Electronic Colloquium on Computational Complexity, Report No. 22 (2013)

Strong LTCs with inverse poly-log rate and constant soundness

Michael Viderman*
Computer Science Department
Technion — Israel Institute of Technology
Haifa, 32000, Israel.
viderman@cs.technion.ac.il

February 5, 2013

Abstract

An error-correcting code $C \subseteq \mathbb{F}^n$ is called (q, ϵ)-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 ϵ is called soundness.

In this paper we solve an open question raised by Goldreich and Sudan (J.ACM 2006) and construct binary linear strong LTCs with query complexity 3, constant relative distance, constant soundness and inverse polylogarithmic rate.

Our result is based on the previous paper of the author (Viderman, ECCC TR12-168), which presented binary linear strong LTCs with query complexity 3, constant relative distance, and inverse polylogarithmic soundness and rate. We show that the "gap amplification" procedure of Dinur (J.ACM 2007) can be used to amplify the soundness of these strong LTCs from inverse polylogarithmic up to a constant, while preserving the other parameters of these codes.

*The research has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 259426.

ISSN 1433-8092

1 Introduction

To explain the results of this work, we start from the necessary definitions. For the introduction we refer a reader to the previous work of the author [9].

A code over a finite alphabet Σ 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 $\text{blocklength}(\mathcal{C})$. 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 Σ is defined to be $\dim(\mathcal{C}) = \log_{|\Sigma|} |\mathcal{C}|$.

The rate of a code $\mathcal{C}$, denoted by $\text{rate}(\mathcal{C})$, 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 $\Delta(\mathcal{C}) = \min_{x \neq y \in \mathcal{C}} \Delta(x, y)$ and its relative distance is defined by $\delta(\mathcal{C}) = \frac{\Delta(\mathcal{C})}{n}$. For $x \in \mathbb{F}^n$ and $\mathcal{C} \subseteq \mathbb{F}^n$, let $\delta(x, \mathcal{C}) = \min_{y \in \mathcal{C}} \{\delta(x, y)\}$ denote the relative distance of x from the code $\mathcal{C}$. If $\delta(x, \mathcal{C}) \geq \rho$, we say that x is ρ-far from $\mathcal{C}$ and otherwise x is ρ-close to $\mathcal{C}$.

Let $[n]$ be the set $\{1, \ldots, n\}$. For $w \in \mathbb{F}^n$, let $\text{supp}(w) = \{i \in [n] \mid w_i \neq 0\}$ and $|w| = |\text{supp}(w)|$. 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 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 $\text{supp}(\mathcal{D}) = \{I \subseteq [n] \mid \mathcal{D}(I) > 0\}$. For $i \in [n]$ we let $N_{\mathcal{D}}(i) = \{I \in \text{supp}(\mathcal{D}) \mid i \in I\}$.

Now we define testers and LTCs (see [7, 9] 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, ϵ, ρ)-weak tester if for all $w \in \mathbb{F}^n$, $\delta(w, \mathcal{C}) \geq \rho$ we have $\mathop{\mathbf{Pr}}_{I \sim \mathcal{D}}[w|_I \notin \mathcal{C}|_I] \geq \epsilon$. A q-query tester $\mathcal{D}$ is a (q, ϵ)-strong tester if for all $w \in \mathbb{F}^n$ we have $\mathop{\mathbf{Pr}}_{I \sim \mathcal{D}}[w|_I \notin \mathcal{C}|_I] \geq \epsilon \cdot \delta(w, \mathcal{C})$.

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

1.1 Main Result

In this paper we resolve the following question raised by Goldreich and Sudan [7].

Question 1.3 ([7]). *Are there exist constants $q \in \mathbb{N}^+$, $d, \epsilon, \gamma > 0$ and a constant size alphabet Σ such that for infinitely many $n \in \mathbb{N}^+$ we have a code $C \subseteq \Sigma^n$, where*

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

Although the requested range of parameters was achieved for the weak LTCs [4, 5, 8], strong LTCs with these parameters were not obtained and this question remained to be a basic open question in the area of LTCs.

Our main theorem (Theorem 1.4) answers positively on Question 1.3.

Theorem 1.4 (Main Theorem). *There exist constants $d, \epsilon, \gamma > 0$ such that for infinitely many $n \in \mathbb{N}^+$ we have a linear code $C \subseteq \mathbb{F}_2^n$, where*

- *C is a $(3, \epsilon)$-strong LTC,*
- *$\delta(C) \geq \gamma$ and $\text{rate}(C) \geq \frac{1}{\log^d n}$.*

We notice that 3 queries are necessary to test non-trivial linear codes [2].[1]

To prove Theorem 1.4, we first present the main observation of this work (Observation 2.4) and its corollary (Corollary 2.5) in Section 2. Then, in Section 3 we recall the main result of [9] and make some immediate corollaries. Finally, in Section 4 we show that a well known "gap amplification" technique of Dinur [5] and its version corresponding to linear codes [8] (see also [3]) can be applied to the codes of [9][2] to obtain relaxed LTCs (see Definition 2.2), which can be converted to the strong LTCs with a desired range of parameters using Corollary 2.5.

2 The Main Observation of this work

Before we present our main observation, we recall some concept used in [9].

Definition 2.1 (A core of the code). Let $C \subseteq \Sigma^n$ be a code. A core of the code C, denoted by $A(C)$, is a nonempty subset of $[n]$ such that if $A(C) \neq [n]$ then any assignment to the entries of $A(C)$ uniquely determines the entries of $[n] \setminus A(C)$ and vice versa. I.e., if $A(C) \neq [n]$ then for any $c \in C$ there is no $c' \in C$ such that $c|_{A(C)} = c'|_{A(C)}$ and $c|_{[n]\setminus A(C)} \neq c'|_{[n]\setminus A(C)}$, or $c|_{[n]\setminus A(C)} = c'|_{[n]\setminus A(C)}$ and $c|_{A(C)} \neq c'|_{A(C)}$.

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 $w, w' \in \Sigma^n$ we let $\delta(w|_{[n]\setminus A(C)}, w'|_{[n]\setminus A(C)}) = \delta(w|_{[n]\setminus A(C)}, C|_{[n]\setminus A(C)}) = 0$.

Our first novelty is the following concept of a relaxed LTC (rLTC).

[1] By "non-trivial" codes we mean codes with a constant relative distance and non-constant dimension.

[2] The codes presented in [9] were very similar to the codes of [8].

Definition 2.2 (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 $\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}[w|_I \notin C|_I] \geq \max\left\{\epsilon_1 \cdot \delta(w|_{A(C)}, c|_{A(C)}), \epsilon_2 \cdot \delta(w|_{-A(C)}, c|_{-A(C)})\right\}$. 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, ϵ_1 is called the first soundness parameter and ϵ_2 is called the second soundness parameter.

Intuitively, think that ϵ_1 is a constant, but ϵ_2 is sub-constant.

Remark 2.3. We note that if $C \subseteq \mathbb{F}^n$ is a (q, ϵ)-strong LTC and $\mathcal{D}$ is its tester, then setting $A(C) = [n]$ it holds that C is a $(q, \epsilon, 1)$-rLTC with regards to the same tester $\mathcal{D}$ because for every $w \in \mathbb{F}^n$ we have

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}[w|_I \notin C|_I] \geq \epsilon \cdot \delta(w, C) = \max\left\{\epsilon \cdot \delta(w|_{[n]}, C|_{[n]}), 1 \cdot 0\right\} =$$

$$= \max\left\{\epsilon \cdot \delta(w|_{A(C)}, C|_{A(C)}), 1 \cdot \delta(w|_{-A(C)}, C|_{-A(C)})\right\}.$$

Our main observation is that a relaxed LTC with sub-constant second soundness parameter can be easily converted to a strong LTC with a constant soundness.

Observation 2.4 (Main observation)**.** *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 $O(n')$.*

Proof. Let $\mathcal{D}$ be the corresponding tester for C. Without loss of generality assume that $A(C)$ is the first $|A(C)|$ indexes, i.e., $A(C) = [|A(C)|]$. If $A(C) = [n]$ then C is a (q, ϵ_1)-strong LTC with regards to the same tester $\mathcal{D}$ and we are done. Thus we assume for the rest of the proof that $A(C) \subsetneq [n]$ and thus $n - |A(C)| \neq 0$.

Let $h = \left\lceil \frac{10}{\epsilon_2} \cdot \frac{n}{|A(C)|} \right\rceil$ and note that $h \geq 1$. Then $\frac{10}{\epsilon_2} \cdot n \leq h \cdot |A(C)| \leq \frac{11}{\epsilon_2} \cdot n$. Now, let C' be a linear code obtained from C by concatenating the symbols of $C|_{A(C)}$ to C exactly h times. I.e., for every $c' \in C'$ we have $c' = (c, (\overbrace{c|_{A(C)}, c|_{A(C)}, \ldots, c|_{A(C)}}^{h}))$ for some $c \in C$. In particular, $C = C'|_{[n]}$ and for every $j \in [h]$ we have $C'|_{n+(j-1)\cdot|A(C)|+[|A(C)|]} = C|_{A(C)}$. We set $A(C') = A(C)$.

We notice that $\dim(C') = \dim(C)$ and the blocklength of C' is $n' = n + h \cdot |A(C)| \leq n + \frac{11}{\epsilon_2} \cdot n \leq \frac{12}{\epsilon_2} \cdot n$. It can be verified that $\delta(C') \geq \frac{(h+1)\cdot|A(C)|\cdot\delta(C|_{A(C)})}{n'} \geq 0.9 \cdot \delta(C|_{A(C)})$. Note also that $\mathrm{rate}(C') = \frac{\dim(C)}{n'} \geq \frac{\epsilon_2}{12} \cdot \mathrm{rate}(C)$.

Let $\mathcal{D}'$ be a tester for C' that on the input word $w \in \mathbb{F}^{(n')}$:

- picks $r \in \{0, 1\}$
- if $r = 1$ then picks random $j_1 \in A(C)$ and $j_2 \in [h]$ and outputs $\{j_1, n + (j_2 - 1) \cdot |A(C)| + j_1\}$
- otherwise ($r = 2$), samples $\mathcal{D}$ on $w|_{[n]}$ and returns its output

We argue that C' is a $(q, \epsilon_1/6)$-strong LTC with respect to its tester $\mathcal{D}'$. Clearly, $\mathcal{D}'$ is a q-query tester. Let $w \in \mathbb{F}^{n'}$ be an input word. We show that

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[w|_I \notin C'|_I\right] \geq \frac{\epsilon_1}{6} \cdot \delta(w, C').$$

Let β be the fraction of bits in $w|_{[n']\setminus[n]}$ that is not equal to the corresponding bits of $w|_{A(C)}$, i.e., $\beta = \mathop{\mathbf{Pr}}_{j_1\in A(C), j_2\in[h]}\left[w|_{j_1} \neq w|_{n+(j_2-1)\cdot|A(C)|+j_1}\right]$. We know that

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[w|_I \notin \mathcal{C}'|_I\right] \geq \frac{\beta}{2} + \frac{\mathbf{Pr}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin C|_I\right]}{2}. \tag{1}$$

By assumption, C is a $(q, \epsilon_1, \epsilon_2)$-rLTC and hence there exists a codeword $c \in C$ such that $\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right] \geq \max\left\{\epsilon_1 \cdot \delta((w|_{[n]})|_{A(C)}, c|_{A(C)}), \epsilon_2 \cdot \delta((w|_{[n]})|_{-A(C)}, c|_{-A(C)})\right\}$. Therefore

$$\delta((w|_{[n]})|_{A(C)}, c|_{A(C)}) \leq \frac{\mathbf{Pr}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right]}{\epsilon_1} \tag{2}$$

and

$$\delta((w|_{[n]})|_{-A(C)}, c|_{-A(C)}) \leq \frac{\mathbf{Pr}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right]}{\epsilon_2}. \tag{3}$$

Let $c' \in C'$ be the corresponding codeword to c (i.e., $c'|_{[n]} = c$). Then (by (1), (2) and (3))

$$\begin{aligned}
\delta(w, c') &\leq \beta + \frac{\delta((w|_{[n]})|_{A(C)}, c|_{A(C)}) \cdot |A(C)| \cdot (h+1) + \delta((w|_{[n]})|_{-A(C)}, c|_{-A(C)}) \cdot (n - |A(C)|)}{n'} \leq \\
&\leq \beta + \frac{(1/\epsilon_1) \cdot |A(C)| \cdot (h+1) + (1/\epsilon_2) \cdot n}{n'} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right] \leq \\
&\leq \beta + \frac{(1/\epsilon_1) \cdot \frac{12}{\epsilon_2} \cdot n + (1/\epsilon_2) \cdot n}{n'} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right] \leq \\
&\leq \beta + \frac{(1/\epsilon_1) \cdot \frac{12}{\epsilon_2} \cdot n + (1/\epsilon_2) \cdot n}{(10/\epsilon_2) \cdot n} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right] \leq \\
&\leq \beta + \frac{3}{\epsilon_1} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin \mathcal{C}|_I\right] \leq \frac{6}{\epsilon_1} \cdot \left(\frac{\beta}{2} + \frac{\mathbf{Pr}_{I\sim\mathcal{D}}\left[(w|_{[n]})|_I \notin C|_I\right]}{2}\right) \leq \frac{6}{\epsilon_1} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[w|_I \notin C'|_I\right].
\end{aligned}$$

That means $\delta(w, c') \leq \frac{6}{\epsilon_1} \cdot \mathbf{Pr}_{I\sim\mathcal{D}'}[w|_I \notin C'|_I]$, where $c' \in C'$. We conclude that

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[w|_I \notin \mathcal{C}'|_I\right] \geq \frac{\epsilon_1}{6} \cdot \delta(w, C').$$

This proves that C' is a $(q, \epsilon_1/6)$-strong LTC with respect to its tester $\mathcal{D}'$, and completes the proof of Observation 2.4. □

Although Observation 2.4 might seem naive, it implies the following corollary that will play a crucial role in the proof of Theorem 1.4.

Corollary 2.5. *Assume that for constants $q \geq 2, \epsilon > 0$ and infinitely many $n \in \mathbb{N}^+$ we have a linear code $C \subseteq \mathbb{F}_2^n$ 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 $\text{rate}(C) = \frac{1}{\text{polylog}(n)}$. Then, there exists $C' \subseteq \mathbb{F}_2^{n'}$ such that $n \leq n' \leq n \cdot \text{polylog}(n)$, C' is a $(q, \epsilon/6)$-strong LTC, $\delta(C') = \Omega(1)$ and $\text{rate}(C') = \frac{1}{\text{polylog}(n')}$ (i.e., Question 1.3 is solved).*

Proof. The construction of the required C' follows immediately from Observation 2.4. □

3 The main result of [9] and its corollaries

In this section we recall the main result of [9]. Then we make some corollaries that will be used later.

Theorem 3.1 ([9]). *For some constant $d \in \mathbb{N}^+$ and infinitely many $n \in \mathbb{N}^+$ there exists a linear code $C \subseteq \mathbb{F}_2^n$ and its tester $\mathcal{D}$ such that*

- *C is a $(3, \frac{1}{\log^d n})$-strong LTC with respect to $\mathcal{D}$,*
- *$\delta(C) = \Omega(1)$,*
- *$\mathrm{rate}(C) = \frac{1}{\log^d n}$,*
- *$|\mathrm{supp}(\mathcal{D})| \leq n \log^d n$ and for every $u \in \mathrm{supp}(\mathcal{D})$ it holds that $\mathcal{D}(u) \leq \frac{\log^d n}{n}$, and*
- *for every $i \in [n]$ we have $|N_{\mathcal{D}}(i)| \leq \log^d n$.*

Remark 3.2. Although in [9] two last bullets were not proved, but one could verify that these bullets hold. The construction in [9] begins from a constant blocklength code C_1 and contained 3 procedures: the star product, the random projection and the distance amplification. These 3 procedures were applied iteratively $\Theta(\log \log n)$ times. Each iteration i is executed on the code C_i that had a blocklength n_i and a tester $\mathcal{D}_i$. The output of each iteration i is the code C_{i+1}.

Initially, the base code $C_1 \subseteq \mathbb{F}_2^{n_1}$ and its tester $\mathcal{D}_1$ satisfied the last two bullets with respect to its blocklength $n_1 = O(1)$. Each iteration, the star product, the random projection and the distance amplification procedures were applied. The random projection does not affect the properties listed in these bullets, but only rearranges the coordinates of the given code in some way. The star product and the distance amplification procedure do affect the properties listed in these bullets, but only by fixed multiplicative constants. E.g., each time the star product and the distance amplification are applied, $\frac{|\mathrm{supp}(\mathcal{D}_i)|}{n_i}$ is increased by a fixed multiplicative constant. If for every $u \in \mathrm{supp}(\mathcal{D}_i)$ it holds that $\mathcal{D}_i(u) \leq \frac{h}{n_i}$, then after the both operations are applied on C_i, for every $u \in \mathrm{supp}(\mathcal{D}_{i+1})$ we have $\mathcal{D}_{i+1}(u) \leq \frac{h \cdot c}{n_{i+1}}$ for a fixed constant c (independent of i). Similarly, $|N_{\mathcal{D}_i}(\cdot)|$ is increased at most by a fixed multiplicative constant each iteration by both procedures. Therefore, after $\Theta(\log \log n)$ iterations the values $\frac{|\mathrm{supp}(\mathcal{D}_i)|}{n_i}$ and $|N_{\mathcal{D}_i}(\cdot)|$ are changed at most by a polylog(n) factor.

We pay attention that one can turn the strong LTCs of Theorem 3.1 to the strong LTCs with a uniform distribution over the tests, and the soundness parameter, roughly speaking, will be preserved.

Corollary 3.3. *For some constant $d \in \mathbb{N}^+$ and infinitely many $n \in \mathbb{N}^+$ there exist a linear code $C' \subseteq \mathbb{F}_2^n$ and its tester $\mathcal{D}'$ which is a uniform distribution over $\mathrm{supp}(\mathcal{D}')$ such that*

- *C' is a $(3, \frac{1}{\log^d n})$-strong LTC with respect to $\mathcal{D}'$,*
- *$\delta(C') = \Omega(1)$,*
- *$\mathrm{rate}(C') \geq \frac{1}{\log^d n}$,*
- *$|\mathrm{supp}(\mathcal{D}')| \leq n \log^d n$, and*

- *for every* $i \in [n]$ *we have* $|N_{\mathcal{D}'}(i)| \le \log^d n$.

Proof. Corollary 3.3 follows from Theorem 3.1 by letting $C' = C$ and $\mathcal{D}'$ be a uniform distribution over $\mathrm{supp}(\mathcal{D})$. Note that $\mathrm{supp}(\mathcal{D}') = \mathrm{supp}(\mathcal{D})$ and for every $u \in \mathrm{supp}(\mathcal{D}')$ we have

$$\mathcal{D}'(u) = \frac{1}{|\mathrm{supp}(\mathcal{D})|} \ge \frac{1}{n \log^d n} = \frac{1}{\log^{2d} n} \cdot \frac{\log^d n}{n} \ge \frac{1}{\log^{2d} n} \cdot \mathcal{D}(u).$$

Then, for any $w \in \mathbb{F}_2^n$ we have

$$\mathop{\mathbf{Pr}}_{u \sim \mathcal{D}'}[\langle u, w\rangle \ne 0] \ge \frac{1}{\log^{2d} n} \cdot \mathop{\mathbf{Pr}}_{u \sim \mathcal{D}}[\langle u, w\rangle \ne 0] \ge \frac{1}{\log^{2d} n} \cdot \frac{1}{\log^d n} \cdot \delta(w, C) = \frac{1}{\log^{3d} n} \cdot \delta(w, C).$$

□

Now, in Corollary 3.4 we show that the 3-query strong LTCs over $\mathbb{F}_2$ from Corollary 3.3 can be easily converted to the 2-query rLTCs over $\mathbb{F}_2^3$ with a similar range of parameters.[3] This conversion is standard (for the case of LTCs, PCPs and assignment testers) and was explained, e.g., in [5, 8].

Corollary 3.4. *For some constant* $d \in \mathbb{N}^+$ *and infinitely many* $n'' \in \mathbb{N}^+$ *there exist a code* $C'' \subseteq (\mathbb{F}_2^3)^{(n'')}$ *and its tester* $\mathcal{D}''$ *which is a uniform distribution over* $\mathrm{supp}(\mathcal{D}'')$ *such that*

- C'' *is linear over* $\mathbb{F}_2$,
- C'' *is a* $(2, \frac{1}{3\log^d n''}, \frac{1}{\log^{3d} n''})$*-rLTC with respect to its core* $A(C'')$ *and* $\mathcal{D}''$,
- $\delta(C''|_{A(C'')}) = \Omega(1)$,
- $\mathrm{rate}(C'') \ge \frac{1}{\log^d n''}$,
- $|\mathrm{supp}(\mathcal{D}'')| \le 3 \cdot n''$, *and*
- *for every* $i \in [n'']$ *we have* $|N_{\mathcal{D}''}(i)| \le \log^d n''$.

Proof. Corollary 3.3 implies the existence of a constant $d \in \mathbb{N}^+$, a linear code $C' \subseteq \mathbb{F}_2^n$ for arbitrary large n and its tester $\mathcal{D}'$ such that C' is a $(3, \frac{1}{\log^d n})$-strong LTC with respect to $\mathcal{D}'$, $\delta(C') = \Omega(1)$, $\mathrm{rate}(C') \ge \frac{1}{\log^d n}$, $|\mathrm{supp}(\mathcal{D}')| \le n \log^d n$, and for every $i \in [n]$ we have $|N_{\mathcal{D}'}(i)| \le \log^d n$.

Let every element of $\mathbb{F}_2^3$ be associated with a tuple of 3 bits. Let also every bit (an element of $\mathbb{F}_2$) be viewed as an element of $\mathbb{F}_2^3$. Let us create a separate code symbol $X_{(i_1,i_2,i_3)}$ for every 3-query test $\{i_1, i_2, i_3\}$ of the original code (similarly, we create $X_{(i_1,i_2)}$ and $X_{(i_1)}$ for every 2-query and 1-query tests). Call the new code C''. A new tester $\mathcal{D}''$ for C'' samples an original tester $\mathcal{D}'$, and if $\mathcal{D}'$ queries i_1, i_2, i_3, then the tester $\mathcal{D}''$ queries the entry $X_{(i_1,i_2,i_3)}$ and a random coordinate $i_j \in \{i_1, i_2, i_3\}$, and accepts iff three bits of $X_{(i_1,i_2,i_3)}$ are summed to 0 and the corresponding bit of $X_{(i_1,i_2,i_3)}$ is equal to i_j.

We let $A(C'') = [n]$ and note that $C'' \subseteq (\mathbb{F}_2^3)^{(n'')}$. One can easily verify that C'' is linear over $\mathbb{F}_2$, $\delta(C''|_{A(C'')}) = \delta(C') = \Omega(1)$, $\mathrm{rate}(C'') \ge \frac{1}{2\log^{2d} n''}$, $|\mathrm{supp}(\mathcal{D}'')| \le 3 \cdot n''$, where $n'' \le n + n \cdot \log^d n$. Moreover, for every $i \in [n'']$ we have $|N_{\mathcal{D}''}(i)| \le \log^d n''$.

[3] During this paper we associate $\mathbb{F}_2^3$ with $\mathbb{F}_{2^3}$.

We prove that C'' is a $(2, \frac{1}{3\log^d n''}, \frac{1}{\log^{2d} n''})$-rLTC with respect to its core $A(C'')$ and $\mathcal{D}''$. Let $w \in (\mathbb{F}_2^3)^{(n'')}$. Since C' is a $(3, \frac{1}{\log^d n})$-strong LTC with respect to $\mathcal{D}'$, there exists $c' \in C'$ such that

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[(w|_{[n]})|_I \notin C'|_I\right] \geq \frac{1}{\log^d n} \cdot \delta(w|_{[n]}, c').$$

Let $c'' \in C''$ be a corresponding codeword to c', i.e., $c''|_{[n]} = c''|_{A(C'')} = c'$. Then,

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right] \geq \frac{1}{3} \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}'}\left[(w|_{[n]})|_I \notin C'|_I\right] \geq \frac{1}{3\log^d n} \cdot \delta(w|_{[n]}, c') = \frac{1}{3\log^d n} \cdot \delta(w|_{A(C'')}, c''|_{A(C'')}).$$

On the other hand,

$$\delta(w|_{-A(C'')}, c''|_{-A(C'')}) \leq \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right] + (\log^d n) \cdot \delta(w|_{[n]}, c''|_{[n]}) \leq$$

$$\leq \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right] + (\log^d n) \cdot (3\log^d n) \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right] \leq (\log^{3d} n) \cdot \mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right],$$

where the last inequality holds for sufficiently large n. Note that we used the fact that every entry indexed by $[n]$ of C'' affects at most $\log^d n$ entries indexed by $[n''] \setminus [n]$. Thus

$$\mathop{\mathbf{Pr}}_{I\sim\mathcal{D}''}\left[w|_I \notin C''|_I\right] \geq \max\left\{\frac{1}{3\log^d n} \cdot \delta(w|_{A(C'')}, c''|_{A(C'')}), \frac{1}{\log^{3d} n} \cdot \delta(w|_{-A(C'')}, c''|_{-A(C'')})\right\}.$$

□

4 Proof of Theorem 1.4

In this section we prove Theorem 1.4.

Dinur [5] suggested the gap amplification procedure to increase the rejection probability of verifiers for PCPs and explained that this procedure fits also for the assignment testers [6] (or alternatively, PCPs of proximity [1]). Then, Meir [8] explained that exploring the fact that the Hadamard code is 3-query strong LTC and 2-query locally correctable one can use it (instead of the PCP composition) to reduce the alphabet size in the composition stage of the gap amplification procedure. In this case the gap amplification procedure can be applied to the linear codes and preserve their linearity. Let us denote by Amplify$(\cdot)$ this version of the gap amplification procedure.

Our contribution here is that we observe that when Amplify$(\cdot)$ is invoked on a relaxed LTC, it improves the first soundness parameter, while does not destroy the second soundness parameter too much. First, we summarize the known affect of Amplify$(\cdot)$ on $\mathbb{F}_2$-linear codes with 2-query testers in Theorem 4.1. Before we state Theorem 4.1. We need to define some auxiliary concepts.

Let $C \subseteq \mathbb{F}^n$ be a code. Let $m \in [n]$ and $J = \{1, 2, \ldots, m\}$. Assume that given a message first of all we compute the values for all coordinates indexed by J, and then based on their values we compute the rest of the coordinates. That means for every $i \in [n] \setminus J$ there exists a function $g_i : \mathbb{F}^{d_i} \to \mathbb{F}$ and $j_1, j_2, \ldots, j_{d_i} \in J$ such that for every $c \in C$ it holds that $c|_i = g_i(c|_{j_1}, c|_{j_2}, \ldots, c|_{j_{d_i}})$. In this case we say that the index i depends only on $j_1, j_2, \ldots, j_{d_i}$. Also, for $k \in [d_i]$ we say that i depends on j_k. We say that i depends on at most d entries of J if $d_i \leq d$. In the given case, for any $k \in [d_i]$ we say that j_k affects i. We say that each entry in J of C affects at most h entries in $[n] \setminus J$ if for all $j \in J$ it holds that j affects at most h entries in $[n] \setminus J$.

Theorem 4.1 (Implicit in [5] and [8]). *Let $\mathbb{F} = \mathbb{F}_{2^3}$. There exist constants $h \in \mathbb{N}^+$ and $\gamma > 0$ such that the following holds. Let $C \subseteq \mathbb{F}^n$ be a code (linear over $\mathbb{F}_2$) with a 2-query tester $\mathcal{D}$ (uniform over its support) such that for every $i \in [n]$ we have $|N_{\mathcal{D}}(i)| \leq g$ for some $g > 0$ (g may depend on n). Then letting $C' = \text{Amplify(C)} \subseteq \mathbb{F}^{n'}$ and $\mathcal{D}'$ be its tester (uniform over its support), where $n \leq n' \leq (g \cdot h) \cdot n$ we have*

- *C' is a code (linear over $\mathbb{F}_2$) and $\text{supp}(\mathcal{D}') = O(g \cdot n')$,*
- *For all $i \in [n']$ we have $|N_{\mathcal{D}'}(i)| \leq h$,*
- *All old entries are preserved: $C'|_{[n]} = C$ and all new entries are computed from the old entries: $\dim(C') = \dim(C)$,*
- *Every entry indexed by $[n]$ affects at most $h \cdot g$ entries indexed by $[n'] \setminus [n]$ in the code C',*
- *All new entries are added as a sequence of blocks (small Hadamard codes) such that all these blocks are of the same constant size and are $(3, \frac{1}{2})$-strong LTCs,*
- *For every $w \in \mathbb{F}^{n'}$: if $\mathbf{Pr}_{I \sim \mathcal{D}}\left[(w|_{[n]})|_I \notin C|_I\right] \geq \epsilon$, then we have $\mathbf{Pr}_{I \sim \mathcal{D}'}[w|_I \notin C'|_I] \geq \min\{10 \cdot \epsilon, \gamma\}$.*
  (Note: a single iteration of gap amplification procedure can improve the rejection probability by any multiplicative constant.)

Now we conclude the following theorem that summarizes the affect of Amplify$(\cdot)$ on the relaxed LTCs.

Theorem 4.2 (Gap Amplification for relaxed LTCs). *Let $\mathbb{F} = \mathbb{F}_{2^3}$. There exist fixed constants $\gamma, d > 0$ such that the following holds. Let $C \subseteq \mathbb{F}^n$ be a $(2, \epsilon_1, \epsilon_2)$-rLTC (linear over $\mathbb{F}_2$), where $\epsilon_2 \leq \epsilon_1$, with regards to its core $A(C)$ and its tester $\mathcal{D}$ (uniform over its support), such that for every $i \in [n]$ we have $|N_{\mathcal{D}}(i)| \leq g$ for some $g > 0$ (g may depend on n). Then letting $C' = \text{Amplify(C)}$ and $A(C') = A(C)$ we have*

- *$C' \subseteq \mathbb{F}^{n'}$ is a $(2, \min\{2\epsilon_1, \gamma\}, \frac{\epsilon_2}{3d \cdot g})$-rLTC with regards to its core $A(C')$ and its new tester $\mathcal{D}''$ (uniform over its support),*
- *for all $i \in [n']$ we have $|N_{\mathcal{D}''}(i)| \leq d$,*
- *$C'|_{[n]} = C$ and thus $\delta(C'|_{A(C')}) = \delta(C|_{A(C)})$, and*
- *$\text{rate}(C') = \Omega(g \cdot \text{rate}(C))$.*

Proof. Theorem 4.1 says that $C'|_{[n]} = C$. The fact that $A(C') = A(C)$ implies that $\delta(C'|_{A(C')}) = \delta(C|_{A(C)})$. Theorem 4.1 also claims that $\dim(C') = \dim(C)$ and $n \leq n' \leq (g \cdot h) \cdot n$ for some constant $h \in \mathbb{N}^+$. Hence $\text{rate}(C') = \Omega(g \cdot \text{rate}(C))$.

Let $w' \in \mathbb{F}^{n'}$ and $w = w'|_{[n]} \in \mathbb{F}^n$. We know that there exists $c \in C$ such that letting $\hat{\epsilon} = \max\left\{\epsilon_1 \cdot \delta(w|_{A(C)}, c|_{A(C)}), \epsilon_2 \cdot \delta(w|_{-A(C)}, c|_{-A(C)})\right\}$ we have

$$\mathbf{Pr}_{I \sim \mathcal{D}}[w|_I \notin C|_I] \geq \hat{\epsilon}.$$

Let $c' = \text{Amplify(c)}$, i.e., the codeword $c' \in C'$ is produced from c by Amplify$(\cdot)$. Note that $c'|_{[n]} = c$. Theorem 4.1 implies that

$$\Pr_{I\sim\mathcal{D}'}\left[w'|_I \notin C'|_I\right] \geq \min\left\{\gamma, 10\cdot\hat{\epsilon}\right\} \geq \min\left\{\gamma, 10\epsilon_1\cdot\delta(w'|_{A(C')}, c|_{A(C')})\right\},$$

where $\mathcal{D}'$ is a tester for C' guaranteed by Theorem 4.1.

Note that all new coordinates that are added form a sequence of the Hadamard blocks of the equal constant size, and each old entry (indexed by $[n]$) affects at most $d\cdot g$ new entries (indexed by $[n']\setminus[n]$). Thus we can define a new tester $\mathcal{D}''$ for C' that invokes the original tester $\mathcal{D}'$ with probability $\frac{1}{2}$, and otherwise pick a random Hadamard block and test it. One can verify that for all $i\in[n']$ we have $|N_{\mathcal{D}''}(i)|\leq d$, and that $\mathcal{D}''$ is uniform over its support (by considering its support as a multiset).

We claim that

$$\Pr_{I\sim\mathcal{D}''}\left[w'|_I \notin C'|_I\right] \geq \frac{1}{2}\cdot\Pr_{I\sim\mathcal{D}'}\left[w'|_I \notin C'|_I\right] \geq \min\left\{\frac{1}{2}\cdot\gamma, 2\epsilon_1\cdot\delta(w'|_{A(C')}, c'|_{A(C')})\right\}$$

and

$$\Pr_{I\sim\mathcal{D}''}\left[w'|_I \notin C'|_I\right] \geq \frac{\epsilon_2}{3d\cdot g}\cdot\delta(w'|_{[n']\setminus A(C')}, c'|_{[n']\setminus A(C')}).$$

The last inequality holds due to the guarantee of the tester $\mathcal{D}'$, and due to the fact that the Hadamard code has soundness parameter $\frac{1}{2}$. Thus C' is a $(2, \min\{2\epsilon_1, \gamma'\}, \frac{\epsilon_2}{3d\cdot g})$-rLTC, where $\gamma' = \gamma/2$, with regards to its core $A(C')$ and its tester $\mathcal{D}''$. □

We are ready to prove Theorem 1.4.

Proof of Theorem 1.4. Let $d\in\mathbb{N}^+$ and $C_1\subseteq(\mathbb{F}_2^3)^n$ be a code from Corollary 3.4. Now we execute the following algorithm.

- For each $i = 1,\ldots,d\cdot\log\log n$ do
  - $C_{i+1} := \text{Amplify(C}_\text{i})$
  - $A(C_{i+1}) := A(C_i) := [n]$

Theorem 4.2 implies that if C_i is a $(2, \frac{2^i}{\log^d n}, \alpha)$-rLTC then C_{i+1} is a $(3, \frac{2^{i+1}}{\log^d n}, \frac{\alpha}{b})$-rLTC for a fixed constant $b > 0$ (besides the first iteration, where b might be polylog(n)). Moreover, $\delta(C_{i+1}|_{A(C_{i+1})}) = \delta(C_i|_{A(C_i)}) = \delta(C)$, $\dim(C_{i+1}) = \dim(C_i)$ and $\text{rate}(C_{i+1}) \geq \beta\cdot\text{rate}(C_i)$ for some constant β (besides the first iteration, where $\text{rate}(C_2) \geq \frac{1}{\text{polylog}(n)}\cdot\text{rate}(C_1)$).

Let $C' = C_{d\cdot\log\log n}$ and $A(C') = A(C_{d\cdot\log\log n})$. Then C' is a $(2,\gamma,\frac{1}{\text{polylog}(n)})$-rLTC, $\text{rate}(C') = \frac{1}{\text{polylog}(n)}$ and $\delta(C'|_{A(C')}) \geq \Omega(1)$.

We claim that C' can be obtained as binary linear $(3,\gamma)$-strong LTC. Observe that the alphabet reduction stage in the gap amplification procedure is done by the encoding every node's assignment of the underlying graph by the binary Hadamard code which is a binary linear 3-query strong LTC. Thus if we don't convert this C' to the 2-query LTC over $\mathbb{F}_2^3$, it will stay binary linear strong LTC (for more details see [8]).

Corollary 2.5 implies the required construction of binary linear $(3,\gamma/6)$-strong LTCs from C'. □

References

[1] 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.

[2] Eli Ben-Sasson, Oded Goldreich, and Madhu Sudan. Bounds on 2-Query Codeword Testing. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization (APPROX-RANDOM)*, volume 2764 of *Lecture Notes in Computer Science*, pages 216–227. Springer, 2003.

[3] Eli Ben-Sasson, Prahladh Harsha, Oded Lachish, and Arie Matsliah. Sound 3-Query PCPPs Are Long. *TOCT*, 1(2), 2009.

[4] Eli Ben-Sasson and Madhu Sudan. Simple PCPs with poly-log rate and query complexity. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC), Baltimore, MD, USA, May 22-24, 2005*, pages 266–275. ACM, 2005.

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

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

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

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

[9] Michael Viderman. Strong LTCs with inverse polylogarithmic rate and soundness. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:168, 2012.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de