Electronic Colloquium on Computational Complexity, Report No. 62 (2024)

Near-Tight Bounds for 3-Query Locally Correctable Binary Linear Codes via Rainbow Cycles

Omar Alrabiah* Venkatesan Guruswami†

Abstract

We prove that a binary linear code of block length n that is locally correctable with 3 queries against a fraction $\delta > 0$ of adversarial errors must have dimension at most $O_\delta(\log^2 n \cdot \log\log n)$. This is almost tight in view of quadratic Reed-Muller codes being a 3-query locally correctable code (LCC) with dimension $\Theta(\log^2 n)$. Our result improves, for the binary field case, the $O_\delta(\log^8 n)$ bound obtained in the recent breakthrough of [KM23a] (and the more recent improvement to $O_\delta(\log^4 n)$ for binary linear codes announced in [Yan24]).

Previous bounds for 3-query linear LCCs proceed by constructing a 2-query locally decodable code (LDC) from the 3-query linear LCC/LDC and applying the strong bounds known for the former. Our approach is more direct and proceeds by bounding the covering radius of the dual code, borrowing inspiration from [IS20]. That is, we show that if $x \mapsto (v_1 \cdot x, v_2 \cdot x, \ldots, v_n \cdot x)$ is an arbitrary encoding map $\mathbb{F}_2^k \to \mathbb{F}_2^n$ for the 3-query LCC, then all vectors in $\mathbb{F}_2^k$ can be written as a $\widetilde{O}_\delta(\log n)$-sparse linear combination of the v_i's, which immediately implies $k \leqslant \widetilde{O}_\delta((\log n)^2)$. The proof of this fact proceeds by iteratively reducing the size of any arbitrary linear combination of at least $\widetilde{\Omega}_\delta(\log n)$ of the v_i's. We achieve this using the recent breakthrough result of [ABS+23] on the existence of rainbow cycles in properly edge-colored graphs, applied to graphs capturing the linear dependencies underlying the local correction property.

1 Introduction

Local correction refers to the notion of correcting a single bit of a received codeword by querying very few other bits of the codeword at random. More concretely, a binary code, which is simply a subset $C \subseteq \{0,1\}^n$, is said to be *locally correctable* using $r \in \mathbb{N}$ queries from a fraction $\delta \in (0,1)$ of errors, abbreviated (r,δ)-LCC, if it can recover any given bit of a codeword $c \in C$ with probability noticeably higher than $1/2$ (say $2/3$) by randomly reading r bits of a received codeword $y \in \{0,1\}^n$ that is at most δn away from c in Hamming distance. Usually, we are interested in the case when δ is a fixed constant bounded away from 0 as the code length $n \to \infty$, and in this case, we refer to such a code as simply a r-LCC.

Throughout this paper, we will restrict our attention to only binary linear codes, particularly binary linear r-LCCs. A binary linear code C of block length n is simply a subspace of $\mathbb{F}_2^n$, where $\mathbb{F}_2$ is the field of two elements. If the dimension of C as a $\mathbb{F}_2$-subspace is k, then one refers to it as an $[n,k]$ code. A generator matrix of C is an $n \times k$ matrix whose columns form a basis of C. Let

*Department of Electrical Engineering and Computer Science, UC Berkeley, Berkeley, CA, 94709, USA. Email: oalrabiah@berkeley.edu. Research supported in part by a Saudi Arabian Cultural Mission (SACM) Scholarship, NSF CCF-2210823 and V. Guruswami's Simons Investigator Award.

†Department of Electrical Engineering and Computer Science, Department of Mathematics, and the Simons Institute for the Theory of Computing, UC Berkeley, Berkeley, CA, 94709, USA. Email: venkatg@berkeley.edu. Research supported by a Simons Investigator Award and NSF grants CCF-2210823 and CCF-2228287.

ISSN 1433-8092

us fix one such choice of generator matrix M, and denote its rows by $v_1, v_2, \dots, v_n \in \mathbb{F}_2^k$. We then have the encoding map $M : \mathbb{F}_2^k \to C$ given by $Mx = [v_1 \cdot x, v_2 \cdot x, \dots, v_n \cdot x]^\top$.

Among its many uses, locally correctable codes play a central role in PCP constructions, where they allow to self-correct a function, purportedly a codeword, after a codeword test ascertains that the function is close to a codeword. They thus allow effective noise-free oracle access to a noisy function, with a small price in the number of queries. We refer the reader to the surveys [Tre04, Yek12, Gop18] for more on the applications and connections of locally correctable codes.

Despite its slew of uses, the best known r-LCCs (even existentially) have $n \approx \exp(k^{1/(r-1)})$, which is achieved by the degree $(r-1)$ Reed-Muller code (evaluations of polynomials of degree $(r-1)$ in $m = O_q(\log n)$ variables at all points in $\mathbb{F}_q^m$).[1] This has remained the case for constant-query local correction since their conception. Indeed, much of the progress on locally correctable codes for a constant number of queries has focused on proving their limitations, specifically for concrete values of r.[2] For $r = 1$, it has long been known that 1-LCCs do not exist [KT00]. For $r = 2$, it has also long been known that one must indeed have $n \geqslant \exp(\Omega_q(k))$ [GKST06, KW04], so the Hadamard code (and the degree one Reed-Muller code) is indeed optimal.

For $r = 3$ and larger, our understanding of r-LCCs is abysmal. The known limitations of r-LCCs, which also apply to r-query locally *decodable* codes (which offer the weaker guarantee of local correction only for the k message symbols encoded by the codeword), stood at the bound $k \leqslant \widetilde{O}(n^{1-1/\lceil 2/r \rceil})$ [KW04, Woo07, Woo12] for a long while. In particular, for 3-LCCs, the quadratic bound $k \leqslant O(\sqrt{n})$ stood for more than a decade. This was recently improved to $k \leqslant \tilde{O}(\sqrt[3]{n})$ in [AGKM23] (with recent logarithmic factor improvements by [HKM+24]), and this bound also applied to 3-query locally decodable codes (LDCs). Then, in a tour de force breakthrough, Kothari and Manohar [KM23a] gave an exponential improvement and showed that $k \leqslant O_q(\log^8 n)$ for 3-query *linear* LCCs (over any field $\mathbb{F}_q$). Since there are beautiful constructions of 3-query linear LDCs of block length sub-exponential in k [Yek08, Rag07, Efr12, DGY11], their bound demonstrated a strong separation between local decodability and local correctability with 3 queries for linear codes. Nonetheless, their result left open the optimality of degree 2 Reed-Muller codes as binary linear 3-LCCs, which have dimension $k = \Theta(\log^2 n)$. Our main result is that they are (almost) optimal.

Theorem 1 (Main). *If C is an $[n, k]$ binary linear $(3, \delta)$-LCC, then $k \leqslant O(\delta^{-2} \log^2 n \cdot \log \log n)$.*

Modulo the $\log \log n$ factor, this settles the dimension versus block length trade-off of 3-query *binary linear* LCCs. Recently, following [KM23a], an improved upper bound of $k \leqslant O(\log^4 n)$ was obtained for binary linear 3-LCCs in [Yan24]. Even more recently, an independent result of [KM24] shows an optimal $k \leqslant O(\log^2 n)$ bound for binary linear *design* 3-LCCs. Such 3-LCCs have the additional property that the linear dependencies of length 4 formed by the query sets (see Definition 2) cover each pair of indices in $[n]$ exactly once. We note that a weaker bound of $k \leqslant O(\log^3 n)$ for binary linear design 3-LCCs was previously shown in [Yan24].

Our proof method additionally sheds some light on the structure of binary linear 3-LCCs. Namely, we prove Theorem 1 by upper bounding the *covering radius of the dual code.*[3] This offers

---

[1] This code requires $q \geqslant r + 1$, but one can also get say binary codes by picking q to be a power of 2 and concatenating the Reed-Muller code over $\mathbb{F}_q$ with the binary Hadamard code.

[2] This statement holds only for the classical constant query regime. Indeed, there have been some great works for when the number of queries r grows with n [GKS13, KSY14, HOW15, KMRZS17, GKO+18] and for relaxed notions of local corrections [GL19, GRR20, AS21, CY22, KM23b, CY23]. There is also a brighter landscape of lower bounds for harsher error models [OPC15, BBG+20, BGGZ21, BCG+22, BBC+23, Gup23].

[3] The covering radius of a linear code $C_0 \subseteq \mathbb{F}_2^n$ is the minimum r such that every point in $\mathbb{F}_2^n$ is within Hamming distance r from some codeword $c \in C_0$. If $H \in \mathbb{F}_2^{m \times n}$ is a parity check matrix of a linear code C_0, then it is the minimum r for which every $s \in \mathbb{F}_2^m$ is the sum of at most r columns of H.

a more direct understanding of the structure and limitations of binary linear 3-LCCs, which can be harder to discern from recent developments [AGKM23, KM23a, HKM+24, Yan24]. Indeed, all such works proceed by constructing a much longer 2-query LDC from the 3-query locally correctable linear code and appealing to the known exponential lower bounds for 2-LDCs [GKST06, KW04].[4]

Our main result on the covering radius of the dual code of a binary linear 3-LCC is the following.

Theorem 2. *Let C be a binary linear $(3, \delta)$-LCC with generator matrix $M \in \mathbb{F}_2^{n\times k}$. Then every $x \in \mathbb{F}_2^k$ can be expressed as the sum of at most $O(\delta^{-2} \log n \cdot \log\log n)$ rows of M.*

Since a generator matrix of C is also a parity check matrix of $C^\perp$, Theorem 2 as stated upper bounds the covering radius of $C^\perp$. Note that Theorem 2 immediately implies Theorem 1, as it shows $2^k \leqslant \sum_{j=0}^{T} \binom{n}{T} \leqslant n^{T+1}$ for $T = O(\delta^{-2} \log n \cdot \log\log n)$. We remark here that the degree 2 Reed-Muller code has a covering radius of $\Theta(\log n)$, which makes our bound in Theorem 2 only a $\log\log n$ factor away from the optimal bound.

Our inspiration for Theorem 2 came from a work of Iceland and Samorodnitsky [IS20], who prove that the dual $C^\perp$ of a binary linear $(2, \delta)$-LCC C has $O(\delta^{-1})$ covering radius (which then immediately implies that $|C| \leqslant n^{O(\delta^{-1})}$).[5] They prove this via analysis of the "discrete Ricci curvature" of the "coset leader graph" associated with C. We develop a more elementary treatment of their ideas and give a similar coupling argument to bound the diameter of the Cayley graph $\mathrm{Cay}(\mathbb{F}_2^k, \{v_1, v_2, \ldots, v_n\})$, which is isomorphic to their coset leader graph. Note that this diameter is precisely the covering radius of $C^\perp$. Using our viewpoint, we produce a new proof of the previously known $k \leqslant O(\log n)$ upper bound for linear 2-query LDCs over any finite field (the proof in [IS20] only applied to LCCs); we present this proof in Appendix A.

Rainbow cycles in properly edge-colored graphs. Our proof of Theorem 2 crucially relies on finding *rainbow cycles* in properly edge-colored graphs. Rainbow cycles are simply cycles where each color appears at most once. There has been numerous works to that end [KMSV07, DLS13, Jan23, JS22, Tom22, KLLT22, ABS+23], culminating in the recent breakthrough of [ABS+23] showing that any properly edge-colored n-vertex graph with average degree $\Omega(\log n \cdot \log\log n)$ must have a rainbow cycle. This bound is tight up to the $O(\log\log n)$ factor—if one colors the edges of the Boolean hypercube with their respective direction, then one obtains a properly edge-colored $\log n$-regular n-vertex graph that has no rainbow cycles.

Our $O(\log n \cdot \log\log n)$ bound in our Theorem 2 is inherited in a black-box fashion from the rainbow cycle bound of [ABS+23]. Should a tight $\Theta(\log n)$ be established for the minimum average degree guaranteeing a rainbow cycle, we would immediately get an asymptotically tight $O(\log^2 n)$ dimension upper bound for binary linear 3-LCCs in Theorem 1. In fact, in our application, the concerned edge-colored graphs have the further property that each color class has $\Omega(n)$ edges. So it would suffice to improve the rainbow cycle bound for such graphs.

LCC lower bounds from rainbow LDC lower bounds. Our 3-LCC result based on rainbow cycles turns out to be a specific instance of a more general reduction from lower bounds for r-LCCs to a "rainbow" form of lower bounds for binary linear $(r-1)$-query LDCs. Our main result is the $r = 3$ case of this phenomenon, where we in fact have such strong "rainbow" bounds for binary linear 2-query LDCs. (This is also why our approach yields concrete results only for the 3-query case.) We outline this reduction in Section 4.

[4] See Appendix B of [AGKM23] and Section 7.6 of [KM23a] for the proper formulation of their blocklength lower bound proofs as reductions to 2-query LDCs.

[5] They also deduce a covering radius upper bound of $O(n^{(r-2)/(r-1)})$ for the r-query case by reducing to the 2-query case. Note that, for $r \geqslant 3$, the resulting bounds for LCCs are weaker than the best-known ones.

1.1 Proof Overview

While our proof of Theorem 1 is rather short (just 2 pages, and self-contained modulo the rainbow cycle bound), we will nonetheless present a proof overview of it to showcase its key ideas. Consider a $(3, \delta)$-LCC whose generator matrix has $v_1, v_2, \dots, v_n \in \mathbb{F}_2^k$ as rows. It is well known that any binary linear $(3, \delta)$-LCC has a collection of hypergraphs $\mathcal{H}_1, \dots, \mathcal{H}_n$ over $[n]$ such that for each $i \in [n]$, the hypergraph $\mathcal{H}_i$ consists of at least $(\delta/3)n$ disjoint subsets of $[n]$ of size 3 each such that for any hyperedge $\{a, b, c\} \in \mathcal{H}_i$, we have that $v_i = v_a + v_b + v_c$ (see Section 2.1). For simplicity, suppose that $\delta \geqslant \Omega(1)$ to ignore any δ dependencies. Our goal is to show that every $x \in \mathbb{F}_2^k$ can be represented as the sum of at most B vectors in $\{v_1, \dots, v_n\}$ for some $B \coloneqq \Theta(\log n \log \log n)$.

Since the v_i's span $\mathbb{F}_2^k$, x can be written as the sum of at most k of the v_i's. Fix any such sum. Our proof proceeds in an iterative fashion: whenever the current representation of x as a sum of the v_i's is longer than B, we will exploit the many local checks expressing each v_i as the sum of many disjoint 3-tuples of other v_j's to produce a shorter representation of x. Applying this compression iteratively yields the desired conclusion.

Now, consider an arbitrary linear combination $\sum_{t \in T} v_t$ with $|T| > B$. For any $t \in T$, we can locally modify $\sum_{t \in T} v_t$ by applying the substitution $v_t = v_a + v_b + v_c$ for any $\{a, b, c\} \in \mathcal{H}_t$. This will increase the length of the sum by (at most) 2, which defeats our initial goal. Nonetheless, since $|\mathcal{H}_t| \geqslant \Omega(n)$ for each $t \in T$, the abundance of choices for the triple $\{a, b, c\} \in \mathcal{H}_t$ presents a possibility for producing cancellations between substituted sums of triples of vectors.

The simplest form of such a cancellation between two substitutions goes as follows: consider any two distinct indices $t_1, t_2 \in T$ such that there are triples $\{a_1, b_1, c_1\} \in \mathcal{H}_{t_1}$ and $\{a_2, b_2, c_2\} \in \mathcal{H}_{t_2}$ satisfying $c_1 = b_2$. Since each hypergraph is a matching of size $\Omega(n)$, such triples do occur whenever $|T| = \omega(1)$. Now, by applying the substitutions $v_{t_1} = v_{a_1} + v_{b_1} + v_{c_1}$ and $v_{t_2} = v_{a_2} + v_{b_2} + v_{c_2}$ in $\sum_{t \in T} v_t$, we obtain a new sum of length at most $|T| + 2 \cdot 2 - 2 \cdot 1 = |T| + 2$ due to v_{c_1} and v_{b_2} canceling each other out.

We can further generalize this form of cancellation to multiple indices as follows: given distinct indices $t_1, \dots, t_m \in T$ such that there exists a "path" of hyperedges $E_s \coloneqq \{a_{E_s}, b_{E_s}, c_{E_s}\} \in \mathcal{H}_{t_s}$ for $s \in [m]$ satisfying $c_{E_s} = b_{E_{s+1}}$ for each $s \in [m-1]$, we can apply the substitutions $v_{t_s} = v_{a_{E_s}} + v_{b_{E_s}} + v_{c_{E_s}}$ for each $s \in [m]$ to the sum $\sum_{t \in T} v_t$ and obtain a new sum of length at most $|T| + 2m - 2(m-1) = |T| + 2$ due to $v_{c_{E_s}}$ and $v_{b_{E_{s+1}}}$ canceling each other out for each $s \in [m-1]$. Thus the length of the new sum hardly deviates from the length of the original sum. Furthermore, by a simple counting argument, one can show that there are such "paths" of length $m = \Omega(|T|)$. However, the length of this new sum is not smaller or even equal to the length of the original sum.

Now, notice that if we had $c_{E_m} = b_{E_1}$ (i.e., the path 'loops back'), then the length of the new sum will now be at most $|T|$. This does not reduce the length of the original sum $\sum_{t \in T} v_t$, but it does 'shift' it to a new sum. In the sequel, we will exploit such 'shifts' to produce a new sum of smaller length. For now, let us consider the feasibility of having $c_{E_m} = b_{E_1}$.

To do so, we will cast our problem in the language of *properly edge-colored graphs* and *rainbow cycles*. Indeed, consider the edge-colored graph G_T with vertices $[n]$ and edges $\{b, c\}$ for $\{a, b, c\} \in \mathcal{H}_t$ (dropping an arbitrary vertex in each triple) forming the t'th color class of edges in G_T for each $t \in T$. As $\mathcal{H}_t$ is a matching, G_T will therefore be a properly edge-colored graph. In this viewpoint, the 'path' of hyperedges $E_1, \dots, E_m$ is in fact a *rainbow path* in G_T with edge colors $t_1, \dots, t_m$ in that order. To have $c_{E_m} = b_{E_1}$, we need this rainbow path to be a *rainbow cycle*. Since the average degree of G_T equals $\Omega(|T|) = \Omega(B) = \Omega(\log n \log \log n)$, we can therefore conclude the existence of a rainbow cycle in G_T by the recent breakthrough result of [ABS+23]. This rainbow cycle gives an alternate representation $\sum_{t \in T'} v_t$ that equals $\sum_{t \in T} v_t$ with $|T'| \leqslant |T|$. Call such an T' a "shift" of

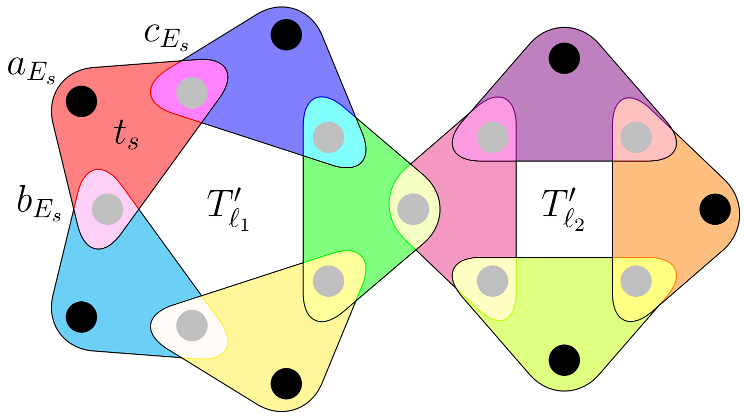

Figure 1: This figure indicates the cancellations that occur in our proof of Theorem 2 via iterative refinement of the representation of an arbitrary vector $x \in \mathbb{F}_2^k$ as a sum more than $p(B+1) = \Omega(\log n \log\log n)$ of the v_i's. The nodes represent indices in $[n]$, with the gray nodes indicating 'canceled' nodes in the sum $\sum_{i\in I} v_i$, while the black nodes represent the 'active' nodes in the sum. The inner gray nodes in the pentagon and the square are cancellations resulting from Lemma 6. The cancellation of the one outer gray node in common is the result of picking a common node between two 'shifts' T'_{ℓ_1} and T'_{ℓ_2} of the sets T_{ℓ_1} and T_{ℓ_2}, which is key idea in the proof of Theorem 2 from Lemma 6. In the figure, a sum of 9 terms (the indices t_s corresponding to each of the 9 colors) is compressed into a sum of 7 terms (the black nodes).

T. Since the hypergraphs $\{\mathcal{H}_t\}_{t\in T}$ are matchings of size $\Omega(n)$, we can in fact extract more from this argument. Specifically, by a more careful selection of the edges of G_T, we can show that the collection of all "shifts" T' of T cover $\Omega(n)$ of the indices in $[n]$. This is the content of Lemma 6.

This now suffices for an actual compression of a somewhat larger sum. Suppose $x = \sum_{i\in I} v_i$ for $|I| > p\cdot(B+1)$ for some large enough constant p (which will depend on δ). Splitting the sum into p disjoint parts $T_1, T_2, \ldots, T_p$, each with more than B terms, the constant fraction 'coverage' of $[n]$ by the "shifts" of each set T_ℓ means (by some simple pigeonholing) that we can find two distinct indices $\ell_1, \ell_2 \in [p]$ and "shifts" T'_{ℓ_1} and T'_{ℓ_2} that intersect. By replacing the sets T_{ℓ_1} and T_{ℓ_2} with their respective "shifts," we end up with a representation of x with at most $|I|-2$ of the v_i's, which concludes our iterative compression argument. See Figure 1 for an illustration.

Proof comparison to [KM23a, Yan24]. One salient common feature in our work and the works of [KM23a, Yan24] is the *chaining* of local checks. However, our implementation of chaining differs fundamentally from [KM23a, Yan24]. In our work, we attempt to chain local checks to form a "cyclical chain" (i.e., rainbow cycles) in order to establish Theorem 2, resulting in a much shorter proof. On the other hand, [KM23a, Yan24] consider a technically involved hypergraph decomposition of a superpolynomial number of chained local checks and then proceed to undertake a highly intricate "row pruning" analysis to ensure that each hypergraph of chained local checks is "spread-out." Admittedly, our proof relies on black-boxing known results from the rainbow cycle literature, some proofs of which are involved. Nonetheless, our proof offers modularity. In particular, any improvement to the result of [ABS⁺23] would immediately yield better lower bounds on binary linear 3-LCC via our proof of Theorem 2. On the other hand, improvements using the methods of [KM23a, Yan24] would likely entail a re-do of their analysis (as was the case in [Yan24]).

1.2 Organization

In Section 2, we state the tools we need for locally correctable codes and edge-colored graphs. In Section 3, we present the proof of Theorem 1 and Theorem 2. In Section 4, we discuss future directions and follow-up questions to our work. Finally, in Appendix A, we present a covering radius upper bound for linear 2-LDCs and discuss how to obtain the exponential blocklength lower bound from our proof.

2 Preliminaries

Let $\mathbb{N} \coloneqq \{0, 1, 2, \ldots\}$, and let $\mathbb{F}_2 = \{0, 1\}$ denote the finite field of size 2. For any positive integer $n \in \mathbb{Z}_+$, we denote $[n] \coloneqq \{1, 2, \ldots, n\}$. For any set X and number $k \in \mathbb{N}$, denote $\binom{X}{k} \coloneqq \{A \mid A \subseteq X\ ,\ |A| = k\}$. Given two sets A and B, let $A \oplus B \coloneqq (A \setminus B) \cup (B \setminus A)$ denote their symmetric difference. Given a vector $x \in \mathbb{F}_2^k$, let $\mathrm{wt}(x)$ denote its Hamming weight (i.e., number of nonzero entries). For any two vectors $x, y \in \mathbb{F}_2^n$, let $d(x, y)$ denote their Hamming distance (i.e., the number of entries that they differ on). We will consider *multi-sets* in this work, which are simply sets that allow elements to repeat. For any multi-set A, the cardinality of A, denoted $|A|$, is the number of elements in A (including repeated elements).

A *hypergraph* is simply a collection of sets $\mathcal{H} \subseteq 2^{[n]}$. We call the sets in the hypergraph *hyperedges* For any $\ell \in \mathbb{Z}_+$, we say that $\mathcal{H}$ is an ℓ-uniform hypergraph if $|A| = \ell$ for all $A \in \mathcal{H}$. We also say that $\mathcal{H}$ is a *matching* if $A \cap B = \varnothing$ for all distinct $A, B \in \mathcal{H}$. If $\mathcal{H}$ is an ℓ-uniform hypergraph and a matching, then we simply call it an *ℓ-uniform matching.*

2.1 Locally Correctable Codes

The following is the usual definition of a linear 3-query locally correctable code C as having a local decoder.

Definition 1 (Binary Linear LCC, local decoder definition)**.** *Given a binary linear code $C \subseteq \mathbb{F}_2^n$, we say that it is a (r, δ)-locally correctable code (abbreviated (r, δ)-LCC) for $r \in \mathbb{N}$ and $\delta \in (0, 1)$ if the following holds: for any received codeword $y \in \mathbb{F}_2^n$ there exists a randomized algorithm $\mathcal{D}^y$ with oracle access to y that takes an index $i \in [n]$ as input and satisfies the following properties: (1) $\mathcal{D}^y(i)$ makes at most r queries to y, and (2) if there exists a codeword $c \in C$ satisfying $d(x, c) \leqslant \delta n$, then $\mathcal{D}^y(i)$ outputs c_i with probability at least $2/3$.*

While Definition 1 is the typical definition of LCCs, we will instead be working with a more combinatorial definition that is amenable to lower bounds.

Definition 2 (Binary Linear LCC, combinatorial definition)**.** *Given a linear code C with generator matrix $M \in \mathbb{F}_2^{n \times k}$ whose columns form a basis for C, let $v_i \in \mathbb{F}_q^k$ be the i'th row of M for $i \in [n]$. The code C is said to be a (r, δ)-locally correctable code (abbreviated (r, δ)-LCC) for $r \in \mathbb{N}$ and $\delta \in (0, 1)$ if there exists r-uniform matchings $\mathcal{H}_1, \ldots, \mathcal{H}_n$ over $[n]$ such that $|\mathcal{H}_i| \geqslant \delta n$ for all $i \in [n]$, and for any $i \in [n]$ and $\{a_1, \ldots, a_r\} \in \mathcal{H}_i$, we have that $v_i = \sum_{s=1}^{r} v_{a_s}$.*

It is well-known from standard reductions [KT00, Yek12, DSW14] that any code satisfying Definition 1 also satisfies Definition 2 for a multiplicative loss of $1/r$ in δ. Therefore, without loss of generality. we will assume throughout the paper that the notion of a binary linear (r, δ)-LCC refers to Definition 2 rather than Definition 1.

Remark 3. *The definition of a linear (r, δ)-LCC in Definition 2 is invariant of the choice of generator matrix M for the code C. Indeed, any generator matrix for C is of the form MB for some invertible matrix $B \in \mathbb{F}_q^{k \times k}$. The rows of MB are $B^\top v_i$ for $i \in [n]$. By linearity, it therefore follows that $B^\top v_i = \sum_{s=1}^{r} B^\top v_{a_s}$ for any $i \in [n]$ and $\{a_1, \ldots, a_r\} \in \mathcal{H}_i$.*

2.2 Edge-colored Graphs

An undirected graph $G = (V, E)$ consists of a set V and a multi-set $E \subseteq \binom{V}{2}$.[6] Given two edges $e_1, e_2 \in E$, we say that e_1 is *incident* to e_2 if they share a common vertex. A subset of edges $E_0 \subseteq E$ is said to be a *matching* if no two different edges in E_0 are incident to each other. Given a set of colors T, we say that a graph G is *edge-colored* if it has an associated function $c : E \to T$, which we call an edge coloring. For graphs with an associated edge coloring, we write them as $G = (V, E, c)$. Given a color $t \in T$, the *color class* of t of G is the multi-set of edges $c^{-1}(t)$. We say that c is a *proper* edge coloring if any two different incident edges $e_1, e_2 \in E$ have different colors. Equivalently, c is a proper edge coloring if $c^{-1}(t)$ is a matching for all $t \in T$.

With all this terminology at hand, we can now define a rainbow cycle.

Definition 3 (Rainbow Cycle)**.** *Given an edge-colored graph $G = (V, E, c)$, a rainbow cycle is a tuple of vertices $(i_1, i_2, \ldots, i_\ell, i_{\ell+1} = i_1) \in V^\ell$ such that $\{i_j, i_{j+1}\} \in E$ for all $j \in [\ell]$ and the multi-set of edges $\{\{i_j, i_{j+1}\} : j \in [\ell]\}$ is each assigned a different color by c.*

We will now rely on the following theorem of [ABS+23]. Note that when the graph is not simple, one can easily find a rainbow cycle of length 2 in the graph (as it is properly edge-colored).

Theorem 4 ([ABS+23], Theorem 1.1)**.** *There exists a universal constant $c_0 > 0$ such that the following holds: any properly edge-colored n-vertex graph G with at least $c_0 n \log n \log \log n$ edges contains a rainbow cycle.*

3 Proof of main 3-LCC result

Let C be an $[n, k]$ binary linear $(3, \delta)$-LCC. Throughout this section, fix a generator matrix $M \in \mathbb{F}_2^{n \times k}$ for C with row vectors $v_1, \ldots, v_n \in \mathbb{F}_2^k$ and associated 3-uniform matchings $\mathcal{H}_1, \ldots, \mathcal{H}_n$ over $[n]$. Our main result for this section is the following theorem, which is just Theorem 2 restated.

Theorem 5. *For any vector $x \in \mathbb{F}_2^k$, there exists a set of indices $I \subseteq [n]$ satisfying $x = \sum_{i \in I} v_i$ and $|I| \leqslant O(\delta^{-2} \log n \log \log n)$.*

Indeed, from Theorem 5, our main result Theorem 1 immediately follows.

Proof of Theorem 1 from Theorem 5. By Theorem 5, for each $x \in \mathbb{F}_2^k$, we know of a set $I_x \subseteq [n]$ of size at most $O(\delta^{-2} \log n \log \log n)$ satisfying $x = \sum_{i \in I_x} v_i$. Now, for distinct $x, y \in \mathbb{F}_2^k$, it follows from the definition of I_x that $I_x \neq I_y$. Since $|I_x| \leqslant O(\delta^{-2} \log n \log \log n)$, then there are at most $n^{O(\delta^{-2} \log n \log \log n)}$ possibilities for any I_x. Thus $2^k \leqslant n^{O(\delta^{-2} \log n \log \log n)}$, from which we conclude that $k \leqslant O(\delta^{-2} \log^2 n \log \log n)$. □

It therefore suffices to establish Theorem 5. For that, we will rely on the following key lemma.

[6] Note that G does not necessarily have to be simple. That is, edges are allowed to repeat.

Lemma 6. *Let c_0 be the absolute constant from Theorem 4. For any set $T \subseteq [n]$ of size at least $2c_0\delta^{-1}\log n \log\log n$, let $W \subseteq [n]$ be the set of indices $j \in [n]$ such that there exists a multi-set T' of indices in $[n]$ with $j \in T'$ satisfying $|T'| \leqslant |T|$ and*

$$\sum_{t\in T} v_t = \sum_{t\in T'} v_t \ .$$

Then $|W| \geqslant (\delta/2)n$.

Indeed, assuming Lemma 6, Theorem 5 follows as argued below.

Proof of Theorem 5 from Lemma 6. Let $I \subseteq [n]$ be a set of minimal cardinality satisfying $x = \sum_{i\in I} v_i$. Such a set exists as the vectors $v_1, \ldots, v_n$ span $\mathbb{F}_2^k$ (as M is full rank). Assume (for the sake of a contradiction) that $|I| \geqslant 10c_0\delta^{-2}\log n \log\log n$. Randomly partition I into $p \coloneqq \lceil 4/\delta \rceil$ sets $T_1 \ldots, T_p$ of equal size. Then $|T_\ell| \geqslant 2c_0\delta^{-1}\log n \log\log n$ for all $\ell \in [p]$. Thus we can apply Lemma 6 to find sets $W_1, \ldots, W_p$ of size at least $(\delta/2)n$ each satisfying the property stated in Lemma 6. Observe that $\sum_{\ell=1}^{p} |W_\ell| \geqslant (4/\delta)\cdot(\delta/2)n = 2n > n$. Thus, we can find distinct $\ell_1, \ell_2 \in [p]$ such that there is an index $j \in W_{\ell_1} \cap W_{\ell_2}$. Without loss of generality, say $(\ell_1, \ell_2) = (1, 2)$. Then by Lemma 6, we can find multi-sets $T_1', T_2' \subseteq [n]$ with $j \in T_1' \cap T_2'$ satisfying $|T_1'| \leqslant |T_1|$, $|T_2'| \leqslant |T_2|$, and

$$\sum_{i\in T_1} v_i = \sum_{i\in T_1'} v_i \ , \quad \text{as well as} \quad \sum_{i\in T_2} v_i = \sum_{i\in T_2'} v_i \ . \tag{1}$$

Now, define the multi-set $I' \coloneqq (T_1' \setminus \{j\}) \cup (T_2' \setminus \{j\}) \cup \cup_{\ell=3}^{p} T_\ell$. From (1), we find that

$$\begin{aligned} x = \sum_{i\in I} v_i &= \sum_{i\in T_1} v_i + \sum_{i\in T_2} v_i + \sum_{\ell=3}^{p}\sum_{i\in T_\ell} v_i \\ &= \sum_{i\in T_1'} v_i + \sum_{i\in T_2'} v_i + \sum_{\ell=3}^{p}\sum_{i\in T_\ell} v_i \\ &= \Big(v_j + \sum_{i\in T_1'\setminus\{j\}} v_i\Big) + \Big(v_j + \sum_{i\in T_2'\setminus\{j\}} v_i\Big) + \sum_{\ell=3}^{p}\sum_{i\in T_\ell} v_i \\ &= \sum_{i\in I'} v_i \ . \end{aligned}$$

Thus $x = \sum_{i\in I'} v_i$. On the other hand, since $|T_1'| \leqslant |T_1|$ and $|T_2'| \leqslant |T_2|$, then we find that

$$|I'| = |T_1' \setminus \{j\}| + |T_2' \setminus \{j\}| + \sum_{\ell=3}^{p} |T_\ell| \leqslant (|T_1| - 1) + (|T_2| - 1) + \sum_{\ell=3}^{p} |T_\ell| = |I| - 2 \ .$$

This contradicts the minimality of I, which is what we wanted to show. □

We now turn to the proof of Lemma 6. For this part, we introduce some notations. For any hyperedge $E \in \cup_{i=1}^{k} \mathcal{H}_i$, write $E = \{a_E, b_E, c_E\}$ for $a_E, b_E, c_E \in [n]$, and let $e_E \coloneqq \{b_E, c_E\}$.

Proof of Lemma 6. Assume (for the sake of a contradiction) that $|W| < (\delta/2)n$. Consider the graph G consisting of $[n]$ as vertices, T as edge colors, and for each $t \in T$, the set $\{e_E : E \in \mathcal{H}_t, a_E \notin W\}$ as the edges of the color class t. Because $\{\mathcal{H}_t\}_{t\in T}$ are 3-uniform matchings, any color class of edges in G will form a matching of edges, meaning that G is properly edge-colored. Furthermore, because $\{\mathcal{H}_t\}_{t\in T}$ are each of size at least δn, each color class has at least $|\mathcal{H}_t| - |W| > \delta n - (\delta/2)n = (\delta/2)n$ edges. Thus G has at least $(\delta/2)n \cdot |T| \geqslant c_0 n \log n \log\log n$ edges.

By Theorem 4, there exists a positive integer $m \geqslant 2$, a subset $U := \{t_1, \ldots, t_m\} \subseteq T$, and hyperedges $E_s \in \mathcal{H}_{t_s}$ for $s \in [m]$ such that the edges $(e_{E_1}, \ldots, e_{E_m})$ form a rainbow cycle in G. This implies that $\oplus_{s=1}^m e_{E_s} = \varnothing$. Now, define the set $T_0 := T \setminus U$. Then we have that

$$\begin{aligned}
\sum_{t\in T} v_t = \sum_{s=1}^{m} v_{t_s} + \sum_{t\in T_0} v_t &= \sum_{s=1}^{m}\left(v_{a_{E_s}} + v_{b_{E_s}} + v_{c_{E_s}}\right) + \sum_{t\in T_0} v_t \\
&= \sum_{s=1}^{m}\left(v_{b_{E_s}} + v_{c_{E_s}}\right) + \sum_{s=1}^{m} v_{a_{E_s}} + \sum_{t\in T_0} v_t \\
&= \sum_{j\in\bigoplus_{s=1}^{m} e_{E_s}} v_j + \sum_{s=1}^{m} v_{a_{E_s}} + \sum_{t\in T_0} v_t \\
&= \sum_{s=1}^{m} v_{a_{E_s}} + \sum_{t\in T_0} v_t \ .
\end{aligned}$$

Thus if we define the multi-set $T' := T_0 \cup \{a_{E_1}, \ldots, a_{E_m}\}$, then we see that $|T'| = |T|$ and $\sum_{t\in T} v_t = \sum_{t\in T'} v_t$. However, since e_{E_s} is an edge in G for each $s \in [m]$, then from the definition of G, we see that $a_{E_s} \notin W$ for all $s \in [m]$. This yields a contradiction by the definitions of W and T'. □

4 Rainbow LDC bounds and higher query LCCs

As we saw in Section 3, our proof of Theorem 1 via Theorem 2 differs fundamentally from recently developed methods [AGKM23, KM23a, HKM+24, Yan24]. However, while our proof method (inspired by [IS20]) produces a nearly tight lower bound for binary linear 3-LCCs, at its current state, it does not improve or even replicate most known lower bounds for r-LCCs/r-LDCs. In this section, we suggest a possible route towards bridging some parts of this gap.

One salient follow-up question to our work is whether our proof of Theorem 2 can be extended to higher query complexities. The proof of Theorem 2 crucially on the results of [ABS+23] (Theorem 4) regarding the existence of rainbow cycles in properly edge-colored graphs, which was only feasible due to the 3-uniformity of the query sets. For higher query complexities, we remedy this obstacle by introducing a hypergraph generalization of Theorem 4, stated below.

Question 1 (Rainbow LDC Lower Bound)**.** *For $\delta > 0$ and $r, n \in \mathbb{N}$ with $r \geqslant 2$, let $k_{\min}^{\text{rainbow}}(r, \delta, n)$ be the smallest possible natural number such that the following holds: for any arbitrary r-matchings $\mathcal{H}_1, \ldots, \mathcal{H}_k$ over $[n]$ with $k \geqslant k_{\min}^{\text{rainbow}}(r, \delta, n)$ satisfying $|\mathcal{H}_i| \geqslant \delta n$ for all $i \in [k]$, there exists a nonempty collection of hyperedges $\mathcal{E} \subseteq \cup_{i=1}^k \mathcal{H}_i$ such that $\bigoplus_{E\in\mathcal{E}} E = \varnothing$ and $|\mathcal{E} \cap \mathcal{H}_j| \leqslant 1$ for all $j \in [k]$. In terms of r, δ, n, what is the smallest upper bound on $k_{\min}^{\text{rainbow}}(r, \delta, n)$?*

Note that in the above notation, we have that $k_{\min}^{\text{rainbow}}(2, \delta, n) \leqslant O(\delta^{-1} \log n \log\log n)$ by Theorem 4. To our knowledge, we are not aware of any non-trivial upper bounds for when $r > 2$.

We dub Question 1 as the *rainbow LDC lower bound* problem. Our specific choice of naming comes from the fact that upper bounds on $k_{\min}^{\text{rainbow}}(r, \delta, n)$ formally prove lower bounds for binary linear r-LDCs. This can be seen from the viewpoint of LDC lower bounds as finding "odd even covers," formally shown in [HKM+24].

Proposition 7. *[HKM+24, Lemma 2.7] For $\delta > 0$ and $r, n \in \mathbb{N}$ with $r \geqslant 2$, let $k_{\min}^{\text{odd}}(r, \delta, n) \in \mathbb{N}$ be the smallest possible natural number such that the following holds: for any arbitrary r-matchings $\mathcal{H}_1, \dots, \mathcal{H}_k$ over $[n]$ with $k \geqslant k_{\min}^{\text{odd}}(r, \delta, n)$ satisfying $|\mathcal{H}_i| \geqslant \delta n$ for all $i \in [k]$, there exists a nonempty collection of hyperedges $\mathcal{E} \subseteq \cup_{i=1}^{k} \mathcal{H}_i$ such that $\bigoplus_{E \in \mathcal{E}} E = \varnothing$ and $|\mathcal{E} \cap \mathcal{H}_j|$ is odd for some $j \in [k]$. Then any binary linear (r, δ)-LDC[7] of block length n has dimension less than $k_{\min}^{\text{odd}}(r, \delta, n)$.*

From the definitions, it follows immediately that $k_{\min}^{\text{odd}}(r, \delta, n) \leqslant k_{\min}^{\text{rainbow}}(r, \delta, n)$. Moreover, using known constructions of r-LDCs [Yek08, Efr12] and Proposition 7, we find that $k_{\min}^{\text{rainbow}}(r, \delta, n) \geqslant k_{\min}^{\text{odd}}(r, \delta, n) \geqslant \exp(\Omega_\delta((\log \log n)^2))$ for $r \geqslant 3$, yielding us non-trivial lower bounds on $k_{\min}^{\text{rainbow}}(r, \delta, n)$ for $r \geqslant 3$. Now, with Question 1 at hand, we can state our generalization of Theorem 1.

Theorem 8. *For $\delta > 0$ and $r, n \in \mathbb{N}$ with $r \geqslant 3$, let $k_{\min}^{\text{rainbow}}(r-1, \delta, n) \in \mathbb{N}$ be as defined in Question 1. Then for any $[n, k]$ binary linear (r, δ)-LCC, we have*

$$k \leqslant O(\delta^{-1} \cdot \log n \cdot k_{\min}^{\text{rainbow}}(r-1, \delta/2, n)) \ .$$

The proof of Theorem 8 follows almost identically the proof of Theorem 1 in Section 3, except that properly edge-colored graphs become properly edge-colored $(r-1)$-uniform hypergraphs.[8] To reduce redundancy, we leave the proof of Theorem 8 as an exercise for the reader.

As for what the value of $k_{\min}^{\text{rainbow}}(r, \delta, n)$ should be, we conjecture that it should asymptotically be the same as $k_{\min}^{\text{odd}}(r, \delta, n)$, despite the much weaker requirement on the even cover in Proposition 7 than in Question 1.

Conjecture 9. *For fixed $\delta > 0$ and $r \in \mathbb{N}$ with $r \geqslant 2$, we have $k_{\min}^{\text{rainbow}}(r, \delta, n) = \Theta_{\delta,r}(k_{\min}^{\text{odd}}(r, \delta, n))$.*

Our reasoning for Conjecture 9 is twofold. First, for the $r = 2$ case, we know that $k_{\min}^{\text{odd}}(r, \delta, n) = \Theta_\delta(\log n)$ [GKST06] and $k_{\min}^{\text{rainbow}}(2, \delta, n) \leqslant O_\delta(\log n \log \log n)$ [ABS+23]. Thus our current state of knowledge already tells us that $k_{\min}^{\text{rainbow}}(2, \delta, n) = \widetilde{\Theta}_\delta(k_{\min}^{\text{odd}}(2, \delta, n))$. Thus, it does not seem too far-fetched to hope for the removal of the logarithmic factor. Furthermore, by Theorem 8, showing that $k_{\min}^{\text{rainbow}}(2, \delta, n) = \Theta_\delta(\log n)$ would immediately imply the optimality of the degree two Reed-Muller code as a binary linear 3-LCC.

Second, for the $r > 2$ case, a heuristic calculation in the second point of Section 9 of [KM23a] suggests that $[n, k]$ r-LCCs should satisfy the bound $k \leqslant \widetilde{O}(n^{1-2/(r-1)})$. Up to polylog factors, this is the same expected bound as what one would get from doing a similar heuristic calculation for $(r-1)$-LDCs using the methods of [AGKM23]. A positive answer to Conjecture 9 along with Theorem 8 would therefore formally prove the heuristic calculation of [KM23a], assuming that the analogous heuristic calculation of [AGKM23] held true for r-LDCs for all $r \geqslant 3$.[9]

Other follow-up questions. There are the obvious challenges of removing the linearity assumption and the binary alphabet assumption. Our crucial reliance on considering the rows of a generator matrix of the 3-LCC in the proof of Theorem 2 makes it unclear how to remove the linearity assumption in our proof. As for extending our proof to larger finite fields, it is easy to extend Theorem 1 to finite fields of characteristic 2 for a poly($|\mathbb{F}|$) loss in the upper bound on k by considering the code defined in Appendix A of [KM23a]. For finite fields of higher characteristic, the presence of negative signs presents a tricky situation for the application of the result of [ABS+23] (Theorem 4) in the proof of Theorem 2. We leave it as an interesting open problem to extend Theorem 2 to linear 3-LCCs over arbitrary finite fields.

[7] See Definition 4 for a formal definition of a linear (r, δ)-LDC.

[8] We say that an edge coloring of a hypergraph $\mathcal{H}$ is *proper* if for any distinct hyperedges $e_1, e_2 \in \mathcal{H}$ satisfying $e_1 \cap e_2 \neq \varnothing$, e_1 and e_2 are assigned different colors.

[9] Note that this heuristic calculation for r-LDCs is formally true for even $r \geqslant 4$ [KW04] and $r = 3$ [AGKM23].

References

[ABS+23] Noga Alon, Matija Bucić, Lisa Sauermann, Dmitrii Zakharov, and Or Zamir. Essentially tight bounds for rainbow cycles in proper edge-colourings. *arXiv preprint arXiv:2309.04460*, 2023. 1, 3, 4, 5, 7, 9, 10

[AGKM23] Omar Alrabiah, Venkatesan Guruswami, Pravesh K Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom CSP refutation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1438–1448, 2023. Full version appears at arXiv:2308.15403. 2, 3, 9, 10

[AS21] Vahid R Asadi and Igor Shinkar. Relaxed locally correctable codes with improved parameters. In *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. 2

[BBC+23] Alexander R Block, Jeremiah Blocki, Kuan Cheng, Elena Grigorescu, Xin Li, Yu Zheng, and Minshen Zhu. On relaxed locally decodable codes for hamming and insertion-deletion errors. In *Proceedings of the conference on Proceedings of the 38th Computational Complexity Conference*, pages 1–25, 2023. 2

[BBG+20] Alexander R Block, Jeremiah Blocki, Elena Grigorescu, Shubhang Kulkarni, and Minshen Zhu. Locally decodable/correctable codes for insertions and deletions. In *40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2020)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2020. 2

[BCG+22] Jeremiah Blocki, Kuan Cheng, Elena Grigorescu, Xin Li, Yu Zheng, and Minshen Zhu. Exponential lower bounds for locally decodable and correctable codes for insertions and deletions. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 739–750. IEEE, 2022. 2

[BGGZ21] Jeremiah Blocki, Venkata Gandikota, Elena Grigorescu, and Samson Zhou. Relaxed locally correctable codes in computationally bounded channels. *IEEE Transactions on Information Theory*, 67(7):4338–4360, 2021. 2

[CY22] Gil Cohen and Tal Yankovitz. Relaxed locally decodable and correctable codes: Beyond tensoring. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 24–35. IEEE, 2022. 2

[CY23] Gil Cohen and Tal Yankovitz. Asymptotically-good RLCCs with $(\log n)^{2+o(1)}$ queries. *Electron. Colloquium Comput. Complex.*, TR23-110, 2023. 2

[DGY11] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011. 2

[DLS13] Shagnik Das, Choongbum Lee, and Benny Sudakov. Rainbow Turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013. 3

[DSW14] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Improved rank bounds for design matrices and a new proof of kelly's theorem. In *Forum of Mathematics, Sigma*, volume 2. Cambridge University Press, 2014. 6

[Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM J. Comput.*, 41(6):1694–1703, 2012. 2, 10

[GKO+18] Sivakanth Gopi, Swastik Kopparty, Rafael Oliveira, Noga Ron-Zewi, and Shubhangi Saraf. Locally testable and locally correctable codes approaching the gilbert-varshamov bound. *IEEE Transactions on Information Theory*, 64(8):5813–5831, 2018. 2

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the Innovations in Theoretical Computer Science Conference*, pages 529–540, 2013. 2

[GKST06] Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Computational Complexity*, 15(3):263–296, 2006. 2, 3, 10, 13

[GL19] Tom Gur and Oded Lachish. A lower bound for relaxed locally decodable codes. *arXiv preprint arXiv:1904.08112*, 2019. 2

[Gop18] Sivakanth Gopi. *Locality in coding theory.* PhD thesis, Princeton University, 2018. 2

[GRR20] Tom Gur, Govind Ramnarayan, and Ron Rothblum. Relaxed locally correctable codes. *Theory of Computing*, 16(1):1–68, 2020. 2

[Gup23] Meghal Gupta. Constant query local decoding against deletions is impossible. *arXiv preprint arXiv:2311.08399*, 2023. 2

[HKM+24] Jun-Ting Hsieh, Pravesh K Kothari, Sidhanth Mohanty, David Munhá Correia, and Benny Sudakov. Small even covers, locally decodable codes and restricted subgraphs of edge-colored Kikuchi graphs. *arXiv preprint arXiv:2401.11590*, 2024. 2, 3, 9, 10

[HOW15] Brett Hemenway, Rafail Ostrovsky, and Mary Wootters. Local correctability of expander codes. *Information and Computation*, 243:178–190, 2015. 2

[IS20] Eran Iceland and Alex Samorodnitsky. On coset leader graphs of structured linear codes. *Discrete and Computational Geometry*, 63:560–576, 2020. 1, 3, 9, 13

[Jan23] Oliver Janzer. Rainbow Turán number of even cycles, repeated patterns and blow-ups of cycles. *Israel Journal of Mathematics*, 253(2):813–840, 2023. 3

[JS22] Oliver Janzer and Benny Sudakov. On the Turán number of the hypercube. *arXiv preprint arXiv:2211.02015*, 2022. 3

[KLLT22] Jaehoon Kim, Joonkyung Lee, Hong Liu, and Tuan Tran. Rainbow cycles in properly edge-colored graphs. *arXiv preprint arXiv:2211.03291*, 2022. 3

[KM23a] Pravesh K Kothari and Peter Manohar. An exponential lower bound for linear 3-query locally correctable codes. *arXiv preprint arXiv:2311.00558*, 2023. 1, 2, 3, 5, 9, 10

[KM23b] Vinayak M Kumar and Geoffrey Mon. Relaxed local correctability from local testing. *arXiv preprint arXiv:2306.17035*, 2023. 2

[KM24] Pravesh Kothari and Peter Manohar. Superpolynomial lower bounds for smooth 3-LCCs and sharp bounds for designs. Preprint, 2024. 2

[KMRZS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *Journal of the ACM (JACM)*, 64(2):1–42, 2017. 2

[KMSV07] Peter Keevash, Dhruv Mubayi, Benny Sudakov, and Jacques Verstraëte. Rainbow Turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007. 3

[KSY14] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. High-rate codes with sublinear-time decoding. *Journal of the ACM (JACM)*, 61(5):1–20, 2014. 2

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 80–86, 2000. 2, 6

[KW04] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. 2, 3, 10, 13

[OPC15] Rafail Ostrovsky and Anat Paskin-Cherniavsky. Locally decodable codes for edit distance. In *Information Theoretic Security: 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings 8*, pages 236–249. Springer, 2015. 2

[Rag07] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. *Electron. Colloquium Comput. Complex.*, TR07-016, 2007. 2

[Tom22] István Tomon. Robust (rainbow) subdivisions and simplicial cycles. *arXiv preprint arXiv:2201.12309*, 2022. 3

[Tre04] Luca Trevisan. Some applications of coding theory in computational complexity. *arXiv preprint cs/0409044*, 2004. 2

[Woo07] David Woodruff. New lower bounds for general locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007. 2

[Woo12] David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *J. Comput. Sci. Technol.*, 27(4):678–686, 2012. 2

[Yan24] Tal Yankovitz. A stronger bound for linear 3-LCC. *Electronic Colloquium on Computational Complexity Report*, ECCC TR122-101, 2024. 1, 2, 3, 5, 9

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *J. ACM*, 55(1):1:1–1:16, 2008. 2, 10

[Yek12] Sergey Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012. 2, 6, 14

A A New Proof of the Exponential Linear 2-LDC Lower Bound

In this appendix, we present a new proof of the well-known exponential lower bound for linear 2-query locally decodable codes [KW04, GKST06] à la [IS20]. We begin by stating the definition of a linear 2-query LDC for general finite fields $\mathbb{F}_q$. Note that this is usually referred to as a linear

2-LDC in normal form, but by known reductions [Yek12], the existence of a linear 2-LDC implies the existence of a linear 2-LDC in normal form. In what follows, the vectors $e_1, \dots, e_k \in \mathbb{F}_q^k$ denote the standard basis.

Definition 4 (Linear LDC)**.** *Given a generator matrix $M \in \mathbb{F}_q^{n\times k}$, let v_i the i'th row of M for $i \in [n]$. For $r \in \mathbb{N}$ and $\delta > 0$, we say that M forms a (r, δ)-locally decodable code (abbreviated (r, δ)-LDC) if there exist r-uniform matchings $\mathcal{H}_1, \dots, \mathcal{H}_k$ over $[n]$ such that $|\mathcal{H}_i| \geqslant \delta n$ for all $i \in [k]$, and for any $i \in [k]$ and $E = \{a_1, \dots, a_r\} \in \mathcal{H}_i$, there exist $\alpha_s^E \in \mathbb{F}_q \setminus \{0\}$ for $s \in [r]$ satisfying $e_i = \sum_{s=1}^r \alpha_s^E v_{a_s}$.*

Remark 10. *While the LCC property (Definition 2) is a property of the code, the LDC property is a property of the generator matrix of the code and not an inherent property of the code. That is, a different choice of generator matrix for the same code would not necessarily fulfill Definition 4.*

We now state the key result driving this section, which is the following *weight contraction* lemma.

Lemma 11. *For any $x \in \mathbb{F}_q^k$, there exist $a_1, a_2 \in [n]$ and $\gamma_1, \gamma_2 \in \mathbb{F}_q \setminus \{0\}$ satisfying*

$$wt(x + \gamma_1 v_{a_1} + \gamma_2 v_{a_2}) \leqslant (1 - 2\delta/q)wt(x) \ .$$

Proof. The proof proceeds via a "path coupling" style argument on the Cayley graph $\mathrm{Cay}(\mathbb{F}_q^k, \{\alpha v_i : \alpha \in \mathbb{F}_q \setminus \{0\}, i \in [n]\})$ but with the Hamming distance acting as the contracted distance. For $q = 2$, if we have $y, z \in \mathbb{F}_2^k$ with $y + z = e_i$, then by evolving (y, z) to $(y + v_a, y + v_b)$ where $a \in [n]$ is uniform, and b is a's matched vertex in $\mathcal{H}_i$ if it exists and $b = a$ otherwise, we can reduce the Hamming distance between y and z with probability $\Omega(\delta)$. For arbitrary $y, z \in \mathbb{F}_q^k$, we consider their shortest path in $\mathrm{Cay}(\mathbb{F}_q^k, \{\alpha e_i : \alpha \in \mathbb{F}_q \setminus \{0\}, i \in [k]\})$ and couple the vertices of each pair of edges along that path accordingly. We now proceed with the formal argument for general q below.

Let $S := \mathrm{supp}(x)$ and $w := |S|$. Write $S = \{i_1, \dots, i_w\}$ and $x = \beta_1 e_{i_1} + \ldots + \beta_w e_{i_w}$ for $\beta_t \in \mathbb{F}_q \setminus \{0\}$. Consider a uniformly randomly and independently chosen $\boldsymbol{\gamma}_0 \in \mathbb{F}_q \setminus \{0\}$ and $\mathbf{a}_0 \in [n]$. For each $t \in [w]$, define $\boldsymbol{\gamma}_t \in \mathbb{F}_q \setminus \{0\}$ and $\mathbf{a}_t \in [n]$ as

$$(\boldsymbol{\gamma}_t, \mathbf{a}_t) = \begin{cases} (-\boldsymbol{\gamma}_{t-1}(\alpha_{\mathbf{a}_{t-1}}^E)^{-1}\alpha_b^E, b) & \text{if } \exists\, b \in [n] \text{ such that } E := \{\mathbf{a}_{t-1}, b\} \in \mathcal{H}_{i_t}, \\ (\boldsymbol{\gamma}_{t-1}, \mathbf{a}_{t-1}) & \text{otherwise.} \end{cases}$$

Note that b is well-defined in the first case as $\mathcal{H}_1, \dots, \mathcal{H}_k$ are matchings. Furthermore, by a simple induction on t, it follows that $(\boldsymbol{\gamma}_t, \mathbf{a}_t)$ is uniformly random on $(\mathbb{F}_q \setminus \{0\}) \times [n]$ for all $t \in \{0, 1, \dots, w\}$. Now, for each $t \in [w]$, define

$$\boldsymbol{\beta}'_t = \begin{cases} \beta_t + \boldsymbol{\gamma}_{t-1}(\alpha_{\mathbf{a}_{t-1}}^E)^{-1} & \text{if } \exists\, b \in [n] \text{ such that } E := \{\mathbf{a}_{t-1}, b\} \in \mathcal{H}_{i_t}, \\ \beta_t & \text{otherwise.} \end{cases}$$

Then from the definitions, it follows that

$$\boldsymbol{\gamma}_{t-1} v_{\mathbf{a}_{t-1}} + \beta_t e_{i_t} = \boldsymbol{\beta}'_t e_{i_t} + \boldsymbol{\gamma}_t v_{\mathbf{a}_t} \tag{2}$$

for all $t \in [w]$. Thus by iteratively applying (2), we deduce that

$$\boldsymbol{\gamma}_0 v_{\mathbf{a}_0} + x = \boldsymbol{\gamma}_0 v_{\mathbf{a}_0} + \beta_1 e_{i_1} + \ldots + \beta_w e_{i_w} = \boldsymbol{\beta}'_1 e_{i_1} + \ldots + \boldsymbol{\beta}'_w e_{i_w} + \boldsymbol{\gamma}_w v_{\mathbf{a}_w} \ .$$

Thus we find that

$$x + \boldsymbol{\gamma}_0 v_{\mathbf{a}_0} - \boldsymbol{\gamma}_w v_{\mathbf{a}_w} = \boldsymbol{\beta}'_1 e_{i_1} + \ldots + \boldsymbol{\beta}'_w e_{i_w} \ . \tag{3}$$

Now, for each $t \in [w]$, let $\mathcal{E}_t$ denote the event that there exists $b \in [n]$ such that $\{\mathbf{a}_{t-1}, b\} \in \mathcal{H}_{i_t}$. Because $\mathbf{a}_{t-1}$ is uniformly random over $[n]$ and $\mathcal{H}_{i_t}$ is a matching of size at least δn, it therefore follows that $\mathbf{Pr}[\mathcal{E}_t] \geqslant 2\delta$. Furthermore, in the event that $\mathcal{E}_t$ occurs, $\boldsymbol{\beta}'_t$ will be uniformly random over $\mathbb{F}_q \setminus \{\beta_t\}$ as $\boldsymbol{\gamma}_{t-1}$ is uniformly random over $\mathbb{F}_q \setminus \{0\}$. This implies that $\mathbf{Pr}\left[\boldsymbol{\beta}'_t = 0 \mid \mathcal{E}_t\right] \geqslant 1/q$. Hence we find that $\mathbf{Pr}\left[\boldsymbol{\beta}'_t = 0\right] \geqslant \mathbf{Pr}\left[\boldsymbol{\beta}'_t = 0 \mid \mathcal{E}_t\right] \mathbf{Pr}[\mathcal{E}_t] \geqslant 2\delta/q$. Now, let $\mathbf{X}$ be the number of $\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_w$ that are equal to zero. By linearity of expectation, we find that $\mathbf{E}[\mathbf{X}] \geqslant (2\delta/q)w$. Thus, there exist $\gamma_0, \gamma_w \in \mathbb{F}_q \setminus \{0\}$ and $a_0, a_w \in [n]$ such that $\mathbf{X} \geqslant (2\delta/q)w$. From (3), we find that

$$\mathrm{wt}(x + \gamma_0 v_{a_0} - \gamma_w v_{a_w}) = \mathrm{wt}(\boldsymbol{\beta}'_1 e_{i_1} + \ldots + \boldsymbol{\beta}'_w e_{i_w}) \leqslant w - \mathbf{X} \leqslant (1 - 2\delta/q)w \ ,$$

which completes our proof. □

By iteratively applying the above lemma an appropriate number of times, one can immediately deduce the following.

Theorem 12. *Suppose that a generator matrix $M \in \mathbb{F}_q^{n \times k}$ with rows $v_1, v_2, \ldots, v_n \in \mathbb{F}_q^k$ forms a $(2,\delta)$-LDC. Then, for some absolute constant $c > 0$, the following holds for every $x \in \mathbb{F}_q^k$:*

- *There exists $I \subseteq [n]$ with $|I| \leqslant cq\delta^{-1} \log k$ such that x is in the $\mathbb{F}_q$-span of $\{v_i\}_{i \in I}$*
- *There exist $J \subseteq [n]$ with $|J| \leqslant cq\delta^{-1}$ and y in the $\mathbb{F}_q$-span of $\{v_j\}_{j \in J}$ such that the Hamming distance between x and y is at most $k/4$.*

The exponential lower bound for 2-LDC now follows by essentially a covering radius argument.

Theorem 13. *Let $M \in \mathbb{F}_q^{n \times k}$ be a generator matrix that forms a $(2,\delta)$-LDC. Then $k \leqslant O_{q,\delta}(\log n)$.*

Proof. Let $v_1, \ldots, v_n \in \mathbb{F}_q^k$ be the n rows of M, and c be the absolute constant from Theorem 12. Define $W \subseteq \mathbb{F}_q^k$ to be the set of vectors which are in the span of at most $cq\delta^{-1}$ vectors amongst the v_i's. Clearly

$$|W| \leqslant (qn)^{cq\delta^{-1}} \ . \tag{4}$$

Let $U \subseteq \mathbb{F}_q^k$ consist of all vectors within Hamming distance $k/4$ from some element of W. By Theorem 12, $U = \mathbb{F}_q^k$. On the other hand,

$$q^k = |U| \leqslant |W| \cdot q^{h_q(1/4)k} \ . \tag{5}$$

where $h_q(x) := x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$ is the q-ary entropy function. Combining (4) and (5), we conclude that $(1 - h_q(1/4))k \leqslant cq\delta^{-1} \log_q(qn)$ so that $k \leqslant O_{q,\delta}(\log n)$ as desired. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il