \frac{q}{q-1} \right) \\
&\leq \frac{\kappa_a}{r + \delta - 1} - \frac{\kappa_a}{\mathcal{G}(\kappa_a, \delta)} && \iff \\
\frac{\kappa_a}{\mathcal{G}(\kappa_a, \delta)} &\left(1 - \frac{\delta_n}{1 - 1/q} \right) \leq \frac{\kappa_a}{r + \delta - 1} (1 - \delta_n)
\end{aligned}$$

\square