Locally Testable Codes Require Redundant Testers *

Eli Ben-Sasson†
Dept. of Computer Science
Technion
Haifa 32000, Israel
eli@cs.technion.ac.il

Venkatesan Guruswami‡
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, 15213, USA.
guruswami@cmu.edu

Tali Kaufman
Dept. of Computer Science and Applied Mathematics
Weizmann Institute
Rehovot, Israel
tali.kaufman@weizmann.ac.il

Madhu Sudan
Microsoft Research
Cambridge, Massachusetts, USA
madhu@microsoft.com

Michael Viderman
Dept. of Computer Science
Technion
Haifa 32000, Israel
viderman@cs.technion.ac.il

October 30, 2010

Abstract

Locally testable codes (LTCs) are error-correcting codes for which membership, in the code, of a given word can be tested by examining it in very few locations. Most known constructions of locally testable codes are linear codes, and give error-correcting codes whose duals have (superlinearly) *many* small weight codewords. Examining this feature appears to be one of the promising approaches to proving limitation results for (i.e., upper bounds on the rate of) LTCs.

Unfortunately till now it was not even known if LTCs need to be non-trivially redundant, i.e., need to have *one* linear dependency among the low-weight codewords in its dual.

In this paper we give the first lower bound of this form, by showing that every positive rate constant query strong LTC must have linearly many redundant low-weight codewords in its dual. We actually prove the stronger claim that the *actual test itself* must use a linear number of redundant dual codewords (beyond the minimum number of basis elements required to characterize the code); in other words, non-redundant (in fact, low redundancy) local testing is impossible.

Our main theorem is a special case of a more general theorem that applies to any tester for an arbitrary linear locally testable code $\mathcal{C}$. The general theorem can be used, for instance, to provide an arguably simpler proof of the main result of [12] which says that testing random low density parity check (LDPC)

*A preliminary version of this paper appeared in the Proceeding of CCC 2009 [10].

†Research of first, fourth and fifth co-authors supported by grant number 2006104 by the US-Israel Binational Science Foundation. Work of the first and last co-authors supported by grant number 679/06 by the Israeli Science Foundation.

‡The work of the second author was done during a visit to the Technion in 2008. Research supported in part by a Packard fellowship, NSF grants CCF-0343672 and CCF-0953155, and US-Israel BSF grant 2008293.

codes requires linear query complexity. Informally, our more general theorem says the following. Take any basis B for the dual code of $\mathcal{C}$ that is comprised of words of small support, i.e., every element of B has very few nonzero entries. Then the dual code of $\mathcal{C}$ must contain many words that are *(i)* not in B, *(ii)* have small support, and most importantly, *(iii)* are a linear combination of a constant fraction of B.

1 Introduction

In this work, we exhibit some *limitations* of locally testable linear codes. A linear code over a finite field $\mathbb{F}$ is a linear subspace $\mathcal{C} \subseteq \mathbb{F}^n$. The dimension of $\mathcal{C}$ is its dimension as a vector space, and its rate is the ratio of its dimension to n. The distance of $\mathcal{C}$ is the minimal Hamming distance between two different codewords. One is typically interested in codes whose distance is a growing function of the block length n, ideally $\Omega(n)$. Such a code is *locally testable* if given a word $x \in \mathbb{F}^n$ one can verify with good accuracy whether $x \in \mathcal{C}$ by reading only a few (say a constant independent of n) chosen symbols from x. More precisely such a code has a *tester*, which is a randomized algorithm with oracle access to the received word x. The tester reads at most q symbols from x and based on this local view decides if $x \in \mathcal{C}$ or not. It should accept codewords with probability one, and reject words that are far (in Hamming distance) from the code with noticeable probability.

Locally Testable Codes (henceforth, LTCs) may be thought of as the combinatorial core of PCP constructions. In recent years, starting with the work of Goldreich and Sudan [20], several surprising constructions of LTCs have been given (see [19] for an extensive survey of some of these constructions). The principal challenge is to understand the largest asymptotic rate possible for LTCs, and to construct LTCs approaching this limit. We now know constructions of LTCs of dimension $n/\log^{O(1)} n$ which can tested with only three queries [13, 16], [25].

One of the outstanding open questions in the subject is whether there are asymptotically good LTCs, i.e., LTCs that have dimension $\Omega(n)$ and distance $\Omega(n)$. Our understanding of the *limitations* of LTCs is, however, quite poor (in fact, practically non-existent), and approaches that may rule out the existence of asymptotically good LTCs have been elusive. Essentially the only negative results on LTCs concern binary codes testable with just 2-queries [9, 22] (which is a severe restriction), random LDPC codes [12], and cyclic codes [5].[1] In fact, we cannot even rule out the existence of binary LTCs meeting the Gilbert-Varshamov bound (which is the best known rate for codes without any local testing restriction). So, for all we know, the strong testability requirement of LTCs may not "cost" anything extra over LDPC codes!

This work is a (modest) initial attempt at addressing our lack of knowledge concerning lower bound results for LTCs. For linear codes, one can assume without loss of generality [12] that the tester picks a low-weight dual codeword $c^{\perp}$ from some distribution, and checks that the input x is orthogonal to $c^{\perp}$. It is thus necessary that if C is a q-query LTC of dimension k, then its dual $C^{\perp}$ has a basis of $n-k$ codewords each of weight at most q.[2] All known constructions of LTCs in fact have duals which have super-linearly many low-weight dual codewords. In other words, there must be a substantial number of linear dependencies amongst the low-weight dual codewords. Examining whether this feature is necessary might be one of the promising approaches to proving limitations (i.e., upper bounds on the rate) of LTCs, as it imposes strong

[1] The last result rules out asymptotically good *cyclic* LTCs; the existence of asymptotically good cyclic codes has been a long-standing open problem, and the result shows the "intersection" of these questions concerning LTCs and cyclic codes has a negative answer.

[2] To be precise, only when $\mathcal{C}$ is a *strong* LTC, as per Definition 2, need $\mathcal{C}^{\perp}$ be spanned by words of weight q. Non-strong LTCs have the property that the set of low-weight words in the dual code must span a large dimensional subspace of $\mathcal{C}^{\perp}$ (see Proposition 28 for an exact statement).

constraints on the dual code.[3] Nevertheless, till now it was not even known if the dual of a LTC has to be non-trivially redundant, i.e., if it must have at least *one* linear dependency among its low-weight codewords.

In this work, we give the first lower bound of this form, by showing that every positive rate constant query LTC must have $\Omega(n)$ redundant low-weight codewords. The result is actually stronger — it shows that the *actual test itself* must use $\Omega(n)$ extra redundant dual codewords (beyond the minimum $n-k$ basis elements). In other words, *non-redundant testing is impossible*. While this might sounds like an intuitively obvious statement, we remark that even for Hadamard codes (whose dual has $\Theta(n^2)$ weight 3 codewords), a non-redundant test consisting of a basis of weight 3 dual codewords was not ruled out prior to our work. Also, without the restriction on number of queries, *every* code does admit a basis tester (which makes at most $k+1$ queries).

We also note that a known upper bound [6, Proposition 11.2] shows that $O(n)$ redundancy suffices for testing. [6] prove this in the context of PCPs, but the technique extends to LTCs as well. For completeness, in Section 6, we include a proof showing that for *every* q-query LTC, there *is* a $O(q)$-query tester that picks a test uniformly from at most $3(n-k)=O(n)$ dual codewords. The quantity $n-k$ (as opposed to n) is significant in that this is the dimension of the dual code, and our lower bound shows that every tester (for any code) must have a support of size at least $n-k$.

We point out that our main theorem (Theorem 5) is actually just a special case of a more general statement given in Theorem 14. For instance, the more general theorem can be used to provide a different and arguably simpler, proof of the main result of [12] that says that testing of random low density parity check (LDPC) codes require linear query complexity (see Section 3.3). But Theorem 14 goes even further and we believe it may be instrumental in proving limitations on the rate of other families of LTCs in the future. We end this section by informally describing this result. Let B be any basis for $\mathcal{C}^\perp$ comprised of words of small support. Such a basis must exist if $\mathcal{C}$ is to be locally testable. Theorem 14 says that *any* tester for $\mathcal{C}$ must use (many) dual words that are each a linear combination of a *constant fraction* of B. In plain words, $\mathcal{C}^\perp$ must have a high level of redundancy and cancelation to allow for large sums of small-support words in B to result in words that are also of small support.

Organization of the paper. In the following section we provide the standard definitions regarding locally testable codes and their testers. In Section 3 we state our main results. This is followed by Section 4, where we prove our main technical theorem (Theorem 14). Then in Section 5 we show that our lower bound shown in Theorem 5 is quite tight (Theorem 14) and in Section 6 we show an upper bound on the redundancy of testers (see Definition 3). Finally, in Section 7 we discuss some open questions.

2 Defining the redundancy of a tester

Preliminary notation Throughout this paper $\mathbb{F}$ is a finite field, $[n]$ denotes the set $\{1,\ldots,n\}$ and $\mathbb{F}^n$ denotes $\mathbb{F}^{[n]}$. For $w=\langle w_1,\ldots,w_n\rangle\in\mathbb{F}^n$ let $\mathrm{supp}(w)=\{i|w_i\neq 0\}$ and $\mathrm{wt}(w)=|w|=|\mathrm{supp}(w)|$. 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 $\delta(x,y)=\frac{\Delta(x,y)}{n}$.

We use the standard notation for describing linear error correcting codes and point out that all codes discussed in this paper are linear. A $[n,k,d]_{\mathbb{F}}$-code is a k-dimensional subspace $\mathcal{C}\subseteq\mathbb{F}^n$ of distance d, defined next. The relative distance of $\mathcal{C}$ is denoted $\delta(\mathcal{C})$ and defined to be the minimal value of $\delta(x,y)$ for

[3] We remark that information on the dual weight distribution is useful, for example, in the linear programming bounds on the rate vs. distance trade-off of a linear code. For LDPC codes whose dual has a low weight basis, stronger upper bounds on distance are known compared to general linear codes of the same rate [7].

two distinct codewords $x, y \in C$. The distance of $\mathcal{C}$ is $\Delta(\mathcal{C}) = \delta(\mathcal{C}) \cdot n$. Let $\delta(x, \mathcal{C}) = \min_{y \in \mathcal{C}}\{\delta(x, y)\}$ denote the relative distance of x from the code C. We say that x is α-far from $\mathcal{C}$ if $\delta(x, \mathcal{C}) \geqslant \alpha$ and otherwise we say x is α-close to $\mathcal{C}$. The inner-product between two vectors u and v in $\mathbb{F}^n$ is $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$.

For a linear code $\mathcal{C}$ let $\mathcal{C}^\perp$ denote its dual code, i.e., $\mathcal{C}^\perp = \{u \in F^n \mid \forall c \in \mathcal{C} : \ \langle u, c\rangle = 0\}$ and recall $\dim(\mathcal{C}^\perp) = n - \dim(\mathcal{C})$. Let $\mathcal{C}^\perp_{<t} = \{u \in \mathcal{C}^\perp \mid |u| < t\}$ and $\mathcal{C}^\perp_{\leqslant t} = \{u \in \mathcal{C}^\perp \mid |u| \leqslant t\}$. For $S \subseteq \mathbb{F}^n$, where S is not a vector space, with some abuse of notation we let $\dim(S) = \dim(\operatorname{span}(S))$.

Definition 1 (Tester)**.** Suppose $\mathcal{C}$ is a $[n, k, d]_{\mathbb{F}}$-code. A *q-query test* for $\mathcal{C}$ is an element $u \in \mathcal{C}^\perp_{\leqslant q}$ and a *q-query tester T* for $\mathcal{C}$ is defined by a distribution p over q-query tests. When $\mathcal{C}$ is clear from context we omit reference to it. The *support* of T, denoted $S = S_T$, is the support of p, i.e., the set $S = S_T = \left\{u \in \mathcal{C}^\perp_{\leqslant q} \,\middle|\, p(u) > 0\right\}$. When p is uniform over a subset of $\mathcal{C}^\perp_{\leqslant q}$ we say the tester is *uniform* and may identify the tester with S.

Invoking the tester T on a word $w \in \mathbb{F}^n$ is done by sampling a test $u \in S_T$ according to the distribution p and outputting accept if $\langle u, w\rangle = 0$, in which case we say that u (and T) accept w, denoted $T[w] = \mathsf{accept}$, and outputting reject, denoted $T[w] = \mathsf{reject}$, if $\langle u, w\rangle \neq 0$. Clearly any such tester always accepts $w \in \mathcal{C}$.

- A (q, ρ')-*strong tester* is a q-query tester T satisfying for all $w \in \mathbb{F}^n$

$$\Pr[T[w] = \mathsf{reject}] \geqslant \rho' \cdot \delta(w, \mathcal{C}).$$

- A (q, ε, ρ)-*tester* is a q-query tester T satisfying for all $w \in \mathbb{F}^n$ that is ε-far from $\mathcal{C}$

$$\Pr[T[w] = \mathsf{reject}] \geqslant \rho.$$

The probability in both equations above is according to the distribution p associated with T.

Definition 2 (Locally Testable Code (LTC) [20])**.** A $[n, k, d]_{\mathbb{F}}$-code $\mathcal{C}$ is said to be a (q, ρ')-*strong locally testable code* if it has a (q, ρ')-strong tester, and $\mathcal{C}$ is a (q, ε, ρ)-*locally testable code* if it has (q, ε, ρ)-tester T. The parameter q is known as the *query complexity* of T, ρ is its *soundness* and ε is its *distance threshold*.

Note that a (q, ρ')-strong LTC is also a $(q, \varepsilon, \rho' \cdot \varepsilon)$ LTC for every $\varepsilon > 0$.

Moreover, if T is a $(q, \rho' > 0)$-strong tester for a $[n, k, d]_{\mathbb{F}}$-code then, letting S_T denote the support of T, we have $\dim(S_T) = \dim(C^\perp) = n - k$.

Remarks on definitions of testers. Our definition of a tester, and an LTC is somewhat different from previous definitions (notably [12] and [20]). We clarify the differences here.

We start with Definition 2. The definition of strong LTCs we use is the same as that in [20]. The weak notion is weaker than their definition of a weak tester (which simply allowed the rejection probability of a weak tester to be smaller by a $o(1)$ additive amount compared to the strong case). Our definition on the other hand only requires rejection probability to be positive when the word is very far (constant relative distance) from the code. Since our goal is to prove "impossibility" results, doing so with weaker definitions makes our result even stronger.

We now discuss Definition 1. For linear LTCs it was shown in [12] (see also references therein) that the tester might as well pick a collection of low-weight dual codewords and verify that the given word w is orthogonal to all of them. On the other hand, our definition (Definition 1) requires the tester to pick only one dual codeword and test orthogonality to it. Our definition is more convenient to use when defining and analyzing the *redundancy* of tests (defined below). We first note that our restricted forms of tests may only alter the soundness of the test by a constant factor. For this we recall the assertion from [12] that showed that without loss of generality a q-query "standard" tester for a $[n, k, d]_{\mathbb{F}}$-code is defined by a distribution over subsets $I \subseteq [n], |I| \leqslant q$. The test associated with I accepts a word w if and only if $\langle w, u \rangle = 0$ for all $u \in \mathcal{C}^{\perp}$ such that $\mathrm{supp}(u) \subseteq I$. (The soundness and the distance threshold of a "standard" tester are defined as in Definition 1.) To convert this "standard" tester to one that only tests one dual codeword, consider a tester that, given I, samples uniformly from the set $U_I = \left\{u \in \mathcal{C}^{\perp} \mid \mathrm{supp}(u) \subseteq I\right\}$ and accepts iff $\langle u, w \rangle = 0$ [4]. This resulting tester conforms to our Definition 1. Furthermore, if the soundness of the "standard" tester is ρ then the soundness of the tester that samples uniformly from U_I is at least $\frac{|\mathbb{F}|-1}{|\mathbb{F}|}\rho \geqslant \frac{1}{2}\rho$. To see this, notice that U_I forms a linear space over $\mathbb{F}$. And the set $\{u \in U_I \mid \langle u, w \rangle = 0\}$ is a linear subspace of U_I. Thus, whenever w is rejected by some $u \in U_I$ we actually know that w is rejected by at least a fraction $\frac{|\mathbb{F}|-1}{|\mathbb{F}|}$ of U_I because the set of rejecting words is the complement of a subspace of U_I. Hence, using our definition of a tester is equivalent to the most general definition of a tester, up to a constant loss in the soundness parameter.

Definition 3 (Linearly independent tester, basis tester and tester redundancy)**.** Suppose $\mathcal{C}$ is a $[n, k, d]_{\mathbb{F}}$-code. A q-query tester T for $\mathcal{C}$ is said to be a *linearly independent tester* if its support $S_T \subseteq \mathcal{C}^{\perp}_{\leqslant q}$ is a set of linearly independent vectors. If T is a linearly independent tester and its support S_T is of size $|S_T| = \dim(\mathcal{C}^{\perp}) = n - k$ then we call it a *basis tester* because S_T forms a basis for $\mathcal{C}^{\perp}$. In case S_T has size larger than $\dim(\mathrm{span}(S_T))$ we define the *redundancy* of T to be $|S_T| - \dim(\mathrm{span}(S_T))$. (Notice that a linearly independent tester has redundancy 0.)

Definition 4 (Expected query complexity)**.** The *expected query complexity* of a tester for $\mathcal{C}$ with distribution p over its support $S \subseteq C^{\perp}$ is defined to be $\mathop{\mathbf{E}}_{u \sim p}[|u|]$.

3 Main results

This section contains four parts. We start by stating our main results — Theorem 5 and Corollaries 8 and 9. Then, we discuss the main technical contribution of this paper — Theorem 14 — which implies all of our main results. We go on to show another application of Theorem 14, namely, a generalization and simplification of the main result from [12] stating that testing random low-density-parity-check (LDPC) codes require linear query complexity. Finally, we provide the proofs of our main results assuming Theorem 14. The proof of Theorem 14 appears in the next section.

3.1 Statement of main results

Theorem 5 (Linearly independent tester)**.** *If a $[n, k, d]_{\mathbb{F}}$-code $\mathcal{C}$ has a $(q, \frac{\delta(\mathcal{C})}{3}, \rho)$-linearly independent tester then*

$$\rho \leqslant \frac{q}{k}.$$

Remark 6. Theorem 5 (Linearly independent tester) holds even for a basis tester that has only *expected* query complexity $\leqslant q$ (and all other parameters are as in the statement of the theorem). Recall that a tester has *expected* query complexity at most q if $\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[|u|] \leqslant q$ where the expectation is taken with respect to the probability $\mathcal{D}$ associated with the tester.

Remark 7. The choice of $\frac{\delta(\mathcal{C})}{3}$ as the distance threshold is crucial in Theorem 14, Theorem 5 and their corollaries.

[4]The idea of choosing a random element of the set U_I is folklore.

The first corollary of our main theorem says that $\Omega(n)$ redundancy is necessary for uniform testing of all codes that have nontrivial (i.e., super-constant) size.

Corollary 8 (Uniform testers for LTCs with super constant size require linear redundancy)**.** *Let $\mathcal{C}$ be a $[n, k, d]$ code that is $(q, \frac{1}{3}\delta(\mathcal{C}), \rho)$-locally testable by a uniform tester using a set $S \subseteq \mathcal{C}^{\perp}_{\leqslant q}$. Then*

$$|S| \geqslant \left(\frac{1 - q/k}{1 - \rho}\right) \cdot \dim(\operatorname{span}(S)) = \left(\frac{1 - q/k}{1 - \rho}\right) \cdot \Omega(n).$$

In words, S has redundancy at least $\frac{\rho - q/k}{1-\rho} \cdot \dim(\operatorname{span}(S))$.

For instance, if $k = \dim(\mathcal{C}) = \omega(1)$ and ρ, q are constants then the previous corollary says that a uniform tester for $\mathcal{C}$ requires a linear amount ($\Omega(n)$) of redundancy. The fact that $\dim(\operatorname{span}(S)) = \Omega(n)$ will be proven below (Proposition 28).

Our second corollary shows that non-trivial redundancy is necessary for general (i.e., for nonuniform) testing.

Corollary 9 (Testers for LTCs with constant rate require linear redundancy)**.** *Let $\mathcal{C}$ be a $[n, k, d]$ code that is $(q, \frac{1}{3}\delta(\mathcal{C}), \rho)$-locally testable by a tester that is distributed over a set $S \subseteq \mathcal{C}^{\perp}_{\leqslant q}$. Then*

$$|S| \geqslant \dim(\operatorname{span}(S)) + \frac{\rho k}{q} - 1.$$

In words, S has redundancy at least $\frac{\rho k}{q} - 1$.

For instance, if $k = \Theta(n)$ and ρ, q are constants (i.e., when $\mathcal{C}$ comes from an asymptotically good family of error correcting codes) then, once again, a linear amount of redundancy is required by any constant-query tester for $\mathcal{C}$. For the state of the art LTCs [13, 16, 25] $k = \Theta(n/\operatorname{poly}(\log n))$ and our result implies that $\Theta(n/\operatorname{poly}(\log n))$ redundancy is necessary in such cases.

Remark 10. Looking at Corollary 9 one might conjecture that a stronger bound on the redundancy of a tester should hold, one that depends on the blocklength n. However, as shown in [15] this conjecture is false. In particular, for every k ranging from $\omega(1)$ to $O(n/poly(\log n))$ [15, Section 4.3] gives a construction of LTCs of dimension k along with constant-query testers with redundancy $k \cdot poly(\log k)$. This result also shows that the lower bound on redundancy is nearly tight, up to a multiplicative $poly(\log k)$ factor. Stated differently, the said result of [15] shows an inherent difference between the redundancy of uniform and non-uniform testers. Namely, the redundancy of a non-uniform tester is at least proportional to the dimension of the code whereas the redundancy of a uniform tester for LTCs with non-constant dimension is at least proportional to the blocklength of the code.

Later on in the paper we show that our main theorem is almost tight in two respects. In Section 5 we show that there do exist codes of constant size that can be strongly tested by a uniform basis tester of $O(1)$ query complexity, and that every code can be strongly tested by a uniform basis tester that has large query complexity. We conclude by showing in Section 6 that if $\mathcal{C} \subseteq F_2^n$ is a (q, ε, ρ)-LTC then it has a $(\frac{10q}{\rho}, \varepsilon, \frac{1}{100})$-tester that is uniform over a multiset S with a small (linear) amount of redundancy, i.e., with $|S| \leqslant 3\dim(\mathcal{C}^{\perp})$ and $\dim(S) \geqslant \dim(\mathcal{C}^{\perp}) - 3\varepsilon n$.

3.2 Main Technical Theorem

Theorem 5 follows from the theorem stated next, which is the main technical contribution of this paper. To state the theorem we need a couple of preliminary definitions.

Definition 11 (Support size of a test)**.** Let T be a tester for $\mathcal{C}$ and $S \subseteq \mathcal{C}^\perp$ be its support. Let $B \subseteq S$ be a basis for S and $u \in S$. Then let $\{u\}_B$ be the subset of B needed to represent u in the basis B. Formally, if $u = \sum_{v \in B} a_v \cdot v$ then

$$\{u\}_B = \{v \in B \mid a_v \neq 0\}.$$

We let $|u|_B = |\{u\}_B|$ be the support size of u with respect to the basis B.

Example 12. For $u \in S$ of the form $u = u_1 + u_2 + u_3$ for $u_1, u_2, u_3 \in B$ we have $\{u\}_B = \{u_1, u_2, u_3\}$ and $|u|_B = |\{u\}_B| = 3$.

It will be convenient to work with the following measure.

Definition 13 (Average weight)**.** Given $u \in S \subseteq C^\perp$ and a basis B we let

$$\operatorname{avg}(\{u\}_B) = \frac{\sum_{u_i \in \{u\}_B} |u_i|}{|u|_B}$$

to denote the average weight of the words in $\{u\}_B$.

Theorem 14 (Main Technical Theorem)**.** *If a $[n,k,d]_{\mathbb{F}}$-code $\mathcal{C}$ has a $(\cdot, \frac{\delta(\mathcal{C})}{3}, \rho)$-tester which is a distribution $\mathcal{D}$ over $S \subseteq C^\perp$ then for every basis B of S it holds that*

$$\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[|u|_B \cdot \operatorname{avg}(\{u\}_B)] \geqslant \rho k.$$

In particular,

- *If for every $u \in S$ we have $|u|_B \leqslant c$ then* $\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[\operatorname{avg}(\{u\}_B)] \geqslant \frac{\rho k}{c}$.
- *If for every $u \in S$ we have $\operatorname{avg}(\{u\}_B) \leqslant q$ then* $\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[|u|_B] \geqslant \frac{\rho k}{q}$.

3.3 A simpler proof of the main result from [12]

Ben-Sasson et al. showed in [12] that a family of randomly chosen low-density-parity-check (LDPC) codes requires, with high probability, linear query complexity. To explain the significance of this result, let us say that that a code $\mathcal{C}$ has *characterization weight* w if $\mathcal{C}^\perp$ is spanned by words of weight at most w. The result of [12] shows a huge gap between characterization weight — which, there, equals 3 — and *query complexity*, which, there, is shown to be linear in the block length of the code. All other upper bounds on the rate of families of locally testable codes are obtained by ruling out a small-weight characterization of the code. For example, the results of [9, 22] that rule out 2-query LTCs do this by (roughly) showing that any code that is *characterized* by 2-query words must be of small size. Similarly, the results of [5] show that any cyclic code with constant rate cannot be *characterized* by constant weight words.

In this section we use our main result to present an arguably simpler proof of the main result of [12]. In particular, we show that to obtain the same qualitative bounds as for random LDPC codes [12], we only replace one of the three conditions required there by a requirement that holds for all LDPC codes, namely, that $\mathcal{C}^\perp$ has constant characterization weight. Now for the details.

We start by stating the following result of [12], which is the combination of Definition 3.4 and Theorem 3.5 there.

Theorem 15 (Some locally-characterized codes require large query complexity)**.** *Let $\mathcal{C}$ be a $[n, k, d]$-code over the two element field $\mathbb{F}_2$ such that $\mathcal{C}^\perp$ has a basis B satisfying the following two conditions for some $0 < \varepsilon, \mu < 1/2$ and some integer q:*

- *Every $w \in \mathbb{F}_2^n$ that is orthogonal to all but one constraint in B satisfies $|w| \geqslant \varepsilon n$.*
- *Every $u \in \mathcal{C}^\perp$ that is the sum of at least $\mu|B|$ constraints of B must satisfy $|u| \geqslant q$.*

Then any tester as per Definition 1 that rejects words that are ε-far from $\mathcal{C}$ with probability at least 2μ must have query complexity $\geqslant q$.

This Theorem implies the main result of [12] because a family of random LDPC codes of constant rate will satisfy the conditions of the previous theorem for some $0 < \varepsilon, \mu < 1/2$ and $q = \delta n$ for some $\delta > 0$.

Our work can be used to replace Theorem 15. In particular, the following statement does not require a basis for $\mathcal{C}^\perp$ (any set S spanning $\mathcal{C}^\perp$ suffices), but we do need S to be comprised of low weight dual words. More importantly, we completely remove the need for the first bullet in Theorem 15.

Theorem 16. *Let $\mathcal{C}$ be a $[n, k, d]$-code over the two element field $\mathbb{F}_2$ such that $\mathcal{C}^\perp$ is spanned by a set $S \subseteq \mathcal{C}^\perp_{\leqslant q^*}$ satisfying the following condition for some $0 < \mu < 1$ and some integer q:*

- *Every $u \in \mathcal{C}^\perp$ that is the sum of at least $\frac{\mu \dim(\mathcal{C})}{q^*}$ words of S must satisfy $|u| \geqslant q$.*

Then any tester as per Definition 1 that rejects words that are $\frac{\delta(\mathcal{C})}{3}$-far from $\mathcal{C}$ with probability at least μ must have query complexity $\geqslant q$.

For instance, in the case of random LDPC codes take S to be the set of rows of the parity check matrix of the code. We get $q^* = O(1)$ and, following the analysis of random expanders as in [12], one can verify that the assumption of the theorem holds for any sufficiently small $\mu > 0$ and for $q = \mu' n$ where $\mu' > 0$ depends only on μ. This implies that testing random LDPC codes requires linear query complexity and thus we recover the main result of [12].

Proof. By way of contradiction assume that for $q' < q$ the code $\mathcal{C}$ is a $(q', \frac{\delta(\mathcal{C})}{3}, \mu)$-LTC with a tester having distribution $\mathcal{D}$. Pick any basis $B \subseteq S$ and then by Theorem 14 it holds that $\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[|u|_B] \geqslant \dfrac{\mu \dim(\mathcal{C})}{q^*}$, where $\mathcal{D}(u) > 0$ implies $|u| \leqslant q' < q$. This implies the existence of $u \in \mathcal{C}^\perp$ such that $|u| < q$ and $|u|_B \geqslant \frac{\mu \dim(\mathcal{C})}{q^*}$. But this contradicts the assumption of our theorem, and the proof is complete. $\square$

3.4 Proofs of main results

We end this section by proving Theorem 5 and its corollaries using Theorem 14.

Proof of Theorem 5. By assumption $\mathcal{C}$ has a $(q, \frac{\delta(\mathcal{C})}{3}, \rho)$-linearly independent tester which is a distribution $\mathcal{D}$ over some set B' (support of the tester). We consider a distribution $\mathcal{D}$ as a function from $C^\perp$ to $[0, 1]$ such that $\mathcal{D}(u) > 0$ iff $u \in B'$. We have $\mathop{\mathbf{E}}_{u \sim \mathcal{D}}[|u|] \leqslant q$. Since B' contains only linearly independent vectors it can be completed to a basis B for $C^{\perp}$ by adding some $u \in C^{\perp}$ such that $\mathcal{D}(u) = 0$. Notice that we still have $\mathop{\mathbf{E}}_{u\sim\mathcal{D}}[|u|] \leqslant q$ since distribution was not changed. We know that for any $u \in B$ it holds that $|u|_B = 1$ and $\operatorname{avg}(u_B) = |u|$. Thus by the first bullet of Theorem 14 we have

$$q \geqslant \mathop{\mathbf{E}}_{u\sim\mathcal{D}}[|u|] \geqslant \rho k.$$

□

Proof of Corollary 8. By assumption $\dim(S) \leqslant \dim(\mathcal{C}^{\perp}) = n - k$. Partition S into $B \cup S'$ where B is a basis for $S \subseteq \mathcal{C}^{\perp}$, $|B| = \dim(S) \leqslant n - k$ and $S' = S \setminus B$ is the set of redundant tests. We bound the size of S' from below.

Consider a basis tester defined by B. By Theorem 5 this tester is not very sound, i.e., there exists a word $w \in \mathbb{F}^n$ that is $(\frac{1}{3}\delta(\mathcal{C}))$-far from $\mathcal{C}$ and is rejected by at most a fraction $\rho_B \leqslant \frac{q}{k}$ of the constraints in B. The overall number of constraints rejecting w is at least $\rho|S| = \rho(|B| + |S'|)$ because S is a uniform tester for $\mathcal{C}$ and w is far from $\mathcal{C}$. Taking the most extreme case that all words in S' reject w we get

$$\rho(\dim(S) + |S'|) \leqslant \rho_B|B| + |S'| \leqslant \frac{q}{k} \cdot \dim(S) + |S'|$$

which implies

$$|S'| \geqslant \frac{\rho - (q/k)}{1-\rho} \cdot (\dim(S)) ,$$

thus completing the proof of Corollary 8. □

Proof of Corollary 9. The high level idea is to partition S into a basis B for $S \subseteq \mathcal{C}^{\perp}$ and a set of redundant tests S' such that, roughly speaking, the probability of sampling from B, according to the distribution p associated with the test T, is large. Then we continue as in the proof of Corollary 8.

To construct the said partition start with an arbitrary partition $S = B \cup S'$ with B a basis for $S \subseteq \mathcal{C}^{\perp}$. Iteratively modify the partition as follows. If there exists $u \in S'$ represented in the basis B as $\sum_{b\in B} \alpha_b b$ and $p(b) < p(u)$ for some $b \in B$ with $\alpha_b \neq 0$, then replace b with u, i.e., set B to be $(B \cup \{u\}) \setminus \{b\}$ and S' to be $(S' \cup \{b\}) \setminus \{u\}$. Repeat the process until no such $u \in S'$ exists. Notice the process must terminate because $\sum_{b\in B} p(b)$ is bounded by 1 and there exists $\gamma > 0$ such that with each iteration this sum increases by at least γ.

At the end of the process we have partitioned S into a basis B for S and a redundant set S' with the following property that will be crucial to our proof. For $u \in S'$,

$$p(u) \leqslant p(b) \text{ for all } b \in \{u\}_B \ .$$

(Recall that $\{u\}_B$ is minimal set of basis vectors in B whose span contains u.)

We continue with our proof. Consider the basis tester T' defined by taking the conditional distribution of the testter T on B. Let p' denote the resulting conditional distribution on B. Note that $p'(b) \geqslant p(b)$ for all $b \in B$. By Theorem 5 there exists w that is $\frac{1}{3}\delta(C)$-far from $\mathcal{C}$ and is rejected by T' with probability at most q/k. Let $B' \subseteq B$ be the set of tests that reject w and notice $p(B') \leqslant p'(B') \leqslant q/k$.

Consider a word $u \in S'$ that rejects w and represent u as a linear combination of elements of $\{u\}_B \subseteq B$. Note that if the test u rejects w then there must be some b in $\{u\}_B$ that also rejects w (and hence belongs to B'). By the special properties of our partition which were discussed in the previous paragraph we have

$$p(u) \leqslant p(b) \leqslant p(B') \leqslant q/k \ .$$

Thus, every test that rejects w from S' has probability at most q/k of being performed and furthermore, the probability of rejecting w using an element of B is at most q/k as well. Summing up, we get

$$\rho \quad \leqslant \quad \Pr[T[w] = \textsf{reject}] \leqslant q/k + |S'| \cdot q/k$$

which after rearranging the terms give $|S'| \geqslant \frac{\rho k}{q} - 1$ as claimed. □

4 Proof of Main Technical Theorem 14

4.1 Overview — Proof of a simple case of Theorem 5

Instead of giving an overview for the proof of Main Theorem 14 we prefer to give an overview to the proof of Theorem 5. To explain what goes on in the proof we focus on a relatively simple case. We say that a tester is *smooth* if it has the property that every bit of the input word w is queried by it with equal probability. Let us sketch how to prove a linear lower bound on the query complexity q of a (q, ρ)-strong smooth and uniform basis tester for a $[n, k = \kappa n, d = \delta n]_{\mathbb{F}_2}$-code $\mathcal{C}$ over the two-element field $\mathbb{F}_2$. Namely, we will show $q = \Omega(n)$.

Let $B = \{u_1, \ldots, u_{n-k}\}$ be the set of tests selected (uniformly) by our smooth basis tester T. By assumption B is a basis for $\mathcal{C}^\perp$ and contains words of size at most q.

The main idea implemented in the proof is to build a special basis for $\mathbb{F}^n$ using the code $\mathcal{C}$ and the basis B. Specifically we define a set $V = \{v_1, \ldots, v_{n-k}\}$ such that for every word $w \in \mathbb{F}^n$ we can find a codeword $c_w \in \mathcal{C}$ and a set $V_w \subseteq V$ such that $w = c_w + \sum_{v \in V_w} v$. (Specifically, we build such a set V by letting $v_j \in \mathbb{F}_2^n \setminus C$ such that v_j has inner product zero with u_i for every $i \neq j$ and inner product one with u_j.)

We note that in this basis, the rejection probability of the basis tester based on B is straightforward to compute. A word w is rejected with probability exactly $|V_w|/|V|$. (This follows from the fact that u_i rejects w iff $v_i \in V_w$.)

Since this applies also to the elements $v_i \in V$ also, we conclude they have small weight. Specifically, using the assumption that B is a (q, ρ)-strong tester we conclude

$$\rho \cdot \frac{|v_i|}{n} = \rho\delta(v_i, C) \leqslant \Pr[T[v_i] = \textsf{reject}]$$

$$= \frac{|\{v_i\}|}{|V|} \leqslant \frac{1}{(1-\kappa)n}$$

which gives $|v_i| \leqslant \frac{1}{\rho(1-\kappa)} = O(1)$.

The non-trivial step now is to consider the probability of rejecting some *low-weight* words. Specifically we consider the probability of rejecting the "unit" vector e_i in the standard basis. I.e., $e_i = 0^{i-1}10^{n-i}$. On the one hand, smoothness implies this word can not be rejected with high-probability if the query complexity is low (since its weight is so low). On the other hand, we note that for some i, the set V_{e_i} has to be large and so it must be rejected with high probability. This leads to a contradiction to the assumption that the query complexity is low. We give more details below.

Note that there must exist a vector e_i whose representation is

$$e_i = c_{e_i} + \sum_{v_j \in V_{e_i}} v_j$$

where c_{e_i} is a *nonzero* codeword. This is because $e_1, \ldots, e_n$ are linearly independent, so they cannot all belong to $\mathrm{span}(V)$ which is a $(n-k)$-dimensional space. The crucial observation is that $|V_{e_i}|$ must be large. This is because $|v_j| \leqslant \frac{1}{\rho(1-\kappa)}$ and $|c_{e_i}| \geqslant \delta n$ so $|V_{e_i}| \geqslant \frac{\delta}{\rho(1-\kappa)} n$. This implies that e_i is rejected with probability

$$\frac{|V_{e_i}|}{|V|} \geqslant \frac{\frac{\delta}{\rho(1-\kappa)} n}{(1-\kappa)n} = \frac{\delta}{\rho}.$$

On the other hand, the assumption of smoothness implies rejection probability of e_i is precisely the probability of querying the ith coordinate which is $\frac{q}{n-k} = \frac{q}{(1-\kappa)n}$. We conclude

$$\frac{\delta}{\rho} \leqslant \frac{|V_{e_i}|}{|V|} = \Pr[T[e_i] = \textsf{reject}] = \frac{q}{(1-\kappa)n}$$

which gives $q \geqslant \frac{\delta(1-\kappa)}{\rho} n = \Omega(n)$ as claimed.

Our proof of Theorem 14 follows the outline laid above. The noticeable differences are that the tester need not be smooth, nor uniform, and the field size may be greater than 2. Furthermore, we think of words represented in an arbitrary basis B for $\mathcal{C}^\perp$ and show that *many* words will be simultaneously far from $\mathcal{C}$ *and* accepted by the tester with high probability. But the overall picture is roughly the same. Now for the details.

4.2 The $(\mathcal{C}, V)$-representation of words in $\mathbb{F}^n$

Let $B = \{u_1, \ldots, u_{n-k}\} \subseteq \mathcal{C}^\perp_{\leqslant q}$ be a basis for $\mathcal{C}^\perp$ obtained by starting with a basis for S and completing it to a basis for $C^\perp$ in an arbitrary manner.

The first part of our proof shows that every word in $\mathbb{F}^n$ can be represented uniquely as the sum of a codeword in $\mathcal{C}$ and a subset of a set of vectors $V = \{v_1, \ldots, v_{n-k}\}$ where the rejection probability of w is related to its representation structure. We start by defining V.

Definition 17. For $i \in [n-k]$ let v_i be a word of minimal weight that satisfies

$$\langle v_i, u_j \rangle = \begin{cases} 1 & i = j \\ 0 & j \in [n-k] \setminus \{i\} \end{cases} \tag{1}$$

and let $V = \{v_1, \ldots, v_{n-k}\}$.

Proposition 18. *For all $v_i \in V$ we have $\frac{\mathrm{wt}(v_i)}{n} = \delta(v_i, \mathcal{C})$.*

Proof. We have $\delta(v_i, \mathcal{C}) \leqslant \frac{\mathrm{wt}(v_i)}{n}$ because $\delta(v_i, 0^n) = \frac{\mathrm{wt}(v_i)}{n}$ and $0^n \in \mathcal{C}$. On the other hand, for every $c \in \mathcal{C}$ we must have $\delta(v_i, c) \geqslant \frac{\mathrm{wt}(v_i)}{n}$ because if $\delta(v_i, c) < \frac{\mathrm{wt}(v_i)}{n}$ then setting $v_i' = v_i - c$ we have $\mathrm{wt}(v_i') < \mathrm{wt}(v_i)$ but v_i' satisfies (1) (with respect to index i), thus contradicting the minimal weight of v_i. □

The following claim states that $\mathbb{F}^n$ is the direct sum of the code $\mathcal{C}$ and $\mathrm{span}(V)$.

Claim 19. $\dim(\mathrm{span}(\mathcal{C} \cup V)) = n$ *and* $\dim(V) = n - k$.

Proof. Let $Z = C \cup V$. To prove both equalities stated in our claim it is sufficient to show that $Z^\perp = \{0^n\}$, i.e., that $\dim(Z^\perp) = 0$, because $\dim(\mathcal{C}) = k$ and $|V| = n-k$. Assume by way of contradiction that $u \in Z^\perp$ is nonzero. Then in particular $u \in \mathcal{C}^\perp$ because $\mathcal{C} \subseteq Z$ which implies $\mathcal{C}^\perp \supseteq Z^\perp$. Thus, u is a nonzero linear combination of vectors from B because B is a basis for $\mathcal{C}^\perp$. Suppose u_i appears in the representation of u under B. Then from (1) we conclude $\langle u, v_i \rangle \neq 0$ which implies $u \notin V^\perp$ which gives $u \notin Z^\perp$, contradicting the assumption $u \in Z^\perp$. So $\dim(Z^\perp) = 0$ and this completes our proof. □

Claim 19 shows that every $w \in \mathbb{F}^n$ has a unique representation as a sum of a single element from C, denoted $c(w)$, and a linear combination of v_j's, denoted $v(w)$. We say $(c(w), v(w))$ is the $(\mathcal{C}, V)$-*representation* of w. We denote by $\Gamma(w) \subseteq [n-k]$ the set of indices (j) of v_j's participating in $v(w)$. Formally, if $v(w) = \sum_{j=1}^{n-k} \alpha_j v_j$ then

$$\Gamma(w) = \{j \mid \alpha_j \neq 0\}.$$

The next claim relates the rejection probability of w by our basis tester to the structure of $v(w)$. For $i \in [n-k]$ let $p(i) = p(u_i)$ denote the probability of u_i under the distribution associated with our basis tester. For $I \subseteq [n-k]$ the set of indices of $B' \subseteq B$ let $p(I) = p(B') = \sum_{i \in I} p(i) = \sum_{u_i \in B'} p(u_i)$.

Claim 20 (Rejection probability is related to $(\mathcal{C}, V)$-representation structure)**.** *For all $w \in \mathbb{F}^n$ we have*

$$\Gamma(w) = \{j \in [n-k] \mid \langle u_j, w \rangle \neq 0\}. \tag{2}$$

Consequently, we have

$$\Pr[T[w] = \textsf{reject}] = p(\Gamma(w)).$$

Proof. Consider the $(\mathcal{C}, V)$-representation of w:

$$w = c(w) + \sum_{j \in \Gamma(w)} \alpha_j v_j, \text{ where } \alpha_j \neq 0.$$

By assumption for all $u_i \in B$ we have $\langle u_i, c(w) \rangle = 0$ and by (1) we have $\langle u_i, v(w) \rangle \neq 0$ if and only if $i \in \Gamma(w)$. This implies (2). The consequence follows because, by definition, the probability of rejecting w is the probability of the event $\langle u_i, w \rangle \neq 0$ where u_i is selected from B with probability $p(i)$. This completes the proof. □

4.3 Main Lemma and Proof of Main Theorem 14

The following lemma is the main part of our proof. Assuming it we shall promptly complete the proof of Theorem 14 and the proof of the lemma comes after the proof of the theorem. In what follows the *singleton vector* $e_i = 0^{i-1}10^{n-i}$ is the characteristic vector of the singleton set $\{i\} \subset [n]$.

Lemma 21 (Main Lemma)**.** *If $\mathcal{C}$ is an $[n, k, d]_{\mathbb{F}}$-code and $B = \{u_1, ..., u_{n-k}\}$ is a basis for $C^{\perp}$, then there exist k distinct indices $i_1, ..., i_k \in [n]$ and k corresponding words $w_{i_1}, ..., w_{i_k} \in \mathbb{F}^n$ such that for every i_j the following two conditions hold:*

- $\delta(w_{i_j}, \mathcal{C}) \geqslant \frac{\delta(\mathcal{C})}{3}$.
- *For every $u \in B$ we have $\langle u, w_{i_j} \rangle \neq 0$ only if $i_j \in \mathrm{supp}(u)$.*

Proof of Theorem 14. We apply Main Lemma 21, and without loss of generality we may assume that $\{i_1, ..., i_k\} = [k]$. Recall that we have $w_1, ..., w_k$ such that for every $i \in [k]$ it holds that $\delta(w_i, C) \geqslant \frac{\delta(\mathcal{C})}{3}$ and for every $u \in B$ we have $\langle u, w_i \rangle \neq 0$ only if $i \in \mathrm{supp}(u)$.
For $i \in [k]$ let $B_{w_i} = \{u \in B \mid \langle u, w_i \rangle \neq 0\}$. Note that $B_{w_i} \subseteq \{u \in B \mid i \in \mathrm{supp}(u)\}$, so we have the following immediate claim (proof omitted), which will be used later in the proof of the theorem.

Claim 22. *For every $i \in [k]$ and $u \in B$ it holds that if $u \in B_{w_i}$ then $i \in \mathrm{supp}(u)$ and thus u can belong to at most $|\mathrm{supp}(u)| = |u|$ different B_{w_j}-s.*

We continue the proof of Theorem 14. For all $i \in [k]$ we have

$$\mathop{\mathbf{Pr}}_{u \in_{\mathcal{D}} S}[\langle u, w_i \rangle \neq 0] \geqslant \rho$$

because $\mathcal{D}$ is $(q, \frac{\delta(\mathcal{C})}{3}, \rho)$ tester of $\mathcal{C}$. Hence for all $i \in [k]$ we have

$$\mathop{\mathbf{Pr}}_{u \in_{\mathcal{D}} S}[|\{u\}_B \cap B_{w_i}| \geqslant 1] \geqslant \rho.$$

So by linearity of expectation:

$$\mathop{\mathbf{E}}_{u \in_{\mathcal{D}} S}[|\{u\}_B \cap B_{w_1}| + |\{u\}_B \cap B_{w_2}| + ... + |\{u\}_B \cap B_{w_k}|] \geqslant \rho k.$$

Let us consider

$$|\{u\}_B \cap B_{w_1}| + |\{u\}_B \cap B_{w_2}| + ... + |\{u\}_B \cap B_{w_k}|.$$

Let $m = |\{u\}_B|$ and $\{u\}_B = \{u_1, ..., u_m\}$. Note that m, u and $u_1, ..., u_m$ are random variables. Let $X_{i,j}$ to be an indicator variable for the event "$u_i \in B_{w_j}$", i.e. $X_{i,j}$ equals 1 if $u_i \in B_{w_j}$ and equals 0 otherwise. Then

$$|\{u\}_B \cap B_{w_1}| + |\{u\}_B \cap B_{w_2}| + ... + |\{u\}_B \cap B_{w_k}| = \sum_{j=1}^{k} \sum_{i=1}^{m} X_{i,j} = \sum_{i=1}^{m} \sum_{j=1}^{k} X_{i,j}$$

Note that $B_{w_1} \cup ... \cup B_{w_k} \subseteq B$. By Claim 22 u_i is contained in at most $|u_i|$ sets B_{w_j} and thus we have

$$|\{u_i\} \cap B_{w_1}| + |\{u_i\} \cap B_{w_2}| + ... + |\{u_i\} \cap B_{w_k}| = \sum_{j=1}^{k} X_{i,j} \leqslant |u_i|$$

So,

$$|\{u\}_B \cap B_{w_1}| + |\{u\}_B \cap B_{w_2}| + ... + |\{u\}_B \cap B_{w_k}| = \sum_{j=1}^{k} \sum_{i=1}^{m} X_{i,j} = \sum_{i=1}^{m} \sum_{j=1}^{k} X_{i,j} \leqslant \sum_{i=1}^{m} |u_i| = \sum_{u_i \in \{u\}_B} |u_i|$$

Thus

$$\mathop{\mathbf{E}}_{u \in_{\mathcal{D}} S}[\mathrm{avg}(\{u\}_B) \cdot |u|_B] = \mathop{\mathbf{E}}_{u \in_{\mathcal{D}} S}\left[\sum_{u_i \in \{u\}_B} |u_i|\right] \geqslant \rho k.$$

This completes the proof of Theorem 14 from Lemma 21. □

Proof of Lemma 21. We start by showing that there exist k distinct singleton vectors, denoted without loss of generality $e_1, \ldots, e_k$, such that $c(e_1), \ldots, c(e_k)$ are linearly independent, hence distinct and nonzero.

Since every word in $\mathbb{F}^n$ has a unique $(\mathcal{C}, V)$-representation we get $e_i \in \{c(e_i) + v \mid v \in \mathrm{span}(V)\}$. This implies

$$\{e_1, \ldots, e_n\} \subseteq \mathrm{span}(\{c(e_1), \ldots, c(e_n)\} \cup V).$$

Counting dimensions, we have

$$\begin{aligned} n &= \dim(\mathrm{span}(\{e_1, \ldots, e_n\})) \\ &\leqslant \dim(\mathrm{span}(\{c(e_1), \ldots, c(e_n)\} \cup V)) \\ &\leqslant \dim(\mathrm{span}(\{c(e_1), \ldots, c(e_n)\})) + \dim(\mathrm{span}(V)). \end{aligned}$$

By Claim 19 we have $\dim(\text{span}(V)) = n-k$, so we conclude that (without loss of generality) $c(e_1), \ldots, c(e_k)$ are linearly independent, as claimed.

Next, we argue that for $i \in [k]$ we have $|v(e_i)| \geqslant d-1$. This is because $e_i = c(e_i) + v(e_i)$ and $|e_i| = 1$ and $|c(e_i)| \geqslant d$ because $c(e_i)$ is a nonzero word in a linear code with minimal distance d.

So far we have shown that every $v(e_i), i \in [k]$ we have $|v(e_i)| \geqslant d-1$

Let $i \in [k]$ and let us show that there exists $w_i \in \text{span}(\{v_j \mid j \in \Gamma(e_i)\})$ such that $\delta(w_i, \mathcal{C}) \geqslant \frac{\delta(\mathcal{C})}{3}$. Note that in this case for all $u \in B$ we have $\langle u, w_i \rangle \neq 0$ only if $i \in \text{supp}(u)$. Now, if $|v_j| \geqslant \frac{1}{3}d$ for some $j \in \Gamma(e_i)$ then setting $w_i = v_j$ completes the proof because Proposition 18 implies that v_j is $\frac{d}{3n}$-far from $\mathcal{C}$.

From here on we assume $|v_j| < \frac{1}{3}d$ for all $j \in \Gamma(e_i)$. Let $t = |\Gamma(e_i)|$ and assume wlog $\Gamma(e_i) = [t]$. Denote the $(\mathcal{C}, V)$-representation of e_i by $c(e_i) + \sum_{j=1}^{t} \alpha_j v_j$ where $\alpha_j \neq 0$. Let $w_\ell = \sum_{j=1}^{\ell} \alpha_j v_j$. We know the following:

- $|w_1| < \frac{1}{3}d$.
- $|w_t| = |v(e_i)| \geqslant d-1$ by the second bullet in Lemma 21.
- $|w_{\ell+1}| \leqslant |w_\ell| + |v_{\ell+1}| < |w_\ell| + \frac{1}{3}d$ for all $1 \leqslant \ell < t$, by the triangle inequality.

This implies the existence of some $\ell \in [t]$ such that $\frac{1}{3}d < |w_\ell| \leqslant \frac{2d}{3}$ and notice $w_i = w_\ell$ is $\frac{d}{3n}$-far from $\mathcal{C}$. To see this note that $\delta(w_\ell, 0) = \frac{|w_\ell|}{n} \geqslant \frac{d}{3n}$ and for every $c \in \mathcal{C}$ such that $c \neq 0^n$ we have $\delta(w, c) \geqslant |c| - |w_\ell| \geqslant d - \frac{2d}{3} = \frac{d}{3}$, where the last inequality follows since $|c| \geqslant \Delta(\mathcal{C}) \geqslant d$. $\square$

5 Tightness of Main Theorem 5

In this section we argue that the bound ($k \leqslant \frac{q}{\rho}$) obtained in Theorem 5 is close to tight. In Proposition 23 we show that there are codes with constant relative distance and constant dimension which have a basis tester, and in Proposition 24 we show in all codes have a basis tester whose query complexity equals to the dimension of the code plus one. Propositions 23 and 24 are folklore.

Proposition 23 (The repetition code has a $(2,1)$-strong uniform basis tester)**.** *For any finite field $\mathbb{F}$ and constant $c \in \mathbb{N}^+$ there exists a $[n = cm, k = c, d = m]_{\mathbb{F}}$-code $\mathcal{C}$ which has a $(2,1)$-strong basis tester.*

Proof. Let $\mathcal{C}$ be the $[n = cm, k = c, d = m]_{\mathbb{F}}$ repetition code where a c-symbol message $a_1, \ldots, a_c$ is encoded by repeating each symbol m times, i.e., $a_1, \ldots, a_c \mapsto a_1^m, \ldots, a_c^m$. Consider the tester that compares a random position in a block to the first symbol in the block. Formally, the tester is defined by the uniform distribution over the following set B of words of weight 2 : $B = \{e_{im+1} - e_{im+j} \mid i \in \{0, \ldots, c-1\}, j \in \{2, \ldots, m\}\}$, where e_ℓ has a 1 in the ℓth coordinate and is zero elsewhere.

It can be readily verified that B is a basis for $\mathcal{C}^\perp$, has query complexity 2 and rejects a word w with probability $\delta(w, \mathcal{C})$ because if the rejection probability is ε this means that at most an ε fraction of symbols need to be changed to reach a word that is constant on each of its c blocks. $\square$

Proposition 24 (Every code has a basis tester with large query complexity)**.** *Let $\mathbb{F}$ be a finite field and $\mathcal{C}$ be a $[n, k, d]_{\mathbb{F}}$ code. Then $\mathcal{C}$ has a $(k+1, 1)$ strong uniform basis tester.*

Proof. Let $G \in \mathbb{F}^{n \times k}$ be a generating matrix for $\mathcal{C}$, i.e., $C = \left\{Gm \mid m \in \mathbb{F}^k\right\}$. Assume without loss of generality that the first k rows of G are linearly independent. This means that after querying the first k symbols of a word $w_1, \ldots, w_k$, one can interpolate to obtain any other symbol of the codeword that is the encoding of the message $m \in \mathbb{F}^k$ such that $(Gm)|_{[k]} = (w_1, \ldots, w_k)$. For $k < i \leqslant n$ let u_i be the constraint that queries the first k symbols of w and accepts iff w_i is equal to the ith symbol of the encoding of m. It can be readily verified that $B = \{u_{k+1}, \ldots, u_n\}$ is a basis for $\mathcal{C}^\perp$ and has query complexity $k + 1$.

Consider the soundness of the uniform tester over B. If $\mathbf{Pr}[T[w] = \mathsf{reject}] \leqslant \rho$ then w is ρ-close to the codeword of $\mathcal{C}$ that is the encoding of m, implying that $\delta(w, \mathcal{C}) \leqslant \delta(w, Gm) \leqslant \rho$. □

6 Bounds on tester support size

We show that every binary linear code $\mathcal{C}$ can be tested with linear redundancy, by proving the following statement. We point out that [6] implicitly showed already that every code can be tested with a linear amount of redundancy. The added value of the following statement is that it shows that the amount of redundancy can be as small as twice the dimension of $\mathcal{C}^\perp$.

Proposition 25. *If $\mathcal{C} \subseteq \mathbb{F}_2^n$ is a $[n, k, d]_{\mathbb{F}_2}$-code that is a (q, ε, ρ)-LTC, then it has $(\frac{10q}{\rho}, \varepsilon, \frac{1}{100})$-tester whose support is over a set U of size at most $3\dim(\mathcal{C}^\perp)$.*

The proof of the above result appears in Section 6.1.

Remark 26. Inspection of the proof of Proposition 25 reveals that $\mathcal{C}$ can be tested by a $(c \cdot q, \varepsilon, 1/c)$-tester whose support is over U of size $\leqslant (4\ln 2 + \eta) \cdot (n - k)$ for any $\eta > 0$, where $c > 1$ is a constant that depends on η and goes to infinity as η goes to 0. Recalling $4\ln 2 = 2.77258\ldots$, we preferred to round this constant up to the closest integer in the statement of the proposition above.

Remark 27. The assumption that the code is binary is not crucial, and Proposition 25 can be stated for the fields of larger cardinality, but the size of U will increase respectively.

While the support of a non-strong tester need not span $\mathcal{C}^\perp$, we can prove that every tester's support must at least span a large subspace of $\mathcal{C}^\perp$.

Proposition 28. *Let T be a (q, ε, ρ)-tester for a linear code $\mathcal{C} \subseteq \mathbb{F}^n$ such that $\varepsilon \leqslant \frac{\delta(C)}{3}$. Let $U \subseteq \mathcal{C}^\perp_{\leqslant q}$ denote the support of T. Then $\dim(U) \geqslant \dim(\mathcal{C}^\perp) - 3\varepsilon n$.*

The proof of the above result appears in Section 6.2.

6.1 Proof of Proposition 25

Let us state a couple of inequalities in probability that will be used later on in the proof.

Claim 29 (Chernoff Bound)**.** *If $X = \sum_{i=1}^m X_i$ is a sum of independent $\{0,1\}$-valued random variables, where $\mathbf{Pr}[X_i = 1] = \gamma$, then*

$$\mathbf{Pr}[X < (1 - \sigma)\gamma m] \leqslant \exp(-\sigma^2 \gamma m/2).$$

Claim 30. *If $X = \sum_{i=1}^m X_i$ is a sum of independent $\{0,1\}$-valued random variables, where $\mathbf{Pr}[X_i = 1] = \gamma$, then*

$$\mathbf{Pr}[X \equiv 0 \pmod 2] \leqslant \frac{1}{2}(1 + \exp(-2\gamma m)).$$

Proof of Proposition 25. Let $t = \frac{10}{\rho}$ and $m = 3\dim(\mathcal{C}^{\perp}) = 3(n-k)$. Let T be the assumed (q, ε, ρ) tester for $\mathcal{C}$. Pick $U = \{u_1, \ldots, u_m\}$ where each u_i is obtained by taking the sum of t independent samples from T. U is a multiset and the distribution p associated with our tester is the uniform distribution over U. The query complexity of U is bounded by $tq = \frac{10q}{\rho}$.

To analyze soundness, fix a word w that is ε-far from $\mathcal{C}$. Let X_i be the indicator random variable for the event $\langle w, u_i \rangle \neq 0$. By Claim 30 it holds that $\mathbf{Pr}[X_i = 0] \leqslant \dfrac{1}{2}(1 + e^{-2\rho t})$ and $\mathbf{Pr}[X_i = 1] \geqslant \dfrac{1}{2}(1 - e^{-2\rho t})$.

Let $U_{bad} = \{u \in U \mid \langle u, w \rangle \neq 0\}$. We know that $\mathbf{E}[|U_{bad}|] \geqslant \frac{m}{2}(1 - e^{-2\rho t}) \geqslant \frac{m}{2}(0.999)$. Let $\sigma = 0.979$ and note that $\frac{m}{2}(0.999)(1-\sigma) = \frac{0.021m}{2}(0.999) > \frac{m}{100}$. Then by the Chernoff bound (Claim 29) we have

$$\mathbf{Pr}\left[\frac{|U_{bad}|}{m} < \frac{1}{100}\right] = \mathbf{Pr}\left[|U_{bad}| < \frac{m}{100}\right] \leqslant e^{-0.979^2(\frac{1}{2}(1-e^{-2\rho t}))m/2}$$

We take a union bound over all words that are ε-far from $\mathcal{C}$. Notice that $\mathbb{F}_2^n$ can be partitioned into 2^{n-k} affine shifts of (the linear space) $\mathcal{C}$. For each such affine shift, which has the form $v + \mathcal{C} = \{v + c \mid c \in \mathcal{C}\}$, the probability of rejecting any two words from $v + \mathcal{C}$ is equal, because they differ only by a word from $\mathcal{C}$ which has inner product 0 with all tests. Thus, it suffices to take a union bound over one representative per affine shift, and there are at most 2^{n-k} of them.

Continuing with the proof, the probability that there exists a ε-far word that is rejected with probability less than $\frac{1}{100}$ is at most $e^{-0.979^2(\frac{1}{2}(1-e^{-2\rho t}))m/2} \cdot 2^{n-k} = e^{-0.979^2(\frac{1}{2}(1-e^{-2\rho t}))m/2 + \ln(2)(n-k)} < 1$.

The inequality follows since by construction we have $m > 2.95(n-k)$ and hence $m > \frac{2.773(n-k)}{0.979^2}$. So $-0.979^2(\frac{1}{2}(1-e^{-2\rho t}))m/2 + \ln(2)(n-k) < 0$ and $e^{-0.979^2(\frac{1}{2}(1-e^{-2\rho t}))m/2 + \ln(2)(n-k)} < 1$.

Hence we showed that there is a positive probability to pick the set U such that every ε-far word is rejected with probability at least $\frac{1}{100}$ and this completes the proof. $\square$

6.2 Proof of Proposition 28

Proof of Proposition 28. Assume by way of contradiction that $\dim(U) < \dim(\mathcal{C}^{\perp}) - 3\varepsilon n$. We call a word w a *coset leader* if w has minimal weight in $w + \mathcal{C} = \{w + c \mid c \in \mathcal{C}\}$. (If there is more than one minimal weight word in $w + \mathcal{C}$ pick arbitrarily one of them to be the coset leader.) The proof of Proposition 18 implies that if w is a coset leader then $\frac{\mathrm{wt}(w)}{n} = \delta(w, C)$.

Let

$$V = \{w \in \mathbb{F}^n \setminus \mathcal{C} \mid \forall u \in U : \langle u, w \rangle = 0$$
$$\text{and } w \text{ is a coset leader of } w + \mathcal{C}\},$$

i.e., V contains all non-codewords that are coset leaders and accepted by all tests in U. Notice that for all $w \in V$ we have $\delta(w, \mathcal{C}) = \delta(w, 0) = |w|$ because w is a coset leader of $w + \mathcal{C}$.

We argue that $\dim(V) \geqslant 3\varepsilon n$ and thus $|\bigcup_{v \in V}(\mathrm{supp}(v))| \geqslant \dim(V) \geqslant 3\varepsilon n$. We show that $\dim(V) \geqslant 3\varepsilon n$. It sufficient to prove that $\dim(V) \geqslant \dim(\mathcal{C}^{\perp}) - \dim(U) > 3\varepsilon n$. Fix any basis $U_B = \{u'_1, ..., u'_m\}$ for U and complete it to a basis for $\mathcal{C}^{\perp}$ in an arbitrary manner obtaining $U_B \cup R_B$, where $R_B = \{u_1, ..., u_h\}$, i.e., $U_B \cap R_B = \emptyset$ and $\mathrm{span}(U_B \cup R_B) = \mathcal{C}^{\perp}$. Note that $h \geqslant 3\varepsilon n$. We argue that $\dim(V) = h$, this is true since for every $u_j \in R_B$ we have unique coset leader $v_j \in \mathbb{F}^n$ (up to multiplication by a nonzero element of $\mathbb{F}$), such that $\langle u_j, v_j \rangle \neq 0$ and $\langle u, v_j \rangle = 0$ for all $u \in (R_B \setminus \{u_j\}) \cup U_B$. Moreover, all these v_j-s are linearly independent by definition.

In addition for all $v \in V$ we have $|\mathrm{supp}(v)| = \delta(w, \mathcal{C}) < \varepsilon n$ because

$$\mathbf{Pr}[T[v] = \text{reject}] = 0.$$

Let $w_1, \ldots, w_s$ be an arbitrary ordering of the elements of V. Let $\mu(\ell)$ the maximal weight of an element in $\mathrm{span}(w_1, \ldots, w_\ell)$. We have $\mu(1) \leqslant \varepsilon n$ and $\mu(s) \geqslant \frac{3}{2}\varepsilon n$ because the expected weight of a word in $\mathrm{span}(V)$ is (exactly) $\frac{|\mathbb{F}|-1}{|\mathbb{F}|}|\bigcup_{w\in V}(\mathrm{supp}(w))|$. To see that the expected weight of a word $w_{exp} \in \mathrm{span}(V)$ is as claimed note that w_{exp} is picked by a random linear combination of vectors in V, where each vector $v \in V$ ia taken independently with probability $1 - \frac{1}{|\mathbb{F}|}$. Hence if $i \in \bigcup_{w\in V}(\mathrm{supp}(w))$ then $i \in \mathrm{supp}(w_{exp})$ with probability $\frac{|\mathbb{F}|-1}{|\mathbb{F}|} \geqslant \frac{1}{2}$.

Finally, we have $\mu(\ell+1) < \mu(\ell) + \varepsilon n$. We conclude there must exist ℓ for which $\varepsilon n < \mu(\ell) \leqslant 2\varepsilon n$. Let w' be a word of maximal weight in $\mathrm{span}(w_1, \ldots, w_\ell)$. We have that $\varepsilon n < |w'| \leqslant 2\varepsilon n$. We conclude that $\Delta(w, 0) = |w'| > \varepsilon n$ and for all $c \in \mathcal{C}$ such that $c \neq 0^n$ we have $\Delta(w, c) \geqslant |c| - |w| \geqslant d - 2\varepsilon n \geqslant d - \frac{2d}{3} = \frac{d}{3}$.

We see that w' is ε-far from $\mathcal{C}$ but accepted by T with probability 1, a contradiction. □

7 Open questions and Discussion

In this section we discuss a possible approach to obtain the upper bounds on the rate of LTCs. Recall that the main open question in the subject of locally testable codes is whether there are asymptotically good LTCs. We believe that there is no asymptotically good LTC and suggest a way that may lead to proving that these codes do not exist. One of the ways to disprove the existence of certain kind of LTCs is to assume its existence and its tester, and then show a counter-example w, which will be far from the code but rejected with very small probability. The main challenge in this way is to construct a specific counter-example that "cheats" the tester for the given LTC. In particular, the problem is to argue that w will satisfy almost all local constraints which can be selected by the tester.

We suggest to consider the basis for a dual code, and to construct many different counter-examples that will be far from each other and far from the code, but each of them will satisfy all but one basis constraint. More formally, we suggest to show that that if $\mathcal{C} \subseteq \mathbb{F}^n$ is an (q, ε, ρ)-asymptotically good LTC and $B \subseteq \mathcal{C}^{\perp}_{\leqslant q}$ is a basis for $\mathcal{C}^{\perp}_{\leqslant q}$ then there exist $\Omega(n)$ different words v_j, where every word is $\Omega(1)$ far from $\mathcal{C}$ and from the other v_j-s, but satisfies all but one constraints from B. This creates a large space of potential counter-examples in $\mathbb{F}^n$. If this approach succeeds then one might show that at least one of v_j-s satisfies almost all local constraints of the tester. Now to the details.

Assume the existence of asymptotically good (q, ε, ρ)-LTC $\mathcal{C} \subseteq \mathbb{F}^n$ whose tester has support S. Note that $k = \dim(C) = \Omega(n)$. Let $B \subseteq \mathcal{C}^{\perp}_{\leqslant q}$ be a corresponding basis.

The technique of the Main Theorem 14 implies that there are $\Omega(k)$ different v_i such that each one is $\Omega(1)$ far from $\mathcal{C}$. To see this note that for each $i \in [k]$ we have:

$$e_i = c_i + \sum_{j\in J_i} v_j; c_i \in C \setminus \{0\}$$

Let $h = \frac{2q(n-k)}{k}$. Note that $h = O(1)$ since $k = \Omega(n)$. We say that $j \in [k]$ has *high degree* if for at least h different $u \in B$ it holds that $j \in \mathrm{supp}(u)$. The number of high degree indices $j \in [k]$ is bounded above by $\frac{q(n-k)}{h} = \frac{k}{2}$. Thus the number of low degree indices $j \in [k]$ is at least $k - k/2 = k/2$. Without loss of generality we assume that all $i \in [k/2]$ have low degree, i.e.

$$e_i = c_i + \sum_{j\in J_i} v_j; c_i \in C \setminus \{0\} \text{ and } |J_i| \leqslant h$$

Hence that all $i \in [k/2]$ we have $\sum_{j \in J_i} |v_j| \geqslant \Delta(\mathcal{C}) - 1$ and there exists v_j ,$|v_j| \geqslant \frac{\Delta(\mathcal{C})-1}{h}$ and $j \in J_i$.

Each v_j can be counted at most q times, since $|\mathrm{supp}(u_j)| \leqslant q$. We conclude that there are at least $\frac{k}{2q}$ different v_j such that for every v_j we have $\Delta(v_j, \mathcal{C}) \geqslant \frac{\Delta(\mathcal{C})-1}{h}$. We believe that a constant fraction of them should be also far from each other and that this should somehow result in additional restrictions on LTCs. Hence, assuming the existence of asymptotically good LTC $\mathcal{C}$, one should get $\Omega(n)$ different v_j, where each one is $\Omega(1)$ far from $\mathcal{C}$ and from the other v_j-s.

Acknowledgments

We would like to thank Oded Goldreich for many valuable discussions including raising the question as to whether one can show LTCs can perform as well as random codes (a question that partly inspired this work, though not resolved yet). We also thank Oded and the anonymous referees for valuable comments on an earlier version of this article. We would like to thank Or Meir for pointers to the literature. We thank Yicheng Pan for pointing out an error in Section 7 in a previous version of the paper.

References

[1] *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*. IEEE Computer Society, 2009.

[2] N. Alon, J. Bruck, J. Naor, M. Naor, and R. M. Roth, "Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs," *IEEE Transactions on Information Theory*, vol. 38, no. 2, p. 509, 1992.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, "Proof verification and the hardness of approximation problems," *Journal of the ACM*, vol. 45, no. 3, pp. 501–555, May 1998.

[4] S. Arora and S. Safra, "Probabilistic checking of proofs: A new characterization of NP," *Journal of the ACM*, vol. 45, no. 1, pp. 70–122, Jan. 1998.

[5] L. Babai, A. Shpilka, and D. Stefankovic, "Locally testable cyclic codes," in *Proceedings: 44th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2003, 11–14 October 2003, Cambridge, Massachusetts*, IEEE, Ed. IEEE Computer Society Press, 2003, pp. 116–125.

[6] M. Bellare, O. Goldreich, and M. Sudan, "Free bits, PCPs, and nonapproximability—towards tight results," *SIAM Journal on Computing*, vol. 27, no. 3, pp. 804–915, Jun. 1998.

[7] Y. Ben-Haim and S. Litsyn, "Upper bounds on the rate of LDPC codes as a function of minimum distance," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2092–2100, 2006. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TIT.2006.872972

[8] E. Ben-Sasson, O. Goldreich, P. Harsha, M. Sudan, and S. P. Vadhan, "Robust PCPs of proximity, shorter PCPs, and applications to coding," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 889–974, 2006.

[9] E. Ben-Sasson, O. Goldreich, and M. Sudan, "Bounds on 2-query codeword testing," in *RANDOM-APPROX*, ser. Lecture Notes in Computer Science, vol. 2764. Springer, 2003, pp. 216–227. [Online].

Available: http://springerlink.metapress.com/openurl.asp?genre=article&issn=0302-9743&volume=2764&spage=216

[10] 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.

[11] E. Ben-Sasson, P. Harsha, O. Lachish, and A. Matsliah, "Sound 3-query PCPPs are long," in *ICALP (1)*, ser. Lecture Notes in Computer Science, vol. 5125. Springer, 2008, pp. 686–697. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-70575-8_56

[12] 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

[13] E. Ben-Sasson and M. Sudan, "Simple PCPs with poly-log rate and query complexity," in *STOC*. ACM, 2005, pp. 266–275. [Online]. Available: http://doi.acm.org/10.1145/1060590.1060631

[14] ——, "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

[15] E. Ben-Sasson and M. Vidermnan, "Low rate is insufficient for local testability," *Electronic Colloquium on Computational Complexity (ECCC)*, no. 004, 2010. [Online]. Available: http://eccc.hpi-web.de/report/2010/004

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

[17] I. Dinur and O. Reingold, "Assignment testers: Towards a combinatorial proof of the PCP theorem," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 975–1024, 2006. [Online]. Available: http://dx.doi.org/10.1137/S0097539705446962

[18] I. Dinur, M. Sudan, and A. Wigderson, "Robust local testability of tensor products of LDPC codes," in *APPROX-RANDOM*, ser. Lecture Notes in Computer Science, vol. 4110. Springer, 2006, pp. 304–315. [Online]. Available: http://dx.doi.org/10.1007/11830924_29

[19] 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

[20] 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.

[21] E. Grigorescu, T. Kaufman, and M. Sudan, "2-transitivity is insufficient for local testability," in *IEEE Conference on Computational Complexity*. IEEE Computer Society, 2008, pp. 259–267. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/CCC.2008.31

[22] V. Guruswami, "On 2-query codeword testing with near-perfect completeness," in *ISAAC*, ser. Lecture Notes in Computer Science, vol. 4288. Springer, 2006, pp. 267–276. [Online]. Available: http://dx.doi.org/10.1007/11940128_28

[23] 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

[24] ——, "Algebraic property testing: the role of invariance," in *STOC*. ACM, 2008, pp. 403–412. [Online]. Available: http://doi.acm.org/10.1145/1374376.1374434

[25] 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

[26] D. A. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1723–1731, 1996.

[27] M. Sudan, "Algorithmic introduction to coding theory, lecture notes," 2001. [Online]. Available: http://theory.csail.mit.edu/~madhu/FT01/

[28] P. Valiant, "The tensor product of two codes is not necessarily robustly testable," in *APPROX-RANDOM*, ser. Lecture Notes in Computer Science, vol. 3624. Springer, 2005, pp. 472–481. [Online]. Available: http://dx.doi.org/10.1007/11538462_40