Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 84 (2020)

Rate Amplification and Query-Efficient Distance Amplification for linear LCC and LDC

Gil Cohen* Tal Yankovitz†

June 29, 2021

Abstract

The main contribution of this work is a rate amplification procedure for LCC. Our procedure converts any q-query linear LCC, having rate ρ and, say, constant distance to an asymptotically good LCC with $q^{\mathrm{poly}(1/\rho)}$ queries.

Our second contribution is a distance amplification procedure for LDC that converts any linear LDC with distance δ and, say, constant rate to an asymptotically good LDC. The query complexity only suffers a multiplicative overhead that is roughly equal to the query complexity of a length $1/\delta$ asymptotically good LDC. This improves upon the $\mathrm{poly}(1/\delta)$ overhead obtained by the AEL distance amplification procedure [AL96, AEL95].

Our work establishes that the construction of asymptotically good LDC and LCC is reduced, with a minor overhead in query complexity, to the problem of constructing a vanishing rate linear LCC and a (rapidly) vanishing distance linear LDC, respectively.

*Tel Aviv University. `gil@tauex.tau.ac.il`. Supported by ERC starting grant 949499 and by the Israel Science Foundation grant 1569/18.

†Tel Aviv University. `talyankovitz@mail.tau.ac.il`. Supported by the Azrieli Faculty Fellowship.

ISSN 1433-8092

Contents

1 Introduction

Coding theory addresses the problem of communicating over an imperfect channel. Classically, the setting is as follows. Alice wishes to communicate a message m to Bob over a channel that can be tampered by an adversary. How should Alice encode m so that if the amount of errors is not excessive, Bob would be able to recover m? To this end, error-correcting codes were first introduced [Sha48]. Recall that a function $C\colon \Sigma^k \to \Sigma^n$ is an *error-correcting code* with distance δ if for every distinct $x, y \in \Sigma^k$, $\mathsf{dist}(C(x), C(y)) \geq \delta$, where dist is the relative Hamming distance.[1] The *rate* of the code C is given by $\rho = k/n$. Using an error-correcting code, Alice can encode her message $m \in \Sigma^k$ and send the resulting codeword $C(m)$. Assuming the fraction of errors is less than $\delta/2$, Bob can decode m from the received z by finding the codeword closest to z. When there is more than one possible message length, we consider a *code family*, which is a family of functions in which each function is a code, and there is one code per message length k. A code family is *asymptotically good* if both the rate and distance of every code in the family are uniformly bounded below by constants $\rho > 0$ and $\delta > 0$, respectively.

1.1 Locally decodable codes and locally correctable codes

Consider the scenario in which Bob is not interested in the entire original message m, but rather in a specific symbol m_i for some $i \in [k]$. A simple, though wasteful solution, is for Bob to decode the entire message m and ignore all symbols but for m_i. However, it is desirable to compute m_i by reading much fewer than n entries of z. *Locally decodable codes (LDC)* are a class of error-correcting codes that have this very strong decoding capability. Another scenario of interest is the one in which Bob needs to know a specific symbol of the *codeword* $C(m)_j$ for some $j \in [n]$, while reading as few symbols as possible. Codes that allow this are called *locally correctable codes (LCC)*. We turn to give the formal definition.

Definition 1.1 (Locally decodable codes (LDC))**.** *A code $C : \Sigma^k \to \Sigma^n$ is (q, δ, ε)*-locally decodable *if there exists a randomized algorithm D, called a* local decoder*, that is given $i \in [k]$ as input and an oracle access to $z \in \Sigma^n$, and has the following guarantee. For every $i \in [k]$, $m \in \Sigma^k$ and $z \in \Sigma^n$ such that $\mathsf{dist}(C(m), z) \leq \delta$ it holds that $\mathbf{Pr}[D^z(i) \neq m_i] \leq \varepsilon$. Moreover, D makes at most q queries to z.*

Definition 1.2 (Locally correctable codes (LCC))**.** *A code $C \subseteq \Sigma^n$ is (q, δ, ε)*-locally correctable *if there exists a randomized algorithm D, called a* local corrector*, that is given $j \in [n]$ as input and an oracle access to $z \in \Sigma^n$, and has the following guarantee. For*

[1] Note that what we call here distance δ is in many cases referred to as *relative* distance.

every $j \in [n]$, $c \in C$ *and* $z \in \Sigma^n$ *such that* $\mathsf{dist}(c, z) \le \delta$ *it holds that* $\mathbf{Pr}[D^z(j) \neq c_j] \le \varepsilon$. *Moreover,* D *makes at most* q *queries to* z*.*

We place z in the upper script in our notation $D^z(i)$ to stress that the number of symbols read from z by D is of importance. The parameter q is called the *query complexity*, and δ is the *local error decoding radius*, in the case LDC, and *local error correction radius* in the case of LCC. However, we also refer to δ, somewhat inaccurately, as the *local distance* of the code. From here on, we do not make any explicit reference to the "global" distance of a code and so we refer to the local distance simply as the distance. Throughout the paper, we only consider non-adaptive LDC and LCC, defined next. Informally, these are code in which the local decoder (or corrector) samples the entries to be read before the querying step takes place. Our results only hold for non-adaptive LDCs and LCCs. For ease of discussion, throughout the introduction we ignore the error parameter ε. More precisely, when stating our results, every LDC or LCC (both in the hypothesis as well as in the LDC or LCC guaranteed by the theorem) has constant error.

A brief history of LDC and LCC. Locally decodable codes were first explicitly defined by Katz and Trevisan [KT00]. However, codes with local guarantees have been used by complexity theorists even before (e.g., [BF90, GLR+91, GS92, BFNW93]) and have been around, implicitly, in the coding theory community almost from the get going [Ree53]. LDC, LCC, and related notions such as locally testable codes (LTC), were intensively studied by theoretical computer scientists motivated by PCPs [ALM+98, AS98, BFLS91, GS06], program checking [BLR90, Lip90, RS96], circuit lower bounds [Dvi11], derandomization [BFNW93, STV01, Tre03], and private information retrieval [CGKS95] to name a few. LDC and LCC are very related notions. Clearly, an LCC with a systematic encoding[2] is also an LDC and so, in particular, linear LCC induce LDC. Of note, it is not yet known in which scenarios LCC are strictly stronger objects compared to LDC.

An intensive research effort is devoted to the construction of local codes (see the excellent survey for LDC [Yek11]). Roughly, the literature can be partitioned to two. The first research path (see e.g., [Yek08, KY09, Efr12, DGY11] and references therein) has the goal of obtaining LDC or LCC with a given, small, number of queries, and an effort is made to maximize the rate while maintaining constant distance. The second research path, which has received much attention in recent years [KSY14, GKS13, HOW15, KMRS17, GKO+18], and is the focus of this paper, insists on asymptotically good codes and aims at minimizing the number of queries.

[2] An encoding from messages to codewords is called *systematic* if the symbols of each message are embedded in its mapped codeword.

It is known [KT00, Woo07] that asymptotically good LDC require $q = \Omega(\log n)$ queries. Whether this bound is tight is a fundamental, major open problem, regardless of explicitness. The Reed Muller code is perhaps the earliest non-trivial example of LDC and LCC. It can achieve query complexity n^{ν} for any desired constant $\nu > 0$. However, the rate deteriorates rapidly as $\nu \to 0$. In fact, up until the introduction of *multiplicity codes* by Kopparty, Saraf and Yekhanin [KSY14] no (non-trivial) LDC or LCC with rate higher than $1/2$ were known. Guo, Kopparty and Saraf [GKS13] introduced the notion of lifting of codes which gave a second high-rate LDC and LCC, also algebraic in nature. A combinatorial high-rate construction of an LCC was obtained by Hemenway, Ostrovsky and Wootters [HOW15] (see also [LW19]).

Despite this exciting sequence of works which allowed for better rate and introduced various interesting techniques, the above constructions all have query complexity $n^{\Theta(1)}$. The fact that three very different constructions were stuck at polynomial query complexity raised the question of whether $n^{o(1)}$-query asymptotically good LDC or LCC exist. This question was resolved in a seminal work by Kopparty, Meir, Ron-Zewi and Saraf [KMRS17] who obtained LCC with query complexity $q = 2^{\widetilde{O}(\sqrt{\log n})} = n^{o(1)}$. To obtain their result, the authors first observed that by instantiating multiplicity codes [KSY14] in a certain regime of parameters, one can get the stated query complexity q above albeit at the cost of having vanishing distance $\delta = 1/(\log n)^{\Theta(1)}$. Then, in order to get codes with constant distance, the authors invoked a distance amplification procedure due to Alon *et al.* [AL96, AEL95]. Kopparty *et al.* [KMRS17] showed that the AEL distance amplification procedure, which was originally introduced in the context of linear-time erasure codes, allows one to convert, in a black-box manner, an LCC with distance δ and query complexity q to an LCC with constant distance and query complexity $q_{\text{new}} = q \cdot \text{poly}(1/\delta)$. This more than sufficed for [KMRS17] as, in their setting, $q = (1/\delta)^{\omega(1)}$, and so the cost of the distance amplification is negligible. The LCC constructed in Kopparty *et al.* [KMRS17] are linear and thus yield LDC as well, and in fact in the same work the state-of-the-art LTC are constructed using the AEL distance amplification procedure.

1.2 Our contribution

Given the pivotal role of the AEL distance amplification procedure in the state-of-the-art constructions of LDC and LCC (as well as LTC) one is prompt to ask whether the $\text{poly}(1/\delta)$ multiplicative cost in query complexity is inherent. If such is the case, when aiming at $\text{poly}(\log n)$-query complexity, a requirement for constant distance can only be relaxed to distance $1/\text{poly}(\log n)$ which, although proved extremely useful [KMRS17], might be restrictive for obtaining better codes.

More generally, the natural question that is raised is to what extent the construction of asymptotically good LDC/LCC can be reduced to the non-asymptotically good variants as they, in turn, may admit low query constructions. The main contribution of this work is the first rate amplification procedure for linear LCC as we elaborate on next (see Section 1.2.1). As our second contribution, we obtain a significantly improved distance amplification procedure (see Section 1.2.2).

1.2.1 Rate amplification

It is unclear to us if rate can be amplified deterministically in general, regardless of locality, in any meaningful formalization. Puncturing is a coding-theoretic technique that allows one to obtain better rates. However, it only seems to work when tailored to specific codes with certain structure or, otherwise, using a randomized encoding. Nonetheless, our main contribution is a devising a rate amplification procedure for linear non-adaptive LCC. To the best of our knowledge, all known constructions of LCC are of this kind. Among these are Reed-Muller codes (and therefore also the Hadamard code) as well as codes obtained by lifting [GKS13], and Multiplicity codes.

Theorem 1.3 (Main result)**.** *Assume one has a non-adaptive linear $(q, \delta = \Omega(1))$-LCC with block-length n_0 having rate $\rho = \rho(n_0)$. Then, for every integers $\ell, c \geq 1$ such that $\ell^2 < c < \log n_0$, one can obtain a non-adaptive linear $(q_{\text{new}}, \delta_{\text{new}})$-LCC with block length $n \approx n_0^\ell$, having rate ρ_{new}, where*

$$\begin{aligned} q_{\text{new}} &= (cq)^{\text{poly}(\ell)}, \\ \delta_{\text{new}} &= (cq)^{-\text{poly}(\ell)}, \\ \rho_{\text{new}} &= 1 - (1-\rho)^\ell - O\left(\frac{\ell^2}{c}\right). \end{aligned}$$

Theorem 1.3, when invoked with $\ell \approx 1/\rho$ and $c \approx 1/\rho^2$, and combined with a distance amplification procedure, yields the following corollary.

Corollary 1.4. *Assume one has a family of constant distance non-adaptive linear LCC with rate $\rho(n) \geq \frac{1}{\sqrt{\log n}}$ and query complexity $q(n)$. Then, for every constant*[3] *$\alpha > 0$ one can obtain asymptotically good LCC with rate $1-\alpha$ on block length n with query complexity $q_{\text{new}} = (q(n)\log n)^{\text{poly}(1/\rho(n))}$.*

[3]The result holds also for sub-constant α, and the assumption is made only for simplicity. See Theorem 4.26 for the formal, more general, version.

1.2.2 Query-efficient distance amplification

The second result of this work is a significantly improved distance amplification procedure for LDC. Roughly speaking, we are able to reduce the $\mathrm{poly}(1/\delta)$ multiplicative factor in query complexity to the query complexity of an asymptotically good LDC on message length $1/\delta$. More precisely,

Theorem 1.5 (Query-efficient distance amplification; informal)**.** *Assume one has a block-length-n LDC with distance δ, constant rate, and query complexity q. Assume further one has a family of asymptotically good LDC where on message length k, the query complexity is q_k. Then, one can obtain asymptotically good LDC with query complexity* [4]

$$q_{\mathrm{new}} = q \cdot q_{O(1/\delta)} \cdot O(\log(1/\delta) \log n). \tag{1.1}$$

Note that by using a standard error-correcting code, which has $q_k = n = \Theta(k)$, Theorem 1.5 gives back the parameters of the AEL distance amplification procedure. However, one can do much better. Indeed, by using the state-of-the-art LDC [KMRS17] which has $q_k = 2^{\widetilde{O}(\sqrt{\log k})}$, one get $q_{\mathrm{new}} = q \cdot (1/\delta)^{o(1)} \log n$. More generally, Theorem 1.5 states that the lower the query complexity of the asymptotically good codes which one starts with is, the more query-efficient is the distance amplification. This "rich getting richer" type of result opens a path to recursive constructions as, indeed, several of our applications are based on. We stress that unlike the AEL distance amplification procedure, ours exploits the local *decodability* requirement and so it works for LDC but not for LCC. The only other technique in the literature that we are aware of that exploits the difference between decodability and correctability, and thus separates LDC from LCC in terms of techniques is matching vectors based constructions. We further remark that, for ease of discussion, Theorem 1.5 is stated without any reference to explicitness. Indeed, we currently lack satisfactory understanding of LDC in the more fundamental information-theoretic level. In any case, explicitness does not cost much in our reduction, and the only change in the theorem's statement when insisting on explicit reductions is replacing Equation (1.1) by roughly $q_{\mathrm{new}} = q \cdot q_{(1/\delta)^{1+\alpha}} \log n$ for any desired constant $\alpha > 0$.

We turn to draw several corollaries of Theorem 1.5, but first set the context. Given the Katz-Trevisan $\Omega(\log n)$ lower bound on the query complexity of asymptotically good LDC, and reassured by [KMRS17] that $n^{o(1)}$-query LDC exist, the next natural goal is to try and construct, or even more fundamentally, prove the existence of LDC with poly-logarithmic (or perhaps a more modest quasi poly-logarithmic $2^{\mathrm{poly}(\log\log n)}$) number of queries. With this goal in mind, the AEL distance amplification procedure allows one

[4] If the family of LDC in the hypothesis has sufficiently low error, the query complexity is even smaller $q_{\mathrm{new}} = q \cdot q_{O(1/\delta)} q_{O(\log(1/\delta))}$.

to relax her effort and construct LDC with distance $\delta = 1/\text{poly}(\log n)$ or slightly lower. Multiplicity codes are indeed a great example where such a relaxation of the distance requirement allows one to obtain much better query complexity. Using Theorem 1.5, we are able to obtain a reduction to LDC having exponentially lower distance $\delta = 1/\text{poly}(n)$.

Corollary 1.6 (Amplifying polynomially-small distance)**.** *Let $0 < \alpha < 1$ be an arbitrary constant. Assume there exists a family of LDC with distance $\delta = n^{-\alpha}$, rate $1 - 1/(\log n)^2$, and query complexity $q(n)$ for block length n. Then, for infinitely many n's, there exists an asymptotically good LDC on block-length n with query complexity $q_{\text{new}} = q(n)^{O(\log\log n)}$.*

Corollary 1.6 implies that for constructing asymptotically good LDC with $q = 2^{\text{poly}(\log\log n)}$ queries, it suffices to construct LDC with extremely poor distance $\delta = 1/\text{poly}(n)$ for the same asymptotic query complexity. In fact, we can even amplify extremely small distance $\delta = n^{-(1-o(1))}$ assuming the rate is slightly larger. One instantiation is as follows.

Corollary 1.7. *Let $c \geq 1$ be any constant. Assume there exists a family of LDC with distance $\delta = n^{-(1-\frac{1}{(\log\log n)^c})}$, rate $\rho = 1 - \frac{1}{(\log n)^{c+2}}$, and query complexity $q(n)$ for block-length n. Then, for infinitely many n's, there exists an asymptotically good LDC on block-length n having query complexity $q_{\text{new}} = q(n)^{O((\log\log n)^{c+1})}$.*

A third interesting application of Theorem 1.5 is when the distance to be amplified is larger than $1/\text{poly}(n)$, though still very small.

Corollary 1.8. *Let $\alpha < 1$ be an arbitrary constant. Assume there exists a distance $\delta = 2^{-(\log n)^\alpha}$ LDC having rate $1 - O(1/\log\log n)$, and query complexity $q(n)$ for block-length n. Then, for infinitely many n's, there exists an asymptotically good LDC on block length n with query complexity $q_{\text{new}} = q(n)^{O(\log\log\log n)}$.*

We conclude this section by noting that the Katz-Trevisan bound [KT00] holds also for sub-constant distance. Quantitatively, the query complexity of constant rate codes with distance δ is $\Omega(\log(\delta n/\log n))$. Thus, even for distance $n^{-\alpha}$, the $\Omega(\log n)$ lower bound holds.

2 Proof overview

In this section we give a brief and informal overview of the ideas that go into our proofs.

2.1 A characterization of non-adaptive linear LCC

To obtain our rate amplification procedure we lay a characterization of non-adaptive linear LCC. We remark that a very similar characterization was given by [KT00] for LDC, who defined the notion of smooth-codes.

Definition 2.1 (Smooth locally recoverable sets; simplified version)**.** *Let Σ, P be arbitrary sets. We say that $C \subseteq \Sigma^P$ is (q,τ)-*smooth locally recoverable *(SLR for short) if there exists a randomized algorithm* Rec*, called a* recovering procedure*, that when given as input $p \in P$ and an oracle access to $c \in C$, outputs $\mathsf{Rec}^c(p) = c_p$ by making at most q queries to c. Moreover, for every $c \in C$ and $p, r \in P$ (not necessarily distinct),*

$$\mathbf{Pr}[\mathsf{Rec}^c(p) \text{ queries } c_r] \leq \tau. \tag{2.1}$$

We will focus on SLR in which Σ is a field and C is a vector space over Σ. In such case we say that C is linear. Of course, it is trivial to construct a $(1,1)$-SLR. Indeed, simply query c_p and output the result. The challenge is to recover c_p without being able to "focus" on any particular entry. This is captured by Equation (2.1) where τ–the *smoothness parameter*–bounds the probability a given entry is allowed to be queried. The formal definition of SLR (see Definition 4.1) also allows the recovering procedure to output a special "failure" symbol $\perp$ with small probability. For ease of discussion, we ignore this here. We have the following easy claim showing that SLR yield LCC. As a result, linear SLR induce LDC.

Claim 2.2. *Let $C \subseteq \Sigma^P$ be a (q,τ)-SLR. Then, C is a (q,δ)-LCC with $\delta = \Omega\left(1/(q\tau|P|)\right)$.*

For the straightforward proof, see Section 4 and, in particular, Claim 4.2. We also have the following (less obvious) claim, showing that, assuming linearity and non-adaptiveness, the other direction also holds, namely, LCC yield SLR.

Claim 2.3. *Let $C \subseteq \Sigma^P$ be a non-adaptive linear (q,δ)-LCC. Then, C is a (linear) (q,τ)-SLR with $\tau = q/(\delta|P|)$.*

This claim and its proof correspond to Theorem 1 of [KT00] with the terminology of smooth-codes. For the more formal statement which also takes into account the error parameter and field size, see Claim 4.3. We remark here that for the proof of Claim 4.3, we construct a recovering procedure based on the local corrector of the given LCC. However, the key idea is to consider the distributions this local corrector induces while ignoring how it reconstruct the symbol after performing the queries.

Note that the lowest sensible value for τ is at about $q/|P|$. Indeed, this will be the case if each of the q queries is marginally uniform over P, and assuming nothing about the correlations between the queries. For such τ, if C is linear then, By Claim 2.2, it yields an LCC with $\delta = \Omega(1/q^2)$. The distance can then be amplified to constant using our distance amplification procedure to yield query complexity $q^{2+o(1)}$ (or using AEL's procedure to get $\mathrm{poly}(q)$ queries).

2.1.1 Dual SLR and their induced SLR

By Claim 2.2 and Claim 2.3, every linear SLR is an LCC, and every linear non-adaptive LCC is a linear SLR. Our rate amplification procedure works for non-adaptive linear SLR, and thus for any non-adaptive linear LCC. In order to amplify the rate of such an SLR, we show that the dual of every non-adaptive linear SLR has a certain structure, which we use to amplify the rate.

Working with dual of codes in the context of LDC or LCC is a very natural approach, and has been explored previously (e.g., [KS07, BIR08]), but to the best of our knowledge, the definition of dual SLR as given below is new. We start by setting some notation. Let P be a set, $\mathbb{F}$ a finite field, and $\mathbb{F}^P$ the set of all functions $\{f : P \to \mathbb{F}\}$. Note that $\mathbb{F}^P$ has a natural $\mathbb{F}$-vector space structure. We consider the natural inner product $\langle \cdot, \cdot \rangle : \mathbb{F}^P \times \mathbb{F}^P \to \mathbb{F}$ that is defined, for $f, g \in \mathbb{F}^P$, by $\langle f, g \rangle = \sum_{p \in P} f(p)g(p)$. For $f \in \mathbb{F}^P$ we denote $|f| = |P \setminus f^{-1}(0)|$. For $p \in P$ define $\mathcal{F}_p = \{f \in \mathbb{F}^P \mid f(p) \neq 0\}$.

The following definition captures the structural properties of the dual of an SLR, which we need for the rate amplification.

Definition 2.4 (Dual SLR; simplified version)**.** *Let P be a set, $\mathbb{F}$ a field. Let $\mathcal{D} = \{D_p \mid p \in P\}$ be a collection of distributions, where for each $p \in P$, $\mathsf{supp}(D_p) \subseteq \mathcal{F}_p$. Set $S \triangleq \bigcup_{p \in P} \mathsf{supp}(D_p)$. The collection $\mathcal{D}$ is said to be a (q, τ, ρ)-*dual SLR *provided the following holds:*

1. *$|f| \leq q$ for all $f \in S$.*

2. *For every pair of distinct $p, r \in P$, it holds that*

$$\Pr_{f \sim D_p}[f(r) \neq 0] \leq \tau.$$

3. *Last, $\dim \operatorname{Span}(S) \leq (1 - \rho)|P|$.*

We call q the *query complexity* of the dual SLR, τ its *smoothness* and ρ its *rate*. The linear subspace $S^\perp$ of $\mathbb{F}^P$ is called the *induced SLR* from $\mathcal{D}$. As the name suggests, the induced SLR $S^\perp$ is indeed an SLR. More precisely, it is a $(q-1, \tau)$ SLR with rate ρ (see Lemma 4.6). It is for the class of dual-induced SLR that we are able to devise our rate amplification procedures. Let p be a prime power. As an example, one can directly show that, say, the two-dimensional Reed-Muller code over $\mathbb{F}_p$ with total-degree $p-2$ is an induced SLR from a $(q = p-1, \tau = \frac{1}{p+1}, \rho = \frac{1}{2} - o(1))$-dual SLR. As mentioned, any linear non-adaptive LCC is a linear SLR, and thus induces a dual SLR.

2.2 Rate amplification for dual-induced SLR

For simplicity, we describe our rate amplification procedure only for $\ell = 2$, where ℓ is as in the notation of Theorem 1.3. We briefly explain how to handle larger ℓ's in Section 2.2.3. Assume $\mathcal{D}$ is a (q, τ, ρ)-dual SLR on $\mathbb{F}^P$ where the rate ρ is the parameter we wish to amplify. Consider the mapping $\Phi : (\mathbb{F}^P)^2 \to \mathbb{F}^{P^2}$ that maps a pair of functions $f_1, f_2 \in \mathbb{F}^P$ to the function $\Phi(f_1, f_2) : P^2 \to \mathbb{F}$ given by $\Phi(f_1, f_2)(p_1, p_2) = f_1(p_1)f_2(p_2)$. Note that this is simply the tensor product.

We now show how to convert our poor-rate dual SLR $\mathcal{D}$ to a new dual-SLR with a better rate. Formally, consider the (q_2, τ_2, ρ_2)-dual SLR $\mathcal{D}^2 = \{D_p^2 \mid p \in P^2\}$, where for every $p = (p_1, p_2) \in P^2$, the distribution D_p^2 is defined as follows. To sample from D_p^2, sample $f_1 \sim D_{p_1}$, $f_2 \sim D_{p_2}$ independently, and return $\Phi(f_1, f_2)$. That $q_2 \le q^2$ is straightforward, and that the new rate $\rho_2 \ge 1 - (1 - \rho)^2$ can be shown using the bilinearity of Φ (see Claim 4.10). As for the smoothness, we prove (see Lemma 4.12) that for every $p, r \in P^2$,

$$\mathbf{Pr}\left[\Phi(f_1, f_2)(r) \neq 0\right] \le \tau^{\Delta(p,r)}, \tag{2.2}$$

where $\Delta(p, r)$ is the non-relative Hamming distance between p and r. In particular, for $r \neq p$, we get the bound $\tau_2 \le \tau$.

Note that as the world is now squared, a bound on the smoothness of merely τ is poor. However, by Equation (2.2), for most points $r \in P^2$ we in fact have a better bound of τ^2. It is only those points of distance one from p that cause the smoothness from "squaring" and, as a result, deteriorate the distance of the induced LCC (recall Claim 2.2). A natural approach would be to "zero out" the problematic points. To make "zero out" formal, for a set $S \subseteq P^2$, let $\nu_S : P^2 \to \mathbb{F}$ be such that $\nu_S(r) = 0$ if $r \in S$ and $\nu_S(r) = 1$ otherwise. Now, instead of $\Phi(f_1, f_2)$ consider the function $\widehat{\Phi}(f_1, f_2) = \Phi(f_1, f_2) \cdot \nu_L$ where

$$L = \{r \in P^2 \mid \Delta(p, r) = 1 \text{ and } \Phi(f_1, f_2)(r) \neq 0\}.$$

By construction, Equation (2.2) implies that the smoothness of dual SLR defined using $\widehat{\Phi}$ is bounded by τ^2. Unfortunately, however, we can no longer guarantee anything about the rate ρ_2 which, recall, is the parameter we set out to improve.

Our key idea is to construct carefully chosen functions in addition to those from $S^2 = \cup_p \mathsf{supp}(D_p^2)$ which allows us to zero out the problematic points while deteriorating the rate only slightly. To describe our solution, let R be a partition of P^2, where each part has size $c + 1$ for some parameter c to be chosen later on. We denote the part, or class, in R containing an element $p \in P^2$ by $[p]$ and write $(p) = [p] \setminus \{p\}$ for the *open class* of p. For each $p \in P^2$ define the function $f_p : P^2 \to \mathbb{F}$ by $f_p(r) = 1$ if $r \in [p]$ and $f_p(r) = 0$ otherwise. We adjoin all $\frac{|P|^2}{c+1}$ functions $\mathcal{L}_R = \{f_p \mid p \in P^2\}$ to S^2 by considering $\mathcal{L}_R^2 = \mathrm{Span}(S^2) + \mathrm{Span}(\mathcal{L}_R)$. That is, our dual-induced SLR is redefined to be $(\mathcal{L}_R^2)^\perp$ rather than $(S^2)^\perp$. This has some cost in rate, but a manageable one. Indeed, note that $\dim(\mathcal{L}_R^2) \leq (1 - \rho_2 + \frac{1}{c+1})|P^2|$. Thus, for sufficiently large c, the rate loss incurred by adding the functions in $\mathcal{L}_R$ can be made small. The advantage we get by adjoining these functions is that we can now zero out any point r we wish by using the points in its open class (r). Indeed, for every $f \in (\mathcal{L}_R^2)^\perp$ and $r \in P^2$ we have $f(r) = -\sum_{w\in(r)} f(w)$. Note that, on top of the $\frac{1}{c+1}$ loss in rate, we expect to pay a multiplicative c cost in query complexity as $|(r)| = c$.

To be more precise, for $p \in P^2$, we define a distribution $(D_R^2)_p$, which will avoid using the problematic points given by L above, as follows. To sample a function $f \sim (D_R^2)_p$ proceed as follows:

1. Sample $g \sim D_p^2$ and let $L = \{r \in P^2 \mid \Delta(p, r) = 1 \text{ and } g(r) \neq 0\}$.

2. For every $r \in L$ and $w \in (r)$ sample $h_{r,w} \sim D_w^2$.

3. Return
$$f = g\nu_L + \sum_{r\in L} g(r) \sum_{w\in(r)} \frac{h_{r,w}\nu_{\{w\}}}{h_{r,w}(w)}. \tag{2.3}$$

Observe that the first summand $g\nu_L$ in Equation (2.3) is the attempt we started with. However, using the partition R, instead of simply zeroing out L (which prevents us from arguing about the rate ρ_2), for every $r \in L$ that was zeroed out, we go over each of the points w in its open class and add a carefully chosen linear combination of the "freshly" sampled functions $\{h_{r,w} \sim D_w^2\}$ to $g\nu_L$ so as to guarantee that $f \in \mathcal{L}_R^2$ (see Claim 4.21).

There is one technical issue the reader should be aware of. It might not be the case that $f(p) \neq 0$, which is the basic requirement of dual SLR. Indeed, while $g(p) \neq 0$ it might be the case $h_{r,w}(p) \neq 0$ for one or more pairs (r, w) as well. As a result, a cancellation may occur, causing $f(p) = 0$. This is where we make use of the $\perp$ symbol in the formal definition of dual SLR. Before outputting f, we check that this cancellation has not occurred and otherwise return $\perp$.

2.2.1 Axis evasive partitions

The above scheme can be implemented with any partition R. However, not every partition will enable us to improve the smoothness. Informally, we would like the partition to have the property that the union of open classes taken over the set of points of distance one from a given point p, is composed of points that are mostly of distance two from one another. To make this precise, we note that the set of points of distance one from a given point p is contained in the union of a horizontal and a vertical line. We refer to such lines, collectively, as axis-parallel lines. The following definition abstracts what we need from the partition so to argue about the smoothness.

Definition 2.5. *Let P be a set. A partition R of P^2 is said to be (c, s)-*axis evasive *if*

1. *For every $p \in P^2$, $|(p)| = c$.*
2. *For every pair of axis-parallel lines ℓ, ℓ' (possibly equal),*
$$\left|\ell \cap \bigcup_{p \in \ell'} (p)\right| \leq s.$$
3. *For every $p \in P^2$ and every axis-parallel line ℓ, $|[p] \cap \ell| \leq 1$.*

We show that by using a (c, s)-axis evasive partition, the dual SLR defined above has smoothness $\tau_2 = O(csq\tau^2)$ (see Claim 4.23). The reader should think of c, s as constants (or slightly sub-constants) and $q \ll \tau^{-1}$, and so $\tau_2 \approx \tau^2 \ll \tau$.

2.2.2 Constructing axis-evasive partitions

Assume $|P| = m$ is an odd prime power, and let c be an even integer such that $c+1 \mid m+1$. Under these assumptions, we are able to give an explicit algebraic construction of (c, s)-axis evasive partitions of P^2 where $s = O(c^2)$ (see Section 5.2). Intuitively, as we want to construct a partition that "breaks" axis-parallel-ness, rotation would be a natural approach. Indeed, for our construction, we identify P with the finite field $\mathbb{F}_m$ and P^2 with $\mathbb{F}_{m^2}$. For every choice of $\alpha \in \mathbb{F}_{m^2} \setminus \mathbb{F}_m$, one can identify $\mathbb{F}_{m^2}$ with $\mathbb{F}_m + \alpha\mathbb{F}_m$. So, informally, $\mathbb{F}_m$ and $\alpha\mathbb{F}_m$ are the horizontal and vertical axes, respectively. To formalize the intuition of rotation, we take an element β of order $c + 1$ in the multiplicative group of $\mathbb{F}_{m^2}$. Being a cyclic group, and since $c + 1 \mid m + 1 \mid m^2 - 1$, such an element exists. Multiplication by β can, informally, be thought of as a rotation by a $\frac{1}{c+1}$ angle. We take the partition of $\mathbb{F}_{m^2} \setminus \{0\}$ according to the cosets of $\langle\beta\rangle$ - the subgroup generated by β (and do not worry much about the origin). We show that, with this construction, properties (1) and (2) of Definition 2.5 are satisfied. Property (3), however, does not and so we need to make a certain modification of the construction to resolve this. We do not delve into the required alternation of the construction here.

2.2.3 Rate amplification for dimension higher than two

Our basic rate amplification procedure can be easily generalized to any $\ell > 2$. On the other hand, our distance-efficient rate amplification procedure is designed for $\ell = 2$.

To go from $\ell = 2$ to higher powers, we more or less do the obvious thing, namely, apply the dual SLR construction iteratively, where in each iteration we square the size of the previously obtained set. The only technical issue is that the divisibility by $c+1$ requirement is not maintained throughout the process. Indeed, 2 is the only nontrivial common factor of $m+1$ and m^2+1. To overcome this, we truncate the resulted set, slightly reducing its size from m^2 to a prime m' that is divisible by $c+1$. The truncation deteriorates the rate and so we would like $m' \approx m^2$. Such prime m' is guaranteed to exist by the Siegel–Walfisz Theorem [Sie35, Wal36] that refines Dirichlet's theorem on primes in arithmetic progressions.

2.3 Query-efficient distance amplification

The AEL distance amplification procedure was originally based on expander graphs [AL96, AEL95]. Kopparty *et al.* [KMRS17] used samplers instead - a point of view that we find fruitful for our needs. Informally, an (ε, δ)-sampler is a bipartite graph on vertex set $L \cup R$ with the following property. For every $T \subseteq R$, having density $\mu(T)$, all but δ-fraction of the left vertices have $\mu(T) \pm \varepsilon$ fraction of their neighbours in T (see Definition 3.1). For simplicity, we assume regularity with left-degree d and right degree D.

Given a code with poor distance δ, AEL amplifies the distance to constant using an (ε, δ)-sampler where, for the reduction, ε is taken to be constant. The AEL procedure has a Dd multiplicative cost in query complexity. Prior works used either expander graphs or "balanced" samplers, namely, samplers with $|L| = |R|$ and $D = d$. With this choice, the lowest possible degree is $d = \Theta(1/(\varepsilon^2\delta))$, which in turn yields a $\Theta((1/\delta)^2)$ multiplicative cost in query complexity.

Our improved distance amplification procedure is based on two simple ideas. Our variant has a lower cost in query complexity: Instead of a Dd factor, our variant has roughly $q_D q_d$ multiplicative cost where, recall, q_k is the query complexity of an asymptotically good LDC on message length k. Our variant also makes use of samplers, and when instantiated with a balanced sampler, the cost is roughly $q_d^2 = q_{1/\delta}^2$. Our second idea allows us to essentially get rid of the square (which is crucial for obtaining our corollaries). It is known that by working with unbalanced samplers, in which $|L| \gg |R|$, one can obtain (ε, δ)-samplers with a much lower left-degree $d = O(\log(1/\delta)/\varepsilon^2)$. We note that, for the original AEL procedure, working with unbalanced samplers cannot yield a significant improvement. Indeed, to achieve this saving in left-degree, the ratio $|L|/|R| = \Omega(1/(\delta \log(1/\delta)))$ which in turn implies $D = |L|d/|R| = \Omega(1/\delta)$. This then only gives a quadratic improvement over AEL. When instantiated with our variant, unbalanced samplers yield query complexity roughly $q_{1/\delta} q_{\log(1/\delta)}$.

3 Preliminaries

Notations and conventions. Unless otherwise stated, all logarithms are taken to the base 2. We denote by the set of natural numbers (of course, including 0). For an integer $c \geq 1$, we let $[c] = \{1, 2, \ldots, c\}$. For ease of readability, we avoid the use of floor and ceiling. This does not affect the stated results. For two strings x, y of equal length over a common alphabet, we denote by $\mathsf{dist}(x, y)$ their relative hamming distance, namely, the fraction of indices on which they disagree. Let $A \neq \emptyset$ be an ambient (finite) set. For $B \subseteq A$, we denote by $\mu(B)$ the density of B in A, namely, $\mu(B) = |B|/|A|$.

Let $G = (V, E)$ be an undirected graph with maximal degree D. Assume that the neighbours of every node $v \in V$ are labeled by distinct numbers from $1, \ldots, \deg(v)$. We define the neighbourhood function $\Gamma_G : V \times [D] \to (V \times [D]) \cup \{\perp\}$ as follows. For $v \in V$ and $i \in [\deg(v)]$ we let $\Gamma_G(v, i) = (u, j)$ where u is the i'th neighbour of v and v is the j'th neighbour of u. For $i \in [D] \setminus [\deg(v)]$ the function is defined to be $\perp$ (though this is only for the sake of formality. We will never use such input i). If G is clear from context we sometimes omit it from the subscript. When interested only on the node u as above and not on j, we make a slight abuse of notation and write $\Gamma(v, i)$ when referring to u. Last, we write $\Gamma(v)$ for the set of all neighbours of v.

3.1 Samplers

Our distance amplification procedure makes use of samplers. These are bipartite graphs with a certain pseudo-random property. Let $G = (L, R, E)$ be a bipartite graph. We say G is *left-regular* if all nodes in L have the same degree.

Definition 3.1 ([BR94]). *Let $0 < \varepsilon, \delta < 1$. A bipartite graph $G = (L, R, E)$ is an* (ε, δ)-sampler *if for every subset $T \subseteq R$, for all but δ-fraction of vertices $v \in L$ it holds that*

$$\left| \frac{|\Gamma(v) \cap T|}{|\Gamma(v)|} - \mu(T) \right| \leq \varepsilon.$$

We will be working with "unbalanced" samplers. These are samplers with $|L| \gg |R|$. The state-of-the-art constructions of these samplers rely on their connection to randomness seeded extractors. We refer the interested reader to the excellent survey by Goldreich [Gol11] for more information. When working with samplers, it is rather typical that the bipartite graph is left-regular, that is, the degree of all vertices in L is the same. A small additional technical property we need is that the degree of every vertex in R is close to the average right-degree. We make use of the following theorem which gives (non-explicit) samplers with near-optimal parameters having the above properties with respect to the degrees. We give a proof sketch for completeness.

Theorem 3.2. *There exists a universal constant $c_{\mathsf{samp}} \geq 1$ such that the following holds. For all integers ℓ, r and all $\varepsilon > 0$, $1/2 > \delta > 0$ for which $\ell \geq \frac{r}{\delta \log(1/\delta)}$, there exists a left-regular (ε, δ)-sampler $G = ([\ell], [r], E)$ with left-degree $d = c_{\mathsf{samp}} \cdot \log(1/\delta)/\varepsilon^2$. Moreover, provided that $\log r < 1/(\delta\varepsilon^2)$, every right vertex has degree in $[D/2, 2D]$ where $D = \ell d/r$ is the average right degree.*

For the proof we need the following well-known lemma.

Lemma 3.3. *For every integers $1 \leq k \leq n$ with $\frac{k}{n} = \delta \leq \frac{1}{2}$ it holds that*

$$\sum_{i=0}^{k} \binom{n}{i} \leq 2^{H(\delta)n},$$

where $H(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function.

Proof sketch for Theorem 3.2. The proof is via the probabilistic method, where for every left vertex we choose d neighbours independently and uniformly at random, and independently across all left vertices (note that in the above we allow for parallel edges, but if that troubles the reader, that can be avoided as well in the regime of interest $d \ll r$ by arguing that the probability of a right neighbor to be selected more than once is small. In any case, our distance amplification procedure works just as well with parallel edges). Fix $T \subseteq [r]$. For $v \in [\ell]$ let F_v be the indicator random variables that is 1 if and only if $||\Gamma(v) \cap T|/d - \mu(T)| > \varepsilon$. By the Chernoff bound, $\mathbf{Pr}[F_v] \leq e^{-\Omega(\varepsilon^2 d)}$. Fix $S \subseteq [\ell]$ with $|S| = \delta\ell$. The probability that for all vertices $v \in S$ it holds that $F_v = 1$ is bounded above by $e^{-\Omega(\varepsilon^2 d \cdot \delta\ell)}$. By taking the union bound over all S and T, we get that except with probability

$$2^r \binom{\ell}{\delta\ell} e^{-\Omega(\varepsilon^2 d\delta\ell)} \leq 2^{r + H(\delta)\ell - c\varepsilon^2 d\delta\ell}, \tag{3.1}$$

the sampled graph is an (ε, δ)-sampler. Note that the last inequality follows by Lemma 3.3, where $c > 0$ is some constant. By taking $c_{\mathsf{samp}} \geq 5/c$, one can verify (using that $H(x) \leq 2x\log(1/x)$ for all $x \leq 1/2$) that the right hand side in Equation (3.1) is bounded by $1/4$.

As for the moreover part, again, by the Chernoff bound, the probability that there exists a right vertex which has degree outside $[D/2, 2D]$ is bounded above by $re^{-\Omega(\ell d/r)}$, and this is bounded by $1/4$ by our choice of parameters and by taking c_{samp} large enough. □

We now turn to state the parameters of the explicit construction of samplers that we use.

Theorem 3.4 ([RVW01], [Gol11]). [5] *For every constant* $\Delta > 0$ *there exists a constant* $c = c(\Delta) \geq 1$ *such that the following holds. For all* $\varepsilon > 0$, $\delta > 0$ [6], *there exists an explicit left-regular* (ε, δ)*-sampler* $G = ([\ell], [r], E)$*. The left-degree of* G *is* $d = ((1/\varepsilon)\log(1/\delta))^c$*. Furthermore, the average right degree* $D = \ell d/r$ *of* G *is in* $[D', 2D']$ *where*

$$D'(\Delta, \varepsilon, \delta) = \frac{d}{2} \cdot \left(\frac{2}{\delta}\right)^{\Delta+1}. \tag{3.2}$$

3.2 Codes

We make use of "standard" error-correcting codes. In this section we gather some known results we use.

Theorem 3.5 (The Gilbert-Varshamov bound). *Let* Σ *be a set of size* $|\Sigma| = q$*. For every* $n \in$*, and* $0 \leq \delta \leq 1 - \frac{1}{q}$ *there exists a code of block-length* n *over* Σ*, with distance at least* δ *and rate* $r \geq 1 - H_q(\delta)$*. Furthermore, if* q *is a prime power and* $\Sigma = \mathbb{F}_q$*, there exists a linear code over* Σ *with rate* $r \geq 1 - H_q(\delta) - g(n)$*, where* $g(n) = O(\frac{1}{n})$*.*

Lemma 3.6. *There exists a constant* $\beta_0 > 0$ *such that the following holds. Let* n *be an integer and* $\frac{1}{\log n} < \beta < \beta_0$ *. Let* $\Sigma = \mathbb{F}_q$ *for* $q \geq 2$ *a prime power. Then, there exists an explicit linear code of block-length* n *over* Σ *with rate* $1 - \beta$ *and relative distance* β^3*.*

The existence of these codes follows from a special case of the Zyablov bound [Zya71], but for completeness we describe a construction which attains the stated parameters. For the proof, we make use of the following easy claim whose proof is omitted.

Claim 3.7. *For every* $x \in (0, 1/2]$ *and* $q \geq 2$, $H_q(x) \leq x \log_q(\frac{q^3}{x})$*.*

Proof of Lemma 3.6. The proof is obtained by taking the code concatenation of two codes, a Reed-Solomon code and a Gilbert-Varshamov code. Let p be the least prime such that $p \geq n$. Recall that $p \leq 2n$. Set C_{RS} to be the Reed-Solomon code over $\mathbb{F}_p$ of block length $n_{\mathrm{RS}} = \frac{(1-\beta^{1.1})n}{\log_q p}$ and message length $k_{\mathrm{RS}} = (1 - \beta^{1.1})n_{\mathrm{RS}}$. So, C_{RS} has rate $1 - \beta^{1.1}$ and relative distance at least $\beta^{1.1}$. Now take C_{GV} to be a linear code of the following parameters. The message length is $k_{\mathrm{GV}} = \log_q p$, the block length is $\frac{1}{1-\beta^{1.1}} k_{\mathrm{GV}}$ (and therefore the rate is $1 - \beta^{1.1}$), and the relative distance is at least $\beta^{1.4}$. We wish to invoke

[5] The sampler in [RVW01] has a mild requirement on ε which we state the theorem without, as it is explained in [Gol11] how this requirement can be relaxed, by using a more recent extractor.

[6] The sampler in [RVW01] has a number of edges z that is a power of two. We state the theorem for a general z as one can take the subgraph of only part of the left vertices, and get a sampler in which δ is at most doubled.

Theorem 3.5 so as to prove that such a code exists. To this end, we must verify that $1 - H_q(\beta^{1.4}) - g(n) \geq 1 - \beta^{1.1}$. Indeed, by Claim 3.7, we have that

$$1 - H_q(\beta^{1.4}) - g(n) \geq 1 - \beta^{1.4} \log_q \left(\frac{q^3}{\beta^{1.4}} \right) - g(n) \geq 1 - \beta^{1.1},$$

where the last inequality holds for all sufficiently small $\beta \geq 0$, and since $g(n) = O(\frac{1}{n})$ and $\beta \geq \frac{1}{\log n}$, by assumption.

Note that C_{GV} is not explicit as Theorem 3.5 only guarantees existence of a code with the stated parameters. However, as the block-length of C_{GV} is $O(\log n)$, such a code can be found by an exhaustive search on generating matrices, in time $2^{O((\log n)^2)}$. To improve on that, we remark that the code C_{GV} can also be found by going only over a limited family of generating matrices (see [GRS12]), and this can be done in time poly(n).

Consider the concatenated code $C_{\text{RS}} \circ C_{\text{GV}}$. It has block length $n_{\text{RS}} \cdot n_{\text{GV}} = n$, rate $(1 - \beta^{1.1})^2$ which is at least $1 - \beta$ for all small enough $\beta > 0$, and relative distance $\beta^{1.2}\beta^{1.4} \geq \beta^3$, completing the proof. □

4 Rate amplification for dual-induced SLR

In this section we introduce the notion of *smooth locally recoverable sets (SLR)* which under non-adaptive and linearity assumptions is shown to be equivalent to LCC. We consider a certain class of SLR, to which we call *dual-induced SLR*. These are SLR that are obtained by the dual of certain structured sets. The structure of these dual-SLR sets allows us to devise a rate amplification procedures for them. Informally, dual-SLR are sets of tuples (or linear spaces of vectors if the alphabet over which we are working is a field) in which every given entry of a tuple in the set can be recovered using only few queries *and* in a "smooth" manner, which is to say that the distribution of every query has high min-entropy.

Definition 4.1 (Smooth locally recoverable sets (SLR))**.** *Let Σ, P be arbitrary non-empty sets. We say that $C \subseteq \Sigma^P$ is (q, τ, ε)-*smooth locally recoverable *(SLR for short) if there exists a randomized algorithm* Rec*, called a* recovering procedure*, that is given as input $p \in P$ and an oracle access to $c \in C$. The recovering procedure outputs either an element of Σ or a symbol $\perp$ which is assumed not to be in Σ. The algorithm* Rec *has the following properties:*

- *For every $(c, p) \in C \times P$, $\mathsf{Rec}^c(p)$ makes at most q queries to c.*
- *For every $c \in C$ and $p, r \in P$ it holds that*

$$\mathbf{Pr}[\mathsf{Rec}^c(p) \text{ queries } c_r] \leq \tau.$$

- *For every* $(c,p) \in C \times P$*, the random variable* $\mathsf{Rec}^c(p) \in \{c_p, \perp\}$*, and*

$$\mathbf{Pr}\left[\mathsf{Rec}^c(p) = \perp\right] \leq \varepsilon.$$

We assume that for every $p \in P$ *whether* $\mathsf{Rec}^c(p) = \perp$ *is independent of* c*, and that it is never the case that* $\mathsf{Rec}^c(p)$ *queries* c_p*. When* Σ *is a field and* C *is a linear subspace of* Σ^P*, we say that* C *is* linear*. In this case, the* rate *of* C *is defined as* $\dim(C)/|P|$*. We will mostly consider* non-adaptive SLR*. These are SLR in which the joint distribution of queries is independent of* c*.*

We remark that the notion of SLR is very similar to the notion of smooth-codes of [KT00] for LDC. We now have the following easy claim showing that SLR yield LCC and, assuming linearity, LDC.

Claim 4.2. *Let* $C \subseteq \Sigma^P$ *be a* (q,τ,ε)*-SLR. Then, for every* $\varepsilon' > 0$*,* C *is a* $(q,\delta,\varepsilon+\varepsilon')$*-LCC with* $\delta = \varepsilon'/(q\tau|P|)$*. As a consequence, if* C *is also linear then* C *is a* $(q,\delta,\varepsilon+\varepsilon')$*-LDC.*

Proof. To show that C is an LCC, we devise a local corrector for C. Given an oracle access to $c \in \Sigma^P$, and $p \in P$ as input, the local corrector computes $z = \mathsf{Rec}^c(p)$. If $z = \perp$ then the local corrector returns some arbitrary element of Σ, and otherwise return z. To analyze this local corrector, let $c' \in \Sigma^P$ be such that $\mathsf{dist}(c,c') \leq \delta|P|$. Denote $B = \{p \in P \mid c_p \neq c'_p\}$. Note that conditioned on $\mathsf{Rec}^c(p) \neq \perp$, the local corrector returns c_p successfully if all q queries do not fall into B. The probability that any given query falls into B is bounded above by $\tau|B|$ and so, by the union bound, the probability that some query falls into B is bounded above by $\tau|B|q \leq \varepsilon'$. This proves that C is a $(q,\delta,\varepsilon+\varepsilon')$-LCC. Note that linear LCC are systematic and so every linear LCC induces an LDC. □

In fact, for linear LCCs, the other direction also holds, meaning that such an LCC is an SLR, as we have in the following claim.

Claim 4.3. *Let* $C \subseteq \mathbb{F}^P$ *be a linear non-adaptive* (q,δ,ε)*-LCC where 1-*$\varepsilon > 1/|\mathbb{F}|$*. Then,* C *is a* (q,τ,ε')*-SLR with* $\tau = q/(\delta|P|)$ *and* $\varepsilon' = 0$*.*

Before we prove the claim, we need to state the following easy to verify fact.

Fact 4.4. *Let* $L \subseteq \mathbb{F}^P$ *be a linear subspace, let* $p \in P$*, let* $Q \subseteq P$ *and let* $x \in \mathbb{F}^{|Q|}$*. Then, one of the following cases must hold.*

1. *There is at most one* $\alpha \in \mathbb{F}$ *for which there exists some* $v \in L$ *satisfying* $v(Q) = x$ [7] *and* $v(p) = \alpha$*;*

[7] For a set $A = \{a_1, \ldots, a_{|A|}\}$, $v(A)$ denotes the sequence $(v(a_1), \ldots, v(a_{|A|}))$.

2. For every $\alpha \in \mathbb{F}$ there is an equal number of $v \in L$ for which $v(Q) = x$ and $v(p) = \alpha$.

In particular, either no function (even randomized) of $v(Q)$ can predict $v(p)$ with probability more than $1/|\mathbb{F}|$, when $v \in L$ is randomly chosen uniformly, or $v(Q)$ always determines $v(p)$.

With that, we now prove Claim 4.3 [8].

Proof for Claim 4.3. To show that C is an SLR, we devise a recovering procedure Rec for it, based on the local corrector promised by it being an LCC. Let D [9] be such a local corrector. For every point $p \in P$, we construct a sequence of disjoint sets $Q_1^p, \ldots, Q_{m_p}^p \subseteq P$, where for every i, $c(Q_i^p)$ determines $c(p)$ while satisfying $|Q_i^p| \leq q$, and $m_p \geq \delta|P|/q$. On $p \in P$ and oracle access to $c \in C$, the procedure $\mathsf{Rec}^c(p)$ acts by uniformly choosing $i \in [m_p]$, querying $c(Q_i^p)$, and using it to deduce and output $c(p)$. The correctness of the result of Rec is immediate (since Rec always succeeds, $\varepsilon' = 0$), and indeed the number of queries is no more than q. Since the sets are disjoint, the probability that a point is queried is no more than $\tau = q/(\delta|P|)$. It only remains to show how the assumed sets can be constructed, to conclude that C is a (q, τ, ε')-SLR, which we now turn to do.

For every $p \in P$, $Q_1^p, \ldots, Q_{m_p}^p$ are constructed as follows. Set $Q_0^p = \emptyset$. For $i = 1, 2, \ldots$, set $S_i = Q_0^p \cup \cdots \cup Q_{i-1}^p$. If $|S_i| > \delta|P|$, halt and set $m_p = i - 1$. Otherwise, it holds that for every $c \in C$, for every modification of the coordinates in S_i to some erroneous values, the decoder D correctly outputs $c(p)$ with probability at least $1 - \varepsilon$. An equivalent description of this case is the following: for every $c \in C$ and $z : S_i \to \mathbb{F}$, define $c_z \in \mathbb{F}^P$ such that for $r \notin S_i$, $c_z(r) = c(r)$, and for $r \in S_i$, $c_z(r) = z(r)$; the decoder D chooses a set of queries $Q \subseteq P$, $|Q| \leq q$, according to a distribution and applies a function f_Q on $c_z(Q)$; with probability at least $1 - \varepsilon$, $f_Q(c_z(Q)) = c(p)$. Since Q is chosen in a manner independent of c and z, one can verify that this implies that there exists some fixed Q for which when $c \in C$ and $z : \mathbb{F} \to S_i$ are chosen randomly in a uniform manner, with probability at least $1 - \varepsilon$ (this time over the choice of c and z), $f_Q(c_z(Q)) = c(p)$. Therefore, we can define another function f'_Q that only gets $c(Q \setminus S_i)$, chooses z at random, and outputs $f_Q(c_z(Q))$. If $c \in C$ is chosen uniformly at random, $f'_Q(Q \setminus S_i) = c(p)$ with probability at least $1 - \varepsilon > 1/|F|$. By Fact 4.4, this implies that $c(Q \setminus S_i)$ determines $c(p)$, for every $c \in C$. We therefore set $Q_i^p = Q \setminus S_i$, and proceed to next i. As this process only halts when $|S_i| > \delta|P|$, and for every i $|S_i| \leq q(i-1)$, we have that indeed $m_p \geq \delta|P|/q$. Further note that by the choice of each Q_i^p, the sets $Q_1^p, \ldots, Q_{m_p}^p$ are disjoint, as required. $\square$

[8] This proof is inspired by the proof of [KT00] of their Theorem 1 and by a proof in [ZDa] for a different claim.

[9] We make the slight assumption that $D^c(p)$ never directly queries $c(p)$. If however it does, then similarly C can be shown to be a (q, τ, ε')-SLR for $\tau = \frac{1}{(\delta|P|/q)-1}$.

4.1 Dual SLR and their induced SLR

Our construction of SLR sets will be via constructing and analyzing sets which we call *dual SLR* sets. The SLR will then be induced from these dual SLR. We start by setting some notation. Let P be a non-empty finite set and $\mathbb{F}$ a finite field. We make use of the standard notation $\mathbb{F}^P$ to denote the set of all functions $\{f : P \to \mathbb{F}\}$. Note that $\mathbb{F}^P$ has a natural $\mathbb{F}$-vector space structure where addition is point-wise, namely, for every $f, g \in \mathbb{F}^P$ and $a, b \in \mathbb{F}$ we have that $af + bg \in \mathbb{F}^P$ is defined by $(af + bg)(p) = af(p) + bg(p)$ for all $p \in P$. We consider the natural inner product map $\langle \cdot, \cdot \rangle : \mathbb{F}^P \times \mathbb{F}^P \to \mathbb{F}$ that is defined, for $f, g \in \mathbb{F}^P$, by $\langle f, g \rangle = \sum_{p \in P} f(p)g(p)$. Given $f \in \mathbb{F}^P$, we let $f^\perp = \{g \in \mathbb{F}^P \mid \langle f, g \rangle = 0\}$. Note that $f^\perp$ is a linear subspace of $\mathbb{F}^P$. More generally, given a set $S \subseteq \mathbb{F}^P$ we define the linear subspace $S^\perp = \bigcap_{f \in S} f^\perp$. For $f \in \mathbb{F}^P$ we denote $|f| = |P \setminus f^{-1}(0)|$.

For the sake of readability, the field $\mathbb{F}$ and the set P will be omitted from the notation that we are about the introduce in this section. Both will be clear from context. For $p \in P$ define $\mathcal{F}_p = \{f \in \mathbb{F}^P \mid f(p) \neq 0\}$. Informally, a dual SLR is a collection of distributions over $\mathbb{F}^P$, one for each point $p \in P$. The distribution D_p, that corresponds to p, outputs a function $g \in \mathcal{F}^P$. We think of g as "passing through" p. We also allow D_p to output a special "failed symbol" $\perp$ with some small probability. A dual SLR has the guarantee that g does not pass through many other points, namely, $|g|$ is bounded, and that the dimension of all functions that can be sampled, when considering all distributions D_p, $p \in P$, is also bounded. Perhaps most importantly is the requirement that for every other fixed $r \in P$, the sampled $g \sim D_p$ is likely to have the property that $g \notin \mathcal{F}_r$. Formally,

Definition 4.5 (Dual SLR)**.** *Let P be a set, $\mathbb{F}$ a field. Let $\mathcal{D} = \{D_p \mid p \in P\}$ be a collection of distributions, where for each $p \in P$, $\mathsf{supp}(D_p) \subseteq \mathcal{F}_p \cup \{\perp\}$. Denote $S = \bigcup_{p \in P} \mathsf{supp}(D_p)$. Let $\mathcal{L}$ be a linear subspace of $\mathbb{F}^P$ such that $S \subseteq \mathcal{L} \cup \{\perp\}$. The pair $(\mathcal{D}, \mathcal{L})$ is said to be a $(q, \tau, \varepsilon, \rho)$-*dual SLR *on $\mathbb{F}^P$ provided the following holds:*

1. *$|g| \leq q$ for all $g \in S \setminus \{\perp\}$.*
2. *For every pair of distinct $p, r \in P$ (not necessarily distinct), it holds that*
$$\Pr_{g \sim D_p} [g(r) \neq 0 \mid g \neq \perp] \leq \tau.$$
3. *For every $p \in P$, $\mathbf{Pr}[D_p = \perp] \leq \varepsilon$.*
4. *$\dim(\mathcal{L}) \leq (1 - \rho)|P|$.*

The linear subspace $\mathcal{L}^\perp$ of $\mathbb{F}^P$ is called the induced SLR *from the dual SLR $(\mathcal{D}, \mathcal{L})$. The parameter τ of a dual-SLR is referred to as its* smoothness*.*

Let $(\mathcal{D}, \mathcal{L})$ be a dual SLR. We turn to show that, as the name suggests, the induced SLR $\mathcal{L}^{\perp}$ is indeed an SLR.

Lemma 4.6. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}, \mathcal{L})$ be $(q, \tau, \varepsilon, \rho)$-dual SLR on $\mathbb{F}^P$. Then the induced SLR $\mathcal{L}^{\perp}$ is a $(q-1, \tau, \varepsilon)$-SLR. Furthermore, $\mathcal{L}^{\perp}$ is linear and has rate ρ or larger.*

Proof. The moreover part readily follows since $\mathcal{L}^{\perp}$ is a linear subspace of $\mathbb{F}^P$ and since

$$\dim(\mathcal{L}^{\perp}) = |P| - \dim(\mathcal{L}) \geq \rho|P|.$$

We describe a recovering procedure for $\mathcal{L}^{\perp}$, namely, a randomized algorithm that is given an oracle access to $f \in \mathcal{L}^{\perp}$ as well as a point $p \in P$ as input. The recovering procedure proceeds as follows:

1. Sample $g \sim D_p$. If $g = \perp$ return $\perp$; Otherwise,
2. Query f on all points $Q = \{r \in P \setminus \{p\} \mid g(r) \neq 0\}$.
3. Return
$$-\frac{1}{g(p)} \sum_{r \in Q} g(r) f(r).$$

The query complexity of Rec is bounded by $q-1$ as $|Q| = |g| - 1 \leq q - 1$. The probability that $\perp$ is returned is at most ε by construction. We turn to prove that $\mathsf{Rec}^f(p) \in \{f(p), \perp\}$. By construction, $\mathsf{Rec}^f(p) = \perp$ if and only if $g = \perp$. Assume than that $g \neq \perp$, hence, $g \in \mathsf{supp}(D_p) \subseteq \mathcal{L}$. As $f \in \mathcal{L}^{\perp}$ we have that $0 = \langle f, g \rangle$, and so

$$0 = \sum_{r \in P} g(r) f(r) = g(p) f(p) + \sum_{r \in Q} g(r) f(r).$$

As $g \in \mathsf{supp}(D_p) \subseteq \mathcal{F}_p$ we have $g(p) \neq 0$, and so

$$f(p) = -\frac{1}{g(p)} \sum_{r \in Q} g(r) f(r) = \mathsf{Rec}^f(p).$$

To conclude the proof, we turn to analyze the smoothness of Rec. First, note that, by construction, f is never queried on p itself. Consider then any $r \in P \setminus \{p\}$. Conditioned on $g \neq \perp$, the function f is queried on r if and only if $g(r) \neq 0$. Thus,

$$\mathbf{Pr}\left[f(r) \text{ is queried}\right] = \mathop{\mathbf{Pr}}_{g \sim D_p}\left[g(r) \neq 0 \mid g \neq \perp\right] \leq \tau,$$

and the proof follows. $\square$

We now show that the opposite holds as well, that any linear, non-adaptive, SLR induces a dual-SLR.

Lemma 4.7. *Let P be a set, $\mathbb{F}$ a field, and let $C \subseteq \mathbb{F}^P$ be a linear non-adaptive (q, τ, ε)-SLR with rate ρ. Then for some set $\mathcal{D}$, $(\mathcal{D}, C^{\perp})$ is a $(q+1, \tau, \varepsilon, \rho)$-dual SLR*

Proof. Let Rec be a recovering procedure promised by C being a SLR. Assume that Rec uses R random bits. For every point $p \in P$, denote by $\mathcal{E}_p$ the set of choices of the random bits $r \in \{0,1\}^R$ for which $\mathsf{Rec}^c(p) = \perp$ (for any $c \in C$). Note that $\mu(\mathcal{E}_p \subseteq \{0,1\}^R) \leq \varepsilon$.

For any $p \in P$ and $r \in \{0,1\}^R$, $r \notin \mathcal{E}_p$, denote by $Q_{p,r} \subseteq P$ the set of query locations which $\mathsf{Rec}(p)$ makes when r is the choice of randomness. Define a function $f_{p,r} : C \to \mathbb{F}$ such that $f(c)$ is the output of $\mathsf{Rec}^c(p)$ when fixing its randomness to r and note that $f_{p,r}(c)$ only depends on $\{c(w) \mid w \in Q_{p,r}\}$, and that $f_{p,r}(c) = c(p)$. Since C is linear, one can easily verify that $f_{p,r}$ is a linear map. Therefore, for some $u_{p,r} \in \mathbb{F}^P$, $f_{p,r}(c) = \langle u_{p,r}, c \rangle$ for every $c \in C$, where $u_{p,r}(w) = 0$ if $w \notin Q_{p,r}$. We have that $c(p) = \langle u_{p,r}, c \rangle$ for every $c \in C$. If we define a function $g_{p,r} \in \mathbb{F}^P$ such that

$$g_{p,r}(w) = \begin{cases} -1, & w = p; \\ u_{p,r}(q), & w \neq p, \end{cases}$$

it follows that $g_{p,r} \in C^{\perp}$. Note that $|g_{\{p,r\}}| \leq q+1$.

For every $p \in P$, define D_p to be the following distribution. To sample from D_p, draw $r \in \{0,1\}^R$ uniformly at random. If $r \in \mathcal{E}_p$, output $\perp$; otherwise, output $g_{p,r}$. Set $\mathcal{D} = \{D_p \mid p \in P\}$ and $\mathcal{L} = C^{\perp}$. It follows trivially by the definitions that $(\mathcal{D}, \mathcal{L})$ is a $(q+1, \tau, \varepsilon, \rho)$-dual SLR. □

4.2 Rate amplification for dual-induced SLR

In this section we describe our first rate amplification procedure for SLR that are induced by dual SLR. Unlike the previous section, it will be more convenient to explicitly state within the notation the set P over which we are working as we will be dealing with several such sets. The field $\mathbb{F}$, however, remains suppressed from the notation as it remains fixed in all SLR under consideration. We start by defining the following map of functions.

Definition 4.8. *Let P be a set and $\mathbb{F}$ a field. For an integer $\ell \geq 1$ we define the map $\Phi \colon (\mathbb{F}^P)^\ell \to \mathbb{F}^{P^\ell}$ as follows. Let $g_1, \ldots, g_\ell \in \mathbb{F}^P$. The function $\Phi(g_1, \ldots, g_\ell) : P^\ell \to \mathbb{F}$ is defined by*

$$\Phi(g_1, \ldots, g_\ell)(p_1, \ldots, p_\ell) = \prod_{i=1}^{\ell} g_i(p_i)$$

for every $(p_1, \ldots, p_\ell) \in P^\ell$.

Observe that Φ is multi-linear. Further, when $\ell = 2$ and g_1, g_2 are viewed as vectors rather than functions, Φ is the outer product of the vectors.

Definition 4.9. *Let P be a set, $\mathbb{F}$ a field. Let $\mathcal{L}^P$ be a linear subspace of $\mathbb{F}^P$. For an integer $\ell \geq 1$, we define*

$$\mathcal{L}^{P^\ell} = \mathrm{Span}\left\{\Phi(g_1, \ldots, g_\ell) \mid g_1, \ldots, g_\ell \in \mathcal{L}^P\right\}.$$

Claim 4.10. *With the notation of Definition 4.9,*

$$\dim\left(\mathcal{L}^{P^\ell}\right) \leq \left(\dim\left(\mathcal{L}^P\right)\right)^\ell.$$

Proof. Let $B = \{g_1, \ldots, g_b\}$ be a basis for $\mathcal{L}^P$, where $b = \dim(\mathcal{L}^P)$. Define

$$B' = \{\Phi(h_1, \ldots, h_\ell) \mid (h_1, \ldots, h_\ell) \in B^\ell\}.$$

Observe that to prove the claim, it suffices to show that for every $f_1, \ldots, f_\ell \in \mathcal{L}^P$ it holds that $\Phi(f_1, \ldots, f_\ell) \in \mathrm{Span}(B')$. As $f_1, \ldots, f_\ell \in \mathcal{L}^P$, for every $i \in [\ell]$ we can write $f_i = \sum_{j=1}^b \lambda_{i,j} g_j$ with $\lambda_{i,j} \in \mathbb{F}$. We have that

$$\begin{aligned}\Phi(f_1, \ldots, f_\ell) &= \Phi\left(\sum_{j_1=1}^b \lambda_{1,j_1} g_{j_1}, \ldots, \sum_{j_\ell=1}^b \lambda_{\ell,j_\ell} g_{j_\ell}\right) \\ &= \sum_{j_1 \ldots, j_\ell \in [b]} \left(\prod_{t=1}^\ell \lambda_{t,j_t}\right) \cdot \Phi(g_{j_1}, \ldots, g_{j_\ell}),\end{aligned}$$

where the last equality follows by the multi-linearity of Φ. □

Definition 4.11. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be $(q, \tau, \varepsilon, \rho)$-dual SLR. Let $\ell \geq 1$ be an integer. For $p \in P^\ell$ we define the distribution $D_p^{P^\ell}$ as follows. Write $p = (p_1, \ldots, p_\ell)$. To sample an element from $D_p^{P^\ell}$ proceed as follows:*

1. *Sample $g_1 \sim D_{p_1}^P, \ldots, g_\ell \sim D_{p_\ell}^P$ independently.*
2. *If there exists $i \in [\ell]$ such that $g_i = \perp$, return $\perp$; Otherwise*
3. *Return $\Phi(g_1, \ldots, g_\ell)$.*

The collection of distributions $\{D_p^{P^\ell} \mid p \in P^\ell\}$ is denoted by $\mathcal{D}^{P^\ell}$.

We have the following lemma.

Lemma 4.12. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. Let $\ell \geq 1$ be an integer and $\mathcal{D}^{P^\ell}$ as in Definition 4.11. Then, for every $p, r \in P^\ell$,*

$$\Pr_{g \sim D_p^{P^\ell}}\left[g(r) \neq 0 \mid g \neq \perp\right] \leq \tau^{\mathsf{dist}(p,r)}.$$

Proof. Write $p = (p_1, \ldots, p_\ell)$, $r = (r_1, \ldots, r_\ell)$. By Definition 4.11, conditioned on $g \neq \perp$ we have that $g = \Phi(g_1, \ldots, g_\ell)$ with $g_i \sim D^P_{p_i}$ for each $i \in [\ell]$ independently. Thus, $g(r) \neq 0$ is the event

$$\Phi(g_1, \ldots, g_\ell)(r_1, \ldots, r_\ell) = \prod_{i=1}^{\ell} g_i(r_i) \neq 0.$$

By the independence of $g_1, \ldots, g_\ell$, and since we are working over a field $\mathbb{F}$ (and so a product is nonzero if and only if each of the terms is nonzero), we get

$$\Pr_{g \sim D^{P^\ell}_p} [g(r) \neq 0 \mid g \neq \perp] = \prod_{i=1}^{\ell} \Pr_{g_i \sim D^P_{p_i}} [g_i(r_i) \neq 0 \mid g_i \neq \perp] . \tag{4.1}$$

Let $T = \{i \in [\ell] \mid p_i \neq r_i\}$. As $\mathcal{D}^P$ is a $(q, \tau, \varepsilon, \rho)$-dual SLR, for each $i \in T$ it holds that

$$\Pr_{g_i \sim D^P_{p_i}} [g_i(r_i) \neq 0 \mid g_i \neq \perp] \leq \tau.$$

Substituting to Equation (4.1), we get

$$\Pr_{g \sim D^{P^\ell}_p} [g(r) \neq 0 \mid g \neq \perp] \leq \tau^{|T|},$$

which completes the proof. $\square$

Definition 4.13. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. For an integer $\ell \geq 1$ let $\mathcal{L}^{P^\ell}$, $\mathcal{D}^{P^\ell}$ be as in Definition 4.9 and Definition 4.11, respectively. We denote the pair $(\mathcal{D}^{P^\ell}, \mathcal{L}^{P^\ell})$ by $(\mathcal{D}^P, \mathcal{L}^P)^\ell$.*

Proposition 4.14. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. Then, for every integer $\ell \geq 1$ we have that $(\mathcal{D}^P, \mathcal{L}^P)^\ell$ is a $(q_\ell, \tau_\ell, \varepsilon_\ell, \rho_\ell)$-dual SLR, where*

$$\begin{aligned} q_\ell &\leq q^\ell, \\ \tau_\ell &\leq \tau, \\ \varepsilon_\ell &\leq \ell\varepsilon, \\ \rho_\ell &\geq 1 - (1 - \rho)^\ell. \end{aligned}$$

Proof. First note that for every $p \in P^\ell$, the distribution $D^{P^\ell}_p$ is supported on $\mathcal{F}^{P^\ell}_p \cup \{\perp\}$. Indeed, if we write $p = (p_1, \ldots, p_\ell)$ then, conditioned on $g \neq \perp$, we have that $g = \Phi(g_1, \ldots, g_\ell)$ where $g_i \in D^P_{p_i}$. Thus,

$$g(p) = \Phi(g_1, \ldots, g_\ell)(p_1, \ldots, p_\ell) = \prod_{i=1}^{\ell} g_i(p_i) \neq 0.$$

Moreover, by Definition 4.9,

$$\bigcup_{p \in P^\ell} \mathsf{supp}(D_p^{P^\ell}) \subseteq \mathcal{L}^{P^\ell} \cup \{\perp\}.$$

We turn to show that $q_\ell \le q^\ell$. Let $p = (p_1, \ldots, p_\ell) \in P^\ell$ and consider any $g \in \mathsf{supp}(D_p^{P^\ell})$. By Definition 4.11, if $g \neq \perp$ then $g = \Phi(g_1, \ldots, g_\ell)$ where $g_i \in \mathsf{supp}(D_{p_i}^P) \setminus \{\perp\}$. Now, for every $r = (r_1, \ldots, r_\ell) \in P^\ell$ we have that

$$g(r) \neq 0 \qquad \Longleftrightarrow \qquad \prod_{i=1}^{\ell} g_i(r_i) \neq 0.$$

Since $\mathbb{F}$ is a field, the above is equivalent to $g_i(r_i) \neq 0$ for all $i \in [\ell]$. Hence there are at most q^ℓ points $r \in P^\ell$ for which $g(r) \neq 0$, and so $q_\ell \le q^\ell$.

The bound on the smoothness readily follows by Lemma 4.12. Indeed, consider any pair of distinct $p, r \in \mathbb{F}^{P^\ell}$. We have that $\mathsf{dist}(p, r) \ge 1$ and so, by Lemma 4.12,

$$\mathop{\mathbf{Pr}}_{g \sim D_p^{P^\ell}} [g(r) \neq 0 \mid g \neq \perp] \le \tau^{\mathsf{dist}(p,r)} \le \tau. \tag{4.2}$$

To bound the probability that $\perp$ is returned, note that the event $D^{P^\ell} = \perp$ holds only if for some $i \in [\ell]$, $g_i = \perp$. Hence, by the union bound, $\mathbf{Pr}[D_p^{P^\ell} = \perp] \le \ell\varepsilon$. We conclude the proof by bounding the dimension of $\mathcal{L}^{P^\ell}$. By assumption, $\dim(\mathcal{L}^P) \le (1-\rho)|P|$. Claim 4.10 then implies that

$$\dim\left(\mathcal{L}^{P^\ell}\right) \le \left(\dim\left(\mathcal{L}^P\right)\right)^\ell \le ((1-\rho)|P|)^\ell = (1-\rho)^\ell |P^\ell|.$$

□

Discussion on the smoothness $\tau_\ell = \tau$. The downside of the rate amplification procedure that was given in this section is that τ_ℓ does not decrease with ℓ (which is bad as, recall, we wish τ to be small as, by Claim 4.2, the distance δ of the resulted LCC is proportional to $1/\tau$). Indeed, with the notation of Proposition 4.14, $\tau_\ell = \tau$. By examining the proof and Lemma 4.12 one natural idea is to consider an SLR not over the entire set P^ℓ but on some subset of it which is a code with distance, say, $d > 1$. This will indeed guarantee that for every two points p, r we have $\mathsf{dist}(p, r) \ge d$ and so the bound in Equation (4.2) will be τ^d rather than τ. While natural, this idea fails to yield better parameters as the rate-loss incurred by using a code (even an MDS) is larger than the improvement on the rate guaranteed via the rate amplification procedure.

In the next sections we give a more elaborate rate amplification procedure (that is based on the one that was given in this section) in which τ does decrease with ℓ. Roughly,

$\tau_\ell = (q \cdot \log|P|)^{\mathrm{poly}(\ell)}\tau^\ell$, and so there is a slight loss in the smoothness, which the reader should think as negligible. The query complexity q_ℓ as well as the rate ρ_ℓ and ε_ℓ are all slightly worse than those obtained in the above rate amplification procedure and so the two techniques are incomparable.

4.3 Distance-efficient rate amplification

Let P be a set, and R a partition of P^2. We denote the part containing p by $[p]_R$ or $[p]$ when R is clear from context. We call $(p) = [p] \setminus \{p\}$ the *open class* of p. For a set $A \subseteq P^2$ we let $(A) = \cup_{p \in A}(p)$. Given $p \in P$ we say that $\{p\} \times P \subseteq P^2$ is *vertical line* and $P \times \{p\}$ is a *horizontal line*. Horizontal and vertical lines are referred to as *axis-parallel lines*, and we denote the set of such lines by

$$\mathcal{X} = \bigcup_{p \in P} \{\{p\} \times P, P \times \{p\}\}.$$

For a point $p = (p_1, p_2) \in P^2$ we denote $S_p = (\{p_1\} \times P) \cup (P \times \{p_2\}) \setminus \{p\}$. That is, S_p is the set of points in P^2 of distance exactly 1 from p. Key to our distance-efficient rate amplification procedure is a partition of the "square" P^2 with certain properties.

Definition 4.15 (Axis-evasive partitions)**.** *Let P be a set. A partition R of P^2 is said to be* (c, s)-axis evasive *if*

1. *For every $p \in P^2$, $|(p)| \le c$.*
2. *For every $\ell, \ell' \in \mathcal{X}$ (possibly equal), $|\ell' \cap (\ell)| \le s$.*
3. *For every $p \in P^2$ and $\ell \in \mathcal{X}$, $|[p] \cap \ell| \le 1$.*

In Section 5 we study such partitions. We prove their existence with certain parameters and give explicit constructions. In this section, however, we work with abstract axis-evasive partitions and analyze our rate amplification procedure with respect to the parameters c, s of the axis-evasive partition as well as the number of parts which we typically denote by t.

Claim 4.16. *Let $p, p' \in P^2$ (possibly equal). Then,*

$$|\{r \in S_p \mid (r) \cap S_{p'} \neq \emptyset\}| \le 4s.$$

Proof. Note that each of $S_p, S_{p'}$ is a subset of the union of two axis-parallel lines. Thus, to prove the claim, it suffices to show that for every $\ell, \ell' \in \mathcal{X}$, not necessarily distinct,

$$|\{r \in \ell \mid (r) \cap \ell' \neq \emptyset\}| \le s.$$

Let $r_1, \dots, r_t \in \ell$ be such that $(r_i) \cap \ell' \neq \emptyset$. Note that for every distinct $i, j \in [t]$ it holds that $((r_i) \cap \ell') \cap ((r_j) \cap \ell') = \emptyset$. Indeed, since R is a partition, if $((r_i) \cap \ell') \cap ((r_j) \cap \ell') \neq \emptyset$ then $r_i \in [r_j]$, but this implies that $|\ell \cap [r_j]| \geq 2$ in contradiction axis evasiveness. Thus,

$$R = \bigcup_{i=1}^{t} ((r_i) \cap \ell')$$

is a disjoint union of size t. However, $R \subseteq (\ell) \cap \ell'$, and so $t \leq |R| \leq |(\ell) \cap \ell'| \leq s$. $\square$

Definition 4.17. *Let P be a set, $\mathbb{F}$ a field. Let R be a (c, s)-axis evasive partition of P^2. For every $p \in P^2$ define the function $g_{[p]} : P^2 \to \mathbb{F}$ as follows:*

$$g_{[p]}(r) = \begin{cases} 1, & r \in [p]; \\ 0, & \textit{otherwise}. \end{cases}$$

We define $\mathcal{L}_R = \{g_{[p]} \mid p \in P^2\}$.

Definition 4.18. *Let P be a set, $\mathbb{F}$ a field. For $S \subseteq P$ define the function $\nu_S : P \to \mathbb{F}$ by*

$$\nu_S(r) = \begin{cases} 0, & r \in S; \\ 1, & \textit{otherwise}. \end{cases}$$

For ease of readability, when S is a singleton $S = \{p\}$, we write ν_p instead of $\nu_{\{p\}}$.

With the notations and definitions above, we are ready to start developing our second rate amplification procedure. We start with the following.

Definition 4.19. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. Let $\mathcal{L}^{P^2}$ be as in Definition 4.9. Let R be a (c, s)-axis evasive partition of P^2. We define for every $p \in P^2$ the distribution $(D_R^{P^2})_p$ as follows. To sample u from $(D_R^{P^2})_p$:*

1. *Sample $g \sim D_p^{P^2}$.*
2. *If $g = \perp$ return $\perp$; Otherwise, denote $L = \{r \in S_p \mid g(r) \neq 0\}$ and proceed as follows.*
3. *For every $r \in L$ and $w \in (r)$ sample $h_{r,w} \sim D_w^{P^2}$.*
4. *If there exist $r \in L$ and $w \in (r)$ such that either $h_{r,w} = \perp$ or $h_{r,w}(p) \neq 0$ return $\perp$. Otherwise return*
$$u = g\nu_L + \sum_{r \in L} g(r) \sum_{w \in (r)} \frac{h_{r,w}\nu_w}{h_{r,w}(w)}. \tag{4.3}$$

Note that, upon reaching Step (4), u is well-defined as $h_{r,w}(w) \neq 0$ for all $r \in L$ and $w \in (r)$. We denote the collection of distributions $\{(D_R^{P^2})_p \mid p \in P^2\}$ by $\mathcal{D}_R^{P^2}$.

We start by analyzing the function u that is given by Equation (4.3) above.

Claim 4.20. *With the notation of Definition 4.19, if $\perp$ is not returned then $u \in \mathcal{F}_p$.*

Proof. As $\perp$ was not returned, for every $r \in L$ and $w \in (r)$ it holds that $h_{r,w} \neq \perp$ and $h_{r,w}(p) = 0$. Substituting to Equation (4.3), we get

$$u(p) = g(p)\nu_L(p) = g(p) \neq 0,$$

where the second equality holds as $p \notin L$ and the last inequality follows since $g \in \mathsf{supp}(D_p^{P^2}) \setminus \{\perp\}$. $\square$

Claim 4.21. *With the notation of Definition 4.19, if $\perp$ is not returned then $u \in \mathcal{L}^{P^2} + \mathcal{L}_R$.*

Proof. Take $f \in (\mathcal{L}^{P^2} + \mathcal{L}_R)^\perp$. To prove the claim, it suffices to show that $\langle u, f\rangle = 0$. Indeed, this would imply $u \in ((\mathcal{L}^{P^2} + \mathcal{L}_R)^\perp)^\perp = \mathcal{L}^{P^2} + \mathcal{L}_R$. As $u \neq \perp$ we have that $g \neq \perp$. Note that

$$\langle g\nu_L, f\rangle = \langle g, f\rangle - \sum_{r\in L} g(r)f(r).$$

Since $g \in \mathsf{supp}(D_p^{P^2})$ we get that $g \in \mathcal{L}^{P^2}$. However, $f \in (\mathcal{L}^{P^2} + \mathcal{L}_R)^\perp \subseteq (\mathcal{L}^{P^2})^\perp$, implying $\langle g, f\rangle = 0$. Thus,

$$\langle g\nu_L, f\rangle = -\sum_{r\in L} g(r)f(r). \tag{4.4}$$

Now, fix $r \in L$ and $w \in (r)$. By Definition 4.19, as $u \neq \perp$ we have that $h_{r,w} \neq \perp$ and so $h_{r,w} \in \mathcal{L}^{P^2}$. However, by the above, $f \in (\mathcal{L}^{P^2})^\perp$ and so $\langle h_{r,w}, f\rangle = 0$. Thus,

$$\langle h_{r,w}\nu_w, f\rangle = \langle h_{r,w}, f\rangle - h_{r,w}(w)f(w) = -h_{r,w}(w)f(w).$$

Therefore, for every fixed $r \in L$ one has that

$$\begin{aligned}\Big\langle \sum_{w\in(r)} \frac{h_{r,w}\nu_w}{h_{r,w}(w)}, f\Big\rangle &= \sum_{w\in(r)} \Big\langle \frac{h_{r,w}\nu_w}{h_{r,w}(w)}, f\Big\rangle \\ &= \sum_{w\in(r)} \frac{1}{h_{r,w}(w)} \langle h_{r,w}\nu_w, f\rangle \\ &= -\sum_{w\in(r)} f(w). \end{aligned} \tag{4.5}$$

Now, $f \in (\mathcal{L}^{P^2} + \mathcal{L}_R)^\perp \subseteq (\mathcal{L}_R)^\perp$ whereas $g_{[r]} \in \mathcal{L}_R$, and so

$$0 = \langle f, g_{[r]}\rangle = \sum_{w\in[r]} f(w).$$

Substituting this to Equation (4.5), we get

$$\left\langle \sum_{w\in(r)} \frac{h_{r,w}\nu_w}{h_{r,w}(w)}, f\right\rangle = f(r).$$

Therefore,

$$\begin{aligned}\left\langle \sum_{r\in L} g(r) \sum_{w\in(r)} \frac{h_{r,w}\nu_w}{h_{r,w}(w)}, f\right\rangle &= \sum_{r\in L} g(r)\left\langle \sum_{w\in(r)} \frac{h_{r,w}\nu_w}{h_{r,w}(w)}, f\right\rangle \\ &= \sum_{r\in L} g(r)f(r).\end{aligned}$$

The above equation together with Equation (4.4) yield $\langle u, f\rangle = 0$. □

Claim 4.22. *With the notation of Definition 4.19, for every $p \in P^2$,*

$$\mathbf{Pr}[(D_R^{P^2})_p = \perp] \le 18csq\tau^2 + 2cq\varepsilon.$$

Proof. First, the probability that $g = \perp$ is bounded by ε. Similarly, the probability that for any specific $r \in L$ and $w \in (r)$, $h_{r,w} = \perp$ is bounded by ε. Thus, by the union bound, and since $|L| \le 2q - 1$ and $|(r)| \le c$, we have that expect with probability $(1 + (2q-1)c)\varepsilon \le 2qc\varepsilon$, the sampling process above will result in $u \ne \perp$.

To complete the analysis, we turn to bound the probability that $h_{r,w}(p) = 0$ for some $r \in L$ and $w \in (r)$. Let $L = \{r_1, \ldots, r_{|L|}\}$. While the random variables in L may be dependent, marginally, it holds that for every $i \in [|L|]$ and every fixed $r \in S_p$, $\mathbf{Pr}[r_i = r] \le \tau$. With this notation, by Definition 4.19, $(D_R^{P^2})_p = \perp$ only if there exist $i \in [|L|]$ and $w \in (r_i)$ such that $h_{r_i,w}(p) \ne 0$.

For a fixed $r \in S_p$ define the event $\mathcal{E}_r$ in which there exists $w \in (r)$ such that $h_{r,w}(p) \ne 0$, (when conditioned on $h_{r,w} \ne \perp$). Note that this event is with respect to the randomness of sampling $h_r = \{h_{r,w} \mid w \in (r)\}$ whereas r is fixed. By the union bound,

$$\mathop{\mathbf{Pr}}_{h_r}[\mathcal{E}_r] \le \sum_{w\in(r)} \mathop{\mathbf{Pr}}_{h_{r,w}}[h_{r,w}(p) \ne 0 \mid h_{r,w} \ne \perp].$$

Observe first that $w \ne p$. Indeed, as $r \in S_p$, both r and p are on some common axis-parallel line $\ell \in \mathcal{X}$. Thus, $w = p$ would imply $|[r] \cap \ell| \ge 2$ which stands in contradiction to the definition of axis-evasiveness. Consider $w \in (r) \setminus S_p$. As $w \ne p$ we have that $\mathsf{dist}(w,p) = 2$. By Lemma 4.12, as $h_{r,w} \sim D_w^{P^2}$ we have that

$$\mathop{\mathbf{Pr}}_{h_{r,w}}[h_{r,w}(p) \ne 0 \mid h_{r,w} \ne \perp] \le \tau^2.$$

If, on the other hand, $w \in (r) \cap S_p$ then $\mathsf{dist}(w,p) = 1$, and Lemma 4.12 then implies that

$$\mathop{\mathbf{Pr}}_{h_{r,w}}[h_{r,w}(p) \ne 0 \mid h_{r,w} \ne \perp] \le \tau.$$

As $|(r)| \leq c$ we conclude that

$$\mathop{\mathbf{Pr}}_{h_r}[\mathcal{E}_r] \leq c\tau^2 + \tau|(r) \cap S_p|.$$

Fix $i \in [|L|]$ and consider the random variable r_i. The above equation, together with $|(r_i)| \leq c$, yields

$$\begin{aligned}\mathop{\mathbf{Pr}}_{r_i,h_{r_i}}[\mathcal{E}_{r_i}] &\leq \mathop{\mathbf{Pr}}_{r_i,h_{r_i}}[\mathcal{E}_{r_i} \mid (r_i) \cap S_p = \emptyset] + \mathop{\mathbf{Pr}}_{r_i,h_{r_i}}[\mathcal{E}_{r_i} \mid (r_i) \cap S_p \neq \emptyset] \mathop{\mathbf{Pr}}_{r_i}[(r_i) \cap S_p \neq \emptyset] \\ &\leq c\tau^2 + (c\tau^2 + c\tau) \mathop{\mathbf{Pr}}_{r_i}[(r_i) \cap S_p \neq \emptyset]. \end{aligned} \tag{4.6}$$

Consider now the set $B = \{r \in S_p \mid (r) \cap S_p \neq \emptyset\}$. As R is (c, s)-axis evasive, Claim 4.16 implies that $|B| \leq 4s$, and so

$$\mathop{\mathbf{Pr}}_{r_i}[(r_i) \cap S_p \neq \emptyset] = \mathbf{Pr}[r_i \in B] \leq 4s\tau.$$

Substituting to Equation (4.6), we get $\mathbf{Pr}[\mathcal{E}_i] \leq 9cs\tau^2$. The proof then follows by taking the union bound over all $i \in [|L|]$ as, indeed, $|L| = 2q - 1$. $\square$

Claim 4.23. *With the notation of Definition 4.19, for every pair of distinct $p, r \in P^2$,*

$$\mathop{\mathbf{Pr}}_{u \sim (D_R^{P^2})_p}[u(r) \neq 0 \mid u \neq \perp] \leq 10csq\tau^2.$$

Proof. By Equation (4.3), u is a linear combination of the (sampled) functions $g\nu_L$, $\{h_{r,w}\nu_w\}$. To prove the claim, we will show that, with high probability, all these functions evaluate to 0 at the point r, implying $u(r) = 0$. We start by bounding $\mathbf{Pr}[(g\nu_L)(r) \neq 0]$. To this end, consider two cases. First, if $r \in P^2 \setminus S_p$ then, as $L \subseteq S_p$, we have that $\nu_L(r) = 1$ and so in such case

$$\mathbf{Pr}[(g\nu_L)(r) \neq 0] = \mathbf{Pr}[g(r) \neq 0] \leq \tau^2, \tag{4.7}$$

where the last inequality follows by Lemma 4.12 and since $\mathsf{dist}(r, p) = 2$ per our assumption $r \notin S_p$ and since $r \neq p$. If, on the other hand, $r \in S_p$ then, by the definition of L,

$$g(r) \neq 0 \implies r \in L \implies \nu_L(r) = 0,$$

and so in this case $(g\nu_L)(r) = 0$.

Let $L = \{r_1, \ldots, r_{|L|}\}$. Consider a fixed $i \in [|L|]$ and denote $(r_i) = \{w_{i,1}, \ldots, w_{i,b}\}$, where $b \leq c$. Fix $j \in [b]$. We turn to bound $\mathbf{Pr}\left[(h_{r_i,w_{i,j}}\nu_{w_{i,j}})(r) \neq 0\right]$. First note that

$$\mathbf{Pr}\left[(h_{r_i,w_{i,j}}\nu_{w_{i,j}})(r) \neq 0 \mid (r_i) \cap S_r = \emptyset\right] \leq \tau^2. \tag{4.8}$$

Indeed, conditioned on the event $(r_i) \cap S_r = \emptyset$, either $w_{i,j} = r$ or $\mathsf{dist}(w_{i,j}, r) = 2$. In the first case,

$$(h_{r_i,w_{i,j}} \nu_{w_{i,j}})(r) = h_{r_i,r}(r)\nu_r(r) = 0.$$

In the second case, the bound follows by Lemma 4.12. Second, note that

$$\mathbf{Pr}\left[(h_{r_i,w_{i,j}} \nu_{w_{i,j}})(r) \neq 0 \mid (r_i) \cap S_r \neq \emptyset\right] \leq \tau. \tag{4.9}$$

Indeed, as before, we may only consider the case $r \neq w_{i,j}$ and then observe that $\mathsf{dist}(r, w_{i,j}) = 1$ and invoke Lemma 4.12. Now, let $B = \{v \in S_p \mid (v) \cap S_r \neq \emptyset\}$. By Claim 4.16, and since R is (c, s)-axis evasive, $|B| \leq 4s$. Recall that $\mathbf{Pr}[r_i = v] \leq \tau$ for every fixed $v \in S_p$, and so

$$\mathbf{Pr}\left[(r_i) \cap S_r \neq \emptyset\right] = \mathbf{Pr}[r_i \in B] \leq 4s\tau. \tag{4.10}$$

By Equations (4.8), (4.9), (4.10) we get

$$\mathbf{Pr}\left[(h_{r_i,w_{i,j}} \nu_{w_{i,j}})(r) \neq 0\right] \leq \tau^2 + 4s\tau^2 \leq 5s\tau^2.$$

The proof then follows by the union bound over all $i \in [|L|]$ and $j \in [|(w_i)|]$. □

Definition 4.24. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. Let $\mathcal{L}^{P^2}$ be as in Definition 4.9. Let R be a (c, s)-axis evasive partition of P^2 and let $\mathcal{D}_R^{P^2}$ be as in Definition 4.19. We denote by $(\mathcal{D}^P, \mathcal{L}^P)_R^2$ the pair $(\mathcal{D}_R^{P^2}, \mathcal{L}^{P^2} + \mathcal{L}_R)$.*

Proposition 4.25. *Let P be a set, $\mathbb{F}$ a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q, \tau, \varepsilon, \rho)$-dual SLR. Let R be a (c, s)-axis evasive partition of P^2 that consists of t parts. Then, $(\mathcal{D}^P, \mathcal{L}^P)_R^2$ is a $(q_R, \tau_R, \varepsilon_R, \rho_R)$-dual SLR with*

$$\begin{aligned}
q_R &\leq 2cq^3 \\
\tau_R &\leq 10csq\tau^2 \\
\varepsilon_R &\leq 18csq\tau^2 + 2cq\varepsilon \\
\rho_R &\geq 1 - (1-\rho)^2 - \frac{t}{|P|^2}.
\end{aligned}$$

Proof. Claim 4.20 implies that for every $p \in P^2$, $\mathsf{supp}((D_R^{P^2})_p) \subseteq \mathcal{F}_p \cup \{\perp\}$. To bound q_R, note that by Equation (4.3),

$$|u| \leq |g\nu_L| + \sum_{r \in L} \sum_{w \in (r)} |h_{r,w}\nu_w|$$

Now, $|g\nu_L| \leq |g| \leq q^2$ and $|h_{r,w}\nu_w| \leq |h_{r,w}| \leq q^2$. Hence, $|u| \leq q^2 + |L|cq^2 \leq 2cq^3$. The stated bounds on τ_R and ε_R readily follows by Claim 4.23 and Claim 4.22, respectively.

As for the rate, we have that

$$\begin{aligned}\dim(\mathcal{L}^{P^2} + \mathcal{L}_R) &\le \dim(\mathcal{L}^{P^2}) + \dim(\mathcal{L}_R) \\ &\le (1-\rho)^2|P|^2 + t,\end{aligned}$$

where the second inequality follows by Proposition 4.14 and since R consists of t parts, implying $|\mathcal{L}_R| = t$. $\square$

4.4 Proofs of Theorem 1.3 and Corollary 1.4

With the machinery developed in the previous section, and using in a black-box manner, the construction of axis-evasive partitions we obtain in Section 5, we are finally ready to prove Theorem 1.3 and Corollary 1.4. We start by giving a more formal statement of Corollary 1.4.

Theorem 4.26. *There exist universal constants $m_0, c' \ge 1$ such that the following holds. Let P be a set of size $m \ge m_0$. Let $\mathbb{F}$ be a field, and let $(\mathcal{D}^P_{\text{in}}, \mathcal{L}^P_{\text{in}})$ be a $(q_{\text{in}}, \tau_{\text{in}}, \varepsilon_{\text{in}}, \rho_{\text{in}})$-dual SLR over $\mathbb{F}^P$. Let $0 < \alpha < 1$ be such that*

$$\rho_{\text{in}} \ge \frac{c'}{\sqrt{\alpha \cdot \log m}} \log\left(\frac{1}{\alpha}\right). \tag{4.11}$$

Then, there exists a $(q_{\text{out}}, \tau_{\text{out}}, \varepsilon_{\text{out}}, \rho_{\text{out}})$-dual SLR $(\mathcal{D}^P_{\text{out}}, \mathcal{L}^P_{\text{out}})$ over $\mathbb{F}^{P_{\text{out}}}$, with $m^\ell/2 \le |P_{\text{out}}| \le m^\ell$, where

$$\ell = \Theta\left(\frac{1}{\rho_{\text{in}}}\log\frac{1}{\alpha}\right), \tag{4.12}$$

having the following parameters:

$$\begin{aligned}q_{\text{out}} &\le q_{\text{in}}^{\text{poly}(\ell)}, \\ \tau_{\text{out}} &\le q_{\text{in}}^{\text{poly}(\ell)}\tau_{\text{in}}^{\ell}, \\ \varepsilon_{\text{out}} &\le q_{\text{in}}^{\text{poly}(\ell)}(\tau_{\text{in}} + \varepsilon_{\text{in}}), \\ \rho_{\text{out}} &\ge 1 - \alpha.\end{aligned}$$

A remark regarding the error. Note that there is another implicit constraint on ρ_{in} and α that originates from the error. Indeed, to make the result non-trivial, one must have $\varepsilon_{\text{out}} < 1$ which, in turn, implies some bound on ℓ and then, through Equation (4.12), a constraint on ρ_{in} and α. However, if that turns out to be a problem for the regime of parameters one is interested in, the probability to output $\perp$ can be reduced by repetition. Thus, by performing an alternating sequence of such error (or failure) reductions and rate amplifications, one can resolve this issue. Note that unlike for LDC, the error reduction has no cost in query complexity, and it certainly has no effect on the smoothness nor on the rate. It does, however, effects the running-time.

As mentioned above, our proof relies on an explicit axis-evasive partition that we construct in Section 5. Formally,

Theorem 4.27. *Let P be a set of size q, where q is an odd prime power. Let c be an even integer such that $c+1 \mid q+1$, and $c \leq \sqrt{q}/10$. Then, there exists a $(c, 4c^2)$-axis evasive partition of P^2 with at most $2q^2/(c+1)$ parts.*

Our proof of Theorem 4.26 is done by applying Proposition 4.25 several times, iteratively, where in each iteration we square the size of the set P obtained by the previous iterative step. Note, however, that Theorem 4.27 requires the set size $|P|$ to be an odd prime power q with the property that $c+1 \mid q+1$. It is best to choose c the same in all applications of Proposition 4.25. However, note that if we start an iteration with a set of size q and so end the iteration with a set of size q^2 then the condition will fail to hold at the beginning of the following iteration. Indeed if $c+1 \mid q+1$ then $q \equiv -1 \pmod{c+1}$ and so $q^2 \equiv 1 \pmod{c+1}$. To overcome this technicality, we do not work with the set obtained by the previous iteration as is. Instead, we find a prime–not much smaller than q^2–that has the desired residue -1 modulo $c+1$. To this end we rely on the Siegel–Walfisz Theorem [Sie35, Wal36] which refines Dirichlet's theorem on primes in arithmetic progressions. The state the Siegel–Walfisz Theorem we set some notation. For an integer $m \geq 1$, we denote Euler's totient function, that counts the positive integers up to m that are relatively prime to m, by $\phi(m)$. For integers n, m, r, we denote the number of (positive) primes less than or equal to n which are congruent to r modulo m by $\pi(n; m, r)$. The *Eulerian logarithmic integral* is given by

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\ln t}.$$

Theorem 4.28 ([Sie35, Wal36])**.** *For every constant $e \geq 1$ there exists a constant $c = c(e)$ such that the following holds. Let n, m, r be positive integers such that $m \leq (\log n)^e$, and m, r coprimes. Then,*

$$\left|\pi(n; m, r) - \frac{\mathrm{Li}(n)}{\phi(m)}\right| = O\left(n \cdot 2^{-c\sqrt{\log n}}\right).$$

We have the following straightforward corollary.

Corollary 4.29. *For every constant $e \geq 1$ there exist constants $c = c(e)$, $n_0 = n_0(e)$ such that the following holds. Let m, r be coprime integers, $m > 0$. Let $n \geq n_0$ be an integer such that $m \leq (\log n)^e$. Then, there exists a prime $p \in [n - \Delta, n]$, where $\Delta = cn/\log n$, such that $p \equiv r \pmod{m}$.*

Proof. To prove the corollary, it suffices to show that $\pi(n; m, r) > \pi(n - \Delta; m, r)$. By Theorem 4.28, there exist constants n_0, c' such that for every $n \geq n_0$,

$$\left|\pi(n; m, r) - \frac{\mathrm{Li}(n)}{\phi(m)}\right| \leq c'n \cdot 2^{-c\sqrt{\log n}}.$$

Thus, it suffices to show that

$$\frac{\mathrm{Li}(n)}{\phi(m)} - c'n \cdot 2^{-c\sqrt{\log n}} > \frac{\mathrm{Li}(n-\Delta)}{\phi(m)} + c'(n-\Delta) \cdot 2^{-c\sqrt{\log{(n-\Delta)}}}.$$

As we may assume that $\Delta \leq n/2$, it suffices to prove that

$$\mathrm{Li}(n) - \mathrm{Li}(n-\Delta) \geq 2c'\phi(m)n \cdot 2^{-c\sqrt{\log(n/2)}}. \tag{4.13}$$

It is well-known that

$$\mathrm{Li}(x) = c_1 + \frac{x}{\ln x} + O\left(\frac{x}{\ln^2 x}\right),$$

where $c_1 = \int_{t=0}^{2} \frac{dt}{\ln t}$ is some constant. Therefore,

$$\mathrm{Li}(n) - \mathrm{Li}(n-\Delta) \geq \frac{\Delta}{\ln(n/2)} - \frac{c''n}{\ln^2 n}.$$

for some constant c''. By our assumption on Δ we can choose the parameter c in the definition of Δ such that the right hand side is bounded below by $n/\ln^2 n$. The proof then follows by Equation (4.13) and noting that $\phi(m) \leq m \leq (\log n)^e = o(2^{-c\sqrt{\log{(n/2)}}})$. ◻

We turn to formally define and analyze the operation of projecting a dual SLR over $\mathbb{F}^P$ on a (large) subset of P.

Definition 4.30. *Let P a set and $P' \subseteq P$. Let $p' \in P'$ and D be a distribution with $\mathsf{supp}(D) \subseteq \mathcal{F}_{p'} \cup \{\perp\}$. We define the $D|_{P'}$ as follows: To sample from $D|_{P'}$, sample $f \sim D$. If $f = \perp$, output $\perp$; if $f \in \mathcal{F}_{p'}$, output $f|_{P'}$. We refer to $D|_{P'}$ as the distribution D projected to P'.*

Definition 4.31. *Let P be a set, $\mathbb{F}$ a field. Let $\mathcal{D} = \{D_p \mid p \in P\}$ be a collection of distributions, where for each $p \in P$, $\mathsf{supp}(D_p) \subseteq \mathcal{F}_p \cup \{\perp\}$. Let $P' \subseteq P$. We define $\mathcal{D}|_{P'}$ to be the collection $\mathcal{D}$ projected to P', that is, $\mathcal{D}|_{P'} = \{D_{p'}|_{P'} \mid p' \in P'\}$.*

Definition 4.32. *Let P be a set, $\mathbb{F}$ a field and let $\mathcal{L}$ be a linear subspace of $\mathbb{F}^P$. Let $P' \subseteq P$. We denote by $\mathcal{L}|_{P'}$ the linear subspace $\mathcal{L}$ projected to P', namely, $\mathcal{L}|_{P'} = \{f|_{P'} \mid f \in \mathcal{L}\}$.*

Claim 4.33. *Let P be a set, $\mathbb{F}$ a field, $(\mathcal{D}, \mathcal{L})$ a $(q, \tau, \varepsilon, \rho)$-dual SLR over $\mathbb{F}^P$, and let $P' \subseteq P$. Then, $(\mathcal{D}|_{P'}, \mathcal{L}_{P'})$ is a $(q, \tau, \varepsilon, \rho')$-dual SLR over $\mathbb{F}^{P'}$, where $\rho' = 1 - \frac{|P|}{|P'|}(1-\rho)$.*

Proof. That the smoothness τ, as well as q and ε, all stay the same after projecting the dual SLR to P', follows immediately from the definitions. The assertion regarding the rate of the induced SLR, ρ', readily follows as we have that

$$\dim(\mathcal{L}|_{P'}) \leq \dim(\mathcal{L}) \leq (1-\rho)|P| = \left(1 - (1 - \frac{|P|}{|P'|}(1-\rho))\right)|P'|.$$

□

Claim 4.34. *There exists a universal constant m_0 such that the following holds. Let P be a set of size $m \geq m_0$. Let $\mathbb{F}$ be a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q_{\text{in}}, \tau_{\text{in}}, \varepsilon_{\text{in}}, \rho_{\text{in}})$-dual SLR over $\mathbb{F}^P$. Let $c \leq \log m$ be an integer. Then, there exists a set P' of size*

$$|P'| \geq \left(1 - O\left(\frac{1}{\log m}\right)\right) m^2,$$

and a $(q_{\text{out}}, \tau_{\text{out}}, \varepsilon_{\text{out}}, \rho_{\text{out}})$-dual SLR $(\mathcal{D}^{P'}, \mathcal{L}^{P'})$ over $\mathbb{F}^{P'}$, where

$$\begin{aligned}
q_{\text{out}} &\leq 2cq_{\text{in}}^3, \\
\tau_{\text{out}} &\leq 40c^3 q_{\text{in}} \tau_{\text{in}}^2, \\
\varepsilon_{\text{out}} &\leq 80c^3 q_{\text{in}} (\tau_{\text{in}}^2 + \varepsilon_{\text{in}}), \\
\rho_{\text{out}} &\geq 1 - (1-\rho_{\text{in}})^2 - O(1/c).
\end{aligned}$$

Proof. By Corollary 4.29 applied with n, m, r in the notation of Corollary 4.29 set to $m, c+1, -1$ in the notation of this claim, respectively, there exists some prime $p \leq m$ such that $m - p = O(\frac{m}{\log m})$, and $c + 1 \mid p + 1$. Take P' to be an arbitrary subset of P of size p. By Claim 4.33, $(\mathcal{D}|_{P'}, \mathcal{L}|_{P'})$ is a $(q_{\text{in}}, \tau_{\text{in}}, \varepsilon_{\text{in}}, \rho')$-dual SLR on P', where

$$\rho' = 1 - \frac{m}{p}(1-\rho_{\text{in}}) \geq \rho_{\text{in}} - O\left(\frac{1}{\log m}\right).$$

By Theorem 4.27 applied to P', which observe is indeed applicable as $c+1 \mid p+1$, there exists an explicit $(c, 4c^2)$-axis evasive partition R of $(P')^2$ with at most $t = 2p^2/(c+1)$ parts. With that partition, we can now apply Proposition 4.25 to $(\mathcal{D}|_{P'}, \mathcal{L}|_{P'})$ and get that $(\mathcal{D}|_{P'}, \mathcal{L}|_{P'})^2_R$ is a $(q_{\text{out}}, \tau_{\text{out}}, \varepsilon_{\text{out}}, \rho_{\text{out}})$-dual SLR with the stated parameters. Note that the assertion regarding the rate follows as $c \leq \log m$, □

The following proposition is a more formal and accurate restatement of Theorem 1.3.

Proposition 4.35. *There exist universal constants $0 < c' < 1$ and $c'', m', \ell' \geq 1$ such that the following holds. Let P be a set of size $m \geq m'$. Let $\mathbb{F}$ be a field, and let $(\mathcal{D}^P, \mathcal{L}^P)$ be a $(q_{\text{in}}, \tau_{\text{in}}, \varepsilon_{\text{in}}, \rho_{\text{in}})$-dual SLR over $\mathbb{F}^P$. Let $\ell = 2^r$ for an integer $r \geq 1$, and assume that*

$\ell \geq \ell'$. Let c be an integer such that $c''\ell^2 \leq c \leq c' \log m$. Then, there exists a set P_ℓ of size $m^\ell/2 \leq |P_\ell| \leq m^\ell$, and a $(q_\ell, \tau_\ell, \varepsilon_\ell, \rho_\ell)$-dual SLR $(\mathcal{D}^{P_\ell}, \mathcal{L}^{P_\ell})$ over $\mathbb{F}^{P_\ell}$, where

$$\begin{aligned}
q_\ell &\leq (2cq_{\text{in}})^{\ell^{\log 3}}, \\
\tau_\ell &= O((c^3 q_{\text{in}})^{\ell^{\log 3}}) \cdot \tau_{\text{in}}^\ell, \\
\varepsilon_\ell &\leq O((c^4 q_{\text{in}})^{\ell^{\log 3}}) \cdot (\tau_{\text{in}} + \varepsilon_{\text{in}}), \\
\rho_\ell &\geq 1 - (1-\rho_{\text{in}})^\ell - O\left(\frac{\ell^2}{c}\right),
\end{aligned}$$

where, recall, the log *function is taken base* 2.

Proof. We construct a sequence of $(q_t, \tau_t, \varepsilon_t, \rho_t)$-dual SLR $(\mathcal{D}^{P_t}, \mathcal{L}^{P_t})$ for $t = 0, 1, \ldots, r = \log \ell$, and show that the last dual-SLR in the sequence has the stated parameters. The first dual-SLR, $(\mathcal{D}^{P_0}, \mathcal{L}^{P_0})$, is taken to be the $(q_{\text{in}}, \tau_{\text{in}}, \varepsilon_{\text{in}}, \rho_{\text{in}})$-dual SLR $(\mathcal{D}^P, \mathcal{L}^P)$ that is given by the hypothesis of the proposition. After constructing $(\mathcal{D}^{P_t}, \mathcal{L}^{P_t})$, we obtain $(\mathcal{D}^{P_{t+1}}, \mathcal{L}^{P_{t+1}})$ by applying Claim 4.34 to $(\mathcal{D}^{P_t}, \mathcal{L}^{P_t})$ with the parameter c taken to be c from the statement of this proposition. Note that, as required by the claim, $c \leq \log m$. Note that, by taking m' to be a large enough constant, all other dual SLR in the sequence will have $|P_t| \geq m$ as well, and so we can apply Claim 4.34 to them. Denote $m_t = |P_t|$. We begin by bounding m_t from below. Indeed, by Claim 4.34, and using that $1 - x \geq e^{-2x}$ for $x \leq 1/2$, we can pick the constant c'' such that

$$m_t \geq e^{-\frac{c''}{\log m_{t-1}}} m_{t-1}^2 \geq e^{-\frac{c''}{\log m_0}} m_{t-1}^2,$$

where the last inequality follows as, for a large enough constant m', the sequence $(m_t)_t$ is monotone increasing. We invoke Claim 6.19 with $a = e^{\frac{c''}{2 \log m_0}}$ and $b = 2$ to conclude that

$$m_t \geq m_0^{2^t} e^{-\frac{c'' 2^t}{\log m_0}} \geq \frac{1}{2} m_0^{2^t},$$

where the last inequality follows as $t \leq r = \log \ell$ and, recall, we take $\ell \leq c' \log m$ for a sufficiently small constant $c' > 0$. In particular, $m_r \geq m^\ell/2$ as stated.

By Claim 4.34, for every $t \geq 1$ we have $q_t \leq 2cq_{t-1}^3$. It is straightforward to prove by that

$$q_t \leq (2cq_{\text{in}})^{3^t}, \tag{4.14}$$

which readily implies the assertion regarding the query complexity. We turn to analyze the rate. Denote $\beta_t = 1 - \rho_t$. Claim 4.34 implies that $\beta_t \leq \beta_{t-1}^2 + c'''/c$, for some constant $c''' > 0$. By induction on t, we get that $\beta_t \leq \beta_0^{2^t} + c''' 4^t/c$. Indeed, the base case $t = 0$ is obvious. Now, by the induction hypothesis,

$$\beta_t \leq \beta_{t-1}^2 + \frac{c'''}{c} \leq \left(\beta_0^{2^{t-1}} + 4^{t-1}\frac{c'''}{c}\right)^2 + \frac{c'''}{c}.$$

One can easily verify that the right hand side is bounded above by the desired bound $\beta_0^{2^t} + c'''4^t/c$ provided that $2^t c'''/c \leq 1$. As $t \leq r$ and $2^r = \ell$, the latter inequality follows assuming $c'''\ell \leq c$. As we assume $c \geq c''\ell^2$, it suffices to choose ℓ' from the statement of the proposition to be a constant larger than the constant c'''/c''. We conclude that,

$$\beta_r \leq \beta_0^\ell + O\left(\frac{4^r}{c}\right) = \beta_0^\ell + O\left(\frac{\ell^2}{c}\right),$$

which implies the assertion regarding the rate.

As for the smoothness, by Claim 4.34, and using Equation (4.14), we have that

$$\tau_t \leq 40c^3 q_{t-1} \tau_{t-1}^2 \leq 40c^3 \left(2cq_{\text{in}}\right)^{3^{t-1}} \tau_{t-1}^2,$$

from which it is easy to verify that

$$\tau_t \leq (40c^3)^{2^t} (2cq_{\text{in}})^{3^t} \tau_0^{2^t},$$

and the assertion regarding the smoothness readily follows. Last is the error which we leave to the reader to verify. □

We can now easily deduce Theorem 4.26

Proof of Theorem 4.26. The proof readily follows from Proposition 4.35 by taking ℓ as defined in Equation (4.12), and with c in the notation of Proposition 4.35 taken to be $c = \Theta(\ell^2/\alpha)$. Note that this choice of parameters satisfies the hypothesis of Proposition 4.35 as indeed implied by Equation (4.11) and by taking c' to be a sufficiently large constant. It is easy to verify that the rate is $1 - \alpha$ with our choice of c, ℓ, and the remaining assertions readily follow by Proposition 4.35. □

5 Axis-evasive partitions

The distance-efficient rate amplification procedure that was developed in the previous section is built on axis-evasive partitions. Note that, by Proposition 4.25, the number of parts t effects the rate, c effects the query complexity and both c, s the deterioration of the distance and error. It is perhaps best to consider the following goal: for a given c we wish to obtain a (c, s)-axis evasive partition with both s, t as small as possible.

We start this section by proving the existence of axis-evasive partitions with great parameters. However, our probabilistic proof does not work for every c but rather, it requires $c = \Omega(\log m)$, where $m = |P|$. Unfortunately, for our distance-efficient rate amplification procedure, we are interested in $c < \log m$ (see Proposition 4.35). Luckily, and perhaps somewhat surprisingly, our explicit construction, described in Section 5.2, does work for every c albeit it requires $c + 1 \mid m + 1$ to hold.

5.1 Existential proof

As mentioned above, while we do not use the following non-constructive proof for the existence of axis-evasive sets, as given by the following lemma, we believe the reader might benefit from reading it still, as it gives an intuition on what is it about axis-evasive partitions which is random and what requires structure.

Lemma 5.1. *Let P be a set of size m, and let c be an integer such that $50 \log m \le c \le \sqrt{m}$. Then, there exists a $(c, s = c)$-axis evasive partition of P^2 with $t \le 5m^2/c$ parts.*

Proof. Let $k = 2m^2/c$. The proof is by a probabilistic argument. We form a partition by assigning to each point $p \in P^2$ a "color" or, more formally, a number in $[k]$. The k parts are then formed by grouping together points with the same color. To this end, for every $p \in P^2$ define a random variable C_p that is uniformly distributed over $[k]$, where $\{C_p \mid p \in P^2\}$ are independent. For $i \in [k]$ let R_i be the number of random variables C_p for which $C_p = i$. Note that R_i is the size of part i, and that $\mathbf{E}[R_i] = c/2$. For every fixed $i \in [k]$, by the Chernoff bound,

$$\mathbf{Pr}\left[R_i \notin [c/4, c]\right] \le 2e^{-c/16}.$$

Thus, by the union bound over $i \in [k]$ and per our assumption $c \ge 50 \log m$, we have that except for probability $1/4$, for every $i \in [k]$, $R_i \in [c/4, c]$.

Now, we would want to claim that this partition satisfies the third condition, meaning that for every $p \in P^2$ and $\ell \in \mathcal{X}$, $|[p] \cap \ell| \le 1$. However, with high probability, this property in fact does not hold. To fix this, we make a slight modification to the random partition above so that it does satisfy the requirement. The change, is simply, given a partition - whenever there is a "collision" on a line $\ell \in \mathcal{X}$, meaning that for some distinct $p, r \in \ell$, $C_p = C_r$, assign new and distinct parts to both p and r. To analyze the number of additional parts we need, we introduce the following notation. For $\ell \in \mathcal{X}$ let

$$\nu(\ell) = \{\{p, r\} \mid p, r \in \ell \text{ and } p \neq r\}.$$

For $v = \{p, r\} \in \nu(\ell)$ define $\mathbb{I}_v^\ell$ to be an indicator for the event that $C_p = C_r$. With this notation, the number of collisions is bounded by $\sum_{\ell \in \mathcal{X}} \sum_{v \in \nu(\ell)} \mathbb{I}_v^\ell$. It holds that

$$\mathbf{E}\left[\sum_{\ell \in \mathcal{X}} \sum_{v \in \nu(\ell)} \mathbb{I}_v^\ell\right] = 2m \binom{m}{2} \frac{1}{k} < \frac{mc}{2}.$$

Therefore, by Markov's inequality, with probability at least $1/2$, the number of collisions is less than mc. In such case, we can add at most mc parts to the partition and be guaranteed that for every $p \in P^2$ and $\ell \in \mathcal{X}$, $|[p] \cap \ell| \le 1$. Recall that since, prior to the procedure above, every part has size at least $c/4$ the total number of parts is now bounded by

$$t \le mc + \frac{m^2}{c/4} \le \frac{5m^2}{c},$$

where the last inequality follows as we assume $c \le \sqrt{m}$.

To conclude the proof, it suffices to show that, with probability larger than $7/8$, it holds that for every $\ell, \ell' \in \mathcal{X}$, $|\ell' \cap (\ell)| \le c$. Note that it suffices to prove this with respect to the partition obtained prior to the procedure above since, by introducing new parts of size one each, one only decrease the intersection size we aim to bound from above. Denote by $C_\ell = \{C_p \mid p \in \ell\} \subseteq [k]$. We have that

$$|\ell \cap (\ell')| = \left|\ell \cap \bigcup_{p \in \ell'} (p)\right| \le 1 + \left|\{p \in \ell' \setminus \ell \mid C_p \in C_\ell\}\right|.$$

Now, by the union bound,

$$\mathbf{Pr}\left[C_p \in C_\ell\right] \le \frac{m}{k} = \frac{c}{2m}.$$

As $\{C_p \mid p \in \ell'\}$ are chosen independently, by the Chernoff bound,

$$\mathbf{Pr}\left[|\{p \in \ell' \setminus \ell \mid C_p \in C_\ell\}| \ge c\right] \le e^{-c/6} \le \frac{1}{m^3},$$

where for the last inequality was used our assumption $c \ge 50 \log m$. The proof then follows by taking the union bound over all $\ell, \ell' \in \mathcal{X}$.

□

5.2 Explicit constructions

In this section we give explicit constructions of axis-evasive partitions (see Definition 4.15). Our constructions are based on quadratic field extensions. We identify a set P of size q–a prime power–with the finite field $\mathbb{F}_q$ in an arbitrary manner, namely, by using an arbitrary bijection which, for ease of readability, we do not make explicit in the notation. We start by giving some basic background on finite fields.

Let $h(x) \in \mathbb{F}_q[x]$ be a degree 2 irreducible monic polynomial. It is a well-known fact that $\mathbb{F}_q[x]/\langle h(x) \rangle$ is a field of size q^2 which we denote, somewhat less informatively, by $\mathbb{F}_{q^2}$. Note that there exists $\alpha \in \mathbb{F}_{q^2}$ such that $h(\alpha) = 0$ (indeed, take $\alpha = x + \langle h(x) \rangle$). Since h is irreducible over $\mathbb{F}_q$ and has degree 2, we can write every element of $\mathbb{F}_{q^2}$ in the form $a + \alpha b$, where $a, b \in \mathbb{F}_q$, in a unique manner. That is, we can identify in the set-theoretic level, $\mathbb{F}_{q^2}$ with $\mathbb{F}_q + \alpha \mathbb{F}_q$. Using this identification, we identify P^2 with $\mathbb{F}_{q^2}$ in the natural way, namely, a point $(a, b) \in P^2$ is identified with $a + \alpha b$ in $\mathbb{F}_{q^2}$. Note that, with this identification, the horizontal lines in P^2 are of the form $b\alpha + \mathbb{F}_q$ where $b \in \mathbb{F}_q$ can be thought of as the fixed height of the line. Similarly, the vertical lines are given by $b + \alpha\mathbb{F}_q$. Given $\delta \in \mathbb{F}_{q^2} \setminus \{0\}$, we say that $\ell_\delta = \delta\mathbb{F}_q \subseteq \mathbb{F}_{q^2}$ is the line through the origin with slope δ.

Our construction of exis-evasive partitions is based on an equivalence relation that we are about to define. The partition is then obtained by considering the respective equivalence classes. We begin the construction by ignoring the "origin" $0 \in \mathbb{F}_{q^2}$ and work only with $\mathbb{F}_{q^2} \setminus \{0\}$. Note that this is the set of invertible elements of $\mathbb{F}_{q^2}$ which has a group structure under the field multiplication. When referring to this multiplicative group we write $(\mathbb{F}_{q^2})^\times$.

Let $\beta \in (\mathbb{F}_{q^2})^\times$. Denote by $o(\beta)$ the order of β in the multiplicative group $(\mathbb{F}_{q^2})^\times$. It will be convenient to denote $c = o(\beta) - 1$. We define an equivalence relation on $(\mathbb{F}_{q^2})^\times$, parameterized by β, as follows: For $\gamma, \delta \in (\mathbb{F}_{q^2})^\times$

$$\gamma \sim \delta \quad \iff \quad \gamma\delta^{-1} \in \langle\beta\rangle, \tag{5.1}$$

where $\langle\beta\rangle$ is the subgroup of $(\mathbb{F}_{q^2})^\times$ that is generated by β. Observe that this is an equivalence relation. Indeed, the classes are the different cosets, that is, the elements of the quotient group $(\mathbb{F}_{q^2})^\times/\langle\beta\rangle$. For completeness, we quickly prove that this is an equivalence relation: as $1 \in \langle\beta\rangle$, we have that $\gamma \sim \gamma$. Secondly, if $\gamma\delta^{-1} \in \langle\beta\rangle$ then $\delta\gamma^{-1} \in \langle\beta^{-1}\rangle = \langle\beta\rangle$ which establishes symmetry. As for transitivity, if $\gamma \sim \delta$ and $\delta \sim \varepsilon$ then

$$\gamma\varepsilon^{-1} = \gamma(\delta^{-1}\delta)\varepsilon^{-1} = (\gamma\delta^{-1})(\delta\varepsilon^{-1}) \in \langle\beta\rangle.$$

One can easily see that the equivalence class of an element $\gamma \in (\mathbb{F}_{q^2})^\times$ is $[\gamma] = \gamma\langle\beta\rangle = \{\gamma, \beta\gamma, \ldots, \beta^c\gamma\}$. Note further that $|[\gamma]| = c + 1$. Indeed, if there are $0 \le j < i \le c$ such that $\beta^i\gamma = \beta^j\gamma$ then $0 = (\beta^i - \beta^j)\gamma = (\beta^{i-j} - 1)\beta^j\gamma$, which is a contradiction as none of the factors in the product is zero.

In the following claim we show that, under some conditions on α, β, the second property of axis-evasiveness is met by the construction above. We mention already here that the third condition in Definition 4.15 is not met by the construction as is (regardless of the choice of α, β), and we will alter it afterwards to meet that property as well.

Claim 5.2. *Assume that $\langle\beta\rangle \cap \ell_\alpha = \langle\beta\rangle \cap \ell_{\alpha^{-1}} = \emptyset$ and that $\langle\beta\rangle \cap \mathbb{F}_q = \{1\}$. Then, for every $\ell, \ell' \in \mathcal{X}$ (not necessarily distinct) it holds that $\left|\ell' \cap (\ell)\right| \le c$.*

Proof. Recall that $(\gamma) = \{\beta\gamma, \ldots, \beta^c\gamma\}$. Thus,

$$\bigcup_{\gamma\in\ell} (\gamma) = \bigcup_{\gamma\in\ell}\bigcup_{i=1}^{c} \{\beta^i\gamma\} = \bigcup_{i=1}^{c} \beta^i\ell.$$

Therefore,

$$\ell' \cap (\ell) = \ell' \cap \bigcup_{\gamma \in \ell} (\gamma) = \bigcup_{i=1}^{c} (\ell' \cap \beta^i \ell). \tag{5.2}$$

Fix $i \in [c]$ and consider two cases. First, if ℓ is vertical, namely, $\ell = b + \alpha \mathbb{F}_q$ for some $b \in \mathbb{F}_q$, then $\beta^i \ell = \beta^i b + \alpha \beta^i \mathbb{F}_q$. Since, by assumption, $\langle \beta \rangle \cap \mathbb{F}_q = \{1\}$ we have that $\alpha \beta^i \mathbb{F}_q \neq \alpha \mathbb{F}_q$ and so the line $\beta^i \ell$ is not vertical. As, by assumption, $\langle \beta \rangle \cap \ell_{\alpha^{-1}} = \emptyset$, we have that $\alpha \beta^i \notin \mathbb{F}_q$ and so the line $\beta^i \ell$ is not horizontal either.

Second, consider the case that ℓ is horizontal $\ell = b\alpha + \mathbb{F}_q$ for some $b \in \mathbb{F}_q$. Then, $\beta^i \ell = b\alpha\beta^i + \beta^i \mathbb{F}_q$. Per our assumption that $\langle \beta \rangle \cap \ell_\alpha = \emptyset$, we have that $\beta^i \mathbb{F}_q \neq \alpha \mathbb{F}_q$ and so the line $\beta^i \ell$ is not vertical. As we assume $\langle \beta \rangle \cap \mathbb{F}_q = \{1\}$, we have that $\beta^i \mathbb{F}_q \neq \mathbb{F}_q$, and so the line $\beta^i \ell$ cannot be horizontal either. To summarize, we have that $\beta^i \ell \notin \mathcal{X}$. However, $\ell' \in \mathcal{X}$ and so $\beta^i \ell$ and ℓ' are two distinct lines. As such, the two lines intersect in at most one point. Equation (5.2) then yield $|\ell' \cap (\ell)| \leq c$.

□

Informal discussion regarding the third property. As mentioned above, the partition of $(\mathbb{F}_{q^2})^\times$ as defined above does not have the third property required for axis-evasiveness. Namely, there are $\gamma \in (\mathbb{F}_{q^2})^\times$ such that $[\gamma]$ intersects some axis-parallel line at more than one point. To get some idea on which equivalence classes $[\gamma]$ are problematic, let us first ask when do $\gamma, \beta\gamma$ are on some common axis-parallel line. We first observe that two points $\delta, \varepsilon \in (\mathbb{F}_{q^2})^\times$ are on a common axis-parallel line if and only if $\delta - \varepsilon \in \{1, \alpha\}\mathbb{F}_q$. Thus, γ and $\beta\gamma$ are on the same axis-parallel line if and only if $\gamma - \beta\gamma = (1-\beta)\gamma \in \{1, \alpha\}\mathbb{F}_q$. This is equivalent to saying that γ is on one of the two lines through the origin with slopes $\frac{1}{1-\beta}, \frac{\alpha}{1-\beta}$.

More generally, $[\gamma]$ intersects with some axis-parallel line in more than one point if and only if $\beta^i\gamma - \beta^j\gamma \in \{1, \alpha\}\mathbb{F}_q$ for some $0 \leq j < i \leq c$. Equivalently, γ is on a line ℓ_δ with

$$\delta \in \left\{ \frac{1}{\beta^i - \beta^j}, \frac{\alpha}{\beta^i - \beta^j} \;\middle|\; 0 \leq j < i \leq c \right\}. \tag{5.3}$$

The key observation is that although there are a fair amount of "bad" points γ, they are all contained in a small number of lines. By "small" here we mean that the number is polynomial in c and is independent of q. Thus, the hope is that by redefining the partition on these few problematic lines we will not harm the previous analysis by much. Indeed, no matter how we alter the partition restricted to these lines, if we make sure none of them is axis-parallel (by requiring more properties from α, β) then each of these lines intersect an axis-parallel line at one point. As a result, the bound obtained in Claim 5.2 will deteriorate proportionally to the number of lines above.

The only small technical issue is that even if $\gamma \in \ell_\delta$ for some slope δ as above, it is not the case that $[\gamma] \subseteq \cup_\varepsilon \ell_\varepsilon$ where ε is taken from the set of slopes given by Equation (5.3). As we wish to alter the partition defined above, it would be cleaner to have all of the points in $[\gamma]$ of a problematic point γ contained in the set of points on which we redefine the partition. Thus, we "close" the set of slopes given by Equation (5.3) to multiplication by β.

Ending the informal discussion and returning to the formal analysis, we consider the set of slopes.

$$\Delta = \left\{ \frac{\beta^k}{\beta^i - \beta^j}, \frac{\alpha\beta^k}{\beta^i - \beta^j} \;\middle|\; 0 \leq j < i \leq c \text{ and } 0 \leq k \leq c \right\} \tag{5.4}$$

Further define the set of all points in $(\mathbb{F}_{q^2})^\times$ covered by the lines with slopes from Δ by

$$U = \bigcup_{\delta \in \Delta} \ell_\delta.$$

This definition of Δ indeed fixes the technical caveat discussed above, as the following claim states.

Claim 5.3. *For every $\gamma \in (\mathbb{F}_{q^2})^\times$ either $[\gamma] \subseteq U$ or $[\gamma] \cap U = \emptyset$.*

Proof. If an element $\varepsilon \in U$ then $\varepsilon \in \ell_\delta$ for some $\delta \in \Delta$. Note that $\beta\varepsilon \in \ell_{\beta\delta}$ and that $\beta\delta \in \Delta$. Hence, $\beta\varepsilon \in U$. Therefore, $\varepsilon \in U \implies \varepsilon\langle\beta\rangle \subseteq U$. Assume now that $[\gamma] \cap U \neq \emptyset$, and take $\gamma\beta^i \in U$. By the above, $\gamma\beta^i\langle\beta\rangle \subseteq U$. The proof then follows as $\gamma\beta^i\langle\beta\rangle = \gamma\langle\beta\rangle = [\gamma]$. □

Define a new partition of $\mathbb{F}_{q^2}$ (including 0) which agrees with the one that is given by Equation (5.1) on $\mathbb{F}_{q^2}^\times \setminus U$. By Claim 5.3, this is well-defined. The new partition, restricted to U, is done as follows. Let $\delta_0 \in \Delta$ be an arbitrary element. Note that

$$U = \ell_{\delta_0} \cup \bigcup_{\delta \in \Delta \setminus \{\delta_0\}} (\ell_\delta \setminus \{0\})$$

is a disjoint union. To partition U, we partition ℓ_{δ_0} as well as each of $\ell_\delta \setminus \{0\}$ where $\delta \in \Delta \setminus \{\delta_0\}$ in an arbitrary way provided it has the least number of parts under the conditions that each part has size at most $c+1$. For ease of readability, we denote by $[\gamma]$ the class with respect to the new partition.

Claim 5.4. *Assume, on top of the assumptions of Claim 5.2 that for every $\delta \in \Delta$, $\ell_\delta \notin \mathcal{X}$. Then, the new partition defined above is $(c, 4c^2)$-axis evasive.*

Proof. First, observe that by construction, every class intersects any axis-parallel line in at most one point. Indeed, classes that are outside of U have this property by the definition of U as can be easily verified (and discussed above). Moreover, by the way we redefined the partition restricted to U, every class that is a subset of U is also a subset of a line ℓ_δ for some $\delta \in \Delta$. As $\ell_\delta \notin \mathcal{X}$ by hypothesis, we have that the line and, as a result, the class it contains, intersects any axis-parallel line in at most one point. This establishes the third property of axis-evasiveness. The second property follows as, by construction, every part has size at most $c+1$.

Moving on to the second property, consider $\ell, \ell' \in \mathcal{X}$, not necessarily distinct. As outside of U the partition is defined as before, Claim 5.3 yields

$$\left| \ell' \cap \bigcup_{\gamma \in \ell \setminus U} (\gamma) \right| \leq c. \tag{5.5}$$

Take $\gamma \in U \cap \ell$. Since, by construction $(\gamma) \subseteq \ell_\delta$ for some $\delta \in \Delta$, and since by hypothesis $\ell_\delta \notin \mathcal{X}$ we have that $|\ell' \cap \ell_\delta| = 1$ and $(\gamma) \cap \ell' \subseteq \ell_\delta \cap \ell'$.Therefore, $|(\gamma) \cap \ell'| \leq 1$. Together with Equation (5.5) we get that $|\ell' \cap (\ell)| \leq c + |U \cap \ell|$. Now, since $\ell \in \mathcal{X}$ and every line ℓ_δ with slope $\delta \in \Delta$ is not in $\mathcal{X}$ we have that $|\ell \cap \ell_\delta| = 1$. Thus, $|U \cap \ell| \leq |\Delta|$ which implies $|\ell' \cap (\ell)| \leq c + |\Delta|$.

To conclude the proof, we turn to bound $|\Delta|$. It is straightforward to give a bound of $O(c^3)$ though one can optimize the bound a bit. Indeed, with the notation of Equation (5.4), by multiplying by $\beta^{-\min(j,k)}$, one can rewrite

$$\Delta = \left\{ \frac{1}{\beta^i - \beta^j}, \frac{\alpha}{\beta^i - \beta^j} \;\middle|\; 0 < j < i \leq c \right\} \bigcup \left\{ \frac{\beta^j}{\beta^i - 1}, \frac{\alpha \beta^j}{\beta^i - 1} \;\middle|\; 0 < i \leq c,\, 0 \leq j \leq c \right\}. \tag{5.6}$$

Thus, $|\Delta| \leq 3c^2$, and the proof follows. $\square$

We summarize the discussion so far.

Proposition 5.5. *Let $\mathbb{F}_q$ be finite field. Let $h(x) \in \mathbb{F}_q[x]$ be a degree 2 irreducible monic polynomial, and consider the field $\mathbb{F}_q[x]/\langle h(x) \rangle$ which we denote by $\mathbb{F}_{q^2}$. Let $\alpha, \beta \in \mathbb{F}_{q^2}$ be two elements satisfying:*

1. $h(\alpha) = 0$,
2. $\langle \beta \rangle \cap \mathbb{F}_q = \{1\}$,
3. $c + 1 = o(\beta) \leq \sqrt{q}/10$,
4. $\langle \beta \rangle \cap \ell_\alpha = \langle \beta \rangle \cap \ell_{\alpha^{-1}} = \emptyset$,

5. $(\langle\beta\rangle - \langle\beta\rangle) \cap \mathbb{F}_q = \{0\}$,

6. $(\langle\beta\rangle - \langle\beta\rangle) \cap \ell_\alpha = (\langle\beta\rangle - \langle\beta\rangle) \cap \ell_{\alpha^{-1}} = \{0\}$.

Then, there exists a partition of $(\mathbb{F}_q)^2$ *that is* $(c, 4c^2)$*-axis-evasive, where* $c = o(\beta) - 1$. *The number of parts in the partition is bounded above by* $2q^2/(c+1)$.

To prove Proposition 5.5 we need the following easy claim.

Claim 5.6. *Let* $\delta \in (\mathbb{F}_{q^2})^\times$ *be such that* $\delta \notin \mathbb{F}_q \cup \ell_\alpha$ *then,* $\ell_\delta \notin \mathcal{X}$.

Proof. Write $\delta = a + \alpha b$ with $a, b \in \mathbb{F}_q$. Then, $\ell_\delta = (a + \alpha b)\mathbb{F}_q$. Observe that if ℓ_δ is vertical then $a = 0$ and so $\delta \in \ell_\alpha$. Similarly, if ℓ_δ is horizontal then $b = 0$ implying $\delta \in \mathbb{F}_q$. □

Proof of Proposition 5.5. To bound the number of parts, recall that in the original partition, each part has size $c+1$. Moreover, in the altered partition we partition each line ℓ_δ with slope $\delta \in \Delta$ (excluding the origin from all but for one of the lines ℓ_{δ_0}) to parts of size $c+1$ each, except for possibly one part. As $|\Delta| \le 3c^2$, the number of parts it bounded by

$$\frac{q^2-1}{c+1} + |\Delta|\left(1 + \frac{q}{c+1}\right) \le \frac{q^2-1}{c+1} + 6cq \le \frac{2q^2}{c},$$

where the last inequality follows by our assumption that $o(\beta) \le \sqrt{q}/10$.

To conclude the proof of the proposition, it suffices to show that for every $\delta \in \Delta$ it holds that $\ell_\delta \notin \mathcal{X}$. By Claim 5.6, it suffices to prove that $\delta \notin \mathbb{F}_q \cup \ell_\alpha = \{1, \alpha\}\mathbb{F}_q$. There are two types of slopes $\delta \in \Delta$, according to whether they appear in the first or second set in Equation (5.6). The first kind is of the form

$$\delta = \frac{\alpha^k}{\beta^i - \beta^j},$$

with $0 < j < i \le c$ and $k \in \{0, 1\}$. If $\delta \in \{1, \alpha\}\mathbb{F}_q$ then $\delta^{-1} \in \{1, \alpha^{-1}\}\mathbb{F}_q$ and so $\beta^i - \beta^j \in \{\alpha^k, \alpha^{k-1}\}\mathbb{F}_q$ in contradiction to our hypothesis. Consider now the other kind of slope

$$\delta = \frac{\alpha^k \beta^j}{\beta^i - 1}$$

where $0 < i \le c$, $0 \le j \le c$ and $k \in \{0, 1\}$. If $\delta \in \{1, \alpha\}\mathbb{F}_q$ then $\delta^{-1} \in \{1, \alpha^{-1}\}\mathbb{F}_q$ and so $(\beta^i - 1)\beta^{-j} \in \{\alpha^k, \alpha^{k-1}\}\mathbb{F}_q$. Note that $(\beta^i - 1)\beta^{-j} = \beta^{i-j} - \beta^{-j} \in \langle\beta\rangle - \langle\beta\rangle$ and so we again get a contradiction.

□

We are now ready to prove Theorem 4.27. For the sake of readability, we repeat its statement here.

Theorem 5.7. *Let P be a set of size q, where q is an odd prime power. Let c be an even integer such that $c+1 \mid q+1$, and $c \leq \sqrt{q}/10$. Then, there exists a $(c, 4c^2)$-axis evasive partition of P^2 with at most $2q^2/(c+1)$ parts.*

Proof. As above, we identify P^2 with $\mathbb{F}_{q^2}$. It is a well-known fact that the multiplicative group $(\mathbb{F}_{q^2})^\times$ is cyclic. A basic result in group theory states that a cyclic group has a (unique) subgroup of every given size which divides the group size. Now, $|(\mathbb{F}_{q^2})^\times| = q^2 - 1 = (q-1)(q+1)$. Thus, as $c+1 \mid q+1$, there exists a subgroup H of $(\mathbb{F}_{q^2})^\times$ of size $c+1$. The subgroup H is cyclic, being a subgroup of a cyclic group. Let β be a generator for H. We first prove that β satisfies those hypothesis of Proposition 5.5 that do not involve α, namely, conditions (2) and (5).

Claim 5.8. $(\langle\beta\rangle - \langle\beta\rangle) \cap \mathbb{F}_q = \{0\}$ *and* $\langle\beta\rangle \cap \mathbb{F}_q = \{1\}$.

Proof. Assume towards a contradiction that $\beta^i - \beta^j \in \mathbb{F}_q$ for some $0 \leq j < i \leq c$. Since $x^q = x$ for every $x \in \mathbb{F}_q$, we get

$$\beta^i - \beta^j = \left(\beta^i - \beta^j\right)^q = \beta^{iq} - \beta^{jq},$$

where the last equality follows since q is divisible by the characteristic of the field. Recall that $o(\beta) = c+1 \mid q+1$ and so $\beta^{i(q+1)} = 1$, implying $\beta^{iq} = \beta^{-i}$. Thus,

$$\beta^i - \beta^j = \frac{1}{\beta^i} - \frac{1}{\beta^j} = \frac{\beta^j - \beta^i}{\beta^{i+j}}.$$

As $\beta^i \neq \beta^j$ the above equation implies $\beta^{i+j} = -1$, and so $-1 \in H$. Since q is odd, the characteristic of the field $\mathbb{F}_{q^2}$ is odd and so $o(-1) = 2$. Lagrange's Theorem then implies that $2 \mid |H| = c+1$, which stands in contradiction to c being even.

To prove that $\langle\beta\rangle \cap \mathbb{F}_q = \{1\}$, take β^i with $0 < i \leq c$. If $\beta^i \in \mathbb{F}_q$ then $\beta^{iq} = \beta^i$. On the other hand, we proved above that $\beta^{iq} = \beta^{-i}$, and so $\beta^i = \beta^{-i}$ implying $\beta^{2i} = 1$. Therefore, $o(\beta) = c+1 \mid 2i$, but this is impossible as $0 < i \leq c$ and, recall, c is even. ☐

We proceed with the proof of Theorem 4.27 by finding $\alpha \in \mathbb{F}_{q^2}$ that, together with the already chosen β, satisfies the remaining conditions in the hypothesis of Proposition 5.5. Since $\mathbb{F}_{q^2}$ is a quadratic field extension of $\mathbb{F}_q$, every element $\gamma \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$ has degree 2. That is, the minimal polynomial h_γ of every such γ over $\mathbb{F}_q$ is of degree 2 (and can be made monic by dividing by the leading coefficient, if necessary). Indeed, $\deg(h_\gamma)$ cannot equal 1 as this would imply $\gamma \in \mathbb{F}_q$. On the other hand,

$$2 = [\mathbb{F}_{q^2} : \mathbb{F}_q] = [\mathbb{F}_{q^2} : \mathbb{F}_q(\gamma)][\mathbb{F}_q(\gamma) : \mathbb{F}_q] = [\mathbb{F}_{q^2} : \mathbb{F}_q(\gamma)]\deg(h_\gamma),$$

which shows that if $\deg(h_\gamma) \neq 1$ then $\deg(h_\gamma) = 2$.

Thus, condition (1) in the hypothesis of Proposition 5.5 holds for every element in $\mathbb{F}_{q^2} \setminus \mathbb{F}_q$. Hence, to prove that all the remaining conditions in the hypothesis of Proposition 5.5 hold, it suffices to prove that there exists $\alpha \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$ which satisfies conditions (4) and (6). To this end, pick a set of "slopes" $\Delta' = \{\delta_1, \ldots, \delta_{q+1}\} \subseteq (\mathbb{F}_{q^2})^\times$ such that $(\mathbb{F}_{q^2})^\times$ is the disjoint union

$$(\mathbb{F}_{q^2})^\times = \bigcup_{\delta \in \Delta'} (\ell_\delta \setminus \{0\}).$$

For example, $\Delta' = \{a + \alpha \mid a \in \mathbb{F}_q\} \cup \{1\}$ will do. For $\delta \in (\mathbb{F}_{q^2})^\times$ let

$$I_\delta = |\langle \beta \rangle \cap \ell_\delta| + |\left(\langle \beta \rangle - \langle \beta \rangle\right) \cap (\ell_\delta \setminus \{0\})|.$$

Since the $\ell_\delta \setminus \{0\}$ with $\delta \in \Delta'$ are disjoint, $0 \notin \langle \beta \rangle$, and since $|\langle \beta \rangle| = c+1$ and $|\langle \beta \rangle - \langle \beta \rangle| \leq (c+1)^2$, we have that

$$\mathop{\mathbf{E}}_{\delta}[I_\delta] \leq \frac{(c+1)^2 + (c+1)}{q+1} \leq \frac{2(c+1)^2}{q+1},$$

where δ is sampled uniformly from Δ'. By Markov's inequality, for at least $3/4$ of the elements $\delta \in \Delta'$ it holds that

$$|I_\delta| \leq \frac{8(c+1)^2}{q+1}.$$

Note that $(\mathbb{F}_{q^2})^\times$ is also a disjoint union of $\{\ell_{\delta^{-1}} \setminus \{0\} \mid \delta \in \Delta\}$. Thus, using the same argument as above, we get that for at least $1/2$ the elements $\delta \in \Delta'$, both $|I_\delta|$ and $|I_{\delta^{-1}}|$ are bounded by $8(c+1)^2/(q+1)$. But, as $c \leq \sqrt{q}/10$, this bound is strictly smaller than 1, implying that $I_i = I_{q+1-i} = 0$. That is, at least half the elements $\delta \in \Delta'$ satisfy conditions (4) and (6). Take α to be any of these elements. To conclude, we found α and β for which all the conditions in the hypothesis of Proposition 5.5 are met, and the proof follows. $\square$

6 Query-efficient distance amplification

In this section we construct our query-efficient distance amplification procedure. We start by giving a somewhat more formal definition of locally decodable codes (compared to Definition 1.1) or, more precisely, a more formal definition of their non-adaptive counterparts. Recall that, informally, these are LDC in which the joint distribution of queries depends solely on the index one wishes to decode and is independent of the received word. By inspection, it is our understanding that the AEL distance amplification procedure also requires non-adaptivity.

Definition 6.1 (Locally decodable codes). *Let (C, Q, R) be a tuple of functions*

$$\begin{aligned}
C &: \Sigma_{\mathsf{in}}^k \to \Sigma_{\mathsf{out}}^n, \\
Q &: [k] \times \{0,1\}^r \to [n]^q, \\
R &: [k] \times \Sigma_{\mathsf{out}}^q \times \{0,1\}^r \to \Sigma_{\mathsf{in}}.
\end{aligned}$$

Define

$$D : [k] \times \Sigma_{\mathsf{out}}^n \times \{0,1\}^r \to \Sigma_{\mathsf{in}}$$

as follows. For $v \in [k]$, $y \in \Sigma_{\mathsf{out}}^n$, and $s \in \{0,1\}^r$, let

$$\begin{aligned}
Q(v, s) &= (u_1, \ldots, u_q), \\
D(v, y, s) &= R(v, y_{u_1}, \ldots, y_{u_q}, s).
\end{aligned}$$

*The tuple (C, Q, R) is called a (q, δ, ε)-*locally decodable code *(or (q, δ, ε)-LDC for short) if the following holds. For every $v \in [k]$, $x \in \Sigma_{\mathsf{in}}^k$, and $y \in \Sigma_{\mathsf{out}}^n$ for which $\mathsf{dist}(y, C(x)) \le \delta$, it holds that*

$$\Pr_{s \sim U_r} [D(v, y, s) = x_v] \ge 1 - \varepsilon.$$

We call the function C the encoding function, *Q the* querying function, *and R the* reconstruction function. *The induced function D is called the* decoding function. *The parameters k, n are referred to as the* message length *and the* block length, *respectively. The sets $\Sigma_{\mathsf{in}}, \Sigma_{\mathsf{out}}$ are called the input alphabet and output alphabet, respectively. We will be interested mostly in locally decodable codes in which $\Sigma_{\mathsf{in}} = \Sigma_{\mathsf{out}}$ in which case we refer to both as the* alphabet *of the code. The parameter r is called the* randomness complexity *of the LDC. We say the LDC is* explicit *if all three functions C, Q, R are polynomial-time computable. Note that then the decoding function D is also polynomial-time computable.*

6.1 The distance amplification procedure

In this section we present our query-efficient distance amplification procedure. We start by describing the building blocks we use and specify their parameters.

Building blocks.

- For $i = 1, 2$ let (C_i, Q_i, R_i) be a $(q_i, \delta_i, \varepsilon_i)$-LDC with message length k_i and block length n_i over the same alphabet Σ. We denote the rate k_i/n_i of C_i by ρ_i.
- Let (C_3, Q_3, R_3) be a family of $(q_3(k_3), \delta_3(k_3), \varepsilon_3(k_3))$-LDC having rate $\rho_3(k_3)$ for message length k_3. The code C_3 is also over the alphabet Σ. We will always work

with functions $q_3, \delta_3, \varepsilon_3, \rho_3$ that are monotone. More precisely, q_3 and ρ_3 are non-decreasing and δ_3, ε_3 are non-increasing. We sometimes write $q_3, \delta_3, \varepsilon_3, \rho_3$ without mentioning explicitly the message length, and by that refer to the largest k_3 considered in the construction for q_3, δ_3 and the smallest k_3 when considering ε_3, ρ_3. In any case, we assume (mostly for simplicity) that $\rho_3(k_3) \geq 1/2$ for all k_3.

- Set $\ell = n_1/k_2$. Let $G = (L, R, E)$ be a $(\delta_2/2, \delta_1)$-sampler with $|L| = \ell$ and $|R| = r$. Assume G is left-regular with left-degree $d = n_2$. Assume further that every right-vertex v of G has degree $\deg(v) \in [D/2, 2D]$, where D is the average right degree $D = \ell d/r = n_1/(r\rho_2)$.

How to think of the parameters? We think of C_1 as the code whose distance δ_1 we wish to amplify. Typically, the code C_2 has a much shorter message length $n_2 \ll n_1$. In all applications in this paper we take δ_2 to be either constant or slightly sub constant in n_1. The code C_3 has a larger block length than C_2 and, depending on the application, it has either a somewhat smaller or much smaller message length than n_1. We typically take $\delta_3 \approx \delta_2$. The rates of all three codes is taken to be constant and even close to one. Note that we take C_3 to be a family of codes, whereas C_1 and C_2 are codes with predetermined message lengths. The reason is that the sampler G is not necessarily right-regular, and in the construction, we associate codes from C_3 with the right vertices of G. Recall, though that the ratio of largest to smallest right-degree is bounded by 4, so that is a minor technicality.

To describe the LDC that is composed of these building blocks, we need to specify the encoding function, querying function and reconstruction function. We start by describing the encoding function.

The encoding function

Let $n = \sum_{v \in R} n_v$ where n_v is the block length of the code from the family C_3 having message length $k_v = \deg(v)$. We define the function $C : \Sigma^{k_1} \to \Sigma^n$ as follows. Let $x \in \Sigma^{k_1}$.

1. Compute $y = C_1(x) \in \Sigma^{n_1}$.
2. Partition y to $y = y^{(1)} \circ \cdots \circ y^{(\ell)}$ consecutive blocks, each of length k_2. Recall that, indeed, $n_1 = \ell k_2$.
3. For every $u \in [\ell]$ compute $z^{(u)} = C_2(y^{(u)}) \in \Sigma^{n_2}$.

4. For every $v \in [r]$ and $j \in [\deg(v)]$ let $(u, j') = \Gamma(v, j) \in [\ell] \times [n_2]$. Define the string $w^{(v)} \in \Sigma^{\deg(v)} = \Sigma^{k_v}$ as follows: for $j \in [\deg(v)]$, $(w^{(v)})_j = (z^{(u)})_{j'}$.
5. For every $v \in [r]$ compute $t^{(v)} = C_3(w^{(v)}) \in \Sigma^{n_v}$.
6. The output of the encoding function on input x is then defined by $C(x) = t^{(1)} \circ \cdots \circ t^{(r)} \in \Sigma^n$, where as usual we identify R with $[r]$.

By the construction of the encoding function, the message length and block length of the resulted code are k_1 and n, respectively. From here on we denote $k = k_1$.

The querying function

We denote the randomness complexity of C_1, C_2, C_3 by r_1, r_2, r_3, respectively. The randomness complexity of the resulting querying function will be $r = r_1+r_2+r_3$, and the query complexity will be $q \leq q_1 q_2 q_3$, where q_3 is taken to be the maximum query complexity taken over all right vertices. We turn to define the querying function $Q : [k] \times \{0,1\}^r \to [n]^q$ as follows. On inputs $p \in [k], s \in \{0,1\}^r$ we proceed as follows.

1. Partition $s = s_1 \circ s_2 \circ s_3$ where $|s_1| = r_1$, $|s_2| = r_2$, $|s_3| = r_3$.
2. Compute $(a_1, \ldots, a_{q_1}) = Q_1(p, s_1) \in [n_1]^{q_1}$.
3. For $i = 1, \ldots, q_1$
   (a) Set $u_i = \lceil a_i / k_2 \rceil$ and $b_i = 1 + ((a_i - 1) \bmod k_2)$. Informally, u_i is the "bucket" in which a_i resides and b_i is its location within the bucket. Note that we start the counting from 1 rather than 0, hence the slightly annoying addition and subtraction by 1 in the definition of b_i.
   (b) Compute $(t_1^{(i)}, \ldots, t_{q_2}^{(i)}) = Q_2(b_i, s_2) \in [n_2]^{q_2}$.
   (c) For $j = 1, \ldots, q_2$
      i. Let $(v^{(i,j)}, \hat{t}_j^{(i)}) = \Gamma(u_i, t_j^{(i)}) \in [r] \times [k_{v^{(i,j)}}]$.
      ii. Compute $(e_1^{(i,j)}, \ldots, e_{q_3}^{(i,j)}) = Q_3(\hat{t}_j^{(i)}, s_3) \in [n_{v^{(i,j)}}]^{q_3}$.
      iii. As before, we endow the right vertices of the sampler in a fixed (arbitrary) order by identifying R with $[r]$. For $h = 1, \ldots, q_3$ set $c^{(i,j,h)}$ to be the absolute location of $e_h^{(i,j)}$ in the ordering of R. That is, $c^{(i,j,h)} = e_h^{(i,j)} + \sum_{v < v^{(i,j)}} n_v$.
4. The result is then given by $Q(p, s) = (c^{(i,j,h)})_{(i,j,h) \in [q_1] \times [q_2] \times [q_3]}$.

Note that, indeed, the query complexity q of the querying function defined above is at most $q_1 q_2 q_3$ where, recall, $q_3 = q_3(2D)$. From here on we identify $[q]$ with $[q_1] \times [q_2] \times [q_3]$.

The reconstruction procedure

We define the reconstruction procedure $R : [k] \times \Sigma^q \times \{0,1\}^r \to \Sigma$ as follows. On inputs $p \in [k]$, $\sigma = (\sigma^{(i,j,h)})_{(i,j,h) \in [q_1] \times [q_2] \times [q_3]} \in \Sigma^q$, and $s \in \{0,1\}^r$, we proceed as follows.

1. Partition $s = s_1 \circ s_2 \circ s_3$ where $|s_1| = r_1$, $|s_2| = r_2$, $|s_3| = r_3$ as in the querying function.
2. For $i = 1, \ldots, q_1$
   (a) For $j = 1, \ldots, q_2$
      i. Denote $(z_1, \ldots, z_{q_3}) = (\sigma^{(i,j,1)}, \ldots, \sigma^{(i,j,q_3)})$.
      ii. Compute $y_j^{(i)} = R_3(\hat{t}_j^{(i)}, z_1, \ldots, z_{q_3}, s_3)$, where $\hat{t}_j^{(i)} = \hat{t}_j^{(i)}(p, s)$ as defined in the querying function.
   (b) Set $x_i = R_2(b_i, y_1^{(i)}, \ldots, y_{q_2}^{(i)}, s_2)$ where $b_i = b_i(p, s)$ as defined in the querying function.
3. The output is then given by $R(p, \sigma, s) = R_1(p, x_1, \ldots, x_{q_1}, s_1)$.

6.2 Analysis

In this section we analyze the LDC obtained above. We prove

Proposition 6.2. *With the notation of the previous section, C is a (q, δ, ε)-LDC, where*

$$\begin{aligned} q &\le q_1 q_2 q_3, \\ \delta &\ge \frac{\delta_2 \delta_3}{16}, \\ \varepsilon &\le \varepsilon_1 + (\varepsilon_2 + \varepsilon_3) n. \end{aligned}$$

Furthermore, C has rate $\rho_1 \rho_2 \rho_3$, where ρ_1, ρ_2 are as defined in the building blocks paragraph, and per our convention set above, $\rho_3 = \rho_3(D/2)$.

Remark regarding the distance. Note that the distance δ of the resulted code C is independent of δ_1, the poor distance of C_1 we set out to amplify. This is the key feature of the AEL distance amplification procedure (which our variant above, of course, maintains). It is yet another instance of a general strategy in pseudo-randomness that combines objects in such a way that the resulted object enjoys the upsides of the different parts and avoid their shortcomings. The Zig-Zag product is another classic example. But, of course, δ_1 has some effect on the resulted code. The effect δ_1 has on the code is via the query complexity.

Indeed, as the analysis will show, the smaller δ_1 is (i.e., the weaker the guarantee we have on the distance of C_1), the larger $k_2 = k_2(\delta_1)$ and $k_3 = k_3(\delta_1)$ must be, with a far stronger effect on k_3. More quantitatively, roughly speaking, by taking a sufficiently good sampler (e.g., the one that is given by Theorem 3.2), $k_2 \approx \operatorname{poly}\log(1/\delta_1)$ and $k_3 \approx \operatorname{poly}(1/\delta)$. This, in turn, effects the query complexities $q_2 = q_2(k_2)$ and $q_3 = q_3(k_3)$.

Proof. That the query complexity is $q \le q_1 q_2 q_3$ readily follows by the querying function, where recall that per our convention $q_3 = q_3(2D)$. To analyze the rate, recall that ρ_3 is a non-decreasing function. Further, our convention dictates that by writing ρ_3 without explicitly mentioning the message length, we refer to ρ_3 applied with the smallest message length taken by the construction, namely, $\rho_3 = \rho_3(D/2)$. Thus,

$$n = \sum_{v \in R} n_v = \sum_{v \in R} \frac{k_v}{\rho_3(k_v)} \le \frac{1}{\rho_3} \sum_{v \in R} k_v = \frac{\ell n_2}{\rho_3} = \frac{n_1}{\rho_2 \rho_3}.$$

Recall that $k = k_1 = \rho_1 n_1$ which shows that $\rho = k/n \ge \rho_1 \rho_2 \rho_3$.

We turn to analyze the distance δ and error ε. Let $x \in \Sigma^k$ and let $\widetilde{C}(x) \in \Sigma^n$ be such that $\mathsf{dist}(\widetilde{C}(x), C(x)) \le \delta$. Define the set of "errors", namely, the disagreements between $C(x)$ and $\widetilde{C}(x)$ by

$$B = \{i \in [n] \mid \widetilde{C}(x)_i \ne C(x)_i\}.$$

By assumption, $\mu(B) \le \delta$. The error set B induces errors "backwards" throughout the construction. We proceed by analyzing these induced errors. Recall that, in the encoding function, we defined for each $v \in [r]$ an element $t^{(v)} = t^{(v)}(x) \in \Sigma^{n_v}$. Partition $\widetilde{C}(x)$ to r substrings $\widetilde{C}(x) = \widetilde{t}^{(1)} \circ \cdots \circ \widetilde{t}^{(r)}$, where $\widetilde{t}^{(v)}$ has length n_v, and define the set

$$B_t = \left\{ v \in [r] \mid \mathsf{dist}\left(t^{(v)}, \widetilde{t}^{(v)}\right) \ge \delta_3 \right\}.$$

Informally, $v \in B_t$ if the adversary has introduced too many errors on the respective block to allow for correct decoding via D_3.

Claim 6.3. $\mu(B_t) \le 8\delta/\delta_3$.

Proof. For $v \in R$ let $e_v = \mathsf{dist}(t^{(v)}, \widetilde{t}^{(v)})$. We have that $\sum_{v \in R} e_v n_v \le \delta n$. On the other hand,

$$\sum_{v \in R} e_v n_v \ge \delta_3 \sum_{v \in B_t} n_v \ge \frac{\delta_3 D |B_t|}{2}.$$

But, per our assumption that $\rho_3 \ge 1/2$, and since $k_v \le 2D$ for all $v \in R$,

$$n = \sum_{v \in R} n_v \le 2 \sum_{v \in R} k_v \le 4Dr.$$

The proof follows by the above three inequalities. $\square$

For convenience we also denote $B_w = B_t$. Next, we define

$$B_z = \{u \in [\ell] \mid |\Gamma(u) \cap B_w| \geq \delta_2 n_2\}. \tag{6.1}$$

Claim 6.4. $\mu(B_z) \leq \delta_1$.

Proof. By Claim 6.3 and by our assumption on δ,

$$\mu(B_w) \leq \frac{8\delta}{\delta_3} = \frac{\delta_2}{2}.$$

Recall that G is a $(\delta_2/2, \delta_1)$-sampler. Thus, at most δ_1-fraction of the left vertices $u \in [\ell]$ satisfy $\mu(\Gamma(u) \cap B_w) \geq \mu(B_w) + \delta_2/2$. The proof then follows since $\mu(B_w) \leq \delta_2/2$. $\square$

Lastly, define

$$B_y = \left\{a \in [n_1] \;\middle|\; \left\lceil \frac{a}{k_2} \right\rceil \in B_z \right\}.$$

For $v \in [r]$, $b \in [k_v]$ we define the function $\widetilde{w}_b^{(v)} : \{0,1\}^{r_3} \to \Sigma$ as follows: on input $s_3 \in \{0,1\}^{r_3}$

$$\widetilde{w}_b^{(v)}(s_3) = D_3(b, \widetilde{t}^{(v)}, s_3).$$

Claim 6.5. *There exists a set $\mathcal{E}_3 \subseteq \{0,1\}^{r_3}$ with $\mu(\mathcal{E}_3) \leq \varepsilon_3 n$ such that for every $s_3 \in \{0,1\}^{r_3} \setminus \mathcal{E}_3$, $v \in [r] \setminus B_t$, and $b \in [k_3]$ it holds that $\widetilde{w}_b^{(v)}(s_3) = w_b^{(v)}$.*

Proof. Fix $v \in [r] \setminus B_t$. By the definition of B_t, one has that $\mathsf{dist}\left(t^{(v)}, \widetilde{t}^{(v)}\right) \leq \delta_3$. By the encoding function, $t^{(v)} = C_3(w^{(v)})$. Therefore, for every $b \in [k_3]$,

$$\Pr_{s_3 \sim U_{r_3}}\left[D_3(b, \widetilde{t}^{(v)}, s_3) \neq w_b^{(v)}\right] \leq \varepsilon_3.$$

The proof then follows by taking the union bound over all $v \in [r] \setminus B_t$ and $b \in [k_v]$ as indeed $\sum k_v \leq n$. $\square$

For $(u, j) \in [\ell] \times [n_2]$ we define the function $\widetilde{z}_j^{(u)} : \{0,1\}^{r_3} \to \Sigma$ as follows. For $s_3 \in \{0,1\}^{r_3}$ we have $\widetilde{z}_j^{(u)}(s_3) = \widetilde{w}_{j'}^{(v)}(s_3)$, where $(v, j') = \Gamma(u, j) \in [r] \times [k_v]$. Further define the function $\widetilde{z}^{(u)} : \{0,1\}^{r_3} \to \Sigma^{n_2}$ by

$$\widetilde{z}^{(u)}(s_3) = \widetilde{z}_1^{(u)}(s_3) \circ \cdots \circ \widetilde{z}_{n_2}^{(u)}(s_3).$$

Claim 6.6. *For every $u \notin B_z$ and $s_3 \in \{0,1\}^{r_3} \setminus \mathcal{E}_3$ it holds that*

$$\mathsf{dist}\left(\widetilde{z}^{(u)}(s_3), z^{(u)}\right) \leq \delta_2.$$

Proof. Fix $s_3 \in \{0,1\}^{r_3} \setminus \mathcal{E}_3$ and consider any $u \in [\ell] \setminus B_z$. By the encoding function, for every $j \in [n_2]$ it holds that $z_j^{(u)} = w_{j'}^{(v)}$, where $(v, j') = \Gamma(u, j)$. As $v \notin B_z$, at most $\delta_2 n_2$ of $j \in [n_2]$ satisfy $v \in B_w$. For every other j,

$$\widetilde{z}_j^{(u)} = \widetilde{w}_{j'}^{(v)}(s_3) = w_{j'}^{(v)} = z_j^{(u)},$$

proving the claim.

□

For $u \in [\ell], a \in [k_2]$ we define the function $\widetilde{y}_a^{(u)} : \{0,1\}^{r_2} \times \{0,1\}^{r_3} \to \Sigma$ as follows. On $(s_2, s_3) \in \{0,1\}^{r_2} \times \{0,1\}^{r_3}$,

$$\widetilde{y}_a^{(u)}(s_2, s_3) = D_2(a, \widetilde{z}^{(u)}(s_3), s_2).$$

Claim 6.7. *There exists a set $\mathcal{E}_2 \subseteq \{0,1\}^{r_2}$ with $\mu(\mathcal{E}_2) \le \varepsilon_2 n$ such that for every $u \in [\ell] \setminus B_z$, $a \in [k_2]$, and $(s_2, s_3) \in (\{0,1\}^{r_2} \setminus \mathcal{E}_2) \times (\{0,1\}^{r_3} \setminus \mathcal{E}_3)$ it holds that $\widetilde{y}_a^{(u)}(s_2, s_3) = y_a^{(u)}$.*

Proof. Fix $u \in [\ell] \setminus B_z$. By the encoding function $z^{(u)} = C_2(y^{(u)})$. Recall that

$$\widetilde{y}_a^{(u)}(s_2, s_3) = D_2(a, \widetilde{z}^{(u)}(s_3), s_2).$$

As $s_3 \notin \mathcal{E}_3$, $u \notin B_z$, Claim 6.6 implies $\mathsf{dist}(\widetilde{z}^{(u)}(s_3), z^{(u)}) \le \delta_2$. Therefore

$$\Pr_{s_2 \sim U_{r_2}}\left[D_2(a, \widetilde{z}^{(u)}(s_3), s_2) \neq y_a^{(u)}\right] \le \varepsilon_2.$$

The proof then follows by taking the union bound over all $a \in [k_2]$ and $u \in [\ell] \setminus B_z$, and noting that $k_2 \ell = n_1 \le n$. □

Claim 6.8. *For every $(s_2, s_3) \in (\{0,1\}^{r_2} \setminus \mathcal{E}_2) \times (\{0,1\}^{r_3} \setminus \mathcal{E}_3)$, it holds that*

$$\mathsf{dist}(\widetilde{y}(s_2, s_3), y) \le \delta_1,$$

where $\widetilde{y}(s_2, s_3)$ is the concatenation of the k_2-length strings $\left(\widetilde{y}^{(u)}(s_2, s_3) \mid u \in [\ell]\right)$.

Proof. Note that by Claim 6.7, the projection of the two strings $\widetilde{y}(s_2, s_3)$, y to a block corresponding to $u \notin B_z$ are in full agreement. The proof then follows by Claim 6.4. □

We now conclude the proof of Proposition 6.2. Let $p \in [k]$, by Claim 6.8, for every $(s_2, s_3) \in (\{0,1\}^{r_2} \setminus \mathcal{E}_2) \times (\{0,1\}^{r_3} \setminus \mathcal{E}_3)$, we have that $\mathsf{dist}(\widetilde{y}(s_2, s_3), y) \le \delta_1$. Since by the encoding function $y = C_1(x)$, it holds

$$\Pr_{s_1 \sim U_{r_1}}\left[D_1(p, \widetilde{y}(s_2, s_3), s_1) \neq x_p\right] \le \varepsilon_1.$$

The proof then follows since $\mu(\mathcal{E}_2) \le \varepsilon_2 n$ and $\mu(\mathcal{E}_3) \le \varepsilon_3 n$.

□

6.2.1 Proof of Theorem 1.5

In this short section we prove Theorem 1.5. We focus on the version that is based on non-explicit samplers, yielding non-explicit reductions. The explicit reduction, which entails a bit more technical work, is deferred to Section 6.3 and Section 6.6. We choose to focus on the non-explicit version first because we believe that understanding LDC in the information-theoretic level is, at present, a deeper and more urgent problem than the question of explicitness. Also, the parameters are easier to work with. For the information-theoretic version, we make use of the sampler that is given by Theorem 3.2. From here on we refer to the constant $c_{\mathsf{samp}} \geq 1$ that appears in that theorem.

Theorem 6.9. *Let C be a block-length-n $(q, \delta, 1/5)$-LDC over alphabet Σ having a constant rate. Let C' be a family of asymptotically good $(q'_n, \delta', 1/5)$-LDC, where q'_n is the query complexity when the code from the family is taken with block length n. Then, there exists an asymptotically good LDC over Σ, with constant error, having block length $\Theta(n)$ and query complexity*

$$q_{\text{new}} = O\left(q \cdot q'_{O(1/\delta)} \log(1/\delta) \log n\right). \tag{6.2}$$

Proof. Take C_1 to be the code C in the hypothesis of the theorem, namely, a code with block length $n_1 = n$ and distance $\delta_1 = \delta$. Recall that in the distance amplification procedure from Section 6.1, we make use of a $(\delta_2/2, \delta_1)$ sampler $G = ([\ell], [r], E)$ with $\ell = n_1/k_2$ and left-degree $d = n_2$. For the proof, we will instantiate the distance amplification procedure with the sampler that is given by Theorem 3.2. We take C_2 to be an asymptotically good code over Σ set with block length

$$n_2 = c_{\mathsf{samp}} \cdot \frac{\log(1/\delta_1)}{(\delta_2/2)^2} = O(\log(1/\delta)),$$

where δ_2 is the (constant) distance of C_2, having rate at least $1/2$. Note that this choice of parameters is as required by Theorem 3.2 from the left degree of the sampler. Clearly, C_2 has query complexity $O(\log(1/\delta))$ and error $\varepsilon_2 = 0$. As for the degree D_v of any given right vertex v of the sampler, note that the average right degree is

$$D = \frac{\ell d}{r} = \frac{d}{\delta \log(1/\delta)} = \frac{4c_{\mathsf{samp}}}{\delta_1 \delta_2^2} = \Theta\left(\frac{1}{\delta}\right).$$

Recall that, by Theorem 3.2, $D_v \in [D/2, 2D]$. For every length in this range, we take a code from the family C' having the required message length. We would like take the family of codes C_3 to be C' though we must reduce the error first. Indeed, note that the error ε of the code obtained by Proposition 6.2 is $\varepsilon_1 + n(\varepsilon_2 + \varepsilon_3)$. As mentioned in the introduction, one can reduce the error from $1/5$ to $1/(10n)$ by applying the decoding procedure for

$c \log n$ times, where c is some large enough constant, and output the symbol according to plurality. This has no effect on the rate or distance of C', and has a multiplicative $O(\log n)$ cost in query complexity. That is, the query complexity of C_3 is $O(q'_{O(1/\delta_1)} \log n)$. The proof then readily follows by Proposition 6.2. $\square$

Improving the query complexity further given low-error LDC. We remark that, if C' has error $O(1/n)$ to begin with, n being the block length of C, then one can skip the error reduction in the proof of Theorem 6.9, and get a slightly better query complexity. Indeed, this will save the $\log n$ factor in Equation (6.2). Moreover, observe that C_2 can be taken to be an LDC as well, rather than a standard code, which will reduce its deterioration on the query complexity from $O(\log(1/\delta))$ to $q'_{O(\log(1/\delta))}$. However, for that, one need the error of C_2 to be $O(1/n)$ as well. Assuming one can obtain such low-error LDC (note that an error of $1/n$ is at least exponentially-small in the length of C_2 since $\delta > 1/n$), the query complexity can be improved further to

$$q_{\text{new}} \leq q \cdot q'_{O(1/\delta)} q'_{O(\log(1/\delta))}.$$

We conclude this section by instantiating Theorem 6.9 with C' taken to be the state-of-the-art construction of asymptotically good LDC.

Theorem 6.10 ([KMRS17]). *Let Σ be a finite alphabet. Then, there exist constants δ, ρ and an explicit infinite family of $(q_k, \delta, 1/5)$-LDC, k being the message length, having query complexity $q_k = 2^{O(\sqrt{\log(k) \log\log k})}$.*

Using it, one gets query complexity

$$q_{\text{new}} \leq q \log(n) \cdot 2^{O(\sqrt{\log(1/\delta) \cdot \log\log(1/\delta)})} = q \log(n) (1/\delta)^{o(1)}.$$

6.3 Relaxing the assumption on the sampler G

In the distance amplification procedure described in Section 6.1, the sampler G is assumed to be a left-regular $(\delta_2/2, \delta_1)$-sampler in which every right degree is in $[D/2, 2D]$. In order for the reduction to result in an explicit code, we want to be able to plug in an explicit sampler in the distance amplification procedure, for which the bounds on the right degree may not hold. We now describe how a sampler that does not satisfy this assumption can be used even so. The change to the construction is detailed as follows.

Modified construction.

- For $i = 1, 2, 3$ let (C_i, Q_i, R_i) be as in Section 6.1. Assume further that $\delta_1 \leq \delta_2/8$.

- Set $\ell = n_1/k_2$. Let $G = (L, R, E)$ be a $(\delta_2/8, \delta_1)$-sampler with $|L| = \ell$ and $|R| = r$. Assume G is left-regular with left-degree $d = n_2$, and denote by $D = \frac{\ell d}{r}$ the average right degree (the right degrees may be arbitrary).

- The encoding function $C : \Sigma^{k_1} \to \Sigma^n$ is the same as in Section 6.1, but for the following change: if $v \in [r]$ has degree outside $[D/2, 2D]$ then discard it.

- The querying function is the same as in Section 6.1, but for the following change: if $v^{(i,j)}$ is a vertex with degree not in $[D/2, 2D]$, then set $(c^{(i,j,h)})_{h\in[q_3]}$ to be an empty tuple.

- The reconstruction procedure is the same as in Section 6.1, but for the following change: if i, j are such that $v^{(i,j)}$ is a vertex with degree not in $[D/2, 2D]$, then set $y_j^{(i)} = \perp$ (or, if one prefers to avoid the use of $\perp$, any $\sigma \in \Sigma$ can be used).

The amendments above have the effect that when encoding the blocks corresponding to right vertices, that are either too big or too small, the encoding discards such blocks and their contents, as if they were deleted. The querying function is changed so that whenever a location in these blocks needs to be queried, that query is skipped. The reconstruction procedure is accordingly changed so that whenever a location was not queried on the account of it residing in a block too big or too small, some arbitrary symbol (or $\perp$) is passed on instead. To analyze the altered distance-amplification procedure we start by proving two simple statements about samplers.

Lemma 6.11. *Let $G = ([\ell], [r], E)$ be a left-regular (ε, δ)-sampler with average right-degree D. Assume $\delta \le 1/4$. Then, G has at most $3\varepsilon r$ right vertices with degree less than $D/2$.*

Proof. Denote by d the left-degree of G. Define $A = \{v \in [r] \mid \deg(v) < D/2\}$. Since G is an (ε, δ) sampler, at least $(1-\delta)$ fraction of the left vertices have (at least) $(\frac{|A|}{r} - \varepsilon)d$ neighbors in A. Hence, A has at least $(1-\delta)\ell(\frac{|A|}{r} - \varepsilon)d$ edges entering it. Therefore, it must hold that

$$\frac{(1-\delta)\ell\left(\frac{|A|}{r} - \varepsilon\right)d}{|A|} < \frac{D}{2}.$$

As the average right degree is $D = \frac{\ell d}{r}$, and since by assumption $\delta \le 1/4$, we conclude that the average right-degree of A is at least

$$\frac{(1-\delta)\ell\left(\frac{|A|}{r} - \varepsilon\right)d}{|A|} = (1-\delta)D\left(1 - \frac{r\varepsilon}{|A|}\right) \ge \frac{3D}{4}\left(1 - \frac{r\varepsilon}{|A|}\right).$$

By the above two equation it follows that $|A| < 3\varepsilon r$. $\square$

Lemma 6.12. *Let $G = ([\ell], [r], E)$ be an (ε, δ)-sampler, which is d-left-regular and has average right-degree D. Assume $\varepsilon \geq \delta$. Then, G has at most $2\varepsilon r$ right vertices with degree larger than $2D$.*

Proof. Define $B = \{v \in [r] \mid \deg(v) > 2D\}$. At least $(1-\delta)$-fraction of the left vertices have at least $(1 - \frac{|B|}{r} - \varepsilon)d$ neighbors in $[r] \setminus B$, so $[r] \setminus B$ has at least $(1-\delta)\ell(1 - \frac{|B|}{r} - \varepsilon)d$ edges going into it. We therefore have that

$$2D|B| + (1-\delta)\ell\left(1 - \frac{|B|}{r} - \varepsilon\right)d \leq rD.$$

As $rD = \ell d$, it follows that

$$|B| \leq \left(\frac{\varepsilon + \delta - \delta\varepsilon}{1+\delta}\right) r \leq 2\varepsilon r.$$

□

We now wish to state the correctness of the changed construction.

Proposition 6.13. *The encoding function C of the modified construction is a (q, δ, ε)-LDC, where*

$$q \leq q_1 q_2 q_3,$$
$$\delta \geq \frac{\delta_2 \delta_3}{32},$$
$$\varepsilon \leq \varepsilon_1 + (\varepsilon_2 + \varepsilon_3)n.$$

Furthermore, C has rate $\rho_1\rho_2\rho_3$, where ρ_1, ρ_2 are as defined in the building blocks paragraph, and per our convention set above, $\rho_3 = \rho_3(D/2)$.

Proof. That the rate and query complexity are as stated is trivial, since the rate and query complexity can only be improved by this modification to the construction in which we discard some of the codeword symbols, and skip some of the queries. We now discuss the distance δ and error ε. Since the proof is almost identical to the proof of Proposition 6.2, we only state how to change the proof above to get a proof for the current proposition. Let

$$X = \{v \in R \mid \deg(v) \notin [D/2, 2D]\}$$

be the set of right vertices with "bad" degrees. Recall that these vertices are ignored by the modified construction. In particular, $n = \sum_{v \in R \setminus X} n_v$. The proof of Proposition 6.2 starts by defining the set

$$B = \{i \in [n] \mid \widetilde{C}(x)_i \neq C(x)_i\},$$

which is the set of "errors". It then goes on by defining another set, B_t, which is the set of "bad" right vertices, for which the adversary has introduced too many errors on the respective block. This is where we make a slight modification, ignoring the vertices in X. Formally, we define

$$B_t = \left\{ v \in R \setminus X \mid \mathsf{dist}\left(t^{(v)}, \widetilde{t}^{(v)}\right) \geq \delta_3 \right\}.$$

In the following claim we bound the density of B_t with respect to the set R (rather than with respect to $R \setminus X$).

Claim 6.14. $\mu_R(B_t) \leq \frac{8\delta}{\delta_3}$.

Proof. The proof is similar to the proof of Claim 6.3 though it takes into account our modifications as described above. For $v \in R \setminus X$ let $e_v = \mathsf{dist}(t^{(v)}, \widetilde{t}^{(v)})$. We have that $\sum_{v \in R \setminus X} e_v n_v \leq \delta n$. On the other hand,

$$\sum_{v \in R \setminus X} e_v n_v \geq \delta_3 \sum_{v \in B_t} n_v \geq \frac{\delta_3 D |B_t|}{2},$$

where the last inequality follows as for every $v \in B_t \subseteq R \setminus X$ it holds that $\deg(v) \geq D/2$. We also have, per our assumption, that $\rho_3 \geq 1/2$, and since $k_v \leq 2D$ for all $v \in R \setminus X$,

$$n = \sum_{v \in R \setminus X} n_v \leq 2 \sum_{v \in R \setminus X} k_v \leq 4Dr.$$

The proof follows by the above three inequalities, $\square$

As in Proposition 6.2, we also denote $B_w = B_t$. The definition of the set B_z is the same as in the proof of Proposition 6.2 with the modification that it "treats" the vertices in X as errors. Formally,

$$B_z = \{u \in [\ell] \mid |\Gamma(u) \cap (B_w \cup X)| \geq \delta_2 n_2\}, \tag{6.3}$$

Claim 6.15. $\mu(B_z) \leq \delta_1$.

Proof. By Claim 6.14, $\mu_R(B_w) \leq \frac{8\delta}{\delta_3}$. Now, G is a $(\delta_2/8, \delta_1)$-sampler. Thus, by Lemma 6.11 and Lemma 6.12 (which are applicable as $\delta_1 \leq \delta_2/8$ per our assumption), $\mu_R(X) \leq \frac{5\delta_2}{8}$. Hence, the density of $B_w \cup X$ with respect to R is

$$\mu_R(B_w \cup X) \leq \frac{8\delta}{\delta_3} + \frac{5\delta_2}{8} \leq \frac{7\delta_2}{8},$$

where the last inequality holds per our assumption $\delta \leq \delta_2\delta_3/32$. Recall that G is a $(\delta_2/8, \delta_1)$-sampler. Thus, at most δ_1-fraction of the left vertices $u \in [\ell]$ satisfy

$$\mu_{\Gamma(u)}(\Gamma(u) \cap (B_w \cup X)) \geq \mu_R(B_w \cup X) + \frac{\delta_2}{8},$$

and the proof follows. □

The rest of the proof is identical to the proof of Proposition 6.2. □

6.4 Reduction to LDC with polynomially-small (and even smaller) distance

In this section we prove the following corollary of Proposition 6.2. We then deduce from it Corollary 1.6 and Corollary 1.7 from the introduction.

Corollary 6.16. *There exists a universal constant c' such that the following holds. Let $c \geq 1$ be any constant. Let $\alpha :\to (0,1)$, $\beta :\to (0,1)$ be two monotone non-increasing functions that satisfy*

$$\alpha(n^{1.01}) \geq c'\beta(\log n) \cdot \log\log n. \tag{6.4}$$

Assume further that $\alpha(n) \leq 0.009$ and that $\beta(n) \leq 0.1$ for all $n \geq 1$. Assume there exists a family of $(q_\alpha(n), n^{-(1-\alpha(n))}, 1/5)$-LDC over alphabet Σ having rate $1-\beta(n)$. Then, for every sufficiently large n there exists a $(q, \delta, 1/5)$-LDC on block length $m \in [n, n^{1.01}]$ [10] *over Σ, where*

$$q = (q_\alpha(n)\log n)^{O\left(\frac{\log\log n}{\alpha(n^{1.01})}\right)},$$
$$\rho = 1 - O\left(\frac{\beta(\log n)\log\log n}{\alpha(n^{1.01})}\right),$$
$$\delta = \beta(\log n)^{O\left(\frac{\log\log n}{\alpha(n^{1.01})}\right)}.$$

To prove Corollary 6.16, we prove the following claim. In its statement, we refer to the constant $c_{\mathsf{samp}} \geq 1$ that is given by Theorem 3.2.

Claim 6.17. *Let $\beta_2 < 1/2$. Assume there exists a $(q_{\text{in}}, \delta_{\text{in}}, \varepsilon_{\text{in}})$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D/2, 2D]$ where*

$$D = \frac{4c_{\mathsf{samp}}n^{1-\alpha(n)}}{\beta_2^6}, \tag{6.5}$$

[10] The constant 1.01 in the exponent, which determines the density of lengths for which we can construct the stated codes, can be replaced by any constant strictly larger than 1, and even by $1 + o(1)$ for a "sufficiently large" $o(1)$. However, for ease of presentation, we stick with this fixed choice.

having rate $\rho_{\text{in}} \geq 1/2$. Then, under the hypothesis of Corollary 6.16, there exists a $(q_{\text{out}}, \delta_{\text{out}}, \varepsilon_{\text{out}})$-LDC over Σ with block-length n having rate ρ_{out}, where

$$\begin{aligned}
\frac{q_{\text{out}}}{q_{\text{in}}} &\leq \frac{4c_{\mathsf{samp}} \log n}{\beta_2^6} \cdot q_{\alpha}(n), \\
\frac{\delta_{\text{out}}}{\delta_{\text{in}}} &\geq \frac{\beta_2^3}{16}, \\
\frac{\rho_{\text{out}}}{\rho_{\text{in}}} &\geq (1-\beta_2)\left(1-\beta(n)\right), \\
\varepsilon_{\text{out}} &\leq \frac{1}{5} + n\varepsilon_{\text{in}}.
\end{aligned}$$

Proof. Let C_1 be the LDC from the hypothesis of Corollary 6.16 taken with block length $n_1 = n$. Let C_2 be a code set with message length $k_2 = \frac{4c_{\mathsf{samp}} \log n}{\beta_2^6}$, over Σ having rate $1-\beta_2$ and distance $\delta_2 = \beta_2^3$. A code with such parameters exists, over any alphabet, by the Gilbert-Varshamov bound.

Recall that in the distance amplification procedure (Section 6.1), we make use of a $(\delta_2/2, \delta_1)$-sampler $G = ([\ell], [r], E)$ with $\ell = n_1/k_2$ and left-degree n_2. For the proof of the claim, we will instantiate the distance amplification procedure with the sampler that is given by Theorem 3.2. To be able to use this sampler, we must verify that the left-degree is indeed large enough with respect to the parameters of the sampler. As, in our case, the left degree is n_2, we need to verify that

$$n_2 \geq c_{\mathsf{samp}} \cdot \frac{\log\left(1/\delta_1\right)}{(\delta_2/2)^2} = \frac{4c_{\mathsf{samp}}(1-\alpha(n))\log n}{\beta_2^6}. \tag{6.6}$$

However,

$$\frac{4c_{\mathsf{samp}}(1-\alpha(n))\log n}{\beta_2^6} \leq \frac{4c_{\mathsf{samp}} \log n}{\beta_2^6} = k_2,$$

and so, Equation (6.6) holds.

As for the degree D_v of any given right vertex v of the sampler, we have by Theorem 3.2 that $D_v \in [D/2, 2D]$, where

$$D = \frac{\ell d}{r} = \frac{4c_{\mathsf{samp}} n^{1-\alpha(n)}}{\beta_2^6},$$

which equals to D as defined in Equation 6.5. Thus, we may use C_{in} as in the hypothesis of the claim. We are therefore in a position to apply Proposition 6.2. The assertions regarding the query complexity, distance and rate readily follow by Proposition 6.2. That the error is bounded as stated readily follows by noting that $\varepsilon_2 = 0$. □

It will be more convenient to have no error loss in the reduction that is given by Claim 6.17. This is easily achievable by amplifying the error of the input code before applying the previous claim.

Corollary 6.18. *Let $\beta_2 < 1/2$. Assume there exists a $(q_{\text{in}}, \delta_{\text{in}}, 1/4)$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D/2, 2D]$, where D is as in Equation (6.5), having rate $\rho_{\text{in}} \geq 1/2$. Then, under the hypothesis of Corollary 6.16, there exists a $(q_{\text{out}}, \delta_{\text{out}}, 1/4)$-LDC over Σ with block-length n having rate ρ_{out}, where*

$$\frac{q_{\text{out}}}{q_{\text{in}}} \leq \frac{100 c_{\mathsf{samp}} \log^2 n}{\beta_2^6} \cdot q_\alpha(n),$$
$$\frac{\delta_{\text{out}}}{\delta_{\text{in}}} \geq \frac{\beta_2^3}{16},$$
$$\frac{\rho_{\text{out}}}{\rho_{\text{in}}} \geq (1-\beta_2)\,(1-\beta(n))\,.$$

Proof. Let r be a parameter we set later on. Define the code C' to be the code C_{in} though with the following decoder. To decode C', apply the decoder of C_{in} for r times and return the symbol according to plurality. Clearly, the rate and distance remain intact. By a simple application of the Chernoff bound, one can show that the error of C' is $2^{-\Omega(r)}$. The query complexity of C' is then rq_{in}. Thus, by taking $r = c\log n$ for a sufficiently large constant c, we can get a code with error $1/n^2$. The query complexity is then increased by a multiplicative $O(\log n)$ factor. The proof then follows by applying Claim 6.17 to C'. $\square$

With Corollary 6.18 we are ready to prove Corollary 6.16.

Proof of Corollary 6.16. The construction of the asserted code is obtained by devising a sequence of LDC $C'_0, C'_1, C'_2, \ldots$ where C'_0 is taken to be a code over Σ with block length

$$n_0 = 2\left(\frac{16 c_{\mathsf{samp}}}{\beta(\log n)^6}\right)^{8/\alpha(n^{1.01})}, \tag{6.7}$$

having rate $\rho_0 = 1 - \beta(\log n)$ and distance $\beta(\log n)^3$. A code with such parameters exists, over any alphabet, by the Gilbert-Varshamov bound. Clearly, as an LDC, this code has error $\varepsilon_0 = 0$ and query complexity n_0. For $t > 0$, the code C'_t is obtained by applying Corollary 6.18 with the code C'_{t-1} as C_{in} in the notations of the corollary and using $\beta_2 = \beta(\log n)$. Denote the message length and block length of C'_t by k_t and n_t, respectively. By construction, for every integer $t \geq 1$ such that $n_t \leq n^{1.01}$ we have that

$$k_{t-1} \leq \frac{8 c_{\mathsf{samp}} n_t^{1-\alpha(n_t)}}{\beta(\log n)^6} \leq \frac{8 c_{\mathsf{samp}} n_t^{1-\alpha(n^{1.01})}}{\beta(\log n)^6}, \tag{6.8}$$

where we used the fact that $\alpha(n)$ is non-increasing. By Corollary 6.18,

$$\rho_t = \frac{k_t}{n_t} \geq (1-\beta(\log n))^2 \rho_{t-1},$$

and so

$$\rho_t \geq (1-\beta(\log n))^{2t}\rho_0 = (1-\beta(\log n))^{2t+1}.$$

In particular, for every $t \leq \frac{1}{4\beta(\log n)}$ we get

$$\rho_t \geq (1-\beta(\log n))^{1+\frac{1}{2\beta(\log n)}} \geq \frac{1}{2}.$$

The the last inequality follows since the function $(1-x)^{1+\frac{1}{2x}} \geq \frac{1}{2}$ for all $x \leq 0.1$ and, recall, we assume that the function β is bounded above by 0.1. By Equation (6.8) we have that for every $t \leq \frac{1}{4\beta(\log n)}$,

$$n_{t-1} \leq 2k_{t-1} \leq \frac{16c_{\mathsf{samp}} n_t^{1-\alpha(n^{1.01})}}{\beta(\log n)^6}.$$

Thus,

$$n_t \geq \left(\frac{n_{t-1}\beta(\log n)^6}{16c_{\mathsf{samp}}}\right)^{\frac{1}{1-\alpha(n^{1.01})}}. \tag{6.9}$$

One can prove the following easy claim by induction.

Claim 6.19. *Let $(n_t)_{t\in}$ be a sequence of positive integers such that $n_t \geq (n_{t-1}/a)^b$ for some $a, b > 1$. Then, for every $t \geq 1$ we have that $n_t \geq (n_0/a^{h(b,t)})^{b^t}$, where $h(b,t) = \sum_{i=0}^{t-1} \frac{1}{b^i}$.*

With the notation of Claim 6.19, we have

$$h\left(\frac{1}{1-\alpha(n^{1.01})}, t\right) = \sum_{i=0}^{t-1}(1-\alpha(n^{1.01}))^i \leq \frac{1}{\alpha(n^{1.01})}.$$

By applying Claim 6.19 with $a = 16c_{\mathsf{samp}}/\beta(\log n)^6$ and $b = \frac{1}{1-\alpha(n^{1.01})}$ we get that for every t such that $n_t \leq n^{1.01}$ it holds

$$n_t \geq \left(\frac{n_0}{\left(\frac{16c_{\mathsf{samp}}}{\beta(\log n)^6}\right)^{1/\alpha(n^{1.01})}}\right)^{\left(\frac{1}{1-\alpha(n^{1.01})}\right)^t} \geq 2^{\left(\frac{1}{1-\alpha(n^{1.01})}\right)^t},$$

where for the last equality we used our of n_0 given in Equation (6.7). We now wish to take t' to be the least integer for which the right hand side is larger or equal than n. However, we must make sure that such t' exists. Indeed, the above analysis only works for t such that both $n_t \leq n^{1.01}$ and $t \leq \frac{1}{4\beta(\log n)}$ holds. So, one must verify that there exists a $t' \leq \frac{1}{4\beta(\log n)}$ for which $n \leq n_{t'} \leq n^{1.01}$. To see this, recall that $k \in [D/2, 2D]$ where D is as given by Equation (6.5). Hence,

$$n_{t-1} \geq k_{t-1} \geq \frac{2c_{\mathsf{samp}} n_t^{1-\alpha(n)}}{\beta_2^6} \geq n_t^{1-\alpha(n^{1.01})},$$

Hence, if $n_{t-1} < n$ then

$$n_t < n^{\frac{1}{1-\alpha(n^{1.01})}} < n^{1.01},$$

where the last inequality follows as $\alpha(n_t) \le 0.009$. Thus,

$$t' = \Theta\left(\frac{\log\log n}{\log\left(\frac{1}{1-\alpha(n^{1.01})}\right)}\right) = \Theta\left(\frac{\log\log n}{\alpha(n^{1.01})}\right),$$

and we can thus see that $t' \le \frac{1}{4\beta(\log n)}$ per our assumption that is given by Equation (6.4).

It is easy to verify that the query complexity $q_{t'}$ of and distance $\delta_{t'}$ of $C'_{t'}$ are

$$q_{t'} = \left(\frac{\log n}{\beta(\log n)}\right)^{\Theta(t')},$$
$$\delta_{t'} = \beta(\log n)^{\Theta(t')}.$$

As for the rate,

$$\rho_{t'} \ge (1-\beta(\log n))^{\Theta(t')} = 1 - O\left(\frac{\beta(\log n)\log\log n}{\alpha(n^{1.01})}\right),$$

where the last equality follows by Equation (6.4). Finally, the error of $C'_{t'}$ can be reduced from $1/4$ to $1/5$ with no asymptotic overhead in query complexity, and so $C'_{t'}$ has all the asserted properties. $\square$

6.4.1 Proofs of Corollary 1.6 and Corollary 1.7

In this short section prove Corollary 1.6 and Corollary 1.7.

Proof of Corollary 1.6. With the hypothesis of the corollary, we may apply Corollary 6.16 with $\alpha(n)$ and $\beta(n)$ in the notation of Corollary 6.16 set to $\alpha(n) = \min(\alpha, 0.009)$ and $\beta(n) = \frac{1}{\log^2 n}$ (and, in fact, taking $\beta(n) = \frac{c}{\log n}$ for sufficiently small constant $c > 0$ will do as well). Note that Equation (6.4) holds with this choice. Corollary 6.16 then yields a $(q_1, \delta_1, \varepsilon_1 = 1/5)$-LDC, where

$$q_1 = (q_\alpha(n) \cdot \log n)^{O(\log\log n)},$$
$$\delta_1 = 2^{-O(\log\log(n)\log\log\log n)},$$
$$\rho_1 = 1 - O\left(\frac{1}{\log\log n}\right).$$

Recall that by the Katz-Trevisan bound [KT00], constant rate LDC with distance δ have query complexity $\Omega(\log(\delta n/\log n))$ (see, e.g., [ZDb]). Thus, $q_\alpha(n) = \Omega(\log n)$ and so, in fact, $q_1 = q_\alpha(n)^{O(\log\log n)}$. The resulted code is obtained by amplifying the distance from δ_1 to constant. Indeed, one can invoke, say, the AEL distance amplification procedure. Since $1/\delta = o(q_1)$, the proof follows. $\square$

Proof of Corollary 1.7. With the hypothesis of the corollary, we may apply Corollary 6.16 with $\alpha(n) = 1/(\log\log n)^c$ and $\beta(n) = 1/(\log n)^{c+2}$ in the notation of Corollary 6.16. Note that Equation (6.4) holds with this choice. Corollary 6.16 then yields a $(q_1, \delta_1, \varepsilon_1 = 1/5)$-LDC, where

$$\begin{aligned} q_1 &= (q_\alpha(n) \cdot \log n)^{O((\log\log n)^{c+1})}, \\ \delta_1 &= 2^{-O((\log\log n)^{c+1} \cdot \log\log\log n)}, \\ \rho_1 &= 1 - O\left(\frac{1}{\log\log n}\right). \end{aligned}$$

By the Katz-Trevisan bound [KT00], $q_\alpha(n) = \Omega(\log n)$ and so, in fact, $q_1 = q_\alpha(n)^{O((\log\log n)^{c+1})}$. The resulted code is obtained by amplifying the distance from δ_1 to constant. By invoking the AEL distance amplification procedure. □

6.5 Proof of Corollary 1.8

In this section we prove Corollary 1.8 based on Proposition 6.2. We start by prove thing following.

Corollary 6.20. *There exists a constant $c \geq 1$ such that the following holds. Let $0 < \alpha < 1$ be an arbitrary constant, and $\beta :\to (0, 1)$ a monotone non-increasing function that satisfy*

$$2^{-\frac{1}{6}(\log n)^\alpha} \leq \beta(n) \leq \frac{c}{\log\log n} \tag{6.10}$$

Assume there exists a family of $(q_\alpha(n), 2^{-(\log n)^\alpha}, 1/5)$-LDC over alphabet Σ having rate $1-\beta(n)$. Then, for every sufficiently large n there exists a $(q, \delta, 1/5)$-LDC on block length m over Σ, for which $\log m \in [\log n, (\log n)^{1/(1-\alpha)}]$, and

$$\begin{aligned} q &= q_\alpha(n)^{O(\log\log\log n)}, \\ \rho &= 1 - O\left(\beta(\log n) \log\log\log n\right), \\ \delta &= \beta(\log n)^{O(\log\log\log n)}. \end{aligned}$$

To prove Corollary 6.20, we prove the following claim. In its statement we refer to the constant $c_{\mathsf{samp}} \geq 1$ that is given by Theorem 3.2.

Claim 6.21. *Let $\beta_2 < 1/2$. Assume there exists a $(q_{\text{in}}, \delta_{\text{in}}, \varepsilon_{\text{in}})$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D/2, 2D]$ where*

$$D = \frac{4c_{\mathsf{samp}} 2^{(\log n)^\alpha}}{\beta_2^6}, \tag{6.11}$$

having rate $\rho_{\text{in}} \geq 1/2$. Then, under the hypothesis of Corollary 6.20, there exists a $(q_{\text{out}}, \delta_{\text{out}}, \varepsilon_{\text{out}})$-LDC over Σ with block length n having rate ρ_{out}, where

$$\frac{q_{\text{out}}}{q_{\text{in}}} \leq \frac{8c_{\mathsf{samp}}(\log n)^{\alpha}}{\beta_2^6} \cdot q_{\alpha}(n),$$
$$\frac{\delta_{\text{out}}}{\delta_{\text{in}}} \geq \frac{\beta_2^3}{16},$$
$$\frac{\rho_{\text{out}}}{\rho_{\text{in}}} \geq (1-\beta_2)\,(1-\beta(n))\,,$$
$$\varepsilon_{\text{out}} \leq \frac{1}{5} + n\varepsilon_{\text{in}}.$$

Proof. Let C_1 be the LDC from the hypothesis of Corollary 6.20 taken with block length $n_1 = n$. Let C_2 be a code set with message length $k_2 = \frac{4c_{\mathsf{samp}}(\log n)^{\alpha}}{\beta_2^6}$, over Σ having rate $1-\beta_2$ and distance $\delta_2 = \beta_2^3$. A code with such parameters exists, over any alphabet, by the Gilbert-Varshamov bound.

In the distance amplification procedure (Section 6.1), we make use of a $(\delta_2/2, \delta_1)$ sampler $G = ([\ell],[r],E)$ with $\ell = n_1/k_2$ and left-degree $d = n_2$. For the proof of the claim, we will instantiate the distance amplification procedure with the sampler that is given by Theorem 3.2, and so we must verify that the left-degree is indeed large enough with respect to the parameters of the sampler. As, in our case, the left degree is n_2, we need to verify that

$$n_2 \geq c_{\mathsf{samp}} \cdot \frac{\log(1/\delta_1)}{(\delta_2/2)^2} = \frac{4c_{\mathsf{samp}}(\log n)^{\alpha}}{\beta_2^6}, \tag{6.12}$$

which indeed holds as the right hand side equals k_2.

As for the degree D_v of any given right vertex v of the sampler, we have by Theorem 3.2 that $D_v \in [D/2, 2D]$, where

$$D = \frac{\ell d}{r} = \frac{4c_{\mathsf{samp}} n^{1-\alpha(n)}}{\beta_2^6},$$

is as defined in Equation 6.11. Thus, we may use C_{in} as in the hypothesis of the claim. We are therefore in a position to apply Proposition 6.2, and the proof readily follows. $\square$

As in the previous section, it will be convenient to have no error loss in the reduction that is given by Claim 6.17. This is easily achievable by amplifying the error of the input code before applying the previous claim. We state the following corollary whose proof is similar to the proof of Corollary 6.18 and so we omit it.

Corollary 6.22. *Let $\beta_2 < 1/2$. Assume there exists a $(q_{\text{in}}, \delta_{\text{in}}, 1/4)$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D/2, 2D]$ where D is as defined in Equation (6.11),* *having rate $\rho_{\text{in}} \geq 1/2$. Then, under the hypothesis of Corollary 6.20, there exists a $(q_{\text{out}}, \delta_{\text{out}}, 1/4)$-LDC over Σ with block length n having rate ρ_{out}, where*

$$\frac{q_{\text{out}}}{q_{\text{in}}} \leq \frac{\log^2 n}{\beta_2^6} \cdot q_\alpha(n),$$
$$\frac{\delta_{\text{out}}}{\delta_{\text{in}}} \geq \frac{\beta_2^3}{16},$$
$$\frac{\rho_{\text{out}}}{\rho_{\text{in}}} \geq (1-\beta_2)\,(1-\beta(n))\,.$$

With Corollary 6.22 we are ready to prove Corollary 6.20.

Proof of Corollary 6.20. The construction of the asserted code starts by devising a sequence of LDC $C'_0, C'_1, C'_2, \ldots$ where C'_0 is taken to be a code over Σ with block length $n_0 = \log n$, having rate $1 - \beta(\log n)$ and distance $\beta(\log n)^3$. We obtain such code using Lemma 3.6. Clearly, as an LDC, this code has error $\varepsilon_0 = 0$ and query complexity n_0. For $t > 0$, the code C'_t is obtained by applying Corollary 6.22 with the code C'_{t-1} as C_{in} in the notations of the corollary and using $\beta_2 = \beta(\log n)$. Denote the message length and block length of C'_t by k_t and n_t, respectively. By construction, for every integer $t \geq 1$ such that $n_t \leq 2^{(\log n)^{1/(1-\alpha)}}$ we have that

$$k_{t-1} \leq \frac{8c_{\mathsf{samp}} 2^{(\log n_t)^\alpha}}{\beta_2^6}$$

By Corollary 6.18,

$$\rho_t = \frac{k_t}{n_t} \geq (1-\beta(\log n))^2 \rho_{t-1},$$

and so

$$\rho_t \geq (1-\beta(\log n))^{2t}\rho_0 = (1-\beta(\log n))^{2t+1}.$$

In particular, for every $t \leq \frac{1}{4\beta(\log n)}$ we get

$$\rho_t \geq (1-\beta(\log n))^{1+\frac{1}{2\beta(\log n)}} \geq \frac{1}{2}.$$

The the last inequality follows since the function $(1-x)^{1+\frac{1}{2x}} \geq \frac{1}{2}$ for all $x \leq 0.1$. Note that, indeed, by our assumption on β if follows that for a large enough n, $\beta(n)$ is bounded above by 0.1. Therefore,

$$n_{t-1} \leq 2k_{t-1} \leq \frac{8c_{\mathsf{samp}} 2^{(\log n_t)^\alpha}}{\beta_2^6}.$$

Now, per our assumption that is given by Equation (6.10), we have that

$$\beta_2 = \beta(\log n) \geq 2^{-\frac{1}{6}(\log\log n)^\alpha} \geq 2^{-\frac{1}{6}(\log n_t)^\alpha},$$

where the last inequality follows as $n_0 = \log n$. Thus, we get

$$n_{t-1} \leq 8c_{\mathsf{samp}} 2^{2(\log n_t)^\alpha} \leq 8^{(\log n_t)^\alpha}.$$

Thus, $\log n_t \geq \left(\frac{\log n_{t-1}}{3}\right)^{1/\alpha}$. By Claim 6.19, we get

$$\log n_t \geq \left(\frac{\log n_0}{3^{\frac{1}{1-\alpha}}}\right)^{\frac{1}{\alpha^t}} \geq 2^{\frac{1}{\alpha^t}}.$$

We now take t' to be the least integer for which the right hand side is larger or equal than $\log n$. Note that $t' = \Theta(\log\log\log n)$. However, the above analysis only holds only for $t \leq \frac{1}{4\beta(\log n)}$ and so one must verify that $t' \leq \frac{1}{4\beta(\log n)}$ which does indeed hold per our assumption that is given by Equation (6.10).

By the above, we get that $C'_{t'}$ is a $(q', \delta', 1/4)$-LDC having rho ρ' where

$$\begin{aligned}
q' &= (q_\alpha(n)\log n)^{O(\log\log\log n)}, \\
\rho' &= 1 - O\left(\beta(\log n)\log\log\log n\right), \\
\delta' &= \beta(\log n)^{O(\log\log\log n)}.
\end{aligned}$$

By [KT00], $q_\alpha(n) = \Omega(\log n)$ and so, in fact, $q' = q_\alpha(n)^{O(\log\log\log n)}$. The final code is obtained by amplifying the distance from δ' to constant. By invoking, say, the AEL distance amplification procedure. □

6.6 Explicit reduction to LDC with polynomially-small distance

In this section we show a result similar to the one proven in Section 6.4, but with an explicit reduction that yields an explicit code. Throughout this section we assume $\Sigma = \mathbb{F}_p$ for some prime power p (this is needed for the existence of explicit base codes). We prove the following corollary of Proposition 6.13

Corollary 6.23. *Let $\alpha > 0$ be a constant. Let $\beta :\rightarrow (0,1)$ be a monotone non-increasing function that satisfies*

$$\frac{1}{n} \leq \beta(n) \leq \frac{\log(1/\alpha)}{24\log n}. \tag{6.13}$$

Assume there exists a family of explicit $(q_\alpha(n), n^{-\alpha}, 1/5)$-LDC over alphabet Σ having rate $1-\beta(n)$ for block-length n. Then, for every sufficiently large n there exists an explicit $(q, \delta, 1/5)$-LDC on block length $\mathrm{poly}(n)$ over Σ, where

$$\begin{aligned}
q &= (q_\alpha(n)\log n)^{O(\log\log n)}, \\
\rho &= 1 - O\left(\beta(\log n)\log\log n\right), \\
\delta &= \beta(\log n)^{O(\log\log n)}.
\end{aligned}$$

Note that the distance δ above can then be further amplified to a constant, at the expense of lowering the rate from $1 - o(1)$ to some constant, without asymptotic cost in query complexity. Indeed, in the above corollary, $1/\delta = \mathrm{poly}(q)$ per our assumption that $\beta(\log n) \geq 1/\log n$.

To prove Corollary 6.23, we prove the following claim. In what follows, we refer to $c = c(\Delta)$ - the function that appears in the statement of Theorem 3.4.

Claim 6.24. *There exists a universal constant $\beta_0 \leq \frac{1}{2}$ such that the following holds. Let n be an integer, and $\beta_2 \in (\frac{1}{\log n}, \beta_0)$. Assume there exists an explicit $(q_{\text{in}}, \delta_{\text{in}}, \varepsilon_{\text{in}})$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D'/2, 4D']$ where $D' = D'(1/\sqrt{\alpha}, \delta_2/8, \delta_1)$ is as defined in Equation (3.2), having rate $\rho_{\text{in}} \geq 1/2$. Then, under the hypothesis of Corollary 6.23, there exists an explicit $(q_{\text{out}}, \delta_{\text{out}}, \varepsilon_{\text{out}})$-LDC over Σ with block-length n having rate ρ_{out}, where*

$$\begin{aligned}
\frac{q_{\text{out}}}{q_{\text{in}}} &\leq (\log n)^{10c(1/\sqrt{\alpha})} \cdot q_\alpha(n), \\
\frac{\delta_{\text{out}}}{\delta_{\text{in}}} &\geq \frac{\beta_2^3}{16}, \\
\frac{\rho_{\text{out}}}{\rho_{\text{in}}} &\geq (1 - \beta_2)\,(1 - \beta(n))\,, \\
\varepsilon_{\text{out}} &\leq \frac{1}{5} + n\varepsilon_{\text{in}}.
\end{aligned}$$

Proof. Let C_1 be the LDC from the hypothesis of Corollary 6.23 taken with block length $n_1 = n$. Set $\delta_2 = \beta_2^3$. By Theorem 3.4, invoked with $\Delta = 1/\sqrt{\alpha}$, there exists an explicit $(\delta_2/8, \delta_1)$-sampler with $z = n/(1 - \beta_2)$ edges. By Theorem 3.4, G has left-degree

$$d = \left(\frac{8}{\delta_2} \log \frac{1}{\delta_1}\right)^c = \left(\frac{8}{\beta_2^3} \alpha \log n\right)^c,$$

where $c = c(\Delta) = c(1/\sqrt{\alpha})$ is the constant as defined in Theorem 3.4. Note that since $\beta_2 \geq 1/\log n$ we have that $d \leq (\log n)^{10c}$. We also have that the average right-degree D is in $[D', 2D']$, where

$$D' = \frac{d}{2} \cdot \left(\frac{2}{\delta_1}\right)^{\Delta+1} \leq n^{2\sqrt{\alpha}},$$

where the inequality holds for all sufficiently large n.

Let C_2 be an explicit code set with message length $k_2 = (1 - \beta_2)d$ over Σ having rate $1 - \beta_2$ and distance $\delta_2 = \beta_2^3$. An explicit code with such parameters exists, by Lemma 3.6, as we can choose β_0 to be smaller than the least β for which the lemma holds.

We now want to instantiate the distance amplification procedure with C_1, C_2, the sampler G, and the code family C_{in} as C_3. Note that since the right degrees of the sampler

G are not necessarily bounded, we use the relaxed distance amplification of Section 6.3. Recall that it is a prerequisite of the distance amplification procedure that the sampler has n_1/k_2 left vertices, and that $n_2 = d$, the degree of the sampler. Both of these hold, as indeed, the block length of C_2 is $\frac{1}{1-\beta_2}(1-\beta_2)d = d$, and the number of left vertices of the sampler is $\frac{z}{d} = \frac{n}{d(1-\beta_2)} = n_1/k_2$. Further note that the distance amplification procedure requires that the family C_3 contains a code with message length k_3 for every $k_3 \in [D/2, 2D]$, and this is indeed satisfied by the assumption regarding the message lengths of the code family C_{in}, of the hypothesis of the claim.

With C_1, C_2, G and C_{in} at hand, we can now apply Proposition 6.13 of the distance amplification procedure. The assertions regarding the query complexity, distance and rate readily follow by Proposition 6.2. That the error is bounded as stated readily follows by noting that $\varepsilon_2 = 0$. □

As in the previous sections, it will be convenient to have no error loss in the reduction that is given by Claim 6.24. This is easily achievable by amplifying the error of the input code before applying the previous claim. We state the following corollary whose proof is similar to the proof of Corollary 6.18 and so we omit it.

Corollary 6.25. *There exists a universal constant $\beta_0 \le \frac{1}{2}$ for which the following holds. Let $\beta_2 \in (\frac{1}{\log n}, \beta_0)$. Assume there exists an explicit $(q_{\text{in}}, \delta_{\text{in}}, 1/4)$-LDC C_{in} over alphabet Σ for every message length $k_{\text{in}} \in [D'/2, 4D']$ where $D' = D'(1/\sqrt{\alpha}, \delta_2/8, \delta_1)$ is as defined in Equation (3.2), having rate $\rho_{\text{in}} \ge 1/2$. Then, under the hypothesis of Corollary 6.23, there exists an explicit $(q_{\text{out}}, \delta_{\text{out}}, 1/4)$-LDC over Σ with block-length n having rate ρ_{out}, where*

$$\frac{q_{\text{out}}}{q_{\text{in}}} \le (\log n)^{10c(1/\sqrt{\alpha})} \cdot q_\alpha(n),$$
$$\frac{\delta_{\text{out}}}{\delta_{\text{in}}} \ge \frac{\beta_2^3}{16},$$
$$\frac{\rho_{\text{out}}}{\rho_{\text{in}}} \ge (1-\beta_2)\,(1-\beta(n))\,.$$

With Corollary 6.25 we are ready to prove Corollary 6.23.

Proof of Corollary 6.23. The construction of the asserted code is obtained by devising a sequence of LDC $C'_0, C'_1, C'_2, \ldots$ where C'_0 is taken to be a code over Σ with block length $n_0 = \log n$ having rate $\rho_0 = 1 - \beta(\log n)$ and distance $\beta(\log n)^3$. By Lemma 3.6 such an explicit code exists, for every large enough n (the lemma holds for every small enough β, and indeed by Equation (6.13), $\beta(n)$ is decreasing). Clearly, as an LDC, this code has error $\varepsilon_0 = 0$ and query complexity n_0. For $t > 0$, the code C'_t is obtained by applying Corollary 6.25 with the code C'_{t-1} as C_{in} in the notations of the corollary and using $\beta_2 = \beta(\log n)$. Note that per our assumption given by Equation (6.13), this choice satisfies $\beta_2 \geq \frac{1}{\log n}$, and for large enough n, $\beta(n) \leq \beta_0$, and so we can apply the corollary. Denote the message length and block length of C'_t by k_t and n_t, respectively. By construction, for every integer $t \geq 1$ we have that

$$k_{t-1} \leq n_t^{2\sqrt{\alpha}} \leq n_t^{\alpha^{1/4}}, \tag{6.14}$$

where the last inequality holds for all large enough n. By Corollary 6.25,

$$\rho_t = \frac{k_t}{n_t} \geq (1 - \beta(\log n))^2 \rho_{t-1},$$

and so

$$\rho_t \geq (1 - \beta(\log n))^{2t} \rho_0 = (1 - \beta(\log n))^{2t+1}.$$

In particular, for every $t \leq \frac{1}{4\beta(\log n)}$ we get

$$\rho_t \geq (1 - \beta(\log n))^{1 + \frac{1}{2\beta(\log n)}} \geq \frac{1}{2}.$$

The last inequality follows since the function $(1-x)^{1+\frac{1}{2x}} \geq \frac{1}{2}$ for all $x \leq 0.1$, and for every large enough n, $\beta(n) \leq 0.1$. By Equation (6.14) we have that for every $t \leq \frac{1}{4\beta(\log n)}$,

$$n_{t-1} \leq 2k_{t-1} \leq 2n_t^{\alpha^{1/4}} \leq n_t^{\alpha^{1/5}}.$$

Thus,

$$n_t \geq n_0^{\frac{1}{\alpha^{t/5}}}. \tag{6.15}$$

It follows that by taking $t' = \lceil \frac{5 \log\log n}{\log(1/\alpha)} \rceil$ we get that $n_{t'} \geq n$. However we need to verify that this choice satisfies $t' \leq \frac{1}{4\beta(\log n)}$ for the above analysis to hold. Indeed per our assumption given by Equation (6.13), it holds that $\frac{6 \log\log n}{\log(1/\alpha)} \leq \frac{1}{4\beta(\log n)}$.

It is easy to verify that the query complexity $q_{t'}$ of and distance $\delta_{t'}$ of $C'_{t'}$ are

$$\begin{aligned} q_{t'} &= \left((\log n) q_\alpha(n)\right)^{\Theta(t')}, \\ \delta_{t'} &= \beta(\log n)^{\Theta(t')}. \end{aligned}$$

As for the rate,

$$\rho_{t'} \geq (1 - \beta(\log n))^{\Theta(t')} = 1 - O\left(\beta(\log n) \log\log n\right).$$

Finally, the error of $C'_{t'}$ can be reduced from $1/4$ to $1/5$ with no asymptotic overhead in query complexity, and so $C'_{t'}$ has all the asserted properties. □

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.

[AL96] Noga Alon and Michael Luby. A linear time erasure-resilient code with nearly optimal recovery. *IEEE Transactions on Information Theory*, 42(6):1732–1736, 1996.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM (JACM)*, 45(1):70–122, 1998.

[BF90] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 37–48. Springer, 1990.

[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.

[BFNW93] Laszlo Babai, Lance Fortnow, Noam Nisan, and Avi Wigderson. BPP has subexponential time simulations unless EXPTIME has publishable proofs. *Computational Complexity*, 3(4):307–318, 1993.

[BIR08] Omer Barkol, Yuval Ishai, and Ronny Roth. *Locally decodable codes and their applications*. PhD thesis, Computer Science Department, Technion, 2008.

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 73–83, 1990.

[BR94] Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994*, pages 276–287. IEEE, 1994.

[CGKS95] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 41–50. IEEE, 1995.

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

[Dvi11] Zeev Dvir. On matrix rigidity and locally self-correctable codes. *computational complexity*, 20(2):367–388, 2011.

[Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM Journal on Computing*, 41(6):1694–1703, 2012.

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

[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.

[GLR+91] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *STOC*, volume 91, pages 32–42. Citeseer, 1991.

[Gol11] Oded Goldreich. A sample of samplers: A computational perspective on sampling. In *Studies in Complexity and Cryptography*, pages 302–332. 2011.

[GRS12] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. *Draft available at http://www. cse. buffalo. edu/ atri/courses/coding-theory/book*, 2012.

[GS92] Peter Gemmell and Madhu Sudan. Highly resilient correctors for polynomials. *Information processing letters*, 43(4):169–174, 1992.

[GS06] Oded Goldreich and Madhu Sudan. Locally testable codes and PCPs of almost-linear length. *Journal of the ACM (JACM)*, 53(4):558–655, 2006.

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

[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.

[KS07] Tali Kaufman and Madhu Sudan. Sparse random linear codes are locally decodable and testable. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 590–600. IEEE, 2007.

[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.

[KY09] Kiran S Kedlaya and Sergey Yekhanin. Locally decodable codes from nice subsets of finite fields and prime factors of mersenne numbers. *SIAM Journal on Computing*, 38(5):1952–1969, 2009.

[Lip90] Richard J. Lipton. Efficient checking of computations. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 207–215. Springer, 1990.

[LW19] Ray Li and Mary Wootters. Lifted multiplicity codes and the disjoint repair group property. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[Ree53] Irving S Reed. A class of multiple-error-correcting codes and the decoding scheme. Technical report, Massachusetts inst of tech Lexington Lincoln lab, 1953.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[RVW01] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 18, 2001. https://eccc.weizmann.ac.il/report/2001/018/.

[Sha48] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001. Originally appeared in *Bell System Tech. J.* 27:379–423, 623–656, 1948.

[Sie35] Carl Siegel. Über die classenzahl quadratischer zahlkörper. *Acta Arithmetica*, 1(1):83–86, 1935.

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

[Tre03] Luca Trevisan. List-decoding using the XOR lemma. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 126–135. IEEE, 2003.

[Wal36] Arnold Walfisz. Zur additiven zahlentheorie. ii. *Mathematische Zeitschrift*, 40(1):592–607, 1936.

[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.

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

[ZDb] Kalina Petrova Zeev Dvir. Lecture 4: Lower bounds for r-query LDCs. Lecture notes: https://www.cs.princeton.edu/~zdvir/LDCnotes/LDC4.pdf, year=2016,.

[Zya71] Victor Vasilievich Zyablov. An estimate of the complexity of constructing binary linear cascade codes. *Problemy Peredachi Informatsii*, 7(1):5–13, 1971.

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