

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

January 28, 2010

Abstract

Locally testable codes (LTCs) 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. Kaufman and Sudan [4] proved that sparse, low-bias linear codes are locally testable and decodable, in particular sparse random codes are locally testable and decodable. Kopparty and Saraf [6] conjectured that all sparse linear codes, i.e., codes with logarithmic dimension, are locally testable.

In this paper we refute this conjecture by showing that for every $\epsilon > 0$ there exists a family of codes $\{C_\epsilon^{(n)} \subset \mathbb{F}_2^n\}_{n \in \mathbb{Z}}$ with relative distance $(1/2 - \epsilon)$ and $\dim(C_\epsilon^{(n)}) = \Theta(\log(n))$ which are not locally testable and not locally decodable. Moreover, for every $d(n)$ ranging from $\omega(1)$ to $\Omega(n)$ we show a family of codes $\{C_\epsilon^{(n)} \subset \mathbb{F}_2^n\}_{n \in \mathbb{Z}}$ with relative distance $(1/2 - \epsilon)$ and $\dim(C_\epsilon^{(n)}) = \Theta(d(n))$ which are not locally testable (decodable). This also shows that the requirement of “low-bias” in the work of Kaufman and Sudan [4] is necessary.

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 [8].

Kaufman and Sudan [4] showed that local testability and decodability exists in random *sparse* linear codes, where a linear code is called sparse if its dimension is logarithmic in its blocklength. A linear code $C \subset \mathbb{F}^n$ has low-bias if all nonzero codewords of C have relative weight in the range $(\frac{1}{2} - n^{-\gamma}, \frac{1}{2} + n^{-\gamma})$ for

*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.

some constant $\gamma > 0$. They showed that all sparse low-bias linear codes are locally testable (decodable)¹. In particular, sparse random linear codes have low-bias and hence are locally testable (decodable).

This result was later generalized by Kopparty and Saraf [7] to the problem known as “local list-decoding in the high error regime” (see [7] for the definition and discussion of the problem).

Kopparty and Saraf [6] conjectured that all sparse linear codes are locally testable. We refute this conjecture in Theorem 3.1 by showing a family of sparse linear codes with distance close to $1/2$ which are not locally testable (and not locally decodable). This result also shows that the “low-bias” requirement in the work of Kaufman and Sudan [4] is necessary.

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 result (Theorem 3.1). Section 4 contains some useful propositions and in Section 5 we prove Theorem 3.1.

2 Preliminaries

Let \mathbb{F} be a finite field and $[n]$ be the set $\{1, \dots, n\}$. 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 $\text{supp}(w) = \{i \in [n] \mid w_i \neq 0\}$ and $|w| = |\text{supp}(w)|$. 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 $\Delta(C) = \min_{c \in C \setminus \{0\}} \{\text{wt}(c)\}$. 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 $u = (u_1, u_2, \dots, u_n), v = (v_1, v_2, \dots, v_n) \in \mathbb{F}^n$ 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 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 $C_{\leq t}^\perp = \{u \in C^\perp \mid |u| \leq t\}$ and $C_t^\perp = \{u \in C^\perp \mid |u| = t\}$. For $w \in \mathbb{F}^n$ and $S = \{j_1, j_2, \dots, j_m\} \subseteq [n]$ we let $w|_S = (w_{j_1}, w_{j_2}, \dots, w_{j_m})$, where $j_1 < j_2 < \dots < j_m$, 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 $\text{span}(C_{\leq q}^\perp) = C^\perp$.

For $w \in \mathbb{F}^n$ 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 [3] (see that paper for a justification of the definition).

Note that given a code $C \subseteq \mathbb{F}^n$, 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 .

¹In fact, Kaufman and Sudan [4] proved a stronger result. They showed that sparse low-bias linear codes are self-correctable and thus are locally decodable.

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 $I \subseteq [n]$ s.t. $|I| \leq q$ and the following holds.

- For all $w \in \mathbb{F}^n$ s.t. $\delta(w, C) \geq \delta$ 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 $I \subseteq [n]$ s.t. $w|_I \in C|_I$ and otherwise output **reject**.

We say that a family of codes $\{C^{(n)} \mid n \in \mathbb{Z}\}$ 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 [2] 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 $I \subseteq [n]$, s.t. $|I| \leq q$ 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 $E_C : \mathbb{F}_2^k \rightarrow \mathbb{F}_2^n$ be its encoding function, i.e., $C = \{E_C(x) \mid x \in \mathbb{F}_2^k\}$. 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 $\Pr[\mathbb{D}^{\hat{c}}[i] = x_i] \geq \frac{1}{2} + \epsilon$, 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 say that a family of codes $\{C^{(n)} \mid n \in \mathbb{Z}\}$ 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 Result

In this section we present our main theorem 3.1. Recall that Kaufman and Sudan [4] showed that sparse binary linear codes with low bias are locally testable and self-correctable (and in particular locally decodable). The next theorem refutes the conjecture of Kopparty and Saraf [6], which says that all sparse codes are locally testable. Moreover, this theorem shows that the “low-bias” requirement in the work of Kaufman and Sudan [4] is necessary, i.e., it shows that there exists a family of sparse codes with relative distance ≥ 0.49 which are not locally testable (decodable).

One could conjecture that sparse 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, but are not locally testable (decodable).

Finally we show that there are sparse codes with dual distance > 2 which are not locally testable. We believe that this case is interesting since it shows that non-redundant² sparse code can be not locally testable. 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 [4] 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.

All our constructions can achieve any dimension from $w(1)$ up to linear.

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

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 $I = \{i_1, i_2, \dots, i_q\} \subseteq [n]$ let $I \bmod m = \{i_1 \bmod m, i_2 \bmod m, \dots, i_q \bmod m\} \subseteq [m]$.

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.

²The 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.

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 s.t. $\delta(w, C) \geq 2\delta$. If $\delta(w|_{[m]}, C|_{[m]}) = \delta(w|_{[m]}, R) \geq \delta$ we are done, since

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

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

But $\delta(w, C) \geq 2\delta$ implies that

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

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

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 $I \subseteq [m]$ s.t. $|I| \leq q$.

We argue that \mathbf{D}_R is a (q, ϵ, δ) -tester for R . Let $w \in \mathbb{F}_2^m$ be a word s.t. $\delta(w, R) \geq \delta$. 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

$\Pr_{I \sim \mathbf{D}_C} [w^{(im)}|_I \in C|_I] < \epsilon$ 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.*

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

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 $w \in \mathbb{F}_2^m$, $\delta(w, R) \leq \delta$ s.t. the decoder \mathbb{D}_R recovers some message bit i with probability less than $1 - \epsilon$. 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. \square

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 Theorem 3.1

In this section we give the proofs of the three bullets 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 [5] imply that R_ϵ is not q -query locally decodable code.

Let $C_\epsilon \subset \mathbb{F}_2^n$ 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, Proposition 4.2 implies that C_ϵ is not locally testable with $o(m)$ queries and not locally decodable with q queries. \square

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 $R \subset \mathbb{F}_2^m$ s.t. $\dim(R) = \Theta(d(n))$, $\Delta(R) \geq \Omega(m)$, $\text{span}(R_{\leq 3}^\perp) = R^\perp$ 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. \square

5.1 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 $u \in (C_1^\perp \cap C_2^\perp)$ \square

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 $c_1 + c_2$. \square

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 $C_1 + C_2$. \square

Proof of the third bullet. Let $m = \log(n)$ and $R \subset \mathbb{F}_2^m$ be a linear code with $\Delta(R) \geq \Theta(m)$ and $\Delta(R^\perp) \geq \Theta(m)$ and $\dim(R) = \Theta(m)$ (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 s.t. $\delta(w, R) \geq \frac{1}{3}$ (a random $w \in \mathbb{F}_2^m$ satisfies this condition with high probability). Notice that for all $u \in R_{\leq o(m)}^\perp$ we have $\langle u, w \rangle = 0$ because R^\perp has no words of weight less than $o(m)$, i.e., $R_{\leq o(m)}^\perp = \emptyset$.

Let $C_1 \subset \mathbb{F}_2^n$ be a (n/m) -repetition code of R . Then $\delta(w^{(n/m)}, C_1) = \delta(w, R) \geq \frac{1}{3}$. Notice that by construction for all $u \in C_{1, \leq o(m)}^\perp$ we have $\langle u, w^{(n/m)} \rangle = 0$. We also have $\delta(C_1) = \Theta(1)$ and $\dim(C_1) = \Theta(\log(n))$. Let $C_2 \subseteq F_2^n$ be the Hadamard code³ (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 $w = (w_1, w_2, \dots, w_n) \in \mathbb{F}_2^n$ let $\pi(w) = (w_{\pi(1)}, w_{\pi(2)}, \dots, w_{\pi(n)})$ be a π -permuted word. Let $\pi(C_2) = \{\pi(c) \mid c \in C_2\}$ be a set of all permuted 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 $\Delta((\pi(C_2))^\perp) = \Delta(C_2^\perp) > 2$.

³Instead 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 $\delta(w^{(n/m)}, C_1 + C_2) < 1/3$ 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 $(C_1 + C_2)_{\leq o(m)}$ 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 $(C_1 + C_2)$ is not locally testable with $o(m)$ queries.

The proof for local decodability is almost the same. R is not q -query locally decodable by the lower bound of [5]. 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. \square

6 Acknowledgement

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

References

- [1] E.Ben-Sasson, P.Harsha and S.Rashodnikova. Some 3CNF properties are hard to test. SICOMP 2005, pp. 1-21.
- [2] E.Ben-Sasson, V. Guruswami, T.Kaufman, M.Sudan and M.Videman. Locally testable codes require redundant testers. In Proceedings of CCC 2009.
- [3] E.Ben-Sasson and M.Videman. Composition of semi-LTCs by two-wise Tensor Products. RANDOM 2009.
- [4] T.Kaufman and M.Sudan. Random Sparse linear codes are locally testable and decodable. FOCS 2007.
- [5] Katz and L.Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In: STOC 2000: Proceedings of the 32nd Annual Symposium on the Theory of Computing. pp. 8086.
- [6] S.Kopparty and S.Saraf. Tolerant linearity testing and local testability. RANDOM 2009.
- [7] S.Kopparty and S.Saraf. Local List-Decoding and Testing of Sparse Random Linear Codes from High-Error. Note in ECCS - TR09-115.

[8] Luca Trevisan. Some Applications of Coding Theory in Computational Complexity. Note in ECCC - TR04-043.

7 Appendix

The next folklore claim (due to [1]) states that the small dual distance of the code is necessary for its local testing.

Claim 7.1 (Folklore). *Let C be a linear code such that $\Delta(C^\perp) \geq (q+1)$, where $q \geq 1$ then C is not locally testable with q queries.*

The other folklore claim (stated e.g. in [2]) says that every code is testable with query complexity equal to its dimension. Intuitively, C is not trivially testable if it can be testable with $o(\dim(C))$. In this paper we will show that a family of codes that can not be non-trivially testable.

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

Let us state the central theorem (which we rephrase) from [1]. Ben-Sasson et al. [1] 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 [5].

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 $\text{span}((C_m)_3^\perp) = (C_m)^\perp$. Moreover, C_m is not locally testable with $o(m)$ queries and not locally decodable with q queries.*