

Locally Recoverable Codes correcting many erasures over small fields

E. Ballico¹

Received: 14 January 2021 / Revised: 16 June 2021 / Accepted: 17 June 2021 / Published online: 6 July 2021 © The Author(s) 2021

Abstract

We define linear codes which are s-Locally Recoverable Codes (or s-LRC), i.e. codes which are LRC in s ways, the case s=1 roughly corresponding to the classical case of LRC codes. We use them to describe codes which correct many erasures, although they have small minimum distance. Any letter of a received word may be corrected using s different local codes. We use the Segre embedding of s local codes and then a linear projection.

Keywords Locally recoverable codes · Erasures · MDS

Mathematics Subject Classification Primary: 11T71 · Secondary: 94B27 · 94B65

1 Introduction

Fix a prime q. The goal is to find long Locally Recoverable Codes (also known as Locally Reparable Codes or LRC) over \mathbb{F}_q ([1–3,8,15–17,20]). We do this using s layers of MDS codes in which the layers are structured in the shape of an MDS code. Our codes have not good minimum distance, but (in our opinion) they are excellent for erasures, because they have the following additional structure.

Definition 1.1 Fix positive integers s and n. Let \mathcal{C} be a length n code. We say that \mathcal{C} is s-recoverable or s-LRC if for each letter $x \in \{1, \ldots, n\}$ there are codes $\mathcal{C}_1(x), \ldots, \mathcal{C}_s(x)\}$ with each $\mathcal{C}_i(x)$ acting on a subset $S_i(x)$ of $\{1, \ldots, n\}$ and $S_i(x) \cap S_j(x) = \{x\}$ for all $i \neq j$. We say the \mathcal{C} is a linear s-LRC code if all codes \mathcal{C} and all $\mathcal{C}_i(x)$ are linear.

In our construction we will also get additional properties on the local codes $C_i(x)$, $1 \le i \le s$, $x \in \{1, ..., n\}$. Fix $i \in \{1, ..., s\}$. All local codes $C_i(x)$, $x \in \{1, ..., n\}$ are isomorphic and in particular they have the same parameters. We call the structure $\{C_i(x)\}_{x \in \{1, ..., n\}}$ the

Communicated by J. W. P. Hirschfeld.

The author was partially supported by MIUR and GNSAGA of INdAM (Italy).

⊠ E. Ballico ballico@science.unitn.it



Department of Mathematics, University of Trento, 38123 Povo, TN, Italy

2158 E. Ballico

i-th layer of the *s*-LRC code C. The *i*-th layer is associated to a single code C_i with the parameters of each C_i .

In some sense as an input we are also using well-known LRC codes, the case s=1 corresponding to the classical case of Locally Recoverable Codes, because each codeword belongs to s local codes C_1, \ldots, C_s , and if at least one local code, say C_i , has only a few erasures or errors we may correct them using only C_i (Remark 2.4). Still, these codes have only length up to $(q+1)^s$ (except a few cases for q even described in Remark 2.3). If we only take s=1 we only have small, but not too small, well-known codes. For $s\gg 0$ we get arbitrarily long codes, but at the expense of the simplicity of the codes. In any case, even for large s, the local codes have small length. Since the construction works for arbitrary linear codes C_1, \ldots, C_s we state the result for $s\geq 2$ arbitrary codes. We stress that the minimum distance of these codes is low, but we find that this type of long codes structured in layers may correct a huge number of erasures. Given an evaluation code C and a received word $w\in \mathbb{F}_q^n$ we do not need to receive other letters to get the true codeword \tilde{w} received as w (assuming there are no errors, only erasures). To get \tilde{w} from w we may work off-line. To get linear codes with arbitrarily high length, even using only s=2, it is sufficient for instance to use an AG code as one of the codes ([14,18,19,21]) and use s-1 smaller codes, say s-1 MDS codes.

In the next theorem we speak about evaluation codes ([6,7]). A non-degenerate linear code is an evaluation code ([21, Theorem 1.1.6]) and the converse holds, since every linear code is isomorphic to its bidual, but we consider directly the dual C_i and its minimum distance d_i , because it is not easy (unless C_i is an MDS) to compute d_i only knowing the minimum distance of C_i^{\perp} .

Theorem 1.2 Fix a prime power q, an integer $s \ge 2$ and s linear evaluation codes C_1, \ldots, C_s over \mathbb{F}_q with C_i an $[n_i, k_i, d_i]$ evaluation code. Set $n := \prod_{i=1}^n n_i$ and $k := \prod_{i=1}^s k_i$. Let t be the maximal non-negative integer such that

$$\frac{q^{t+1}-1}{q-1} + q^{t-1}\sqrt{q} \ge \binom{n}{2}(q-1) + \sum_{i=1}^{s} (n/n_i) \frac{q^{k_i}-1}{q-1} \tag{1}$$

Then there exists an [n, k-t] linear code over \mathbb{F}_q which corrects at least $-1 + \prod_{i=1}^s (d_i - 1)$ erasures (assuming that no error occurs).

Remark 1.3 More precisely for each integer w such that $0 \le w \le t$ we construct an [n, k-w] evaluation code \mathcal{D}_w which corrects at least $-1+\prod_{i=1}^s (d_i-1)$ erasures (assuming that no error occurs). The code \mathcal{D}_0 is just the dual of the tensor product of the codes $\mathcal{C}_1^{\perp}, \ldots, \mathcal{C}_s^{\perp}$, while all other codes \mathcal{D}_w are linear projections of \mathcal{D}_0 which preserves the length, n, the structure of s-LRC code and the ability to correct at least $-1+\prod_{i=1}^s (d_i-1)$ erasures (assuming that no error occurs). Suppose that for some $x \in \{1, \ldots, n\}$ the local codes $\mathcal{C}_i(x), 1 \le i \le s$, are not sufficient to get the value of the received word at x. Call $S \subset \{1, \ldots, n\}$ the set of all such letters x. If #S is at most the minimum distance of the global code, then we may correct the received word.

Remark 1.4 If each C_i is an evaluation MDS cods, i.e. if $d_i = k_i + 1$, each code D_w , $0 \le w \le t$, corrects k - 1 erasures (if no error occurs). If all but one local code, say C_h , has no error or erasure and $s \ge 2$, just assuming $d_h \ge 2$ for all $h \ne i$ is sufficient to correct every received word of each D_w (Remark 2.4).



2 The proof

In many papers the LRC is used with respect to disjoint supports, i.e. if we have an [n, k]-code \mathcal{U} with m local codes $\mathcal{U}_1, \ldots, \mathcal{U}_m$ we assume the existence of a partition $I_1 \sqcup \cdots \sqcup I_m$ of $\{1, \ldots, n\}$ such that each \mathcal{U}_h has as letters the elements of I_h . This is the set-up of [1,3,5] and of many other schemes for locally recoverable codes and we were inspired by these papers.

We follow [5, Sect. 3.1], [6,7] and [21, Sect. 1.1.6] to use projective geometry to handle linear codes over \mathbb{F}_q . Recall that there is a bijection equivalence classes of non-degenerate linear $[n, k, d]_q$ codes and the set of equivalence classes of projective $[n, k, d]_q$ systems, the latter being described by a set $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ spanning $\mathbb{P}^{k-1}(\mathbb{F}_q)$ and with cardinality n ([21, Theorem 1.1.6]). In the description of [21, Theorem 1.1.6] a dual vector space occurs and this explain our reason to call them evaluation codes, not just linear codes.

We recall that any linear code over a finite field is isomorphic to the dual of its dual and hence each linear code is equivalent to a dual code over the same field. It is also equivalent to a code constructed from a finite set $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ ([5]). For a finite set $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ the integer #S is the length of the equivalent code \mathcal{C} , k is an upper bound for the integer k' such that \mathcal{C} is an [n,k'] code. More precisely, $k'-1=\dim\langle S\rangle$, where $\langle \ \rangle$ denotes the linear span. Thus \mathcal{C} is an [n,k]-code if and only if S spans $\mathbb{P}^{k-1}(\mathbb{F}_q)$. We say that S or $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ is an *evaluation code*. Let $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ be an evaluation code. Its minimum distance is the smallest integer d such that there is $S' \subseteq S$ such that #S' = d and $\dim\langle S' \rangle \leq d-2$. Fix an integer $s \geq 2$. In the set-up of evaluation codes the dual of the product $\mathcal{C}_1^\perp \otimes \cdots \otimes \mathcal{C}_s^\perp$ (or tensor product as it is known in linear algebra) of the codes \mathcal{C}_i , $1 \leq i \leq s$, obtained using the set $S_i \subset \mathbb{P}^{k_i-1}(\mathbb{F}_q)$ is a code with length $\prod_{i=1}^s \#S_i$ and obtained evaluating the image $v(S) \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$, where $k := \prod_{i=1}^s k_i$, and $v : \prod_{i=1}^s \mathbb{P}^{k_i-1}(\mathbb{F}_q) \to \mathbb{P}^{k-1}(\mathbb{F}_q)$ is the Segre embedding ([10]). Note that for a fixed prime power q for $s \gg 0$ the length of $\mathcal{C}_1 \otimes \cdots \otimes \mathcal{C}_s$ goes to $+\infty$.

Remark 2.1 Let \mathcal{C} be a linear [n, k]-code associated to a finite set $S \subseteq \mathbb{P}^{k-1}(\mathbb{F}_q)$. For any $A \subseteq S$ set $\delta(A) := \#S - 1 - \dim(A)$. A set $A \subseteq S$ is the support of a codeword of \mathcal{C} if and only if $\delta(A') < \delta(A)$ for all $A' \subseteq A$, i.e. if and only if $\langle B \rangle = \langle A \rangle$ for all $B \subset A$ such that #B = #A - 1.

Take $A \subseteq S$ such that $\delta(A') < \delta(A)$ for all $A' \subseteq A$. This is the case if and only if $\langle \nu(E) \rangle = \langle \nu(A) \rangle$ for all $E \subset A$ such that #E = #A - 1. If one of the previous equivalent conditions is satisfied we say that A (or $\nu(A)$) is *equally dependent*.

Remark 2.2 Let \mathcal{C} be an [n,k,d] code over the field \mathbb{F}_q . Its dual \mathcal{C}^\perp with respect to the Hamming metric is an [n,n-k]-code over \mathbb{F}_q . It is complicated to know the minimum distance of \mathcal{C}^\perp , unless we know more about \mathcal{C} , but it would be the case in the following two important cases. If \mathcal{C} is an MDS, i.e. if d=n-k+1, then \mathcal{C}^\perp is an MDS, i.e. \mathcal{C}^\perp has minimum distance k+1. Now assume that \mathcal{C} is an AG code over a smooth genus g curve X. In this case we will not use the true minimum distance of \mathcal{C} , but only that $d \geq n-k+1-g$ ([14, Theorem 5.2.2], [19,21]). In this case \mathcal{C}^\perp is a AG code over the same smooth curve X ([19, §8.1]) and hence \mathcal{C}^\perp has minimum distance at least k+1-g. In our paper we will only use \mathcal{C}^\perp and its minimum distance k+1 or its lower bound k+1-g. This is the case because we use evaluation codes induced by finite sets $S \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$, where #S = n and S spans $\mathbb{P}^{k-1}(\mathbb{F}_q)$.

An [n, k, d] code \mathcal{C} corrects d-1 erasures (assuming as always in this paper that no error occurs) ([12, pp. 44-45], [13, p. 55]). Then \mathcal{C}^{\perp} corrects k erasures if \mathcal{C} is an MDS and at least k-g erasures if \mathcal{C} is an AG code on a smooth curve of genus g.



2160 E. Ballico

Remark 2.3 If q is even and $k_i = 3$ for some i, we may take $n_i = q + 2$ and take as C_i an MDS code for the following reason. It was classically known the existence of a set $S_2 \subset \mathbb{P}^2(\mathbb{F}_q)$ such that $\#S_2 = q + 2$ and no 3 of the point of S_2 are collinear (an arc in the terminology of [9,11]).

Remark 2.4 A linear code with minimum distance d > 2a + b corrects at least a errors and b erasures. Suppose you receive a word τ for one of the codes \mathcal{D}_w , $0 \le w \le t$, considered in Remark 1.3 and let τ_i the part of τ corresponding to the letters of the local code \mathcal{C}_i . Suppose that τ_i contains at most a errors and b erasures with $2a + b \le d_i - 1$. Then we correct τ_i using only the local code \mathcal{C}_i . Hence in the set-up of Theorem 1.2 we are allowed in each local code a small number of errors (if the local code has very few erasures), but we must correct all the errors before running the algorithm to correct the erasures, otherwise we only get garbage.

Remark 2.5 We recall that a t-blocking set of a projective space $\mathbb{P}^r(\mathbb{F}_q)$ is a set $B \subset \mathbb{P}^r(\mathbb{F}_q)$ such that $B \cap V \neq \emptyset$ for all $V \in Gr(r-t+1,r+1)(\mathbb{F}_q)$. If $B \supseteq B'$ and B' is a t-blocking set, then B is a blocking set. Any $W \in Gr(t+1,r+1)(\mathbb{F}_q)$ is a t-blocking set. A. Beutelspacher proved that $\#B \ge \frac{q^{t+1}-1}{q-1} + q^{t-1}\sqrt{q}$ if B is a t-blocking set containing no $W \in Gr(t+1,r+1)(\mathbb{F}_q)$ ([4]). We mention that blocking sets and [4] were used for the construction of LRC in [5,17].

Proof of Theorem 1.2: First of all we need to prove that some x>0 exists. The proof will also show that the code \mathcal{D}_t is layered so that it may be used as an s-LRC. Set $Y:=\prod_{i=1}^s \mathbb{P}^{k_i-1}$ and take $S_i\subset \mathbb{P}^{k_i-1}(\mathbb{F}_q)$ inducing the evaluation code \mathcal{C}_i . Let $v:Y\to \mathbb{P}^{k-1}$ denote the Segre embedding. Fix $i\in\{1,\ldots,s\}$. Let T_i be the union of the n/n_i linear subspaces of $\mathbb{P}^{k-1}(\mathbb{F}_q)$ spanned by the sets $A_1\times\cdots\times A_s$ with $A_i=S_i$ and $\#A_j=1$ for all $j\neq 0$. For any integer $w\geq 0$ let $Gr(w+1,k)(\mathbb{F}_q)$ denote the Grassmannian of all w-dimensional \mathbb{F}_q -linear subspaces of $\mathbb{P}^{k-1}(\mathbb{F}_q)$. For any $W\in Gr(w+1,k)$ let $\ell_W:\mathbb{P}^{k-1}(\mathbb{F}_q)\setminus W\to \mathbb{P}^{k-1-w}(\mathbb{F}_q)$ denote the linear projection from W. We want to find w and $w\in Gr(w+1,k)(\mathbb{F}_q)$ such that the linear projection from w induces an [n,k-w-2] codes which corrects at least δ errors and preserve the layered structure so that the corresponding code \mathcal{D}_t corrects at least $\prod_{i=1}^s (d_i-1)$ erasures. Fix an integer $w\geq 0$. We want to find a w-dimensional linear subspace w such that the linear projection from w gives an [n,k-2-w] code layered into w codes with the parameters of the codes $\mathcal{C}_1,\ldots,\mathcal{C}_s$. To get that the length of the code is w it is sufficient to find w such that $w\in S$ and w intersects no line spanned by two points of w.

Then we want to preserve the layers. We need that the projection ℓ_W preserves the layers, i.e. it sends the linear spaces giving the layers onto same dimensional linear spaces, i.e. that $W \cap M = \emptyset$ for all linear spaces giving the layers. We do not need that after the projection these linear spaces are different and we do not need that the images of linear spaces belonging to the same layer (which before the linear projection were disjoint) are mapped by ℓ_W to disjoined linear spaces. We need that the $\sum_{i=1}^s n/n_i$ linear subspace of $\mathbb{P}^{k-1}(\mathbb{F}_q)$ used do not meet W (we do not need that their images by ℓ_W are disjoint, not even that all these images are different linear subspaces).

Let B' the union of all linear spaces in the layers. For $1 \le i \le s$ there are n/n_i layers belonging to the layer corresponding to C_i and each of these linear spaces has dimension $k_i - 1$ and hence it has $\frac{q^{k_i - 1}}{q - 1}$ points. Let B'' be the union of the lines spanned by 2 of the points of S. Set $B := B' \cup B''$.

Claim 1:
$$\#B < \binom{n}{2}(q-1) + \sum_{i=1}^{s} (n/n_i) \frac{q^{k_i}-1}{q-1}$$
.



Proof of Claim 1: There are at most $\binom{n}{2}$ lines spanned by the points of S. Any line $L \subset \mathbb{P}^{k-1}(\mathbb{F}_q)$ has cardinality q+1. Since any line occurring in the definition of B'' contains 2 points of S, $\#B'' \leq \binom{n}{2}(q-1)+n$. Since $S \subset B'$, $\#(B \setminus B') \leq \binom{n}{2}(q-1)$. B'' is the union of n/n_i (k_i-1) -dimensional linear spaces, $1 \leq i \leq s$. Strict inequality holds in Claim 1 because $s \geq 2$ and linear spaces belonging to different rulings of Y meets at points defined over \mathbb{F}_q .

To conclude the proof it is sufficient to find $W \in Gr(n+1-t, n+1)(\mathbb{F}_q)$ such that $B \cap W = \emptyset$, i.e. to prove that B is not a t-blocking set in the sense of [4].

Claim 2: *B* does not contain any *t*-dimensional linear subspace.

Proof of Claim 2: Since S_i spans $\mathbb{P}^{k_i-1}(\mathbb{F}_q)$, the set S spans $\mathbb{P}^{k-1}(\mathbb{F}_q)$ and hence Claim 2 follows from the inclusion $S \subset B$.

By [4] (Remark 2.5) to conclude the proof it is sufficient to quote Claims 1 and 2.

Funding Open access funding provided by University degli Studi di Trento within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Ballico, E., Marcolla, C.: Higher Hamming weights for locally recoverable algebraic curves. Finite Fields Its Appl. 40, 61–72 (2016)
- Barg, A., Haymaker, K., Howe, E., Everett, W., Matthews, G.L., Várilly-Alvarado, A.: Locally recoverable codes from algebraic curves and surfaces. In: Howe, E.W., Lauter, K., Walker, J. (eds.) Algebraic Geometry for Coding Theory and Cryptography. Association for Women in Mathematics Series, vol. 9, pp. 95–127. Springer, Cham (2017)
- 3. Barg, A., Tamo, I., Vládut, S.: Locally recoverable codes on algebraic curves. IEEE Trans. Inform. Theory 63(8), 4928–4939 (2017)
- Beutelspacher, A.: Blocking sets and partal spreads in finite projective spaces. Geom. Dedicata 9, 425–449 (1980)
- Bogart, T., Horlemann-Trautmann, A.L., Karpuk, D., Neri, A., Velasco, M.: Constructing partial MDS codes from reducible algebraic curves, arXiv:2007.14829.
- Couvreur, A.: The dual minimum distance of arbitrary dimensional algebraic-geometric codes. J. Algebra 350(1), 84–107 (2012)
- Couvreur, A., Duursma, I.: Evaluation codes from smooth quadric surfaces and twisted Segre varieties. Des. Codes Cryptogr. 66(1–3), 291–303 (2013). https://doi.org/10.1007/s10623-012-9692-4
- Gopalan, P., Huang, C., Jenkins, B., Yekhanin, S.: Explicit maximally recoverable codes with locality. IEEE Trans. Inform. Theory 65(6), 3692–3701 (2018)
- 9. Hirschfeld, J.W.P.: Projective Geometries Over Finite Fields. Clarendon Press, Oxford (1979)
- Hirschfeld, J.W.P., Thas, J.A.: General Galois Geometries. Oxford Mathematical Monographs. Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, New York (1991)
- Hirschfeld, J.W.P., Thas, J.A.: Open problems in finite projective spaces. Finite Fields Appl. 32, 44–81 (2015)
- Huffman, W.C., Press, V.: Fundamentals of Error Correcting Codes. Cambridge University Press, Cambridge, New York (2003)
- 13. Morelos-Zaragoza, R.H.: The Art of Error Correcting Coding. Wiley, West Sussex (2002)
- Niederreiter, H., Xing, C.P.: Algebraic Geometry in Coding Theory and Cryptography. Princeton University Press, Princeton, NJ (2009)



2162 E. Ballico

 Papailiopoulos, D.S., Dimakis, A.G.: Locally reparaible codes. IEEE Trans. Inform. Theory 60(10), 5843–5855 (2014)

- 16. Rawat, A.S., Koyluoglu, O.O., Silberstein, N., Vishwanath, S.: Optimal locally reparable and secure codes for distributed storage systems. IEEE Trans. Inform. Theory **60**(1), 212–236 (2013)
- Silberstein, N., Zeh, A.: Anticode-based locally repairable codes with high availability. Des. Codes. Cryptogr. 86, 419–445 (2018)
- 18. Stepanov, S.A.: Codes on Algebraic Curves. Springer, Berlin (1999)
- 19. Stichtenoth, Algebraic: function Fields and Codes, 2nd edn. Springer, Berlin (2009)
- Tamo, I., Ye, M., Barg, A.: The repair problem for Reed-Solomon codes: optimal repair of single and multiple erasures with almost optimal node size. IEEE Trans. Inform. Theory 65(5), 2673–2695 (2019)
- Tsfasman, M., Vlădut, S., Nogin, D.: Algebraic Geometric Codes: Basic Notions, Mathematical Surveys and Monographs, vol 139. American Mathematical Society, Washington

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

