Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 20 (2015)

# Explicit Strong LTCs with inverse poly-log rate and constant soundness

Michael Viderman[*]

April 2, 2018

## Abstract

An error-correcting code $C \subseteq \mathbb{F}^n$ is called $(q, \epsilon)$-strong locally testable code (LTC) if there exists a tester that makes at most $q$ queries to the input word. This tester accepts all codewords with probability 1 and rejects all non-codewords $x \notin C$ with probability at least $\epsilon \cdot \delta(x, C)$, where $\delta(x, C)$ denotes the relative Hamming distance between the word $x$ and the code $C$. The parameter $q$ is called the query complexity and the parameter $\epsilon$ is called soundness.

Goldreich and Sudan (J.ACM 2006) asked about the existence of strong LTCs with constant query complexity, constant relative distance, constant soundness and inverse polylogarithmic rate. They also asked about the explicit constructions of these codes.

Strong LTCs with the required range of parameters were obtained recently in the works of Viderman (CCC 2013, FOCS 2013) based on the papers of Meir (SICOMP 2009) and Dinur (J.ACM 2007). However, the construction of these codes was *probabilistic*.

In this work we show that codes presented in the works of Dinur (J.ACM 2007) and Ben-Sasson and Sudan (SICOMP 2005) provide the *explicit* construction of strong LTCs with the above range of parameters. Previously, such codes were proven to be weak LTCs. Using the results of Viderman (CCC 2013, FOCS 2013) we prove that such codes are, in fact, strong LTCs.

[*] Yahoo Research, Haifa, Israel. Email: `viderman@oath.com`

ISSN 1433-8092

## Contents

# 1 Introduction

Probabilistically Checkable Proof (PCP) systems [2, 3, 21] (a.k.a. Holographic Proofs [4]) are proof systems that allow efficient probabilistic verification of a claim by reading few symbols of the proof. The celebrated PCP theorem [2, 3] is one of the main breakthrough results in complexity theory. This theorem asserts that for every language in $\mathcal{NP}$ there exists a polynomial-time PCP verifier that queries the proof in a constant number of locations. The verifier is guaranteed to always accept valid proofs of true statements, and to accept any claimed proof of false assertions with low probability. The theorem has found many applications in theoretical computer science, especially in establishing lower bounds for approximation algorithms [6, 5, 21, 28].

Informally, most of the PCP constructions were achieved using error-correcting codes, possessing nice properties. Let us first give some auxiliary definitions regarding error-correcting codes.

A code over a finite alphabet $\Sigma$ is a subspace $\mathcal{C} \subseteq \Sigma^n$. A linear code over a finite field $\mathbb{F}$ is a linear subspace $\mathcal{C} \subseteq \mathbb{F}^n$. In this case, $n$ is the blocklength of the code $\mathcal{C}$, denoted by blocklength$(\mathcal{C})$. The dimension of a linear code $\mathcal{C}$, denoted by $\dim(\mathcal{C})$, is its dimension as a vector space and is equal to $\log_{|\mathbb{F}|} |\mathcal{C}|$. The dimension of a non-linear code $\mathcal{C}$ over the alphabet $\Sigma$ is defined to be $\dim(\mathcal{C}) = \log_{|\Sigma|} |\mathcal{C}|$. The rate of a code $\mathcal{C}$, denoted by rate$(\mathcal{C})$, is defined to be $\frac{\dim(\mathcal{C})}{\text{blocklength}(\mathcal{C})} = \frac{\dim(\mathcal{C})}{n}$.

We define the distance between two words $x, y \in \mathbb{F}^n$ to be $\Delta(x, y) = |\{i \mid x_i \neq y_i\}|$ and the relative distance to be $\delta(x, y) = \frac{\Delta(x,y)}{n}$. The distance of $\mathcal{C}$ is defined by $\Delta(\mathcal{C}) = \min\limits_{x \neq y \in \mathcal{C}} \Delta(x, y)$ and its relative distance is defined by $\delta(\mathcal{C}) = \frac{\Delta(\mathcal{C})}{n}$. We note that if $\mathcal{C}$ is linear then $\Delta(\mathcal{C}) = \min\limits_{c \in \mathcal{C} \setminus \{0\}} \{|c|\}$. One is typically interested in codes whose distance is linear to the blocklength of $\mathcal{C}$, i.e., $\Omega(n)$.

For $x \in \mathbb{F}^n$ and $\mathcal{C} \subseteq \mathbb{F}^n$, let $\delta(x, \mathcal{C}) = \min\limits_{y \in \mathcal{C}} \{\delta(x, y)\}$ denote the relative distance of $x$ from the code $\mathcal{C}$. If $\delta(x, \mathcal{C}) \geq \rho$, we say that $x$ is $\rho$-far from $\mathcal{C}$ and otherwise $x$ is $\rho$-close to $\mathcal{C}$.

## 1.1 Locally Testable Codes

Most of the PCP constructions (e.g., [7, 11, 18, 26]) are tightly related to a special kind of error-correcting codes possessing some testability properties. These codes are called *locally testable*.

In other words, locally testable codes (LTCs) are error correcting codes that have a tester, which is a randomized algorithm with oracle access to the received word $x$. The tester reads a sublinear amount of information from $x$ and based on this "local view" decides if $x \in C$ or not. It should accept codewords with probability one, and reject words that are far (in Hamming distance) from the code with noticeable probability. Such codes are of interest in computer science due to their numerous connections to probabilistically checkable proofs (PCPs) and property testing (see the surveys [39, 23] for more information). LTCs were implicit already in [4] (cf. [23, Sec. 2.4]) and they were explicitly studied by Goldreich and Sudan [26].

By now several different constructions of LTCs are known including codes based on low-degree polynomials over finite fields and affine-invariant codes [1, 2, 16, 9, 8, 15, 27, 30, 32, 29, 37], constructions based on PCPs of proximity/assignment testers [7, 19, 18][1], sparse random linear codes [14, 31, 34] and tensor products of codes [20, 13, 12, 35, 41, 33].

Basically, there are two kinds of LTCs: weak and strong. A code $\mathcal{C}$ is said to be $(q, \epsilon, \rho)$-weak LTC if there exists a randomized algorithm $T$, called tester, that makes at most $q$ queries to the

[1] As was pointed out in [26], not all PCP constructions are known to yield LTCs, but some of them (e.g., PCPs of proximity/assignment testers) can be adapted to yield LTCs.

input word $w$. If $w \in \mathcal{C}$ then $T$ accepts $w$ with probability 1, but if $w$ is $\rho$-far from $C$ the tester $T$ rejects $w$ with probability at least $\epsilon$. Let us notice that the tester is not required to reject when $0 < \delta(w, \mathcal{C}) < \rho$. This is the reason why such codes are called *weak* LTCs.

In contrast to weak LTCs, the testers for strong LTCs are required to reject all non-codewords with corresponding probability. More formally, a code $\mathcal{C}$ is called $(q, \epsilon)$-strong LTC if there exists a tester $T$ that makes at most $q$ queries to the input word $w$. If $w \in \mathcal{C}$ then $T$ accepts $w$ with probability 1, but if $w \notin \mathcal{C}$ then $T$ rejects $w$ with probability at least $\epsilon \cdot \delta(w, \mathcal{C})$. The parameter $q$ is called the query complexity and the parameter $\epsilon$ is called soundness.

Informally, we say that a code $\mathcal{C}$ is a weak LTC if it has a linear distance and there exist constants $q, \epsilon > 0$ and $\rho \leq \delta(\mathcal{C})/3$ such that $C$ is a $(q, \epsilon, \rho)$-weak LTC. [2] Similarly, we say that a code $\mathcal{C}$ is a strong LTC if it has a linear distance and there exist constants $q, \epsilon > 0$ such that $\mathcal{C}$ is a $(q, \epsilon)$-strong LTC.

LTCs were explicitly studied in the work of Goldreich and Sudan [26], who presented probabilistic construction of strong LTCs. These LTCs achieve constant query complexity, constant soundness and rate $\dfrac{1}{\exp(\tilde{O}(\sqrt{\log n}))}$, where $n$ denotes the blocklength.

Later, other constructions of LTCs [11, 18, 35] succeeded to obtain the rate $\dfrac{1}{\text{polylog}(n)}$ together with constant query complexity and soundness, however these codes were weak LTCs. It can be verified that every strong LTC is also a weak LTC, but some weak LTCs are not strong LTCs [43]. So, strong LTCs are strictly stronger objects than weak LTCs. As was pointed out by Goldreich [22], strong LTCs correspond to proximity oblivious testers [25] whereas weak LTCs are even weaker than ordinary testers, i.e., the testers for weak LTCs are supposed to work only for a fixed value of the proximity parameter. In the journal version of [26], the authors pointed out that all known LTCs that achieve inverse polylogarithmic rate are weak LTCs, and asked about the existence of strong LTCs with polylogarithmic rate and, in particular, about the *explicit* construction of such codes [26, Section 6].

The previous papers of the author [43, 42] showed a *probabilistic* construction of binary linear 3-query strong LTCs with inverse polylogarithmic rate, constant soundness and constant relative distance. In this paper (Section 1.4), we show the explicit construction of linear strong LTCs with constant query complexity, constant soundness, polylogarithmic rate and constant relative distance over a fixed field, therefore resolving a question raised by Goldreich and Sudan [26].[3] We would like to stress that the codes we refer to were, in fact, only a special case of PCPs and PCPs of proximity constructed in [11, 18]. Therefore, this work discovers strong local testability properties in the objects tightly related to the short PCPs and so, might be useful for the future PCPs related applications.

As was mentioned previously, we prove that the codes of [11, 18] yield explicit strong LTCs. To do that we want to reuse the arguments of [43, 42] to prove that the codes of [11, 18] are explicit strong LTCs. These codes (as well as codes of [35, 43, 42]) involve two kind of symbols: code symbols and proof symbols (called also core symbols and non-core symbols, respectively, in

[2] The parameter $\rho$ is required to be less than $\delta(\mathcal{C})/2$ to avoid trivial solutions like claiming that every perfect code $\mathcal{C}$ is a $(0, 1, \delta(\mathcal{C})/2)$-weak LTC. Recall that a code $\mathcal{C} \subseteq \mathbb{F}^n$ is called perfect if there are no words in $\mathbb{F}^n$ that are $(\delta(\mathcal{C})/2)$-far from $\mathcal{C}$. So, in this case one could say that no queries are needed and all $(\delta(\mathcal{C})/2)$-far words are rejected with probability 1 vacuously.

[3] A suggestion to show such explicit construction was raised in personal discussion with Or Meir, and later was asked in [42].

[43]). [4] However, the codes of [11] have good distance only on the code symbols with no guarantee on the proof symbols, while the arguments used in [43] require both good distance on the code coordinates and on the proof coordinates. Therefore, our main technical ingredient in this work is observing that the results of [43] can be reproved with only requirement of good distance on the code coordinates.

The rest of the paper is organized as follows. We provide the necessary definitions in Section 1.2 and state our main result (Theorem 1.5) in Section 1.4. The main ideas and the overview of the proof of Theorem 1.5 are given in Section 2. The proof of Theorem 1.5 is postponed to Section A.

## 1.2 Preliminaries

Let $[n]$ be the set $\{1, \ldots, n\}$. For $w \in \mathbb{F}^n$, let $\operatorname{supp}(w) = \{i \in [n] \mid w_i \neq 0\}$ and $|w| = |\operatorname{supp}(w)|$. For $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n) \in \mathbb{F}^n$ let $\langle u, v\rangle$ denote the bilinear function from $\mathbb{F}^n \times \mathbb{F}^n$ to $\mathbb{F}$ defined by $\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$. The dual code is defined by $\mathcal{C}^{\perp} = \{u \in \mathbb{F}^n \mid \forall c \in \mathcal{C} : \ \langle u, c\rangle = 0\}$. Similarly, we define $\mathcal{C}^{\perp}_{\leq t} = \left\{u \in \mathcal{C}^{\perp} \mid |u| \leq t\right\}$. For $w \in \mathbb{F}^n$ and $S = \{j_1, j_2, \ldots, j_m\} \subseteq [n]$ we let $w|_S = (w_{j_1}, w_{j_2}, \ldots, w_{j_m})$, where $j_1 < j_2 < \ldots < j_m$, be the restriction of $w$ to the subset $S$. Similarly, we let $\mathcal{C}|_S = \{c|_S \mid c \in \mathcal{C}\}$ denote the projection of the code $\mathcal{C}$ onto $S$. We define $\mathcal{C}|_{-S} = \mathcal{C}|_{[n]\setminus S}$, i.e., projection of the code $\mathcal{C}$ to all coordinates besides $S$. For $A \subseteq \mathbb{N}$ and $b \in \mathbb{N}$ we let $A + b = b + A = \{a + b \mid a \in A\}$. For a code $\mathcal{C}$ we let $\operatorname{coord}(\mathcal{C})$ to be a coordinate set of the code, e.g., if $\mathcal{C} \subseteq \mathbb{F}^n$ then $\operatorname{coord}(\mathcal{C}) = [n]$.

For the distribution $\mathcal{D}$ over the subsets of $[n]$ we let $\mathcal{D}(I)$ to denote the probability that a subset $I \subseteq [n]$ is selected by $\mathcal{D}$ and $\operatorname{supp}(\mathcal{D}) = \{I \subseteq [n] \mid \mathcal{D}(I) > 0\}$. For $i \in [n]$ we let $N_{\mathcal{D}}(i) = \{I \in \operatorname{supp}(\mathcal{D}) \mid i \in I\}$.

Now we define testers and LTCs (see [26, 43] for the justification of this definition).

**Definition 1.1** (LTCs and Testers)**.** A $q$-query tester for a code $\mathcal{C} \subseteq \mathbb{F}^n$ is a distribution $\mathcal{D}$ over subsets $I \subseteq [n]$ such that $|I| \leq q$. A $q$-query tester $\mathcal{D}$ is a $(q, \epsilon, \rho)$-weak tester if for all $w \in \mathbb{F}^n$, $\delta(w, \mathcal{C}) \geq \rho$ we have $\Pr_{I\sim\mathcal{D}}[w|_I \notin \mathcal{C}|_I] \geq \epsilon$. A $q$-query tester $\mathcal{D}$ is a $(q, \epsilon)$-strong tester if for all $w \in \mathbb{F}^n$ we have $\Pr_{I\sim\mathcal{D}}[w|_I \notin \mathcal{C}|_I] \geq \epsilon \cdot \delta(w, \mathcal{C})$.

A code $\mathcal{C} \subseteq \mathbb{F}^n$ is a $(q, \epsilon, \rho)$-*weak LTC* if it has a $(q, \epsilon, \rho)$-weak tester. A code $\mathcal{C} \subseteq \mathbb{F}^n$ is a $(q, \epsilon)$-*strong LTC* if it has a $(q, \epsilon)$-strong tester.

**Remark 1.2.** Although the tester in Definition 1.1 does not output accept or reject, the way a standard tester does, it can be converted to output accept, reject as follows. Whenever the task is to test whether $w \in \mathcal{C}$ and a subset $I \subseteq [n]$ is selected by the tester, the tester can output accept if $w|_I \in \mathcal{C}|_I$ and otherwise output reject. In this manner, the tester always accepts the codewords of $\mathcal{C}$.

It is not hard to see that every strong LTC is a weak LTC, but not vice versa [43, Proposition B.1].

The support of the tester $\mathcal{D}_C$ for the code $\mathcal{C} \subseteq \mathbb{F}^n$ is denoted by $\operatorname{supp}(\mathcal{D}_C)$ and defined to be $\operatorname{supp}(\mathcal{D}_C) = \{I \subseteq [n] \mid \mathcal{D}_C(I) > 0\}$. We say that $\mathcal{D}_C$ is uniform over its support if for each

[4] The concepts like 'code' and 'proof' coordinates appear in PCP related literature, but might appear under different names (statement and proof) and with slightly different meaning as well.

$I_1, I_2 \in \operatorname{supp}(\mathcal{D}_C)$ we have $\mathcal{D}_C(I_1) = \mathcal{D}_C(I_2)$. The test neighbors of the coordinate $i \in [n]$ are defined by $N_{\mathcal{D}_C}(i) = \{I \subseteq [n] \mid \mathcal{D}_C(I) > 0, i \in I\}$.

### 1.2.1 Tensor Product of Codes

The definitions appearing here are standard in the literature on tensor-based LTCs (e.g., [10, 35, 43, 20, 41, 12]). For $x \in \mathbb{F}^{n_1}$ and $y \in \mathbb{F}^{n_2}$ we let $x \otimes y$ denote the tensor product of $x$ and $y$ (i.e., the matrix $M$ with entries $M(i,j) = x_j \cdot y_i$ where $(i,j) \in [n_2] \times [n_1]$). Let $\mathcal{R} \subseteq \mathbb{F}^{n_1}$ and $\mathcal{C} \subseteq \mathbb{F}^{n^2}$ be linear codes. We define the tensor product code $\mathcal{R} \otimes \mathcal{C}$ to be the linear space spanned by words $r \otimes c \in \mathbb{F}^{n_2 \times n_1}$ for $r \in \mathcal{R}$ and $c \in \mathcal{C}$. Some known facts regarding the tensor products (see e.g., [20]):

- The code $\mathcal{R} \otimes \mathcal{C}$ consists of all $n_2 \times n_1$ matrices over F whose rows belong to $\mathcal{R}$ and columns belong to $\mathcal{C}$,
- $\dim(\mathcal{R} \otimes \mathcal{C}) = \dim(\mathcal{R}) \cdot \dim(\mathcal{C})$,
- $\operatorname{rate}(\mathcal{R} \otimes \mathcal{C}) = \operatorname{rate}(\mathcal{R}) \cdot \operatorname{rate}(\mathcal{C})$ and
- $\delta(\mathcal{R} \otimes \mathcal{C}) = \delta(\mathcal{R}) \cdot \delta(\mathcal{C})$.

We let $\mathcal{C}^{\otimes 1} = \mathcal{C}$ and $\mathcal{C}^{\otimes m} = \mathcal{C}^{\otimes (m-1)} \otimes \mathcal{C}$ for $m > 1$. Note by this definition, $\mathcal{C}^{\otimes 2^0} = \mathcal{C}$ and $\mathcal{C}^{2^m} = \mathcal{C}^{\otimes 2^{m-1}} \otimes \mathcal{C}^{\otimes 2^{m-1}}$ for $t > 0$. We also notice that for a code $\mathcal{C} \subseteq \mathbb{F}^n$ and $m \geq 1$ it holds that $\operatorname{rate}(\mathcal{C}^{\otimes m}) = (\operatorname{rate}(\mathcal{C}))^m$, $\delta(\mathcal{C}^{\otimes m}) = (\delta(\mathcal{C}))^m$ and the blocklength of $C^{\otimes m}$ is $n^m$. We notice that if $\operatorname{coord}(\mathcal{C}) = [n]$ then the coordinate set of $\mathcal{C} \otimes \mathcal{C}$ is $\operatorname{coord}(\mathcal{C} \otimes \mathcal{C}) = [n] \times [n]$.

## 1.3 Robust Testing

In this section we define some properties of codes that are sufficient for robust testing. We start this section by defining the notion of robustness (Definition 1.4) as was introduced in [10]. To do that we provide the definition of local distance (Definition 1.3), which will be used in Definition 1.4 and later in our proofs. In this section we use $n$ to denote the blocklength of the code $C$, i.e., $n = |\operatorname{coord}(C)|$. Without loss of generality we assume that $\operatorname{coord}(C) = [n]$.

**Definition 1.3** (Local distance)**.** Let $C \subseteq \mathbb{F}^n$ be a code and $w|_I$ be the view on the coordinate set $I \subseteq [n]$ obtained from the word $w \in \mathbb{F}^n$. The local distance of $w$from $C$ with respect to $I$ is $\Delta(w|_I, C|_I) = \min c \in C\{\Delta(w|_I, c|_I)\}$ and similarly the relative local distance of $w$ from $C$ with respect to $I$ is $\delta(w|_I, C|_I) = \min_{c \in C} \{\delta(w|_I, c|_I)\}$.

Informally, we say that a tester is robust if for every word that is far from the code, the tester view is far on average from any consistent view. This notion was defined for LTCs following an analogous definition for PCPs [7, 18]. We are ready to provide a general definition of robustness.

**Definition 1.4** (Robustness)**.** Given a tester (i.e., a distribution) $\mathcal{D}$ for the code $C \subseteq \mathbb{F}^n$, we let

$$\rho^{\mathcal{D}}(w) = \mathop{\mathbf{E}}_{I \sim \mathcal{D}}[\delta(w|_I, C|_I)] \quad \text{be the expected relative local distance of input } w.$$

We say that the tester $D$ has robustness $\rho^{\mathcal{D}}(C)$ on the code $C$ if for every $w \in \mathbb{F}^n$ it holds that $\rho^{\mathcal{D}}(w) \geq \rho^{\mathcal{D}}(C) \cdot \delta(w, C)$. Let $\{C_n\}_n$ be a family of codes where $C_n$ is of blocklength $n$ and $D_n$ is a tester for $C_n$. A family of codes $\{C_n\}_n$ is robustly testable with respect to testers $\{D_n\}_n$ if there exists a constant $\alpha > 0$ such that for all $n$ we have $\rho^{\mathcal{D}_n}(C_n) \geq \alpha$.

### 1.4 Main Result

In this paper we show the *explicit* construction of strong LTCs over a fixed field with a range of parameters asked by Goldreich and Sudan [26]. Although the requested range of parameters was achieved for the *probabilistic* construction of strong LTCs [43, 42], explicit strong LTCs with this range of parameters was not obtained.

**Theorem 1.5** (Main Theorem)**.** *There exist constants $q, d, \epsilon, \gamma > 0$ and a constant size field $\mathbb{F}$ such that for infinitely many $n \in \mathbb{N}^+$ we have an* explicit *construction of a linear code $C \subseteq \mathbb{F}^n$, where*

- *$C$ is a $(q, \epsilon)$-strong LTC,*
- *$\delta(C) \geq \gamma$ and* $\text{rate}(C) \geq \frac{1}{\log^d n}$.

The proof of Theorem 1.5 is given in Section A. Before we are going over the proof let us present the overview and the main ideas of this proof in Section 2.

## 2 Overview of the Proof

Roughly speaking, to prove Theorem 1.5 we apply the arguments of [43, 42] to the construction of [18] (which was obtained by applying the gap amplification procedure on the codes of [11]). However, it is impossible to use the observations of [43, 42] as is since there are considerable differences between the construction of [11] and the construction of [43, 42] (based on [35]).

The first obstacle is that the codes of [11] were obtained from iterative polynomial constructions, while the codes appeared in [43, 42] (based on [35]) are tensor products of general error correcting codes. This issue is resolved since the construction of [11] can be viewed as a kind of tensoring over a large field (see [35, Section 7.2]). It is worth to mention that while [43, 35] used iterative 3-wise tensor products for the code construction, [11] can be viewed as iterative 2-wise tensor products (see Section A.2). In general, codes composed as 2-wise tensor products might be not testable [17, 40, 24], however, there are still ways to use such kind of code products to obtain locally testable codes, e.g., [20, 12, 13].

The second difficulty is that both kind of constructions have the 'code' coordinates and 'proof' coordinates. While there is no distance guarantee on the proof coordinates in [11], the constant relative distance on these coordinates is required in the works of [43, 42]. More formally, consider a code $C \subseteq \mathbb{F}^n$ from [11] and assume that $[n]$ is partitioned to code coordinates set $s_1$ and proof coordinates set $s_2$, i.e., $s_1 \cup s_2 = [n]$, $s_1 \cap s_2 = \emptyset$. So, while nothing could be guaranteed regarding $\delta(C|_{s_2})$, to use the arguments of [43, 42] we need to know that $\delta(C|_{s_2})$ is constant, or in words, $C|_{s_2}$ has constant relative distance.

In [43] there was suggested a way to slightly modify the construction of [35] to ensure the good distance on the proof coordinates. However, it would be much more problematic to modify a part of coordinates in the construction of [11] since it has very concrete and non-flexible polynomial-based structure. Therefore, the main technical ingredient in our work is the modification of the arguments of [43, 42] to avoid 'good distance' requirement from the proof coordinates. We succeed to prove that it was redundant requirement and only a good distance on the code coordinates is sufficient.

So far, the proof (presented in Section A) is starting from presenting central concepts which play crucial role in the proof. After these concepts are presented, the rest of the proof is done in 3 stages. The first stage (Section A.2) argues that a construction of [11] can be viewed as iterative tensoring with some “smart” projection procedure, where in every iteration a tensoring is done only over the code coordinates and a part of the symbols are moved from the code coordinates part to the proof coordinates part.

The second stage (Section A.3) is to recall the arguments of [43] to argue that codes of [11] are COLTCs (Definition A.3). This stage reproves one of the technical ingredients of [43] without “good” distance requirement from the proof coordinates, and in particular, Corollary A.14 shows that the codes of [11] are COLTCs (and hence strong LTCs) even though no distance is guaranteed on the proof coordinates.

Finally, in Section A.4 we recall the arguments of [42] to argue that when the gap amplification procedure of [18] is applied to the codes of [11] it yields strong LTCs. In words, this section recalls a relaxed LTC (rLTC) concept (Definition A.15) from [42], and the observation that strong LTCs are also relaxed LTCs with the corresponding range parameters. Theorem A.17 and its Corollary A.18 claim that the gap amplification of Dinur [18] can be applied to the codes of [11], proven to be COLTCs in Corollary A.14 (and thus are strong LTCs and rLTCs), and yields rLTCs with the constant first soundness parameter and inverse poly-log second soundness parameter. Corollary A.20 implies that such rLTCs can be explicitly converted to the required strong LTCs with constant soundness.

The fact that the codes construction of [11] and the gap amplification procedure are deterministic, yields *explicit* strong LTCs.

Theorem 1.5 will follow from Corollary A.18 and Corollary A.20.

### Acknowledgements

The author thanks Or Meir for raising the suggestion to obtain an explicit construction of strong LTCs by applying the arguments of [43, 42] to the codes of [11]. We would like to thank the anonymous reviewers for their valuable comments.

## References

[1] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

[2] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof Verification and the Hardness of Approximation Problems. *Journal of the ACM*, 45(3):501–555, May 1998.

[3] Sanjeev Arora and Shmuel Safra. Probabilistic Checking of Proofs: A New Characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

[4] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking Computations in Polylogarithmic Time. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing (STOC), May 5-8, 1991, New Orleans, Louisiana, USA*, pages 21–31. ACM, 1991.

[5] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing (STOC)*, pages 294–304, New York, 1993. ACM SIGACT, ACM Press.

[6] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free Bits, PCPs, and Nonapproximability—Towards Tight Results. *SIAM Journal on Computing*, 27(3):804–915, June 1998.

[7] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Robust PCPs of Proximity, Shorter PCPs, and Applications to Coding. *SIAM Journal on Computing*, 36(4):889–974, 2006.

[8] Eli Ben-Sasson, Elena Grigorescu, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. On sums of locally testable affine invariant properties. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 6845 of *Lecture Notes in Computer Science*, pages 400–411. Springer, 2011.

[9] Eli Ben-Sasson, Noga Ron-Zewi, and Madhu Sudan. Sparse affine-invariant linear codes are locally testable. In *FOCS*, pages 561–570. IEEE Computer Society, 2012.

[10] Eli Ben-Sasson and Madhu Sudan. Robust locally testable codes and products of codes. *Random Struct. Algorithms*, 28(4):387–402, 2006.

[11] Eli Ben-Sasson and Madhu Sudan. Short PCPs with polylog query complexity. *SIAM J. Comput*, 38(2):551–607, 2008.

[12] Eli Ben-Sasson and Michael Viderman. Composition of Semi-LTCs by Two-Wise Tensor Products. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 5687 of *Lecture Notes in Computer Science*, pages 378–391. Springer, 2009.

[13] Eli Ben-Sasson and Michael Viderman. Tensor Products of Weakly Smooth Codes are Robust. *Theory of Computing*, 5(1):239–255, 2009.

[14] Eli Ben-Sasson and Michael Viderman. Low rate is insufficient for local testability. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 6302 of *Lecture Notes in Computer Science*, pages 420–433. Springer, 2010.

[15] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of reed-muller codes. In *FOCS*, pages 488–497. IEEE Computer Society, 2010.

[16] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, December 1993.

[17] Don Coppersmith and Atri Rudra. On the Robust Testability of Product of Codes. *Electronic Colloquium on Computational Complexity (ECCC)*, (104), 2005.

[18] Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3):12:1–12:44, June 2007.

[19] Irit Dinur and Omer Reingold. Assignment Testers: Towards a Combinatorial Proof of the PCP Theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006.

[20] Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust Local Testability of Tensor Products of LDPC Codes. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 4110 of *Lecture Notes in Computer Science*, pages 304–315. Springer, 2006.

[21] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[22] Oded Goldreich. Home page.

[23] Oded Goldreich. Short Locally Testable Codes and Proofs (Survey). *Electronic Colloquium on Computational Complexity (ECCC)*, (014), 2005.

[24] Oded Goldreich and Or Meir. The tensor product of two good codes is not necessarily robustly testable. *Inf. Process. Lett*, 112(8-9):351–355, 2012.

[25] Oded Goldreich and Dana Ron. On proximity-oblivious testing. *SIAM J. Comput*, 40(2):534–566, 2011.

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

[27] Elena Grigorescu, Tali Kaufman, and Madhu Sudan. Succinct representation of codes with applications to testing. *SIAM J. Discrete Math*, 26(4):1618–1634, 2012.

[28] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[29] Tali Kaufman and Shachar Lovett. New Extension of the Weil Bound for Character Sums with Applications to Coding. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, (FOCS)*, pages 788–796. IEEE, 2011.

[30] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. Comput*, 36(3):779–802, 2006.

[31] Tali Kaufman and Madhu Sudan. Sparse Random Linear Codes are Locally Decodable and Testable. In *FOCS*, pages 590–600. IEEE Computer Society, 2007.

[32] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC), Victoria, British Columbia, Canada, May 17-20, 2008*, pages 403–412. ACM, 2008.

[33] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):11:1–11:42, 2017.

[34] Swastik Kopparty and Shubhangi Saraf. Local list-decoding and testing of random linear codes from high error. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 417–426. ACM, 2010.

[35] Or Meir. Combinatorial Construction of Locally Testable Codes. *SIAM J. Comput*, 39(2):491–544, 2009.

[36] Alexander Polishchuk and Daniel A. Spielman. Nearly-linear size holographic proofs. In *Proceedings of the 26th Annual Symposium on the Theory of Computing*, pages 194–203, New York, May 1994. ACM Press.

[37] Noga Ron-Zewi and Madhu Sudan. A new upper bound on the query complexity for testing generalized reed-muller codes. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 7408 of *Lecture Notes in Computer Science*, pages 639–650. Springer, 2012.

[38] Madhu Sudan. Algorithmic introduction to coding theory, Lecture notes, 2001.

[39] Luca Trevisan. Some Applications of Coding Theory in Computational Complexity, September 23 2004.

[40] Paul Valiant. The Tensor Product of Two Codes Is Not Necessarily Robustly Testable. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 3624 of *Lecture Notes in Computer Science*, pages 472–481. Springer, 2005.

[41] Michael Viderman. A combination of testability and decodability by tensor products. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 7408 of *Lecture Notes in Computer Science*, pages 651–662. Springer, 2012.

[42] Michael Viderman. Strong LTCs with inverse poly-log rate and constant soundness. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS, Berkeley, CA, USA, 26-29 October, 2013*, pages 330–339, 2013.

[43] Michael Viderman. Strong LTCs with inverse polylogarithmic rate and soundness. In *Proceedings of the 28th Conference on Computational Complexity, CCC, Palo Alto, California, USA, 5-7 June, 2013*, pages 255–265, 2013.

# A Proof of Theorem 1.5

In this section we present the proof of Theorem 1.5 which contains three parts. Let us first present in Section A.1 the central concepts from [43] which play the important part in this work as well.

## A.1 Preliminary notations: Core Oriented LTCs and Core Oriented Robustness

**Definition A.1** (A core of the code)**.** Let $C \subseteq \mathbb{F}^n$ be a linear code. A core of the code $C$, denoted by $A(C)$, is a nonempty subset of $[n]$ such that $\dim(C) = \dim(C|_{A(C)})$, i.e., any assignment to the entries of $A(C)$ uniquely determines the entries of $[n] \setminus A(C)$. In particular, for any $c \in C$ there is no $c' \in C$ such that $c|_{A(C)} = c'|_{A(C)}$ and $c|_{[n]\setminus A(C)} \neq c'|_{[n]\setminus A(C)}$.

We say that $A(C)$ is a $\gamma$-core of the code $C$ if $A(C)$ is a core of $C$, $\delta(C|_{A(C)}) = \frac{\Delta(C|_{A(C)}}{|A(C)|} \geq \gamma$.

Clearly, there might be many options for $A(C)$, and in this case we fix only one such option. If $A(C) = [n]$ then for any $w, w' \in \mathbb{F}^n$ we let $\delta(w|_{[n]\setminus A(C)}, w'|_{[n]\setminus A(C)}) = \delta(w|_{[n]\setminus A(C)}, C|_{[n]\setminus A(C)}) = 0$.

We let residue(C) = [n] \ A(C) to be the non-core coordinates of $C$.

Usually, in the locally testable codes the distance is measured exactly as in the general error-correcting codes, i.e., with respect to the entire blocklength. However, when we consider a specific subset of coordinates, called the core of the code, we need to define a new concept of distance (used in [43]).

**Definition A.2** (Core oriented distance)**.** Assume $C \subseteq \mathbb{F}^n$ is a linear code and $A(C)$ is its core. We define a core oriented distance between two words $w, w'F^n$ to be

$$\delta_{A(C)}(w, w') = \max\left\{\delta(w, w'), \delta(w|_{A(C)}, w'|_{A(C)})\right\},$$

and a core oriented distance between the word $wF^n$ and the code $C$ to be

$$\delta_{A(C)}(w, C) = \min_{c \in C}(\delta_{A(C)}(w, c)).$$

We note that for every code $C \subseteq \mathbb{F}^n$ with a core $A(C)$ and $w \in F^n$ it holds that $\delta_{A(C)}(w, C) \geq \delta(w, C)$. In particular, if $w$ is $\delta$-close to $C$ with respect to the core oriented distance then $w$ is $\delta$-close to $C$ with respect to the "standard" distance $\delta(w, C)$.

An important building block in the proof is a new kind of local testable code defined in [43].

**Definition A.3** (Core Oriented LTC (COLTC))**.** Let $C \subseteq \mathbb{F}^n$ be a linear code and let $\mathcal{D}$ be a distribution over subsets $I \subseteq [n]$ such that $|I| \leq q$. A $\mathcal{D}$ is a $(q, \epsilon)$-COLTC tester if, given that $A(C)$ is a core of $C$, for all $w \in \mathbb{F}^n$ we have

$$\Pr_{I \sim D}[w|_I \notin C|I] \geq \epsilon \cdot \delta_{A(C)}(w, C).$$

A code $C \subseteq \mathbb{F}^n$ is called a $(q, \epsilon)$-COLTC if it has a $(q, \epsilon)$-COLTC tester.

Let $C$ be a linear code and $A(C)$ be its core. If $C$ is a $(q, \epsilon)$-COLTC (with respect to the tester $D_C$) then $C$ is a $(q, \epsilon)$-strong LTC. To see this let $w \in \mathbb{F}^n$ and note that

$$\Pr_{I \sim D_C}[w|_I \in C|_I] \geq \epsilon \cdot \delta_{A(C)}(w, c) \geq \epsilon \cdot \delta(w, C).$$

### A.1.1 Core Oriented Robust Testing of COLTCs

Now we present one of the central concepts in [43] called "core robustness". Before that recall the more standard notion of robust testing (Definition 1.4) used e.g., in [10, 20, 12, 35].

In contrast to robust testing, in a core oriented robust testing (Definition A.4) we pay a special attention on the core of the code and consider a core oriented distance rather than a standard distance as in Definition 1.4.

**Definition A.4** (Core robustness)**.** Assume that $\mathcal{D}$ is a tester (i.e., a distribution) for the linear code $C \subseteq \mathbb{F}^n$ with a core $A(C)$. Assume that for every subset $I \subseteq [n]$ such that $\mathcal{D}(I) > 0$ it holds that $C|_I$ is a linear code with a core $A(C|_I)$. We let

$$\rho^{\mathcal{D}}_{A(C)}(w) = \mathop{\mathbf{E}}_{I \sim \mathcal{D}}\left[\delta_{A(C|_I)}(w|_I, C|_I)\right]$$

be the expected core oriented relative local distance of input $w$. We say that the tester $\mathcal{D}$ for the code $C$ has core robustness $\rho^{\mathcal{D}}_{A(C)}(C)$ if for every $w \in \mathbb{F}^n$ it holds that $\rho^{\mathcal{D}}_{A(C)}(w) \geq \rho^{D}_{A(C)}(C) \cdot \delta_{A(C)}(w, C)$.

It turns out that a combination of core robustness with COLTCs is highly useful.

**Claim A.5** (Claim 7.2 in [43])**.** *Let $C$ be a $(q, \epsilon)$-COLTC and let $D_C$ be its tester. Let $\hat{C} \subseteq \mathbb{F}^{\mathrm{coord}(\hat{C})}$ be a linear code with a core $A(\hat{C})$ and let $\mathcal{D}_{\hat{C}}$ be its tester. Assume that $\rho^{D_{\hat{C}}}_{A(\hat{C})}(\hat{C}) \geq \alpha$ and for every local view $I \subseteq \mathrm{coord}(\hat{C})$ such that $\mathcal{D}_{\hat{C}}(I) > 0$ it holds that $\hat{C}|_I = C$. Then $\hat{C}$ is a $(q, \alpha \cdot \epsilon)$-COLTC.*

### A.1.2 Star Products

Now we provide a definition of star products from [43]. These products of codes are very similar to ones used in [35, Section 4], although there exist minor differences. We notice that in this work we are interested mainly in such products of second power (and will not consider $m$-wise products for $m \geq 3$ as in [43]).

Informally, 2-star product of $C \in \mathbb{F}^n$ is defined by taking tensor product on its core coordinates $A(C)$ and appending its non-core parts (residues) $[n] \setminus A(C)$ to every row/column of such tensor product. This tensor product becomes a core of the obtained start product, and all appended residues are the residue of the star product.

**Definition A.6** (Star Products - $C^{\star 2}$)**.** . Let $C \subseteq \mathbb{F}^n$ be a linear code with a $\gamma$-core $A(C)$, where $\gamma > 0$ is a constant.

We let $C^{\star 2}$ to be a linear code over $\mathbb{F}$ such that its core and residue will be defined as follows.

we let the core coordinates $A(C^{\star 2}) = A(C) \times A(C)$, and the projection of the code on the core coordinates is $C^{\star 2}|_{A(C^{\star 2})} = C|_{A(C)} \otimes C|_{A(C)}$. The residue of $C^{\star 2}$ (residue(C$^{\star 2}$)) is defined by the residue of every codeword $c \in C^{\star 2}$ as follows:

view every such codeword $c|_{A(C^{\star 2})} \in C^{\star 2}|_{A(C^{\star 2})}$ as a matrix of size $|A(C)| \times |A(C)|$. Its residue(c) is defined by appending for every row $r$ of $c|_{A(C^{\star 2})}$ attach residue residue(r) such that both parts together is a codeword in $C$, i.e., $(r, \text{residue(r)}) \in$ C. Since $A(C)$ is a core, the residue(r) is defined uniquely given $r$.

The blocklength of $C^{\star 2}$ is $|A(C)|^2 + 2 \cdot A(C) \cdot (n - A(C))$.

Now, let us define formally the tester for the star product.

**Definition A.7** (Tester for Star Product)**.** Let $\mathcal{D}_C$ be a tester for a linear code $C \in \mathbb{F}^n$ with core $A(C)$. Then, a tester for $C^{\star 2}$, denoted by $\mathcal{D}_{C^{\star 2}}$, is defined by:

- pick random $r \in \{0, 1\}$
  - if $r = 0$ pick random row of $A(C^{\star 2})$ and its residue.
  - else pick random column of $A(C^{\star 2})$ and its residue.

Notice that for every local view $I \subseteq \mathrm{coord}(C^{\star 2})$ such that $\mathcal{D}_{C^{\star 2}}(I) > 0$ it holds that $C^{\star 2}|_I = C$. I.e., a local view of such a tester on any codeword of $C^{\star 2}$ is always a codeword of $C$.

The rest of the proof is organized as follows. A first part of the proof (Section A.2) argues that codes of [11] can be viewed as tensoring over sufficiently large field, together with the *explicit* projection step. The purpose of this part is to define the required abstraction over the codes of [11] needed for our work.

A second part is presented in Section A.3, where it is explained that the codes of Ben-Sasson and Sudan [11] are COLTCs, however, their parameters range is not sufficiently nice. Here we use the fact that their construction can be viewed as a repetitive tensoring as was explained in Section A.2.

Finally, Section A.4 recalls the notion of a relaxed LTC and some related results [42]. It shows that a work of Dinur [18] can be used to improve that soundness of the relaxed LTCs. It turns out that this improvement is sufficient to change the codes of [11] to relaxed LTCs with good enough parameters (which is what we need by Corollary A.20). Since the gap amplification technique is explicit, it yields explicit construction of relaxed LTCs. Corollary A.20 proves that relaxed LTCs with sufficiently nice parameters can be converted to strong LTCs with constant soundness. That proves Theorem 1.5.

## A.2 Abstraction over the codes of [11]

Let us recall an auxiliary procedure that will be used in the code constructions.

**Definition A.8** (Projection of core symbols)**.** Given a linear code $C \subseteq \mathbb{F}^n$ with a code $A(C)$, the $\mathrm{projection(C)} \subseteq \mathrm{\mathbb{F}^n}$ is a linear code with a new core $A' = A'(C) \subseteq A(C)$, i.e., some core symbols moved to the non-core part.

The abstraction assumed in this Section described in [35, Section 7.2]. For the sake of completeness we give the required details in this section.

Let $\delta, \gamma > 0$ be constants and $\mathbb{F}$ be sufficiently large field. [5]

The explicit construction of [11] can be viewed in the following way. It started from an error-correcting code of constant size blocklength $C_1 \subseteq \mathbb{F}^{n_1}$ such that $A(C_1) = [n_1]$, and for $i = 2 \ldots \log\log n$ the following holds.

- $T_i = C_{i-1}^{\star 2}$, i.e., tensoring is done over the core coordinates and residue coordinates are just appended to every row/column code.
- $C_{i+1} = \mathrm{projection(T_i)}$, where a constant fraction of the code coordinates are moved from the "core" part ($A(T_i)$) to the "residue" part ($\mathrm{residue(T_i)}$). Notice that projection is the explicitly defined procedure.
- $A(C_{i+1})$ is a $\gamma$-core of $C_{i+1}$, and in particular, $\delta(C_i|_{A(\mathrm{projection(C_i)})}) \geq \gamma$.
- $\dim(C_{i+1}) = \dim(C_i)^2$
- It is known that $\mathrm{rate}(C_i|_{A(\mathrm{projection(C_i)})})) \geq \delta$
- It is known that $\mathrm{rate}(C_{i+1}) \geq \delta \cdot \mathrm{rate}(C_i)$

It was proved in [11] that if $w \in \mathbb{F}^{n_i}$ and $w|_{A(C_i)} \notin C_i|_{A(C_i)}$ it holds that the local view of the tester $\mathcal{D}_{C_i}$ is far from the corresponding code, and this distance is proportional to $\delta(w|_{A(C_i)}, C_i|_{A(C_i)})$. For more details please see Section B.

[5] The codes of [11] are constructed over fields of size linear in a blocklength of the code. The folklore statement (Theorem C.1) says it is not hard to explicitly convert such strong LTCs to be over a constant size field with similar parameters.

**Claim A.9** ([11]). *Let $i$ be the number of iteration in the code construction of [11]. Let $C_i \subseteq \mathbb{F}^{n_i}$ be the [11] code from $i^{th}$ iteration and $A(C_i)$ its $\gamma$-core for a constant $\gamma > 0$. There exists a constant $\epsilon_\gamma$ depending only on $\gamma$ such that for every $M \in \mathbb{F}^{|A(C_i)|\times|A(C^i)|}$ the random row/column of $M$ is $\epsilon \cdot \delta(M, C_i|_{A(C_i)} \otimes C_i|_{A(C_i)})$ close to $C_i|_{A(C_i)}$.*

In the classical PCPP structure [11], i.e., the distance is measured only between the core coordinates, and proof (non-core) coordinates help only to test, without being involved into the distance calculations. In our scenario (strong locally testable codes) we need to take the proof coordinates into consideration as well.

It is not hard to verify that the above codes construction has the following properties.

**Claim A.10.** *Let $C \subseteq \mathbb{F}^n$ be the code from the above construction and $\mathcal{D}_C$ be its tester. Then,*

- *$\mathcal{D}_C$ is uniform over $\mathrm{supp}(\mathcal{D}_C)$, and*
- *for each $i \in [n]$ we have $|N_{\mathcal{D}_C}(i)| \leq O(\mathrm{poly}\log(n))$.*

*Proof.* The construction has $O(\log\log(n))$ iterations. Every iteration, the tester has uniform distribution over all its local views, and this invariant holds during the entire construction. Every iteration the maximal number of tests querying a coordinate is at most doubled. Hence, after $O(\log\log(n))$ iterations, it holds that for any $i \in [n]$ we have $|N_{\mathcal{D}_C}(i)| \leq 2^{O(\log\log(n))} = \mathrm{polylog}(n)$. □

### A.3 The codes of [11] are COLTCs — Main Technical ingredient

Now we reprove Theorem A.12 ([43]) showing that the star products are robustly testable with respect to "core robustness" (see Definition A.4). Then one can conclude that if $C$ is a $q$-query COLTC, then $C^{\star 2}$ is a $q$-query COLTC.

**Claim A.11** (Unique decoding w.r.t. core). *Let $C \subseteq \mathbb{F}^n$ be a linear code and let $A(C)$ be its $\gamma$-core. Assume $x \in \mathbb{F}^n$ such that $\delta_{A(C)}(x, C) < \gamma/2$ and $\delta(x|_{A(C)}, c|_{A(C)}) < \gamma/2$ for some $c \in C$. Then, $c$ is the unique closest codeword of $C$ to $x$ w.r.t. core distance and in particular, it holds that $\delta(x|_{A(C)}, c|_{A(C)}) \leq \delta_{A(C)}(x, c) < \gamma/2$.*

*Proof.* Assume to the contrary, i.e., there exists $c' \in C$ such that $c' \neq c$ and $\delta(x|_{A(C)}, c'|_{A(C)}) \leq \delta(x|_{A(C)}, c|_{A(C)}) < \gamma/2$.

Since $A(C)$ is a $\gamma$-core Definition A.1 implies that $\delta(C|_{A(C)}) \geq \gamma$ and hence $c'|_{A(C)} = c|_{A(C)}$. Then, again Definition A.1 implies that $c' = c$ since any assignment to the entries of $A(C)$ uniquely determines the entries of $[n] \setminus A(C)$. Contradiction.

We claim it holds that $\delta_{A(C)}(x, c) < \gamma/2$ because otherwise, by Definition A.2 we have $\delta(x, c) >= \gamma/2$. However, by the assumption of Claim A.11 there exists $c'' \in C$ such that $\delta_{A(C)}(x, c'') < \gamma/2$, but then, again $\delta(x|_{A(C)}, c''|_{A(C)}) < \gamma/2$ and $c'' = c$. Contradiction.

Finally, notice that by Definition A.2 we have $\delta(x|_{A(C)}, c|_{A(C)}) \leq \delta_{A(C)}(x, c)$. □

We are ready to prove that the star product of codes has the core robustness.

**Theorem A.12** (Star Products has core robustness). *Let $C$ be a linear code with a $\gamma$-core $A(C)$. Assume that $\mathcal{D}$ is the star product tester for the code $C^{\star 2}$. Let $\epsilon_\gamma$ be a constant from Claim A.9. Then,*

$$\rho^{\mathcal{D}}_{A(C^{\star 2})}(C^{\star 2}) \geq \frac{\epsilon_\gamma \cdot \gamma}{10}.$$

*Proof.* Let $M \in \mathbb{F}^{\text{coord}(C^{\star 2})}$ be an input word and $\alpha = \rho^{\mathcal{D}}_{A(C^{\star 2})}(M)$. If $\alpha \geq \frac{\epsilon_\gamma \cdot \gamma}{10}$ we are done. Otherwise, assume that $\alpha < \frac{\epsilon_\gamma \cdot \gamma}{10}$ for the rest of the proof. We need to prove that $\delta_{A(C^{\star 2})}(M, C^{\star 2}) \leq \frac{\alpha}{\frac{\epsilon_\gamma \cdot \gamma}{10}}$.

Recall that a local view of the tester $\mathcal{D}$ is either row or column of $M|_{A(C^{\star 2})}$ with its residue. We say the local view $x$ of the tester $\mathcal{D}$ is bad if $\delta(x, C) \geq \gamma/2$. Therefore, the fraction of bad local views is at most $\frac{\alpha}{\gamma/2}$.

By Definition A.4 it holds that the average local view (row/column with its residue) is $\alpha$-close to $C$ with respect to core distance $\delta_{A(C)}(\cdot,\cdot)$. By Claim A.9 we have $\delta(M|_{A(C^{\star 2})}, C^{\star 2}|_{A(C^{\star 2})}) \leq \alpha/\epsilon_\gamma$. Let $X$ be the closest codeword to $M|_{A(C^{\star 2})}$.

That means $\delta(M|_{A(C^{\star 2})}, X) \leq \alpha/\epsilon_\gamma$. Namely, at most $(\alpha/\epsilon_\gamma)$-fraction of symbols of $M|_{A(C^{\star 2})}$ should be changed in order to get $X$. Since the coordinates of $M|_{A(C^{\star 2})}$ form a matrix, it holds that a typical row is $(\alpha/\epsilon_\gamma)$-close to corresponding row in $X$, and a typical column is $(\alpha/\epsilon_\gamma)$-close to the corresponding row in $X$. The number of rows/columns that are $\gamma/2$-far from the corresponding row/column of $X$ is at most $\frac{\alpha}{\epsilon_\gamma \cdot \gamma/2}$.

Claim A.11 implies that $\delta_{A(C^{\star 2})}(M, C^{\star 2}) \leq \frac{\alpha}{\gamma/2} + \alpha/(\epsilon_\gamma \cdot \gamma/2) + \delta(M|_{A(C^{\star 2})}, C^{\star 2}|_{A(C^{\star 2})}) \leq \alpha/(\epsilon \cdot \gamma/8) \leq \frac{\alpha}{\frac{\epsilon_\gamma \cdot \gamma}{10}}$. We are done. □

The following simple claim from [43] says that COLTC properties are preserved after projection.

**Claim A.13.** *Let $C \subseteq \mathbb{F}^n$ be a linear code and $A(C)$ be its $\gamma$-core. Let $\beta > 0$ be a constant. Let $A'(C) \subseteq A(C)$ be a new $\gamma$-core of $C$, s.t. $|A'(C)| \geq |A(C)| \cdot \beta$. Then, if $C$ is a $(q, \epsilon)$-COLTC w.r.t. the core $A(C)$, then $C$ is a $(q, \epsilon \cdot \beta)$-COLTC w.r.t. the core $A'(C)$.*

*Proof.* The fact that $C$ is a $(q, \epsilon)$-COLTC w.r.t. the core $A(C)$ guarantees that there exists a $(q, \epsilon)$-COLTC tester $\mathcal{D}_C$ w.r.t. the core $A(C)$. The same tester $\mathcal{D}_C$ is also a $(q, \epsilon \cdot \beta)$-COLTC tester w.r.t. the core $A'(C)$ since for every $w \in \mathbb{F}^n$ the rejection probability of the tester remains the same regardless of the core (since it is the same tester), and on the other hand, $\beta \cdot \delta_{A(C)}(w, C) \leq \delta_{A'(C)}(w, C)$. The last statement holds since $\beta \cdot \delta(w|_{A(C)}, C|_{A(C)}) \leq \delta(w|_{A'(C)}, C|_{A'(C)})$. □

We are ready to claim that the codes of [11] are COLTCs with constant query complexity and inverse poly-log soundness.

**Corollary A.14** (The codes of [11] are COLTCs)**.** *Let $C \subseteq \mathbb{F}^n$ be obtained from the construction in Section A.2 after $O(\log \log n)$ iterations. Then, $C$ is a $(q, \frac{1}{\text{poly} \log n})$-COLTC and as a consequence a $(q, \frac{1}{\text{poly} \log n})$-strong LTC. Moreover, by Claim A.10 we have*

- *$\mathcal{D}_C$ is uniform over* $\text{supp}(\mathcal{D}_C)$*, and*
- *for each $i \in [n]$ we have $|N_{\mathcal{D}_C}(i)| \leq O(\text{poly} \log(n))$.*

*Proof.* The statement follows since the query complexity remains the same during the iterations, and the constructions start with a constant blocklength code, i.e., constant query complexity. The soundness parameter is reduced by a constant every iteration. Therefore, after $O(\log \log n)$ iterations, the soundness parameter of a $C$ is $\frac{1}{\text{poly} \log n}$. □

### A.4 Relaxed LTCs and the Gap Amplification

First, we recall a notion of *relaxed LTCs* (rLTCs) Definition A.15 from [42]. Intuitively, relaxed LTCs have two kind of coordinates: those with good testability and those which worse (but non-trivial) testability (see Definition A.15). Then, we state Theorem A.17 and its Corollary A.18 saying that the gap amplification of [18] when applied on the codes of [11] yields rLTCs with constant first soundness parameter and inverse poly-log second soundness parameter.

Finally, we recall an Observation A.19 and its Corollary A.20 from [42] saying that such relaxed LTCs can be easily converted to strong LTCs with constant soundness. Hence Theorem 1.5 follows.

**Definition A.15** (Relaxed LTC)**.** A $q$-query tester $\mathcal{D}$ is a $(q, \epsilon_1, \epsilon_2)$-rLTC tester for a linear code $C \subseteq \mathbb{F}^n$ with a core $A(C)$, if for every $w \in \mathbb{F}^n$ there exists $c \in C$ such that $\mathbf{Pr}_{I \sim \mathcal{D}}[w|_I \notin C|_I] \geq \max\left\{\epsilon_1 \cdot \delta(w|_{A(C)}, c|_{A(C)}), \epsilon_2 \cdot \delta(w|_{-A(C)}, c|_{-A(C)})\right\}$. A code $C \subseteq \mathbb{F}^n$ with a core $A(C)$ is a $(q, \epsilon_1, \epsilon_2)$-*rLTC* if it has a $(q, \epsilon_1, \epsilon_2)$-rLTC tester.

The parameter $q$ is called the query complexity, $\epsilon_1$ is called the first soundness parameter and $\epsilon_2$ is called the second soundness parameter.

Intuitively, think that $\epsilon_1$ is a constant, but $\epsilon_2$ is sub-constant.

The following simple observation [42] says that any strong LTC is also a relaxed LTC with similar parameters.

**Observation A.16** (Strong LTCs are relaxed)**.** *If $C \subseteq \mathbb{F}^n$ is a $(q, \epsilon)$-strong LTC then it is also a $(q, \epsilon, 1)$-rLTC with regards to the code $A(C) = [n]$.*

The observation follows immediately from the definition of relaxed LTCs (Definition A.15).

In [42] it was explained that the gap amplification of Dinur can be applied to strong LTCs with inverse poly-log soundness, which can be considered as relaxed LTCs (Observation A.16). In this case, the gap amplification preserves it to be an rLTC where first soundness parameter increased by a constant, while the second parameter decreased by a constant. In particular, the arguments of Section 4 and Section 5 in [42] can be summarized to the following theorem describing the affect of the gap amplification technique when applied to the codes from Section A.2.

**Theorem A.17** (Stated in [42], Section 5)**.** *Let $q >= 2$ be a constant. Assume $C \subseteq \mathbb{F}^n$ is a $(q, \frac{1}{\text{polylog}(n)})$-strong LTC and $\mathcal{D}_C$ its tester such that it holds that*

- $\text{rate}(C) = \frac{1}{\text{polylog}(n)}$ *and* $\delta(C) \geq \Omega(1)$
- $\mathcal{D}_C$ *is uniform over* $\text{supp}(\mathcal{D}_C)$ *and for each* $i \in [n]$ *we have* $|N_{\mathcal{D}_C}(i)| \leq O(\log n)$*.*

*Then the following holds.*

*For constant $q \geq 2, \epsilon > 0$, a fixed field $\mathbb{F}$ and infinitely many $n \in \mathbb{N}^+$ we have explicit construction for a linear code $C' \subseteq \mathbb{F}^n$ with a core $A(C')$ such that $C'$ is a $(q, \epsilon, \frac{1}{\text{polylog}(n)})$-rLTC, $\delta(C'|_{A(C')}) = \Omega(1)$ and $\text{rate}(C') = \frac{1}{\text{polylog}(n)}$.*

Since Corollary A.14 satisfies the assumption of Theorem A.17, we conclude the following corollary.

**Corollary A.18.** *For constant $q \geq 2, \epsilon > 0$, a fixed field $\mathbb{F}$ and infinitely many $n \in \mathbb{N}^+$ we have explicit construction for a linear code $C \subseteq \mathbb{F}^n$ with a core $A(C)$ such that $C$ is a $(q, \epsilon, \frac{1}{\text{polylog}(n)})$-rLTC, $\delta(C|_{A(C)}) = \Omega(1)$ and $\text{rate}(C) = \frac{1}{\text{polylog}(n)}$.*

An observation made in [42] was that a relaxed LTC with sub-constant second soundness parameter can be easily converted to a strong LTC with a constant soundness.

**Observation A.19** (A conversion of rLTCs to strong LTCs)**.** *Let $q \geq 2$ and $C \subseteq \mathbb{F}^n$ be a linear $(q, \epsilon_1, \epsilon_2)$-rLTC with a core $A(C)$. Then there exists a linear $(q, \epsilon_1/6)$-strong LTC $C' \subseteq \mathbb{F}^{n'}$, where $n \leq n' \leq \frac{12}{\epsilon_2} \cdot n$, $\dim(C') = \dim(C)$, $\mathrm{rate}(C') \geq \frac{\epsilon_2}{12} \cdot \mathrm{rate}(C)$ and $\delta(C') \geq 0.9 \cdot \delta(C|_{A(C)})$. Moreover, the construction of $C'$ from $C$ is explicit and done in time $O(n')$.*

Based on Observation A.19, [42] proved the following corollary that will play a crucial role in the proof of Theorem 1.5.

**Corollary A.20.** *Assume that for constants $q \geq 2, \epsilon > 0$, field $\mathbb{F}$ and infinitely many $n \in \mathbb{N}^+$ we have a linear code $C \subseteq \mathbb{F}^n$ with a core $A(C)$ such that $C$ is a $(q, \epsilon, \frac{1}{\mathrm{polylog}(n)})$-rLTC, $\delta(C|_{A(C)}) = \Omega(1)$ and $\mathrm{rate}(C) = \frac{1}{\mathrm{polylog}(n)}$. Then, there exists $C' \subseteq \mathbb{F}^{n'}$ such that $n \leq n' \leq n \cdot \mathrm{polylog}(n)$, $C'$ is a $(q, \epsilon/6)$-strong LTC, $\delta(C') = \Omega(1)$ and $\mathrm{rate}(C') = \frac{1}{\mathrm{polylog}(n')}$. Moreover, $C'$ is constructed explicitly from $C$.*

Theorem 1.5 follows from Corollary A.18 and Corollary A.20.

## B A building block for robustness result in [11]

To present more detailed explanation, let us recall the definition of the Reed-Solomon codes.

**Definition B.1** (Reed-Solomon codes)**.** Let $K$ denote a finite field, let $S \subseteq K$ and let $d < |S|$ denote a natural number. The Reed-Solomon code $RS_{K,S,d} : K^{d+1} \to K^{|S|}$ is defined as follows: Suppose we wish to encode a message $a \in K^{d+1}$ with $RS_{K,S,d}$. We define the polynomial $P_a(X) = \sum_{i=0}^{d} a_i X^i$, and set the codeword $RS_{K,S,d}(a)$ to consist of the evaluations of $P_a$ at each of the elements of $S$. The relative distance of $RS_{K,S,d}$ is $1 - \frac{d+1}{|S|}$ (see [38, Lecture 4]).

The work of Ben-Sasson and Sudan [11] used as one of the building blocks the following result of Polishchuk and Spielman [36].

**Theorem B.2** (Theorem 9 in [36])**.** *Let $\mathbb{F}$ be a field, let $X = \{x_l, ..., x_n\} \subseteq \mathbb{F}$, and let $Y = \{y_l, ..., y_n\} \subseteq \mathbb{F}$. Let $R(x, y)$ be a polynomial over $\mathbb{F}$ of degree $(d, n)$ and let $C(x, y)$ be a polynomial over $\mathbb{F}$ of degree $(n, d)$. If*

$$\Pr_{(x,y)\in X\times Y}[R(x, y) \neq C(x, y)] < \delta^2,$$

*and $n > 2\delta n + 2d$, then there exists a polynomial $Q(x, y)$ of degree $(d, d)$ such that*

$$\Pr_{(x,y)\in X\times Y}\left[R(x, y) \neq Q(x, y) \quad \text{or} \quad C(x, y) \neq Q(x, y) < 2\delta^2.\right]$$

This result was one of the main technical ingredients which allowed a composition in the work of Ben-Sasson and Sudan [11]. Informally, it says that given a word, which is candidate to be tensoring of two single polynomials, if a typical "row" is close to $d$-degree polynomial, and a typical "column" is close to $d$-degree polynomial, then the candidate word is close to be the tensoring of two $d$-degree polynomials. That means that the tensor product of RS codes can be tested using the row/column verifier.

**Main Result of [11].** In [11] it was shown that certain Reed-Solomon codes can provide weak LTCs (using repetitions of some coordinates). More precisely, they showed the following result.

**Theorem B.3** (Theorem 4 in [11])**.** *Let $K = GF(2^l)$ and let $L \subseteq K$ be a $GF(2)$-linear subspace of $K$. Then for any $d < |L|$ the code $RS_{K,L,d}$ is a PCP of proximity with query complexity $O(1)$, rejection ratio $1/\text{poly}(\log |L|)$, randomness complexity $\log |L| + O(\log \log |L|)$ and proof length $|L| \cdot \text{poly}(\log |L|)$.*

We note that PCP of proximity gives immediately a weak LTC with similar parameters range by repetition of code coordinates a number of times [11].

The parameters range we are interested in is obtained by choosing $d = O(|L|)$, so one gets a PCP of proximity with constant query complexity and inverse poly-log rejection ratio, with an alphabet of a super-constant size (since a Reed-Solomon code of block length $n$ must be over an alphabet of size at least $n$).

Although the construction of [11] is done over a large field ($|\mathbb{F}| = O(n)$), it can be reduced to the field of constant size, where the blocklength is increased by poly-log factor and the soundness parameter is decreased by poly-log factor (see the folklore Theorem C.1). This preserves the strong LTC to have inverse poly-log rate and soundness.

## C Auxiliary statements

**Theorem C.1** (Folklore)**.** *Let $C \subseteq \mathbb{F}_{p^t}^n$ be an explicit linear $(q, \epsilon)$-strong LTC, where $q \geq 3$. Then the code $C' \subseteq \mathbb{F}_p^{tn}$ can be explicitly constructed from $C$ such that*

- *$C'$ is a $(q, \epsilon')$-strong LTC, where $\epsilon' = \Omega(\epsilon/t)$*
- $\delta(C') = \Omega(\delta(C))$
- $\text{rate}(C') \geq \Omega(\text{rate}(C)/t)$

*Proof Sketch:* The code $C'$ is constructed from $C$ by encoding every element of $\mathbb{F}_{p^t}$ by an LDPC code $R \subseteq \mathbb{F}_p^{O(t)}$, i.e., by $O(t)$ elements of $\mathbb{F}_p$, such that $\dim(R) = t$, and the blocklength of $R$ is $O(t)$. Since $R$ is an $LDPC$ it has a set $U = \{u_1, \ldots, u_{b-r}\} \subseteq R^{\perp}_{\leq q}$ such that $\text{span}(U) = R^{\perp}$. The tester for $R$ on the input word $w$ picks a random $u \in U$ and accepts if and only if $\langle w, u \rangle = 0$. It is not hard to see that the soundness of this tester is at least $1/|U| \geq 1/b$ and its query complexity is at most $q$. The tester of $C'$ combines the testers of $R$ and $C$. □