Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 36 (2024)

A stronger bound for linear 3-LCC

Tal Yankovitz*

Abstract

A q-locally correctable code (LCC) $C : \{0,1\}^k \to \{0,1\}^n$ is a code in which it is possible to correct every bit of a (not too) corrupted codeword by making at most q queries to the word. The cases in which q is constant are of special interest, and so are the cases that C is linear.

In a breakthrough result Kothari and Manohar (STOC 2024) showed that for linear 3-LCC $n = 2^{\Omega(k^{1/8})}$. In this work we prove that $n = 2^{\Omega(k^{1/4})}$. As Reed-Muller codes yield 3-LCC with $n = 2^{O(k^{1/2})}$, this brings us closer to closing the gap. Moreover, in the special case of design-LCC (into which Reed-Muller fall) the bound we get is $n = 2^{\Omega(k^{1/3})}$.

*Tel Aviv University. talyankovitz@mail.tau.ac.il. Supported by ERC starting grant 949499.

ISSN 1433-8092

Contents

1 Introduction **1**

1.1 Our result 1

1.1.1 The special case of designs 2

1.1.2 Larger alphabets 2

1.2 Proof overview 3

1.3 Comparison with [KM23] 6

1.4 Organization 6

2 Preliminaries **7**

3 Proof **7**

3.1 Decoding sequences 7

3.2 Extended decoding sequences 9

3.3 Decoding sequences and random sets 14

3.4 Deducing the bound 23

A The case of design LCC **26**

B Easy claims - proofs **30**

C The case of larger alphabets **31**

C.0.1 Preliminaries for this section 31

C.0.2 A bound for special-form LCC 32

C.0.3 From LCC to special-form LCC 35

D Proof sketch for Fact 2.3 and Fact C.4 **36**

1 Introduction

A q-locally correctable code (q-LCC) is a code in which every bit of the codeword - veiled by access to a noisy version of it - can be corrected by making at most q queries to the noisy word. A q-locally decodable code (q-LDC) is a code in which every bit of the message can be decoded by making at most q queries to the accessible word. More formally,

Definition 1.1. *An injective $C : \{0,1\}^k \to \{0,1\}^n$ is a (q, δ, ε)-LCC ((q, δ, ε)-LDC), for $\varepsilon < 1/2$, if there exists a randomized procedure that takes as input $j \in [n]$ (respectively, $i \in [k]$), gets oracle access to $z \in \{0,1\}^n$ at relative Hamming distance at most δ from $C(x)$ for some x, and in making at most q queries to z, and with probability at least $1-\varepsilon$: its output is equal to $C(x)_j$ (respectively, x_i). We say that C is linear if it is a linear map.*

LCCs in the regime in which q and δ are constant are of special interest, and the central question is how small n can be compared to k. Within this regime, in the case that $q = 2$ there are tight upper and lower bounds [GKST02, KdW04], showing that $n = 2^{\Theta(k)}$. For every larger $q \geq 3$, polynomial lower bounds are known [KT00, KdW04, Woo07] while the best upper bounds are exponential. In an exciting development, [KM23] proved a much stronger lower bound in the case that $q = 3$, showing that

Theorem 1.2 ([KM23])**.** *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear $(3, \delta, \varepsilon)$. Then $n = 2^{\Omega(\delta^2 k^{1/8})}$.*

Besides the strong bound their result also established a separation between 3-LCCs and 3-LDCs, as 3-LDCs with $n = 2^{k^{o(1)}}$ are known [Yek08, Efr09].

The methods [KM23] use in obtaining the bound are based on spectral refutations via Kikuchi matrices constructed from XOR formulas obtained by long chain derivations. The Kikuchi matrix method was also used in obtaining a better bound in the case of 3-LDC [AGKM23].

As noted by [KM23] the state of the art upper bound for $q = 3$ LCC is achieved by binary Reed-Muller codes, yielding

$$n = 2^{O(k^{1/2})},$$

while having a constant δ.

1.1 Our result

In this work we prove the following.

Theorem 1.3. *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear $(3, \delta, \varepsilon)$-LCC. Then $n = 2^{\Omega(\delta^{1/2} k^{1/4})}$.*

We remark that besides k the expression has a stronger dependence on δ. The proof only relies on elementary facts and the [GKST02] bound. While inspired by clever ideas of [KM23] the viewpoint of the proof is different and diverging from [KM23], our approach goes via constructing *asymmetric decoding sequences.*

1.1.1 The special case of designs

In the special case of *design* 3-LCCs we get a better bound. We define design LCC as follows.

Definition 1.4. *We say that a (q, δ, ε)-LCC is a* design LCC *if there exists a randomized procedure satisfying the requirements in Definition 1.1 and further: For every distinct pair of coordinates $a, c \in [n]$, there are at most $O(1)$ j's such that a and c can both be queried by invoking (once) the procedure to correct j.*

In other words, design LCCs are LCCs in which learning about any pair of coordinates sampled to be queried by the correction procedure almost reveals the identity of the coordinate being corrected (upto $O(1)$ options).[1]

Fact 1.5. *Reed-Muller codes are design LCCs.*[2]

Theorem 1.6. *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear $(3, \delta, \varepsilon)$-design LCC. Then $n = 2^{\Omega(\delta^{2/3} k^{1/3})}$.*

Thus in the state of affairs we get, in the case of 3-design LCC, the gap is smaller.

1.1.2 Larger alphabets

The definition of LCC can be naturally generalized to fields other than $\mathbb{F}_2$. [KM23] give a generalized version of their bound for the cases of fields other than $\mathbb{F}_2$. We do so as well, and get the following generalization.

Theorem 1.7. *For any field $\mathbb{F}$, let $C : \mathbb{F}^k \to \mathbb{F}^n$ be an $\mathbb{F}$-linear $(3, \delta, \varepsilon)$-LCC. Then $n = 2^{\Omega(\frac{\delta^{1/2} k^{1/4}}{|\mathbb{F}|^{1/2}})}$.*

The generalized LCC definition and the proof for Theorem 1.7 are in the appendix.

[1] For fields with characteristic 0 (and fields with very large characteristic) there are strong bounds [BDYW11, DSW14, DGOS18] on equivalents of linear design q-LCC, over these fields.

[2] In our definition: for constant q's.

1.2 Proof overview

We turn to give a high level overview of the elements of the proof. Assume that $C : \{0,1\}^k \to \{0,1\}^n$ is a linear $(3, \delta, \varepsilon)$-LCC.

Decoding sets. For a coordinate $j \in [n]$ and a set $Q \subseteq [n]$ we say that Q *determines* j if $C(x)_j = \sum_{j' \in Q} C(x)_{j'}$ $\forall x \in \{0,1\}^k$. It is well known that a linear (q, δ, ε)-LCC induces $m = \frac{\delta n}{q}$ sets $\{Q^j_w\}_{j \in [n],[w \in m]}$ such that for every j, $\{Q^j_w\}_{[w \in m]}$ are disjoint subsets of $[n]$ of size at most q, which determine j. We call the sets $\{Q^j_w\}_{j \in [n],[w \in m]}$ *decoding sets.*

Decoding sequences. We can use the decoding sets to generate many more *decoding sequences* of any length s. We assume without loss of generality that every query set is of size exactly 3[3], and we arbitrarily divide each query set Q^j_w into three designated parts $Q^j_w = \{A(j;w), B(j;w), c(j;w)\}$. Given $r \in [m]^s$, which is a "set of instructions", we construct two longer *decoding sequences* $A(j;r) \in [n]^s$, $B(j;r) \in [n]^s$, and $c(j;r) \in [n]$ to which we call a *reminder.* These are defined by $A(j;r) = A(j;r_1) \circ A(c(j;r_1);r_2,\ldots,r_s)$, $B(j;r) = B(j;r_1) \circ B(c(j;r_1);r_2,\ldots,r_s)$ and $c(j;r) = c(c(j;r_1);r_2,\ldots,r_s)$. It is easy to see that because of the promise that the query sets $\{A(j;w), B(j;w), c(j;w)\}_w$ determine j,

$$C(x)_{A(j;r)} + C(x)_{B(j;r)} + C(x)_{c(j;r)} = C(x)_j, \ \forall x \in \{0,1\}^k, \tag{1.1}$$

where for a sequence $D = (d_1, \ldots, d_s) \in [n]^s$, $C(x)_D := C(x)_{d_1} + \ldots + C(x)_{d_s}$. From here on we fix some s to be the length of the sequences. We say that the sequences are asymmetric because when using them A and c will be used as one part, and B will be the other.

Extended decoding sequences. For the argument to work we will need to handle the case of *repeated suffixes.* A suffix of a decoding sequence involves only the "A part" decoding sequence and the reminder c. For any $g \geq 1$, $\tilde{A} \in [n]^{s-g}$ and $c \in [n]$, the degree of the suffix $(\tilde{A}; c)$ is the number of j''s such that $A(j';r') = \tilde{A}$ and $c(j';r') = c$ for some $r' \in [m]^{s-g}$. It can be checked that if the query sets are such that every pair of distinct $a, c \in [n]$ is contained in at most one query set Q^j_w, then the degree of every $(\tilde{A}; c)$ is at most 1, however this may not be the case with the query sets of C. The aim in the proof is to argue that there are many "different" ways to deduce each $j \in [n]$, and while repeated suffixes may pose a problem, they can also be useful. If there is a specific decoding sequence which contains a suffix $(\tilde{A}; c)$ for $\tilde{A} \in [n]^{s-g}$ and $c \in [n]$ with a high degree, we can use this fact to generate on base of this decoding sequence - more decoding sequences - as

[3] As we can add zero coordinates and use them to extend the sets, at most doubling n.

many more as the degree, having each one of them posses a different reminder (this raises the probability that a random set of coordinates is useful for decoding j). More specifically, if we start with the decoding sequence $A = A(j;r)$, $B = B(j;r)$, $c = c(j;r)$, for $r \in [m]^s$, and $A = (A_1, \ldots, A_g, A_{g+1}, \ldots, A_s)$ where $\tilde{A} = (A_{g+1}, \ldots, A_s; c)$ is of high degree, then we can take any $j' \in [n]$ for which there exists $r' \in [m]^{s-g}$ such that $A(j';r') = (A_{g+1}, \ldots, A_s)$ and $c(j';r') = c$, and replace in $A(j;r)$ the part $(A_{g+1}, \ldots, A_s)$ with $B(j';r')$, and replace $c(j;r)$ with j'. That is, part of the A-part of the suffix turns into a B-part of a different decoding sequence, and the reminder is switched with the starting point of a sequence. Note that the length of the sequence is unchanged. We see that every choice among the different j''s gives us a different reminder. It can be checked that as the suffix was shared, we maintained the property stated in (1.1) - that is, that the obtained sequence still determines j. We remark, however, that there is some cost to doing this - if we do this for suffixes with too low degree, we may end up making the reminder more predictable rather than less, and so there is benefit in doing this only if the degree is above a certain threshold.

The asymmetric graph. For the high level explanation of this part we will ignore and rest aside the problem of repeated suffixes and so we will assume decoding sequences rather than extended decoding sequences. Let ℓ be a parameter. For every $j \in [n]$ we construct a bipartite graph G_j which on the left side has a vertex set $\binom{[n]}{\ell}^s \times [n]$ and on the right side has a vertex set $\binom{[n]}{\ell}^s$. The set of vertices will be the same for every j, but the edges will be j-dependent. Fix some j. The edges of G_j are colored with colors $r \in [m]^s$[4] and the total number of edges is the sum of the number of edges of each color. For a certain color $r \in [m]^s$, we put an edge of color r between $(L, h) = ((L_1, \ldots, L_s), h)$ of the left side and $L' = (L'_1, \ldots, L'_s)$ of the right side if $A(j;r)_1 \in L_1, \ldots, A(j;r)_s \in L_s$, $c(j;r) = h$, $B(j;r)_1 \notin L_1, \ldots, B(j;r)_s \notin L_s$[5], and

$$L' = ((L_1 \setminus \{A(j;r)_1\}) \cup \{B(j;r)_1\}, \ldots, (L_s \setminus \{A(j;r)_s\}) \cup \{B(j;r)_s\}).$$

Notice that such an edge $(((L_1, \ldots, L_s), h), (L'_1, \ldots, L'_s))$ will satisfy that

$$C(x)_{L_1} + \ldots + C(x)_{L_s} + C(x)_{L'_1} + \ldots + C(x)_{L'_s} + C(x)_h = C(x)_j \ \forall x \in \{0,1\}^k \qquad (1.2)$$

by Equation (1.1).

[4] In fact, if there are repeated suffixes, we also have colors that say which reminder we take when "switching" according to the repeated suffix. But we ignore this in describing the graph in this overview.

[5] We can assume that the color satisfies $A(j;r)_t \neq B(j;r)_t$ so the condition can be true, but in fact we do things a bit differently in the technical part and so this assumption will not be needed.

We would like to argue that there are many edges in G_j and that they are "different" in some way from one another. We cannot hope to show that there is a large matching, since the graph is asymmetric. However, as a first step, we will argue that there are indeed many edges, $\Theta\left(m^s \left(\frac{\ell}{n}\right)^s \binom{n}{\ell}^s\right)$, and that - in a good choice of ℓ and s - it holds that a constant fraction among them are edges that only touch two vertices whose degree is close to the average degree of their side. On the left side the average degree is $\Theta(m^s \frac{1}{n} \left(\frac{\ell}{n}\right)^s)$ and on the right side it is $\Theta(m^s \left(\frac{\ell}{n}\right)^s)$.[6] Arguing that there are many edges touching vertices with degree close to the average degree on the left part will require care, and is where the problem with repeated suffixes arises. The key to showing this is to argue that this holds separately within each color.

Finally, we construct another bipartite graph G'_j from G_j, for every $j \in [n]$, as follows. We maintain the left vertices of G'_j, and for the right side - we duplicate every vertex to have n copies - so that G'_j is balanced. As for the edges, we only consider edges of G_j whose both endpoints have a degree that is close to the average. For every vertex on the right side in G_j, we distribute evenly its edges - that we considered - among the n copies of the vertex in G'_j. In that way, a vertex which originally had degree close to $m^s \left(\frac{\ell}{n}\right)^s$ will now have degree close to $\frac{1}{n} m^s \left(\frac{\ell}{n}\right)^s$, like the vertices on the left side which already had degree close to $m^s \frac{1}{n} \left(\frac{\ell}{n}\right)^s$. Since we considered $\Theta\left(m^s \left(\frac{\ell}{n}\right)^s \binom{n}{\ell}^s\right)$ edges of G_j, and in defining their induced edges in G'_j - we used each edge once, then G'_j has $\Theta\left(m^s \left(\frac{\ell}{n}\right)^s \binom{n}{\ell}^s\right)$ edges, and maximal left and right degrees close to $m^s \frac{1}{n} \left(\frac{\ell}{n}\right)^s$, it follows that G'_j contains a large matching.

Bounding k. Recall that the vertices of G'_j are the same for every $j \in [n]$, on the left side each vertex is of the form $((L_1, \ldots, L_s), h)$, on the right side each vertex is $((L'_1, \ldots, L'_s), t)$ where t is the copy number, and that for every edge e in G'_j and every $C(x)$, the total sum of the coordinates of $C(x)$ corresponding to both endpoints of e is equal to $C(x)_j$ (note that t is just a copy number, not a coordinate, and it doesn't affect the summation). Hence, we will define a code $C' : \{0,1\}^k \to \{0,1\}^N$ "on" the vertices of $\{G'_j\}_j$, with $N = 2\binom{n}{\ell}^s n$, given by

$$C'(x)_{(L_1,\ldots,L_s,h)} = C(x)_{L_1} + \ldots + C(x)_{L_s} + C(x)_h$$

and

$$C'(x)_{(L'_1,\ldots,L'_s,t)} = C(x)_{L'_1} + \ldots + C(x)_{L'_s}.$$

[6] Notice that, in particular, it cannot be true that there are many such edges if $m^s \frac{1}{n} \left(\frac{\ell}{n}\right)^s \ll 1$, as in such a case every edge contradicts the requirement. Since $m \approx \delta n$, we see that $(\delta \ell)^s$ cannot be much smaller than n, and so $\max(\ell, s) \gtrapprox \log n$. Notice that this is a result of the $\frac{1}{n}$ factor - which is present because of the "imbalanced" nature of the graph - but when $s \geq \log n$ the "imbalance" becomes less meaningful, and smaller choices of ℓ become possible.

As C is injective we can assume without loss of generality that for every $i \in [k]$, $C(x)_i = x_i$[7]. Since there is a large matching in each G'_j for every $j \in [n]$, there is in particular a large matching in G'_i for every $i \in [k]$. And so, for every $i \in [k]$ there is a large number of disjoint pairs of coordinates of C' which determine i (and thus, x_i), by Equation (1.2), and it follows that C' is a 2-LDC. It is only left to apply the bound of [GKST02] to the get the result.[8]

1.3 Comparison with [KM23]

It is somewhat hard to compare exactly the proof of [KM23] and the proof we give here since the two proofs differ in viewpoints. The [KM23] Kikuchi method proof utilizes some tools that we do not use here (such representing via XOR instances, bounding the spectral norm, computing some partial derivatives, etc). The proof in the viewpoint we give here is combinatorical and only uses elementry facts. Yet there are similar points made in both arguments and our argument is inspired by the clever ideas of [KM23]. The long chain-derivation idea employed by [KM23] is of course similar to the decoding sequences we use. The issue with heavy pairs which arises in the [KM23] argument is like the issue with repeated suffixes here. The handling of [KM23] for the issue is by their constructing of contiguously regular partitions, and while there are differences in the handling, we employed a threshold check similar to the one incorporated there. We add that we do not partition $[k]$ into two sets as is done in [KM23] (in a part that could be interpreted as partitioning the *message* into two sets and zeroing one of the sets). The "asymmetric" part here is taking a different approach compared to [KM23].

1.4 Organization

In Section 3.1 we define Decoding sequences. In Section 3.2 we define Extended decoding sequences and prove needed claims regarding them. In Section 3.3 we define the asymmetric graph, and argue for the existence of a large matching in the final graph. In Section 3.4 we define a 2-LDC code on top of the constructed graph, and deduce the bound. In the appendix, Appendix A, we analyze the case of 3-design LCC. In Appendix C we prove the generalized bound for larger alphabets.

[7] It is a well known fact that a linear code can be made systematic.

[8] If the "large matching" is in fact of size $\Omega(N)$, the [GKST02] bound implies that k is logarithmic in the length of the code, N, and so we want $\log N \approx \ell \cdot s \cdot \log n$ to be small. One limitation on how small s and ℓ can be is discussed in a previous footnote. Moreover, in fact we will argue for the existence of a matching of a size that is slightly smaller than $\Omega(N)$.

2 Preliminaries

Notations. All logarithms in this paper are taken to the base 2. The set of natural numbers is $\mathbb{N} = \{0, 1, 2, \ldots\}$. For $n \in \mathbb{N}$, $n \geq 1$, we use $[n]$ to denote the set $\{1, \ldots, n\}$. We denote by $\binom{[n]}{\ell}$ the set of subsets of $[n]$ of size ℓ. We use $\mathbb{F}$ to denote a field, and $\mathbb{F}_q$ to specify that it is of size q. For a sequence $D \in \{0,1\}^s$ and a set $S = \{h_1, \ldots, h_t\} \subseteq [s]$ for $h_1 < \ldots < h_t$ we denote by D_S the sequence at locations $h_1, \ldots, h_t$. If $S = \emptyset$ then $D_S = \varepsilon$ is the empty sequence. We will also write $D_{h_1,\ldots,h_t}$ as short for $D_{\{h_1,\ldots,h_t\}}$. We use $\circ$ to denote the concatenation of two sequences. We use $\mathbb{I}_A$ to denote an indicator random variable, supported on $\{0,1\}$, for the event A.

We will need the following fact, that says that for a linear code there is a systematic encoding.

Fact 2.1. *For every linear injective $C : \mathbb{F}^k \to \mathbb{F}^n$ there is a linear $\tilde{C} : \mathbb{F}^k \to \mathbb{F}^n$ such that $Img(C) = Img(\tilde{C})$, and there are $j_1, \ldots, j_k \in [n]$ such that for every $i \in [k]$ $\tilde{C}(x)_{j_i} = x_i \forall\, x$.*

We will make use of the [GKST02] bound for 2-LDCs.

Theorem 2.2 ([GKST02])**.** *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear map such that for every $i \in [k]$ there is a set of $\delta_c n$ disjoint pairs of coordinates $\{u, v\} \subseteq [n]$ such that $x_i = C(x)_u + C(x)_v$ $\forall x$. Then $k = O(\frac{1}{\delta_c} \log n)$.*

We will also need the following well known fact regarding LCCs.[9]

Fact 2.3. *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear (q, δ, ε)-LCC. Then there exist sets $\{Q^j_w\}_{j\in[n],w\in[m]}$ for $m \geq \delta n/q$ such that for every $j \in [n]$ the sets $Q^j_w \mid_{w\in[m]}$ are disjoint subsets of $[n]$ of size at most q, each satisfying that $C(x)_j = \sum_{j' \in Q^j_w} C(x)_{j'}$ $\forall x \in \{0,1\}^k$.*

A proof sketch for Fact 2.3 is found in the appendix, for completeness.

3 Proof

3.1 Decoding sequences

We start by defining *decoding sequences*, which are composed of two sequences of length s, the *A-part*, the *B-part*, which are both sequences of n coordinates, and of a *reminder* c which is a single coordinate.

[9] See a similar statement by [KT00] for the case of LDCs.

First, we assume that we are given $n \in \mathbb{N}$ and $m \leq n$ and nm sets $Q_w^j \subseteq [n] \mid_{j \in [n], w \in [m]}$, such that each set is of size exactly 3.[10] We will assume that it's possible to order each set such that it is composed of three designated elements, $\{A(Q_w^j), B(Q_w^j), c(Q_w^j)\} = Q_w^j$, satisfying the following guarantee:

(*) For every $j \in [n]$, $w \neq w'$ and $D \in \{A, B, c\}$, $D(Q_w^j) \neq D(Q_{w'}^j)$.

Note that if for every j the sets $\{Q_w^j\}_{w \in [m]}$ are known to be disjoint, in particular the guarantee is satisfied by any arbitrary ordering.

Definition 3.1 (Decoding sequences). *For every $s \geq 0$ we will define three functions*

$$\begin{aligned} A &: [n] \times [m]^s \to [n]^s, \\ B &: [n] \times [m]^s \to [n]^s, \\ c &: [n] \times [m]^s \to [n]. \end{aligned}$$

The definition is inductive. For $s = 0$, $r \in [m]^s$ and $j \in [n]$

$$\begin{aligned} A(j; r) &= \varepsilon, \\ B(j; r) &= \varepsilon, \\ c(j; r) &= j, \end{aligned}$$

where ε is the empty sequence. For $s > 0$, $r \in [m]^s$ and $j \in [n]$

$$\begin{aligned} A(j; r) &= A(Q_{r_1}^j) \circ A(c(Q_{r_1}^j); r_{2,\dots,s}), \\ B(j; r) &= B(Q_{r_1}^j) \circ B(c(Q_{r_1}^j); r_{2,\dots,s}), \\ c(j; r) &= c(c(Q_{r_1}^j); r_{2,\dots,s}). \end{aligned}$$

We will require the following easy claim which considers how many sequences satisfy a set of *constraints*.

Claim 3.2. *For every $j \in [n]$, $S \subseteq [s]$ and $E \in [n]^s$,*

$$\begin{aligned} |\{r \in [m]^s \mid A(j; r)_S = E_S\}| &\leq m^{s - |S|}. \\ |\{r \in [m]^s \mid B(j; r)_S = E_S\}| &\leq m^{s - |S|}. \end{aligned}$$

In particular, for every $j \in [n]$, $A(j; \cdot)$ and $B(j; \cdot)$ are injective.

Proof. Follows directly from Definition 3.1 and the guarantee (*). □

[10] In this part (and the two following) we only make structural definitions and claims based on the given sets $\{Q_w^j\}$. But it will be good to notice that the definition of *decoding sequences* that we give next is tailored to be useful for decoding, assuming that $\{Q_w^j\}$ are.

3.2 Extended decoding sequences

We turn to define extended decoding sequences, in order to handle repeated suffixes. For what comes next, we fix some s and some $f \geq 1$. We will need the following definitions.

Definition 3.3. *Let $h \geq 0$, $A \in [n]^h$ and $c \in [n]$. We define*

$$J(A; c) = \{j \in [n] \mid \exists r \in [m]^h \; A(j; r) = A, c(j; r) = c\}$$

and

$$\deg(A; c) = |J(A; c)|.$$

That is, J defines the set of indices for which there is a sequence whose A-part and c are equal to a given suffix $(A; c)$, and the degree of $(A; c)$ is the number of these indices.

Definition 3.4. *Define for every $g \in [s]$*

$$R^g(j) = \left\{ r \in [m]^s \mid {\scriptstyle \deg(A(j;r)_{g+1,\ldots,s};c(j;r)) \geq f^{s-g} \wedge \atop \forall g' \in \{1,\ldots,g-1\} : \deg(A(j;r)_{g'+1,\ldots,s};c(j;r)) \leq f^{s-g'}} \right\}.$$

In words, $R^g(j)$ is the set of instructions which result in a sequence with a common suffix starting from position $g+1$, and for every location closer to the start of the sequence, the suffix is uncommon. By common we mean that the degree of the suffix crosses a threshold which depends on where the suffix starts: we check if the degree is at most f^{s-g}.

We argue that there is (at least) one R^g than contains many of the possible instructions. That is, there is an R^g which induces many sequences.

Claim 3.5. *For every $j \in [n]$ there exists some $g \in [s]$ such that $|R^g(j)| \geq \frac{1}{s} m^s$.*

Proof. For every $r \in [m]^s$, there is at least one $g = g(r) \in [s]$ such that $r \in R^g(j)$. This holds as for $g = 1$ the second condition

$$\forall g' \in \{1, \ldots, g-1\} = \emptyset : \deg(A(j; r;)_{g'+1,\ldots,s}, c(j; r)) \leq f^{s-g'}$$

is always (trivially) met. Hence, if $g = 1$ also satisfies the first condition

$$\deg(A(j; r;)_{g+1,\ldots,s}, c(j; r)) \geq f^{s-g}$$

we can take $g(r) = 1$. If otherwise, then $g = 2$ always satisfies the second condition, and if it also satisfies the first we can take $g(r) = 2$. And so on. Notice that for $g = s$ the first condition is always met, since

$$1 = \deg(\varepsilon; c(j; r)) = \deg(A(j; r)_{g+1,\ldots,s}; c(j; r)) \geq f^{s-g} = 1,$$

and therefore the described procedure must halt.

Since there are m^s r's and each one is a member of at least one of the s $R^g(j)$'s, the claim follows. □

When we will use extended decoding sequences (which we have not yet defined), we will do it with respect to one specific g (for each $j \in [n]$) that satisfies the above claim.

Before we turn to define extended decoding sequences we need to handle another matter as set up. In the above we defined $R^g(j)$ to correspond to sequences in which that g-th suffix is repeated at least f^{s-g} times (among suffix sequences of length $s-g$). We wish to reduce to the case that each such suffix is repeated exactly f^{s-g} times, and towards that we will require the following definitions. We will later cut a part of $J(A;c)$ to achieve this - but we start with assuming that we are given a subset of $J(A;c)$ which is of size that is a multiple of f^{s-g} (we will next denote this subset by$J^g(j;A;c)$).

A remark for first time reading. The next definition will make a few somewhat long notations. A first time reader may be advised to only skim through this definition (and the related claim Claim 3.7 that follows it) and to go back to it after reading Section 3.3. For the reader who opts to doing so it should be helpful to know that in the claim following the definition, Claim 3.7, we define a set of "good enough" instruction sets $\tilde{R}^g(j) \subseteq R^g(j) \subseteq [m]^s$ and argue that it is large enough - we will only consider sequences induced by instructions r from this set (rather than $[m]^s$).

Definition 3.6. *Let $j \in [n]$, $g \in [s]$, $A \in [n]^{s-g}$ and $c \in [n]$ be such that* $\deg(A;c) \geq f^{s-g}$*, and further let $J^g(j;A;c) \subseteq J(A;c)$ be a subset (chosen specifically for j and g) of size which is a multiple of f^{s-g}.*

- *We arbitrarily partition $J^g(j;A;c)$ into parts of size f^{s-g}, and we assume that each part has an arbitrary fixed order.*
- *For every $c' \in J^g(j;A;c)$, we define $P(j;c';A;c)$ to be the part of the partition to which c' belongs.*
- *Furthermore, for every $z \in [f^{s-g}]$ we denote by $P(j;c';A;c)_z$ the z-th element of the part $P(j;c';A;c)$.*
- *Lastly, we will denote by $T(j;c';A;c)_z$ the (unique) sequence $r \in [m]^{s-g}$ for which $A(P(j;c';A;c)_z;r) = A$(and $c(P(j;c';A;c)_z;r) = c$).*

Intuitively, splitting each $J^g(j;A;c)$ into parts $\{P(j;c';A;c)\}_{c'}$ of size exactly f^{s-g} will allow us to fall back to the case that $J(A;c)$ had been of size f^{s-g} to begin with. We will need to be able to address a specific index within each such part, and so we set the notation $P(j;c';A;c)_z$. As for the last defined notation $T(j;c';A;c)_z$, recall that the way $J(A;c)$ is defined is by taking all indices c' for which there is some instruction sequence

r that results in a specific common suffix $(A;c)$ - we want to be able to address this r by virtue of which c' is in $J(A;c)$, and in $P(j;c';A;c)$.

We are almost done with the setup - it is only left to explain how we cut $J(A;c)$ to be of a right size, as we have assumed, without losing too many of our sequences: We will assume that the subsets $J^g(j;A;c) \subseteq J(A;c)$ mentioned in the previous definition are such who satisfy the following claim.[11]

Claim 3.7. *For every $j \in [n]$ and $g \in [s]$, there exist subsets $J^g(j;A;c) \subseteq J(A;c)\ \forall A, c$, of sizes that are a multiple of f^{s-g}, such that the set*

$$\tilde{R}^g(j) := \{r \in R^g(j) \mid c(j;r_{1,\ldots,g}) \in J^g(j;A(j;r)_{g+1,\ldots,s};c(j;r))\}$$

is of size at least $\frac{1}{2}|R^g(j)|$ (Recall that by the definition of $J(A;c)$ it is known that $c(j;r_{1,\ldots,g}) \in J(A(j;r)_{g+1,\ldots,s};c(j;r))$).

Proof. Fix $j \in [n]$ and $g \in [s]$. For every A, c and every $c' \in [n]$ set $\text{count}_{A,c}(c') = |\{r \in R^g(j) \mid c(j;r_{1,\ldots,g}) = c' \wedge A(j;r)_{g+1,\ldots,s} = A \wedge c(j;r) = c\}|$. For every A, c we sort $J(A;c)$ in descending order of $\text{count}_{A,c}(c')$ (for $c' \in J(A;c)$) and take $J^g(j;A;c)$ to be the first $|J(A;c)| - (|J(A;c)| \mod f^{s-g})$ elements. Note that for every A, c such that $|J(A;c)| \geq f^{s-g}$, $|J^g(j;A;c)| \geq \frac{1}{2}|J(A;c)|$. As for every $r \in R^g(j)$, $|J(A(j;r)_{g+1,\ldots,s};c(j;r))| \geq f^{s-g}$, and as we sorted $J(A;c)$ from the most used (as $c(j;r_{1,\ldots,g})$ for $r \in R^g(j) \mid A(j;r)_{g+1,\ldots,s} = A \wedge c(j;r) = c$) to the least used, the claim follows. □

We can now finally define *extended decoding sequences*, which are dependent on which g is chosen to be used.

A remark for first time reading. A reader that only skimmed through Definition 3.6 (see previous remark) may also only skim through the definition of c^g and A^g below (which depend on Definition 3.6).

Definition 3.8 (Extended decoding sequences)**.** *We extend Definition 3.1 by defining three more functions for every $g \in [s]$*

$$\begin{aligned} A^g &: [n] \times \tilde{R}^g(j) \times [f^{s-g}] \to [n]^s, \\ B^g &: [n] \times \tilde{R}^g(j) \times [f^{s-g}] \to [n]^s, \\ c^g &: [n] \times \tilde{R}^g(j) \times [f^{s-g}] \to [n]. \end{aligned}$$

[11] The reason that we "lose" sequences in cutting $J(A;c)$ is that it will come with disallowing sequences that pass through the cut indices (dependent on the suffix $(A;c)$; see next definition).

as follows

$$\begin{aligned} c^g(j;r;z) &= P\big(j;c(j;r_{1,\dots,g});A(j;r)_{g+1,\dots,s};c(j;r)\big)_z, \\ A^g(j;r;z) &= A(j;r)_{1,\dots,g} \circ B\Big(c^g(j;r;z);T\big(j;c(j;r_{1,\dots,g});A(j;r)_{g+1,\dots,s};c(j;r)\big)_z\Big) \\ B^g(j;r) &= B(j;r). \end{aligned}$$

[12]

In our use of extended decoding seqeunces in the next section, we will need to be able to bound the number of sequences that are "close" to some sequence, and for that we have the following definitions and two claims. In the case of the A-part, we will specifically consider the case that the sequence is close to some other sequence, and that the reminders are equal.

Definition 3.9. *For $j \in [n]$, $g \in [s]$, $A \in [n]^s, B \in [n]^s$, $c \in [n]$ and $S \subseteq [s]$ define*

$$\begin{aligned} R^g(j,A,S,c) &= \{(r,z) \in \tilde{R}^g(j) \times [f^{s-g}] \mid A^g(j;r;z)_S = A_S \wedge c^g(j;r;z) = c\}, \\ R^g(j,B,S) &= \{r \in \tilde{R}^g(j) \mid B^g(j;r)_S = B_S\}. \end{aligned}$$

That is, given a set S of "constraints" we count either how many sequences there are that agree with a given A on S, and have specific reminder, or how many sequences there are whose B-part agrees with a given sequence on B on S (without a requirement on the reminder).

Claim 3.10.

$$|R^g(j,A,S,c)| \leq \begin{cases} m^{s-|S|} & \textit{if } [g] \subseteq S \\ m^{s-|S|-1} f^{s-\max([g]\setminus S)} & \textit{else.} \end{cases}$$

A remark for first time reading. A reader that only skimmed through Definition 3.6 and the definitions of c^g and A^g may also only skim through the proof for Claim 3.10. After reading Section 3.3 the motivation for Claim 3.10 should be clear - and so Definition 3.6, the definitions of c^g and A^g and the proof Claim 3.10 can be thoroughly read afterwards.

[12] In words, opening up the definitions of $P(\cdot)_z$ and $T(\cdot)_z$: c^g and A^g take as input an extra instruction $z \in [f^{s-g}]$, beside the instructions r. This instruction determines which member j' of the part $P\big(j;c(j;r_{1,\dots,g});A(j;r)_{g+1,\dots,s};c(j;r)\big)$ is taken to be the new reminder. The new A-part is achieved by maintaining the original A-part upto location g, and the locations $g+1,\dots,s$ are replaced with the B-part of the sequence starting from j', using the same instruction sequence r' for which $A(j';r') = A(j,r)_{g+1,\dots,s}$ and $c(j';r') = c(j;r)$. B^g remains the same as the original B and doesn't take an extra instruction z.

Proof for Claim 3.10. Consider $r \in \tilde{R}^g(j), z \in [f^{s-g}]$ such that $A^g(j;r;z)_S = A_S$ and $c^g(j;r;z) = c$. Note that it is enough to bound the number of options for r, since for every possible r, the constraint $c^g(j;r;z) = c$ determines z.

Note first that:

(**) There are at most $m^{s-g-|S\cap\{g+1,\dots,s\}|}$ options for $B\Big(c^g(j;r;z); T\big(j; c(j;r_{1,\dots,g}); A(j;r)_{g+1,\dots,s}; c(j;r)\big)_z\Big)$ and $A(j;r)_{g+1,\dots,s}$.

Indeed, this follows by Claim 3.2, as we have a B-part sequence of length $s-g$, subjected to $|S \cap \{g+1,\dots,s\}|$ constraints, and $c^g(j;r;z) = c$ is fixed. Moreover, for every such option for $B\Big(c^g(j;r;z); T\big(j; c(j;r_{1,\dots,g}); A(j;r)_{g+1,\dots,s}; c(j;r)\big)_z\Big)$ there is only one option for $A(j;r)_{g+1,\dots,s} = A\Big(c^g(j;r;z); T\big(j; c(j;r_{1,\dots,g}); A(j;r)_{g+1,\dots,s}; c(j;r)\big)_z\Big)$ and $c(j;r) = c\Big(c^g(j;r;z); T\big(j; c(j;r_{1,\dots,g}); A(j;r)_{g+1,\dots,s}; c(j;r)\big)_z\Big)$.

We proceed by analyzing the two cases of the argued inequality.

1. If $[g] \subseteq S$. Since $A(j;r)_{[g]}$ is fixed, every option for $A(j;r)_{g+1,\dots,s}$ determines $A(j;r)$, and by Claim 3.2, it also determines r. Hence in this case by (**)
$$|R^g(j,A,S,c)| \le m^{s-g-|S\cap\{g+1,\dots,s\}|} = m^{s-|S|}.$$

2. If $[g] \not\subseteq S$, set $g' = \max([g] \setminus S)$. First, note that again by Claim 3.2, there are at most $m^{g'-1-|S\cap\{1,\dots,g'-1\}|}$ options for $A(j;r)_{1,\dots,g'-1} = A^g(j;r;z)_{1,\dots,g'-1}$ as it is an A-part sequence of length $g'-1$ subjected to $|S\cap\{1,\dots,g'-1\}|$ constraints. Secondly, we argue that for every option for $A(j;r)_{g+1,\dots,s}$ and $c(j;r)$ (and recall that by (**) there are at most $m^{s-g-|S\cap\{g+1,\dots,s\}|}$ such) the number of options for $c(j;r_{1,\dots,g'})$ is at most $f^{s-g'}$.

   To see this, we first note that by the choice of g', $A(j;r)_{g'+1,\dots,g} = A^g(j;r;z)_{g'+1,\dots,g}$ is fixed as $\{g'+1,\dots,g\} \subseteq S$, and so every option for $A(j;r)_{g+1,\dots,s}$ determines $A(j;r)_{g'+1,\dots,s}$. Secondly we, again, consider two cases - for the location of g', and see that the number of options for $c(j;r_{1,\dots,g'})$ is at most $f^{s-g'}$ in both of them:

   (a) If $g' < g$. Since $\tilde{R}^g(j) \subseteq R^g(j)$, $\deg(A(j;r)_{g'+1,\dots,s}; c(j;r)) \le f^{s-g'}$, by the definition of $R^g(j)$. Furthermore, $c(j;r_{1,\dots,g'}) \in J(A(j;r)_{g'+1,\dots,s}; c(j;r))$ which is of size $\deg(A(j;r)_{g'+1,\dots,s}; c(j;r))$. Hence, there are at most $f^{s-g'}$ options for $c(j;r_{1,\dots,g'})$.

(b) If $g' = g$. Since $c(j; r_{1,\ldots,g}) \in P(j; c; A(j;r)_{g+1,\ldots,s}; c(j;r))$ (since it is a part of a partition to which both c and $c(j; r_{1,\ldots,g})$ belong), and $|P(j; c; A(j;r)_{g+1,\ldots,s}; c(j;r))| = f^{s-g}$, there are at most f^{s-g} options for $c(j; r_{1,\ldots,g})$.

Thus, we conclude that there are at most $m^{g'-1-|S\cap\{1,\ldots,g'-1\}|}$ options for $A(j;r)_{1,\ldots,g'-1}$, at most $m^{s-g'-|S\cap\{g'+1,\ldots,s\}|}$ options for $A(j;r)_{g'+1,\ldots,s}$ - and for every fixing of these options - at most $f^{s-g'}$ options for $c(j; r_{1,\ldots,g'})$. Notice that if $A(j;r)_{1,\ldots,g'-1}$ is given, then $c(j; r_{1,\ldots,g'-1})$ is known (as $r_{1,\ldots,g'-1}$ is known). If $c(j; r_{1,\ldots,g'-1})$ is known then for every option for $c(j; r_{1,\ldots,g'})$ there is one option for $A(j;r)_{g'}$ (since knowing $c(j; r_{1,\ldots,g'})$ further determines $r_{g'}$). Hence, there are at most $m^{g'-1-|S\cap\{1,\ldots,g'-1\}|}$ options for $A(j;r)_{1,\ldots,g'-1}$, at most $m^{s-g'-|S\cap\{g'+1,\ldots,s\}|}$ options for $A(j;r)_{g'+1,\ldots,s}$ - and for every fixing of these - at most $f^{s-g'}$ options for $A(j;r)_{g'}$. We conclude that in the case that $[g] \nsubseteq S$:

$$\begin{aligned} |R^g(j,A,S,c)| \leq\, & m^{g'-1-|S\cap\{1,\ldots,g'-1\}|} m^{s-g'-|S\cap\{g'+1,\ldots,s\}|} f^{s-g'} \\ =& m^{s-1-|S\cap\{1,\ldots,s\}|} f^{s-g'} \\ & m^{s-1-|S|} f^{s-g'}, \end{aligned}$$

where we used the fact that $g' \notin S$ by its definition.

We have thus shown the two cases of the inequality, and the claim follows. □

The following claim bounds the sequences that are close to the B-part.

Claim 3.11.

$$|R^g(j,B,S)| \leq m^{s-|S|}.$$

Proof. follows directly from Claim 3.2. □

3.3 Decoding sequences and random sets

In this part we will define a bipartite graph for every $j \in [n]$, whose edges will correspond to extended decoding sequences of j and so they can be used in correcting j. The aim is to show that each such graph contains a large matching.

Before we define these graphs, we require some set up. Again in this subsection we assume $n \in \mathbb{N}$ and $m \leq n$. We will also assume nm sets $\{Q_r^j\}_{j\in[n],r\in[m]}$ which are as in Section 3.1, and satisfy guarantee (*). Further, we fix s and ℓ to be some parameters to be chosen later, and we will assume that $\ell = o(n)$ and

$$s = O\left(\frac{n}{\ell}\right). \tag{3.1}$$

We set

$$f = e_0 \left(1 + \frac{m\ell}{n}\right) \tag{3.2}$$

for a small enough universal constant $e_0 < 1$. Moreover we will assume that s, ℓ satisfy

$$\left(e_0 \left(1 + \frac{m\ell}{n}\right)\right)^s > m. \tag{3.3}$$

For every $j \in [n]$ we choose $g(j) \in [s]$ to be one that satisfies Claim 3.5 with respect to s, f (and j). In a slight abuse of notation we will write g as short for $g(j)$ but it will always be in contexts where j is specific.

We now define a relation which we will use in defining the edges of the graphs, and a couple of notations.

Definition 3.12. *Let*

$$\begin{aligned}&A = (a_1, \dots, a_s) \in [n]^s, B = (b_1, \dots, b_s) \in [n]^s, c \in [n],\\ &L = (L_1, \dots, L_s, L_{s+1}) \in \binom{[n]}{\ell}^s \times [n], \quad L' = (L'_1, \dots, L'_s) \in \binom{[n]}{\ell}^s.\end{aligned}$$

We write $A \subseteq L$ if $a_1 \in L_1, \dots, a_s \in L_s$, $B \subseteq L'$ if $b_1 \in L'_1, \dots, b_s \in L'_s$; $c \subseteq L$ if $L_{s+1} = c$. Further, we write $L \sim_{A,B,c} L'$ if $A, c \subseteq L$ and

$$L'_1 = (L_1 \setminus \{a_1\}) \cup \{b_1\}, \dots, L'_s = (L_s \setminus \{a_s\}) \cup \{b_s\} \tag{3.4}$$

(which implies $B \subseteq L'$). Moreover, we write $(B \setminus A) \cap L = \emptyset$ if for every $h \in [s]$, if $a_h \neq b_h$, $b_h \notin L_h$. Similarly, we write $(A \setminus B) \cap L' = \emptyset$ if for every $h \in [s]$, if $a_h \neq b_h$, $a_h \notin L'_h$. We note that if $L \sim_{A,B,c} L'$ and L, L' are from the above sets, it follows that $(B \setminus A) \cap L = \emptyset$ as otherwise some L'_h would have been of size smaller than ℓ, and also that $(A \setminus B) \cap L' = \emptyset$ (which follows directly from Equation (3.4)).

We also define two probabilities.

Definition 3.13. *For $r \in \tilde{R}^g(j)$, $z \in [f^{s-g}]$,*

$$\begin{aligned}p^g_A(j; r; z) &= \Pr_L[A^g(j; r; z), c^g(j; r; z) \subseteq L \wedge (B^g(j; r) \setminus A^g(j; r; z)) \cap L = \emptyset]\\ p^g_B(j; r; z) &= \Pr_{L'}[B^g(j; r) \subseteq L' \wedge (A^g(j; r; z) \setminus B^g(j; r)) \cap L' = \emptyset],\end{aligned}$$

where $L \in \binom{[n]}{\ell}^s \times [n]$, $L' \in \binom{[n]}{\ell}^s$ are uniformly random.

It is easy to bound these probabilities.

Claim 3.14. *For any $r \in \tilde{R}^g(j)$, $z \in [f^{s-g}]$,*

$$\frac{1}{n}\left(\frac{\ell}{n}\right)^s \geq p_A^g(j;r;z) \geq \frac{1}{n}\left(\frac{\ell}{n}\right)^s\left(1-\frac{\ell-1}{n-1}\right)^s,$$

$$\left(\frac{\ell}{n}\right)^s \geq p_B^g(j;r) \geq \left(\frac{\ell}{n}\right)^s\left(1-\frac{\ell-1}{n-1}\right)^s.$$

The simple proof for Claim 3.14 is in the appendix. We continue with considering two more conditional probabilities and bound them in the two following claims.

Definition 3.15. *For $A \in [n]^s, B \in [n]^s$, $c \in [n]$, $r \in \tilde{R}^g(j)$ and $z \in [f^{s-g}]$ define*

$$\begin{aligned} p_A^g(j;r;z \mid A,B,c) = \Pr_L[&A^g(j;r;z), c^g(j;r;z) \subseteq L \wedge (B^g(j;r) \setminus A^g(j;r;z) \cap L) = \emptyset \\ &\mid A, c \subseteq L \wedge (B \setminus A) \cap L = \emptyset], \\ p_B^g(j;r;z \mid A,B) = \Pr_{L'}[&B^g(j;r) \subseteq L' \wedge (A^g(j;r;z) \setminus B^g(j;r)) \cap L' = \emptyset \\ &\mid B \subseteq L' \wedge (A \setminus B) \cap L' = \emptyset], \end{aligned}$$

where $L \in \binom{[n]}{\ell}^s \times [n]$, $L' \in \binom{[n]}{\ell}^s$ are uniformly random.

Claim 3.16. *Let $A \in [n]^s$, $c \in [n]$, $r \in \tilde{R}^g(j)$, $z \in [f^{s-g}]$, and $S \subseteq [s]$ be such that for $h \notin S$, $A^g(j;r;z)_h \neq A_h$. Then*

$$p_A^g(j;r;z \mid A,B,c) \leq \begin{cases} \left(\frac{\ell}{n}\right)^{s-|S|} & \text{if } c^g(j;r;z) = c \\ 0 & \text{if } c^g(j;r;z) \neq c. \end{cases}$$

Claim 3.17. *Let $B \in [n]^s$, $r \in \tilde{R}^g(j)$ and $S \subseteq [s]$ be such that for $h \notin S$, $B^g(j;r)_h \neq B_h$. Then for every $z \in [f^{s-g}]$*

$$p_B^g(j;r;z \mid A,B) \leq \left(\frac{\ell}{n}\right)^{s-|S|}.$$

The proofs for the two claims are in the appendix.

We can now define the graphs for every $j \in [n]$.

Definition 3.18. *For every $j \in [n]$ we define two bipartite graphs. The first graph $G_j = (U,V,E_j)$ is defined as follows. The left and right vertices are $U = \binom{[n]}{\ell}^s \times [n]$ and the right vertices are $V = \binom{[n]}{\ell}^s$. For every $r \in \tilde{R}^g(j)$, $z \in [f^{s-g}]$, we define the following set of edges*

$$E_{j,r,z} = \{(L,L') \in U \times V \mid L \sim_{A^g(j;r;z),B^g(j;r),c^g(j;r;z)} L'\}.$$

We say that the edges in $E_{j,r,z}$ are colored *with r, z. The set of edges E_j of G_j is achieved by appending all edges $E_{j,r,z}$ of each color r, z, allowing multiple edges.*

The second graph is $\tilde{G}_j = (U, \tilde{V}, \tilde{E}_j)$ and it is obtained by duplicating each right vertex of G_j so that it will have n copies, duplicating each edge into as many copies as well. Denote $N = |U| = |\tilde{V}|$.

We note that for every $j \in [n]$ we can characterize the set of edges of each color r, z, as we have the following claim.

Claim 3.19. *For every r, z, $E_{j,r,z}$ is a perfect matching between*

$$U_{j,r,z} = \{L \mid A^g(j;r;z), c^g(j;r;z) \subseteq L \wedge (B^g(j;r) \setminus A^g(j;r;z)) \cap L = \emptyset\} \subseteq U$$

and

$$V_{j,r,z} = \{L' \mid B^g(j;r) \subseteq L' \wedge (A^g(j;r;z) \setminus B^g(j;r)) \cap L' = \emptyset\} \subseteq V$$

of size $p^g_A(j;r;z)n\binom{n}{\ell}^s = p^g_B(j;r)\binom{n}{\ell}^s$.

Proof. Denote $A = (a_1, \dots, a_s) = A^g(j;r;z)$, $B = (b_1, \dots, b_s) = B^g(j;r)$ and $c = c^g(j;r;z)$. The claim follows immediately by the definitions as, first, for every $L = (L_1, \dots, L_{s+1}) \in U_{j,r,z}$ there is exactly one $L' = (L'_1, \dots, L'_s)$ such that $(L, L') \in E_{j,r,z}$: $L' = ((L_1 \setminus \{a_1\}) \cup \{b_1\}, \dots, (L_s \setminus \{a_s\}) \cup \{b_s\})$, and indeed $L' \in V_{j,r,z}$, since $B \subseteq L'$, and for every $h \in [s]$ if $a_h \neq b_h$ then $a_h \notin L'$, and lastly $L' \in \binom{[n]}{\ell}^s$ since for every h we removed a_h and added b_h, and $b_h \notin L_h$ if $a_h \neq b_h$. Similarly, for every $L' = (L'_1, \dots, L'_s) \in V_{j,r,z}$ there is exactly one $L = (L_1, \dots, L_{s+1})$ such that $(L, L') \in E_{j,r,z}$: $L = ((L'_1 \setminus \{b_1\}) \cup \{a_1\}, \dots, (L'_s \setminus \{b_s\}) \cup \{a_s\}, c)$, and indeed $L \in U_{j,r,z}$. Lastly, for every $(L, L') \in E_{j,r,z}$, $L \in U_{j,r,z}$ and $L' \in V_{j,r,z}$.

The claimed size of the matching follows as $|U_{j,r,z}| = p^g_A(j;r;z)\binom{n}{\ell}^s n$ and $|V_{j,r,z}| = p^g_B(j;r)\binom{n}{\ell}^s$ by the definitions of $p^g_A(j;r;z)$ and $p^g_B(j;r)$. □

Using this claim, we can see what are the average degrees of the graphs. Set for $j \in [n]$[13]

$$\Delta(j) = f^{s-g}\left(\frac{m\ell}{n}\right)^s. \tag{3.5}$$

We will write Δ as short for $\Delta(j)$ but it will always be in contexts where j is specific. For every $j \in [n]$, Claim 3.19, Claim 3.14 with Equation (3.1) and Claim 3.5 imply that the average degree of the left side of G_j is

$$\sum_{r \in \tilde{R}^g(j), z \in [f^{s-g}]} p^g_A(j;r;z) = |\tilde{R}^g(j)| f^{s-g} \Theta\left(\frac{1}{n}\left(\frac{\ell}{n}\right)^s\right) \in \left[\Omega\left(\frac{1}{sn}\Delta\right), O\left(\frac{1}{n}\Delta\right)\right], \tag{3.6}$$

[13] Recall that $g = g(j)$ is j-specific.

by Claim 3.5 and Claim 3.7, and the average degree of the right side is

$$\sum_{r\in\tilde{R}^g(j),z\in[f^{s-g}]} p_B^g(j;r;z) = |\tilde{R}^g(j)|f^{s-g}\Theta\left(\left(\frac{\ell}{n}\right)^s\right) \in \left[\Omega\left(\frac{1}{s}\Delta\right), O(\Delta)\right].$$

In what follows we argue that there are many edges that touch vertices whose degree is close to the average degree (of their side).

Definition 3.20. *For every $j \in [n]$ we define*

$$E'_j = \left\{(L,L') \in E_j \mid \deg_{G_j}(L) \leq w_0\left(1+\frac{n}{m\ell}\right)^s \frac{1}{m}\Delta\right\}$$
$$E''_j = \{(L,L') \in E_j \mid \deg_{G_j}(L') \leq w_0\left(1+\frac{n}{m\ell}\right)^s \Delta\}$$

for some large enough universal constant w_0.

The fact that there are many edges touching vertices with degree close to average will follow by the fact that this holds within each color.

Proposition 3.21. *Define for every r, z,*

$$E'_{j,r,z} = \left\{(L,L') \in E_{j,r,z} \mid \deg_{G_j}(L) \leq w_0\left(1+\frac{n}{m\ell}\right)^s \frac{1}{m}\Delta\right\}.$$

Then $|E'_{j,r,z}| \geq \frac{2}{3}|E_{j,r,z}|$.

Proposition 3.22. *Define for every r, z,*

$$E''_{j,r,z} = \{(L,L') \in E_{j,r,z} \mid \deg_{G_j}(L') \leq w_0\left(1+\frac{n}{m\ell}\right)^s \Delta\}.$$

Then $|E''_{j,r,z}| \geq \frac{2}{3}|E_{j,r,z}|$.

Before we prove Proposition 3.21 and Proposition 3.22 we conclude that they imply that there exists a large matching in $\tilde{G}_j$. First we note that indeed the above bounds for each color, and for each side separately, imply many such edges in G_j that satisfy the requirement on both of their sides.

Claim 3.23. *For every j, $|E'_j \cap E''_j| \geq \frac{1}{3}|E_j|$.*

Proof. For every r, z, by Proposition 3.21 and Proposition 3.22, $|E'_{j,r,z} \cap E''_{j,r,z}| \geq \frac{1}{3}|E_{j,r,z}|$. As all edges in $E'_{j,r,z} \cap E''_{j,r,z}$ are in $E'_j \cap E''_j$, and edges corresponding to different r, z have different colors, the claim follows. □

We conclude that there is indeed a large matching in each $\tilde{G}_j$.

Lemma 3.24. *$\tilde{G}_j$ contains a matching $M_j \subseteq \tilde{E}_j$ of size $\Omega\left(\frac{1}{\left(1+\frac{n}{m\ell}\right)^s}\frac{m}{sn}N\right)$.*

Proof. To show that it contains a large matching, we won't use all the edges of $\tilde{G}_j$. Rather, we will consider a subset $\tilde{E}'_j \subseteq \tilde{E}_j$, which is chosen as follows. Recall that every edge of $\tilde{G}_j$ is induced by an edge of G_j. First, we will only consider edges induced from an edge e in G_j such that $e \in E'_j \cap E''_j$. Secondly, we will only use one of the n copies of e in $\tilde{G}_j$. Specifically, for every L' which is the right end of such edge, we have that $\deg_{G_j}(L') \leq w_0\left(1+\frac{n}{m\ell}\right)^s\Delta$, and so we will arbitrarily split the set of edges touching L' into at most n parts, indexed by $1,2,\ldots,$ of size at most $\frac{1}{n}w_0\left(1+\frac{n}{m\ell}\right)^s\Delta$. For every such part i we add to $\tilde{E}'_j$ the induced edges touching $(i,L') \in \tilde{V}$. In that way, we ensure that the maximal right degree in $\tilde{E}'_j$ is at most $\frac{1}{n}w_0\left(1+\frac{n}{m\ell}\right)^s\Delta$. As we take every edge in $E'_j \cap E''_j$ exactly once, $|\tilde{E}'_j| = |E'_j \cap E''_j|$, and the maximal left degree is, like in E'_j, at most $w_0\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta$. Notice that as $m \leq n$, both the left and right degrees in $\tilde{E}'_j$ are at most $w_0\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta$. Hence, there is a matching $M_j \subseteq \tilde{E}'_j \subseteq \tilde{E}_j$ of size at least

$$\begin{aligned}\Omega\left(\frac{|E'_j \cap E''_j|}{\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta}\right) &= \Omega\left(\frac{\frac{1}{3}|E_j|}{\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta}\right)\\ &= \Omega\left(\frac{\frac{1}{sn}\Delta N}{\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta}\right)\\ &= \Omega\left(\frac{1}{\left(1+\frac{n}{m\ell}\right)^s}\frac{m}{sn}N\right)\end{aligned}$$

by Equation (3.6) and Claim 3.23. As required. □

We now prove Proposition 3.21 and Proposition 3.22.

Proof for Proposition 3.21. Denote $A = A^g(j;r;z)$, $B = B^g(j;r)$ and $c = c^g(j;r;z)$. Using Claim 3.19 $|E_{j,r,z}| = |U_{j,r,z}|$ for $U_{j,r,z} = \{L \mid A,c \subseteq L \wedge (B \setminus A) \cap L = \emptyset\}$, and $|E'_{j,r,z}| = |\{L \in U_{j,r,z} \mid \deg_{G_j}(L) \leq w_0\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta\}|$, where w_0 is a large enough constant. Thus, by Markov's inequality, to conclude the proposition it is enough to show that

$$\mathop{\mathbf{E}}_{L\mid A,c\subseteq L\wedge(B\setminus A)\cap L=\emptyset}[\deg_{G_j}(L)] = O\left(\left(1+\frac{n}{m\ell}\right)^s\frac{1}{m}\Delta\right).$$

Indeed, using Claim 3.16,

$$\begin{aligned}
\mathop{\mathbf{E}}_{\substack{L \mid A, c \subseteq L \\ \wedge (B \setminus A) \cap L = \emptyset}} [\deg_{G_j}(L)] &= \mathop{\mathbf{E}}_{\substack{L \mid A, c \subseteq L \\ \wedge (B \setminus A) \cap L = \emptyset}} \left[\sum_{r \in \tilde{R}^g(j), z \in [f^{s-g}]} \mathbb{I}_{\substack{A^g(j;r;z), c^g(j;r;z) \subseteq L \\ \wedge (B^g(j;r) \setminus A^g(j;r;z)) \cap L = \emptyset}} \right] \\
&= \sum_{r \in \tilde{R}^g(j), z \in [f^{s-g}]} p_A^g(j; r; z \mid A, B, c) \\
&= \sum_{S \subseteq [s]} \Bigg(\sum_{\substack{r, z \in \{r, z \mid \{h \in [s] \mid A^g(j;r;z)_h = A_h\} = S, \\ c^g(j;r;z) = c\}}} p_A^g(j; r; z \mid A, B, c) \\
&\quad + \sum_{c' \neq c} \sum_{\substack{r, z \in \{r, z \mid \{h \in [s] \mid A^g(j;r;z)_h = A_h\} = S, \\ c^g(j;r;z) = c\}}} p_A^g(j; r; z \mid A, B, c) \Bigg) \\
&\le \sum_{S \subseteq [s]} \sum_{\substack{r, z \in \{r, z \mid \{h \in [s] \mid A^g(j;r;z)_h = A_h\} = S, \\ c^g(j;r;z) = c\}}} \left(\frac{\ell}{n} \right)^{s - |S|} \\
&\le \sum_{S \subseteq [s]} |R^g(j, A, S, c)| \left(\frac{\ell}{n} \right)^{s - |S|} ,
\end{aligned}$$

where recall that the definition of $R^g(j, A, S, c)$ as per Definition 3.9 is all the instructions that result in agreement on the reminder c, and on A_S (the last transition is an inequality because in $R^g(j, A, S, c)$ we don't insist on disagreement outside S). We invoke Claim 3.10 to bound the above sum. We first consider the part of the sum, which is over sets that contain $[g]$, that is $S = [g] \cup S'$ for some $S' \subseteq [s] \setminus [g]$,

$$\begin{aligned}
\sum_{S' \subseteq [s] \setminus [g]} \left(|R^g(j, A, [g] \cup S', c)| \left(\frac{\ell}{n} \right)^{s - g - |S'|} \right) &\le \sum_{S' \subseteq [s] \setminus [g]} m^{s - g - |S'|} \left(\frac{\ell}{n} \right)^{s - g - |S'|} \\
&= \left(1 + \frac{m\ell}{n} \right)^{s - g} .
\end{aligned}$$

The other part of the sum, over sets which don't contain $[g]$, again by Claim 3.10, is bounded by

$$\begin{aligned}
&\sum_{[g]\not\subseteq S\subseteq[s]}\left(|R^g(j,A,S,c)|\left(\frac{\ell}{n}\right)^{s-|S|}\right)\\
\leq&\sum_{[g]\not\subseteq S\subseteq[s]} m^{s-|S|-1} f^{s-\max([g]\setminus S)}\left(\frac{\ell}{n}\right)^{s-|S|}\\
=&\frac{1}{m}\sum_{[g]\not\subseteq S\subseteq[s]} f^{s-\max([g]\setminus S)}\left(\frac{m\ell}{n}\right)^{s-|S|}\\
:=&\alpha(g).
\end{aligned}$$

We continue by bounding $\alpha(g)$ for $g \in [s]$. This time, splitting the sets according to $b = \max([g] \setminus S)$, that is writing $S = S' \cup \{b+1,\dots,g\}$ for some $S' \subseteq [b-1] \cup ([s] \setminus [g])$. Writing $\alpha(g)$ in such manner we get that

$$\begin{aligned}
\alpha(g) =&\frac{1}{m}\sum_{b\in[g]}\sum_{S'\subseteq[b-1]\cup([s]\setminus[g])} f^{s-b}\left(\frac{m\ell}{n}\right)^{s-(g-b)-|S'|}\\
=&\frac{1}{m}\left(\frac{m\ell}{n}\right)\sum_{b\in[g]} f^{s-b}\sum_{S''\subseteq[s-(g-b+1)]}\left(\frac{m\ell}{n}\right)^{|S''|}\\
=&\frac{1}{m}\left(\frac{m\ell}{n}\right)\sum_{b\in[g]} f^{s-b}\left(1+\frac{m\ell}{n}\right)^{s-(g-b)-1}\\
=&\frac{1}{m}\left(\frac{\frac{m\ell}{n}}{1+\frac{m\ell}{n}}\right)\sum_{b\in[g]} f^{s-b}\left(1+\frac{m\ell}{n}\right)^{s-(g-b)}\\
\leq&\frac{1}{m}\sum_{b\in[g]} f^{s-b}\left(1+\frac{m\ell}{n}\right)^{s-(g-b)}\\
=&\frac{1}{m}f^{s-g}\left(1+\frac{m\ell}{n}\right)^{s}\sum_{b\in[g]} f^{g-b}\left(1+\frac{m\ell}{n}\right)^{-(g-b)}\\
=&\frac{1}{m}f^{s-g}\left(1+\frac{m\ell}{n}\right)^{s}\sum_{b=0}^{g-1} f^{b}\left(1+\frac{m\ell}{n}\right)^{-b}.
\end{aligned}$$

And, we continue by plugging Equation (3.2), and we see that

$$\alpha(g) \leq \frac{1}{m}f^{s-g}\left(1+\frac{m\ell}{n}\right)^{s}\cdot O(1),$$

We conclude that, recalling Δ's definition in Equation (3.5),

$$\begin{aligned}
\mathop{\mathbf{E}}_{L \mid A, c \subseteq L \wedge (B \setminus A) \cap L = \emptyset}[\deg_{G_j}(L)] &\leq \left(1 + \frac{m\ell}{n}\right)^{s-g} + \alpha(g) \\
&= \left(1 + \frac{m\ell}{n}\right)^{s-g} + \frac{1}{m} f^{s-g} \left(1 + \frac{m\ell}{n}\right)^{s} \cdot O(1) \\
&= \frac{1}{m} f^{s-g} \left(1 + \frac{m\ell}{n}\right)^{s} \cdot O(1) \\
&= O\left(\left(1 + \frac{n}{m\ell}\right)^{s} \frac{1}{m} \Delta\right),
\end{aligned}$$

where the second equality is by Equation (3.3). The proposition follows. □

Proof for Proposition 3.22. Denote $A = A^g(j;r;z)$ and $B = B^g(j;r)$. Using Claim 3.19 $|E_{j,r,z}| = |V_{j,r,z}|$ for $V_{j,r,z} = \{L' \mid B \subseteq L' \wedge (A \setminus B) \cap L' = \emptyset\}$, and $|E''_{j,r,z}| = |\{L' \in V_{j,r,z} \mid \deg_{G_j}(L') \leq w_0 \left(1 + \frac{n}{m\ell}\right)^s \Delta\}|$, where w_0 is a large enough constant. Thus, by Markov's inequality, to conclude the proposition it is enough to show that

$$\mathop{\mathbf{E}}_{L' \mid B \subseteq L' \wedge (A \setminus B) \cap L' = \emptyset}[\deg_{G_j}(L')] = O\left(\left(1 + \frac{n}{m\ell}\right)^{s} \Delta\right).$$

Indeed,

$$\begin{aligned}
\mathop{\mathbf{E}}_{L' \mid B \subseteq L' \wedge (A \setminus B) \cap L' = \emptyset}[\deg_{G_j}(L)] &= \mathop{\mathbf{E}}_{L' \mid B \subseteq L' \wedge (A \setminus B) \cap L' = \emptyset}\left[\sum_{r \in \tilde{R}^g(j), z \in [f^{s-g}]} \mathbb{I}_{\substack{B^g(j;r) \subseteq L' \\ \wedge (A^g(j;r;z) \setminus B^g(j;r)) \cap L' = \emptyset}}\right] \\
&= \sum_{r \in \tilde{R}^g(j), z \in [f^{s-g}]} p^g_B(j;r;z \mid A, B) \\
&\sum_{S \subseteq [s]} \sum_{z \in [f^{s-g}]} \sum_{\substack{r \in \{r \mid \\ \{h \in [s] \mid B^g(j;r)_h = B_h\} = S\}}} p^g_B(j;r;z \mid A, B) \\
&\leq \sum_{S \subseteq [s]} \sum_{z \in [f^{s-g}]} \sum_{\substack{r \in \{r \mid \\ \{h \in [s] \mid B^g(j;r)_h = B_h\} = S\}}} \left(\frac{\ell}{n}\right)^{s-|S|} \\
&= f^{s-g} \sum_{S \subseteq [s]} \sum_{\substack{r \in \{r \mid \\ \{h \in [s] \mid B^g(j;r)_h = B_h\} = S\}}} \left(\frac{\ell}{n}\right)^{s-|S|},
\end{aligned}$$

where the inequality follows by Claim 3.17. We continue, noting that the above is bounded above by

$$\begin{aligned}
&f^{s-g}\sum_{S\subseteq[s]}|R^g(j,B,S)|\left(\frac{\ell}{n}\right)^{s-|S|}\\
&\leq f^{s-g}\sum_{S\subseteq[s]}m^{s-|S|}\left(\frac{\ell}{n}\right)^{s-|S|},\\
&=f^{s-g}\left(1+\frac{m\ell}{n}\right)^{s}\\
&=\left(1+\frac{n}{m\ell}\right)^{s}\Delta,
\end{aligned}$$

where the inequality is using Claim 3.11. The proposition follows. □

3.4 Deducing the bound

Let $C:\{0,1\}^k\to\{0,1\}^n$ be a linear $(3,\delta,\varepsilon)$-LCC. Set $m=\frac{\delta n}{3}$. Without loss of generality, by Fact 2.1 for every $i\in[k]$, $C(x)_i=x_i\ \forall x$. By Fact 2.3 there exist sets $Q^j_r\mid_{j\in[n],r\in[m]}$ of size at most 3, such that for every $j\in[n]$, the m sets $Q^j_w\mid_{w\in[m]}$ are disjoint, and for every $w\in[m]$,

$$C(x)_j=\sum_{j'\in Q^j_w}C(x)_{j'}\ \forall x\in\{0,1\}^k. \tag{3.7}$$

Without loss of generality the sets $Q^j_r\mid_{j\in[n],r\in[m]}$ are of size exactly 3 (we can add zero coordinates, at worst doubling n).

Let s and ℓ be parameters. Set $U=\{0\}\times\binom{[n]}{\ell}^s\times[n]$ and $V=\{1\}\times\binom{[n]}{\ell}^s\times[n]$. We define a new code $C':\{0,1\}^k\to\{0,1\}^{U\cup V}$ as follows. For every $x\in\{0,1\}^k$

$$\forall u=(0,L_1,\ldots,L_s,c)\in U: C'(x)_u=C(x)_c+\sum_{t\in[s],j\in L_t}C(x)_j.$$

$$\forall v=(1,L'_1,\ldots,L'_s,c')\in V: C'(x)_v=\sum_{t\in[s],j\in L'_t}C(x)_j.$$

That is, C' is defined on top two coordinate sets which are all the possibilities for s subsets of $[n]$ of size ℓ, and one more coordinate c in $[n]$. In each coordinate in the U side, $C'(x)$ has the sum of the elements of $C(x)$ which are contained in the chosen sets, and the element of the extra coordinate. On the V side, $C'(x)$ has the sum of the elements of $C(x)$ which are contained in the chosen sets, and the extra coordinate is not used.

Let $A(j;r)$, $B(j;r)$, $c(j;r)$ and $A^g(j;r;z)$, $B^g(j;r)$, $c^g(j;r;z)\mid_{j,g,r,z}$ be the decoding sequences and extended decoding sequences considered in the previous sections. The following claim can easily be verified, by inspecting the definition of decoding sequences and extended decoding sequences, and relaying on Equation (3.7) (a proof is given in the appendix).

Claim 3.25. *For every $j \in [n]$ and $r \in [m]^s$, $\sum_{h\in[s]} C(x)_{A(j;r)_h} + \sum_{h\in[s]} C(x)_{B(j;r)_h} + C(x)_{c(j;r)} = C(x)_j \; \forall x \in \{0,1\}^k$. Similarly, $\sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + \sum_{h\in[s]} C(x)_{B^g(j;r)_h} + C(x)_{c^g(j;r;z)} = C(x)_j \; \forall x \in \{0,1\}^k$.*

We also observe the following direct implication.

Claim 3.26. *Let $u = (0, L_1, \dots, L_s, c) \in U$ and $v = (1, L'_1, \dots, L'_s, c') \in U$ be such that $(L_1, \dots, L_s, c) \sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)} (L'_1, \dots, L'_s)$ where the $\sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)}$ relation is as defined in the previous part in Definition 3.12. Then $C'(x)_u + C'(x)_v = C(x)_j$ for every x.*

Proof. We argue that

$$C'(x)_u + C'(x)_v = \sum_{h\in S} C(x)_{A^g(j;r;z)_h} + \sum_{h\in[s]} C(x)_{B^g(j;r)_h} + C(x)_{c^g(j;r;z)}.$$

To see this, denote $A^g(j;r;z) = (a_1, \dots, a_s)$ and $B^g(j;r) = (b_1, \dots, b_s)$. By the assumption $(L_1, \dots, L_s, c) \sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)} (L'_1, \dots, L'_s)$ we have that $c = c^g(j;r;z)$ and for every $h \in [s]$

$$C(x)_{L_h} + C(x)_{L'_h} = C(x)_{L_h} + C(x)_{L_h \setminus \{a_h\}} + C(x)_{b_h} = C(x)_{a_h} + C(x)_{b_h}$$

and so

$$C'(x)_u + C'(x)_v = \sum_{h\in[s]} C(x)_{L_h} + c + \sum_{h\in[s]} C(x)_{L'_h} = \sum_{h\in[s]} (a_h + b_h) + c$$

and so the claim follows by applying Claim 3.25. □

For every $j \in [n]$ let $\tilde{G}_j$ be the bipartite graph defined in the previous section. Note that $|U| = |V| = N$ and there is a natural isomorphism between the vertices of $\tilde{G}_j$ and the coordinates of C', given by that every left side vertex $L \in \binom{[n]}{\ell}^s \times [n]$ corresponds to the coordinate $u = (0, L) \in U$, and every right vertex (i, L') for $i \in [n]$ and $L' \in \binom{[n]}{\ell}^s$ corresponds to the coordinate $v = (1, L', i) \in V$. By the previous claim, for every edge $(L, (i, L'))$ in $\tilde{G}_j$ and its corresponding coordinates (u, v), we have that $C'(x)_u + C'(x)_v = C(x)_j$ for every $x \in \{0,1\}^k$.

We can now apply Lemma 3.24 and conclude that for every $j \in [n]$ there is a set of

$$\Omega\left(\frac{1}{(1+\frac{n}{m\ell})^s}\frac{m}{sn}N\right) = \Omega\left(\frac{1}{(1+\frac{1}{\delta\ell})^s}\frac{\delta}{s}N\right)$$

disjoint pairs of coordinates $\{u, v\}$ of C' such that $C'(x)_u + C'(x)_v = C(x)_j \; \forall x$.

We can now conclude the bound.

Theorem 3.27 (Theorem 1.3, rephrased)**.** *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear $(3, \delta, \varepsilon)$-LCC. Then $k = O\left(\frac{1}{\delta^2} \log^4 n\right)$.*

Proof. Consider C' as defined according to C and note that it is a code of length $2N$, where $N = n\binom{n}{\ell}^s$. Set $\delta_{C'} = \frac{1}{\left(1+\frac{1}{\delta\ell}\right)^s}\frac{\delta}{s}$. From the conclusion of the previous paragraph, for every $i \in [k]$ there is a set of $\Omega(\delta_{C'} N)$ disjoint pairs of coordinates $\{u, v\}$ such that $C'(x)_u + C'(x)_v = C(x)_i = x_i \ \forall x \in \{0,1\}^k$. Hence, by Theorem 2.2,

$$k = O\left(\frac{1}{\delta_{C'}} \log N\right) = O\left(\frac{1}{\delta_{C'}} s\ell \log n\right).$$

Thus, if we set $\ell = \Theta(\frac{1}{\delta} \log n)$ and $s = \Theta(\log n)$ then the assumption in Equation (3.3) is met, and as $\left(1 + \frac{1}{\delta\ell}\right)^s = O(1)$, $\delta_{C'} = \Omega(\frac{\delta}{s})$. We get that

$$k = O\left(\frac{1}{\delta} s^2 \ell \log n\right) = O\left(\frac{1}{\delta^2} \log^4 n\right),$$

as required. □

References

[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.

[BDYW11] Boaz Barak, Zeev Dvir, Amir Yehudayoff, and Avi Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 519–528, 2011.

[DGOS18] Zeev Dvir, Ankit Garg, Rafael Oliveira, and József Solymosi. Rank bounds for design matrices with block entries and geometric applications. *Discrete Analysis*, 5(2018):1–24, 2018.

[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, page e4. Cambridge University Press, 2014.

[Efr09] Klim Efremenko. 3-query locally decodable codes of subexponential length. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 39–44, 2009.

[GKST02] Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings 17th IEEE Annual Conference on Computational Complexity*, pages 175–183. IEEE, 2002.

[KdW04] 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.

[KM23] Pravesh K Kothari and Peter Manohar. An exponential lower bound for linear 3-query locally correctable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, number 162, 2023.

[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.

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

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

[Yek11] S. Yekhanin. Locally decodable codes. In *International Computer Science Symposium in Russia*, pages 289–290. Springer, 2011.

[ZD] Kalina Petrova Zeev Dvir. Lecture 1: Introduction. Lecture notes: https://www.cs.princeton.edu/~zdvir/LDCnotes/LDC1.pdf, year=2016,.

A The case of design LCC

In this part we prove Theorem 1.6. We start by restating the definition of design LCC.

Definition A.1. *We say that a (q, δ, ε)-LCC is a* design LCC *if there exists a randomized procedure satisfying the requirements in Definition 1.1 and further: For every distinct pair of coordinates $a, c \in [n]$, there are at most $O(1)$ j's such that a and c can both be queried by invoking (once) the procedure to correct j.*

Let $C : \{0,1\}^k \to \{0,1\}^n$ be a linear $(3, \delta, \varepsilon)$ design LCC. Set $m = \frac{\delta n}{3}$. It follows by the proof for Fact 2.3 that there exist sets $Q_r^j \mid_{j \in [n], r \in [m]}$ of size at most 3, such that for every $j \in [n]$, the m sets $Q_w^j \mid_{w \in [m]}$ are disjoint, and for every $w \in [m]$,

$$C(x)_j = \sum_{j' \in Q_w^j} C(x)_{j'},$$

and further, for every distinct $a, c \in [n]$,

$$|\{j \in [n] \mid \exists w \in [m] : \{a, c\} \subseteq Q_w^j\}| \leq O(1). \tag{A.1}$$

We now notice the following fact, which takes more care in proving than in the non-design case.

Claim A.2. *We can assume without loss of generality that the sets $Q_r^j \mid_{j \in [n], r \in [m]}$ are of size exactly 3.*

Proof. We show that we can add some $O(n)$ coordinates to C and $\{Q_w^j\}$, such that each set is of size 3, without invalidating Equation (A.1) (and the other assumed properties).

We first wish to argue that by adding $O(n)$ zero coordinates we can increase by 1 the size of every Q_w^j for which $|Q_w^j| < 3$ while maintaining Equation (A.1). If indeed we can do that, after doing it once we will do it once more, and be done.

Towards that, we first argue that we can add $n' \in [n, 4n]$ zero coordinates that satisfy the requirements. We do this by adding $n' = 4^{\lceil \frac{1}{2} \log n \rceil}$ zero coordinates, $o_1, \ldots, o_{n'}$ to C. We identify $\{o_1, \ldots, o_{n'}\}$ with $(\mathbb{F}_4)^{\lceil \frac{1}{2} \log n \rceil}$ and we denote by L the set of all lines in $(\mathbb{F}_4)^{\lceil \frac{1}{2} \log n \rceil}$[14]. For every $j \in [n']$, we construct $\{Q_w^{o_j}\}_w$ by taking all lines that pass through o_j, and for each such line $\ell = \{o_j, o_{j_1}, o_{j_2}, o_{j_3}\}$ we define the set $Q_\ell^{o_j} = \{o_{j_1}, o_{j_2}, o_{j_3}\}$. We then take the first m lines $\ell_1, \ldots, \ell_m$ and for every $w \in [m]$ we set $Q_w^{o_j} = Q_{\ell_w}^{o_j}$. Notice that the number of such lines is $\frac{n'-1}{3} \geq \frac{n-1}{3} > m$ and so there are enough lines. Thus, the zero coordinates we added $\{o_1, \ldots, o_{n'}\}$ have sets $\{Q_w^{o_j}\}$ of size exactly 3, for every o_j the sets are disjoint since its lines are disjoint. Moreover, Equation (A.1) holds with regards to $\{o_1, \ldots, o_{n'}\}$ since every pair of coordinates is contained in one line $\ell = \{o_{j_1}, o_{j_2}, o_{j_3}, o_{j_4}\}$.

We use the n coordinates $\{o_1, \ldots, o_n\}$ to increase the size of too small sets, as follows. For every j, w such that $|Q_w^j| < 3$ we arbitrarily choose $a \in Q_w^j$ (if $|Q_w^j| = 1$ there is only one choice, otherwise there are two[15]), and we set $Q_w'^j = Q_w^j \cup \{o_{1+(j+a \mod n)}\}$. We

[14] That is, $L = \{\{At + B \mid t \in \mathbb{F}_4\} \mid A, B \in (\mathbb{F}_4)^{\lceil \frac{1}{2} \log n \rceil}\}$.

[15] We can assume without loss of generality that there are no query sets of size 0 since without loss of generality (the original) C doesn't contain coordinates fixed to zero (removing such coordinates can only improve the parameters of the LCC, so we first remove them and then apply the transformation in the proof).

argue that in doing this, we maintained Equation (A.1). Indeed, every pair x, y such that $x, y \notin \{o_1, \ldots, o_n\}$ satisfied Equation (A.1) before and still does. So is the case if $x, y \in \{o_1, \ldots, o_n\}$. Now, for a pair $x \notin \{o_1, \ldots, o_n\}$ and $y \in \{o_1, \ldots, o_n\}$ such that $\{x, y\} \subseteq Q_w^j$ for some j and w, we have that either $y = o_{1+(x+j \mod n)}$ or $y = o_{1+(z+j \mod n)}$ for $z \in Q_w^j$. The first case can only occur for one j, since $(x+j \mod n) \neq (x+j' \mod n)$ if $j \neq j'$. The second case can only occur for $O(1)$ j's, because for every j that satisfies it, there is some w such that $\{x, z\} \subseteq Q_w^j$, and so by Equation (A.1).

Thus we have shown that we can increase by 1 the size of too small sets, while maintaining the properties, by adding $O(n)$ zero coordinates. The claim follows. □

The saving in the design case will follow by that we argue that we can reduce to the case that for every g, $A \subseteq [n]^{s-g}$, $c \in [n]$, $\deg(A; c) \leq 1$. We first require the following lemma which addresses the degree of suffixes of length 1.

Lemma A.3. *Assume that for every distinct $a, c \in [n]$, $|\{j \in [n] \mid \exists w \in [m] : \{a, c\} \subseteq Q_w^j\}| \leq O(1)$. Then, given that $m = \omega(\log n)$, there is a way to order each Q_w^j as three parts $A(j; w)$, $B(j; w)$, $c(j; w)$ such that for every $j \in [n]$, $|\{w \in [m] \mid \deg(A(j; w); c(j; w)) = 1\}| = \Omega(m)$.*[16] [17]

Proof. The proof is by the probabilistic method.[18] We assume that for every distinct $a, c \in [n]$, $|\{j \in [n] \mid \exists w \in [m] : \{a, c\} \subseteq Q_w^j\}| \leq y_0$ for some constant y_0. For every $j \in [n]$ and $w \in [m]$ we choose a uniformly random ordering $\{A(j; w), B(j; w), c(j; w)\} = Q_w^j$. Notice that given that we chose a specific ordering $\{A(j; w), B(j; w), c(j; w)\}$ for Q_w^j, the probability that $\deg(A(j; w); c(j; w)) > 1$ is bounded above by a constant smaller then 1. Indeed, there are at most $y_0 - 1$ $j' \in [n] \setminus \{j\}$ such that $\{A(j; w), c(j; w)\} \subseteq Q_{w'}^{j'}$ for some w'. Note that $\deg(A(j; w); c(j; w)) > 1$ only if for one of those we chose $A(j'; w') = A(j; w)$ and $c(j'; w') = c(j; w)$, and the probability that this occurs for (j', w') is less than $\frac{1}{q}$ (it is $\frac{1}{6}$ in the case that $q = 3$). Since there are at most $y_0 - 1$ such j's, and their corresponding events are independent, the probability that $\deg(A(j; w); c(j; w)) > 1$ is bounded by $\alpha := 1 - (1 - \frac{1}{q})^{y_0 - 1} < 1$, which is a constant.

Fix some $j \in [n]$. We bound the probability p_j that $|\{w \in [m] \mid \deg(A(j; w); c(j; w)) > 1\}| > \sqrt{\alpha} m$. From the above, $\mathbf{E}[\sum_{w \in [m]} \mathbb{I}_{\deg(A(j;w);c(j;w))>1}] \leq \alpha m$. Notice that for $w \neq w'$, the events $\deg(A(j; w); c(j; w)) > 1$ and $\deg(A(j; w'); c(j; w')) > 1$ are not necessarily independent, rather, they are negatively correlated, as $A(j; w') \neq A(j; w)$ and $c(j; w') \neq c(j; w)$. By the Chernoff bound for negatively correlated random variables, $p_j = \mathbf{Pr}[\sum_{w \in [m]} \mathbb{I}_{\deg(A(j;w);c(j;w))>1} > \frac{1}{\sqrt{\alpha}} \alpha m] \leq 2^{-\Omega(m)}$.

[16] This can be generalized for larger q's as well.

[17] The $\Omega(m)$ bound can be meaningful only for m large enough.

[18] Note that a naive greedy approach could lead to some j's losing many (or all) or their sets.

Hence, taking a union bound over all $j \in [n]$, the probability that for some $j \in [n]$, $|\{w \in [m] \mid \deg(A(j;w);c(j;w)) > 1\}| > \sqrt{\alpha}m$, is bounded by $n2^{-\Omega(m)} = o(1)$ per the assumption on m. The lemma follows. □

Notice that we can indeed assume $m = \omega(\log n)$ as in the hypothesis of the lemma, since if $m = O(\log n)$ then $\delta = O(\frac{\log n}{n})$ and the bound in Theorem 1.6 holds trivially. Therefore, without loss of generality, we will assume that the ordering of $\{Q_w^j\}$ from which the decoding sequences are defined satisfies Lemma A.3.

We now note that the above implies that the degree of every suffix is bounded by 1 (and not only for suffixes of length 1).

Claim A.4. *If for every $a, c \in [n]$, $\deg(a;c) \le 1$, then for every $g \in [s]$, $A \in [n]^{s-g}$ and $c \in [n]$, $\deg(A;c) \le 1$.*

Proof. The proof is by induction. The base case for $g = s$ holds trivially by the definitions. As for the induction step, for every $g < s$, for every j such that there exists $r \in [m]^{s-g}$ for which $A(j;r) = A$ and $c(j;r) = c$, we have that for $j' = c(j;r_1)$ it holds that $A(j';r_{2,\dots,s-g}) = A_{2,\dots,s-g}$ and $c(j';r_{2,\dots,s-g}) = c$. By the induction hypothesis, there is at most one j' for which this holds. But, we also must have that $A(j;r)_1 = A(j;r_1) = A_1$. So j must satisfy $A(j;r_1) = A_1$ and $c(j;r_1) = j'$, and so as $\deg(A_1;j') \le 1$ per the assumption, there is at most one such j. □

We can now deduce the theorem.

Proof for Theorem 1.6. Exactly the same as the proof for Theorem 3.27, except for the following. In the proof for Theorem 3.27 we got that there is a set of disjoint pairs of size $\Omega\left(\frac{1}{\left(1+\frac{1}{\delta\ell}\right)^s}\frac{\delta}{s}N\right)$. The $\frac{1}{s}$ factor in the expression was inherited from applying Claim 3.5 and Claim 3.7 by Section 3.3 to argue that $|\tilde{R}^g(j)| \ge \frac{1}{2}|R^g(j)| = \Omega(\frac{1}{s}m^s)$ for every j (in Equation (3.6)). In our current case, we note that we can just take $g = s$ for every j and get larger $\tilde{R}^g(j)$'s. Indeed, since $\deg(A;c) \le 1$ for every suffix $A;c$, $R^s = [m]^s$, where R^s is as defined in Definition 3.4, and so $|\tilde{R}^s(j)| = \Omega(m^s)$.

Thus, in our case, we do not lose the $\frac{1}{s}$ factor and for every $i \in [k]$ there is a set of disjoint pairs of size $\Omega\left(\frac{1}{\left(1+\frac{1}{\delta\ell}\right)^s}\delta N\right)$. Setting the same ℓ and s as in the proof for Theorem 3.27, we get that

$$k = O\left(\frac{1}{\delta}s\ell\log n\right) = O\left(\frac{1}{\delta^2}\log^3 n\right).$$

The theorem follows. □

B Easy claims - proofs

Proof for Claim 3.14. Denote $A^g(j;r;z) = (a_1, \dots, a_s)$, $B^g(j;r) = (b_1, \dots, b_s)$ and $c^g(j;r;z) = c$. Since $L = (L_1, \dots, L_s, L_{s+1}) \in \binom{[n]}{\ell} \times [n]$ is a product we can bound separately for $h \in [s]$ that probability p_h that $a_h \in L_h$ and $b_h \notin L_h$ if $a_h \neq b_h$. We see that $\frac{\ell}{n}\left(1 - \frac{\ell-1}{n-1}\right) \leq p_h \leq \frac{\ell}{n}$: the probability that $a_h \in L_h$ is $\frac{\ell}{n}$ and if $a_h \neq b_h$, $p_h = \frac{\ell}{n} \cdot \mathbf{Pr}_{L_h}[b_h \notin L \setminus \{a_h\} \mid a_h \in L] = \frac{\ell}{n} \cdot \left(1 - \frac{\ell-1}{n-1}\right)$. Hence, as the probability that $L_{s+1} = c$ is $\frac{1}{n}$, $\frac{1}{n}\left(\frac{\ell}{n}\right)^s \cdot \left(1 - \frac{\ell-1}{n-1}\right)^s \leq p^g_A(j;r;z) \leq \frac{1}{n}\left(\frac{\ell}{n}\right)^s$. Similarly for $p^g_B(j;r;z)$. □

Proof for Claim 3.16. Clearly if $c^g(j;r;z) \neq c$ the conditional probability is zero since we conditioned on $L_{s+1} = c$. If $c^g(j;r;z) = c$, then $p^g_A(j;r;z \mid A, B, c) \leq \prod_{h \in [s] \setminus S} p_h$ where for $h \in [s] \setminus S$, $p_h := \mathbf{Pr}_{L_h}[A^g(j;r;z)_h \in L_h \mid A_h \in L_h] = \frac{\ell-1}{n-1} < \frac{\ell}{n}$, as $A^g(j;r;z)_h \neq A_h$ for $h \notin S$. Thus $p^g_A(j;r;z \mid A, B, c) \leq \left(\frac{\ell}{n}\right)^{|S|-s}$. □

Proof for Claim 3.17. Identical to the case that $c^g(j;r;z) = c$ in the previous proof. □

Proof for Claim 3.25. The proof that $\sum_{h \in [s]} C(x)_{A(j;r)_h} + \sum_{h \in [s]} C(x)_{B(j;r)_h} + C(x)_{c(j;r)} = C(x)_j$ is by inspecting Definition 3.1 and by induction on s. For the base case $s = 0$, the two summations are empty, and $C(x)_{c(j;r)} = C(x)_{c(j;\varepsilon)} = C(x)_j$ by definition. As for the induction step, for every $s > 0$ we have that

$$
\begin{aligned}
&\sum_{h \in [s]} C(x)_{A(j;r)_h} + \sum_{h \in [s]} C(x)_{B(j;r)_h} + C(x)_{c(j;r)} \\
=&C(x)_{A(j;r)_1} + C(x)_{B(j;r)_1} + \sum_{h \in \{2,\dots,s\}} C(x)_{A(j;r)_h} + \sum_{h \in \{2,\dots,s\}} C(x)_{B(j;r)_h} + C(x)_{c(j;r)} \\
=&C(x)_{A(j;r_1)} + C(x)_{B(j;r_1)} + \\
&\sum_{h \in [s-1]} C(x)_{A(c(j;r_1);r_{2,\dots,s})_h} + \sum_{h \in [s-1]} C(x)_{B(c(j;r_1);r_{2,\dots,s})_h} + C(x)_{c(c(j;r_1);r_{2,\dots,s})} \\
&= C(x)_{A(j;r_1)} + C(x)_{B(j;r_1)} + C(x)_{c(j;r_1)} \\
&= C(x)_j,
\end{aligned}
$$

where the penultimate equality is by the induction hypothesis, and the last equality is as $\{A(j;r_1), B(j;r_1), c(j;r_1)\} = Q^j_{r_1}$ and by Equation (3.7).

The proof that $\sum_{h \in [s]} C(x)_{A^g(j;r;z)_h} + \sum_{h \in [s]} C(x)_{B^g(j;r)_h} + C(x)_{c^g(j;r;z)}$ follows by inspecting Definition 3.8 noting that for $j' = P(j; c(j; r_{1,\dots,g}); A(j;r)_{g+1,\dots,s}; c(j;r))_z$ and

$r' = T(j; c(j; r_{1,\ldots,g}); A(j;r)_{g+1,\ldots,s}; c(j;r))_z$ we have that

$$\begin{aligned}
&\sum_{h\in[s]} C(x)_{B^g(j;r)_h} + \sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + C(x)_{c^g(j;r;z)}\\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} + \sum_{h\in[g]} C(x)_{A(j;r)_h} + \sum_{h\in[s-g]} C(x)_{B(j';r')_h} + C(x)_{j'}\\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} + \sum_{h\in[g]} C(x)_{A(j;r)_h} + \sum_{h\in[s-g]} C(x)_{A(j';r')_h} + \sum_{h\in[s-g]} C(x)_{B(j';r')_h}\\
&\quad + C(x)_{c(j';r')} + C(x)_{j'} - \sum_{h\in[s-g]} C(x)_{A(j';r')_h} - C(x)_{c(j';r')}\\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} + \sum_{h\in[g]} C(x)_{A(j;r)_h} + \sum_{h\in[s-g]} C(x)_{A(j';r')_h} + C(x)_{c(j';r')}\\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} + \sum_{h\in[g]} C(x)_{A(j;r)_h} + \sum_{h\in\{g+1,\ldots,s\}} C(x)_{A(j;r)_h} + C(x)_{c(j;r)}\\
&= C(x)_j,
\end{aligned}$$

where the third equality is as

$$C(x)_{j'} = \sum_{h\in[s-g]} C(x)_{A(j';r')_h} + \sum_{h\in[s-g]} C(x)_{B(j';r')_h} + C(x)_{c(j';r')}$$

and the penultimate equality is by the definitions of $P(j; c(j; r_{1,\ldots,g}); A(j;r)_{g+1,\ldots,s}; c(j;r))_z$ and $T(j; c(j; r_{1,\ldots,g}); A(j;r)_{g+1,\ldots,s}; c(j;r))_z$. □

C The case of larger alphabets

In this section we prove Theorem 1.7. The proof is composed of two steps. In the first step we define *special-form LCC* which, informally, are LCC in which the decoding procedure of each coordinate is done by multiplying some queried coordinates by either 1 or -1, and then summing up the obtained values. The idea is to show that in the case of special-form LCC the same bound on k holds as in the case of binary LCC. In the second step we argue that a linear LCC over any field $\mathbb{F}$ can be converted to a special-form LCC, at the cost of a $|\mathbb{F}|$-dependent deterioration in δ. This second step is identical to the preconditioning done by [KM23] in their handling of the case $\mathbb{F} \neq \mathbb{F}_2$.

C.0.1 Preliminaries for this section

We start by formally extending the definition of LCC to alphabets other than binary.

Definition C.1. *An injective $C : \mathbb{F}^k \to \mathbb{F}^n$ is a (q, δ, ε)*-LCC*, for $\varepsilon < 1 - \frac{1}{|\mathbb{F}|}$, if there exists a randomized procedure that takes as input $j \in [n]$, gets oracle access to $z \in \mathbb{F}^n$ at relative Hamming distance at most δ from $C(x)$ for some x, and in making at most q queries to z, and with probability at least $1 - \varepsilon$: its output is equal to $C(x)_j$. We say that C is linear if it is a linear map.*

We will make use of the following special case of the [GKST02] bound for linear 2-LDCs over any field.

Theorem C.2 ([GKST02])**.** *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a linear map such that for every $i \in [k]$ there is a set of $\delta_c n$ disjoint pairs of coordinates $\{u, v\} \subseteq [n]$ such that $x_i = C(x)_u + C(x)_v$ $\forall x \in \mathbb{F}^k$. Then $k = O(\frac{1}{\delta_c}(\log n + \log |\mathbb{F}|))$.*

Definition C.3. *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a linear map, let $j \in [n]$ and let $Q = \{Q_1, \ldots, Q_{|Q|}\} \subseteq [n]$. We say that Q* determines *j* (in *C*) *if there exists $\alpha_1, \ldots, \alpha_{|Q|} \in \mathbb{F}$ such that $C(x)_j = \sum_{t \in [|Q|]} \alpha_t C(x)_{Q_t}$ $\forall x \in \mathbb{F}^k$.*

Fact C.4. *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a linear (q, δ, ε)-LCC. Then there exist sets $\{Q^j_w\}_{j \in [n], w \in [m]}$ for $m \geq \delta n / q$ such that for every $j \in [n]$ the sets $Q^j_w \mid_{w \in [m]}$ are disjoint subsets of $[n]$ of size at most q, such that each Q^j_w determines j in C.*

We include for completeness a proof sketch for Fact C.4 in Appendix D.

C.0.2 A bound for special-form LCC

Definition C.5. *We say that an injective linear map $C : \mathbb{F}^k \to \mathbb{F}^n$ is a special-form $(3, \delta)$-LCC if for every $j \in [n]$ there exist $m = \frac{\delta n}{3}$ disjoint sets $Q^j_r = \{A(j; r), B(j; r), c(j; r)\} \mid_{r \in [m]}$ of size 3 such that for every $r \in [m]$*

$$C(x)_j = C(x)_{B(j;r)} - C(x)_{A(j;r)} + C(x)_{c(j;r)} \ \forall x \in \mathbb{F}^k. \tag{C.1}$$

Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a special-form $(3, \delta)$-LCC. Set $m = \frac{\delta n}{3}$. By Fact 2.1 we can assume without loss of generality that for every $i \in [k]$, $C(x)_i = x_i$ $\forall x \in \mathbb{F}^k$.

Let s and ℓ be parameters. Set $U = \{0\} \times \binom{[n]}{\ell}^s \times [n]$ and $V = \{1\} \times \binom{[n]}{\ell}^s \times [n]$. We define a new code $C' : \mathbb{F}^k \to \mathbb{F}^{U \cup V}$ as follows. For every $x \in \mathbb{F}^k$

$$\forall u = (0, L_1, \ldots, L_s, c) \in U : C'(x)_u = C(x)_c - \sum_{t \in [s], j \in L_t} C(x)_j.$$

$$\forall v = (1, L'_1, \ldots, L'_s, c') \in V : C'(x)_v = \sum_{t \in [s], j \in L'_t} C(x)_j.$$

Notice that the only change from the way that C' is defined in Section 3.4 is the minus sign in the definition of $C'(x)_u$.

Let $A(j;r)$, $B(j;r)$, $c(j;r)$ and $A^g(j;r;z)$, $B^g(j;r)$, $c^g(j;r;z)\ |_{j,g,r,z}$ be the decoding sequences and extended decoding sequences defined in Section 3.1 and Section 3.2, according to the sets $Q_r^j = \{A(j;r), B(j;r), c(j;r)\}\ |_{j\in[n],r\in[m]}$ by virtue of which C is a special-from $(3,\delta)$-LCC. We argue that the following claim, similar to Claim 3.25 of Section 3.4, holds.

Claim C.6. *For every $j \in [n]$ and $r \in [m]^s$, $\sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[s]} C(x)_{A(j;r)_h} + C(x)_{c(j;r)} = C(x)_j\ \forall x \in \mathbb{F}^k$. Similarly, $\sum_{h\in[s]} C(x)_{B^g(j;r)_h} - \sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + C(x)_{c^g(j;r;z)} = C(x)_j\ \forall x \in \mathbb{F}^k$.*

Proof. The proof that $\sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[s]} C(x)_{A(j;r)_h} + C(x)_{c(j;r)} = C(x)_j$ is by inspecting Definition 3.1 and by induction on s. For the base case $s = 0$, the two sums are empty, and $C(x)_{c(j;r)} = C(x)_{c(j;\varepsilon)} = C(x)_j$ by definition. As for the induction step, for $s > 0$ we have that

$$\begin{aligned}
&\sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[s]} C(x)_{A(j;r)_h} + C(x)_{c(j;r)} \\
=&C(x)_{B(j;r)_1} - C(x)_{A(j;r)_1} + \sum_{h\in\{2,\ldots,s\}} C(x)_{B(j;r)_h} - \sum_{h\in\{2,\ldots,s\}} C(x)_{A(j;r)_h} + C(x)_{c(j;r)} \\
=&C(x)_{B(j;r_1)} - C(x)_{A(j;r_1)} \\
&+ \sum_{h\in[s-1]} C(x)_{B(c(j;r_1);r_{2,\ldots,s})_h} - \sum_{h\in[s-1]} C(x)_{A(c(j;r_1);r_{2,\ldots,s})_h} + C(x)_{c(c(j;r_1);r_{2,\ldots,s})} \\
&= C(x)_{B(j;r_1)} - C(x)_{A(j;r_1)} + C(x)_{c(j;r_1)} \\
&= C(x)_j,
\end{aligned}$$

where the penultimate equality is by the induction hypothesis, and the last equality is as $\{A(j;r_1), B(j;r_1), c(j;r_1)\} = Q_{r_1}^j$ and by Equation (C.1).

The proof that $\sum_{h\in[s]} C(x)_{B^g(j;r)_h} - \sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + C(x)_{c^g(j;r;z)}$ follows by inspecting Definition 3.8 noting that for $j' = P(j; c(j;r_{1,\ldots,g}); A(j;r)_{g+1,\ldots,s}; c(j;r))_z$ and

$r' = T(j; c(j; r_{1,\dots,g}); A(j; r)_{g+1,\dots,s}; c(j; r))_z$ we have that

$$
\begin{aligned}
&\sum_{h\in[s]} C(x)_{B^g(j;r)_h} - \sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + C(x)_{c^g(j;r;z)} \\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[g]} C(x)_{A(j;r)_h} - \sum_{h\in[s-g]} C(x)_{B(j';r')_h} + C(x)_{j'} \\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[g]} C(x)_{A(j;r)_h} + \sum_{h\in[s-g]} C(x)_{A(j';r')_h} - \sum_{h\in[s-g]} C(x)_{B(j';r')_h} - C(x)_{c(j';r')} + C(x)_{j'} \\
&- \sum_{h\in[s-g]} C(x)_{A(j';r')_h} + C(x)_{c(j';r')} \\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[g]} C(x)_{A(j;r)_h} - \sum_{h\in[s-g]} C(x)_{A(j';r')_h} + C(x)_{c(j';r')} \\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[g]} C(x)_{A(j;r)_h} - \sum_{h\in\{s-g\}} C(x)_{A(j;r)_{g+h}} + C(x)_{c(j;r)} \\
&= \sum_{h\in[s]} C(x)_{B(j;r)_h} - \sum_{h\in[s]} C(x)_{A(j;r)_h} + C(x)_{c(j;r)} \\
&= C(x)_j.
\end{aligned}
$$

□

We also observe the following direct implication.

Claim C.7. *Let $u = (0, L_1, \dots, L_s, c) \in U$ and $v = (1, L'_1, \dots, L'_s, c') \in U$ be such that $(L_1, \dots, L_s, c) \sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)} (L'_1, \dots, L'_s)$ where the $\sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)}$ relation is as defined in Section 3.3 in Definition 3.12. Then $C'(x)_u + C'(x)_v = C(x)_j$ for every x.*

Proof. We argue that

$$C'(x)_u + C'(x)_v = \sum_{h\in[s]} C(x)_{B^g(j;r)_h} - \sum_{h\in[s]} C(x)_{A^g(j;r;z)_h} + C(x)_{c^g(j;r;z)}.$$

To see this, denote $A^g(j; r; z) = (a_1, \dots, a_s)$ and $B^g(j; r) = (b_1, \dots, b_s)$. By the assumption $(L_1, \dots, L_s, c) \sim_{A^g(j;r;z), B^g(j;r), c^g(j;r;z)} (L'_1, \dots, L'_s)$ we have that $c = c^g(j; r; z)$ and for every $h \in [s]$

$$C(x)_{L'_h} - C(x)_{L_h} = C(x)_{L'_h \setminus L_h} - C(x)_{L_h \setminus L'_h} = C(x)_{b_h} - C(x)_{a_h}.$$

Therefore, by the definition of $C'(x)_u$ and $C'(x)_v$,

$$C'(x)_u + C'(x)_v = c - \sum_{h\in[s]} C(x)_{L_h} + \sum_{h\in[s]} C(x)_{L'_h} = \sum_{h\in[s]} (b_h - a_h) + c.$$

The claim now follows by applying Claim 3.25. □

For every $j \in [n]$ let $\tilde{G}_j$ be the bipartite graph defined in Section 3.3. Note that $|U| = |V| = N$ and there is a natural isomorphism between the vertices of $\tilde{G}_j$ and the coordinates of C', given by that every left side vertex $L \in \binom{[n]}{\ell}^s \times [n]$ corresponds to the coordinate $u = (0, L) \in U$, and every right vertex (i, L') for $i \in [n]$ and $L' \in \binom{[n]}{\ell}^s$ corresponds to the coordinate $v = (1, L', i) \in V$. By the previous claim, for every edge $(L, (i, L'))$ in $\tilde{G}_j$ and its corresponding coordinates (u, v), we have that $C'(x)_u + C'(x)_v = C(x)_j$ for every $x \in \mathbb{F}^k$.

We can now apply Lemma 3.24 and conclude that for every $j \in [n]$ there is a set of

$$\Omega\left(\frac{1}{\left(1+\frac{n}{m\ell}\right)^s}\frac{m}{sn}N\right) = \Omega\left(\frac{1}{\left(1+\frac{1}{\delta\ell}\right)^s}\frac{\delta}{s}N\right)$$

disjoint pairs of coordinates $\{u, v\}$ of C' such that $C'(x)_u + C'(x)_v = C(x)_j \ \forall x \in \mathbb{F}^k$.

We can now conclude the bound.

Proposition C.8. *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a special-form $(3, \delta)$-LCC, and assume that $|\mathbb{F}| \leq n$. Then $k = O\left(\frac{1}{\delta^2}\log^4 n\right)$.*

Proof. The proof is the same as the proof for Theorem 3.27, except that instead of invoking Theorem 2.2 we apply Theorem C.2. Combined with the assumption $|\mathbb{F}| \leq n$, and therefore in particular $\log(|\mathbb{F}|) = O(\log N)$ (as $n \leq N$), we still get that

$$k = O\left(\frac{1}{\delta_{C'}}\log N\right)$$

in this case as well. The rest of the proof is identical. □

C.0.3 From LCC to special-form LCC

Claim C.9. *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a linear $(3, \delta, \varepsilon)$-LCC. Then, there exists a code $C' : \mathbb{F}^k \to \mathbb{F}^{n'}$ which is a special-form $(3, \delta')$-LCC, for $n' = O(|\mathbb{F}|n)$ and $\delta' = \Omega(\frac{\delta}{|\mathbb{F}|})$.*

Proof. Set $m = \frac{\delta n}{3}$. By Fact C.4 there exist sets $Q_r^j \mid_{j \in [n], r \in [m]}$ of size at most 3, such that for every $j \in [n]$, the m sets $Q_r^j \mid_{r \in [m]}$ are disjoint, and every Q_r^j determines j in C. Without loss of generality the sets $Q_r^j \mid_{j \in [n], r \in [m]}$ are of size exactly 3 (we can add zero coordinates, at worst doubling n). Denote $Q_r^j = \{A(j; r), B(j; r), c(j; r)\}$ and let $\{\alpha_{r,1}^j, \alpha_{r,2}^j, \alpha_{r,3}^j \in \mathbb{F} \setminus \{0\}\}_{j \in [n], r \in [m]}$ be such that for every j, r

$$C(x)_j = \alpha_{r,1}^j C(x)_{A(j;r)} + \alpha_{r,2}^j C(x)_{B(j;r)} + \alpha_{r,3}^j C(x)_{c(j;r)} \ \forall x \in \mathbb{F}^k.$$

Set $N' = [n] \times (\mathbb{F} \setminus \{0\})$ and we define $C' : \mathbb{F}^k \to \mathbb{F}^{N'}$ as follows. For every $j \in [n]$ and $\beta \in \mathbb{F} \setminus \{0\}$

$$C'(x)_{j,\beta} = \beta C(x)_j.$$

We argue that C' is a special-form $(3,\delta')$-LCC for $\delta' = \frac{\delta}{|\mathbb{F}|-1}$, of length $n' = (|\mathbb{F}|-1)n$. Indeed, for every $j \in [n]$ and $\beta \in \mathbb{F} \setminus \{0\}$, we consider the sets $Q_r^j |_{r \in [m]}$. For every $r \in [m]$ we have that

$$\begin{aligned} C'(x)_{j,\beta} &= \beta C(x)_j \\ &= \beta\alpha_{r,1}^j C(x)_{A(j;r)} + \beta\alpha_{r,2}^j C(x)_{B(j;r)} + \beta\alpha_{r,3}^j C(x)_{c(j;r)} \\ &= -C'(x)_{A(j;r),-\beta\alpha_{r,1}^j} + C'(x)_{B(j;r),\beta\alpha_{r,2}^j} + C'(x)_{c(j;r),\beta\alpha_{r,3}^j} \ \forall x \in \mathbb{F}^k. \end{aligned}$$

Thus, we define $Q_r'^{j,\beta} = \{A'(j,\beta;r), B'(j,\beta;r), c'(j,\beta;r)\}$ for $A'(j,\beta;r) = (A(j;r), -\beta\alpha_{r,1}^j) \in N'$, $B'(j,\beta;r) = (B(j;r), \beta\alpha_{r,2}^j) \in N'$ and $c'(j,\beta;r) = (c(j;r), \beta\alpha_{r,3}^j) \in N'$ in order to satisfy Equation (C.1) of Definition C.5. It remains to check that the sets $Q_r'^{j,\beta} |_{r \in [m]}$ are disjoint. Indeed, this follows immediately from that $Q_r^j |_{r \in [m]}$ are disjoint (if we consider only the first part of the name of each coordinate in $Q_r'^{j,\beta}$ we will get disjoint sets). Finally, for every j, β there are $\frac{\delta n}{3}$ sets $\{Q_r'^{j,\beta}\}$ and $\frac{\delta n}{3} = \frac{\delta' |N'|}{3}$ for $\delta' = \frac{\delta}{|\mathbb{F}|-1}$. The claim follows. □

We can now conclude the bound for a general $\mathbb{F}$.

Theorem C.10 (Theorem 1.7, rephrased)**.** *Let $C : \mathbb{F}^k \to \mathbb{F}^n$ be a linear $(3,\delta,\varepsilon)$-LCC. Then $k = O\left(\frac{|\mathbb{F}|^2}{\delta^2} \log^4 n\right)$.*

Proof. By Claim C.9 there exists a linear map $C' : \mathbb{F}^k \to \mathbb{F}^{n'}$ which is a special-form $(3,\delta')$-LCC, for $n' = O(|\mathbb{F}|n)$ and $\delta' = \Omega(\frac{\delta}{|\mathbb{F}|})$. Since if $|\mathbb{F}| > k$ the argued bound holds trivially, we can assume $|\mathbb{F}| \le k$. In particular, $|\mathbb{F}| \le n \le n'$, and so we can invoke Proposition C.8 with regards to C'. We get that

$$k = O\left(\frac{1}{\delta'^2} \log^4 n'\right) = O\left(\frac{|\mathbb{F}|^2}{\delta^2} (\log(n) + \log(|\mathbb{F}|)^4)\right) = O\left(\frac{|\mathbb{F}|^2}{\delta^2} \log^4 n\right),$$

as required. □

D Proof sketch for Fact 2.3 and Fact C.4

For completeness we give in this part a proof sketch for Fact 2.3 and Fact C.4. See also [KT00, Yek11, ZD].

Definition D.1. *We say that a randomized procedure A that gets oracle access to $z \in \mathbb{F}^n$ and makes queries to z is* non-adaptive *if the distribution of the set of coordinates that it queries does not depend on z.*

Fact D.2. *Let $C : \mathbb{F}^k \to \mathbb{F}^k$ and $\ell : \mathbb{F}^k \to \mathbb{F}$ be two linear maps. Further let $Q \subseteq [n]$ and $f \in \mathbb{F}^Q$. Then, one of the following cases must hold.*

1. *There is at most one $\alpha \in \mathbb{F}$ for which there exists some $x \in \mathbb{F}^k$ satisfying $C(x)_Q = f$ and $\ell(x) = \alpha$.*
2. *For every $\alpha \in \mathbb{F}$ there is an equal number of $x \in \mathbb{F}^k$ for which $C(x)_Q = f$ and $\ell(x) = \alpha$.*

In particular, either no function (even randomized) of $C(x)_Q$ can predict $\ell(x)$ with probability larger than $\frac{1}{|\mathbb{F}|}$, when $x \in \mathbb{F}^k$ is randomly chosen uniformly, or $C(x)_Q$ determines $\ell(x)$ for all $x \in \mathbb{F}^k$. Also in particular, in the second case, there is a linear *function from $C(x)_Q$ to $\ell(x)$, and thus a vector $v \in \mathbb{F}^Q$ such that $\ell(x) = \sum_{h \in Q} v_h C(x)_h$.*

It is enough to give a proof sketch for Fact C.4 since Fact 2.3 is a special case of it.

Proof sketch for Fact C.4. Let $j \in [n]$, and let A_j be a randomized procedure that satisfies the requirements of Definition C.1, with an input coordinate fixed to j. Without loss of generality A_j makes exactly q queries always (we can add queries and ignore them).

Since we only require that $\varepsilon < 1 - \frac{1}{|\mathbb{F}|}$, we can assume without loss of generality that A_j is non-adaptive. Indeed, consider the algorithm A'_j which first guesses q values $h_1, \ldots, h_q \in \mathbb{F}$, and then starts to simulate A_j. At each time $t \in [q]$ that A_j makes a query, A'_j queries that coordinate of z and checks if its value is equal to h_t. If it is different, A'_j halts the simulation of A_j and outputs a random value in $\mathbb{F}$. Otherwise, if all guesses turned up to be correct, it answers like A_j. A'_j is non-adaptive because it can make all queries in advance. The success probability of A'_j is larger than $\frac{1}{|\mathbb{F}|}$: there is a positive chance that A'_j guesses correctly the queries that A_j is going to make, and conditioned on that, the success probability is $1 - \varepsilon > \frac{1}{|\mathbb{F}|}$; if, on the other hand, the guesses of A'_j are incorrect, the success probability is $\frac{1}{|\mathbb{F}|}$. Therefore the assumption that A_j is non-adaptive is indeed without loss of generality.

We start by showing that there exists one set $Q \in \binom{[n]}{q}$ such that Q determines j in C. Let X be a uniformly random message in $\mathbb{F}^k$. We have that

$$\mathbf{Pr}[A_j(C(X)) = C(X)_j] \geq 1 - \varepsilon$$

where the probability is over the sampling of X and over the randomness of A_j, since for every fixed $x \in \mathbb{F}^k$ the above holds. Moreover,

$$\begin{aligned}&\mathbf{Pr}[A_j(C(X)) = C(X)_j] \\ &= \sum_{Q \in \binom{[n]}{q}} \mathbf{Pr}[A_j(C(X)) \text{ queries on } Q] \cdot \mathbf{Pr}[A_j(C(X)) = C(X)_j \mid A_j(C(X)) \text{ queries on } Q].\end{aligned}$$

Hence there exists some fixed $Q \in \binom{[n]}{q}$ such that

$$\mathbf{Pr}[A_j(C(X)) = C(X)_j \mid A_j(C(X)) \text{ queries on } Q] \geq 1 - \varepsilon > \frac{1}{|\mathbb{F}|}.$$

Since A_j is non-adaptive, there is a randomized procedure $A_j^Q : \mathbb{F}^Q \to \mathbb{F}$ such that

$$\mathbf{Pr}[A_j(C(X)) = C(X)_j \mid A_j(C(X)) \text{ queries on } Q] = \mathbf{Pr}[A_j^Q(C(X)_Q) = C(X)_j].$$

Hence, by Fact D.2, Q determines j in C (with the function ℓ being C's j-th coordinate). To see that there is another set, Q', disjoint from Q, which also determines j in C, consider the code $C' : \mathbb{F}^k \to \mathbb{F}^n$ defined by $C'(x)_t = C(x)_t$ for $t \notin Q$ and $C'(x)_t = 0$ for $t \in Q$. It is not hard to see, following the same reasoning, that if $|Q| \leq \delta n$, then there exists a set Q' and a procedure $A_j^{Q'}$ such that $\mathbf{Pr}[A_j^{Q'}(C'(X)_{Q'}) = C(X)_j] > \frac{1}{|\mathbb{F}|}$, and without loss of generality $Q' \cap Q = \emptyset$ (as $C'(X)_Q$ is known to be 0). Hence again by Fact D.2 Q' is another set, disjoint from Q, that determines j in C. We can continue in this manner as long as the size of the union of the sets is at most δn, and hence there are at least $\frac{\delta n}{q}$ disjoint sets for every $j \in [n]$. □