

# LCC and LDC: Tailor-made distance amplification and a refined separation

Gil Cohen\*      Tal Yankovitz†

September 13, 2021

## Abstract

The Alon-Edmonds-Luby distance amplification procedure (FOCS 1995) is an algorithm that transforms a code with vanishing distance to a code with constant distance. AEL was invoked by Kopparty, Meir, Ron-Zewi, and Saraf (J. ACM 2017) for obtaining their state-of-the-art LDC, LCC and LTC. Cohen and Yankovitz (CCC 2021) devised a procedure that can amplify inverse-polynomial distances, exponentially extending the regime of distances that can be amplified by AEL. However, the improved procedure only works for LDC and assuming rate  $1 - \frac{1}{\text{poly log } n}$ .

In this work we devise a distance amplification procedure for LCC with inverse-polynomial distances even for vanishing rate  $\frac{1}{\text{poly log log } n}$ . For LDC, we obtain a more modest improvement and require rate  $1 - \frac{1}{\text{poly log log } n}$ . Thus, the tables have turned and it is now LCC that can be better amplified. Our key idea for accomplishing this, deviating from prior work, is to tailor the distance amplification procedure to the code at hand.

Our second result concerns the relation between linear LDC and LCC. We prove the existence of linear LDC that are not LCC, qualitatively extending a separation by Kaufman and Viderman (RANDOM 2010).

---

\*Department of Computer Science, Tel Aviv University. Supported by the ERC starting grant 949499 and the Israel Science Foundation grant 1569/18. Email: [gil@tauex.tau.ac.il](mailto:gil@tauex.tau.ac.il).

†Department of Computer Science, Tel Aviv University. Supported by the Azrieli Faculty Fellowship. Email: [talyankovitz@mail.tau.ac.il](mailto:talyankovitz@mail.tau.ac.il).

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Distance amplification . . . . .	1
1.2	LDC and LCC . . . . .	1
1.3	Improved distance amplification for LDC . . . . .	2
<b>2</b>	<b>Our contribution</b>	<b>2</b>
2.1	Tailor-made distance amplification procedure . . . . .	2
2.2	Refined separation between LDC and LCC . . . . .	5
<b>3</b>	<b>Preliminaries</b>	<b>8</b>
3.1	Notations and conventions . . . . .	8
3.2	Error correcting codes . . . . .	8
3.3	Locally decodable codes and locally correctable codes . . . . .	10
<b>4</b>	<b>Tailor made distance amplification</b>	<b>11</b>
4.1	Characterization of LCC . . . . .	11
4.2	Splitters for query sets . . . . .	14
4.3	The distance amplification procedure . . . . .	15
4.4	Deriving the Corollaries . . . . .	19
<b>5</b>	<b>LDC are not LCC via random weighted tensor codes</b>	<b>29</b>
5.1	Preliminaries for this section . . . . .	29
5.2	A necessary condition for local correction . . . . .	30
5.3	Weighted tensors . . . . .	31
5.4	Local decodability of weighted tensors . . . . .	33
5.5	Local correctability of random weighted tensors . . . . .	35
5.6	Deriving the theorem . . . . .	38
<b>A</b>	<b>From smooth LCC to good LCC</b>	<b>41</b>

# 1 Introduction

## 1.1 Distance amplification

It is a recurrent theme in coding theory that the construction of a code is done in two steps. In the first step, a code with weak parameters is constructed, and typically it is the distance of the code that is unsatisfactory. In the second step, one transforms the code obtained in the first step to a code with the desired parameters, where typically, in the process, the other parameters deteriorate only slightly. When the distance is the unsatisfactory parameter, the second step is referred to as a distance amplification step.

Examples that fall into this framework include the breakthrough constructions of near-optimal small-bias sets by Ta-Shma [TS17], and the state-of-the-art construction of locally decodable codes (LDC), locally correctable codes (LCC), and locally testable codes (LTC) by Kopparty, Meir, Ron-Zewi, and Saraf [KMRS17]. A prominent example from the (adjacent) PCP literature is Dinur’s celebrated proof of the PCP Theorem by gap amplification [Din06]. It is interesting to note that in all the above cases the first step is done using algebraic machinery whereas the second step is based on combinatorial arguments.

## 1.2 LDC and LCC

Informally, a linear  $(q, \delta)$  *locally decodable code (LDC)* is a code, given by an  $\mathbb{F}$ -linear encoding function  $\text{Enc} : \mathbb{F}^k \rightarrow \mathbb{F}^n$ , where  $\mathbb{F}$  is a finite field, that is also equipped with a “local decoder”. The latter is a randomized algorithm, denoted by  $\text{Dec}$ , with the following guarantee. Given an oracle access to  $z \in \mathbb{F}^n$  that is within relative Hamming distance  $\delta$  from some codeword  $\text{Enc}(x)$ , and given  $i \in [k]$ ,  $\text{Dec}^z(i) = x_i$  with high probability. Moreover,  $\text{Dec}$  makes at most  $q$  queries to  $z$ . That is, every message symbol can be decoded, with high probability, by querying only few symbols of a corrupted codeword. A  $(q, \delta)$  *locally correctable code (LCC)* is the variant in which one wishes to decode (or, more precisely, correct) the codeword symbols rather than the message symbols.

Locally decodable codes were defined by Katz and Trevisan [KT00] who proved that asymptotically good LDC require  $q = \Omega(\log n)$  queries. Whether or not this bound is tight is a major open problem. An intensive research effort is devoted to the study and construction of LDC and LCC. Of particular interest is the study of asymptotically good LDC and LCC [KSY14, GKS13, HOW15, KMRS17, GKO<sup>+</sup>18, CY21] where the goal is to minimize the query complexity.

In their seminal work, Kopparty, Meir, Ron-Zewi and Saraf [KMRS17] contained LDC

and LCC with sub-polynomial query complexity. For the first step, a code with vanishing distance  $\delta = \frac{1}{\text{poly}(\log n)}$  was used [KSY14], having the desired query complexity, namely,  $q = 2^{\tilde{O}(\sqrt{\log n})}$ . Then, in the second step the authors invoked a distance amplification procedure due to Alon, Edmonds and Luby [AL96, AEL95], which was originally introduced in the context of linear-time erasure codes, and observed that it converts an LDC (resp. LCC) with distance  $\delta$  and query complexity  $q$  to an LDC (resp. LCC) with constant distance and query complexity  $q_{\text{new}} = q \cdot \text{poly}(\frac{1}{\delta})$ .

### 1.3 Improved distance amplification for LDC

Motivated by the key role that the distance amplification procedure plays in [KMRS17], Cohen and Yankovitz [CY21] asked whether much lower distances can be amplified. Indeed, AEL’s procedure is mostly relevant in the regime  $\delta = \frac{1}{\text{poly}(\log n)}$ . In [CY21], the authors devised an improved procedure that can amplify distances as low as  $\frac{1}{n^\alpha}$  for any constant  $\alpha < 1$  with a fairly low cost in query complexity,  $q_{\text{new}} = q^{O(\log \log n)}$ <sup>1</sup> (and even for  $\alpha = 1 - o(1)$  at a small additional cost in query complexity). However, their improved distance amplification procedure has two drawbacks:

1. Unlike the AEL distance amplification procedure, the improved procedure was only shown to work for LDC (and it may or may not work for LCC).
2. Second, to amplify the distance, the original LDC must have rate close to one, more precisely, rate  $1 - \frac{1}{\text{poly}(\log n)}$ .

## 2 Our contribution

We turn to present the two results of this work.

### 2.1 Tailor-made distance amplification procedure

Our first contribution is a distance amplification procedure for LCC that can amplify distances as low as those handled by [CY21] (for LDC). Moreover, our procedure works even for vanishing rate LCC.

---

<sup>1</sup> $\text{poly}(\log \log n)$  factors in the exponent of the query complexity can be safely ignored given that, at present, the lowest known query complexity is  $2^{\tilde{\Theta}(\sqrt{\log n})}$ . Such an overlook will matter only when (and if) the query complexity will go below quasi-poly-logarithmic.

**Theorem 2.1** (Distance amplification for LCC; informal). *Let  $h \geq 1 \geq \alpha > 0$  be any constants. There exists a transformation that takes a  $q$ -query LCC with distance  $\frac{1}{n^\alpha}$  and rate  $\frac{1}{(\log \log n)^h}$  to an asymptotically good LCC with query complexity*

$$q_{\text{new}} = q^{O((\log \log n)^{2h+2})}.$$

We chose to state our result in a somewhat informal manner. For the formal statement, see Corollary 4.19. We turn to give further details on the result.

**Explicitness.** In the statement of Theorem 2.1 we ignore the issue of explicitness. Indeed, understanding LDC and LCC is already interesting in the information-theoretic level. Having said that, our transformation is fairly explicit: It is a zero error randomized transformation that runs in polynomial-time. More precisely, for every “failure” parameter  $\varepsilon > 0$ , our transformation runs in time  $\text{poly}(n) \cdot \log \frac{1}{\varepsilon}$  and produces an LCC with probability at least  $1 - \varepsilon$ ; otherwise, it declares failure. We find this aspect to be a minor issue as, recall, LCC are anyhow randomized in nature. Nonetheless, it will be interesting to obtain a deterministic transformation with matching parameters.

**Codes vs. family of codes.** A second issue that we chose to sweep under the rag in the statement of Theorem 2.1 is that the transformation operates on the level of family of codes rather than on the level of individual codes. That is, in order to produce an asymptotically good LCC of a given block-length  $n$ , our transformation requires as input a sufficiently dense family of codes. By that we mean that the consecutive block-lengths in the family are not too far apart. The density of the resulted family of codes is the same as that of the original family.

**Amplifying lower distances.** Like [CY21], we can even amplify sub-polynomial distances, in particular, distances of the form  $1/n^{1-1/g(n)}$  for an increasing function  $g$ , and assuming a certain technical relation between  $g$  and the rate. In particular, for every constant  $m \geq 1$  we can handle  $g(n) = (\log \log n)^m$ , and end up with query complexity

$$q_{\text{new}} = q^{O((\log \log n)^{2h+2m+2})}.$$

We note that constructing a code for  $g(n) = \log n$  is trivial.

**Amplifying the distance of LDC.** We also obtain an improvement for LDC by devising a distance amplification procedure that requires rate  $1 - \frac{1}{\text{poly}(\log \log n)}$ , modestly improving upon the  $1 - \frac{1}{\text{poly}(\log n)}$  rate required by [CY21]. The reason that we can do much better

for LCC is due to the rate amplification procedure of [CY21] that, informally, can amplify rate  $\rho$  LCC with  $q$  queries to constant rate LCC with query complexity  $q_{\text{new}} = q^{\text{poly}(\frac{1}{\rho})}$ . Such a transformation is not known for LDC.

### 2.1.1 Proof idea

In this section we give a short and informal account on our proof technique, and start by contrasting our technique with prior work. Both the AEL distance amplification procedure, as was used in [KMRS17], and the one given by [CY21] are based on samplers and further involve a “small” code, that is, a code with logarithmic block-length. The latter improves upon the former by using unbalanced samplers (rather than balanced ones, or expander graphs as was used originally [AEL95, AL96]) and using a recursive construction. To obtain our result, we deviate from prior work and tailor the distance amplification procedure to the LCC at hand. That is, our procedure is “white box” - it produces a new code with improved distance by first examining the structure of the given code. To tailor the procedure to the LCC at hand, we do not work directly with the definition of LCC as it lacks sufficient structure to work with. Instead, we work with a more combinatorial characterization of LCC as was used in [CY21]. We turn to elaborate on this.

Let  $C \subseteq \mathbb{F}^n$  be a linear  $(q, \delta)$ -LCC. One can prove the following structural result. With every coordinate  $i \in [n]$  one can associate a set, called a *query set*,  $A_i = \{Q_1^i, \dots, Q_m^i\}$  of  $m = \delta n/q$  disjoint subsets of  $[n]$ , each of size at most  $q$ , such that the following holds: For every  $c \in C$  and  $t \in [m]$ ,  $c_i$  can be deduced from  $c_{Q_t^i}$ . Assume from here on, for simplicity, that  $\delta = 1/\sqrt{n}$  and so  $m = \sqrt{n}/q$ . Denote  $\bar{A}_i = \bigcup_{t=1}^m Q_t^i$  and note that  $|\bar{A}_i| \leq \sqrt{n}$ .

For our distance amplification procedure, we make use of a special partition  $\pi$  of  $[n]$  into  $\sqrt{n}$  parts  $P_1, \dots, P_{\sqrt{n}}$ , each of size  $\sqrt{n}$ . We say that such a partition is a *d-splitter* for  $C$  (more precisely, a *d-splitter* for the query sets  $A_1, \dots, A_n$  obtained from  $C$ ) if for every  $s \in [\sqrt{n}]$  and  $i \in [n]$ ,  $|P_s \cap \bar{A}_i| \leq d$ . We wish to minimize  $d$  and thus consider a max load balls into bins like problem: For every  $i \in [n]$  we place a ball with color  $i$  at each of the coordinates in  $\bar{A}_i$ . Note that a coordinate  $j \in [n]$  may contain many balls of different colors. Indeed, the average number of balls at coordinate  $j \in [n]$  is  $\sqrt{n}$ . Our goal is to choose the partition  $\pi$  in such a way that every part  $P_t$  will contain at most  $d$  balls of the same color. It is easy to show that a *d-splitter* for  $C$  exists with  $d = O(\frac{\log n}{\log \log n})$ .

We construct a new code  $C' \subseteq \mathbb{F}^n$  as follows. We take  $C'$  to be the code  $C' \subseteq C$  with the property that for every part  $P_s$  of  $\pi$ , when  $C'$  is projected to the coordinate set  $P_s$ , the obtained vectors consist of codewords of a code  $C_{\sqrt{n}}$  having block length  $\sqrt{n}$ , which is a  $q'$ -query LCC. That is to say, we require that for every  $c \in C'$  and  $s \in [\sqrt{n}]$ ,  $c_{P_s} \in C_{\sqrt{n}}$ .

Observe that  $C'$  can be constructed by adjoining to the parity checks of  $C$ , the parity checks of  $C_{\sqrt{n}}$  when restricted to each block in  $\pi$ .

We show that if  $C_{\sqrt{n}}$  is a smooth LCC, which means that it queries each coordinate with roughly the same probability, then so is  $C'$ . Moreover,  $C'$  has query complexity  $qq'$ . Thus,  $C$  can be transformed into a smooth LCC of length  $n$  given that a smooth LCC of length  $\sqrt{n}$  is at hand. This calls for a recursive construction which results with a smooth LCC with query complexity  $q^{O(\log \log n)}$ . After obtaining a smooth code, the final step is to invoke the AEL distance amplification to end up with a good LCC. This final step has a minor effect on the query complexity.

The above recursive construction must start with LCC of rate  $1 - \frac{1}{\text{poly}(\log \log n)}$ . This is due to the rate deterioration throughout the  $\log \log n$  recursive calls. For amplifying rate  $\frac{1}{\text{poly}(\log \log n)}$  LCC, as stated in Theorem 2.1, we invoke the rate amplification procedure of [CY21] before running the recursive construction described above. This has some effect on the density of the LCC family that the recursion has access to which requires some care.

## 2.2 Refined separation between LDC and LCC

Understanding the relation between LDC and LCC is fundamental. Currently the only regime in which the state of affairs is better understood is the 2-query regime [BDYW11, BDSS11, BGT16]. In the constant-query regime for  $q \geq 3$ ,  $q$ -LDC with sub-exponential length are known [Yek08, Efr12, DGY11] whereas it is not known if this can be matched for  $q$ -LCC. Recall that in the constant-rate regime, the state of the art result of [KMRS17] achieves sub-polynomial query complexity and holds for LDC and LCC alike.

In the general case, clearly, a systematic LCC is an LDC. As every linear code can be made systematic (by applying Gaussian elimination to its generating matrix), a linear LCC induces a linear LDC with the same parameters. Thus, informally, LCC are stronger than LDC, at least for linear codes.

**Are LDC and LCC “equivalent”?** As for the converse, Kaufman and Videman [KV10] observed that an LDC is not necessarily an LCC. Their proof starts with an LDC. If it is not an LCC to begin with, we are done. If it is an LCC, the proof goes on by transforming it to a new code by appending to it one additional entry that does not involve low-weight constraints (namely, every vector in the dual code that does not vanish on the new entry is of large weight). In this way, one obtains an LDC with an entry that cannot be corrected with few queries. Such an entry can be shown to exist by a counting argument. This argument can be extended to produce many new bits that cannot be corrected.

While, formally, the argument above establishes the existence of LDC that are not LCC, it has a drawback which makes it somewhat less appealing. In the resulted code, the adjoined bits that cannot be corrected are not needed for decoding the original bits. This means that if one is given a code that is not an LCC because of the above transformation, with the task of taking such a code and “convert” it to an LCC, this could be done easily: simply by removing these coordinates, and this clearly would not harm the code’s dimension. This raises the question: Can any linear LDC be so “easily” converted to an LCC of similar dimension and query complexity?

The thought that the answer to this question may turn out to be in the affirmative is not far fetched in the case of linear codes. Indeed, we know that the locality features of linear codes “come from” linear relations between different bits of the codeword and of the message. For example, if a linear code  $\text{Enc} : \mathbb{F}^k \rightarrow \mathbb{F}^n$  is a  $q$ -query LDC, and in particular the  $i$ -th bit of each message  $m$  can be deduced from a subset  $Q \subseteq [n]$  that consists of at most  $q$  coordinates of  $c = \text{Enc}(m)$ , then there exists a linear map  $f_{i,Q}$  which satisfies  $m_i = f_{i,Q}(c_Q)$  for any  $m$ . Likewise, if  $m_i$  can as well be deduced from another subset  $Q' \subseteq [n]$ ,  $|Q'| \leq q$  (as is expected due to the distance guarantee), then there is a linear map  $f_{i,Q'}$  satisfying  $m_i = f_{i,Q'}(c_{Q'})$  for every  $m$ . It follows that in such a case, for every codeword  $c$ ,  $f_{i,Q}(c_Q) = f_{i,Q'}(c_{Q'})$ . Since  $f_{i,Q}$  and  $f_{i,Q'}$  are linear maps (that, we may assume, depend on all their parameters) this means that for every  $j \in Q \Delta Q'$ , there exists a linear map  $g_j$  satisfying  $c_j = g_j(c_{Q_j})$  for every codeword  $c$ , where  $Q_j = (Q \cup Q') \setminus \{j\}$ .

Therefore, by the mere fact that  $j \in [n]$  is sometimes used in the local decoding process of  $i \in [k]$ , it is implied that it is possible to “correct” the  $j$ -th coordinate by reading only a few locations of the codeword (at most  $2q - 1$ ). Thus, the question of whether local decoding implies local correction is in place, in the case of linear codes, and especially so in the setting where  $k$  is close to  $n$ .

In light of this, the fact that in the separating result of [KV10] between linear LDC and LCC, the coordinates which are shown to be uncorrectable are not used by the local decoding process, calls for the question of whether there exists a linear LDC with uncorrectable coordinates that are crucial for the decoding process.

**Our result.** The second contribution of this work is a proof for the existence of an LDC that is not an LCC in the following stronger sense: It contains entries that cannot be corrected which are crucial for the local decoder. This raises the question of what we mean by coordinates that are “crucial”. The mere fact that it is possible for a set of coordinates to be queried by the local decoding process should not qualify them as such, as what allows for a code to be locally decodable or locally correctable is that there are many options to decode or correct each symbol. Thus, a more suitable interpretation



for a “crucial” set of coordinates  $J \subseteq [n]$  is the following: If every coordinate  $j \in J$  is “zeroed out” from the code (i.e., for every codeword  $c$  we override  $c_j$  with zero) then the transformed code is no longer locally decodable. With this we are ready to present our separation.

**Theorem 2.2** (Separation of LDC and LCC; Informal). *Let  $C : \mathbb{F}^k \rightarrow \mathbb{F}^n$  for  $|\mathbb{F}| > 2$  and  $k = \Theta(n)$  be a linear  $q$ -query LDC. Then, there exists a linear  $q^2$ -query LDC  $\hat{C} : \mathbb{F}^{k^2} \rightarrow \mathbb{F}^{n^2}$  with the following property. There exists a subset of coordinates  $J \subseteq [n^2]$  in which every coordinate cannot be locally corrected with query complexity  $\sqrt{n}$  and correction radius  $1/\sqrt{n}$ . Moreover, if every coordinate  $j \in J$  is zeroed out from the code, then the relative distance of the obtained code is  $\tilde{O}(1/\sqrt{n})$  (and so it is certainly not an LDC).*

For the formal, more general, statement, see Theorem 5.19. Note that our result does not cover the binary field and it is an interesting question whether it can be extended to include that case.

**Proof idea.** The underlying idea of the proof of Theorem 2.2 is an operation on two codes to which we call *weighted tensoring*. The weighted tensoring of codes is similar to the standard tensoring of codes. In the case of standard tensoring, the encoding of the tensor of two codes is done by taking a matrix as input and applying the first code to each column and then applying the second code to each row in the resulted matrix. In the encoding of a weighted tensor, before the second step, each entry of the matrix is multiplied by a non-zero field element, or weight.

We consider the case of weighted tensoring which is done with *random* weights. We show that while the code resulted from this is an LDC (assuming that the two input codes were so), with high probability there is a set of coordinates in the code that cannot be locally corrected, while being crucial for the decoding. The analysis showing that the set of coordinates cannot be locally corrected is done by considering the affect of the weights on the dual code. A probabilistic argument is then used to show that the argued codes exist.

**Discussion.** We end this section with a short discussion to clarify a potentially confusing point. While LCC are, in a sense, more powerful than LDC (indeed, our second contribution, Theorem 2.2, attempts to formalize that better), our first result, given by Theorem 2.1, transforms a vanishing rate LCC with polynomially-small distance to an asymptotically good LCC—a result that is not known for LDC. So, how can it be that we can do this for LCC and not for the weaker LDC?

Of course, this should cause no confusion as the latter is a *transformation* that works for LCC and not LDC, not a *construction* nor it is even a proof of existence. Put differently, although the transformation generates the stronger object, the transformation is also given it as its input.

## 3 Preliminaries

### 3.1 Notations and conventions

Unless stated otherwise, all logarithms are taken to the base 2. For  $n \in \mathbb{N}$ , we use  $[n]$  to denote the set  $\{1, \dots, n\}$ . For ease of readability, we sometimes avoid the use of floor and ceiling. This does not affect the stated results. We use  $\mathbb{F}$  to denote a field, and any referenced field is assumed to be finite and of a constant size. When  $n$  and  $\mathbb{F}$  are clear from context, we use  $e_i \in \mathbb{F}^n$  to denote the  $i$ -th vector of the standard basis. For  $q \in \mathbb{N}$ , we use  $H_q$  to denote the  $q$ -ary entropy function, and  $H$  to denote the binary entropy function. For a vector  $v \in \mathbb{F}^n$ , we denote by  $|v|$  the *hamming weight* of  $v$ , which is the number of its non-zero coordinates  $|v| = |\{j \in [n] \mid v_j \neq 0\}|$ , and the *support* of  $v$  is  $\text{supp}(v) = \{j \in [n] \mid v_j \neq 0\}$ . For two vectors  $u, v \in \mathbb{F}^n$ , we denote their (absolute) hamming distance by  $\text{dist}(u, v)$ . For a linear subspace  $L \subseteq \mathbb{F}^n$ , we denote by  $L^{\leq q}$  the set of vectors of weight at most  $q$ . For two vector  $u, v \in \mathbb{F}^n$ , we use  $\langle u, v \rangle$  to denote the inner product of  $u$  and  $v$ ,  $\sum_{i=1}^n u_i v_i \in \mathbb{F}$ . For a vector  $v \in \mathbb{F}^n$  and a sequence  $I = (i_1, \dots, i_m) \in [n]^m$ , we denote by  $v_I$  the vector  $(v_{i_1}, \dots, v_{i_m}) \in \mathbb{F}^m$ . For a linear subspace  $L \subseteq \mathbb{F}^n$  and a sequence  $I = (i_1, \dots, i_m) \in [n]^m$ , we denote by  $L_I$  the subspace  $\{v_I \mid v \in L\}$ . Note that  $L_I$  is indeed a subspace as it is given by a suitable projection.

A *partition*  $\pi$  of size  $k$  of  $[n]$  is a set  $\{P_1, \dots, P_k\}$  of disjoint subsets of  $[n]$ , such that  $P_1 \cup \dots \cup P_k = [n]$ . A partition  $\{P_1, \dots, P_k\}$  is *ordered* if each  $P_i$  is a sequence rather than a set (and the sequences, when viewed as sets, satisfy the same requirements). Throughout this paper, any partition of  $[n]$  will be an ordered partition (though we may not state it explicitly) with the sequences defined by the natural increasing order of  $\mathbb{N}$ .

### 3.2 Error correcting codes

We start by recalling the definition of an error correcting code, and of a family of error correcting codes. In this work we only consider linear codes.

**Definition 3.1.** *For  $n \in \mathbb{N}$  and  $\mathbb{F}$  a field, a code of length  $n$  over  $\mathbb{F}$  is a linear subspace*

$C \subseteq \mathbb{F}^n$ .<sup>2</sup> The dimension of the code, denoted by  $k$ , is the dimension of  $C$  over  $\mathbb{F}$ ,  $\dim_{\mathbb{F}} C$ . The (non-local) distance of the code, denoted by  $d$ , is  $\min_{c \in C, c \neq 0} |c|$ . The rate of the code, denoted by  $\rho$ , is  $k/n$ . The (non-local) relative distance of the code, denoted by  $\Delta$ , is  $d/n$ . The elements of  $C$  are called codewords.

We will also need to consider encodings of codes.

**Definition 3.2.** We call a function  $\text{Enc} : \mathbb{F}^k \rightarrow \mathbb{F}^n$  an encoding of a code  $C$  if it is an injective linear map and  $C = \text{Im}(\text{Enc})$ .

**Definition 3.3.** For a field  $\mathbb{F}$ , a code family over  $\mathbb{F}$  is a set of codes  $C = \{C^n\}$ , which contains at most one code  $C^n$  of length  $n$  over  $\mathbb{F}$ , for every possible length  $n \in \mathbb{N}$ . For every  $n \in \mathbb{N}$ , we denote by  $\lceil n \rceil^C$  the minimal length of a code in the family  $C$  of length at least  $n$ , and by  $\lfloor n \rfloor^C$  the maximal length of a code in the family of length at most  $n$ . For constants  $n_0 \in \mathbb{N}$ ,  $c \geq 1$  and  $d \leq 1$ , we say that the family is  $(n_0, c, d)$ -dense if for every  $n \geq n_0$ ,  $\lceil n \rceil^C \leq cn$  and  $\lfloor n \rfloor^C \geq dn$ .

**Definition 3.4.** For a field  $\mathbb{F}$ , a code-encoding family over  $\mathbb{F}$  is a set of pairs of codes and corresponding encodings  $C = \{(C^k, \text{Enc}^k)\}$ , which contains at most one code  $C^k$  of dimension  $k$  over  $\mathbb{F}$ , for every possible dimension  $k \in \mathbb{N}$ . For every  $k \in \mathbb{N}$ , we denote by  $\lceil k \rceil^C$  the minimal dimension of a code in the family  $C$  of dimension at least  $k$ , and by  $\lfloor k \rfloor^C$  the maximal dimension of a code in the family of dimension at most  $k$ . For constants  $k_0 \in \mathbb{N}$ ,  $c \geq 1$  and  $d \leq 1$ , we say that the family is  $(k_0, c, d)$ -dense if for every  $k \geq k_0$ ,  $\lceil k \rceil^C \leq ck$  and  $\lfloor k \rfloor^C \geq dk$ .

**Definition 3.5.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$ . The dual code of  $C$  is defined to be its orthogonal subspace  $C^\perp$ .

**Definition 3.6.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$ , let  $i \in [n]$  and  $B \subseteq [n]$ . We say that  $B$  determines  $i$  in  $C$  if there exists a function  $f : \mathbb{F}^{|B|} \rightarrow \mathbb{F}$  such that for every  $c \in C$ ,  $c_i = f(c_B)$ .

We also need the following property of linear codes.

**Fact 3.7.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$ . Further let  $i \in [n]$ ,  $Q \subseteq [n]$  and  $x \in \mathbb{F}^{|Q|}$ . Then, one of the following cases must hold.

1. There is at most one  $\alpha \in \mathbb{F}$  for which there exists some  $c \in C$  satisfying  $c_Q = x$  and  $c_i = \alpha$ .

---

<sup>2</sup>We may omit the phrase “over  $\mathbb{F}$ ” if the underlying field is clear from context.

2. For every  $\alpha \in \mathbb{F}$  there is an equal number of  $c \in C$  for which  $c_i = \alpha$ .

In particular, either no function (even randomized) of  $c_Q$  can predict  $c_i$  with probability larger than  $1/|\mathbb{F}|$ , when  $c \in C$  is randomly chosen uniformly, or  $c_Q$  determines  $c_i$  for all  $c \in C$ .

### 3.3 Locally decodable codes and locally correctable codes

**Definition 3.8.** For  $C \subseteq \mathbb{F}^n$ , we say that a procedure  $f : A \rightarrow B$  is with oracle access to  $c \in C$  if when  $f$  is run, it gets besides an input  $a \in A$ , access to  $c \in C$ :  $f$  can query  $c_i$  for indices  $i \in [n]$ . To describe a specific run of  $f$  with input  $a \in A$  and oracle access to  $c \in C$ , we either say that  $f(a)$  is run with oracle access to  $c$ , or write  $f^c(a)$  for short. We say that  $f$  is non-adaptive if the queries it makes are independent of  $c \in C$ .

**Definition 3.9.** For a code  $C$  of length  $n$  and dimension  $k$  over  $\mathbb{F}$ , and  $\text{Enc}$  an encoding of it,  $(C, \text{Enc})$  is called a  $(q, \delta, \varepsilon)$ -LDC (locally decodable code, abbreviated) if there exists a randomized procedure  $\text{Dec} : [k] \rightarrow \mathbb{F}$  that is given an oracle access to  $z \in \mathbb{F}^n$ , and has the following guarantee. For every  $i \in [k]$ ,  $x \in \mathbb{F}^k$  and  $z \in \mathbb{F}^n$  satisfying  $\text{dist}(z, \text{Enc}(x)) \leq \delta n$ ,  $\text{Dec}^z(i) = x_i$  with probability at least  $1 - \varepsilon$ . Furthermore,  $\text{Dec}^z(i)$  always makes at most  $q$  queries to  $z$ . We further require that  $\text{Dec}$  is non-adaptive. We call  $\text{Dec}$  a local decoder (or decoder) for  $(C, \text{Enc})$ , and the parameter  $q$  is called the query complexity of  $(C, \text{Enc})$ .

**Definition 3.10.** A code-encoding family  $C = \{(C^k, \text{Enc}^k)\}$  of codes over  $\mathbb{F}$  is called a family of good  $q(k)$ -LDC, or a family of good LDC with query complexity  $q(k)$ , if every code  $C^k$  in the family is a code with rate at least  $\rho(k)$ , which is a  $(q(k), \delta(k), \varepsilon(k))$ -LDC, for  $\rho(k) = \Omega(1)$ ,  $\delta(k) = \Omega(1)$ , and  $\varepsilon(k) \leq 1/3$ .

We have the following easy fact.

**Fact 3.11.** If  $C$  is a code of length  $n$  and dimension  $k > 0$  over  $\mathbb{F}$  and  $\text{Enc}$  is an encoding of it, and if  $(C, \text{Enc})$  is a  $(q, \delta, \varepsilon)$ -LDC, then, provided that  $\varepsilon < 1 - 1/|\mathbb{F}|$ , the (non-local) relative distance of  $C$ ,  $\Delta$ , satisfies  $\Delta > \delta$ .<sup>3</sup>

*Proof.* Assume towards a contradiction that  $\Delta \leq \delta$ . Then, there exists some  $x \in \mathbb{F}^k$  and  $i \in [k]$  such that  $|\text{Enc}(x)| \leq \delta n$  and  $x_i \neq 0$ , and we may assume without loss of generality that  $x_i = 1$ . For every  $\gamma \in \mathbb{F}$  we define  $x_\gamma = \gamma x$ . Note that  $(x_\gamma)_i = \gamma$ . Consider the following scenario: We randomly and uniformly sample  $X \in \{x_\gamma \mid \gamma \in \mathbb{F}\}$ . Since for every  $\gamma \in \mathbb{F}$ , the distance of  $\text{Enc}(x_\gamma)$  from the zero codeword satisfies  $\text{dist}(\text{Enc}(x_\gamma), 0) =$

<sup>3</sup>Note that in the case that  $\varepsilon < 1/2$  a stronger bound  $\Delta > 2\delta$  holds.

$|\text{Enc}(x_\gamma)| = |\gamma \cdot \text{Enc}(x)| \leq \delta n$ , it is always the case that  $\text{dist}(\text{Enc}(X), 0) \leq \delta n$ . Therefore, the probability that  $\text{Dec}^0(i) = X_i$  (over the choice of  $X$  and the randomness of  $\text{Dec}$ ) is at least  $1 - \varepsilon > 1/|\mathbb{F}|$ , where  $\text{Dec}$  is a local decoder promised by the fact that  $(C, \text{Enc})$  is a  $(q, \delta, \varepsilon)$ -LDC. This is clearly a contradiction, since  $\text{Dec}^0(i)$  is independent of  $X$  and  $X_i$  is uniformly distributed over  $\mathbb{F}$ .  $\square$

**Definition 3.12.** A code  $C$  of length  $n$  over  $\mathbb{F}$  is called a  $(q, \delta, \varepsilon)$ -LCC (locally correctable code, abbreviated) if there exists a randomized procedure  $\text{Cor} : [n] \rightarrow \mathbb{F}$  that is given an oracle access to  $z \in \mathbb{F}^n$ , and has the following guarantee. For every  $i \in [n]$ ,  $y \in C$  and  $z \in \mathbb{F}^n$  satisfying  $\text{dist}(z, y) \leq \delta n$ ,  $\text{Cor}^z(i) = y_i$  with probability at least  $1 - \varepsilon$ . Furthermore,  $\text{Cor}^z(i)$  always makes at most  $q$  queries to  $z$ . We further require that  $\text{Cor}$  is non-adaptive and that  $\text{Cor}(i)$  never queries  $i$ <sup>4</sup>. We call  $\text{Cor}$  a local corrector (or corrector) for  $C$ , and the parameter  $q$  is called the query complexity of  $C$ .

**Definition 3.13.** For a code  $C$  of length  $n$  over  $\mathbb{F}$  (not necessarily a  $(q, \delta, \varepsilon)$ -LCC), and  $i \in [n]$ , we say that  $i$  is a  $(\delta, q, \varepsilon)$ -correctable coordinate in  $C$  if there exists a procedure  $\text{Cor} : [n] \rightarrow \mathbb{F}$  such that  $\text{Cor}(i)$  satisfies the requirements in Definition 3.12.

**Definition 3.14.** A family  $C = \{C^n\}$  of codes over  $\mathbb{F}$  is called a family of good  $q(n)$ -LCC, or a family of good LCC with query complexity  $q(n)$ , if every code  $C^n$  in the family is a code with rate at least  $\rho(n)$ , which is a  $(q(n), \delta(n), \varepsilon(n))$ -LCC, for  $\rho(n) = \Omega(1)$ ,  $\delta(n) = \Omega(1)$ , and  $\varepsilon(n) \leq 1/3$ .

The following well-known fact is an implication of the fact that every linear code has a systematic encoding<sup>5</sup>.

**Fact 3.15.** If a code  $C$  is a  $(q, \delta, \varepsilon)$ -LCC, then there exists an encoding  $\text{Enc}$  such that  $(C, \text{Enc})$  is a  $(q, \delta, \varepsilon)$ -LDC.

## 4 Tailor made distance amplification

### 4.1 Characterization of LCC

In this section, we will need to use two characterizations of LCC, as was given by Definition 3.12. The first, given next in Definition 4.1, is of a  $(q, \tau)$ -LCC, and resembles the

---

<sup>4</sup>The assumption that  $\text{Cor}(i)$  never queries  $i$  is only for simplicity. Any LCC which defies this assumption can be easily converted to one which does not, with a negligible effect on  $\delta$ .

<sup>5</sup>An encoding  $\text{Enc}$  is a *systematic* encoding if for some  $f : [k] \rightarrow [n]$ , for all  $x \in \mathbb{F}^k$  and  $i \in [k]$ ,  $\text{Enc}(x)_{f(i)} = x_i$ .

definition of smooth codes given by [KT00] for LDC. A  $(q, \tau)$ -LCC differs from a  $(q, \delta, \varepsilon)$ -LCC in that its local correction is only required to succeed if it is given a codeword of the code, rather than a possible corrupted codeword. Accordingly, the correction of a  $(q, \tau)$ -LCC has no “distance” guarantee, but instead it is required not to query any coordinate with too high probability, i.e., probability larger than  $\tau$ . When we will construct an LCC, it will be easier to first argue that it is a  $(q, \tau)$ -LCC and use that to show it can be made into a  $(q, \delta, \varepsilon)$ -LCC for any  $\varepsilon$  and  $\delta = \varepsilon/(\tau n)$ .

The second characterization, which will be given in Definition 4.5, is of what we call a  $(q, \tau)$ -query-set LCC. Informally, a code is  $(q, \tau)$ -query-set LCC if for every coordinate we have a large enough set of disjoint subsets of  $[n]$ , from which it can be decoded. The distance amplification procedure that we define utilizes these query sets and so the properties of the input code that we will use are that of its characterization as a  $(q, \tau)$ -query-set LCC. This is, in a sense, a more “combinatorial” characterization of LCC which can be more conveniently used when a manipulation of these objects is needed.

The three characterizations of LCC all imply each other, but some of the transitions are at some cost to the parameters. Indeed, Claim 4.2 will show that a  $(q, \tau)$ -LCC is a  $(q, \delta, \varepsilon)$ -LCC for  $\delta = \varepsilon/(\tau n)$ , Claim 4.6 will show that a  $(q, \tau)$ -query-set is a  $(q, \tau)$ -LCC, and Claim 4.7 will complete the cycle and show that a  $(q, \delta, \varepsilon)$ -LCC is a  $(q, \tau)$ -query-set LCC for  $\tau = q/(\delta n)$ .

**Definition 4.1.** *A code  $C$  of length  $n$  over  $\mathbb{F}$  is called a  $(q, \tau)$ -LCC if there exists a randomized procedure  $\text{Cor} : [n] \rightarrow \mathbb{F}$  that is given an oracle access to  $c \in C$ , and has the following guarantee. For every  $i \in [n]$  and  $c \in C$ ,  $\text{Cor}^c(i) = c_i$ , with probability 1. Furthermore,  $\text{Cor}^c(i)$  always makes at most  $q$  queries to  $c$ , and for every  $j \in [n]$ , the probability that  $c_j$  is queried by  $\text{Cor}^c(i)$  is at most  $\tau$ . We further require that  $\text{Cor}$  is non-adaptive and that  $\text{Cor}(i)$  never queries  $i$ . We call the parameter  $q$  the query complexity and the parameter  $\tau$  the smoothness of the LCC.*

**Claim 4.2.** *Let  $C$  be a code of length  $n$  which is a  $(q, \tau)$ -LCC. Then, for any  $\varepsilon > 0$ ,  $C$  is a  $(q, \delta, \varepsilon)$ -LCC with  $\delta = \varepsilon/(\tau n)$ .*

*Proof.* Let  $\varepsilon > 0$  and let  $\text{Cor}$  be a corrector of  $C$ . Let  $c \in C$  and  $z \in \mathbb{F}^n$  such that  $\text{dist}(c, z) \leq \delta n = \varepsilon/\tau$ , and set  $B = \{j \in [n] \mid z_j \neq c_j\}$ . Fix  $i \in [n]$ . By the union bound over  $j \in B$ , except with probability  $\varepsilon$ , when  $\text{Cor}(i)$  is run with oracle access to  $c \in C$ , it does not make a query to an index in  $B$ . If this is the case, then if  $\text{Cor}$  was given access to  $z$  instead of  $c$ , it would successfully output  $c_i$ , as well. Thus,  $C$  is indeed a  $(q, \delta, \varepsilon)$ -LCC as the same corrector  $\text{Cor}$  can be used with oracle access to strings  $z \in \mathbb{F}^n$ , and given that  $\text{dist}(c, z) \leq \delta n$ ,  $\text{Cor}(i)$  is promised to output  $c_i$  with probability at least  $1 - \varepsilon$ .  $\square$

**Definition 4.3.** A set  $\mathcal{A} = \{A_1, \dots, A_n\}$  is called an  $n$ -query-set if for every  $i \in [n]$ ,  $A_i$  is a set of disjoint subsets of  $[n] \setminus \{i\}$ . For every  $i \in [n]$  we define  $\overline{A}_i = \bigcup_{B \in A_i} B$ .

**Definition 4.4.** Let  $C$  be a code of length  $n$  and let  $\mathcal{A} = \{A_1, \dots, A_n\}$  be an  $n$ -query-set.  $\mathcal{A}$  is said to be a query-set for  $C$  if for every  $i \in [n]$  and  $B \in A_i$ ,  $B$  determines  $i$  in  $C$  (see Definition 3.6).

**Definition 4.5.** Let  $C$  be a code of length  $n$ .  $C$  is said to be a  $(q, \tau)$ -query-set-LCC if there exists a set  $\mathcal{A} = \{A_1, \dots, A_n\}$  which is a query-set for  $C$ , such that for every  $i \in [n]$ ,  $|A_i| \geq 1/\tau$  and for every  $B \in A_i$ ,  $|B| \leq q$ .

**Claim 4.6.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$  which is a  $(q, \tau)$ -query-set LCC. Then  $C$  is a  $(q, \tau)$ -LCC.

*Proof.* Let  $\mathcal{A} = \{A_1, \dots, A_n\}$  be a query set that corresponds to  $C$  being a  $(q, \tau)$ -query-set LCC. The following corrector  $\text{Cor}$  shows that  $C$  is a  $(q, \tau)$ -LCC. Given  $i \in [n]$ , and oracle access to  $c \in C$ ,  $\text{Cor}(i)$  samples uniformly at random some  $B \in A_i$  and queries  $c_B$ . As  $B$  determines  $i$  in  $C$ , there exists a function  $f$  satisfying  $f(c_B) = c_i$  for every  $c \in C$ , and so  $\text{Cor}(i)$  uses such a function and outputs its result. Thus, for every  $c \in C$ , the output of  $\text{Cor}(i)$  is always equal to  $c_i$ , and note that as any sampled  $B \in A_i$  satisfies  $|B| \leq q$ ,  $\text{Cor}(i)$  always makes at most  $q$  queries. Since  $A_i$  is of size at least  $1/\tau$  and is composed of disjoint subsets of  $[n] \setminus \{i\}$ , any coordinate is queried by  $\text{Cor}(i)$  with probability at most  $\tau$ , and  $\text{Cor}(i)$  never queries  $i$ . Thus,  $C$  is a  $(q, \tau)$ -LCC.  $\square$

**Claim 4.7.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$  which is a  $(q, \delta, \varepsilon)$ -LCC, for  $\varepsilon < 1 - 1/|\mathbb{F}|$ . Then,  $C$  is a  $(q, \tau)$ -query-set-LCC for  $\tau = q/(\delta n)$ .

The proof for the claim is similar to the proof in [KT00] to their Theorem 1 and to the proof in [ZD] for Theorem 1.1.

*Proof for Claim 4.7.* To prove the claim, we need to show that there exists a set  $\mathcal{A} = \{A_1, \dots, A_n\}$  which is a query-set for  $C$ , such that for every  $i \in [n]$ ,  $|A_i| \geq 1/\tau = \delta n/q$  and for every  $B \in A_i$ ,  $|B| \leq q$ . We construct  $\mathcal{A}$  with the required properties by constructing each of the subsets separately. Let  $\text{Cor}$  denote a corrector promised by the fact that  $C$  is a  $(q, \delta, \varepsilon)$ -LCC, and let  $i \in [n]$ . To construct  $A_i$ , we construct a sequence of disjoint sets  $B_1^i, \dots, B_{m_i}^i \subseteq [n] \setminus \{i\}$ , in an iterative manner. We will eventually set  $A_i = \{B_1^i, \dots, B_{m_i}^i\}$ . It will hold that for every  $j$ ,  $B_j^i$  determines  $i$  in  $C$ , while satisfying  $|B_j^i| \leq q$ , and that  $m_i \geq \delta n/q$ , which will conclude the proof.

The construction of  $B_1^i, \dots, B_{m_i}^i \subseteq [n]$  is done by the following procedure. Start by setting  $B_0^i = \emptyset$ . For  $j = 1, 2, \dots$ , set  $S_j^i = B_0^i \cup \dots \cup B_{j-1}^i$ . If  $|S_j^i| > \delta n$  halt and set

$m_i = j - 1$  and  $A_i = \{B_1^i, \dots, B_{m_i}^i\}$ . Otherwise, it holds that for every  $c \in C$ , for every modification of the coordinates in  $S_j^i$  to some erroneous values,  $\text{Cor}(i)$  correctly outputs  $c_i$  with probability at least  $1 - \varepsilon$ . An equivalent description of this case is the following: for every  $c \in C$  and  $z : S_j^i \rightarrow \mathbb{F}$ , define  $c^z \in \mathbb{F}^n$  such that for every  $r \notin S_j^i$ ,  $c_r^z = c_r$  and for  $r \in S_j^i$ ,  $c_r^z = z(r)$ . The corrector  $\text{Cor}$  chooses a set of queries  $Q \subseteq [n] \setminus \{i\}$ ,  $|Q| \leq q$ , according to some distribution<sup>6</sup> and applies some function  $f_Q$  on  $c_Q^z$ . We know that with probability at least  $1 - \varepsilon$ ,  $f_Q(c_Q^z) = c_i$ . Since  $Q$  is sampled in a manner that is independent of  $c$  and  $z$ , by an averaging argument, there exists some fixed  $Q$  for which when  $c \in C$  and  $z : S_j^i \rightarrow \mathbb{F}$  are chosen randomly in a uniform manner, with probability at least  $1 - \varepsilon$  (this time over the choice of  $c$  and  $z$ ),  $f_Q(c_Q^z) = c_i$ . Therefore, we can define another function  $f'_Q$  that only gets  $c_{Q \setminus S_j^i}$ , chooses  $z$  uniformly at random, and outputs  $f_Q(c_Q^z)$ . If  $c \in C$  is chosen uniformly at random,  $f'_Q(c_{Q \setminus S_j^i}) = c_i$  with probability at least  $1 - \varepsilon > 1/|\mathbb{F}|$ . By Fact 3.7, this implies that  $Q \setminus S_j^i$  determines  $i$  in  $C$ . We therefore set  $B_j^i = Q \setminus S_j^i$ <sup>7</sup>, and proceed to the next  $j$ .

As this process only halts when  $|S_j^i| > \delta n$ , and for every  $j$ ,  $|S_j^i| \leq q(j - 1)$ , we have that  $m_i \geq \delta n/q$ . Further note that by the choice of each  $B_j^i$ , the sets  $B_1^i, \dots, B_{m_i}^i$  are disjoint, and of size at most  $q$ , as required. This thus shows how each  $A_i$  can be constructed, and the claim follows.  $\square$

## 4.2 Splitters for query sets

Splitters for query sets, that are defined as follows, are key ingredients in our distance amplification procedure. Informally, a  $c$ -splitter for a query set  $\mathcal{A} = \{A_1, \dots, A_n\}$  is partition of  $[n]$  which satisfies that for every  $i$ , the intersection between  $\overline{A}_i$ , the union all the sets in  $A_i$  that correspond to an index  $i$ , and each part of the partition, is not too large, i.e., of size at most  $c$ . In the distance amplification procedure, we will describe a corrector which samples a set  $B \in A_i$ , in some query set  $\mathcal{A}$ , and then makes queries according to which parts of the  $c$ -splitter intersect with  $B$ . For the resulted queries to be smooth, we will need the partition to “split”  $A_1, \dots, A_n$ , meaning that no part of the partition is too common within any certain  $A_i$ .

**Definition 4.8.** *Let  $n \in \mathbb{N}$ ,  $\mathcal{A}$  an  $n$ -query-set and  $c \in \mathbb{N}$ . A partition  $\pi$  of  $[n]$  is called a  $c$ -splitter of  $\mathcal{A}$  if for every  $i \in [n]$  and  $P \in \pi$ ,  $|P \cap \overline{A}_i| \leq c$ .*

The next claim shows that if each  $A_i$  is of size at most  $k$ , then  $c$ -splitters with  $k$  parts

---

<sup>6</sup>As the corrector in non-adaptive,  $\text{Cor}(i)$  naturally induces a distribution on subsets of  $[n]$  which correspond to the possible query sets.

<sup>7</sup>Note that  $i \notin B_j^i$ , as  $i \notin Q$ , since  $\text{Cor}(i)$  by definition never queries  $i$ .



exist, for  $c$ , the bound on the maximal intersection, being equal to roughly the minimal intersection that is possible, up to a constant factor.

**Claim 4.9.** *Let  $n, k, q \in \mathbb{N}$  such that  $k/n \leq 1$  and  $q \geq \log n$ . Further let  $\mathcal{A} = \{A_1, \dots, A_n\}$  be an  $n$ -query-set such that for every  $i \in [n]$ ,  $|A_i| \leq k$  and for every  $B \in A_i$ ,  $|B| \leq q$ . Then, there exists a partition  $\pi$  of  $[n]$  with  $k$  parts, each of size  $n/k$ , which is a  $c$ -splitter of  $\mathcal{A}$  for  $c = 2eq$ .*

*Proof.* The proof is by a probabilistic argument. We randomly choose a partition  $\pi$  with  $k$  equally-sized parts in a uniform manner among all such partitions. We bound the probability that  $\pi$  is not a  $c$ -splitter for  $\mathcal{A}$ : this is the case if  $|\overline{A}_i \cap P| > c$  for some  $i \in [n]$  and  $P$  a part of  $\pi$ . Towards this end, we first fix some  $i \in [n]$  and  $t \in [k]$ , and let  $P_t$  denote the  $t$ -th part of  $\pi$ . We have that for every  $j \in \overline{A}_i$  the probability that  $j \in P_t$  is  $1/k$ , and for every fixed subset of  $\overline{A}_i$  of size  $c$ , the probability that it is contained in  $P_t$  is at most  $(1/k)^c$  (since for distinct  $j, j' \in \overline{A}_i$ , the events that  $j \in P_t$  and  $j' \in P_t$  are negatively correlated). By a union bound over the possible subsets of size  $c$ , the probability that  $|\overline{A}_i \cap P_t| > c$  is at most

$$\begin{aligned} \binom{|\overline{A}_i|}{c} (1/k)^c &\leq \left( \frac{e|\overline{A}_i|}{ck} \right)^c \\ &\leq \left( \frac{eq}{c} \right)^c \\ &= \left( \frac{1}{2} \right)^{2eq}. \end{aligned}$$

By taking a union bound over all possible  $i, t$ , the probability that there exist  $i \in [n]$  and  $t \in [k]$  such that  $|\overline{A}_i \cap P_t| > c$  is at most  $nk \left( \frac{1}{2} \right)^{2eq} \leq n^2 \left( \frac{1}{2} \right)^{2eq}$ , which is less than 1 a  $q \geq \log n$ , and the claim follows.  $\square$

### 4.3 The distance amplification procedure

We now turn to define the basic operation behind our distance amplification procedure. This operation “composes”<sup>8</sup> two codes of different lengths, a big code and a small code, in a way that is parameterized by some partition of  $[n]$ . The result is a code of the same length as the big code, with an improved smoothness (if the partition satisfies certain requirements), as we will have in the claims that follow the definition. The distance amplification procedure (or perhaps, more directly, the smoothness amplification procedure) will be an iterative application of this composition.

---

<sup>8</sup>Note that the term “composition” here is used in a different sense than the usual composition of two codes in coding theory, which is achieved from the composition of the encoding functions.

**Definition 4.10.** Let  $C_1$  be a code of length  $n_1$ ,  $C_2$  a code of length  $n_2$ ,  $\pi$  a partition of  $[n_1]$  into  $n_1/n_2$  parts of size  $n_2$ . We define the  $\pi$ -composition of  $C_1$  and  $C_2$ , which we denote by  $C_1 \odot_\pi C_2$ , to be the code  $\{c \in C_1 \mid \forall P \in \pi \quad c_P \in C_2\}$ .

A bound on the rate of the composition of two codes is given in the following claim.

**Claim 4.11.** If  $C_1, C_2$  are codes with of lengths  $n_1, n_2$  and rates  $\rho_1, \rho_2$  respectively, then  $C = C_1 \odot_\pi C_2$  is a code of length  $n_1$  and rate at least  $\rho_1 + \rho_2 - 1$ .

*Proof.* That the length of  $C$  is  $n_1$  follows from the definition. As for the rate, by inspecting the code dual to  $C$ , it can be seen that the dimension of  $C^\perp$  is at most

$$d = (1 - \rho_1)n_1 + \frac{n_1}{n_2}(1 - \rho_2)n_2.$$

From that, the rate of  $C$  is at least  $1 - d/n_1 = \rho_1 + \rho_2 - 1$ .  $\square$

We now show that if the partition used in the composition is a  $c$ -splitter for a query set of the big code, the resulted code has smoothness roughly equal to the product of the two smoothnesses.

**Claim 4.12.** Let  $C_1$  be a code of length  $n_1$  and  $C_2$  a code of length  $n_2$  which is a  $(q_2, \tau_2)$ -LCC. Let  $\mathcal{A} = \{A_1, \dots, A_{n_1}\}$  be a query-set for  $C_1$  such that for every  $i$ ,  $|A_i| \geq 1/\tau_1$  and for every  $B \in A_i$ ,  $|B| \leq q_1$ . If  $\pi$  is a  $c$ -splitter for  $\mathcal{A}$ , then  $C = C_1 \odot_\pi C_2$  is a  $(q, \tau)$ -LCC for  $q = q_1 q_2$  and  $\tau = c\tau_1 \tau_2$ .

*Proof.* To show that  $C$  is a  $(q, \tau)$ -LCC we need to show a corrector  $\text{Cor}$  for it. We first set up some notations. Let  $\text{Cor}_2$  be a corrector promised by the fact that  $C_2$  is a  $(q_2, \tau_2)$ -LCC. For every  $j \in [n]$ , let  $P_j$  denote the part of  $\pi$  that contains  $j$ , and let  $\bar{j}$  denote the index of  $j$  in  $P_j$  with respect to the natural order. For  $i \in [n]$ , and  $B \in A_i$ , let  $f_{i,B} : \mathbb{F}^{|B|} \rightarrow \mathbb{F}$  denote a function satisfying  $f_{i,B}(c_B) = c_i$  for every  $c \in C_1$ . Such  $f_{i,B}$  is guaranteed to exist as  $\mathcal{A}$  is a query-set for  $C_1$ .

For  $i \in [n]$ ,  $\text{Cor}(i)$  with oracle access to  $c \in C$  acts as follows: it first samples  $B \in A_i$  uniformly at random. Secondly, for every  $j \in B$ , the procedure obtains  $c_j$  by invoking  $\text{Cor}_2(\bar{j})$  with oracle access to  $c_{P_j}$ . After obtaining  $c_j$  for every  $j \in B$ ,  $\text{Cor}(i)$  outputs  $f_{i,B}(c_B)$ .

That  $\text{Cor}(i)$  successfully outputs  $c_i$  for every  $c \in C$  is immediate, and follows from the fact that for every  $j$ ,  $c_{P_j}$  is a codeword of  $C_2$  and so  $\text{Cor}_2(\bar{j})$  with access to  $c_{P_j}$  correctly outputs  $c_j$ , and from the fact  $c \in C_1$  and so  $f_{i,B}(c_B) = c_i$ . Moreover,  $\text{Cor}(i)$  makes at most  $q_1 q_2$  queries to  $c$ , since  $|B| \leq q_1$  by assumption, and  $\text{Cor}_2$  makes at most  $q_2$  queries.

It remains to bound the probability that a coordinate  $r \in [n]$  is queried by  $\text{Cor}(i)$  for  $i \in [n]$ . Let  $p$  be the probability that  $\text{Cor}(i)$  queries  $r$ . Fix  $B \in A_i$ . Conditioned on the event that  $B$  was sampled by  $\text{Cor}(i)$  in the first step,  $r$  is queried by  $\text{Cor}(i)$  if one of the calls to  $\text{Cor}_2(\bar{j})$ , with oracle access to  $c_{P_j}$ , queries  $c_r$  for some  $j \in B$ . That probability is at most  $|B \cap P_r| \tau_2$ . Indeed, this follows by taking the union bound over the different  $j \in B$ , noting that if  $j \notin P_r$ ,  $c_r$  cannot be queried by  $\text{Cor}_2(\bar{j})$ , and using that  $\text{Cor}_2$  queries any coordinate with probability bounded above by  $\tau_2$ . Therefore,

$$\begin{aligned}
p &\leq \sum_{B \in A_i} \Pr[B \text{ is sampled by } \text{Cor}(i)] \cdot |B \cap P_r| \tau_2 \\
&= \sum_{B \in A_i} \frac{1}{|A_i|} \cdot |B \cap P_r| \tau_2 \\
&\leq \sum_{B \in A_i} \tau_1 \cdot |B \cap P_r| \tau_2 \\
&= \tau_1 \tau_2 |P_r \cap \overline{A_i}| \\
&\leq c \tau_1 \tau_2.
\end{aligned}$$

Note that we used the assumptions that  $|A_i| \geq 1/\tau_1$ , and that  $\pi$  is a  $c$ -splitter for  $\mathcal{A}$ . We thus have that  $p \leq c \tau_1 \tau_2$ , which concludes the proof.  $\square$

The following lemma concludes the properties of the code that is achieved by the composition of two codes, when done with the  $c$ -splitter that is given by Claim 4.9.

**Lemma 4.13.** *Let  $n \in \mathbb{N}$ . Assume there exists a code  $C_1$  of length  $n$  over  $\mathbb{F}$ , with rate  $\rho_1$ , which is a  $(q_1, \tau_1)$ -query-set-LCC for  $q_1 \geq \log n$ . Further assume that there exists a code  $C_2$  of length  $n\tau_1$  over  $\mathbb{F}$ , with rate  $\rho_2$ , which is a  $(q_2, \tau_2)$ -LCC. Then, there exists a code  $C$  of length  $n$ , with rate  $\rho_1 + \rho_2 - 1$ , which is a  $(q_1 q_2, 2eq_1 \tau_1 \tau_2)$ -LCC.*

*Proof.* As  $C_1$  is a  $(q_1, \tau_1)$ -query-set-LCC, there exists an  $n$ -query-set  $\mathcal{A} = \{A_1, \dots, A_n\}$  in which for every  $i$ ,  $|A_i| \geq 1/\tau_1$  and for every  $B \in A_i$ ,  $|B| \leq q_1$ . In particular, there exists a query set  $\mathcal{A}' = \{A'_1, \dots, A'_n\}$  in which every  $A'_i$  is of size exactly  $1/\tau_1$  (which is achieved by, for each  $A_i$ , arbitrarily removing sets  $B \in A_i$  until it is of size  $1/\tau_1$ ). By Claim 4.9 invoked with  $k = 1/\tau_1$ , there exists a partition  $\pi$  of  $[n]$ , in which every part is of size  $\tau_1 n$ , which is a  $c$ -splitter for  $\mathcal{A}'$ , with  $c = 2eq_1$ . We take  $C = C_1 \odot_\pi C_2$  to be the code with the claimed properties. Indeed, by Claim 4.11,  $C$  is of length  $n$ , and has rate at least  $\rho_1 + \rho_2 - 1$ . Furthermore, by applying Claim 4.12, and using that  $\pi$  is a  $c$ -splitter for  $\mathcal{A}'$ , we get that  $C$  is a  $(q, \tau)$ -LCC for  $q = q_1 q_2$  and  $\tau = 2eq_1 \tau_1 \tau_2$ , and the lemma follows.  $\square$

The following lemma, or more precisely, its proof, composes the distance amplification procedure. It assumes a family of codes which are LCC, and describes the properties of

the code that is obtained by an iterative application of the composition, where at each iteration a code of the family is composed with the “current” code.

**Lemma 4.14.** *Assume there exists a family of codes  $C = \{C^n\}$  over  $\mathbb{F}$ , in which every code  $C^n$  of length  $n$  in the family is a code of rate  $\rho(n) = 1 - r(n)$ , which is a  $(q(n), \tau(n))$ -query-set-LCC for  $q(n) \geq \log n$ . Then, for every  $t \in \mathbb{N}$ , there exists a code family  $C' = \{(C')^n\}$  over  $\mathbb{F}$  which has a code  $(C')^n$  of length  $n$  for every  $n$  which is a code length in  $C$ , and  $(C')^n$  has the following properties. Define  $n_1 = n$  and for  $i = 2, \dots, t + 1$  let  $n_i = \lceil \tau(n_{i-1})n_{i-1} \rceil^C$ . Then,  $(C')^n$  has rate  $\rho'(n) = 1 - \sum_{i=1}^t r(n_i)$ , and is a  $(q'(n), \tau'(n))$ -LCC for  $q'(n) = \prod_{i=1}^t q(n_i)$  and*

$$\tau'(n) = (2e)^{t-1} \frac{n_{t+1}}{n} \prod_{i=1}^{t-1} q(n_i).$$

*Proof.* To show the existence of a code family with the claimed properties, we describe how for every  $n$  that is a length of a code in the family  $C$ , a code of the same length, of the family  $C'$ , can be constructed. Let  $C^n$  be a code of length  $n$  of the family  $C$ . Set  $n_1 = n$  and for  $i = 2, \dots, t + 1$ ,  $n_i = \lceil \tau(n_{i-1})n_{i-1} \rceil^C$ , as defined in the claim. We construct a sequence of codes  $C'_1, \dots, C'_t$ , where for each  $i \in [t]$ ,  $C'_i$  is a code of length  $n_i$  and rate  $\rho'_i$ , which is a  $(q'_i, \tau'_i)$ -LCC. We start by setting  $C'_t = C^{n_t}$ , and for  $i = t - 1, \dots, 1$ , we take  $C'_i$  to be a code which is the result of applying Lemma 4.13 on  $C^{n_i}$  and  $C'_{i+1}$ . Note that  $C^{n_i}$  is a  $(q(n_i), \tau(n_i))$ -query-set-LCC and  $C'_{i+1}$  is a code of length  $n_{i+1} \geq \tau(n_i)n_i$ , and so in particular  $C^{n_i}$  is indeed of smoothness  $n_{i+1}/n_i$ , as required for the lemma to be applicable. From Lemma 4.13 it follows that  $C'_i$  is a code of rate

$$\rho'_i = \rho(n_i) + \rho'_{i+1} - 1 = \rho'_{i+1} - r(n_i)$$

which is a  $(q'_i, \tau'_i)$ -LCC for

$$\begin{aligned} q'_i &= q(n_i)q'_{i+1}, \\ \tau'_i &= 2eq(n_i)\tau'_{i+1} \frac{n_{i+1}}{n_i}. \end{aligned}$$

Recall that  $C'_t = C^{n_t}$  and so  $\rho'_t = 1 - r(n_t)$ ,  $q'_t = q(n_t)$  and  $\tau'_t = \tau(n_t)$ . It follows inductively that for every  $i \in [t]$ ,

$$\begin{aligned} \rho'_i &= 1 - \sum_{j=i}^t r(n_j), \\ q'_i &= \prod_{j=i}^t q(n_j), \end{aligned}$$

and

$$\begin{aligned}\tau'_i &= (2e)^{t-i} \left( \prod_{j=i}^t \frac{n_{j+1}}{n_j} \right) \left( \prod_{j=i}^{t-1} q(n_j) \right) \\ &= (2e)^{t-i} \frac{n_{t+1}}{n_i} \left( \prod_{j=i}^{t-1} q(n_j) \right).\end{aligned}$$

We set  $C'_1$ , which is indeed a code of length  $n$ , to be the code  $(C')^n$  of  $C'$ , and from the account given above it follows that its rate, query complexity and smoothness are as stated, i.e., that  $q'_1 = q'(n)$ ,  $\rho'_1 = \rho'(n)$  and  $\tau'_1 = \tau'(n)$ . We thus have that  $C'$  is a family of codes with rate at least  $\rho(n)$  that are  $(q(n), \tau(n))$ -LCC, and the lemma follows.  $\square$

## 4.4 Deriving the Corollaries

In this part we deduce two corollaries of our distance amplification procedure that is given by Lemma 4.14. As a special case of the first corollary, Corollary 4.16, we will have that if one has a sufficiently dense code family of  $(q(n), \delta(n), \varepsilon(n))$ -LCC which is of high rate, meaning that each code has rate  $\rho(n)$  that approaches 1 “fast enough”, but with  $\delta(n)$  that is only polynomially small in  $n$ ,  $\delta(n) = 1/n^\alpha$ , for some constant  $\alpha \in (0, 1)$ , then there exists a good family of LCC with query complexity  $q(n)^{O(\log \log n)}$ . In the general case, a weaker guarantee on  $\delta(n)$  can also be handled by Corollary 4.16, meaning that a sub-polynomial  $\delta(n)$  can also be amplified. More precisely, Corollary 4.16 will state that if  $\delta(n) = 1/n^{1-1/g(n)}$  for a (non-decreasing) function  $g(n)$ , then a family of good LCC can be constructed, with query complexity  $q(n)^{O(g(n) \log \log n)}$ . The requirement of the rate function  $\rho(n)$ , which we described as approaching 1 “fast enough”, in more detail comes down to the requirement that  $\rho(n) \geq 1 - 1/(g(n)(\ln \ln n)^2)$ .

The second corollary, Corollary 4.19, addresses the case that the family of  $(q(n), \delta(n), \varepsilon(n))$ -LCC one starts with is of a much smaller rate, either of a constant rate or of a vanishing rate of  $(1/\ln \ln n)^h$  for some constant  $h$ . In the case that  $\delta(n) = 1/n^\alpha$  for some constant  $\alpha \in (0, 1)$  and  $\rho(n) \geq (1/\ln \ln n)^h$ , as a special case Corollary 4.19 we will have that there exists a family of good LCC with query complexity  $q(n)^{\text{poly}(\log \log n)}$ . Here too, sub-polynomial  $\delta(n)$  can also be handled by the corollary, as in a more general case, it is shown by Corollary 4.19 that if  $\delta(n) = 1/n^{1-1/g(n)}$  for a non-decreasing  $g(n) \leq \log n$ , and if  $\rho(n)$  is at least  $(1/\ln \ln n)^h$  for some constant  $h$ , then a family of good LCC can be constructed, with query complexity  $q(n)^{g(n) \text{poly}(\log \log n)}$ . The precise statement Corollary 4.19 is more generally stated and handles a few more cases that may be of interest.

We remark that while in any case that Corollary 4.16 can be applied so can Corollary 4.19 be used, the reason that we state both corollaries is that if one starts with

an LCC that satisfies the requirement of Corollary 4.16 then using it, rather than using Corollary 4.19, would result in a better bound on the resulted query complexity. We further remark that the proof for Corollary 4.19 builds on Corollary 4.16. Lastly, another reason that Corollary 4.16 is of interest is that it has an analogous corollary in the case of LDC (see Corollary 4.17), unlike Corollary 4.19 (whose proof relies on properties specific to LCC).

#### 4.4.1 From high rate and low distance LCC to good LCC

To prove the first corollary, we will need the following lemma which states that any family of  $(q, \tau)$ -LCC with constant rate can be converted to a family of good LCC by paying a multiplicative factor of  $\text{poly}(\tau n)$  in query complexity. This lemma follows from the AEL distance amplification procedure [AL96, AEL95] and from the adaptation of it by [KMRS17] for LDC and LCC. To derive this lemma with certain parameters, some adaptations to these techniques are needed, and so we provide a full proof for Lemma 4.15 in the appendix (Section A), for completeness.

**Lemma 4.15.** *Let  $C = \{C^n\}$  be a code family over  $\mathbb{F}$  in which every code  $C^n$  is a  $(q(n), \tau(n))$ -LCC with rate  $\rho(n) = \Omega(1)$ . Then, there exists a code family  $C' = \{(C')^n\}$  over  $\mathbb{F}$  which has a code  $(C')^n$  of length  $n$  for every  $C^n$  in  $C$ , such that  $(C')^n$  is a  $(q'(n), \delta'(n), \varepsilon)$ -LCC for  $q'(n) = O(q(n)(n\tau(n))^2)$ ,  $\delta'(n) = \Omega(1)$  and  $\varepsilon = 1/3$ , with rate  $\rho'(n) = \Omega(1)$ .*

We now state our first corollary.

**Corollary 4.16.** *Let  $q(n) \geq \log n^9$  and  $g(n) > 1$  be two non-decreasing functions. Assume there exists a family of codes  $C = \{C^n\}$  over  $\mathbb{F}$  that is  $(n_0, c, d)$ -dense, in which every code  $C^n$  of length  $n$  has rate*

$$\rho(n) \geq 1 - \frac{1}{g(n)(\ln \ln n)^2},$$

*and either  $C^n$  is a  $(q(n), \delta(n), \varepsilon(n))$ -LCC, for  $\varepsilon(n) < 1 - 1/|\mathbb{F}|$  and  $\delta(n) = 1/n^{1-1/g(n)}$ , or it is a  $(q(n), \tau(n))$ -query-set-LCC, for  $\tau(n) = q(n)/n^{1/g(n)}$ . Then, there exists a family of codes  $C' = \{(C')^n\}$  over  $\mathbb{F}$  that is  $(n_0, c, d)$ -dense, which is a family of good LCC with query complexity  $q_{\text{new}}(n) = q(n)^{O(g(n) \ln \ln n)}$ .*

Note that Corollary 4.16 allows for the code family  $C$  in the hypothesis to be one of two types, either a family of  $(q, \delta, \varepsilon)$ -LCC or a family of  $(q, \tau)$ -query-set-LCC. For the

---

<sup>9</sup>We remark that while we assume for simplicity that  $q(n) \geq \log n$ , by the Katz-Trevisan bound (instantiated for the case of rate and distance as specified by the corollary), lifting this assumption would not yield an improvement in the obtained query complexity in any case.

proof, what we actually need is that  $C$  is of the second type. However, if one starts with a family  $C$  which is known to be of the first (more standard) type, with the specified  $\delta(n)$ , by Claim 4.7 it will follow that  $C$  is a family of query-set-LCC with the same smoothness  $\tau(n)$  that is stated in the corollary in the second case. The corollary explicitly allows both of the types because it is also possible that the base code is already known to be a query-set-LCC, as would be the case in the proof of Corollary 4.19, which uses Corollary 4.16. It is preferable to avoid going back and forth between the types, as this has some cost in the resulted parameters.

Before giving the proof for Corollary 4.16, we state a corollary analogous to it, that holds in the case of LDC. The proof for this corollary is straightforward given the result regarding LCC, and follows the same lines.<sup>10</sup>

**Corollary 4.17.** *Let  $n(k) > k$ ,  $q(k) \geq \log n(k)$  and  $g(k) > 1$  be non-decreasing functions. Assume there exists a code-encoding family  $C = \{(C^k, \text{Enc}^k)\}$  over  $\mathbb{F}$  that is  $(k_0, c, d)$ -dense, in which every code  $C^k$  of dimension  $k$  has rate*

$$\rho(k) \geq 1 - \frac{1}{g(k)(\ln \ln k)^2} > \frac{1}{2},$$

*and either  $(C^k, \text{Enc}^k)$  is a  $(q(k), \delta(k), \varepsilon(k))$ -LDC, for  $\varepsilon(k) < 1 - 1/|\mathbb{F}|$  and  $\delta(k) = 1/n(k)^{1-1/g(k)}$ , or it is a  $(q(k), \tau(k))$ -query-set-LDC, for  $\tau(k) = q(k)/n(k)^{1/g(k)}$ . Then, there exists a code-encoding family  $C' = \{((C')^k, (\text{Enc}')^k)\}$  over  $\mathbb{F}$  that is  $(k_0, c, d)$ -dense, which is a family of good LDC with query complexity  $q_{\text{new}}(k) = q(k)^{O(g(k) \ln \ln k)}$ .*

*Proof of Corollary 4.16.* We argue that a code family of the claimed properties can be constructed, and specifically we will show such a code family that has a code of length  $n$  for every  $n$  which is a code length of the family  $C$ . The underlying idea of the proof consists of applying, for every  $n$  which is a code length in  $C$ , Lemma 4.14 with the family  $C$  and some appropriate  $t \in \mathbb{N}$  which depends on  $n$ , to get that there exists a code family  $C'$  in which  $(C')^n$  is a code of length  $n$  with desired properties.

With that plan in mind, let  $C^n$  be a code in  $C$  of length  $n$ . First, note that either by assumption, or by Claim 4.7 (if  $C^n$  is given as a  $(q(n), \delta(n), \varepsilon(n))$ -LCC), we have that  $C^n$  is a  $(q(n), \tau(n))$ -query-set-LCC, as required by the hypothesis of Lemma 4.14. Now, we invoke Lemma 4.14 with the code family  $C$  and with  $t$  to be chosen later, to get a code family  $C'$  and a code  $(C')^n$  within it. We follow the notation of Lemma 4.14 and set  $n_1 = n$  and for  $i = 2, \dots, t + 1$ ,  $n_i = \lceil \tau(n_{i-1})n_{i-1} \rceil^C$ . We have that

$$n_i = \lceil q(n_{i-1})n_{i-1}^{(1-1/g(n_{i-1}))} \rceil^C, \quad (4.1)$$

---

<sup>10</sup>A separate proof will appear in the full version of this paper.

and if we set  $c' = n_0 + c$ , as  $C$  is  $(n_0, c, d)$ -dense,

$$n_i \leq c' \cdot q(n_{i-1}) n_{i-1}^{(1-1/g(n_{i-1}))}.$$

Using the facts that both  $q(n)$  and  $g(n)$  are non-decreasing, it follows that

$$n_i \leq (c' \cdot q(n))^{\sum_{j=0}^{i-1} ((1-1/g(n))^j)} n^{\prod_{j=1}^{i-1} (1-1/g(n_j))},$$

and as  $\sum_{j=0}^{i-1} (1-1/g(n))^j \leq g(n)$ , we get that

$$n_i \leq (c' \cdot q(n))^{g(n)} n^{\prod_{j=1}^{i-1} (1-1/g(n_j))}.$$

We need to choose  $t$  so that  $n_{t+1}$  is minimized (as  $n_{t+1}$  affects the resulted smoothness). We choose  $t$  to be the minimal integer satisfying

$$n^{\prod_{j=1}^t (1-1/g(n_j))} \leq e(c' \cdot q(n))^{g(n)}.$$

Note that this choice implies

$$n_{t+1} \leq e(c' \cdot q(n))^{2g(n)} \tag{4.2}$$

and

$$n^{\prod_{j=1}^{t-1} (1-1/g(n_j))} > e(c' \cdot q(n))^{g(n)}. \tag{4.3}$$

In order to verify that with that choice of  $t$ ,  $(C')^n$  has the claimed properties, we need to bound from above the value of  $t$  which attains this (this would also imply that our choice of  $t$  is well defined, i.e., that such  $t$  exists). For any  $t' \geq g(n) \ln \ln n$ , we have that

$$\begin{aligned} n^{\prod_{j=1}^{t'} (1-1/g(n_j))} &\leq n^{(1-1/g(n))^{t'}} \\ &\leq n^{e^{-t'/g(n)}} \\ &\leq n^{e^{-(g(n) \ln \ln n)/g(n)}} \\ &= e \\ &\leq e(c' \cdot q(n))^{g(n)}, \end{aligned}$$

where the second inequality follows from that  $e^x \geq 1 + x$  for any  $x \in \mathbb{R}$ . Therefore, we have that  $t \leq g(n) \ln \ln n$ . By the conclusion of Lemma 4.14,  $(C')^n$  is a code with rate

$$\rho'(n) \geq 1 - \sum_{i=1}^t \frac{1}{g(n_i) (\ln \ln n_i)^2},$$

which is a  $(q'(n), \tau'(n))$ -LCC, for

$$q'(n) \leq q(n)^t = q(n)^{O(g(n) \ln \ln n)},$$



and

$$\begin{aligned}
\tau'(n) &= (2e)^{t-1} \cdot \frac{n_{t+1}}{n} \cdot \prod_{i=1}^{t-1} q(n_i) \\
&\leq q(n)^{O(g(n) \ln \ln n)} \cdot \frac{n_{t+1}}{n} \\
&\leq q(n)^{O(g(n) \ln \ln n)} \cdot \frac{1}{n},
\end{aligned}$$

where the last inequality follows as  $q(n) \geq \log n$  and by Equation (4.2).

If it is the case that  $\rho'(n) = \Omega(1)$  we can conclude the proof by invoking Lemma 4.15. Taking the set of codes  $(C')^n$  (for every  $n$  a code length in  $C$ ) to be the code family of the hypothesis of Lemma 4.15, we would get that there exists a code family  $C'' = \{(C'')^n\}$  which satisfies the following.  $C''$  is a family of codes in which every  $(C'')^n$  is a  $(q''(n), \delta''(n), \varepsilon)$ -LCC, for  $q''(n) = O(q'(n)(n\tau'(n))^2) = q(n)^{O(g(n) \ln \ln n)}$ ,  $\delta''(n) = \Omega(1)$ , and  $\varepsilon = 1/3$ . Thus,  $C''$  is a family of a good LCC with query complexity  $q_{\text{new}}(n) = q(n)^{O(g(n) \log \log n)}$ , the argued query complexity. Moreover, this code family has the same code lengths as  $C$ , and is thus  $(n_0, c, d)$ -dense as well.

It only remains to show that  $\rho'(n) = \Omega(1)$ . To bound  $\rho'(n)$  from below we need to bound  $\sum_{i=1}^t \frac{1}{g(n_i)(\ln \ln n_i)^2}$  from above. First, we define  $\bar{n}_1 = n$  and for  $i > 1$ ,  $\bar{n}_i = \bar{n}_{i-1}^{1-1/g(n_{i-1})}$ . As by Equation (4.1),  $n_i \geq n_{i-1}^{1-1/g(n_{i-1})}$ , it follows by induction that  $n_i \geq \bar{n}_i$ , and note that  $\bar{n}_i = n^{\prod_{j=1}^{i-1} (1-1/g(n_j))}$ . We thus have that

$$\sum_{i=1}^t \frac{1}{g(n_i)(\ln \ln n_i)^2} \leq \sum_{i=1}^t \frac{1}{g(n_i)(\ln \ln \bar{n}_i)^2}.$$

Since

$$\ln \ln \bar{n}_i = \ln \ln n + \sum_{j=1}^{i-1} \ln \left( 1 - \frac{1}{g(n_j)} \right),$$

we have that

$$\begin{aligned}
g(n_i)(\ln \ln \bar{n}_i) - g(n_i)(\ln \ln \bar{n}_{i+1}) &= -g(n_i) \ln \left( 1 - \frac{1}{g(n_i)} \right) \\
&\geq g(n_i) \frac{1}{g(n_i)} \\
&= 1,
\end{aligned}$$

where we used that  $x \geq \ln(1+x)$ , for every  $x > 0$ . With that, we can deduce the following

$$\begin{aligned} \frac{1}{g(n_i)(\ln \ln \bar{n}_i)^2} &\leq \int_{g(n_i) \ln \ln \bar{n}_{i+1}}^{g(n_i) \ln \ln \bar{n}_i} \frac{1}{g(n_i)(\ln \ln \bar{n}_i)^2} dx \\ &= \int_{\ln \ln \bar{n}_{i+1}}^{\ln \ln \bar{n}_i} \frac{g(n_i)}{g(n_i)(\ln \ln \bar{n}_i)^2} dx \\ &\leq \int_{\ln \ln \bar{n}_{i+1}}^{\ln \ln \bar{n}_i} \frac{1}{x^2} dx, \end{aligned}$$

where the last inequality follows from that  $\bar{n}_{i+1} \leq \bar{n}_i$ . This implies that

$$\begin{aligned} \sum_{i=1}^t \frac{1}{g(n_i)(\ln \ln \bar{n}_i)^2} &\leq \int_{\ln \ln \bar{n}_{t+1}}^{\ln \ln \bar{n}_1} \frac{1}{x^2} dx \\ &= \frac{\ln \ln \bar{n}_1 - \ln \ln \bar{n}_{t+1}}{(\ln \ln \bar{n}_1)(\ln \ln \bar{n}_{t+1})} \\ &\leq \frac{1}{\ln \ln \bar{n}_{t+1}}. \end{aligned}$$

Furthermore, by Equation (4.3), we have that  $\bar{n}_t \geq e(c' \cdot q(n))^{g(n)}$ , and so

$$\bar{n}_{t+1} = \bar{n}_t^{1-1/g(n_t)} \geq e(c' \cdot q(n))^{(1-1/g(n_t))g(n)}.$$

Moreover, we have that

$$\left(1 - \frac{1}{g(n_t)}\right) g(n) = \Omega(g(n)),$$

as  $g(n)$  is non-decreasing and  $g(n) > 1$ , and so it follows that

$$\ln \ln \bar{n}_{t+1} = \Omega(\ln \ln q(n) + \ln g(n)).$$

Therefore,

$$\begin{aligned} \rho'(n) &\geq 1 - \frac{1}{\ln \ln \bar{n}_{t+1}} \\ &= 1 - O\left(\frac{1}{\ln \ln q(n) + \ln g(n)}\right), \end{aligned}$$

which establishes that indeed  $\rho'(n) = \Omega(1)$ , and the claim follows.  $\square$

#### 4.4.2 From low rate and low distance LCC to good LCC

For the proof of our second corollary we will need the following proposition from [CY21]. This proposition is basically Proposition 4.14 in [CY21] but for  $(q, \tau)$ -query-set-LCC rather than for a different object<sup>11</sup>. That the proposition indeed applies to  $(q, \tau)$ -query-set-LCC is quite immediate with the account given in [CY21].

<sup>11</sup>“dual SLR” in the terminology of [CY21].

**Proposition 4.18** (Implicit in [CY21]). *Let  $C$  be a code of length  $n$  over  $\mathbb{F}$  with rate  $\rho$  that is a  $(q, \tau)$ -query-set-LCC. Then, for every  $\ell \in \mathbb{N}$ , there exists a code  $C'$  of length  $n' = n^\ell$  with rate  $1 - (1 - \rho)^\ell$ , which is a  $(q', \tau)$ -query-set-LCC for  $q' = q^\ell$ .*

We remark that while the rate of the resulted code of Proposition 4.18 is improved compared to that of the starting code of the hypothesis, it's smoothness is quite bad. Note that this is true even though the smoothness of the obtained code is  $\tau$ , i.e., the same as that of the initial code, because the length of the code has increased. Indeed, even if originally the smoothness of  $C$  was the best possible,  $\tau = \Theta(q/n)$ , for  $\ell > 1$  the smoothness of  $C'$  is, at best, polynomially small in its length. Nonetheless, for our purposes these codes will do. We point out that while we do not need to use it here, in [CY21] a procedure that amplifies rate while maintaining the smoothness is given, as in [CY21] the objective is to show that LCC with vanishing rate that can be as small as  $1/\sqrt{\log n}$  implies an LCC with constant rate, and for that such a procedure is crucial. In our case the codes of Proposition 4.18 are satisfactory.

We now state our second corollary.

**Corollary 4.19.** *Let  $h \geq 1$  be an arbitrary constant,  $q(n) \geq \log n$  and  $g(n) \in [1, \log n]$  non-decreasing functions, and  $\rho(n)$  a non-increasing function, satisfying*

$$\frac{1}{(\ln \ln n)^h} \leq \rho(n) \leq 1 - \frac{1}{g(n)(\ln \ln n)^2}$$

for every  $n$ . Assume further that

$$\frac{1}{\rho(n+1)}(\ln g(n+1) + \ln \ln \ln(n+1)) - \frac{1}{\rho(n)}(\ln g(n) + \ln \ln \ln n) = O\left(\frac{1}{\log n}\right).$$

Assume there exists a family of codes  $C = \{C^n\}$  over  $\mathbb{F}$  that is  $(n_0, 1, 1)$ -dense<sup>12</sup>, in which every code  $C^n$  of length  $n$  is a code of rate  $\rho(n)$ , which is a  $(q(n), \delta(n), \varepsilon(n))$ -LCC, for  $\varepsilon(n) < 1 - 1/|\mathbb{F}|$  and

$$\delta(n) = \frac{1}{n^{1-1/g(n)}}.$$

Then, there exists a family of codes  $C' = \{(C')^n\}$  over  $\mathbb{F}$ , which is a family of good LCC with query complexity  $q_{\text{new}}(n) = q(n)^{e(n)}$  for

$$e(n) = O\left(\frac{1}{\rho(n)^2}(\ln g(n) + \ln \ln \ln n)^2 g(n) \ln \ln n\right).$$

---

<sup>12</sup>Note that if one starts with a code family  $C$  that is  $(n_0, c, d)$  for some constants  $c, d$ , then it can be easily converted to a  $(n_0, 1, 1)$ -dense family, with a constant multiplicative cost to the rate and with little affect to the obtained parameters.

*Proof.* To show that the family  $C$  can be converted to a family of good LCC, the idea is to apply Corollary 4.16. In order to be able to do so, we first need to show that the family  $C$  can be converted to a family  $C'$  with high enough rate. We now explain how such a family  $C'$  can be constructed. First we define, for every  $n$  which is a code length in  $C$ ,

$$\ell(n) = 10h \frac{1}{\ln\left(\frac{1}{1-\rho(n)}\right)} (\ln g(n) + \ln \ln \ln n), \quad (4.4)$$

and  $N(n) = n^{\ell(n)}$ . We have that

$$\ell(n) = \Theta\left(\frac{1}{\rho(n)} (\ln g(n) + \ln \ln \ln n)\right), \quad (4.5)$$

and per our assumption that  $\rho(n) \leq 1 - \frac{1}{g(n)(\ln \ln n)^2}$ , it follows that  $\ell(n) > 1$ , and thus

$$N(n) > n. \quad (4.6)$$

Further note that the function  $N(n)$  is strictly increasing, and so it can be seen that there exists a function  $\hat{n} : \mathbb{N} \rightarrow \mathbb{N}$  which is strictly increasing as well, and which satisfies that for every  $n$  which is a code length in  $C$ ,  $\hat{n}(N(n)) = n$ . Let  $\hat{n}$  be any such function. Now, by Claim 4.7 every  $C^n$  a code of length  $n$  of the family  $C$ , is a  $(q(n), \tau(n))$ -query-set-LCC for

$$\tau(n) = \frac{q(n)}{n^{1/g(n)}}.$$

For every such  $C^n$ , we apply Proposition 4.18 with  $\ell = \ell(n)$ , to get that there exists a code  $(C')^{N(n)}$  over  $\mathbb{F}$ , of length  $N(n)$ , with the parameters detailed by the theorem. We define the code family  $C'$  to be  $\{(C')^{N(n)} \mid C^n \in C \text{ and } n \geq \max(n_0, n_1)\}$ , for some constant  $n_1 \in \mathbb{N}$  to be chosen later. Note that as the function  $N(n)$  is strictly increasing,  $C'$  has at most one code of every length. Define for every  $N \in \mathbb{N}$

$$\begin{aligned} g'(N) &= \ell(\hat{n}(N))g(\hat{n}(N)), \\ \rho'(N) &= 1 - (1 - \rho(\hat{n}(N)))^{\ell(\hat{n}(N))}, \\ q'(N) &= q(\hat{n}(N))^{\ell(\hat{n}(N))}, \\ \tau'(N) &= \frac{q'(N)}{N^{1/g'(N)}}. \end{aligned}$$

By Theorem 4.18, every  $(C')^N$  is a code of length  $N$  of  $C'$  with rate  $\rho'(N)$ . Further it is

a  $(q'(N), \tau(\hat{n}(N)))$ -query-set-LCC. We have that

$$\begin{aligned}
\tau(\hat{n}(N)) &= \frac{q(\hat{n}(N))}{\hat{n}(N)^{1/g(\hat{n}(N))}} \\
&= \frac{q(\hat{n}(N))}{N^{1/(\ell(\hat{n}(N))g(\hat{n}(N)))}} \\
&= \frac{q(\hat{n}(N))}{N^{1/g'(N)}} \\
&\leq \frac{q'(N)}{N^{1/g'(N)}} \\
&= \tau'(N).
\end{aligned}$$

It follows that every  $(C')^N$  is in particular a  $(q'(N), \tau'(N))$ -query-set-LCC.

With the family  $C'$  at hand we wish to invoke Corollary 4.16, but before we can do that, we need to verify that it satisfies the corollary's hypotheses. First, we need to verify that for every code length  $N$  which is a code length of  $C'$ ,  $q'(N) \geq \log N$ , and indeed as  $q'(N) = q(\hat{n}(N))^{\ell(\hat{n}(N))}$ ,  $\log N = \ell(\hat{n}(N)) \log \hat{n}(N)$  and  $q(\hat{n}(N)) \geq \log \hat{n}(N)$  this holds for every  $\hat{n}(N) \geq 4$ . Secondly, we need to verify that  $1/(1 - \rho'(N)) \geq g'(N)(\ln \ln N)^2$  for every  $N$  which is a code length of  $C'$ . Equivalently, we need to verify that for every  $n \geq n_1$  a code length of  $C$ ,  $1/(1 - \rho'(N(n))) \geq g'(N(n))(\ln \ln N(n))^2$ . Indeed, on the one hand we have that

$$\begin{aligned}
\frac{1}{1 - \rho'(N(n))} &= \frac{1}{(1 - \rho(n))^{\ell(n)}} \\
&= \frac{1}{(1 - \rho(n))^{\frac{10h}{\ln\left(\frac{1}{1-\rho(n)}\right)}(\ln g(n) + \ln \ln \ln n)}} \\
&= (g(n) \ln \ln n)^{10h}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
g'(N(n))(\ln \ln N(n))^2 &= \ell(n)g(n)(\ln \ln n + \ln \ell(n))^2 \\
&\leq \ell(n)^2 g(n)(\ln \ln n)^2 \\
&= \left(10h \frac{1}{\ln\left(\frac{1}{1-\rho(n)}\right)} (\ln g(n) + \ln \ln \ln n)\right)^2 g(n)(\ln \ln n)^2 \\
&\leq \left(10h \frac{1}{\ln(1 + \rho(n))}\right)^2 g(n)^3 (\ln \ln n)^4 \\
&= O\left(\frac{1}{\rho(n)^2} g(n)^3 (\ln \ln n)^4\right) \\
&= O(g(n)^3 (\ln \ln n)^{2h+4}),
\end{aligned}$$

where the penultimate equality holds since  $\ln(1 + \rho(n)) = \Omega(\rho(n))$  (as  $\rho(n)$  is non-increasing), and the last equality is due to the hypothesis  $\rho(n) \geq 1/(\ln \ln n)^h$ . Thus we have that

$$\frac{1}{1 - \rho'(N(n))} = \omega(g'(N(n))(\ln \ln N(n))^2).$$

We therefore set  $n_1$  to be the minimal satisfying that for every  $n \geq n_1$ ,  $1/(1 - \rho'(N(n))) \geq g'(N(n))(\ln \ln N(n))^2$ , and  $n_1 \geq 4$ . Note that  $n_1$  is some (well defined) constant. With that choice we indeed have that for every  $(C')^N \in C'$ ,  $1/(1 - \rho'(N)) \geq g'(N)(\ln \ln N)^2$ .

Another thing that we need to verify is that for some constants, the family  $C'$  is  $(n'_0, c', d')$ -dense, and observe that this holds if  $\lceil N \rceil^{C'}/N = O(1)$ . We have that for every  $N \geq N(\max(n_0, n_1))$ ,  $\lceil N \rceil^{C'} \leq N(\lceil \hat{n}(N) \rceil)$ . This holds since, notice,  $\hat{n}(N)$  is defined for every  $N \in \mathbb{N}$ , and we have that  $\lceil \hat{n}(N) \rceil \in \mathbb{N}$  and  $\hat{n}(N) \geq n_0$  (as  $N \geq N(n_0)$ ), and so by the fact that  $C$  is  $(n_0, 1, 1)$ -dense,  $\lceil \hat{n}(N) \rceil$  is a code length of  $C$ . Thus,  $N(\lceil \hat{n}(N) \rceil)$  is a code length of  $C'$  which satisfies  $N(\lceil \hat{n}(N) \rceil) \geq N$  which shows that indeed  $\lceil N \rceil^{C'} \leq N(\lceil \hat{n}(N) \rceil)$ . Furthermore, as  $N(n)$  is increasing,  $N(\lceil \hat{n}(N) \rceil) \leq N(\hat{n}(N) + 1)$ , and therefore it is enough to verify that

$$\frac{N(n+1)}{N(n)} = O(1). \tag{4.7}$$

We have that

$$\begin{aligned} \frac{N(n+1)}{N(n)} &= \frac{(n+1)^{\ell(n+1)}}{n^{\ell(n)}} \\ &= n^{\ell(n+1) - \ell(n)} \left(1 + \frac{1}{n}\right)^{\ell(n+1)}, \end{aligned}$$

and so for Equation (4.7) to hold it must be the case that  $\ell(n+1) - \ell(n) = O(1/\log n)$  and  $\ell(n+1) = O(n)$ . Indeed, it follows by Equation (4.4) that

$$\begin{aligned} \ell(n+1) - \ell(n) &= O\left(\frac{1}{\rho(n+1)}(\ln g(n+1) + \ln \ln \ln(n+1)) - \frac{1}{\rho(n)}(\ln g(n) + \ln \ln \ln n)\right) \\ &= O\left(\frac{1}{\log n}\right), \end{aligned}$$

the second equality holds per our assumption regarding  $g(n)$  and  $\rho(n)$ . By Equation (4.5), and by the assumptions  $g(n) \leq \log n$ ,  $\rho(n) \geq 1/(\ln \ln n)^h$ , it follows that  $\ell(n) = o(n)$ . We can thus conclude that Equation (4.7) holds and that  $C'$  is  $(n'_0, c', d')$ -dense for some constants.

After verifying that  $C'$  withstands its requirements, we can now apply Corollary 4.16. We get that there exists a code family  $C'' = \{(C'')^n\}$  over  $\mathbb{F}$ , in which every code is a good

LCC, with query complexity

$$\begin{aligned}
q_{\text{new}}(N) &= q'(N)^{O(g'(N) \ln \ln N)} \\
&= \left(q(\widehat{n}(N))^{\ell(\widehat{n}(N))}\right)^{O(\ell(\widehat{n}(N))g(\widehat{n}(N)) \ln \ln N)} \\
&= q(\widehat{n}(N))^{O(\ell(\widehat{n}(N))^2 g(\widehat{n}(N)) \ln \ln N)} \\
&\leq q(N)^{O(\ell(N)^2 g(N) \ln \ln N)} \\
&= q(N)^{O\left(\frac{1}{\rho(N)^2} (\ln g(N) + \ln \ln \ln N)^2 g(N) \ln \ln N\right)},
\end{aligned}$$

where the first inequality is justified by Equation (4.6), and the last equality is due to Equation (4.5). Thus,  $C''$  is a family with the desired query complexity, from which the claim follows.  $\square$

## 5 LDC are not LCC via random weighted tensor codes

In this section we prove Theorem 2.2. We show that there exist linear codes which are LDC but not LCC, in the following strong sense. What we prove is that not only are these codes LDC while not being LCC even for a weak requirement of very high query complexity and very low correction radius, moreover, this negative property that local correction with such parameters is impossible is maintained in any puncturing of the code. We will be able to show this to be the case because in the codes that we construct the uncorrectable coordinates are crucial for the distance of the code, and in particular for the LDC feature of the code, thus any attempt to remove them while keeping these properties, fails.

### 5.1 Preliminaries for this section

**Notation.** In what follows we will sometimes need to conveniently convert a pair of indices  $i_1 \in [m_1]$ ,  $i_2 \in [m_2]$  to an index  $i \in [m_1 m_2]$ , and so we set the following convention. Where  $m_1, m_2 \in \mathbb{N}$  are clear from context and  $i_1 \in [m_1]$ ,  $i_2 \in [m_2]$ , we denote by  $(i_1; i_2)$  the index  $(i_2 - 1)m_1 + i_1 \in [m_1 m_2]$ .

**Definition 5.1.** For a code  $C$  of length  $n$  over  $\mathbb{F}$ , we say that a coordinate  $j \in [n]$  is trivial (in  $C$ ) if for every  $c \in C$ ,  $c_j = 0$ .

We define the operation of puncturing of codes.

**Definition 5.2.** Let  $C$  be a code of length  $n$  and dimension  $k$  over  $\mathbb{F}$  and let  $J \subseteq [n]$ . For every codeword  $c \in C$ , we define the vector  $(y_1, \dots, y_n) \in \mathbb{F}^n$ , where  $y_j = c_j$  if  $j \notin J$  and  $y_j = 0$  otherwise, to be the  $J$ -puncturing of  $c$ , and we denote it by  $c_{\setminus J}$ . We define  $\{c_{\setminus J} \mid c \in C\}$  to be the  $J$ -punctured code  $C$ , and denote it by  $C_{\setminus J}$ . Note that  $C_{\setminus J}$  is indeed a code. Furthermore, given an encoding  $\text{Enc}$  of  $C$ , we define  $\text{Enc}_{\setminus J} : \mathbb{F}^k \rightarrow \mathbb{F}^n$  by  $\text{Enc}_{\setminus J}(x) = \text{Enc}(x)_{\setminus J}$  for all  $x \in \mathbb{F}^k$ .

We have the following easy claim regarding the puncturing and the dual operators.

**Claim 5.3.** Let  $J \subseteq [n]$ ,  $C$  a code of length  $n$ , and  $C_{\setminus J}$  its  $J$ -punctured code. Then, for every  $w \in (C_{\setminus J})^\perp$ ,  $w_{\setminus J} \in C^\perp$ .

*Proof.* Let  $w \in (C_{\setminus J})^\perp$ . We have that for all  $c \in C$ ,  $\langle c_{\setminus J}, w \rangle = 0$ . As  $\langle c_{\setminus J}, w \rangle = \langle c, w_{\setminus J} \rangle$ , we have that for all  $c \in C$ ,  $\langle c, w_{\setminus J} \rangle = 0$ , and so  $w_{\setminus J} \in C^\perp$ .  $\square$

## 5.2 A necessary condition for local correction

We start this section by stating a necessary condition for a coordinate of a code to be locally correctable. Using this condition we will be able to prove that some coordinates of a code are not locally correctable. The condition is a certain property of the dual code. We will call coordinates satisfying the property *dual correctable* coordinates.

**Claim 5.4.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$ ,  $j \in [n]$  and  $Q \subseteq [n]$  a set of size  $q$ , satisfying  $j \notin Q$ . If  $Q$  determines  $j$  in  $C$ , then there exists some  $w \in (C^\perp)_{\leq q+1}$  such that  $j \in \text{supp}(w) \subseteq Q \cup \{j\}$ .

*Proof.* As  $Q$  determines  $j$  in  $C$ , there exists a function  $f : \mathbb{F}^{|Q|} \rightarrow \mathbb{F}$  such that for every  $c \in C$ ,  $c_j = f(c_Q)$ . As  $C$  is a vector space, it readily follows that  $f$  can be taken to be a linear map. Therefore, there exists a vector  $w' \in \mathbb{F}^n$  such that for every  $c \in C$ ,  $\langle w', c \rangle = c_j$ , and  $\text{supp}(w') = Q$ . It follows that if we take  $w = w' - e_j$ , we have that for every  $c \in C$ ,  $\langle w, c \rangle = 0$ , and so  $w \in C^\perp$  and  $j \in \text{supp}(w) \subseteq Q \cup \{j\}$ , as required.  $\square$

**Definition 5.5.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$ , and let  $j \in [n]$ . We say that  $j$  is  $(q, \delta)$ -dual correctable coordinate in  $C$  if for  $m \geq \delta n/q$  there exist some  $w_1, \dots, w_m \in (C^\perp)_{\leq q+1}$ , with the following guarantee. For every  $i \in [m]$ ,  $j \in \text{supp}(w_i)$  and for every  $i, i' \in [m]$ ,  $i \neq i'$ ,  $\text{supp}(w_i) \cap \text{supp}(w_{i'}) = \{j\}$ .

**Claim 5.6.** Let  $C$  be a code of length  $n$  over  $\mathbb{F}$  and let  $j \in [n]$ . If  $j$  is a  $(q, \delta, \varepsilon)$ -correctable coordinate in  $C$  for  $\varepsilon < 1 - 1/|\mathbb{F}|$ , then  $j$  is a  $(q, \delta)$ -dual correctable coordinate.



*Proof.* First, note that there exists a set  $A_j = \{B_1, \dots, B_m\}$  for  $m \geq \delta n/q$  such that for every  $i \neq i'$ ,  $B_i \cap B_{i'} = \emptyset$ , and for every  $i \in [m]$ ,  $|B_i| \leq q$ ,  $j \notin B_i$ , and  $B_i$  determines  $j$  in  $C$ . This is exactly proven by the argument in the proof of Claim 4.7, and simply it follows by taking  $A_j$  to be the set  $A_j$  constructed in that proof. Note that while Claim 4.7 is stated for codes which are  $(q, \delta, \varepsilon)$ -LCC, the construction of the set  $A_j$  in the proof only uses the property that  $j$  is a  $(q, \delta, \varepsilon)$ -locally-correctable coordinate in  $C$ .

Secondly, for each  $i \in [m]$  we apply Claim 5.4 with respect to the set  $B_i$ , to conclude the existence of some  $w_i \in (C^\perp)_{\leq q+1}$  such that  $j \in \text{supp}(w_i) \subseteq B_i \cup \{j\}$ . It immediately follows that  $j$  is a  $(q, \delta)$ -dual correctable coordinate in  $C$ .  $\square$

### 5.3 Weighted tensors

We turn next to define an operation to which we call the *weighted tensor* of two codes and state several of its properties. The codes of Theorem 2.2 will be constructed using a weighted tensor. This operation gets two input codes (more precisely, two codes and respective encodings), and a matrix of non-zero entries, and results in a new code. To define the result of the operation, we will define a new encoding function which depends on the encodings of the two input codes and on the weight matrix. We will then take the resulted code to be the image of that encoding.

We thus begin by describing the encoding function of the *weighted tensor*.

**Inputs.** Let

- $\text{Enc}_1 : \mathbb{F}^{k_1} \rightarrow \mathbb{F}^{n_1}$  be a linear map.
- $\text{Enc}_2 : \mathbb{F}^{k_2} \rightarrow \mathbb{F}^{n_2}$  be a linear map.
- $B \in \mathbb{F}^{n_1 \times k_2}$  be a matrix with non-zero entries.

We define the following function  $\text{Enc} : \mathbb{F}^{k_1 k_2} \rightarrow \mathbb{F}^{n_1 n_2}$  that acts as follows on input  $x \in \mathbb{F}^{k_1 k_2}$ .

**Action of  $\text{Enc}$  on  $x$ .**

1. Identify  $x$  with a matrix  $X \in \mathbb{F}^{k_1 \times k_2}$  where for  $i_1 \in [k_1]$ ,  $i_2 \in [k_2]$ ,  $X_{i_1, i_2} = x_{(i_1, i_2)}$ .
2. Use  $\text{Enc}_1$  to encode each column of  $X$  and set  $X'$  to be the resulted matrix,  $X' \in \mathbb{F}^{n_1 \times k_2}$ .

3. For each  $j_1 \in [n_1], i_2 \in [k_2]$  multiply the element  $X'_{j_1, i_2}$  by  $B_{j_1, i_2}$  and set  $X''$  to be the resulted matrix.
4. Use  $\text{Enc}_2$  to encode each row of  $X''$  and set  $X'''$  to be the resulted matrix,  $X''' \in \mathbb{F}^{n_1 \times n_2}$ .
5. Output  $x' \in \mathbb{F}^{n_1 n_2}$  where for  $j_1 \in [n_1], j_2 \in [n_2]$ ,  $x'_{(j_1; j_2)} = X'''_{j_1, j_2}$ .

**Properties of Enc.** We turn to state a few properties of the function  $\text{Enc}$ .

**Claim 5.7.** *If  $\text{Enc}_1$  and  $\text{Enc}_2$  are injective then so is  $\text{Enc}$ .*

*Proof.* Follows trivially as  $B$  is a matrix with no zero entries. □

**Claim 5.8.** *Let  $A^1 \in \mathbb{F}^{n_1 \times k_1}$  and  $A^2 \in \mathbb{F}^{n_2 \times k_2}$  be the generating matrices of  $\text{Enc}_1$  and  $\text{Enc}_2$ , respectively. Then, for every  $x \in \mathbb{F}^{k_1 k_2}$ ,  $\text{Enc}(x) = Ax$ , where  $A \in \mathbb{F}^{n_1 n_2 \times k_1 k_2}$  is the matrix where for  $i_1 \in [k_1], i_2 \in [k_2], j_1 \in [n_1], j_2 \in [n_2]$ ,*

$$A_{(j_1; j_2), (i_1; i_2)} = A^1_{j_1, i_1} A^2_{j_2, i_2} B_{j_1, i_2}. \quad (5.1)$$

*In particular,  $\text{Enc}$  is a linear map.*

*Proof.* Let  $x \in \mathbb{F}^{k_1 k_2}$ , and let  $X', X'', X''', x'$  be as in the encoding described above. As  $\text{Enc}(x) = x'$ , we need to show that  $Ax = x'$ . Let  $j_1 \in [n_1], j_2 \in [n_2]$ . For every  $i_2 \in [k_2]$ , the  $i_2$ -nd column of  $X'$  is given by  $A^1 X e_{i_2}$  since  $X'$  is the result of applying  $\text{Enc}_1$  on each column of  $X$ . Therefore, for  $i_2 \in [k_2]$ , we have that

$$X''_{j_1, i_2} = (A^1 X e_{i_2})_{j_1} B_{j_1, i_2} = \sum_{i_1 \in [k_1]} A^1_{j_1, i_1} X_{i_1, i_2} B_{j_1, i_2}, \quad (5.2)$$

as  $X''$  is the result of multiplying  $X'$  and  $B$  entry-wise. For  $j_1 \in [n_1]$ , the  $j_1$ -st row of  $X'''$  is given by  $(A^2 (X'')^\top e_{j_1})^\top$ , since  $X'''$  is the result of applying  $\text{Enc}_2$  to each row of  $X''$ , and so

$$\begin{aligned} x'_{(j_1; j_2)} &= X'''_{j_1, j_2} \\ &= (A^2 (X'')^\top e_{j_1})_{j_2} \\ &= \sum_{i_2 \in [k_2]} A^2_{j_2, i_2} ((X'')^\top e_{j_1})_{i_2} \\ &= \sum_{i_2 \in [k_2]} A^2_{j_2, i_2} X''_{j_1, i_2}. \end{aligned}$$

Therefore, by Equation (5.2) and Equation (5.1),

$$\begin{aligned}
x'_{(j_1; j_2)} &= \sum_{i_2 \in [k_2]} A_{j_2, i_2}^2 \sum_{i_1 \in [k_1]} A_{j_1, i_1}^1 X_{i_1, i_2} B_{j_1, i_2} \\
&= \sum_{\substack{i_1 \in [k_1] \\ i_2 \in [k_2]}} A_{(j_1; j_2), (i_1; i_2)} X_{i_1, i_2} \\
&= (Ax)_{(j_1; j_2)}.
\end{aligned}$$

Thus  $Ax = x'$ , as required.  $\square$

**The weighted tensor operation.** We can now define the weighted tensor operation.

**Definition 5.9.** Let  $\text{Enc}_1, \text{Enc}_2, B$  and  $\text{Enc}$  be as above. Let  $C_1$  be a code of length  $n_1$  and dimension  $k_1$  over  $\mathbb{F}$  such that  $\text{Enc}_1$  is an encoding of it, and let  $C_2$  be a code of length  $n_2$  and dimension  $k_2$  over  $\mathbb{F}$  such that  $\text{Enc}_2$  is an encoding of it. Let  $C$  be the image of  $\text{Enc}$ . We define the  $B$ -weighted tensor of  $(C_1, \text{Enc}_1)$  and  $(C_2, \text{Enc}_2)$  to be the pair  $(C, \text{Enc})$ , and denote  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$ .

**Claim 5.10.** Let  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$ . Then  $C$  is a code of length  $n = n_1 n_2$  and dimension  $k = k_1 k_2$  over  $\mathbb{F}$ , and  $\text{Enc}$  is an encoding of it.

*Proof.* Follows immediately by the definition of  $\text{Enc}$ , and since  $\text{Enc}$  is injective by Claim 5.7, linear by Claim 5.8, and by the fact that  $C$  is defined to be its image.  $\square$

## 5.4 Local decodability of weighted tensors

In this part we show that the weighted tensor of two LDC is an LDC with comparable parameters, regardless of the weight matrix.

**Claim 5.11.** Let  $(C_1, \text{Enc}_1)$  be a  $(q_1, \delta_1, \varepsilon_1)$ -LDC, where  $C_1$  is a code of length  $n_1$  and dimension  $k_1$  over  $\mathbb{F}$ . Let  $(C_2, \text{Enc}_2)$  be a  $(q_2, \delta_2, \varepsilon_2)$ -LDC, where  $C_2$  is a code of length  $n_2$  and dimension  $k_2$  over  $\mathbb{F}$ , and let  $B \in \mathbb{F}^{n_1 \times k_2}$  be a matrix with no zero entries. Then,  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$  is a  $(q_1 q_2, \delta_1 \delta_2, 1 - (1 - \varepsilon_1)(1 - \varepsilon_2)^{q_1})$ -LDC.

*Proof.* Let  $\text{Dec}_1$  be a decoder promised by the fact that  $(C_1, \text{Enc}_1)$  is a  $(q_1, \delta_1, \varepsilon_1)$ -LDC and let  $\text{Dec}_2$  be a decoder promised by the fact that  $(C_2, \text{Enc}_2)$  is a  $(q_2, \delta_2, \varepsilon_2)$ -LDC. To show that  $(C, \text{Enc})$  is a  $(q_1 q_2, \delta_1 \delta_2, (1 - \varepsilon_1)(1 - \varepsilon_2)^{q_1})$ -LDC, we describe a decoder  $\text{Dec}$  for it. For every  $i = (i_1; i_2)$ ,  $i_1 \in [k_1]$ ,  $i_2 \in [k_2]$ ,  $\text{Dec}$  acts as follows on input  $i$  and oracle access to  $z \in \mathbb{F}^{n_1 n_2}$ .

1. Identify  $z$  with a matrix  $Z \in \mathbb{F}^{n_1 \times n_2}$  where for  $j_1 \in [n_1]$ ,  $j_2 \in [n_2]$ ,  $Z_{j_1, j_2} = z_{(j_1; j_2)}$ .
2. Simulate  $\text{Dec}_1(i_1)$ . Instead of giving  $\text{Dec}_1(i_1)$  direct oracle access to a word  $y \in \mathbb{F}^{n_1}$  do the following. For every index  $j_1 \in [n_1]$  that  $\text{Dec}_1(i_1)$  needs to query:
  - (a) Simulate  $\text{Dec}_2(i_2)$  with oracle access to the  $j_1$ -st row of  $Z$ ,  $Z_{j_1}$ .
  - (b) Divide the result of  $\text{Dec}_2(i_2)$  by  $B_{j_1, i_2} \in \mathbb{F}$  and feed it to  $\text{Dec}_1$ .<sup>13</sup>
3. Output the result of  $\text{Dec}_1(i_1)$ .

We turn to analyze the above decoder. First, it is immediate that given that  $\text{Dec}_1$  and  $\text{Dec}_2$  are non-adaptive, so is  $\text{Dec}$ . Secondly, it is also immediate that  $\text{Dec}$  makes at most  $q_1 q_2$  queries in any case. Thirdly, we need to show that the output of  $\text{Dec}(i)$  is correct with probability at least  $(1 - \varepsilon_1)(1 - \varepsilon_2)^{q_1}$ . Towards that, let  $z \in \mathbb{F}^{n_1 n_2}$  be such that  $\text{dist}(\text{Enc}(x), z) \leq \delta_1 \delta_2 n_1 n_2$  for some  $x \in \mathbb{F}^{k_1 k_2}$  and assume that  $\text{Dec}(i)$  is run with oracle access to  $z$ . By the bound on  $\text{dist}(\text{Enc}(x), z)$ , it follows that at most  $\delta_1 n_1$  rows of  $Z$  have more than  $\delta_2 n_2$  erroneous entries, i.e., entries  $(j_1, j_2)$  such that  $Z_{j_1, j_2} \neq \text{Enc}(x)_{(j_1; j_2)}$ . Let  $E \subseteq [n_1]$  be the set of indices of these “bad” rows.

Recall that in the definition of  $\text{Enc}$ ,  $x' = \text{Enc}(x)$  corresponds to a matrix  $X''$ , and the rows of  $X''$  are codewords of  $C_2$  as  $X''$  is the result of applying  $\text{Enc}_2$  on every row of  $X'$ . We thus have that for every  $j_1 \notin E$ , it holds that  $\text{Dec}_2(i_2)$ , when run with oracle access to  $Z_{j_1}$ , outputs  $X''_{j_1, i_2}$  with probability at least  $(1 - \varepsilon_2)$ . Further recall that the matrix  $X''$  is the result of multiplying, entry-wise, a matrix  $X'$  with the values of  $B$ , and the matrix  $X'$  is the result of applying  $\text{Enc}_1$  on each column of the matrix  $X$  that corresponds to  $x$ . Let  $p$  be the probability of the event that for every index  $j_1 \notin E$  which  $\text{Dec}(i_1)$  requests to query it is fed with  $X'_{j_1, i_2}$ . Conditioned on this event,  $\text{Dec}(i_1)$  outputs  $X_{i_1, i_2}$  with probability at least  $1 - \varepsilon_1$ .<sup>14</sup> It follows that in general,  $\text{Dec}(i_1)$  outputs  $X_{i_1, i_2}$  with probability at least  $p(1 - \varepsilon_1)$ .

It only remains to bound  $p$  from below. Clearly, for every index  $j_1 \notin E$  which  $\text{Dec}_1(i_1)$  requests to query, it is fed by  $\text{Dec}$  with  $X'_{j_1, i_2}$  if and only if  $\text{Dec}_2(i_2)$  with oracle access to  $Z_{j_1}$  outputs  $X''_{j_1, i_2}$ , as  $\text{Dec}$  divides that result by the same weight  $B_{j_1, i_2}$  which is used by the encoding to multiply  $X'_{j_1, i_2}$ . As mentioned, the probability for the output of  $\text{Dec}_2(i_2)$

<sup>13</sup>This can be thought of as giving  $\text{Dec}_1(i_1)$  oracle access to a “virtual” word, i.e., a word  $Y \in \mathbb{F}^{n_1}$  which is a random variable that satisfies  $Y_{j_1} = \text{Dec}_2^{Z_{j_1}}(i_2)/B_{j_1, i_2}$  for every index  $j_1 \in [n_1]$ .

<sup>14</sup>This holds as conditioned on the described event, the situation is equivalent to the case that  $\text{Dec}_1$  was given direct oracle access to a string  $Y \in \mathbb{F}^{n_1}$  which satisfies that for every  $j_1 \notin E$ ,  $Y_{j_1} = \text{Enc}_1(m)_{j_1}$  (in our case  $m$  is the  $i_2$ -nd column of  $X'$ ) and for  $j_1 \in E$ ,  $Y_{j_1}$  can have any value and may depend on the random choices of  $\text{Dec}_1$ . We have that  $\Pr[\text{Dec}_1^Y(i_1) = m_{i_1}] = \sum_{s \in \mathbb{F}^{|E|}} \Pr[Y_E = s] \Pr[\text{Dec}_1^Y(i_1) = m_{i_1} \mid Y_E = s] \geq \sum_{s \in \mathbb{F}^{|E|}} \Pr[Y_E = s](1 - \varepsilon_1) = 1 - \varepsilon_1$ .

to satisfy this is at least  $1 - \varepsilon_2$ . Thus, as  $\text{Dec}_1$  makes at most  $q_1$  queries and as different calls to  $\text{Dec}_2$  are independent, the probability  $p$  that all the queries made by  $\text{Dec}_1$  to indices not in  $E$  are met with the correct values of  $X'$  satisfies  $p \geq (1 - \varepsilon_2)^{q_1}$ , and so it follows that  $\text{Dec}_1$  outputs  $X_{i_1, i_2}$  with probability at least  $(1 - \varepsilon_1)(1 - \varepsilon_2)^{q_1}$ . As the output of  $\text{Dec}$  is equal to the result of the simulation of  $\text{Dec}_1(i_1)$  and as  $X_{i_1, i_2} = x_{(i_1, i_2)} = x_i$ , this shows that  $\text{Dec}(i)$  is correct with the stated probability and the claim follows.  $\square$

## 5.5 Local correctability of random weighted tensors

In this part we show that the weighted tensor of two codes, when performed with a randomly chosen weight matrix is, with high probability, not locally correctable. In particular, we show that a subset of the coordinates cannot be locally corrected even with a small correction radius guarantee, and cannot be removed from the code either if its decodability is to be preserved.

Let  $(C_1, \text{Enc}_1)$  be a  $(q_1, \delta_1, \varepsilon_1)$ -LDC for a code  $C_1$  of length  $n_1$  and dimension  $k_1$  over  $\mathbb{F}$  and  $\varepsilon_1 < 1 - 1/|\mathbb{F}|$ . Let  $(C_2, \text{Enc}_2)$  be a  $(q_2, \delta_2, \varepsilon_2)$ -LDC for a code  $C_2$  of length  $n_2$  and dimension  $k_2$  over  $\mathbb{F}$  and  $\varepsilon_2 < 1 - 1/|\mathbb{F}|$ . We assume that  $C_1$  and  $C_2$  are free of trivial coordinates.<sup>15</sup> Let  $B \in \mathbb{F}^{n_1 \times k_2}$  a random, uniformly and independently sampled, matrix of non-zero weights. Let  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$  be the  $B$ -weighted tensor of the two codes, and denote by  $n$  the length of  $C$  and by  $k$  its dimension. By Claim 5.10,  $n = n_1 n_2$  and  $k = k_1 k_2$ .

We will need the following definition.

**Definition 5.12.** For  $j_1^* \in [n_1]$ ,  $j_2^* \in [n_2]$  and  $\tilde{q} \in \mathbb{N}$ , we say that  $(j_1^*, j_2^*)$  is  $\tilde{q}$ -possibly correctable if there exists  $w \in (C^\perp)_{\leq \tilde{q}+1}$  such that  $w_{(j_1^*, j_2^*)} \neq 0$  and for every  $j_2 \in [n_2] \setminus \{j_2^*\}$ ,  $w_{(j_1^*, j_2)} = 0$ .

**Claim 5.13.** For  $\tilde{q} \in \mathbb{N}$  and  $j \in [n]$ , if  $j = (j_1^*, j_2^*)$  and  $(j_1^*, j_2^*)$  is not  $\tilde{q}$ -possibly correctable, then for every  $\delta \geq \tilde{q}/n_1$ ,  $j$  is not a  $(\tilde{q}, \delta)$ -dual correctable coordinate of  $C$ .

*Proof.* Let  $\tilde{q} \in \mathbb{N}$  and  $j \in [n]$  be such that  $j = (j_1^*, j_2^*)$  and  $(j_1^*, j_2^*)$  is not  $\tilde{q}$ -possibly correctable. Assume towards contradiction that  $j$  is a  $(\tilde{q}, \delta)$ -dual correctable coordinate of  $C$ , for  $\delta \geq \tilde{q}/n_1$ . Then, there exist  $w_1, \dots, w_m \in (C^\perp)_{\leq \tilde{q}+1}$ ,  $m \geq \delta n / \tilde{q} \geq n_2$ , such that for every  $i \in [m]$ ,  $j \in \text{supp}(w_i)$  and for every  $i, i' \in [m]$ ,  $i \neq i'$ ,  $\text{supp}(w_i) \cap \text{supp}(w_{i'}) = \{j\}$ . Thus, the sets  $\text{supp}(w_1) \setminus \{j\}, \dots, \text{supp}(w_m) \setminus \{j\}$  are disjoint. Set  $R = \{(j_1^*, r) \mid r \in [n_2] \setminus \{j_2^*\}\}$ . Since  $j$  is not  $\tilde{q}$ -possibly correctable, the sets  $R \cap \text{supp}(w_1) \setminus \{j\}, \dots, R \cap \text{supp}(w_m) \setminus \{j\} \subseteq R$  are all disjoint and non-empty. Thus,  $m \leq |R| = n_2 - 1$ , in contradiction.  $\square$

<sup>15</sup>This assumption is for simplicity, clearly any trivial coordinate can be removed from a code and this would only improve the parameters of the code.

We will now describe a set of coordinates which we will argue are not  $\tilde{q}$ -possibly correctable, with high probability over the choice of  $B$ . We show this to be the case for every coordinate  $(j_1; j_2)$  such that  $j_1 \in [n_1]$ , and such that  $j_2$  is in some subset  $I \subseteq [n_2]$ . Let  $A^1$ ,  $A^2$  and  $A$  denote the matrices that correspond to  $\mathbf{Enc}_1$ ,  $\mathbf{Enc}_2$  and  $\mathbf{Enc}$ , respectively. Let  $t < k_2$  be a parameter. We write  $t = \alpha k_2$  for  $\alpha < 1$ . Define  $I \subseteq [n_2]$  to be the set of indices of  $C_2$  which in  $\mathbf{Enc}_2$  depend on at least  $t$  message bits, i.e.,  $I = \{j_2 \in [n_2] \mid |A_{j_2}^2| \geq t\}$ , where  $|A_{j_2}^2|$  is the weight of the  $j_2$ -th row of  $A^2$ . We set

$$J = [n_1] \times I \quad (5.3)$$

and

$$\bar{J} = \{(j_1; j_2) \mid (j_1, j_2) \in J\},$$

the corresponding set of coordinates in  $[n]$ . We will now state a few claims which imply that, with high probability, over the choice of  $B$ ,  $\bar{J}$  is a set of coordinates that are not correctable in  $C$ .

**Claim 5.14.** *For  $(j_1^*, j_2^*) \in J$  and  $\tilde{q} \in \mathbb{N}$ ,  $(j_1^*, j_2^*)$  is  $\tilde{q}$ -possibly correctable with probability at most  $\binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$ , over the choice of  $B$ .*

*Proof.* Let  $(j_1^*, j_2^*) \in J$  and  $\tilde{q} \in \mathbb{N}$ . Note that  $(j_1^*, j_2^*)$  is  $\tilde{q}$ -possibly correctable if there exists some  $w \in \mathbb{F}_{\leq \tilde{q}+1}^{n_1 n_2}$ , such that  $w_{(j_1^*; j_2^*)} \neq 0$  and  $w_{(j_1^*; j_2)} = 0$  for every  $j_2 \in [n_2] \setminus j_2^*$ ,  $w \in C^\perp$ . Further note that in such a case, we may assume  $w_{(j_1^*; j_2^*)} = 1$  without loss of generality. We therefore fix  $w \in \mathbb{F}_{\leq \tilde{q}+1}^{n_1 n_2}$  with the aforementioned properties and consider the probability that  $w \in C^\perp$ . Note that  $w \in C^\perp$  if and only if  $w^\top A = 0$ , and that  $w^\top A = 0$  if and only if for every  $i_1 \in [k_1]$  and  $i_2 \in [k_2]$ ,  $(w^\top A)_{(i_1; i_2)} = 0$ . Hence, we are interested in the probability, over the choice of  $B$ , that  $(w^\top A)_{(i_1; i_2)} = 0$  for specific  $i_1, i_2$ . We fix  $i_1 \in [k_1]$  such that  $A_{j_1^*, i_1}^1 \neq 0$ . Note that such  $i_1$  exists as  $C_1$  is assumed not to have trivial coordinates. We fix  $i_2 \in [k_2]$  such that  $A_{j_2^*, i_2}^2 \neq 0$ . There are at least  $t$  possible choices of such  $i_2$ , since  $j_2^* \in I$ . We have that

$$\begin{aligned} (w^\top A)_{(i_1; i_2)} &= \sum_{\substack{j_1 \in [n_1] \\ j_2 \in [n_2]}} w_{(j_1; j_2)} A_{(j_1; j_2), (i_1; i_2)} \\ &= \sum_{\substack{j_1 \in [n_1] \\ j_2 \in [n_2]}} w_{(j_1; j_2)} A_{j_1, i_1}^1 A_{j_2, i_2}^2 B_{j_1, i_2} \\ &= w_{(j_1^*; j_2^*)} A_{j_1^*, i_1}^1 A_{j_2^*, i_2}^2 B_{j_1^*, i_2} + \sum_{\substack{j_1 \in [n_1] \setminus \{j_1^*\} \\ j_2 \in [n_2]}} w_{(j_1; j_2)} A_{j_1, i_1}^1 A_{j_2, i_2}^2 B_{j_1, i_2}, \end{aligned}$$

where the second equality is due to Claim 5.8, and the last equality is per our assumption on  $w$ . Also by that assumption, and by the choice of  $i_1, i_2$ , we have that

$w_{(j_1^*, j_2^*)} A_{j_1^*, i_1}^1 A_{j_2^*, i_2}^2 \neq 0$ . Further note that  $B_{j_1^*, i_2} \in \mathbb{F} \setminus \{0\}$  is uniformly chosen and is independent of the disjoint distribution of  $\{B_{j_1, i_2} \mid j_1 \in [n_1] \setminus \{j_1^*\}\}$ , and thus

$$\Pr_B [(w^\top A)_{(i_1; i_2)} = 0] \leq \frac{1}{|\mathbb{F}| - 1}.$$

Finally, note that this holds for all  $i_2$  for which  $A_{j_2^*, i_2}^2 \neq 0$  – there are at least  $t$  such, and since the weights that correspond to different  $i_2$ 's are chosen independently, we get that

$$\Pr_B [w^\top A = 0] \leq \left( \frac{1}{|\mathbb{F}| - 1} \right)^t.$$

By taking the union bound over all possible choices of  $w \in \mathbb{F}_{\leq \tilde{q}+1}^{n_1 n_2}$  such that  $w_{(j_1^*, j_2^*)} = 1$ , we get that the probability that  $(j_1^*, j_2^*)$  is  $\tilde{q}$ -possibly correctable is at most  $\binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$ , as required.  $\square$

**Claim 5.15.** *With probability at least  $1 - n_1 n_2 \binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$ , over the choice of  $B$ , for all  $(j_1^*, j_2^*) \in J$ ,  $(j_1^*, j_2^*)$  is not  $\tilde{q}$ -possibly correctable.*

*Proof.* The proof follows from Claim 5.14 by taking the union bound over all possibilities for  $(j_1^*, j_2^*) \in J$ .  $\square$

We can now conclude that, with high probability over  $B$ , the coordinates corresponding to  $J$ , as defined by Equation (5.3), are not locally correctable. From that it will easily follow that any puncturing that leaves out such a coordinate (to remain in the code), remains not an LCC, as we have in the following claim.

**Claim 5.16.** *For  $\tilde{q} \in \mathbb{N}$ , with probability at least  $1 - n_1 n_2 \binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$ , over the choice of  $B, C$  satisfies the following. Every  $j \in \bar{J}$ ,  $j$  is not  $(\tilde{q}, \delta, \varepsilon)$ -locally correctable in  $C$ , for any  $\delta \geq \tilde{q}/n_1$  and  $\varepsilon < 1 - 1/|\mathbb{F}|$ .*

*Proof.* Let  $\tilde{q} \in \mathbb{N}$ . By Claim 5.15, with probability at least  $1 - n_1 n_2 \binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$ , it is the case that every  $(j_1^*, j_2^*) \in J$  is not  $\tilde{q}$ -possibly correctable. Assume that this is indeed the case. Let  $j \in \bar{J}$  be such that  $j = (j_1; j_2)$  for  $(j_1, j_2) \in J$  and  $j \notin J'$ . Per our assumption,  $(j_1, j_2)$  is not  $\tilde{q}$ -possibly correctable. Assume towards a contradiction that for  $\delta \geq \tilde{q}/n_1$  and  $\varepsilon < 1 - 1/|\mathbb{F}|$ ,  $j$  is  $(\tilde{q}, \delta, \varepsilon)$ -locally correctable coordinate in  $C$ . By Claim 5.6, it follows that  $j$  is a  $(\tilde{q}, \delta)$ -dual correctable coordinate of  $C$ . As  $(j_1, j_2)$  is not  $\tilde{q}$ -possibly correctable in  $C$ , by Claim 5.13, this is a contradiction.  $\square$

In the next claim we show that if the coordinates corresponding to  $J$  are removed from the code, the resulted code is not an LDC (and in particular, not an LCC).

**Claim 5.17.** *Let  $\Delta$  be the (non-local) relative distance of  $C_{\setminus \bar{J}}$ . Then  $\Delta < t/k_2$ .*

*Proof.* Recall that  $J = [n_1] \times I$ , where  $I \subseteq [n_2]$  is the set of coordinates of  $C_2$  which depend on at least  $t$  message bits in the encoding  $\text{Enc}_2$ , and that  $\bar{J}$  is the corresponding subset of  $[n]$ . One can verify that

$$(C_{\setminus \bar{J}}, \text{Enc}_{\setminus \bar{J}}) = (C_1, \text{Enc}_1) \otimes_B ((C_2)_{\setminus I}, (\text{Enc}_2)_{\setminus I}).$$

Let  $\tilde{A}^2$  denote the matrix that corresponds to  $(\text{Enc}_2)_{\setminus I}$ , and note that  $\tilde{A}^2$  is achieved by setting each row  $j_2 \in I$  of  $A^2$  to a zero row. By the definition of  $I$ , it follows that  $\tilde{A}^2$  has less than  $n_2 t$  non-zero entries. Let  $\tilde{A}$  denote the matrix that corresponds to  $\text{Enc}_{\setminus \bar{J}}$ . By Claim 5.8 we have that

$$\tilde{A}_{(j_1; j_2), (i_1; i_2)} = A_{j_1, i_1}^1 \tilde{A}_{j_2, i_2}^2 B_{j_1, i_2}$$

for every  $j_1 \in [n_1]$ ,  $j_2 \in [n_2]$ ,  $i_1 \in [k_1]$ ,  $i_2 \in [k_2]$ . From that, it follows that  $\tilde{A}$  has less than  $n_1 k_1 n_2 t$  non-zero entries, and therefore there exists a column of  $\tilde{A}$ ,  $v \in \mathbb{F}^{n_1 n_2}$ , which has less than  $n_1 n_2 t / k_2$  non-zero entries. As we have that  $v \in C_{\setminus \bar{J}}$ , it follows that  $\Delta < t / k_2$ .  $\square$

We have the following claim to conclude this part.

**Claim 5.18.** *Let  $(C_1, \text{Enc}_1)$  be a  $(q_1, \delta_1, \varepsilon_1)$ -LDC of length  $n_1$  and dimension  $k_1$  over  $\mathbb{F}$ , and let  $(C_2, \text{Enc}_2)$  be a  $(q_2, \delta_2, \varepsilon_2)$ -LDC of length  $n_2$  and dimension  $k_2$  over  $\mathbb{F}$ . Assume that  $C_1$  and  $C_2$  have no non-trivial coordinates. Let  $B \in \mathbb{F}^{n_1 \times k_2}$  be a random matrix of non-zero weights, chosen uniformly and independently, and let  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$ . For every  $t < k_2$  and  $\tilde{q}, \tilde{q}' \in \mathbb{N}$ ,  $\delta \geq \tilde{q} / n_1$ ,  $\delta' \geq t / k_2$  and  $\varepsilon < 1 - 1 / |\mathbb{F}|$ , with probability at least  $1 - n_1 n_2 \binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t$  over the choice of  $B$ ,  $C$  satisfies the following. There exists a set  $\bar{J} \subseteq [n]$  such that every  $j \in \bar{J}$  is not  $(\tilde{q}, \delta, \varepsilon)$ -locally correctable in  $C$ . Further, the relative (non-local) distance of  $C_{\setminus \bar{J}}$  is less than  $t / k_2$ .*

*Proof.* The proof follows by Claim 5.17 and Claim 5.16.  $\square$

## 5.6 Deriving the theorem

We are now ready to derive the main result of this section.

**Theorem 5.19.** *Let  $(C_0, \text{Enc}_0)$  be a  $(q_0, \delta_0, \varepsilon_0)$ -LDC for a code  $C_0$  of dimension  $k_0$  and length  $n_0$  over  $\mathbb{F}$  for  $|\mathbb{F}| > 2$ , such that  $\varepsilon_0 < 1 - 1 / |\mathbb{F}|$ ,  $k_0^{1/2} > 10 \log n_0$ , and assume that  $C_0$  has no trivial coordinates. Then, there exists a  $(q_0^2, \delta_0^2, 1 - (1 - \varepsilon_0)^{q_0 + 1})$ -LDC<sup>16</sup>  $(C, \text{Enc})$  for a code  $C$  of dimension  $k = k_0^2$  and length  $n = n_0^2$  over  $\mathbb{F}$  satisfying the following property.*

<sup>16</sup>Note that if  $(C_0, \text{Enc}_0)$  has error parameter  $\varepsilon_0 \leq 1/2 - \Omega(1)$  then the error parameter of  $(C, \text{Enc})$  can be made to match it by first reducing the error of  $(C_0, \text{Enc}_0)$  through repetitions. For example, if  $\varepsilon_0 \leq 1/3$ , a decoder for  $(C_0, \text{Enc}_0)$  which makes  $\ell = \Theta(\log q_0)$  simulations of Dec and outputs the majority vote can be used to show that  $(C, \text{Enc})$  is a  $(O(q_0 \log q_0)^2, \delta_0^2, 1/3)$ -LDC.



There exists a set  $J \subseteq [n]$  of coordinates such that every  $j \in J$ ,  $j$  is not  $(k^{1/4}, k^{1/4}/n^{1/2}, \varepsilon)$ -locally correctable in  $C$ , for any  $\varepsilon < 1 - 1/|\mathbb{F}|$ . Moreover, the relative distance of  $C_{\setminus J}$  is less than  $5 \log(n)/k^{1/4}$  (in particular for any  $\tilde{q} \in \mathbb{N}$  and  $\varepsilon < 1 - 1/|\mathbb{F}|$ ,  $C_{\setminus J}$  is not a  $(\tilde{q}, 5 \log(n)/k^{1/4}, \varepsilon)$ -LDC).<sup>17</sup>

*Proof.* Let  $(C_0, \text{Enc}_0)$  which is a  $(q_0, \delta_0, \varepsilon_0)$ -LDC for a code  $C_0$  of dimension  $k_0$  and length  $n_0$  over  $\mathbb{F}$ , such that  $\varepsilon_0 < 1 - 1/|\mathbb{F}|$ . Set  $C_1 = C_2 = C_0$ ,  $\text{Enc}_1 = \text{Enc}_2 = \text{Enc}_0$ , and for convenience also set  $n_1 = n_2 = n_0$  and  $k_1 = k_2 = k_0$ . Let  $B \in \mathbb{F}^{n_1 \times n_2}$  be a matrix of non-zero weights, sampled uniformly and independently at random, and set  $(C, \text{Enc}) = (C_1, \text{Enc}_1) \otimes_B (C_2, \text{Enc}_2)$  to be the  $B$ -weighted tensor of  $C_1$  and  $C_2$ , and denote its dimension by  $k$  and its length by  $n$ . By Claim 5.10,  $k = k_0^2$  and  $n = n_0^2$ . By Claim 5.11,  $(C, \text{Enc})$  is a  $(q, \delta, \varepsilon)$ -LDC for  $q = q_0^2$ ,  $\delta = \delta_0^2$  and  $\varepsilon = (1 - \varepsilon_0)^{q_0+1}$ . Thus, indeed  $(C, \text{Enc})$  is an LDC with the claimed properties, for any sampled  $B$ .

It remains to show that  $C$  satisfies the claimed negative LCC and LDC properties. Towards this end, set  $\tilde{q} = \sqrt{k_0} = k^{1/4}$  and  $t = 10\tilde{q} \log n_0$ . Note that  $t < k_0$  as by assumption  $k_0^{1/2} > 10 \log n_0$ . Further note that with that choice of  $t$  we have that

$$1 - n_1 n_2 \binom{n_1 n_2}{\tilde{q}} |\mathbb{F}|^{\tilde{q}} / (|\mathbb{F}| - 1)^t > 0.$$

Therefore, by Claim 5.18, with a probability greater than zero, there exists a set  $J \subseteq [n]$ , satisfying the following. For every

$$\delta \geq \tilde{q}/n_0 = \frac{k^{1/4}}{n_0} = \frac{k^{1/4}}{n^{1/2}},$$

$\varepsilon < 1 - 1/|\mathbb{F}|$  and  $j \in J$ ,  $j$  is not a  $(\tilde{q}, \delta, \varepsilon)$ -locally correctable coordinate of  $C$ . Moreover, the relative distance of  $C_{\setminus J}$  is less than

$$\frac{t}{k_0} = \frac{10\tilde{q} \log n_0}{k_0} = \frac{5 \log n}{k^{1/4}}.$$

By Fact 3.11, this implies that for every encoding  $\text{Enc}'$  of  $C_{\setminus J}$ ,  $\varepsilon < 1 - 1/|\mathbb{F}|$ , and  $\tilde{q} \in \mathbb{N}$ ,  $(C_{\setminus J}, \text{Enc}')$  is not a  $(\tilde{q}, 5 \log n/k^{1/4}, \varepsilon)$ -LDC.  $\square$

<sup>17</sup>Clearly the fact that there exist such a set  $J$  implies that no puncturing of the code at a set  $J' \subsetneq [n]$  can make it a  $(k^{1/4}, \max(k^{1/4}/n^{1/2}, (5 \log n)/k^{1/4}), \varepsilon)$ -LCC. Indeed, if  $J' \subseteq J$  then  $C_{\setminus J'}$  is not a  $(k^{1/4}, (5 \log n)/k^{1/4}, \varepsilon)$ -LDC, and in particular not an  $(k^{1/4}, \max(k^{1/4}/n^{1/2}, (5 \log n)/k^{1/4}), \varepsilon)$ -LCC, and if  $J' \not\subseteq J$  then  $C_{\setminus J'}$  is not a  $(k^{1/4}, k^{1/4}/n^{1/2}, \varepsilon)$ -LCC.

## 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.
- [BDSS11] Arnab Bhattacharyya, Zeev Dvir, Amir Shpilka, and Shubhangi Saraf. Tight lower bounds for 2-query lccs over finite fields. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 638–647. IEEE, 2011.
- [BDYW11] Boaz Barak, Zeev Dvir, Amir Yehudayoff, and Avi Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 519–528, 2011.
- [BGT16] Arnab Bhattacharyya, Sivakanth Gopi, and Avishay Tal. Lower bounds for 2-query lccs over large alphabet. *arXiv preprint arXiv:1611.06980*, 2016.
- [CY21] Gil Cohen and Tal Yankovitz. Rate amplification and query-efficient distance amplification for linear lcc and ldc. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.
- [DGY11] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011.
- [Din06] Irit Dinur. The PCP theorem by gap amplification. In *Proc. 38th ACM Symp. on Theory of Computing*, pages 241–250, 2006.
- [Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM Journal on Computing*, 41(6):1694–1703, 2012.
- [Gil52] Edgar N Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.
- [GKO<sup>+</sup>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.
- [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.
- [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.
- [KV10] Tali Kaufman and Michael Viderman. Locally testable vs. locally decodable codes. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 670–682. Springer, Berlin, 2010.
- [TS17] A. Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251. ACM, 2017.
- [Var57] R.R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akad. Nauk, SSSR*, 117:739–741, 1957.
- [Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM (JACM)*, 55(1):1–16, 2008.
- [ZD] Kalina Petrova Zeev Dvir. Lecture 1: Introduction. Lecture notes: <https://www.cs.princeton.edu/~zdvir/LDCnotes/LDC1.pdf>, year=2016,.

## A From smooth LCC to good LCC

In this part we provide a proof for Lemma 4.15.

**Theorem A.1** (The Gilbert-Varshamov bound, [Gil52, Var57]). *For any  $n \in \mathbb{N}$ , a field  $\mathbb{F}$  of size  $q$ , and  $0 \leq \delta \leq 1 - 1/q$ , there exists a code of length  $n$  over  $\mathbb{F}$  with relative distance at least  $\delta$  and rate  $r \geq 1 - H_q(\delta) - g(n)$ , where  $g(n) = 2/n$ .*

**Definition A.2.** *A linear subspace  $L \subseteq \mathbb{F}^n$  is called a  $(q, \delta, \alpha)$ -local-amplifier if there exists a deterministic procedure  $\text{Cor} : [n] \rightarrow \mathbb{F}$  that is given oracle access to  $z \in \mathbb{F}^n$  and has the following guarantee. For every  $y \in L$  and  $z \in \mathbb{F}^n$  such that  $\text{dist}(z, y) \leq \delta n$ ,  $\text{Cor}(i)$  outputs  $y_i$  when given oracle access to  $z$ , for at least  $\alpha$ -fraction of the indices  $i \in [n]$ . Furthermore,  $\text{Cor}$  always makes at most  $q$  queries to  $z$ .*

**Claim A.3.** *For every  $n \in \mathbb{N}$ ,  $\mathbb{F}$  a field, and  $\delta, \alpha \in (0, 1)$  such that  $\delta \leq 1/25$ , there exists a linear subspace  $L \subseteq \mathbb{F}^n$  which is a  $(q, \delta, \alpha)$ -local-amplifier for  $q = 25/(\delta(1 - \alpha)^2)$ , such that  $\dim L \geq (1 - 2H_{|\mathbb{F}|}(5\sqrt{\delta}) - 4\sqrt{\delta}(1 - \alpha)/5)n$ .*

*Proof.* We prove the existence of such a subspace with a probabilistic argument. Set  $d = 5/((1 - \alpha)\sqrt{\delta})$ , and let  $C_d \subseteq \mathbb{F}^d$  be a code of length  $d$ , with relative distance  $\Delta = 5\sqrt{\delta}$  and rate  $r \geq 1 - H_{|\mathbb{F}|}(5\sqrt{\delta}) - g(d)$  where  $g$  is as in Theorem A.1. The existence of such a code  $C_d$  follows from Theorem A.1. Let  $\pi$  be a partition of  $[n]$  into  $n/d$  blocks  $P^1, \dots, P^{n/d}$  of size  $d$ , chosen uniformly at random. Further let  $\pi'$  be the fixed partition of  $[n]$  into  $n/d$  subsequent blocks of size  $d$ :  $\pi' = \{(1, \dots, d), \dots, (n - d + 1, \dots, n)\}$ . We argue that the following subspace  $L = \{y \in \mathbb{F}^n \mid \text{for every } P \in \pi, P' \in \pi', y_P \in C_d \text{ and } y_{P'} \in C_d\}$  is of the claimed properties, with probability greater than 0. Note that indeed, as  $L$  is defined by at most  $2(n/d)(1 - r)d$  constraints,

$$\begin{aligned} \dim L &\geq (2r - 1)n = (1 - 2H_{|\mathbb{F}|}(5\sqrt{\delta}) - 2g(d))n \\ &= (1 - 2H_{|\mathbb{F}|}(5\sqrt{\delta}) - 4/d)n \\ &= \left(1 - 2H_{|\mathbb{F}|}(5\sqrt{\delta}) - \frac{4}{5}\sqrt{\delta}(1 - \alpha)\right)n \end{aligned}$$

To show that  $L$  is a  $(q, \delta, \alpha)$ -local-amplifier, we need to describe a corrector  $\text{Cor}$  for it, and towards that we first set up some notation. For every  $i \in [n]$ , let  $P'_i$  denote the part of  $\pi'$  which satisfies  $i \in P'$  and let  $r'_i$  denote the index of  $i$  in  $P'$ . Similarly, for  $j \in [n]$ , let  $P_j$  denote the part of  $\pi$  which satisfies  $j \in P_j$  and let  $r_j$  denote the index of  $j$  in  $P_j$ . On input  $i \in [n]$ , and oracle access to  $z \in \mathbb{F}^n$ ,  $\text{Cor}(i)$  acts as follows. For every  $j \in P'_i = \{j_1^i, \dots, j_d^i\}$ ,  $\text{Cor}(i)$  queries  $z_{P_j}$ , finds a word  $c_j \in C_d$  closest to  $z_{P_j} \in \mathbb{F}^d$ , and sets  $t_j = (z_{P_j})_{r_j}$ ;  $\text{Cor}(i)$  then finds a word  $c'_i \in C_d$  closest to  $(t_{j_1^i}, \dots, t_{j_d^i}) \in \mathbb{F}^d$ , and outputs  $(c'_i)_{r'_i}$ .

We inspect the described procedure  $\text{Cor}$ . Note first that the number of queries  $\text{Cor}$  makes is exactly  $q = d^2 = 25/((1 - \alpha)^2\delta)$ , as required. Secondly, it is immediate that

with access to  $z$  which satisfies  $\text{dist}(z, y) \leq \delta n$ , the output of  $\text{Cor}(i)$  is equal to  $y_i$  if  $\text{dist}((t_{j_1^i}, \dots, t_{j_d^i}), y_{P_i'}) < \frac{1}{2}\Delta d$ ; this, in turn, holds if we have that for less than  $\Delta d/2$  indices  $j \in P_i'$ ,  $\text{dist}(z_{P_j}, y_{P_j}) \geq \Delta d/2$ . Further note that for every  $y \in L$  and  $z \in \mathbb{F}^n$  such that  $\text{dist}(z, y) \leq \delta n$ , it is immediate that for at most a  $\delta/(\Delta/2)$  fraction of the parts  $P^1, \dots, P^{n/d}$  of  $\pi$ , it holds that  $\text{dist}(z_{P^i}, y_{P^i}) \geq \Delta d/2$ . Therefore,  $\text{Cor}(i)$  always succeeds on at least an  $\alpha$ -fraction of the indices  $i \in [n]$ , if it is the case that the following property holds: for every set  $I \subseteq [n/d]$  of “bad” parts indices (among  $P^1, \dots, P^{n/d}$ ),

$$|I| \leq \frac{\delta}{\frac{1}{2}\Delta} \cdot \frac{n}{d}, \quad (\text{A.1})$$

we have that for less than  $(1 - \alpha)n$  indices  $i \in [n]$ , at least  $\Delta d/2$  of the indices  $j \in P_i'$  satisfy  $j \in P^t$  for  $t \in I$ . We denote by  $p$  the probability, over the choice of  $\pi$ , that the requirement is not met, and we wish to show that it is less than 1.

We thus turn to bound  $p$ . We first fix some  $I \subseteq [n/d]$  satisfying Equation (A.1). For any  $D \subseteq [n]$ , the probability that for all  $j \in D$ ,  $j \in P^t$  for  $t \in I$ , is at most  $(2\delta/\Delta)^{|D|}$ . We have that for every subset of parts  $B \subseteq \pi'$ , the probability that for all  $P' \in B$ , at least  $\Delta d/2$  of the indices  $j \in P'$  satisfy  $j \in P^t$  for  $t \in I$ , by a union bound over the possible subsets of size  $\Delta d/2$  of each  $P' \in B$ , is at most

$$\binom{d}{\frac{1}{2}\Delta d}^{|B|} \left(\frac{2\delta}{\Delta}\right)^{|B|\Delta d/2}.$$

Again taking a union bound, this time over the possible subsets  $B \subseteq \pi'$ , of size  $(1 - \alpha)n/d$ , the probability that there exists such a set  $B$  of size at least  $(1 - \alpha)n/d$  is at most

$$\binom{n}{(1 - \alpha)n/d} \binom{d}{\frac{1}{2}\Delta d}^{(1 - \alpha)n/d} \left(\frac{2\delta}{\Delta}\right)^{((1 - \alpha)n/d)\frac{1}{2}\Delta d}.$$

By taking another union bound, over the possible choices of  $I$ , we can bound the probability that for some set  $I \subseteq [n/k]$  satisfying Equation (A.1), there exists such a set  $B$  of size at least  $(1 - \alpha)n/d$ , and get that

$$p \leq \binom{n/d}{(2\delta/\Delta)n/d} \binom{n/d}{(1 - \alpha)n/d} \binom{d}{\frac{1}{2}\Delta d}^{(1 - \alpha)n/d} \left(\frac{2\delta}{\Delta}\right)^{((1 - \alpha)n/d)\Delta d/2}.$$

One can verify that when plugging  $\Delta = 5\sqrt{\delta}$  and  $d = 5/(\sqrt{\delta}(1 - \alpha))$  results in that right hand side is indeed smaller than one, as required.  $\square$

We the claim, we can now prove Lemma 4.15.

*Proof for Lemma 4.15.* Let  $C^n$  be a code of the family  $C$ . We show that for every such  $C^n$ , there exists a code  $(C')^n$  with the desired properties, and we will finally take the family  $C'$  to be  $\{(C')^n\}$ . By Claim 4.2 used with  $\varepsilon = 1/3$ ,  $C^n$  is a  $(q(n), \delta(n), 1/3)$ -LCC for  $\delta(n) = 1/(3n\tau(n))$ . Let  $L \subseteq \mathbb{F}^n$  be a  $(q''(n), \delta'(n), \alpha(n))$ -local-amplifier, for  $\alpha(n) = 1 - \delta(n)$  where  $\delta'(n)$  is defined to be the maximal value in  $(0, 1/25]$  which satisfies

$$2H_{\mathbb{F}}(5\sqrt{\delta'(n)}) + 4\sqrt{\delta'(n)}\delta(n)/5 \leq \rho(n)/2. \quad (\text{A.2})$$

One can see that since  $\rho(n) = \Omega(1)$ ,  $\delta'(n) = \Omega(1)$ . By Claim A.3, for

$$\begin{aligned} q''(n) &= 25 \frac{25}{(1 - \alpha(n))^2 \delta'(n)} \\ &= 25 \frac{25}{\delta(n)^2 \delta'(n)} \\ &= O((n\tau(n))^2), \end{aligned}$$

there exists such a subspace  $L$ , satisfying

$$\dim L \geq (1 - 2H_{\mathbb{F}}(5\sqrt{\delta'(n)}) - 4\sqrt{\delta'(n)}\delta(n)/5)n \leq \rho(n)n/2,$$

by Equation (A.2). We take  $(C')^n = C^n \cap L$  to be the code of the claimed properties, and note that since the co-dimension of  $(C')^n$  is at most  $(1 - \rho(n) + \rho(n)/2)n$  the rate of  $(C')^n$  is at least  $\rho(n)/2 = \Omega(1)$ .

It remains to show that  $(C')^n$  is a  $(q'(n), \delta'(n), \varepsilon)$ -LCC, and towards that we describe a corrector  $\text{Cor}'$  for it. Let  $\text{Cor}$  be a corrector for  $C^n$  promised by it being a  $(q(n), \delta(n), \varepsilon)$ -LCC, and let  $\text{Cor}''$  be a corrector promised by that  $L$  is a  $(q''(n), \delta'(n), \alpha(n))$ -local-amplifier. On input  $i \in [n]$ , and oracle access to  $z \in \mathbb{F}^n$  such that  $\text{dist}(z, c) \leq \delta'(n)n$  for  $c \in C'$ ,  $\text{Cor}'(i)$  as follows.  $\text{Cor}'(i)$  simulates  $\text{Cor}(i)$ , and whenever it needs to query  $z_j$  for some  $j \in [n]$ ,  $\text{Cor}'(i)$  simulates  $\text{Cor}''(j)$  with access to  $z$ , and feeds  $\text{Cor}(i)$  the result.

It is immediate the number of queries that  $\text{Cor}'(i)$  makes is at most  $q(n) \cdot q''(n) = O(q(n)(n\tau(n))^2)$ . As for the correctness, for any  $c \in L$  and thus for any  $c \in (C')^n$ , given that  $\text{dist}(z, c) \leq \delta'(n)n$ , we are promised that for at least an  $\alpha(n)$ -fraction of the indices  $j \in [n]$ ,  $\text{Cor}''(j) = c_j$ , when  $\text{Cor}''(j)$  is run with oracle access to  $c$ . Thus, from the point of view of the procedure  $\text{Cor}$ , it is given access to a string  $z'$  which satisfies  $\text{dist}(z', c) \leq (1 - \alpha(n))n = \delta(n)n$ , and thus  $\text{Cor}(i)$  correctly outputs  $c_i$  with probability at least  $1 - \varepsilon$ , by its promise. Thus,  $\text{Cor}'(i)$  also outputs  $c_i$  with probability at least  $1 - \varepsilon$ , as required.  $\square$