Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 4 (2010)

Low Rate Is Insufficient for Local Testability

Eli Ben-Sasson* Michael Viderman
Computer Science Department
Technion — Israel Institute of Technology
Haifa, 32000, Israel.
{eli, viderman}@cs.technion.ac.il

February 16, 2010

Abstract

Locally testable codes are error-correcting codes for which membership of a given word in the code can be tested probabilistically by examining it in very few locations. A linear code $C \subseteq \mathbb{F}_2^n$ is called sparse if $\dim(C) = O(\log(n))$. We say that a code $C \subseteq \mathbb{F}_2^n$ is ϵ-biased if all nonzero codewords of C have relative weight in the range $(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)$, where ϵ may be a function of n.

Kaufman and Sudan [10] proved that for sparse linear codes with relative distance $\frac{1}{2} - n^{-\Omega(1)}$ are locally testable. Moreover, they showed that all sparse $n^{-\Omega(1)}$-biased linear codes are locally decodable. In particular sparse random codes are locally testable and are locally decodable with probability $1-o(1)$.

Kopparty and Saraf [11] conjectured that all sparse linear codes (even with a bad distance) are locally testable.

In this paper we refute this conjecture by showing that for every $d(n)$ ranging from $\omega(1)$ to $\Omega(n)$ there exists a family of codes $\left\{C^{(n)} \subset \mathbb{F}_2^n\right\}_{n\in\mathbb{Z}}$ with linear distance and $\dim(C^{(n)}) = \Theta(d(n))$ which are not locally testable (decodable).

Furthermore, we show that there exists a family of codes $\left\{C^{(n)} \subset \mathbb{F}_2^n\right\}_{n\in\mathbb{Z}}$ with bias $n^{-o(1)}$ and $\dim(C^{(n)}) = \log(n)$ which are not locally testable (decodable). This also shows that the results of Kaufman and Sudan [10] were surprisingly tight.

1 Introduction

Locally testable codes (LTCs) are error correcting codes for which distinguishing, when given oracle access to a purported word w, between the case that w is a codeword and the case that it is very far from all codewords, can be accomplished by a randomized algorithm, called a *tester*, which reads a constant amount of information from w.

On the other hand, *locally decodable codes* (LDCs) allow to recover each message entry with high probability by reading only a few entries of the codeword even if a constant fraction of it is adversely corrupted. Both families of error correcting codes are explicitly studied, see e.g. the survey [14].

Given a linear code $C \subseteq \mathbb{F}^n$, the dimension of $\mathcal{C}$, denoted by $\dim(\mathcal{C})$, is its dimension as a vector space and its distance, denoted by $\Delta(\mathcal{C})$, is the minimal Hamming distance between two closest different

*The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 240258. Research of both authors supported by grant number 2006104 by the US-Israel Binational Science Foundation and by grant number 679/06 by the Israeli Science Foundation.

ISSN 1433-8092

codewords. A linear code $C \subseteq \mathbb{F}_2^n$ is called *sparse* if its dimension is $O(\log(n))$. We say that a code C is ϵ-biased if all nonzero codewords of C have relative weight in the range $(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)$, where ϵ may be a function of n. Notice that the ϵ-bias property implies that the relative distance is at least $\frac{1}{2} - \epsilon$.

Kaufman and Sudan [10] showed that local testability and decodability exists in random sparse linear codes. They showed that for any constant $\gamma > 0$ all sparse linear codes with relative distance $\frac{1}{2} - n^{-\gamma}$ are (strongly) locally testable. They also showed that for any constant $\gamma > 0$ all sparse $n^{-\gamma}$-biased linear codes are locally decodable. [1]. In particular, sparse random linear codes have low-bias and hence are locally testable and decodable.

This result was later generalized by Kopparty and Saraf [12] to the problem known as "local list-decoding and testing in the high error regime" (see [12] for the definition and discussion of the problem), i.e., they proved that sparse $n^{-\Omega(1)}$-biased linear codes are locally testable and locally list-decodable even in the high error regime.

Kopparty and Saraf [11] conjectured that all sparse linear codes (even with a bad distance) are locally testable. In particular, this conjecture says that the result of Kaufman and Sudan could be extended to all sparse linear codes and the $n^{-\Omega(1)}$-bias requirement is not necessary.

We refute this conjecture in Theorem 3.1 by showing that for any $d(n)$ ranging from $\omega(1)$ to $\Omega(1)$ there exists a family of linear codes $\left\{C^{(n)} \subset \mathbb{F}_2^n\right\}_{n\in\mathbb{Z}}$ with linear distance and $\dim(C^{(n)}) = \Theta(d(n))$ which are not locally testable (decodable).

Furthermore, in Theorem 3.4 we show that there exists a family of sparse $n^{-o(1)}$-biased linear codes which are not locally testable (decodable), i.e., for any computable function $h(n) = o(1)$ we can construct a family of sparse $n^{-h(n)}$-biased linear codes which are not locally testable (decodable).

In particular, this theorem shows that the $n^{-\Omega(1)}$-bias requirement in the work of Kaufman and Sudan [10] is necessary, and has a surprising tightness.

Organization of the paper. In the following section we provide the standard definitions regarding locally testable and locally decodable codes. In Section 3 we state our main results (Theorems 3.1, 3.4). Section 4 contains some useful propositions and in Section 5 we prove Theorems 3.1 and 3.4.

2 Preliminaries

Let $\mathbb{F}$ be a finite field and $[n]$ be the set $\{1, \ldots, n\}$. In this work, we consider only linear codes. We start with a few definitions.

Let $C \subseteq \mathbb{F}^n$ be a linear code over $\mathbb{F}$. For $w \in \mathbb{F}^n$, let $\mathrm{supp}(w) = \{i \in [n] \mid w_i \neq 0\}$ and $|w| = |\mathrm{supp}(w)|$. We define the relative distance between two words $x, y \in \mathbb{F}^n$ to be $\delta(x, y) = \frac{\Delta(x,y)}{n}$ and let $\Delta(x, y) = \delta(x, y) \cdot n$. The distance of a code is denoted by $\Delta(C)$ and defined to be the minimal value of $\Delta(x, y)$ for two distinct codewords $x, y \in C$. Similarly, the relative distance of the code is denoted $\delta(C) = \frac{\Delta(C)}{n}$. For $x \in \mathbb{F}^n$ and $C \subseteq \mathbb{F}^n$, let $\delta(x, C) = \delta_C(x) = \min_{y\in C} \{\delta(x, y)\}$ denote the relative distance of x from the code C. We note that $\Delta(C) = \min_{c\in C\setminus\{0\}} \{\mathrm{wt}(c)\}$. For two linear codes $C_1, C_2 \subseteq \mathbb{F}^n$ we let $\delta(C_1, C_2) = \min_{c_1\in C_1\setminus\{0\}} \{\delta(c_1, C_2)\}$.

If $\delta(x, C) \geq \epsilon$, we say that x is ϵ-far from C and otherwise x is ϵ-close to C. Let $\dim(C)$ be the dimension of C. 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 bilin-

[1] In fact, Kaufman and Sudan [10] proved a stronger result. They showed that sparse "low-bias" linear codes are self-correctable and thus are locally decodable.

ear 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 $C^\perp$ is defined as $C^\perp = \{u \in \mathbb{F}^n \mid \forall c \in C : \langle u, c\rangle = 0\}$. In a similar way we define $C^\perp_{\leq t} = \left\{u \in C^\perp \mid |u| \leq t\right\}$ and $C^\perp_t = \left\{u \in C^\perp \mid |u| = t\right\}$. For $w \in \mathbb{F}^n$ and $S = \{j_1, j_2, ..., j_m\} \subseteq [n]$ we let $w|_S = (w_{j_1}, w_{j_2}, ..., w_{j_m})$, where $j_1 < j_2 < \ldots < j_m$, be the restriction of w to the subset S. Similarly, we let $C|_S = \{c|_S \mid c \in C\}$ denote the projection of the code C onto S. We say that a code C has a q-characterization if $\mathrm{span}(C^\perp_{\leq q}) = C^\perp$.

For $w \in F^n$ and $i \in \mathbb{N}$ let $w^{(i)} \in \mathbb{F}^{ni}$ be the concatenation of w to itself i times.

2.1 Background on LTCs and LDCs

We formally define locally testable codes, using the definition provided in [4] (see that paper for a justification of the definition).

Note that given a code $C \subseteq F^n$, the subset $I \subseteq [n]$ uniquely defines $C|_I$. The linearity of C implies that $C|_I$ is a linear subspace of $\mathbb{F}^I$.

Definition 2.1 (LTCs and Testers)**.** Let $C \subseteq \mathbb{F}^n$ be a linear code. A (q, ϵ, δ)-tester T for C is a distribution $\mathbf{D}$ over subsets $I \subseteq [n]$ such that $|I| \leq q$ and the following holds.

- For all $w \in \mathbb{F}^n$ such that $\delta(w, C) \geq \delta$ we have $\Pr_{I \sim \mathbf{D}}[w|_I \notin C|_I] \geq \epsilon$.

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

The tester outputs accept on the given word w whenever it selects a subset $I \subseteq [n]$ such that $w|_I \in C|_I$ and otherwise output reject. Notice that a tester for a linear code is non-adaptive [2].

We say that a family of codes $\left\{C^{(n)} \;\middle|\; n \in \mathbb{Z}\right\}$ is locally testable if there exist constants $q, \epsilon, \delta > 0$ such that for infinitely many n it holds that $C^{(n)} \subseteq \mathbb{F}^n$ is a (q, ϵ, δ)-LTC.

Remark 2.2. Usually we assume that $\delta \leq 1/3$ (see the discussion in [1] regarding this issue). Hence, we say that the family of codes is **not** locally testable if for every constants $q, \epsilon > 0$, large enough n and distribution $\mathbf{D}_n$ over subsets $I \subseteq [n]$, such that $|I| \leq q$ there is a word w with $\delta(w, C^{(n)}) \geq 1/3$ and $\Pr_{I \sim \mathbf{D}_n}[w|_I \notin C|_I] < \epsilon$.

Next we define locally decodable codes.

Definition 2.3 (LDCs and Decoders)**.** Let $C \subseteq \mathbb{F}_2^n$ and $E_C : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be its encoding function, i.e., $C = \left\{E_C(x) \mid x \in \mathbb{F}_2^k\right\}$. Then C is a (q, ϵ, δ)-LDC if there exists a randomized decoder ($\mathbb{D}$) such that:

- In every invocation, $\mathbb{D}$ makes at most q queries.
- For all $x \in \mathbb{F}^k$, $i \in [k]$ and $\hat{c} \in \mathbb{F}^n$ such that $\Delta(E_C(x), \hat{c}) \leq \delta n$ we have $\mathbf{Pr}\left[\mathbb{D}^{\hat{c}}[i] = x_i\right] \geq \frac{1}{2} + \epsilon$, i.e., with probability at least $\frac{1}{2} + \epsilon$ entry x_i will be recovered correctly.

Note that the definition implies that $\delta < \delta(C)/2$. We can assume without loss of generality that the decoder for a linear code is non-adaptive [5]. We say that a family of codes $\left\{C^{(n)} \;\middle|\; n \in \mathbb{Z}\right\}$ is locally decodable if there exist constants $q, \epsilon, \delta > 0$ such that for infinitely many n it holds that $C^{(n)} \subseteq \mathbb{F}^n$ is a (q, ϵ, δ)-LDC.

It is well-known that q-query self-correctable codes are q-query locally decodable, so if we show that a code C is not q-query LDC then it is not q-query self correctable.

Remark 2.4. Notice that a message space for LDC can be a linear subspace of $M \subset \mathbb{F}^k$, i.e., the messages are of the length k but not every word in $\mathbb{F}^k$ is a (legal) message. In this case the dimension of the code is $\dim(M)$. Note that because of linearity of C the message space M must be a linear subspace (closure to addition).

3 Main Results

In this section we present our main results, i.e., Theorem 3.1 and Theorem 3.4. Theorem 3.1 refutes the conjecture of Kopparty and Saraf [11], which says that all sparse linear codes are locally testable. It shows that there exists a family of sparse linear codes with relative distance ≥ 0.49 which are not locally testable (decodable).

One could conjecture that all sparse linear codes with small characterization are locally testable or decodable. So, Theorem 3.1 also shows that there are sparse linear codes which have a small characterization and linear distance, but are not locally testable (decodable).

Finally we show that there are sparse linear codes with dual distance > 2 which are not locally testable (decodable). We believe that this case is interesting since it shows that non-redundant[2] sparse code can be not locally testable (decodable). Equivalently, picking a large fraction of columns from the Hadamard generating matrix ($G \in F_2^{[n]\times[2^n]}$) may result in a generating matrix of a code which is *not* locally testable (decodable). This contrasts with the result of [10] which says that taking a large *random* fraction of columns from the Hadamard generating matrix gives, with high probability, a generating matrix of a locally testable (decodable) code.

Theorem 3.1 (Low Rate does not imply LTC or LDC)**.** *For every $q, \epsilon > 0$, function $w(1) \leq d(n) \leq O(n)$ and infinitely many $n > 0$*

1. *There exists $C_\epsilon \subset \mathbb{F}_2^n$ such that $\delta(C) \geq \frac{1}{2} - \epsilon$, $\dim(C_\epsilon) = \Theta(d(n))$ and C_ϵ is not locally testable with $o(d(n))$ queries and is not locally decodable with q queries.*
2. *There exists $C \subset \mathbb{F}_2^n$ such that $\dim(C) = \Theta(d(n))$, $\Delta(C) \geq \Omega(n)$, $\mathrm{span}(C^\perp_{\leq 3}) = C^\perp$ and C is not locally testable with $o(d(n))$ queries and is not locally decodable with q queries.*
3. *There exists $C \subset \mathbb{F}_2^n$ such that $\dim(C) = 1.1\log(n)$, $\Delta(C) \geq \Omega(n)$, $\Delta(C^\perp) > 2$ and C is not locally testable with $o(\log(n))$ queries and is not locally decodable with q queries.*

Remark 3.2. Folklore claim 7.2 states that every linear code C is testable by $\dim(C)+1$ queries. Intuitively, C is non-trivially testable if it can be testable with $o(\dim(C))$ queries. So, Theorem 3.1 shows the families of linear codes that can not be non-trivially testable.

Remark 3.3. The construction in the bullet three of Theorem 3.1 can not achieve dimension lower than $\log(n)$, since every code $C \subseteq \mathbb{F}_2^n$ such that $\dim(C) < \log(n)$ has $\Delta(C^\perp) \leq 2$.

The next theorem shows a surprising tightness of the results in [10]. Recall that Kaufman and Sudan showed that sparse linear codes with bias $n^{-\Omega(1)}$ are locally testable (decodable). We show that for every $h(n) = o(1)$ there exists a family of sparse linear codes with bias $n^{-h(n))}$ which are not locally testable (decodable).

[2]The term "non-redundant" here means the dual distance of the code is at least 3. Note that if the dual distance of a code is 1 then some of its bits are identically 0, and if the dual distance is 2 then some pairs of bits are equal to each other and hence one of each pair is redundant.

Theorem 3.4 (Tightness of [10])**.** *For every constant $q > 0$, computable function $h(n) = o(1)$ and infinitely many $n > 0$ there exists $C \subset \mathbb{F}_2^n$ such that $\dim(C) = \log(n)$, C is $n^{-h(n)}$-biased and C is not locally testable (decodable) with q queries.*

4 Repetition does not affect the LTCs and LDCs

4.1 Testers and Decoders modulo m

Definition 4.1 (Repetition Code)**.** Let $R \subseteq \mathbb{F}^m$ be a linear code and $i > 0$. We say that $C \subseteq \mathbb{F}^{(im)}$ is the i-repetition code of R if it holds that $c \in C$ if and only if $c = r^{(i)}$, where $r \in R$, i.e., every codeword of C is a codeword of R repeated i times.

Notice that the linearity of R implies the linearity of its repetition code.

Let $R \subset \mathbb{F}_2^m$ be a linear code and $C \subset \mathbb{F}_2^{(mi)}$ be its (i)-repetition code. For $I = \{i_1, i_2, \ldots, i_q\} \subseteq [n]$ let $I \bmod m = \{i_1 \bmod m, i_2 \bmod m, \ldots, i_q \bmod m\} \subseteq [m]$.

Given a tester $\mathbf{D}_C$ for C we define a (modulo-m)-tester $\mathbf{D}_R$ for R as follows.

- Pick a test $I \subseteq [mi]$ according to the tester $\mathbf{D}_C$.
- Choose test $I \bmod m$.

The modulo-m decoder is defined in the similar way, i.e., invokes the decoder of C and take modulo m over all queried indices.

Note that if $w \in \mathbb{F}_2^m$ then the tester (decoder) for C will get the same "queried entries" on $w^{(i)}$ as modulo-m tester (decoder) for R on w. In this way, if tester for C rejects then the tester for R rejects and if the decoder for C recovers correctly the message bit then the decoder for R recovers correctly the same message bit.

4.2 Main Propositions

We prove the simple (but important) propositions, which say that the repetition does not affect the testability and the decodability. These propositions (Proposition 4.2 and Proposition 4.3) are shown for binary codes but can be easily extended for any field.

Proposition 4.2. *Let $R \subset \mathbb{F}_2^m$ be a linear code and $i > 0$ be an integer. Let $C \subset \mathbb{F}_2^{(mi)}$ be an (i)-repetition code of R. Then,*

- *if R is a (q, ϵ, δ)-LTC then C is $(q, \min\{\epsilon/2, \delta/2\}, 2\delta)$-LTC*
- *if C is (q, ϵ, δ)-LTC then R is a (q, ϵ, δ)-LTC.*

Proof. Note that $C|_{[m]} = R$.

Proof of the first bullet. Assume that R is a (q, ϵ, δ)-LTC and let $\mathbf{D}_R$ be a (q, ϵ, δ)-tester for R. We define a tester $\mathbf{D}_C$ for C as following.

- flip a coin
- **If** "heads,"
  - pick $j \in [m]$ and $l_1 \in [i-1]$ independently at random,

– pick $I = \{j, j + m \cdot l_1\}$ (note that $I \subseteq [mi]$);

- **Else** pick $I \sim \mathbf{D}$ (note that $I \subseteq [m]$).

We argue that $\mathbf{D}_C$ is a $(q, \min\{\epsilon/2, \delta/2\}, 2\delta)$-tester for C. Let $w \in \mathbb{F}^{(mi)}$ be a word such that $\delta(w, C) \geq 2\delta$. If $\delta(w|_{[m]}, C|_{[m]}) = \delta(w|_{[m]}, R) \geq \delta$ we are done, since

$$\Pr_{I \sim \mathbf{D}_C}[w|_I \notin C|_I] \geq \frac{1}{2} \cdot \Pr_{I \sim \mathbf{D}_R}\left[(w|_{[m]})|_I \notin R|_I\right] \geq \frac{\epsilon}{2}.$$

Otherwise we have $\delta(w|_{[m]}, C|_{[m]}) = \delta(w|_{[m]}, R) < \delta$.
But $\delta(w, C) \geq 2\delta$ implies that

$$\mathbf{E}_{j \in [i-1]}\left[\delta(w|_{\{jm+1,\ldots,(j+1)m\}}, w|_{[m]})\right] \geq 2\delta - \delta = \delta.$$

Hence $\Pr_{I \sim \mathbf{D}_C}[w|_I \notin C|_I] \geq \frac{1}{2} \Pr_{j \in [m], l_1 \in [i-1]}\left[w|_{\{j, j+m \cdot l_1\}} \notin C|_{\{j, j+m \cdot l_1\}}\right] \geq \delta/2.$ □

Proof of the second bullet. Assume that C is a (q, ϵ, δ)-LTC and let $\mathbf{D}_C$ be its (q, ϵ, δ)-tester. Let $\mathbf{D}_R$ be a modulo-m tester of C. Note that $\mathbf{D}_R$ is a distribution over subsets $I \subseteq [m]$ such that $|I| \leq q$.

We argue that $\mathbf{D}_R$ is a (q, ϵ, δ)-tester for R. Let $w \in F_2^m$ be a word such that $\delta(w, R) \geq \delta$. Assume by way of contradiction that $\Pr_{I \sim \mathbf{D}_R}[w|_I \notin R|_I] < \epsilon$. Notice that $\delta(w^{(im)}, C) = \delta(w, R) \geq \delta$. We have $\Pr_{I \sim \mathbf{D}_C}\left[w^{(im)}|_I \in C|_I\right] < \epsilon$ since if for $I \subset [im]$ it holds that $w^{(im)}|_I \notin C|_I$ then $w|_{I \bmod m} \notin R$. We conclude that $\mathbf{D}_C$ is not a (q, ϵ, δ)-distribution for C. Contradiction. □

□

Proposition 4.3. *Let $R \subset \mathbb{F}_2^m$ be a code such that the first $\dim(R)$ bits of R are message bits. Let $i > 0$ and C be an (i)-repetition of R. Then,*

- *If R is not (q, ϵ, δ)-LDC then C is not (q, ϵ, δ)-LDC.*
- *If R is a (q, ϵ, δ)-LDC then C is a $(q, \epsilon/2, \frac{\epsilon\delta}{2})$-LDC.*

Proof. Let $k = \dim(R) = \dim(C)$. Note that the first k bits of C are message bits.

Proof of the first bullet. If R is not (q, ϵ, δ)-LDC then for every q-query decoder there exist a word $w \in \mathbb{F}_2^m$ such that $\delta(w, R) \leq \delta$, but there exists $i \in [k]$ such that the probability that the decoder recovers correctly the message bit i is less then $1 - \epsilon$.

Assume by way of contradiction that C is a (q, ϵ, δ)-LDC and has the decoder $\mathbb{D}_C$. Let $\mathbb{D}_R$ be a modulo-m decoder of $\mathbb{D}_C$. But then there exist $w \in \mathbb{F}^m, \delta(w, R) \leq \delta$ such that the decoder $\mathbb{D}_R$ recovers some message bit i with probability less then $1 - \epsilon$. But $\mathbb{D}_C$ will get always the same information on $w^{(n/m)}$ as $\mathbb{D}_R$ on w, and moreover, $\delta(w^{(n/m)}, C) = \delta(w, R) \leq \delta$. Thus $\mathbb{D}_C$ is not a (q, ϵ, δ)-decoder for C. Contradiction. □

Proof of the second bullet. Assume R is (q, ϵ, δ)-LDC and let D_R be its (q, ϵ, δ)-decoder. For $j \in [i]$ and $I = \{i_1, i_2, \ldots, i_q\} \subseteq [m]$ let $jm + I = \{jm + i_1, jm + i_2, \ldots, jm + i_q\}$. Note that $jm + I \subseteq \{mj + 1, \ldots, m(j + 1)\}$. The decoder D_C for C recovers message bit l from the given word $w \in \mathbb{F}^{im}$ as following.

- pick $j \in [i-1]$ and select $I = jm + [m]$,
- return $D_R^{(w|_I)}[l]$, i.e., recover the lth message bit as the decoder D_R on $w|_I$.

We argue that if $\delta(w, C) \leq \frac{\epsilon\delta}{2}$ then D_C recovers correctly the lth message bit of C with probability at least $\frac{1}{2} + \frac{\epsilon}{2}$. Let $r^{(i)} \in C$ be a closest codeword of C to w, i.e., $\delta(w, r^{(i)}) = \delta(w, C) \leq \frac{\epsilon\delta}{2}$.

For $j \in [i]$ we say that $w|_{(jm+[m])}$ is a j-block of w. We say that j-block is corrupted if $\delta(w|_{jm+[m]}, r) > \delta$. The fraction of corrupted blocks is bounded by $\frac{\epsilon}{2}$, because otherwise we have $\delta(w, r^{(i)}) > \frac{\epsilon\delta}{2}$. Recall that the decoder D_C for C picks random $j \in [i-1]$ and invokes the decoder for R on the j-block of w. The probability that D_R will be invoked on the non-corrupted block and will recover a message bit correctly is at least $(\frac{1}{2} + \epsilon) \cdot (1 - \frac{\epsilon}{2}) \geq \frac{1}{2} + \epsilon - \frac{\epsilon}{2} = \frac{1}{2} + \frac{\epsilon}{2}$. □

□

From the previous propositions we conclude the next corollary.

Corollary 4.4. *Let $R \subset \mathbb{F}_2^m$ be a linear code and $i > 0$ be an integer. Let $C \subset \mathbb{F}_2^{(mi)}$ be an (i)-repetition code of R. Then,*

- *if R is not (q, ϵ, δ)-LTC then C is not (q, ϵ, δ)-LTC.*
- *if R is not (q, ϵ, δ)-LDC then C is not (q, ϵ, δ)-LDC.*

5 Proof of Main Results (Theorems 3.1,3.4)

We first prove Theorem 3.1 in Section 5.1. Then we prove Theorem 3.4 in Section 5.2.

5.1 Proof of Theorem 3.1

Proof of the first bullet. Let $\epsilon > 0$ be a constant and $m = d(n) \geq w(1)$. Let $R_\epsilon \subset \mathbb{F}_2^m$ be a linear code such that $\delta(R_\epsilon) \geq \frac{1}{2} - \epsilon$ and $\delta(R_\epsilon{}^\perp) \geq \Omega(1)$ and $\dim(R_\epsilon) = \Theta(m)$ (e.g., a random linear code of constant rate will have these properties). Claim 7.1 implies that R_ϵ is not locally testable with $o(m)$ queries. Lower bounds on locally decodable codes from [9] imply that R_ϵ is not q-query locally decodable code.

Let $C_\epsilon \subset \mathbb{F}^n$ be a (n/m)-repetition code of R_ϵ. We have $\dim(C_\epsilon) = \Theta(m) = \Theta(\log(n))$ and $\delta(C_\epsilon) \geq 1/2 - \epsilon$. Furthermore, Corollary 4.4 implies that C_ϵ is not locally testable with $o(m)$ queries and not locally decodable with q queries. □

One could conjecture that sparse codes, which are characterized by small weight dual words and have linear distance are locally testable. We refute this next.

Proof of the second bullet. Let $m = d(n) \geq w(1)$. Then, for sufficiently large m, Theorem 7.3 implies the existence of a linear code $R \subset \mathbb{F}_2^m$ such that $\dim(R) = \Theta(d(n))$, $\Delta(R) \geq \Omega(m)$, $\mathrm{span}(R^\perp_{\leq 3}) = R^\perp$ and R is not locally testable with $o(m)$ queries and not locally decodable with q queries.

Let $C \subset \mathbb{F}_2^n$ be a (n/m)-repetition code of R. Then Proposition 4.2 implies that C is not locally testable with $o(m) = o(d(n))$ queries. Proposition 4.3 implies that C is not locally decodable with q queries. □

5.1.1 Proof of the Third Bullet

Given two linear codes with the same blocklength ($C_1, C_2 \subseteq \mathbb{F}^n$) let $C_1+C_2 = \{c_1 + c_2 \mid c_1 \in C_1, c_2 \in C_2\}$. Notice that $(C_1 + C_2)$ is a linear code.

We start from a straightforward proposition that will be useful in the next theorem.

Proposition 5.1. *Let C_1 and C_2 be two linear binary codes with the same blocklength. Then $(C_1 + C_2)^\perp = {C_1}^\perp \cap {C_2}^\perp$ and so $(C_1 + C_2)^\perp \subseteq {C_1}^\perp$ and $(C_1 + C_2)^\perp \subseteq {C_2}^\perp$.*

Proof. We have $u \in (C_1 + C_2)^\perp$ iff ($u \in {C_1}^\perp$ and $u \in {C_2}^\perp$) iff $u \in ({C_1}^\perp \cap {C_2}^\perp)$ □

It follows that if C_1 is a repetition code but C_2 is not then $C_1 + C_2$ is not repetition code. Moreover, if there is a "small-size" intersection between low-weight dual words of C_1 and of C_2 then $C_1 + C_2$ will have a small number of low-weight dual words and hence, intuitively will not be a locally testable.

Claim 5.2. *Let $C \subseteq \mathbb{F}^n$ be a linear code and a (q, ϵ, δ)-LDC. Assume that $C' \subset C$ is a linear code (subcode of C). Then C' is a (q, ϵ, δ)-LDC.*

Proof. Assume C has the (linear) message space $S \subseteq F^k$ and has the decoder $\mathbb{D}$. Let $S' \subset S$ be a (linear) message space for C'. We argue that C' has the same decoder $\mathbb{D}$. Let w be δ-close to C' (δ-close to the encoding of some message $m \in S'$). Then w is δ-close to C and thus for all $i \in [k]$ the decoder $\mathbb{D}$ recovers correctly the message entry (m_i) with probability at least $\frac{1}{2} + \epsilon$.

Notice that the message space S' of the subcode C' will have smaller dimension than S, i.e., $\dim(S') < \dim(S)$. S' is a linear vector space because for every two messages $x_1, x_2 \in S'$ which are encoded to $c_1, c_2 \in C'$, respectively, we have $(x_1 + x_2) \in S'$ and $(x_1 + x_2)$ is encoded to $c_1 + c_2$. □

Proposition 5.3. *Let $C_1, C_2 \subseteq \mathbb{F}_2^n$ be linear codes. If C_1 is not locally decodable with q queries then $C_1 + C_2$ is not locally decodable with q queries.*

Proof. If $C_1 + C_2$ is locally decodable with q queries then C_1 is locally decodable with q queries by Claim 5.2, because C_1 is a subcode of $C_1 + C_2$. □

Proof of the third bullet. Let $m = \log(n)$ and $R \subset \mathbb{F}_2^m$ be a linear code with $\Delta(R) \geq m/5$, $\Delta(R^\perp) \geq \Theta(m)$ and $\dim(R) = m/10$ (e.g., a random linear code of constant rate will have these properties). Claim 7.1 implies that R is not locally testable with $o(m)$ queries, and in particular, not locally testable with q queries.

Let $w \in \mathbb{F}_2^m$ be a word such that $\delta(w, R) \geq \frac{1.1}{3}$ (a random $w \in \mathbb{F}^m$ satisfies this condition with high probability). Notice that for all $u \in R^\perp_{\leq o(m)}$ we have $\langle u, w \rangle = 0$ because $R^\perp$ has no words of weight less then $o(m)$, i.e., $R^\perp_{\leq o(m)} = \emptyset$.

Let $C_1 \subset \mathbb{F}_2^n$ be a (n/m)-repetition code of R. Then $\delta(w^{(n/m)}, C_1) = \delta(w, R) \geq \frac{1.1}{3}$. Notice that by construction for all $u \in {C_1}^\perp_{\leq o(m)}$ we have $\langle u, w^{(n/m)} \rangle = 0$. We also have $\delta(C_1) \geq 1/5$ and $\dim(C_1) = \log(n)/10$. Let $C_2 \subseteq F_2^n$ be the Hadamard code[3] (assume w.l.o.g. that n is a power of 2). Note that $\Delta({C_2}^\perp) > 2$.

Let $\pi : [n] \mapsto [n]$ be a permutation. With some abuse of notation, for $w = (w_1, w_2, \ldots, w_n) \in \mathbb{F}^n$ let $\pi(w) = (w_{\pi(1)}, w_{\pi(2)}, \ldots, w_{\pi(n)})$ be a π-permuted word. Let $\pi(C_2) = \{\pi(c) \mid c \in C_2\}$ be a set of all permutated codewords of C_2. Note that for every permutation $\pi : [n] \mapsto [n]$ it holds that $\delta(\pi(C_2)) = \delta(C_2)$, $\dim(\pi(C_2)) = \dim(C_2)$ and $\Delta((\pi(C_2))^\perp) = \Delta({C_2}^\perp) > 2$.

[3] Instead of the Hadamard code we could take any binary, sparse code with linear distance and dual distance > 2.

Recall that $\delta(C_1, C_2) = \min_{c_1 \in C_1 \setminus \{0\}} \{\delta(c_1, C_2)\}$. We say that a permutation $\pi : [n] \mapsto [n]$ is *good* if $\delta(w^{(n/m)}, C_1 + \pi(C_2)) \geq \frac{1}{3}$ and $\delta(C_1, \pi(C_2)) \geq 1/10$.

We argue that a random permutation $\pi : [n] \mapsto [n]$ is good with probability at least $1 - o(1)$. It is sufficient to show that a random permutation $\pi : [n] \mapsto [n]$ is bad with probability at most $o(1)$.

By the Chernoff inequality the probability that for some $c_1 \in (C_1 \setminus \{0\})$ and $c_2 \in (C_2 \setminus \{0\})$ we get $\delta(w^{(n/m)}, c_1 + \pi(c_2)) < 1/3$ with probability at most $\frac{1}{2^{\Omega(n)}}$. Note that if $c_2 = 0$ then $\delta(w^{(n/m)}, c_1 + \pi(c_2)) = \delta(w^{(n/m)}, c_1) \geq \frac{1.1}{3}$ by construction.

Take a union bound over all $c_1 \in (C_1 \setminus \{0\})$ and $c_2 \in C_2$ to get that $\delta(w^{(n/m)}, C_1 + \pi(C_2)) < 1/3$ with probability at most $\frac{O(n^2)}{2^{\Omega(n)}} = o(1)$. Moreover, the probability that for given $c_1 \in (C_1 \setminus \{0\})$ and $c_2 \in C_2$ we have $\delta(c_1, \pi(c_2)) \leq 1/10$ is bounded by $\frac{1}{2^{\Omega(n)}}$. Take a union bound over all $c_1 \in (C_1 \setminus \{0\})$ and $c_2 \in C_2$ to get that the probability that $\delta(C_1, \pi(C_2)) \leq 1/10$ is bounded by $\frac{O(n^2)}{2^{\Omega(n)}} = o(1)$.

We conclude that a random permutation $\pi : [n] \mapsto [n]$ is bad with probability at most $o(1)$. So, let $\pi : [n] \mapsto [n]$ be a good permutation. Then, we have $\delta(C_1 + \pi(C_2)) = \delta(C_1, \pi(C_2)) \geq 1/10$ and $\delta(w^{(n/m)}, C_1 + C_2) \geq 1/3$. Proposition 5.1 implies that $w^{(n/m)}$ satisfies all constraints in $(C_1 + C_2)^{\perp}_{\leq o(m)}$ and thus $w^{(n/m)}$ will be accepted with probability 1 by any tester for $C_1 + \pi(C_2)$ with query complexity $\leq o(m)$. We conclude that $(C_1 + \pi(C_2))$ is not locally testable with $o(m)$ queries. Notice that $\dim(C_1 + \pi(C_2)) = \dim(C_1) + \dim(C_2) = 1.1 \log(n)$.

The proof for local decodability is almost the same. R is not q-query locally decodable by the lower bound of [9]. Corollary 4.4 implies that C_1 is not q-query locally decodable. Proposition 5.3 implies that $C_1 + \pi(C_2)$ is not q-query locally decodable. □

5.2 Proof of Theorem 3.4

Proof. In this proof we always assume that $n > 2$. Let $h(n) = o(1)$ such that $h(n) > 0$ for all $n > 2$. Without loss of generality we can assume that $h(n) \geq \frac{1}{3\log(\log(n))}$ (otherwise let $h(n) = \frac{1}{3\log(\log(n))}$) because if $h(n) < \frac{1}{3\log(\log(n))}$ then $n^{-h(n)} > n^{-\frac{1}{3\log(\log(n))}}$, so we will prove the Theorem even for a lower bias than $n^{-h(n)}$. Hence we assume that $h(n) \geq \frac{1}{3\log(\log(n))}$.

Let $g(n) = 3h(n)$ and then $g(n) \geq \frac{1}{\log(\log(n))}$ for all $n > 2$. Let $f(n) = \frac{1}{g(n)} = \omega(1)$, note that $f(n) \leq \log(\log(n))$. Let $m = n^{g(n)}$ and $R \subseteq \mathbb{F}_2^m$ be a random linear code such that $\dim(R) = \log(m) \cdot f(n) = \log(n)$. The probability that at least one nonzero codeword of R has relative weight less than $\frac{1}{2} - m^{-1/3}$ or more than $\frac{1}{2} + m^{-1/3}$ is bounded by $\frac{2 \cdot 2^{(\log(m) \cdot f(n))}}{2^{\Omega(m^{1/3})}} = o(1)$, and this follows from the Chernoff inequality and the Union bound. Moreover, the probability that $\Delta(R^{\perp}) < \log(f(n)) = \omega(1)$ is bounded by $\frac{m^{\log(f(n))}}{2^{\dim(R)}} = o(1)$, and this follows from the Union bound. So, let R be a $m^{-1/3}$-biased code such that $\dim(R) = \log(m) \cdot f(n)$ and $\Delta(R^{\perp}) \geq \log(f(n)) = \omega(1)$, i.e., $R^{\perp}_{\leq \log(f(n))} = R^{\perp}_{\leq \omega(1)} = \emptyset$. Assume without loss of generality that the first $\dim(R)$ bits of R are message bits.

Notice that $\log(m) \cdot f(n) = \log(n)$ and $m^{-1/3} = n^{-(1/3)g(n)} = n^{-h(n)}$. Claim 7.1 implies that R is not locally testable (decodable) with $q = O(1)$ queries.

Let $C \subseteq \mathbb{F}_2^n$ be the (n/m)-repetition code of R. We have $\dim(C) = \dim(R) = \log(n)$ and C has the same bias as R, i.e., C is $n^{-h(n)}$-biased. In particular we have $\delta(C) \geq \frac{1}{2} - n^{-h(n)}$. Furthermore, Corollary 4.4 implies that C is not locally testable (decodable) with $q = O(1)$ queries. The Theorem follows. □

6 Acknowledgement

We thank Tali Kaufman, Swastik Kopparty and Shubhangi Saraf for helpful discussions.

References

[1] E. Ben-Sasson, V. Guruswami, T. Kaufman, M. Sudan, and M. Viderman, "Locally testable codes require redundant testers," in *IEEE Conference on Computational Complexity*. IEEE Computer Society, 2009, pp. 52–61. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/CCC.2009.6

[2] E. Ben-Sasson, P. Harsha, and S. Raskhodnikova, "Some 3CNF properties are hard to test," *SIAM Journal on Computing*, vol. 35, no. 1, pp. 1–21, 2005. [Online]. Available: http://epubs.siam.org/SICOMP/volume-35/art_44544.html

[3] E. Ben-Sasson and M. Sudan, "Robust locally testable codes and products of codes," *Random Struct. Algorithms*, vol. 28, no. 4, pp. 387–402, 2006. [Online]. Available: http://dx.doi.org/10.1002/rsa.20120

[4] E. Ben-Sasson and M. Viderman, "Composition of semi-LTCs by two-wise tensor products," in *APPROX-RANDOM*, ser. Lecture Notes in Computer Science, I. Dinur, K. Jansen, J. Naor, and J. D. P. Rolim, Eds., vol. 5687. Springer, 2009, pp. 378–391. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03685-9

[5] A. Deshpande, R. Jain, T. Kavitha, S. V. Lokam, and J. Radhakrishnan, "Lower bounds for adaptive locally decodable codes," *Random Struct. Algorithms*, vol. 27, no. 3, pp. 358–378, 2005. [Online]. Available: http://dx.doi.org/10.1002/rsa.20069

[6] I. Dinur, "The PCP theorem by gap amplification," *Journal of the ACM*, vol. 54, no. 3, pp. 12:1–12:44, Jun. 2007.

[7] O. Goldreich, "Short locally testable codes and proofs (survey)," *Electronic Colloquium on Computational Complexity (ECCC)*, no. 014, 2005. [Online]. Available: http://eccc.hpi-web.de/eccc-reports/2005/TR05-014/index.html

[8] O. Goldreich and M. Sudan, "Locally testable codes and PCPs of almost-linear length," *Journal of the ACM*, vol. 53, no. 4, pp. 558–655, Jul. 2006.

[9] J. Katz and L. Trevisan, "On the efficiency of local decoding procedures for error-correcting codes," in *STOC*, 2000, pp. 80–86. [Online]. Available: http://doi.acm.org/10.1145/335305.335315

[10] T. Kaufman and M. Sudan, "Sparse random linear codes are locally decodable and testable," in *FOCS*. IEEE Computer Society, 2007, pp. 590–600. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/FOCS.2007.65

[11] S. Kopparty and S. Saraf, "Tolerant linearity testing and locally testable codes," in *APPROX-RANDOM*, ser. Lecture Notes in Computer Science, I. Dinur, K. Jansen, J. Naor, and J. D. P. Rolim, Eds., vol. 5687. Springer, 2009, pp. 601–614. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03685-9

[12] S. Kopparty and S. Saraf, "Local list-decoding and testing of random linear codes from high-error," in *STOC* 2010.

[13] O. Meir, "Combinatorial construction of locally testable codes," in *STOC*. ACM, 2008, pp. 285–294. [Online]. Available: http://doi.acm.org/10.1145/1374376.1374419

[14] L. Trevisan, "Some applications of coding theory in computational complexity," Sep. 23 2004. [Online]. Available: http://arxiv.org/abs/cs/0409044

7 Appendix

The next folklore claim states that the small dual distance of the linear code $C \subseteq \mathbb{F}^n$ is necessary for its local testing and local decoding. We explain this claim now.

Ben-Sasson et al. [2] showed that a q-query tester for a locally testable code is (w.l.o.g.) a distribution over dual codewords of weight at most q. In particular, if $\Delta(C^\perp) \geq q+1$ we conclude that C is not locally testable with q queries. Now, assume that the first $\dim(C)$ entries of the code C are message entries and $\Delta(C^\perp) \geq q+1$. Then any local decoder which makes only $q-1$ queries always obtains a "local view" that contains no information about the message entries and hence message entries can not be recovered with non-trivial probability.

Claim 7.1 (Folklore)**.** *Let $C \subseteq \mathbb{F}^n$ be a linear code such that $\Delta(C^\perp) \geq (q+1)$, where $q \geq 1$. Assume that the first $\dim(C)$ entries of C are message entries. Then C is not locally testable with q queries and not locally decodable with $q-1$ queries.*

The other folklore claim (stated e.g. in [1]) says that every linear code is testable with query complexity equal to its dimension plus one.

Claim 7.2 (Folklore 2)**.** *Every linear code C is testable by $\dim(C)+1$ queries.*

Let us state the central theorem (which we rephrase) from [2]. Ben-Sasson et al. [2] showed a family of codes $C_m \subset \mathbb{F}_2^m$ which has linear distance, constant rate and was characterized by 3 weight dual words. They proved that this family is not locally testable with $o(m)$ queries. Note that this family of codes is not local decodable with constant number of queries (q) because of the lower bound on the blocklength of locally decodable codes due to Katz and Trevisan [9].

Theorem 7.3. *Let $q > 0$ be a constant integer. For infinitely many $m > 0$ there exists a family of codes $C_m \subset \mathbb{F}_2^m$ which has $\delta(C_m) = \Theta(1)$, $\dim(C_m) = \Theta(m)$ and $\mathrm{span}((C_m)_3^\perp) = (C_m)^\perp$. Moreover, C_m is not locally testable with $o(m)$ queries and not locally decodable with q queries.*

ECCC ISSN 1433-8092

http://eccc.hpi-web.de