Electronic Colloquium on Computational Complexity, Report No. 168 (2025)

Asymptotically good large-alphabet LDCs with polylogarithmic query complexity

Tal Yankovitz*

Abstract

A large alphabet Locally Decodable Code (LDC) $C : \Sigma^k \to \Sigma'^n$, where Σ' may be large, is a code where each symbol of x can be decoded by making few queries to a noisy version of $C(x)$. The rate of C is its information rate, namely $\frac{k\log(|\Sigma|)}{n\log(|\Sigma'|)}$. We construct the first constant-rate large alphabet LDC C making a polylogarithmic number of queries (in k and n), while satisfying $\log|\Sigma'| \leqslant k^{\varepsilon}$ for any chosen constant $\varepsilon < 1$. We add that in fact we show a code with a property stronger than being a large alphabet LDC, which we dub *block-wise Locally Correctable Code (block-wise LCC)*, implying LDC.

Our construction is a variant of multivariate Multiplicity codes which were introduced in the seminal work of Kopparty, Saraf and Yekhanin (STOC '11). However we remark that our definition of the code and its analysis are taking a somewhat different approach, considering specific linear relations that are required for our purposes. While the resulting rate is akin to the one obtained through standard multiplicity codes analysis, this dual-based analysis extends to other families of linear-constraint codes of the same flavor and may be of independent technical interest.

To get the polylogarithmic query complexity we observe a correction process for which very few random lines suffice in order to correct an element, as opposed to an exponential number of lines as is usually required in decoding Multiplicity codes. This seems to be the first non-trivial case where the lower-bound for LDC due to Katz and Trevisan (STOC '00), which in particular implies that for constant rate the number of queries is at least logarithmic in the code's length, is close to tight.

*UT Austin. talyanko@utexas.edu. Supported by NSF Grant CCF-2312573.

ISSN 1433-8092

Contents

1 Introduction **1**
1.1 Our main contribution 2
1.2 Technical overview, and second contribution 3
1.2.1 Setting the ground: a naive attempt 3
1.2.2 The line-constraints subspace 4
1.2.3 Our second contribution - bounding the dimension of the line-constraints subspace 5
1.2.4 On bounding the dimension of the line-constraints subspace 6
1.2.5 From the line-constraints subspace to Theorem 1.4 7

2 Preliminaries **10**
2.1 Notation. 10
2.2 Facts. 10

3 The line constraints subspace **11**
3.1 The desired functions and the bound on their dimension 11
3.2 Expressing the desired functions 12
3.3 From a span of functions in t, x, H to a span of polynomials in a, b 13
3.4 Bounding the span of polynomials over the Reals 15
3.5 Concluding Theorem 3.2 18

4 Good blockwise LCCs with polylog query complexity **19**
4.1 High-rate blockwise LCCs 19
4.2 Applying the AEL distance amplification to get asymptotically good blockwise LCCs 23

1 Introduction

Locally decodable codes were first defined by Katz and Trevisan [KT00]. They, in particular, allow for sublinear decoding algorithms in the case that a part of data is required.

Definition 1.1 (LDC)**.** *$C : \Sigma^k \to \Sigma'^n$ is a (q, δ)-LDC* (locally decodable code, abbreviated) *if there exists a randomized procedure* $\mathsf{Dec} : [k] \to \Sigma$ *that is given an oracle access to $z \in \Sigma'^n$ and has the following guarantee. For every $i \in [k]$, $x \in \Sigma^k$ and $z \in \Sigma'^n$ satisfying* $\mathsf{HammingDistance}(z, C(x)) < \delta n$, $\mathsf{Dec}^z(i) = x_i$ *with probability at least $\frac{2}{3}$. Furthermore* $\mathsf{Dec}^z(i)$ *always makes at most q queries to z.*

Their study, and the study of the closely related *locally correctable codes*, has attracted substantial attention. For a comprehensive exposition, the reader may consult the excellent survey of Yekhanin [Yek11]. Locally decodable codes have abundant applications, including in error correcting codes, complexity theory, PCPs, error reduction, cryptography, hardness amplification, data structures, and more.

A central question in the area of locally decodable codes is the optimal tradeoff between the information rate of the code, $\frac{k \log |\Sigma|}{n \log |\Sigma'|}$ and the number of needed queries q, and a rich line of work has been dedicated to shedding light on this question, yet much has remained unknown.

Katz and Trevisan [KT00] have proved that[1]

$$n \geqslant \left(\frac{1}{6} \cdot \delta\right)^{\frac{1}{q-1}} \cdot \left(\frac{1}{q^2}\right)^{\frac{1}{q-1}} \cdot \left(\frac{2}{3} \cdot \frac{k \cdot \lfloor \log |\Sigma| \rfloor}{\log |\Sigma'|}\right)^{1+\frac{1}{q-1}}. \tag{1.1}$$

In particular, whenever $\delta = \Omega(1)$, if $q = O(1)$[2] then $n = \left(\frac{k \log |\Sigma|}{\log |\Sigma'|}\right)^{1+\Omega(1)}$, and if the information rate is constant, i.e., $n = O\left(\frac{k \log |\Sigma|}{\log |\Sigma'|}\right)$, then $q = \Omega(\log n)$.

As for constructions, in the case that $q = O(1)$ is needed, sub-exponential constructions are known due to Yekhanin and Efremenko [Yek08, Efr09], the state of the art giving binary codes of a length n which is exponential in $2^{\log^\varepsilon k}$ (and thus subexponential in k) [Efr09]. In the other regime which aims for constant rate, after several works [GKS13, KSY14, HOW15] improved on the rate of codes with polynomial query complexity $q = n^\varepsilon$, Kopparty, Meir, Ron-Zewi and Saraf [KMRS17] achieved high-rate binary codes with $q = 2^{O\left(\sqrt{\log(n) \log\log(n)}\right)}$.

[1] We remark that in fact they state their bound for the case that the input alphabet $\Sigma = \{0, 1\}$, but it is easy to see that it extends to the case of any Σ, by choosing an injective mapping $\{0, 1\}^{\lfloor \log |\Sigma| \rfloor} \to \Sigma$.

[2] The case $q = 1$ is handled separately in [KT00], where it is shown that it is impossible to have $q = 1$ with a nontrivial code alphabet Σ'.

Significant work was also put in attempt to improve the Katz-Trevisan bound. In the case $q = 2$ tight exponential bounds were provided [Gol11, KDW03, DS05, BGT16]. For larger constant q's the work of [KDW03] gave a polynomial improvement in Equation (1.1) for small alphabets, with improvements in [Woo07, AGKM23, BHKL25, JM25].[3] We add that in the constant rate regime no improvements upon the $q = \Omega(\log n)$ that follows from Equation (1.1) were discovered.

1.1 Our main contribution

In this work we construct the first constant-rate polylogarithmic query locally decodable code with non-trivial code alphabet. Prior to this work, the best known locally decodable code with $\log|\Sigma'| = O(n^{\varepsilon})$ for any constant $\varepsilon < 1$ had an inverse polynomial rate [BFLS91].

In fact a stronger object is achieved, which we now define. Since our constructed code is going to be linear, we define it over a finite field $\mathbb{F}$.

Definition 1.2 ((q, δ)-blockwise LCC)**.** *For two sets P, S where $|P| = n$, a code $C \subseteq \mathbb{F}^{P\times S}$ is a* (q, δ)-blockwise LCC *if there exists a randomized procedure* $\mathsf{Cor} : P \times S \to \mathbb{F}$ *that is given an oracle access to $z \in \mathbb{F}^{P\times S}$ and has the following guarantee. For every $(p, a) \in P \times S$, $c \in C$ and $z \in \mathbb{F}^{P\times S}$ such that $|\{p \in P \mid z(p, \cdot) \neq c(p, \cdot)\}| < \delta n$, $\mathsf{Cor}^z(p, a) = c(p, a)$ with probability at least $\frac{2}{3}$. Furthermore, $\mathsf{Cor}^z(p, a)$ always makes at most q queries to z, where* each query of Cor consists of obtaining $c(p', \cdot)$ for some $p' \in P$.

We think of the set P as a set of n *points* where "on" each point $p \in P$ there is a *block* $(c(p, a))_{a\in S}$ which is of size $|S|$; the overall length will usually be denoted by $N := n|S|$. We proceed to make the following remark.

Remark 1.3 (From linear blockwise LCC, to LDC)**.** *If $C \subseteq \mathbb{F}^{P\times S}$ is a $\mathbb{F}$-vector space of dimension k which is a (q, δ)-blockwise LCC then C induces $\widetilde{C} : \mathbb{F}^k \to (\mathbb{F}^S)^P$ which is a (q, δ)-LDC. Indeed, we can choose any systematic mapping $C' : \mathbb{F}^k \to C$ where by systematic we mean that the symbols of x are embedded in the symbols of $C'(x)$. Since we can correct the symbols of $C'(x)$, we can decode the symbols of x. We thus take $\widetilde{C} : \mathbb{F}^k \to (\mathbb{F}^S)^P$ to be such that for $p \in P$, $\widetilde{C}(x)(p) = C'(x)(p, \cdot)$.*

Our main result is the following.

Theorem 1.4. *For every $\sigma \geqslant 4$ the following holds. For every $n' \in \mathbb{N}$ there exists $n \geqslant n'$ for which the following holds. There is a linear code $C \subseteq \mathbb{F}_q^{P\times S}$, where $q = O(\log^{\sigma+3}(n))$,*

[3] Also, for bounds on 3-query locally correctable codes see [KM23] and followups [Yan24, AG24, KM24].

$|P| = n$ *and* $|S| \leqslant n^{3/\sigma}$*, with rate* $\rho = \frac{\dim_{\mathbb{F}_q}(C)}{n|S|} = \Omega(1)$*, which is a* $(\log^{3\sigma+9}(n), \Omega(1))$*-blockwise LCC.*

One interesting conclusion from Theorem 1.4 is that attempts to significantly strengthen the Katz-Trevisan bound should rely on the code alphabet being small.

Remarks. We did not optimize the polylogarithmic factor in the query complexity; we chose a simpler exposition, and a more careful analysis should further reduce this factor. Secondly, while we do not highlight it in the paper an explicit construction of C follows naturally.

We turn to give a technical overview of the construction and analysis. In Section 1.2.3 we describe a technical contribution.

1.2 Technical overview, and second contribution

We construct a code which we view as a variant of multivariate multiplicity codes which were introduced in the most influential work of Kopparty, Saraf and Yekhanin [KSY14]. However, we take a somewhat different approach in defining the code and in the analysis (we will not explicitly mention derivatives). Instead of considering an encoding (of polynomials into evaluations of their derivatives) we define a set of linear constraints, sufficient for local correction, and prove an upper bound on the dimension of the linear subspace spanned by these constraints. After giving the technical details we address the connection to (normally defined) multiplicity codes, in Remark 1.7.

1.2.1 Setting the ground: a naive attempt

It is well known that a Reed-Muller code $C_{\mathsf{RM}} \subseteq \mathbb{F}_q^{\mathbb{F}_q^m}$ consisted of the evaluations of m-variate polynomials in $\mathbb{F}_q[x_1, \ldots, x_m]$ of total degree at most d - while possessing wanted local-correction features - are of rapidly vanishing rate whenever $m = \omega(1)$. These correction features stem from the dual code $C_{\mathsf{RM}}^{\perp}$ which contains linear constraints $\ell(x_1, \ldots, x_m) \in \mathbb{F}_q^{\mathbb{F}_q^m}$, in particular constraints supported on lines of $\mathbb{F}_q^m$ (that is, they are 0 outside of the line), giving rise to equations

$$\sum_{\alpha_1, \ldots, \alpha_m \in \mathbb{F}_q} \ell(\alpha_1, \ldots, \alpha_m) \cdot c(\alpha_1, \ldots, \alpha_m) = 0$$

which hold for every $c \in C_{\mathsf{RM}}$. A very naive attempt at increasing the rate of the code while preserving the wanted correction features is to define a new code C' over *copies* of the coordinate-sets of C_{RM}, say s copies indexed by $h \in [s]$, while keeping the same constraints.

That is - for every $\ell(x_1, \ldots, x_m) \in C_{\mathsf{RM}}^{\perp}$ that was of need for the local correction, we take $\ell'(x_1, \ldots, x_m, h) = \ell(x_1, \ldots, x_m)$ to be in the space orthogonal to C'. Since the code length increased from $n = q^m$ to $s \cdot n$ while the co-dimension remained as before, in particular, at most n, the rate of C' is at least $1 - \frac{1}{s}$. However, this rate seems too good to be useful and indeed it is, since by ignoring the copy number h in our added constraints, we made the constraints of the code only dependent on the sum of the copies. That is,

$$\sum_{h \in [s]} \sum_{\alpha_1, \ldots, \alpha_m \in \mathbb{F}_q} \ell(\alpha_1, \ldots, \alpha_m, h) \cdot c(\alpha_1, \ldots, \alpha_m, h) = 0$$

is what we have for $c \in C'$, and thus we cannot ever correct a specific coordinate, rather only the sum of its "copies". However, if we could make a more clever choice for our $\ell'(x_1, \ldots, x_m, h)$ – one which *does* depend on h, hopefully while still keeping the dimension required in order to span all these constraints more close to n than to $s \cdot n$, then possibly we would gain something. This is going to be what we aim towards doing, as we explain next.

1.2.2 The line-constraints subspace

Continuing the approach of the previous discussion we will construct a subspace of low-weight constraints, adding "copies" of the coordinate-set $\mathbb{F}_q^m$, with the choice that each copy will be indexed by $H \in \mathcal{H}$ where $\mathcal{H} \subseteq (\mathbb{N} \cup \{0\})^m$. That is, the constraints, and the induced code, are subspaces of $\mathbb{F}_q^{\mathbb{F}_q^m \times \mathcal{H}}$.

Now, some technical details. First, to ignore sign $+1$ or -1 nuances in this informal overview we will assume that $\mathbb{F}_q$ is of characteristic 2. Second, notation wise, we will write x as short for $(x_1, \ldots, x_m)$, and x^I as short for $\prod_{i \in [m]} x_i^{I_i}$. Third, it will be convenient for us to have another designated variable - which we will denote by t - and we will only consider *ordered* lines which are indexed by t. That is, our space is $\mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}}$, and we will consider all lines corresponding to direction $a \in \mathbb{F}_q^m$ and offset $b \in \mathbb{F}_q^m$: the set of points $\{(\tau, a_1\tau + b_1, \ldots, a_m\tau + b_m) \mid \tau \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^{m+1}$.

We will define the following constraints. For every $\tau \in \mathbb{F}_q$, $\alpha \in \mathbb{F}_q^m$ and $H \in \mathcal{H}$

$$L^{a,b}(\tau, \alpha, H) = \begin{cases} a^H & \text{if } \alpha = a\tau + b \\ 0 & \text{otherwise} \end{cases} \qquad \in \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}} \tag{1.2}$$

We will thus take our constraints subspace to be

$$\mathcal{L} = \mathrm{Span}\{L^{a,b} \mid a, b \in \mathbb{F}_q^m\}.$$

We call $\mathcal{L}$ *the line-constraints subspace*, and we see that the defined constraints do depend on the copy H.

Two questions arise: is $\mathcal{L}$ useful for local correction, and what can we say on its dimension, especially what's its dependence on the number of copies $|\mathcal{H}|$. We first discuss the second question, in our analysis we show that $\dim(\mathcal{L})$ can be related to the structure of the set of copies $\mathcal{H}$. After that, we will discuss the first question.

1.2.3 Our second contribution - bounding the dimension of the line-constraints subspace

Recall that $\mathcal{L} \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}}$, and we define $N := |\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}| = q^{m+1}|\mathcal{H}|$. We now make a definition regarding the structure of $\mathcal{H}$ which we show is key in the bound on the dimension of $\mathcal{L}$.

Definition 1.5.

$$\mathit{Boundary}(\mathcal{H}) := \{H \in \mathcal{H} \mid \exists i \in [m] : H + e_i \notin \mathcal{H}\},$$

where e_i is the i-th unit vector.

With the definition of the boundary of $\mathcal{H}$ we can present our second contribution, which is a bound on the dimension of the line-constraints subspace, related to $\text{Boundary}(\mathcal{H})$.

Theorem 1.6. *For $m = o(q)$,*

$$\dim(\mathcal{L}) \leqslant N \cdot m \cdot |\mathit{Boundary}(\mathcal{H})| + o(N).$$

That is, while we added $|\mathcal{H}|$ "copies" of the coordinates to our code, we only paid for that in dimension proportional to $m \cdot |\text{Boundary}(\mathcal{H})|$, so whenever $|\text{Boundary}(\mathcal{H})| \ll \frac{1}{m}|\mathcal{H}|$, we profit.

Before overviewing the elements of the proof for Theorem 1.6, we pause to discuss instantiations of it.

Instantiations of Theorem 1.6. One natural choice for the set $\mathcal{H}$, for a parameter $s \in \mathbb{N}$, is $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s-1\}$, where $|H| := \sum_{i=1}^m H_i$, and note that $|\mathcal{H}| = \binom{m+s-1}{m}$. In fact, this choice corresponds to the normally defined multiplicity codes where the encoding outputs evaluations of derivatives up-to order $s-1$. For this choice, $\text{Boundary}(\mathcal{H}) = \{H \in \mathcal{H} \mid |H| = s-1\}$, which is of size $\binom{m+s-2}{m-1}$. Using that $\frac{i}{j}\binom{i-1}{j-1} = \binom{i}{j}$, we see that

$$\frac{|\text{Boundary}(\mathcal{H})|}{|\mathcal{H}|} = \frac{m}{m+s-1},$$

and thus defining a code by taking $C = \mathcal{L}^\perp$ with this choice for $\mathcal{H}$, results by Theorem 1.6 in a code with rate $1 - \frac{m^2}{m+s-1} - o(1)$, or more precisely $1 - \frac{m^2}{m+s-1} - \frac{m+1}{q}$, using the more detailed bound from the technical section. We remark that this bound on the rate is quite similar to the bound $(1-\frac{m^2}{s})(1-\frac{2}{q})^m$ on the rate of multiplicity codes which follows from the rate bound in [KSY14][4].

However, there are also other possible choices for $\mathcal{H}$. Another possible example is taking $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid H \leqslant \overline{s-1}\}$, where $\overline{s-1}$ denotes $(s-1,\dots,s-1)$, and by $\leqslant$ we mean that the inequality holds at every *individual* entry. We remark, without getting into the details, that whenever $s \leqslant q$ this choice also allows correction. In this case

$$\frac{|\mathrm{Boundary}(\mathcal{H})|}{|\mathcal{H}|} = \frac{s^m-(s-1)^m}{s^m} = 1-\left(1-\frac{1}{s}\right)^m \leqslant \frac{m}{s},$$

and thus by Theorem 1.6 for this choice $C = \mathcal{L}^\perp$ would have a similar rate as the previous option. This example does not seem to be equivalent to multiplicity codes, and it seems interesting to wonder what different options for $\mathcal{H}$ can give with respect to local correction, where the choice of $\mathcal{H}$ does matter (specifically to get the low-query of Theorem 1.4 we will need the firstly discussed, multiplicity-like $\mathcal{H}$).

1.2.4 On bounding the dimension of the line-constraints subspace

We now turn to give a technical overview on the proof for Theorem 1.6. It turns out that we can algebraically express the line-functions defined in Equation (1.2) above by relying only on Fermat's Little Theorem. We define the following (H-dependent function times a) polynomial for every $a, b \in \mathbb{F}_q^m$:

$$L^{a,b}(t,x,H) = a^H \prod_{i\in[m]} (1-(x_i+a_it+b_i)^{q-1}). \tag{1.3}$$

and it is an easy check that for every $\tau \in \mathbb{F}_q$ and $\alpha \in \mathbb{F}_q^m$, $L^{a,b}(\tau,\alpha,H)$ evaluates exactly to our wanted function. This is useful when we look for a small basis for $\mathcal{L}$. If we open up the product in Equation (1.3), then we can see (the full details in Section 3) that in the case that $m = o(q)$, the challenge boils down to bounding the dimension of the span of functions

$$\widetilde{L}^{a,b}(t,x,H) = a^H \prod_{i\in[m]} (x_i+a_it+b_i)^{q-1},$$

i.e., those containing the "heavy", degree $m(q-1)$, product. We will thus denote $\widetilde{\mathcal{L}} = \mathrm{Span}\{\widetilde{L}^{a,b} \mid a,b \in \mathbb{F}_q^m\}$ and turn our focus to bounding its dimension since it will dominate the dimension of $\mathcal{L}$.

[4]When choosing the maximal degree of the evaluated polynomial to be $d = s(q-1)-1$ to give a comparable setting to ours.

Now, one can check that

$$\widetilde{L}^{a,b}(t,x,H) = \sum_{\bar{0} \leqslant I \leqslant \overline{q-1}} \sum_{j=0}^{|I|} \underbrace{\left(\sum_{\substack{J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \right)}_{:=D^{a,b}(j,I,H)} \cdot \binom{\overline{q-1}}{I} t^j x^{\overline{q-1}-I}.$$

where $I, J \in (\mathbb{N} \cup \{0\})^m$, $\bar{0} := \underbrace{(0,\ldots,0)}_{m \text{ times}}$, $\overline{q-1} := \underbrace{(q-1,\ldots,q-1)}_{m \text{ times}}$, $|I| := \sum_{r=1}^m I_r$, $J \leqslant I \iff J_1 \leqslant I_1 \wedge \cdots \wedge J_m \leqslant I_m$, $\binom{I}{J} := \prod_{r=1}^m \binom{I_r}{J_r}$ and $x^I := \prod_{r=1}^m x_r^{I_r}$.

The next step is to consider the defined above $D^{a,b}(j,I,H)$ for every $I \leqslant \overline{q-1}$, $j \leqslant |I|$ and $H \in \mathcal{H}$, and we will inspect them as *polynomials in* a, b – i.e., while $\widetilde{L}^{a,b}$ is a function of t, x and H, where a and b are some fixed elements of $\mathbb{F}_q^m$, we will analyze the family of polynomials $D^{a,b}(j,I,H) \in \mathbb{F}_q[a,b]$ *defined according to all possible* I, j, H. Doing so, we define $\mathcal{D} = \mathrm{Span}\{D^{a,b}(j,I,H) \mid I \leqslant \overline{q-1}, j \leqslant |I|, H \in \mathcal{H}\} \subseteq \mathbb{F}_q[a,b]$, and in Section 3 we prove that $\dim \widetilde{\mathcal{L}} \leqslant \dim \mathcal{D}$, so it turns out that it suffices to consider these polynomials. In fact, we bound the dimension of $\mathrm{Span}\{D^{a,b}(j,I,H)\}$ as polynomials *over the reals*, which suffices in order to bound $\dim \mathcal{D}$.

It may look daunting to analyze the dimension of $\mathrm{Span}\left\{\sum_{J \leqslant I, |J|=j} \binom{I}{J} a^{H+J} b^{I-J} \mid j, I, H\right\}$ as polynomials in a, b since the coefficients are specific sums of m-wise products of binomial coefficients. However, it turns out that all is needed in order to do so is the fact that $\frac{i}{j}\binom{i-1}{j-1} = \binom{i}{j}$[5]. Using this fact, we show in Section 3 that for any $j > 0$

$$jD^{a,b}(j,I,H) = \sum_{r \in [m] | I_r > 0} I_r D^{a,b}(j-1, I-e_r, H+e_r), \tag{1.4}$$

In particular, in order to span the entire space, it suffices to take a set which consists of $\{D^{a,b}(0,I,H) \mid I \leqslant \overline{q-1}, H \in \mathcal{H}\}$ (which is a small set when $q = \omega(1)$ since fixing j to 0 corresponds the size being divided by q), and of $\{D^{a,b}(j,I,H) \mid I \leqslant \overline{q-1}, j \leqslant |I|, H \in \mathrm{Boundary}(\mathcal{H})\}$ since $\mathrm{Boundary}(\mathcal{H})$ consists exactly of the H's where we can't apply Equation (1.4) in order to span them using "higher" H's. This is enough to deduce Theorem 1.6, for the full statement and proofs see Section 3.

1.2.5 From the line-constraints subspace to Theorem 1.4

As mentioned above, we will instantiate Theorem 1.6 with $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s-1\}$, and take $C = \mathcal{L}^\perp \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}} = \mathbb{F}_q^{P \times \mathcal{H}}$ where $P = \mathbb{F}_q \times \mathbb{F}_q^m$ and $n = |P|$. Assume that we wish to correct a coordinate $(\tau^*, \alpha^*, H^*) \in \mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}$, that is, to recover $c(\tau^*, \alpha^*, H^*)$ amid some $c \in C$.

[5] Which was already used one time in the discussion following Theorem 1.6.

It will be sufficient to show that C has a *smooth* local correction procedure, where by smooth we roughly mean that coordinate of the code is queried with about the same probability. If C satisfies this, since we have good and query efficient distance amplification procedures [AEL95, KMRS17, CY21, CY22], that would yield a constant correction radius code as desired.

The codewords $c \in C$ by definition satisfy the line constraints, which are, observe from Equation (1.2), that for every direction $a \in \mathbb{F}_q^m$ and offset $b \in \mathbb{F}_q^m$,

$$\sum_{\tau \in \mathbb{F}_q} \sum_{H \in \mathcal{H}} a^H c(\tau, a_1\tau + b_1, \ldots, a_m\tau + b_m, H) = 0.$$

In particular imagine that we choose $a \in \mathbb{F}_q^m$ uniformly at random, and set $b = \alpha^* - a\tau^*$. It is not too hard to see every point (beside (τ^*, α^*)) has probability at most $\frac{q}{n}$ to be on the ordered line with direction a and this offset b, and that (τ^*, α^*) is the τ^*-th point of the ordered line. Thus,

$$\sum_{H \in \mathcal{H}} a^H c(\tau^*, \alpha^*, H) = \sum_{\tau \in \mathbb{F}_q \setminus \{\tau^*\}} \sum_{H \in \mathcal{H}} a^H c(\tau, a\tau + b, H), \tag{1.5}$$

and recall that we ignore plus/minus signs by assuming $\mathsf{char}(\mathbb{F}_q) = 2$ in this overview. As we are interested in $c(\tau^*, \alpha^*, H)$, and by querying the sampled line, we would only get one equation involving it but also other unknowns, we can, like in the decoding of multiplicity codes [KSY14] choose roughly $|\mathcal{H}| < s^m$ such lines and solve the system of equations. In fact in [Kop15] it is shown that in the case of multiplicity codes the number of lines can even be reduced to $2^{O(m)}$. However, for our needs, even $2^{O(m)}$ is far too large since we aim for a polylogarithmic number of queries.

We take a moment to inspect Equation (1.5). If we define for every $(\tau, \alpha) \in \mathbb{F}_q \times \mathbb{F}_q^m$ the polynomial

$$p^c_{\tau,\alpha} = \sum_{H \in \mathcal{H}} c(\tau, \alpha, H) \cdot y^H \qquad \in \mathbb{F}_q[y] = \mathbb{F}_q[y_1, \ldots, y_m],$$

then rewriting Equation (1.5),

$$p^c_{\tau^*,\alpha^*}(a) = \sum_{\tau \in \mathbb{F}_q \setminus \{\tau^*\}} p^c_{\tau, a\tau+b}(a).$$

That is, *querying the line at direction a gives us the evaluation of $p^c_{\tau^*,\alpha^*}$ on point a.* Yet, by itself, this does not paint a better way to obtain $c(\tau^*, \alpha^*, H^*)$. However, we can make the following observation, which is that we know something about the polynomial $p^c_{\tau^*,\alpha^*}$: that by our choice of $\mathcal{H}$, *it is of total degree at most $s-1$.* This means that directly getting its evaluation on a is not the only way to deduce $p^c_{\tau^*,\alpha^*}(a)$. Rather, we can "locally correct" $p^c_{\tau*,\alpha*}(a)$ by obtaining any s evaluation points of $p^c_{\tau*,\alpha*}$, on a line which passes through a (recall that a in itself was the direction of a line chosen in order to correct $c(\tau^*, \alpha^*, H^*)$).

Did we make any progress by observing that we can "locally correct" $p^c_{\tau*,\alpha*}(a)$? This would have helped us, in case we needed to obtain $p^c_{\tau*,\alpha*}(a)$ instead of $c(\tau^*, \alpha^*, H^*)$. This is because it suggests a way to obtain $p^c_{\tau*,\alpha*}(a)$ *smoothly*, opposed to only having one deterministic way (querying exactly the line at direction a). Instead, to query all the lines corresponding to a set of s directions $a^{(1)}, \ldots, a^{(s)} \in \mathbb{F}_q^m$ which lie on a line of $\mathbb{F}_q^m$ which passes through a – suffices in order to obtain $p^c_{\tau*,\alpha*}(a)$. One can observe that we can choose such a line uniformly at random, and the s directions on it uniformly at random, resulting in a smooth decoding procedure for $p^c_{\tau*,\alpha*}(a)$, since marginally each direction $a^{(i)}$ is uniform.

But again, we did not set out to obtain $p^c_{\tau*,\alpha*}(a)$ for some $a \in \mathbb{F}_q^m$; rather, our goal was to locally correct our code C, that is to recover $c(\tau^*, \alpha^*, H^*)$. The final trick, then, is to change that goal. Since we have, for each point $(\tau, \alpha) \in \mathbb{F}_q \times \mathbb{F}_q^m$, a good correction procedure for evaluations of the polynomial $p^c_{\tau,\alpha}$, why not replace each block $(c(\tau, \alpha, H))_{H\in\mathcal{H}}$ with $(p_{\tau,\alpha}(a))_{a\in S}$, where $S \subseteq \mathbb{F}_q^m$ is some chosen set of evaluation points (one needs to verify that this is a linear transformation and indeed it is). In fact, this is what we do. We accordingly construct from C a code $C' \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times S}$, and by choosing S to be any interpolating set of $\mathbb{F}_q^m$ for degree at most $s-1$ polynomials (that is, no such polynomial evaluates to 0 on all of S) of size $|\mathcal{H}|$, this doesn't change the length of the code, and keeps Equation (1.5) useful for our decoding, since we can deduce each $p^c_{\tau,\alpha}$ by querying $(p^c_{\tau,\alpha}(a))_{a\in S}$.

To conclude, we constructed a block-wise locally correctable code $C' \subseteq \mathbb{F}_q^{P\times S}$, $|P| = n$, which corrects each coordinate by smoothly querying $s \cdot (q-1)$ blocks, corresponding to the s line directions we sample, and the $q-1$ blocks we query on each such line. In our choice of parameters we will set, for any chosen σ: $q \approx \log^\sigma n$, $s = \log^{O(1)}(n)$ and $m \approx \frac{1}{\sigma} \cdot \frac{\log n}{\log\log n} - 1$ (to be constant with that $|P| = q^{m+1} = n$). This choice, by Theorem 1.6, assures that C (and therefore C') has a high rate. The block-length is $|S| = |\mathcal{H}| \leqslant s^m \approx (\log n)^{O(1)\cdot\frac{1}{\sigma}\cdot\frac{\log n}{\log\log n}} = n^{\frac{O(1)}{\sigma}}$, while the query complexity is less than $sq \approx (\log n)^{\sigma+O(1)}$, as wanted. The exact details, as well as the distance amplification step, are found in Section 4.

Remark 1.7 (On the connection to multiplicity codes.)**.** *The line-constraint subspace underlying our construction is closely related to the structure of classical multiplicity codes [KSY14]. In fact, one can view our C defined above as a restricted version of a multiplicity code, where we retain only a subset of the linear relations that arise from* *taking directional derivatives along lines (though the final C' is a different code). We believe that standard multiplicity codes themselves would have sufficed, but here we isolate only the minimal portion of the structure that suffices for our decoding argument. The more rich structure of standard multiplicity codes is very useful, while in our view focusing only on the linear relations considered by us here has the advantage of making the steps described in Section 1.2.5 follow somewhat more naturally.*

2 Preliminaries

2.1 Notation.

All logarithms are taken base 2. $\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers. For $m \in \mathbb{N}$, $[m] = \{1, 2, \ldots, m\}$. For a prime power q, $\mathbb{F}_q$ is the finite field with q elements. For two vector spaces $A = \mathbb{F}_q^U$ and $B = \mathbb{F}_q^V$ their tensor product $A \otimes B \subseteq \mathbb{F}_q^{U \times V}$ is the space $\mathrm{Span}\{f \in \mathbb{F}_q^{U \times V} \mid \exists g \in \mathbb{F}_q^U, h \in \mathbb{F}_q^V \text{ such that } \forall x, y\ f(x, y) = g(x) \cdot h(y)\}$.

Abbreviated m-wise notation. Fix $m \in \mathbb{N}$. For vectors $u = (u_1, \ldots, u_m)$ over a ring/field and a multi-index $I = (i_1, \ldots, i_m) \in (\mathbb{N} \cup \{0\})^m$, write

$$u^I := \prod_{r=1}^{m} u_r^{i_r}.$$

For $k \in \mathbb{N} \cup \{0\}$, let $\bar{k} := (k, \ldots, k) \in (\mathbb{N} \cup \{0\})^m$ and abbreviate $u^k := u^{\bar{k}}$. For $I = (i_1, \ldots, i_m), J = (j_1, \ldots, j_m) \in (\mathbb{N} \cup \{0\})^m$, define

$$|I| := \sum_{r=1}^{m} i_r, \qquad I \leqslant J \iff i_1 \leqslant j_1, \ldots, i_m \leqslant j_m, \qquad \binom{I}{J} := \prod_{r=1}^{m} \binom{i_r}{j_r}.$$

For $i \in [m]$, e_i denotes the i-th unit vector.

For indeterminates $x = (x_1, \ldots, x_m)$, we write the monomial $x^I := \prod_{r=1}^m x_r^{i_r}$. For a subset $W = \{w_1, \ldots, w_{|W|}\} \subseteq [m]$ where $w_1 < \cdots < w_{|W|}$, we define $x_W = (x_{w_1}, \ldots, x_{w_{|W|}})$.

2.2 Facts.

We will use the following easy fact.

Fact 2.1. *Let $v_1, \ldots, v_t \in \mathbb{Z}^n$ be integral vectors such that $\dim_{\mathbb{R}}(v_1, \ldots, v_t) = k$ and let p be a prime number. Then, $\dim_{\mathbb{F}_p}(v_1^p, \ldots, v_t^p) \leqslant k$ where v_i^p is the vector v_i with all of its elements reduced modulo p.*

Proof. Over any field $\mathbb{F}$, $\dim_{\mathbb{F}}(v_1, \ldots, v_t) \leqslant k$ if and only if every $k+1$-subset of $\{v_1, \ldots, v_t\}$ is $\mathbb{F}$-linearly dependent. Let $v_{i_1}^p, \ldots, v_{i_{k+1}}^p$ be a $k+1$-subset of $v_1^p, \ldots, v_t^p$. Since $\dim_{\mathbb{R}}(v_1, \ldots, v_t) = k$, $v_{i_1}, \ldots, v_{i_{k+1}}$ are linearly dependent over $\mathbb{R}$. Thus there exist not-all-zero $\gamma_1, \ldots, \gamma_{k+1}$ such that $\sum \gamma_j v_{i_j} = \bar{0}$ and since $v_{i_1}, \ldots, v_{i_{k+1}}$ are integral we can assume without loss of generality that $\gamma_1, \ldots, \gamma_{k+1} \in \mathbb{Z}$. Moreover, we can further assume without loss of generality that it is not the case that p divides all of $\gamma_1, \ldots, \gamma_{k+1}$ (otherwise we divide them by their largest common divisible power of p). Thus, $\sum(\gamma_j \mod p)v_{i_j}^p = \bar{0}$ over $\mathbb{F}_p^n$ is a zero non-trivial linear combination of $v_{i_1}, \ldots, v_{i_{k+1}}$. The fact follows. □

Fact 2.2. *For every $i, j \in \mathbb{N}$*

$$\frac{i}{j}\binom{i-1}{j-1} = \binom{i}{j}.$$

Fact 2.3. *For every e_f unit vector for $f \in [m]$, and $I, J \in (\mathbb{N} \cup \{0\})^m$ such that $I, J \geqslant e_f$*

$$\frac{i_f}{j_f}\binom{I-e_f}{J-e_f} = \binom{I}{J}.$$

Proof. Follows trivially from Fact 2.2. □

3 The line constraints subspace

In the following section q is a prime power, $m \in \mathbb{N}$ and $s \in \mathbb{N}$ are some parameters. $\mathcal{H} \subseteq (\mathbb{N} \cup \{0\})^m$ is a finite set. We define $n := q^{m+1}$ and $N := n|\mathcal{H}|$.

3.1 The desired functions and the bound on their dimension

We define a linear subspace

$$\mathcal{L} = \mathrm{Span}\{L^{a,b} \mid a, b \in \mathbb{F}_q^m\}$$

which is to contain all functions which correspond to lines of direction (minus) a and offset (minus) b. For every $a, b \in \mathbb{F}_q^m$,

$$L^{a,b} : \mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H} \to \mathbb{F}_q$$

is defined as follows. For every $\tau \in \mathbb{F}_q$, $\alpha \in \mathbb{F}_q^m$ and $H \in \mathcal{H}$

$$L^{a,b}(\tau, \alpha, H) = \begin{cases} a^H & \text{if } \alpha = -a\tau - b \\ 0 & \text{otherwise.} \end{cases}$$

In words, $L^{a,b}(\tau, \alpha, H)$ takes value a^H if (τ, α) is on the ordered line $\{(\tau, -a\tau - b) \mid \tau \in \mathbb{F}_q\}$, and 0 outside of it.

In Theorem 3.2 we state a bound on the dimension of $\mathcal{L}$ over $\mathbb{F}_q$. Prior to that, we make the following important definition.

Definition 3.1.

$$\mathit{Boundary}(\mathcal{H}) := \{H \in \mathcal{H} \mid \exists i \in [m] : H + e_i \notin \mathcal{H}\}.$$

Theorem 3.2 (Theorem 1.6, rephrased)**.**

$$\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant (m(q-1)+1) \cdot q^m \cdot |\mathit{Boundary}(\mathcal{H})| + \frac{m+1}{q} N.$$

In particular, for the choice $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s-1\}$,

$$\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant \frac{(m(q-1)+1)}{q} \cdot \frac{m}{m+s-1} \cdot N + \frac{m+1}{q} \cdot N.$$

We defer the proof for Theorem 3.2 to the end of this section and we first set up some needed claims and definitions.

3.2 Expressing the desired functions

The following claim states that each function $L^{a,b}$ can be expressed as a product of a polynomial in $\mathbb{F}_q[t, x]$ where $x = (x_1, \ldots, x_m)$, and the function a^H.

Claim 3.3. *For every* $a, b \in \mathbb{F}_q^m$

$$L^{a,b}(t, x, H) = a^H \prod_{i \in [m]} (1 - (x_i + a_i t + b_i)^{q-1}).$$

Proof. For every $\alpha \in \mathbb{F}_q$, $\alpha^{q-1} = 1$ if $\alpha \neq 0$ and 0 otherwise. Thus, for every $i \in [m]$, $1 - (x_i + a_i t + b_i)^{q-1}$ is 1 if $x_i = -a_i t - b_i$ and 0 otherwise. Hence, the product over i evaluates to 1 if $x = -at - b$ and to 0 otherwise. It only remains to multiply by a^H per the definition of $L^{a,b}$. $\square$

Thus

$$\begin{aligned} L^{a,b}(t, x, H) &= a^H \sum_{W \subseteq [m]} (-1)^{|W|} (x_W + a_W t + b_W)^{\overline{q-1}} \\ &= a^H \sum_{W \subsetneq [m]} (-1)^{|W|} (x_W + a_W t + b_W)^{\overline{q-1}} + (-1)^m a^H (x + at + b)^{\overline{q-1}}. \quad (3.1) \end{aligned}$$

We define for every $a, b \in \mathbb{F}_q^m$ the functions

$$\widetilde{L}^{a,b}(t,x,H) = a^H (x + at + b)^{\overline{q-1}},$$
$$\widetilde{\widetilde{L}}^{a,b}(t,x,H) = a^H \sum_{W \subsetneq [m]} (-1)^{|W|} (x_W + a_W t + b_W)^{\overline{q-1}},$$

and the families

$$\widetilde{\mathcal{L}} = \operatorname{Span}\{\widetilde{L}^{a,b} \mid a, b \in \mathbb{F}_q^m\},$$
$$\widetilde{\widetilde{\mathcal{L}}} = \operatorname{Span}\{\widetilde{\widetilde{L}}^{a,b} \mid a, b \in \mathbb{F}_q^m\}.$$

We observe that it is essentially enough to bound only $\dim_{\mathbb{F}_q}(\widetilde{\mathcal{L}})$.

Claim 3.4.

$$\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant \dim_{\mathbb{F}_q}(\widetilde{\mathcal{L}}) + \frac{m}{q} N.$$

Proof. We have that $\mathcal{L} \subseteq \widetilde{\mathcal{L}} + \widetilde{\widetilde{\mathcal{L}}}$, by Equation (3.1), and thus $\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant \dim_{\mathbb{F}_q}(\widetilde{\mathcal{L}}) + \dim_{\mathbb{F}_q}(\widetilde{\widetilde{\mathcal{L}}})$. It remains to observe that for every $a, b \in \mathbb{F}_q^m$, $\widetilde{\widetilde{L}}^{a,b}$ can be expressed as the product of a^H with a sum of m polynomials $g_1^{a,b}(t,x), \ldots, g_m^{a,b}(t,x)$ where for every $i \in [m]$, $g_i^{a,b}$ does not depend on x_i, and thus is spanned by the set of monomials $M_i = \{t^j x^I \mid 0 \leqslant j \leqslant q-1, \overline{0} \leqslant I \leqslant \overline{q-1}, I_i = 0\}$, which is of size n/q. As

$$\widetilde{\widetilde{L}}^{a,b} \in \mathbb{F}_q^{\mathcal{H}} \otimes \operatorname{Span}(\bigcup_{i \in [m]} M_i)$$

we conclude that $\dim_{\mathbb{F}_q}(\widetilde{\widetilde{\mathcal{L}}}) \leqslant |\mathcal{H}| \sum_{i \in [m]} |M_i| \leqslant |\mathcal{H}| mn/q = \frac{m}{q} N$. □

3.3 From a span of functions in t, x, H to a span of polynomials in a, b

We want to show that we can span each $\widetilde{L}^{a,b}(t,x,H) = a^H(x + at + b)^{\overline{q-1}}$ using a low dimension. Notice that

$$a^H (at + b + x)^{\overline{q-1}} = a^H \sum_{0 \leqslant I \leqslant \overline{q-1}} \binom{\overline{q-1}}{I} (at+b)^I x^{\hat{I}},$$

where $\hat{I} := \overline{q-1} - I$. Recall that $\widetilde{\mathcal{L}} \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}}$ is a vector space, while were defined to be the span of a set of functions in variables t, x, H, going over all possible $a, b \in \mathbb{F}_q^m$. We will now observe that the dimension of $\widetilde{\mathcal{L}}$ is in fact related to the dimension of the space of certain polynomials in *formal variables* $a = (a_1, \dots, a_m), b = (b_1, \dots, b_m)$, over $\mathbb{F}_q[a,b]$. Specifically, we define for every $0 \leqslant j \leqslant m(q-1)$, $\bar{0} \leqslant I \leqslant \overline{q-1}$ and $H \in \mathcal{H}$,

$$D^{a,b}(j,I,H) := \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \in \mathbb{F}_q[a,b], \tag{3.2}$$

and we then take

$$\mathcal{D} := \operatorname{Span}\{D^{a,b}(j,I,H) \mid 0 \leqslant j \leqslant m(q-1), \bar{0} \leqslant I \leqslant \overline{q-1}, H \in \mathcal{H}\},$$

and we view $\mathcal{D}$ as a vector space over $\mathbb{F}_q$. We argue that

Claim 3.5.

$$\dim_{\mathbb{F}_q} \widetilde{\mathcal{L}} \leqslant \dim_{\mathbb{F}_q} \mathcal{D}.$$

Proof. Fix $a, b \in \mathbb{F}_q^m$ and expand

$$\widetilde{L}^{a,b}(t,x,H) = a^H \, (x + at + b)^{\overline{q-1}} = \sum_{\bar{0} \leqslant I \leqslant \overline{q-1}} \binom{\overline{q-1}}{I} x^{\hat{I}} \, (at+b)^I \, a^H.$$

Writing $(at+b)^I = \sum_{J \leqslant I} \binom{I}{J} (at)^J b^{I-J}$ and grouping by the power of t,

$$\widetilde{L}^{a,b}(t,x,H) = \sum_{\bar{0} \leqslant I \leqslant \overline{q-1}} \sum_{j=0}^{|I|} \left(\sum_{\substack{J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \right) \cdot \binom{\overline{q-1}}{I} t^j x^{\hat{I}}.$$

By definition,

$$D^{a,b}(j,I,H) = \sum_{\substack{J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \in \mathbb{F}_q[a,b],$$

so we can rewrite the expansion as

$$\widetilde{L}^{a,b}(t,x,H) = \sum_{\bar{0} \leqslant I \leqslant \overline{q-1}} \sum_{j=0}^{|I|} \binom{\overline{q-1}}{I} t^j x^{\hat{I}} \cdot D^{a,b}(j,I,H). \tag{3.3}$$

Let $r = \dim_{\mathbb{F}_q} \mathcal{D}$ and choose a basis $p_1, \dots, p_r$ of $\mathcal{D}$ (as a subspace of $\mathbb{F}_q[a,b]$). For each triple (j, I, H) there exist scalars $\gamma_h(j,I,H) \in \mathbb{F}_q$ such that, in $\mathcal{D}$,

$$D^{a,b}(j,I,H) = \sum_{h=1}^{r} \gamma_h(j,I,H) \, p_h(a,b).$$

Substituting this into (3.3) and interchanging sums gives

$$\widetilde{L}^{a,b}(t,x,H) = \sum_{h=1}^{r} p_h(a,b) \cdot \left(\sum_{\bar{0} \leqslant I \leqslant \overline{q-1}} \sum_{j=0}^{|I|} \gamma_h(j,I,H) \binom{\overline{q-1}}{I} t^j x^{\hat{I}} \right). \tag{3.4}$$

Define the (fixed) functions

$$G_h(t,x,H) = \sum_{\bar{0}\leqslant I\leqslant \overline{q-1}} \sum_{j=0}^{|I|} \gamma_h(j,I,H) \binom{\overline{q-1}}{I} t^j x^{\hat{I}} \;\in\; \mathbb{F}_q^{\mathbb{F}_q\times\mathbb{F}_q^m\times\mathcal{H}}.$$

These $G_1,\ldots,G_r$ do not depend on a,b. Equation (3.4) above says that

$$\widetilde{L}^{a,b} \in \mathrm{Span}_{\mathbb{F}_q}\{G_1,\ldots,G_r\} \qquad \text{for every } a,b\in\mathbb{F}_q^m.$$

Hence $\widetilde{\mathcal{L}} \subseteq \mathrm{Span}_{\mathbb{F}_q}\{G_1,\ldots,G_r\}$ and therefore

$$\dim_{\mathbb{F}_q}\widetilde{\mathcal{L}} \leqslant r = \dim_{\mathbb{F}_q}\mathcal{D}.$$

□

We thus turn our focus to showing that $\dim_{\mathbb{F}_q}\mathcal{D}$ is small. Observe, by considering Equation (3.2), that each $D^{a,b}(j,I,H)$ was defined as a polynomial with *integer* coefficients (then taken modulo the characteristic to get a polynomial over $\mathbb{F}_q$). Hence, we can re-view each $D^{a,b}(j,I,H)$ as being in $\mathbb{R}[a,b]$, and consider $\dim_{\mathbb{R}}$ of $\mathcal{D}$.

Claim 3.6.

$$\dim_{\mathbb{F}_q}\mathcal{D} \leqslant \dim_{\mathbb{R}}\mathcal{D}.$$

Proof. Follows by Fact 2.1. □

3.4 Bounding the span of polynomials over the Reals

Proposition 3.7.

$$\dim_{\mathbb{R}}\mathcal{D} \leqslant q^m|\mathcal{H}| + (m(q-1)+1)\cdot q^m \cdot |\mathit{Boundary}(\mathcal{H})|.$$

Proof. Consider

$$A = \{D^{a,b}(0,I,H) \mid \bar{0}\leqslant I\leqslant \overline{q-1}, H\in\mathcal{H}\},$$
$$B = \{D^{a,b}(j,I,H) \mid 0\leqslant j\leqslant m(q-1), \bar{0}\leqslant I\leqslant \overline{q-1}, H\in \mathrm{Boundary}(\mathcal{H})\}.$$

We prove

$$\mathrm{Span}(A\cup B) = \mathcal{D}. \tag{3.5}$$

Since $|A| = q^m|\mathcal{H}|$ and $|B| = (m(q-1)+1)\, q^m\, |\mathrm{Boundary}(\mathcal{H})|$, the bound on $\dim_{\mathbb{R}}\mathcal{D}$ follows immediately from (3.5).

Fix $j \in \{1, \ldots, m(q-1)\}$, $\bar{0} \leqslant I \leqslant \overline{q-1}$, and $H \in \mathcal{H}$. If $H \in \text{Boundary}(\mathcal{H})$ then $D^{a,b}(j, I, H) \in B$ and we are done. Otherwise, it suffices to show

$$D^{a,b}(j, I, H) \ \in \ \text{Span}_{\mathbb{R}}\Big\{ D^{a,b}(j-1,\, I-e_f,\, H+e_f) \ \mid \ f \in G \Big\}, \tag{3.6}$$

where $G = \{f \in [m] \mid i_f > 0\}$. Once Equation (3.6) is established, we may iterate the step while $j > 0$: either we reach $j' = 0$ (hence a member of A), or at some intermediate time we use an index f with $H' := H + e_f \in \text{Boundary}(\mathcal{H})$, in which case the corresponding term lies in B. In all cases we obtain $D^{a,b}(j, I, H) \in \text{Span}_{\mathbb{R}}(A \cup B)$, proving Equation (3.5).

To show Equation (3.6) recall that

$$D^{a,b}(j, I, H) \ = \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J| = j}} \binom{I}{J} a^{H+J} b^{I-J}.$$

Consider the linear combination

$$(*) := \sum_{f \in G} \frac{i_f}{j} D^{a,b}(j-1,\, I-e_f,\, H+e_f).$$

For each $f \in G$, by the definition of $D^{a,b}$ we have

$$D^{a,b}(j-1,\, I-e_f,\, H+e_f) = \sum_{\substack{\bar{0} \leqslant J' \leqslant I-e_f \\ |J'| = j-1}} \binom{I-e_f}{J'} a^{H+e_f+J'}\, b^{I-e_f-J'}.$$

Substituting this into $(*)$ gives

$$(*) = \sum_{f \in G} \frac{i_f}{j} \sum_{\substack{\bar{0} \leqslant J' \leqslant I-e_f \\ |J'| = j-1}} \binom{I-e_f}{J'} a^{H+e_f+J'}\, b^{I-e_f-J'}. \tag{3.7}$$

Next, for each fixed $f \in G$, we perform a change of variables: let

$$J = J' + e_f.$$

Then note that

$$\bar{0} \leqslant J' \leqslant I - e_f,\ |J'| = j-1 \quad \Longleftrightarrow \quad e_f \leqslant J \leqslant I,\ |J| = j.$$

Further, under this substitution we have

$$\binom{I-e_f}{J'} = \binom{I-e_f}{J-e_f}, \quad a^{H+e_f+J'} = a^{H+J}, \quad b^{I-e_f-J'} = b^{I-J}.$$

Thus we get

$$(*) = \sum_{f \in G} \frac{i_f}{j} \sum_{\substack{e_f \leqslant J \leqslant I \\ |J|=j}} \binom{I - e_f}{J - e_f} a^{H+J} b^{I-J},$$

which already looks more similar to $D^{a,b}(j, I, H)$ that we wish to show is expressed, though we are not quite done yet.

For each $f \in G$ and each J in the range (notice that $I, J \geqslant e_f$), we can apply Fact 2.3 to get

$$i_f \binom{I - e_f}{J - e_f} = j_f \binom{I}{J}.$$

Thus,

$$(*) = \sum_{f \in G} \sum_{\substack{e_f \leqslant J \leqslant I \\ |J|=j}} \frac{j_f}{j} \binom{I}{J} a^{H+J} b^{I-J}.$$

We proceed by noticing that we can extend the inner sum to range over all $\bar{0} \leqslant J \leqslant I$ with $|J| = j$, since for J with $j_f = 0$, the inner term anyhow evaluates to 0.

$$(*) = \sum_{f \in G} \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \frac{j_f}{j} \binom{I}{J} a^{H+J} b^{I-J}.$$

Changing the order of summation and taking out terms which don't depend on f,

$$(*) = \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \sum_{f \in G} \frac{j_f}{j}.$$

Now, by G's definition - for $f \notin G$, $i_f = 0$ and thus also $j_f = 0$ for every J in the summation, we see that

$$\begin{aligned}
(*) &= \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \sum_{f \in [m]} \frac{j_f}{j} \\
&= \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J} \frac{|J|}{j} \\
&= \sum_{\substack{\bar{0} \leqslant J \leqslant I \\ |J|=j}} \binom{I}{J} a^{H+J} b^{I-J}
\end{aligned}$$

which is the definition of $D^{a,b}(j, I, H)$ - as wanted. We thus established Equation (3.6), from which as said, the proposition follows.

$\square$

3.5 Concluding Theorem 3.2

Proof for Theorem 3.2. By Claim 3.4,

$$\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant \dim_{\mathbb{F}_q}(\widetilde{\mathcal{L}}) + \frac{m}{q}N.$$

By Claim 3.5 and Claim 3.6,

$$\dim_{\mathbb{F}_q} \widetilde{\mathcal{L}} \leqslant \dim_{\mathbb{F}_q} \mathcal{D} \leqslant \dim_{\mathbb{R}} \mathcal{D}.$$

By Proposition 3.7,

$$\dim_{\mathbb{R}} \mathcal{D} \leqslant (m(q-1)+1) \cdot q^m \cdot |\mathrm{Boundary}(\mathcal{H})| + \frac{1}{q}N.$$

Thus,

$$\dim_{\mathbb{F}_q}(\mathcal{L}) \leqslant (m(q-1)+1) \cdot q^m \cdot |\mathrm{Boundary}(\mathcal{H})| + \frac{m+1}{q}N,$$

as desired.

As for the in particular part of the theorem, it follows by noting that for $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s-1\}$,

$$|\mathcal{H}| = \binom{m+s-1}{m},$$

whereas

$$\mathrm{Boundary}(\mathcal{H}) = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| = s-1\}$$

and so

$$|\mathrm{Boundary}(\mathcal{H})| = \binom{m+s-2}{m-1}.$$

Appealing to Fact 2.2,

$$\frac{m+s-1}{m} \cdot |\mathrm{Boundary}(\mathcal{H})| = |\mathcal{H}|,$$

and so

$$\begin{aligned}\dim_{\mathbb{F}_q}(\mathcal{L}) &\leqslant (m(q-1)+1) \cdot q^m \cdot |\mathcal{H}| \cdot \frac{m}{m+s-1} + \frac{m+1}{q} \cdot N \\ &\leqslant \frac{(m(q-1)+1)}{q} \cdot \frac{m}{m+s-1} \cdot N + \frac{m+1}{q} \cdot N\end{aligned}$$

as wanted. □

4 Good blockwise LCCs with polylog query complexity

In this part we construct a good blockwise LCC with polylogarithmic query complexity. We will do so in two stages, first in Section 4.1 we will construct one such with rate $1-o(1)$ and a (modestly) vanishing correction radius. Second, in Section 4.2 we will apply the AEL distance amplification to increase the distance, and conclude Theorem 1.4. Such two step approach is similar to the one taken in [KMRS17].

As is pretty standard, it will be more convenient to work with a slightly different definition of local correction, in which we will consider the probability the a point being queried, instead of directly considering corruptions.

Definition 4.1 ((q,μ)-blockwise smooth LCC)**.** *A code $C \subseteq \mathbb{F}^{P\times S}$ is a (q,μ)-blockwise smooth LCC if there exists a randomized procedure* $\mathsf{Cor} : P \times S \to \mathbb{F}$ *that is given oracle access to $c \in C$ and has the following guarantee. For every $(p,a) \in P \times S$ and $c \in C$,* $\mathsf{Cor}^c(p,a) = c(p,a)$ *with probability 1. Furthermore,* $\mathsf{Cor}^c(p,a)$ *always makes at most q queries to c, where each query of* Cor *consists of obtaining $c(p',\cdot)$ for some $p' \in P$. Moreover, for every $(p',\cdot) \in P \times S$ the probability that $c(p',\cdot)$ is queried by* $\mathsf{Cor}^c(p,a)$ *is at most μ.*

We state the simple fact that a smooth-enough (q,μ)-blockwise smooth LCC is a decent-δ blockwise LCC (as defined in Definition 1.2).

Claim 4.2. *If $C \subseteq \mathbb{F}^{P\times S}$ where $|P| = n$ is a (q,μ)-blockwise smooth LCC then it is a blockwise-(q,δ) LCC for $\delta = \frac{1}{3n\mu}$.*

Proof. For $z \in \mathbb{F}^{P\times S}$ such that $|\{p \in P \mid z(p,\cdot) \neq c(p,\cdot)\}| < \delta n$, $c^z(p,a)$ outputs $c(p,a)$ in the case that no points $p' \in P$ where z and c differ were queried. By a union bound, since the probability to query each p' is at most μ, for $\delta = \frac{1}{3n\mu}$, the probability to make an erroneous query is at most $\delta n\mu = \frac{1}{3}$. □

4.1 High-rate blockwise LCCs

We will need to use interpolating sets for $\mathbb{F}_q^m$ which we define as follows.

Definition 4.3. *An s-interpolating set $S \subseteq \mathbb{F}_q^m$ is a set such that for every polynomial $q \in \mathbb{F}_q[y_1,\ldots,y_m]$ of total degree at most $s-1$, there exists $\alpha \in S$ such that $q(\alpha) \neq 0$.*

The following fact is well known.

Fact 4.4. *For every $s \leqslant q-1$ there is an explicit s-interpolating set $S \subseteq \mathbb{F}_q^m$ of size $\binom{m+s-1}{m}$.*

The next claim states that we can, instead of viewing each block as "coefficients" of a degree less than s polynomial, view each block as evaluations of such a polynomial, while still having the same line-wise requirements satisfied. The claim asserts that this results in the same dimension, but we stress that this is not the same code, since the constraints are in fact different.

Claim 4.5. *Let $C \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times \mathcal{H}}$, for $\mathcal{H} = \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s-1\}$, be the largest linear code satisfying the following property. For every $c \in C$ and for every $a, b \in \mathbb{F}^m$*

$$\sum_{\tau \in \mathbb{F}_q} \sum_{H \in \mathcal{H}} a^H c(\tau, a\tau + b, H) = 0. \tag{4.1}$$

Let $S \subseteq \mathbb{F}_q^m$ be an s-interpolating set and let $C' \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times S}$ be the largest linear code satisfying the following property. For every $c' \in C'$ and for every $a, b \in \mathbb{F}^m$

$$\sum_{\tau \in \mathbb{F}_q} q_{\tau, a\tau+b}(a) = 0 \tag{4.2}$$

where for every $\tau \in \mathbb{F}_q$ and $\gamma \in \mathbb{F}_q^m$, $q_{\tau,\gamma}$ is the unique polynomial of degree at most $s-1$ such that $\forall \beta \in S, q_{\tau,\gamma}(\beta) = c'(\tau, \gamma, \beta)$ (notice that these are indeed linear requirements). Then, $\dim C' = \dim C$.

Proof. The proof is straightforward. We show $C \subseteq C'$ by describing an injective $f : C \to C'$ (the other direction is identical). For $c \in C$ we define $c' = f(c)$ to be the word obtained by setting for every $\tau \in \mathbb{F}_q$, $\gamma \in \mathbb{F}_q^m$ and $\beta \in S$ $c'(\tau, \gamma, \beta) = \sum_{H \in \mathcal{H}} \beta^H c(\tau, \gamma, H)$. Indeed f is injective, since S is an s-interpolating set, for every $c_1 \neq c_2$, $c_1' = f(c_1) \neq c_2' = f(c_2)$. Moreover, it is immediate from the definition of $q_{\tau,\gamma}$ that since c satisfied any Equation (4.1), c' satisfies any Equation (4.2), and thus $c' \in C'$. □

The following important proposition asserts that a code which is constructed to satisfy Equation (4.2) is a blockwise smooth LCC with a low query.

Proposition 4.6. *Let $C' \subseteq \mathbb{F}_q^{\mathbb{F}_q \times \mathbb{F}_q^m \times S}$, for $S \subseteq \mathbb{F}_q^m$ an s-interpolating set, be such that for every $c' \in C'$ and for every $a, b \in \mathbb{F}^m$*

$$\sum_{\tau \in \mathbb{F}_q} q_{\tau, a\tau+b}(a) = 0 \tag{4.3}$$

where for every $\tau \in \mathbb{F}_q$ and $\gamma \in \mathbb{F}_q^m$, $q_{\tau,\gamma}$ is the unique polynomial of degree at most $s-1$ such that $\forall \beta \in S, q_{\tau,\gamma}(\beta) = c'(\tau, \gamma, \beta)$. Assume that $s \leqslant q-1$. Then, if we define $P = \mathbb{F}_q \times \mathbb{F}_q^m$ and $n = |P|$, $C' \subseteq \mathbb{F}^{P \times S}$ is a $(sq, \frac{sq}{n})$-blockwise smooth LCC.

Proof. To prove the proposition, we describe the correction procedure Cor.

The correction Cor. On oracle access to $c \in C'$, in order to correct an element $(\tau^*, \alpha^*, a^*) \in \mathbb{F}_q \times \mathbb{F}_q^m \times S$, Cor proceeds as follows.

1. Sample uniformly at random a line direction $\nu \in \mathbb{F}_q^m$.
2. Sample uniformly at random distinct $\sigma_1, \ldots, \sigma_s \in \mathbb{F}_q \setminus \{0\}$.
3. For every $i \in [s]$:
   (a) Set $a^{(i)} = \sigma_i \nu + a^*$.
   (b) Set $b^{(i)} = \alpha^* - \tau^* a^{(i)}$.
   (c) Query the $q-1$ blocks at the points $p \in P$ which are on the ordered line with direction $a^{(i)}$ and offset $b^{(i)}$, except for the τ^*-th point. That is, we query the blocks of the points $\{(\tau, \tau a^{(i)} + b^{(i)}) \mid \tau \in \mathbb{F}_q \setminus \{\tau^*\}\} \subseteq P$. By "query the blocks" we mean that for every such p we query $c'(p, \cdot)$.
   (d) For every such point on the line, p_τ for $\tau \in \mathbb{F}_q \setminus \{\tau^*\}$, denote the resulted block of the query by $B_\tau : S \to \mathbb{F}_q$.
   (e) For each $\tau \in \mathbb{F}_q \setminus \{\tau^*\}$, compute the unique degree less than s polynomial $q_\tau \in \mathbb{F}_q[y_1, \ldots, y_m]$ which agrees with B_τ on S.
   (f) Set $\Delta_i = \sum_{\tau \in \mathbb{F}_q \setminus \{\tau^*\}} q_\tau(a^{(i)})$.
4. Compute the unique univariate polynomial of degree less than s, $r \in \mathbb{F}_q[z]$, such that for every $i \in [s]$, $r(\sigma_i) = -\Delta_i$.
5. Output $r(0)$.

Query analysis. It follows immediately by inspecting Item 3c that Cor queries at most $s(q-1)$ blocks.

Correctness. Let $q^* \in \mathbb{F}_q[y_1, \ldots, y_m]$ denote the unique degree less than s polynomial which agrees with $c'(\tau^*, \alpha^*, \cdot)$ on S. Now, notice that for every $i \in [s]$ the value Δ_i that we compute at Item 3f of the iteration is exactly the sum of the evaluations on $a^{(i)} \in \mathbb{F}_q^m$ of the polynomials in $\mathbb{F}_q[y_1, \ldots, y_m]$ of the blocks along the line $(\tau, \tau a^{(i)} + b^{(i)})_{\tau \in \mathbb{F}_q}$, except for the evaluation of the polynomial which corresponds to the τ^*-th point of the line. Notice that the τ^*-th point of the line, by Item 3b, is exactly

$$(\tau^*, \tau^* a^{(i)} + b^{(i)}) = (\tau^*, \tau^* a^{(i)} + \alpha^* - \tau^* a^{(i)}) = (\tau^*, \alpha^*),$$

which corresponds to the polynomial q^*. Thus, by Equation (4.3),

$$\Delta_i + q^*(a^{(i)}) = 0,$$

and hence $q^*(a^{(i)}) = -\Delta_i$. Now, notice that $a^{(1)}, \ldots, a^{(s)}$ are all on the line $\{\sigma \cdot \nu + a^* \mid \sigma \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^m$. Since the reduction of q^* on that line is a univariate polynomial of degree less than s, which on $0 \in \mathbb{F}_q$ evaluates to $q^*(a^*)$, it readily follows that in Item 4 the computed r is equal to that polynomial, and that the value outputted in Item 5 is $q^*(a^*)$. Since $q^*(a^*) = c'(\tau^*, \alpha^*, a^*)$ by the definition of q^*, it follows that Cor is correct.

Smoothness. Let $p' = (\tau', \alpha') \in P$. Fix $i \in [s]$. We ask what is the probability that the block corresponding to p' is queried in Item 3c. If $\tau' = \tau^*$ this never happens. If $\tau' \neq \tau^*$, this happens if and only if $\tau' a^{(i)} + b^{(i)} = \alpha'$. By inspecting Item 3a and Item 3b one sees that this event, in turn, is equivalent to

$$(\tau' - \tau^*)(\sigma_i \nu + a^*) = \alpha' - \alpha^*.$$

Since σ_i is by choice nonzero, and $\nu \in \mathbb{F}_q^m$ is independent of it and uniformly random, the probability for this to occur is $\frac{1}{q^m}$. Since this was for a fixed $i \in [s]$, the probability that the block corresponding to p' is queried by any of the s iterations is at most $\frac{s}{q^m} = \frac{sq}{n}$, as required. ☐

The following proposition concludes that for infinitely many n's, there exists a high rate blockwise LCC.

Proposition 4.7. *For every $\sigma \geqslant 4$ the following holds. For every $n' \in \mathbb{N}$ there exists $n \geqslant n'$ for which the following holds. There is a code $C \subseteq \mathbb{F}_q^{P \times S}$, with $|P| = n$, $q = \mathrm{poly}_\sigma(\log n)$, $|S| \leqslant n^{3/\sigma}$, which is a $(2\log^{\sigma+3}(n), \frac{2\log^{\sigma+3}(n)}{n})$-blockwise smooth LCC, with $\dim_{\mathbb{F}_q}(C) = (1 - o(1))N$.*

Proof. We set q to be the minimal power of 2 which is larger than $\log^\sigma(n')$. Note that $\log^\sigma(n') \leqslant q \leqslant 2\log^\sigma(n')$. We further set

$$\begin{aligned}
m &= \frac{\log(n')}{\log q} - 1, \\
n &= q^{m+1}, \\
s &= m^3, \\
P &= \mathbb{F}_q \times \mathbb{F}_q^m, \\
\mathcal{H} &= \{H \in (\mathbb{N} \cup \{0\})^m \mid |H| \leqslant s - 1\}, \\
N &= |P| \cdot |\mathcal{H}|.
\end{aligned}$$

Note that $|P| = q^{m+1} = n \geqslant n'$. Let $C \subseteq \mathbb{F}_q^{P\times\mathcal{H}} = \mathbb{F}_q^{\mathbb{F}_q\times\mathbb{F}_q^m\times\mathcal{H}}$ be the maximal linear subspace satisfying all line constraints Equation (4.1). Notice that $C = \mathcal{L}^{\perp}$,[6] where $\mathcal{L}$ is as defined in the previous section. Thus, by the in particular part of Theorem 3.2,

$$\dim_{\mathbb{F}_q} C \geqslant N - \left(\frac{(m(q-1)+1)}{q} \cdot \frac{m}{m+s-1} \cdot N + \frac{m+1}{q} \cdot N\right) = (1-o(1))N,$$

for our choice of q, m and s. Let $S \subseteq \mathbb{F}_q^m$ be an s-interpolating set, which exists by Fact 4.4, and let C' be the maximal linear subspace satisfying all line constraints as in Equation (4.3). By Claim 4.5, $\dim C' = \dim C$. By Proposition 4.6, C', is a $(sq, \frac{sq}{n})$-blockwise smooth LCC, and notice that $sq \leqslant 2\log^{\sigma+3}(n') \leqslant 2\log^{\sigma+3}(n)$ and

$$|S| \leqslant s^m \leqslant (\log^3 n')^{\frac{\log n'}{\log q}} = 2^{\frac{3\log(n')\cdot\log\log(n')}{\log(q)}} \leqslant 2^{\frac{3\log(n')\cdot\log\log(n')}{\sigma\log\log(n')}} \leqslant n^{3/\sigma}.$$

As required. $\square$

4.2 Applying the AEL distance amplification to get asymptotically good blockwise LCCs

[KMRS17] crucially observed that the AEL distance amplification [AEL95] is fit for amplifying the distance of LCCs, and adapted it. A variant of those amplification procedures, which is given in the language of linear constraints, was described in [CY22] (in particular it doesn't increase the length of the code), so we opt to using it, for convenience.

Definition 4.8 ([CY22])**.** *A linear subspace $A \subseteq \mathbb{F}^n$ is called a (q,δ,α)-*local-amplifier *if there exists a deterministic procedure* $\mathsf{Amp} : [n] \to \mathbb{F}$ *that is given oracle access to $z \in \mathbb{F}^n$ and has the following guarantee. For every $y \in A$ and $z \in \mathbb{F}^n$ such that* $\mathsf{Dist}(z,y) \leqslant \delta n$, $\mathsf{Amp}(i)$ *outputs y_i when given oracle access to z, for at least α-fraction of the indices $i \in [n]$. Furthermore,* Amp *always makes at most q queries to z.*

Claim 4.9 ([CY22])**.** *For every $n \in \mathbb{N}$, $\mathbb{F}$ a field, and $\delta, \alpha \in (0,1)$ such that $\delta \leqslant 1/25$, there exists a linear subspace $A \subseteq \mathbb{F}^n$ which is a (q,δ,α)-local-amplifier for $q = 25/(\delta(1-\alpha)^2)$ such that $\dim_{\mathbb{F}}(A) \geqslant (1-2H_{|\mathbb{F}|}(5\sqrt{\delta})-\sqrt{\delta}(1-\alpha))n$. Furthermore, adding to Definition 4.8, for every z, the guaranteed α-fraction set of good indices depends only on the locations where z disagrees with y ("corruptions"), and is a monotone function of these locations (that is, for two sets of corruptions where one is contained in the other, their two respective good indices sets satisfy that the second is contained in the first).*

[6] In fact, not precisely, for the following reason. Equation (4.1) requires that for every $a, b \in \mathbb{F}_q^m$, the sum along the line $(\tau, a\tau + b)_\tau$ is 0. However in Section 3, $\mathcal{L}$ was defined so that for every $a, b \in \mathbb{F}_q^m$, we have a constraint supported on the line $(\tau, -a\tau - b)_\tau$, to make the analysis more nice. However this is of little importance, as by permuting the coordinate names $(\tau,\alpha) \to (\tau,-\alpha)$ we don't change the dimension.

With everything set, we now prove Theorem 1.4.

Proof for Theorem 1.4. Set $\delta = 0.01$. Let $C \subseteq \mathbb{F}_q^{P\times S}$ be the code guaranteed by Proposition 4.7, when invoked with σ and n' of the hypothesis, to be a $(2\log^{\sigma+3}(n), \frac{2\log^{\sigma+3}(n)}{n})$-blockwise smooth LCC, with

$$\dim_{\mathbb{F}_q}(C) \geqslant (1 - o(1)) \cdot |P| \cdot |S|.$$

By Claim 4.2 C is a (qs, δ')-blockwise LCC for

$$\delta' = \frac{n}{3n \cdot 2\log^{\sigma+3}(n)} = \Theta\left(\frac{1}{\log^{\sigma+3}(n)}\right),$$

and let Cor be a corresponding local corrector for it. Set $\alpha = 1 - \delta'$ and let $A \subseteq \mathbb{F}_q^P$ be a (q_A, δ, α)-local-amplifier which exists by Claim 4.9, for

$$\begin{aligned}\dim_{\mathbb{F}_q}(A) &\geqslant (1 - 2H_q(5\sqrt{\delta}) - \sqrt{\delta}(1-\alpha))n = (\Omega(1) - O(\delta'))n = \Omega(n),\\ q_A &= \frac{25}{\delta(\delta')^2} = O(\log^{2\sigma+6}(n)),\end{aligned}$$

and let Amp be its corresponding procedure. We take

$$C' = \{c \in C \mid \forall a \in S : c(\cdot, a) \in A\}.$$

First, to address the dimension of C', by counting constraints we see that

$$\dim_{\mathbb{F}_q}(C') \geqslant \dim_{\mathbb{F}_q}(C) - (n - \dim_{\mathbb{F}_q}(A))|S| = \Omega(n|S|),$$

by the bounds on $\dim_{\mathbb{F}_q}(C)$ and $\dim_{\mathbb{F}_q}(A)$, and thus the rate $\rho = \frac{\dim_{\mathbb{F}_q}(C')}{n|S|} = \Omega(1)$.

Secondly, it readily follows that C' is a blockwise LCC as desired. Indeed, let $z \in \mathbb{F}_q^{P\times S}$ be such that for some $c \in C'$, $|\{p \in P \mid z(p,\cdot) \neq c(p,\cdot)\}| < \delta n$. Since, by Claim 4.9, for every $z_a \in \mathbb{F}_q^P$ (where for $a \in S$, $z_a := z(\cdot, a)$) the guaranteed set of good indices is a monotone function of the corruptions, we can assume without loss of generality that whenever for some $p \in P$ and $a \in S$ $c(p,a) \neq z(p,a)$, then for all $a' \in S$ $c(p,a') \neq z(p,a')$. As the good α-fraction sets induced by A for every z depend only on the location of corruptions, the good indices sets are the same for all $p \in P$, that is, there is an α fraction subset $P_{\text{good}} \subseteq P$ such that for any $p \in P_{\text{good}}$ and for any $a \in S$, $\mathsf{Amp}^{z(\cdot,a)}(p) = c(p,a)$. Hence, simulating Cor and whenever its makes a query $p \in P$ feeding it with the result of $(\mathsf{Amp}^{z(\cdot,a)}(p))\mid_{a\in S}$ has the same result as simulating it on a word z' which is $\alpha = (1-\delta')$-close to c. Notice that for such words z', $\mathsf{Cor}^{z'}(p,a)$ outputs the correct result $c(p,a)$ with probability at least $\frac{2}{3}$, as required. The number of queries of this correction procedure is at most

$$q_A \cdot 2\log^{\sigma+3} = O(\log^{3\sigma+9}(n)),$$

as wanted. $\square$

References

[AEL95] Noga Alon, Jeff Edmonds, and Michael Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519. IEEE, 1995.

[AG24] Omar Alrabiah and Venkatesan Guruswami. Near-tight bounds for 3-query locally correctable binary linear codes via rainbow cycles. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1874–1882. IEEE, 2024.

[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.

[BFLS91] László Babai, Lance Fortnow, Leonid A Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 21–32, 1991.

[BGT16] Arnab Bhattacharyya, Sivakanth Gopi, and Avishay Tal. Lower bounds for 2-query LCCs over large alphabet. *arXiv preprint arXiv:1611.06980*, 2016.

[BHKL25] Arpon Basu, Jun-Ting Hsieh, Pravesh K Kothari, and Andrew D Lin. Improved lower bounds for all odd-query locally decodable codes. *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, 2025.

[CY21] Gil Cohen and Tal Yankovitz. Rate amplification and query-efficient distance amplification for linear lcc and ldc. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[CY22] Gil Cohen and Tal Yankovitz. LCC and LDC: Tailor-made distance amplification and a refined separation. In *49th EATCS International Conference on Automata, Languages, and Programming*, volume 229 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 44, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022.

[DS05] Zeev Dvir and Amir Shpilka. Locally decodable codes with 2 queries and polynomial identity testing for depth 3 circuits. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 592–601, 2005.

[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.

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 529–540. ACM, 2013.

[Gol11] Oded Goldreich. Short locally testable codes and proofs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 333–372. Springer, 2011.

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

[JM25] Oliver Janzer and Peter Manohar. A $k^{\frac{q}{q-2}}$ lower bound for odd query locally decodable codes from bipartite kikuchi graphs. *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, 2025.

[KDW03] Iordanis Kerenidis and Ronald De Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 106–115, 2003.

[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.

[KM24] Pravesh K Kothari and Peter Manohar. Exponential lower bounds for smooth 3-lccs and sharp bounds for designs. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1802–1845. IEEE, 2024.

[KMRS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *Journal of the ACM (JACM)*, 64(2):11, 2017.

[Kop15] Swastik Kopparty. Some remarks on multiplicity codes. *arXiv preprint arXiv:1505.07547*, 2015.

[KSY14] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. High-rate codes with sublinear-time decoding. *Journal of the ACM (JACM)*, 61(5):28, 2014.

[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.

[Yan24] Tal Yankovitz. A stronger bound for linear 3-lcc. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1786–1801. IEEE, 2024.

[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.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il