THEORY OF COMPUTING, Volume 18 (12), 2022, pp. 1–25
www.theoryofcomputing.org

From Local to Robust Testing via Agreement Testing

Irit Dinur[*] Prahladh Harsha[†] Tali Kaufman[‡] Noga Ron-Zewi[§]

Received March 25, 2019; Revised January 1, 2021; Published June 6, 2022

Abstract. A local tester for an error-correcting code is a probabilistic procedure that queries a small (sublinear) subset of coordinates, accepts codewords with probability one, and rejects non-codewords with probability proportional to their distance from the code. The local tester is said to be *robust* if for non-codewords it satisfies the stronger condition that the average distance of local views from accepting views is proportional to the distance from the code. Robust testing is an important component in constructions of locally testable codes and probabilistically checkable proofs as it allows for composition of local tests.

We show that for certain codes, any (natural) local tester can be converted to a robust tester with roughly the same number of queries. Our result holds for the class of *affine-invariant lifted codes* which is a broad class of codes that includes Reed–Muller codes, as well as recent constructions of high-rate locally testable codes (Guo, Kopparty, and Sudan, ITCS 2013). Instantiating this with known local testing results for lifted codes gives a more

A conference version of this paper appeared in the Proceedings of the 10th Innovations in Theoretical Computer Science Conference, 2019 [15].

[*] Supported by ERC-CoG grant number 772839.

[†] Supported by the Department of Atomic Energy, Government of India, under project no. RTI4001 and in part by the UGC-ISF grant and the Swarnajayanti Fellowship.

[‡] Supported by ERC and BSF grants.

[§] Supported by ISF grant 735/20.

ACM Classification: G.3

AMS Classification: 68Q87

Key words and phrases: local testing, robust soundness, agreement testing, lifted codes, affine-invariant codes

DOI: 10.4086/toc.2022.v018a012

direct proof that improves some of the parameters of the main result of Guo, Haramaty, and Sudan (FOCS 2015), showing robust soundness of lifted codes.

To obtain the above transformation, we relate the notions of local testing and robust testing to the notion of *agreement testing* that attempts to find out whether valid partial assignments can be stitched together to a global codeword. We first show that agreement testing implies robust testing, and then show that local testing implies agreement testing. Our proof is combinatorial, and is based on sampling properties of the collection of local views of local testers. Thus, it immediately applies to local testers of lifted codes that query random affine subspaces in $\mathbb{F}_q^m$, and moreover seems amenable to extension to other families of locally testable codes with expanding families of local views.

1 Introduction

Our main result shows a transformation from *local testing* to *robust testing* for the class of *affine-invariant lifted codes*. We start by describing the notions of local testing, robust testing, and lifted codes.

1.1 Local testing and robust testing

A code is a subset $C \subseteq \Sigma^n$. The elements of Σ^n are called *words*, the elements of C are called *codewords*, Σ is the *alphabet* of the code, and n is the *block length*. The *rate* of the code is the ratio $(\log_{|\Sigma|}|C|)/n$. The code is *linear* if $\Sigma = \mathbb{F}_q$ where $\mathbb{F}_q$ is the finite field of q elements, and C is an $\mathbb{F}_q$-linear subspace of $\mathbb{F}_q^n$. It will be convenient to think of codewords in C as functions $f : U \to \Sigma$ where U is a domain of size n. For a pair of functions $f, g : U \to \Sigma$, we let $\mathrm{dist}(f,g)$ denote their normalized Hamming distance, i. e., the fraction of inputs $x \in U$ for which $f(x) \neq g(x)$. By the *minimum distance* $\mathrm{mindist}(C)$ of the code we mean its minimum normalized distance, i. e., the minimum of $\mathrm{dist}(f,g)$ over all pairs of distinct codewords $f, g \in C$. For a function $f : U \to \Sigma$, we let $\mathrm{dist}(f,C)$ denote the minimum of $\mathrm{dist}(f,g)$ over all codewords $g \in C$.

Let $Q \in \mathbb{Z}^+$ and $\alpha > 0$. A *local tester* for the code C with *query complexity* Q and *soundness* α is a probabilistic oracle algorithm that on oracle access to a function $f : U \to \Sigma$ makes at most Q queries to f, and accepts $f \in C$ with probability one *(completeness)*, while rejecting all words $f \in \Sigma^n$ with probability at least $\alpha \cdot \mathrm{dist}(f,C)$. In this article, we shall restrict our attention to local testers that pick a random subset $K \subseteq U$ of cardinality Q according to some distribution, and accept[1] if and only if $f|_K \in C|_K$. Then the completeness requirement is automatically satisfied ($f|_K \in C|_K$ with probability one whenever $f \in C$), and we only need to consider the soundeness condition,

$$\Pr_K[f|_K \notin C|_K] \geq \alpha \cdot \mathrm{dist}(f,C) \tag{1.1}$$

for all words f.

[1] Local testers may generally apply a more complex predicate on $f|_K$. However, natural local testers are typically of the restricted form we consider, and moreover it can be shown that a local tester for a linear code must be of this form [8].

In this article we will be interested in the stronger notion of robust soundness.[2] We say that a local tester is *robustly sound* (or just *robust*) if for all words, the average distance of its local views from accepting views is at least proportional to the distance of the given function from the code. That is, instead of (1.1) we now require that

$$\mathbb{E}_K\left[\operatorname{dist}(f|_K, C|_K)\right] \geq \alpha \cdot \operatorname{dist}(f, C) \tag{1.2}$$

for all words f. We refer to α as the *robust soundness* of the tester.

The notion of robust soundness was introduced by Ben-Sasson and Sudan [10] based on analogous notions for probabilistically checkable proofs [7, 18]. Robust soundness is a natural property of local testers that relates the global distance of a function from the code to its local distance from accepting views on local views. Moreover, robust soundness is also an important ingredient in constructions of locally testable codes and probabilistically checkable proofs as it allows for composition of local tests. Specifically, it follows by definition that if a code C is robustly testable with query complexity Q and soundness α, and additionally each local restriction $C|_K$ is locally testable with query complexity Q' and soundness α', then the code C is locally testable with query complexity Q' and soundness $\alpha \cdot \alpha'$. This property is useful when local restrictions can be tested efficiently, which can happen if the code has many symmetries (as is the case with the class of lifted codes considered in this paper), or can be achieved, in the case of probabilistically checkable proofs, by attaching a short proof of proximity.

One can easily observe that (1.2) implies (1.1) since $f|_K \notin C|_K$ whenever $\operatorname{dist}(f|_K, C|_K) > 0$, so robust soundness is a stronger requirement than (standard) soundness in local testing. For the other direction, note that a local tester with soundness α has robust soundness at least α/Q since $\operatorname{dist}(f|_K, C|_K) \geq 1/Q$ whenever $f|_K \notin C|_K$. A natural question is whether this loss in robust soundness is necessary, and whether robust soundness is a strictly stronger notion than (standard) soundness for local testing.

In this paper we shall show that this loss is unnecessary for the class of lifted codes, discussed below.

1.2 Lifted codes

For sets B, Σ let $\{B \to \Sigma\}$ denote the set of all functions that map B to Σ. We can view this set as the set of strings of length $n = |B|$ over Σ. If $\Sigma = \mathbb{F}_q$ then this set is a vector space over $\mathbb{F}_q$ of dimension $|B|$.

Lifted codes are specified by a *base code* $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ and a dimension $m \geq \ell$. We further assume that the base code C is linear and affine invariant. That is, for any codeword $f \in C$, and for any affine transformation $A : \mathbb{F}_q^\ell \to \mathbb{F}_q^\ell$, it holds that $f \circ A \in C$. Given these we define the *lifted code* $C^{\ell \nearrow m}$ to be the code consisting of all functions $f : \mathbb{F}_q^m \to \mathbb{F}_q$ that satisfy that $f|_L \in C$ for any ℓ-dimensional affine subspace L of the domain $\mathbb{F}_q^m$.

Lifted codes were first introduced by Ben-Sasson, Maatouk, Shpilka and Sudan [9], and their local testability properties were further explored in subsequent work [24, 25, 23]. They are a natural generalization of the well-studied family of Reed–Muller codes, and moreover they also give rise to new families of locally testable codes that outperform Reed–Muller codes in a certain range of parameters [24]. Specifically, lifted codes lead to one of the two known constructions (the other one being tensor codes [10, 11, 32, 29]) of high-rate locally testable codes (i. e., locally testable codes with rate approaching one

[2] This strengthened notion of soundness has been referred to as *robust soundness* [7, 10] or just *robustness* [23] in earlier work. To make precise the contrast with the standard notion of soundness (1.1), we will refer to it as “robust soundness” in this paper.

and locality n^β for any constant $\beta > 0$). Generally, lifted codes form a natural subclass of affine-invariant codes having the 'single-orbit characterization' property that is known to imply local testability, as well as local decodability [28].

There is a natural local test associated with lifted codes: on oracle access to a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$, pick a uniform random ℓ-dimensional affine subspace $L \subseteq \mathbb{F}_q^m$ and accept if and only if $f|_L \in C$. It follows immediately by definition that this test accepts any valid codeword $f \in C^{\ell \nearrow m}$ with probability one, but more work is required to show that this test is sound. Specifically, since the test forms a single-orbit characterization, it follows from the work of Kaufman and Sudan [28] that it has soundness roughly $q^{-2\ell}$. The dependence of the soundness on the dimension ℓ was later eliminated by Haramaty, Ron-Zewi and Sudan [25] who showed soundness that is only a function of q (though an extremely quickly decaying one).

As for robust soundness, the above local testing results, together with the straightforward transformation from local testing to robust testing, immediately give robust soundness that is dependent on the dimension ℓ. This dependence was eliminated recently by Guo, Haramaty and Sudan [23] who showed robust soundness of the form $\mathrm{poly}(\delta)$ (about δ^{74}, where δ is the minimum distance of the code) for the local test that queries subspaces of slightly larger dimension of 2ℓ. Interestingly, Guo, Haramaty and Sudan [23] did not rely on the aforementioned local testing results, but rather relied on viewing lifted codes as the intersection of 'modified tensor codes.' They then proceeded by showing that these modified tensor codes are robustly testable (using Viderman's method [32] showing robust soundness of tensor codes), and that this implies local testability of the lifted code. (See Section 2.4 for more details about the method of Guo, Haramaty and Sudan [23].)

1.3 Our results

Our main result gives a transformation from local testing to robust testing, that does not suffer the factor of Q (the query complexity) loss in robust soundness, for the class of lifted codes. The transformation uses local testability in a 'black-box' manner, and shows that if a code in this family is locally testable (using the natural subspace tester) then it is also robustly testable with roughly the same number of queries and robust soundness.

For $k \geq \ell$, let the *k-dimensional test* denote the local tester that on oracle access to a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ queries a uniform random k-dimensional affine subspace $K \subseteq \mathbb{F}_q^m$ and accepts if and only if $f|_K \in C^{\ell \nearrow k}$.

Main Theorem 1.1. *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. Suppose that $C^{\ell \nearrow m}$ is locally testable using the k-dimensional test with query complexity q^k and soundness α, and let $\delta := \min_{k \leq r \leq m} \mathrm{mindist}(C^{\ell \nearrow r})$. Then $C^{\ell \nearrow m}$ is robustly testable using the $(2k+2+\log_q(4/\delta))$-dimensional test with query complexity $O(q^{2k+2}/\delta)$ and robust soundness $\Omega(\alpha \cdot \delta^3)$.*

Note that if the minimum distance δ is constant, we only incur a constant multiplicative loss in robust soundness and testing dimension.[3] We conjecture that the testing dimension for proving robust soundness may be as small as $k+1$, and leave it as an interesting question for future research.

[3] One natural example for codes with constant δ are the lifted Reed-Solomon codes of Guo, Kopparty and Sudan [24]. Furthermore, it was shown by them [24] that for any $m \geq \ell$, $\mathrm{mindist}(C^{\ell \nearrow m}) \geq \mathrm{mindist}(C) - q^{-\ell}$ (see Proposition 3.2 below).

To apply the above theorem one can instantiate it with the local testing result of Kaufman and Sudan [28] that says that lifted codes are locally testable using the ℓ-dimensional test with soundness $\approx q^{-2\ell}$ (see Theorem 3.4 below). However, to obtain constant robust soundness we need that the soundness of the initial local tester be constant (independent of q and ℓ), and for this we observe (in Proposition 6.1) that the soundness of Kaufman and Sudan [28] can be easily amplified to $\Omega(1)$ at the cost of increasing the testing dimension[4] to $\approx 3\ell$. Using this observation we obtain the following.

Main Corollary 1.2. *There exists an absolute constant $c > 1$ so that the following holds. Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code of minimum distance δ, and $m \geq \ell$. Then $C^{\ell \nearrow m}$ is robustly testable using the $(6\ell + 4 + \log_q(c/\delta))$-dimensional test with robust soundness $\Omega(\delta^3)$.*

Compared to the above corollary, the Guo, Haramaty and Sudan [23] use a lower dimension of 2ℓ, but also obtain a lower robust soundness of $\Omega(\delta^{74})$.

As described next, our proof is combinatorial, relying mainly on sampling properties of the collection of local views. In particular, it uses very little about the algebraic structure of lifted codes or the base code. We thus hope that such techniques may prove useful in the future for showing robust soundness for other families of locally testable codes with similar sampling properties.

2 Overview of the proof

Our proof is based on a new connection between the notions of local testing, robust testing, and *agreement testing*. Specifically, we show that for the class of lifted codes, agreement testing implies robust testing, and local testing implies agreement testing. The combination of these two implications gives our Main Theorem 1.1. Next we elaborate on the notion of agreement testing, followed by an overview of each of the implications.

2.1 Agreement testing

An *agreement test* attempts to find out whether assignments to local views can be stitched together to a single global codeword. Let $C \subseteq \{U \to \Sigma\}$ be a code, and let $\mathcal{S}$ be a collection of subsets of U. An *agreement tester* for $(C, \mathcal{S})$ is a probabilistic oracle algorithm that receives oracle access to a collection of partial assignments $\{f_S : S \to \Sigma \mid S \in \mathcal{S}\}$, where $f_S \in C|_S$ for any $S \in \mathcal{S}$. The tester queries a few of the f_S and is required to accept with probability one any collection $(f_S)_S$ that is consistent with some global codeword $g \in C$ (that is, $g|_S = f_S$ for any $S \in \mathcal{S}$), while rejecting any inconsistent collection $(f_S)_S$ with probability at least proportional to the minimal fraction of the f_S that must be changed in order to be consistent with some global codeword. In this paper, we focus on an agreement tester that picks a pair of sets $S, S' \in \mathcal{S}$ according to some distribution and accepts if and only if f_S and $f_{S'}$ agree on their intersection $S \cap S'$.

Agreement testing has first appeared in PCP constructions [6, 5, 20, 3, 2] as so-called "low-degree tests," and is a key component in the proof of almost all PCP theorems. A prime example is the line

[4] Such an amplification with a similar blow-up in query complexity can be easily obtained by repeating the test and accepting if and only if all invocations accept; we, however, need that the tester be a subspace tester which can be obtained using sampling properties of affine subspaces.

vs. line low-degree test [21, 31] in the proof of the PCP theorem. In the PCP construction, a function on a large vector space is replaced by an ensemble of (supposed) restrictions to all possible affine lines. These restrictions are supplied by a prover and are not a priori guaranteed to agree with any single global function. The "low-degree test" is run by the verifier to check that restrictions on intersecting lines agree with each other, i. e., they give the same value to the point of intersection. The main point of the argument is to show that the passing of the test implies agreement with a single global function. In these early low-degree tests (including the linearity tester of Blum, Luby and Rubinfeld [12]) an agreement test component can be discerned but quite implicitly. Indeed, it was only separated in the work of Raz, Safra, Arora, and Sudan [30, 4] that looked at the so-called list-decoding regime,[5] with the goal of proving a large gap for the PCP.

Goldreich and Safra [22] tried to separate the algebraic aspect of the low-degree test from the combinatorial, and formulated a more general "consistency test" which is also referred to as an agreement test. They also proved a certain local to global result which was too weak to be useful for PCPs. In hindsight it is clear that since their family of subsets consisted of axis-parallel lines, the expansion was not strong enough for a good agreement test. Only recently [16, 13] has the role of expansion underlying the family of subsets begun to be uncovered.

Work on agreement testing then continued the combinatorial direction of Goldreich and Safra [22] mainly in the list-decoding regime for direct product testing [18, 14, 27, 19, 17]. The techniques developed in this line of work turn out to be useful also for agreement testing in the unique-decoding regime (which is the more standard testing regime), and in particular for our work here.

Our proof of agreement testing based on local testing (see Section 5) is significantly simpler than general agreement testing theorems. The reason is that we avoid a major technical difficulty that stems from the varying size of disagreement between pairs of local views. In the general setting this can be anywhere from disagreement on only one point in the intersection to disagreement on the entire intersection, leading to a very subtle argument. In contrast, in our setting the local view restricted to the intersection is itself an error correcting code, so whenever local views disagree, they must disagree on a constant fraction of the intersection. This makes the proof of the agreement theorem quite direct, and much easier than the proofs in the aforementioned papers.

An agreement testing theorem that holds in quite broad generality was proven by Dikstein and Dinur [13] (published after the conference version of the present paper [15]). This result (see Theorems 6.2 and 6.3 in [13]) is similar to what we prove here but the parameters of Dikstein and Dinur are slightly weaker.

2.2 Agreement testing implies robust testing

We begin with an overview of the simpler implication from agreement testing to robust testing. In what follows, we say that a collection $\mathcal{S}$ of subsets of the ground set U *samples well inside U*, if, informally, for all $0 \leq \gamma \leq 1$ and for any subset $A \subseteq U$ of density γ (i. e., $|A|/|U| = \gamma$), it holds that for most subsets $S \in \mathcal{S}$, the density of A inside S is approximately γ (i. e., $|A \cap S|/|S| \approx \gamma$).

Suppose that we have an agreement tester for $(C, \mathcal{S})$ as described above. We would like to show that the local tester that queries a random $S \in \mathcal{S}$ is robustly sound. Let $\mathcal{T}$ be the collection of subsets

[5] In the list decoding regime one would like to reject a function that is $(1-\varepsilon)$-far from the code with very high probability of $1 - O(\varepsilon)$.

of U formed by the pairwise intersections of distinct sets in $\mathcal{S}$. The main properties of $\mathcal{S},\mathcal{T}$ we need are sampling properties, specifically, that $\mathcal{S}$ samples well inside U and that for any $S \in \mathcal{S}$, the sets in $\mathcal{T}$ contained in S sample well inside S. The main property of the code C we need is that for any set $T \in \mathcal{T}$, the code $C|_T := \{c|_T \mid c \in C\}$ has large distance. In the case of lifted codes these properties can be guaranteed by letting $\mathcal{S}$ and $\mathcal{T}$ be families of affine subspaces of fixed dimensions s and t, respectively, where[6] $s \gg t$.

To see that the proposed local tester is indeed robustly sound, suppose we have a function $f : U \to \Sigma$ that is close to $C|_S$ on a typical $S \in \mathcal{S}$. Our goal is to show that f is close to a codeword $g \in C$. We first create an instance $(f_S)_S$ for the agreement tester by letting $f_S \in C|_S$ be the closest valid assignment to $f|_S$. Next observe that since $f|_S$ is typically close to f_S, and by assumption the sets T sample well inside the sets S, for a typical T and S,S' containing T, it holds that both $f_S|_T$ and $f_{S'}|_T$ are close to $f|_T$, and by the property that $C|_T$ has large distance for any $T \in \mathcal{T}$, this implies in turn that typically $f_S|_T = f_{S'}|_T$. Consequently, agreement testability implies the existence of a codeword $g \in C$ that agrees with most f_S, and so $g|_S = f_S$ and f_S is close to $f|_S$ for most S. But since $\mathcal{S}$ samples well inside U we conclude that f must be close to g as required.

2.3 Local testing implies agreement testing

We now turn to the local testing to agreement testing implication which is a bit more involved. Suppose that we have a local testing algorithm for C that queries a random set $K \in \mathcal{K}$ and accepts if and only if $f|_K \in C|_K$. We would like to obtain an agreement tester for C with respect to some collection of subsets $\mathcal{S}$. As before, let $\mathcal{T}$ be the collection of subsets of U formed by the pairwise intersections of sets in $\mathcal{S}$. Once more the main properties of $\mathcal{S},\mathcal{T},\mathcal{K}$ we require are sampling properties. Specifically, we need that $\mathcal{S}$ samples well inside U, and that for any $T \in \mathcal{T}$, all sets in $\mathcal{K}$ contained in T sample well inside T. We also require distance properties of C, specifically that $C, C|_T, C|_S$ all have large distance for any $S \in \mathcal{S}$ and $T \in \mathcal{T}$. Once more, in the case of lifted codes these properties can be guaranteed by letting $\mathcal{S},\mathcal{T},\mathcal{K}$ be families of affine subspaces of fixed dimensions s,t,k, respectively, where $s \gg t \gg k$.

To show agreement testability, let $(f_S)_S$ be a collection of valid assignments to sets in $\mathcal{S}$ (so $f_S \in C|_S$ for any S), and suppose that f_S agrees with $f_{S'}$ on $S \cap S'$ for most pairs S,S'. Our goal will be to find a global codeword $g \in C$ that agrees with most f_S. We find the function g in the following three stages.

Initial stage. In the first stage we define for any $K \in \mathcal{K}$ a 'most popular function' Plur_K by choosing the most common value among $f_S|_K$ going over all $S \in \mathcal{S}$ containing K. We then show, using the assumption that the f_S typically agree on their pairwise intersections, that this most popular function is obtained with overwhelming probability for a typical K.

Local structure stage. In the second stage we define for each $K \in \mathcal{K}$ a function $g_K : U \to \Sigma$ by letting $g_K(x)$ be the most common 'vote' among all f_S where $S \supseteq K$ and $x \in S$ and f_S agrees with Plur_K on K (this function is well defined because of the initial stage). We then show that for a typical K, g_K is close to some function $h_K \in C$, and moreover $h_K|_S = f_S$ for most S containing K.

[6] In this paper, we use the informal notation $s \gg t$ for positive numbers s,t to indicate s is *much greater* than t.

To see why the above holds, first note that by the assumptions that $C|_T$ has large distance for any $T \in \mathcal{T}$, and the sets K sample well inside each T if a pair of partial functions f_S and $f_{S'}$ agree on K then they must typically also agree on their whole intersection. Therefore $g_K(x)$ is also typically defined with overwhelming probability. Consequently, for a typical K, and most K', $g_K|_{K'}$ agrees with some f_S. Recalling that the f_S are valid assignments, local testability then implies the existence of $h_K \in C$ that is close to g_K. The fact that $h_K|_S = f_S$ for most S containing K follows by the assumption that $\mathcal{S}$ samples well inside U, and the distance property on $\mathcal{S}$.

Global structure stage. In the final stage we show that there exists $\hat{K} \in \mathcal{K}$ such that $h_{\hat{K}}$ agrees with f_S for most S (not necessarily containing $\hat{K}$). We can then set our 'global function' g to be equal to $h_{\hat{K}}$. To this end, we first observe that it suffices to show that most functions h_K are in fact identical. This now follows since for typical $S \supseteq K \cup K'$ it holds that $h_K|_S = f_S = h_{K'}|_S$, and consequently since $\mathcal{S}$ samples well inside U, it must typically hold that $h_K = h_{K'}$.

2.4 The proof method of Guo, Haramaty and Sudan

Guo, Haramaty and Sudan's method [23] to show robust soundness of lifted codes is very different from ours. In particular, it relies heavily on the algebraic structure of lifted codes. More specifically, the proof is based on viewing the lifted codes as the intersection of 'modified tensor codes.' The *tensor power* $C^{\otimes m}$ of a code $C \subseteq \{\mathbb{F}_q \to \mathbb{F}_q\}$ can be thought of as the 'axis-parallel lifting' of C. That is, it is the code that consists of all functions $f : \mathbb{F}_q^m \to \mathbb{F}_q$ whose restrictions to any axis-parallel line belong to C. The 'modified tensor code' is a code of the form $C_b^{\otimes m}$ where b is a direction in $\mathbb{F}_q^m$, and $C_b^{\otimes m}$ consists of all functions $f \in C^{\otimes m}$ whose restrictions to lines in direction b also belong to C.

The authors first use Viderman's method [32], showing robust testing of tensor codes, to show that the modified tensor codes are also robustly testable. They then use the fact that the lifted code is the intersection of all codes of the form $C_b^{\otimes m}$ for all directions b (this is true when the dimension of the base code for lifting is $\ell = 1$; when $\ell > 1$ the proof becomes more complicated) to deduce robust testability for the lifted code. However, since the intersection of robustly testable codes is not necessarily robustly testable, non-trivial work is required to show robust testability, which in particular exploits the degree structure of affine-invariant lifted codes.

The above program can be carried out only when the dimension m of the lifted code is a small constant multiple of ℓ, and the authors use the bootstrapping technique of [31, 2, 4, 1] to extend the result to arbitrarily large m.

In contrast, we work directly with lifted codes of large dimension which allows us to exploit the sampling properties of large affine subspaces in $\mathbb{F}_q^m$. To the best of our knowledge, even for the special case of low-degree polynomials, this gives the first analysis of robust soundness that is not based on the two-step approach of first analyzing the constant-dimensional case and only then moving to the general case.

In contrast to the proof of Guo, Haramaty and Sudan [23] who reprove local testability along the way, our proof uses local testability in a black-box manner. Thus, it exhibits a separation between the algebraic properties that are used for showing local testability, and the combinatorial properties that are needed in order to turn local testability into robust testability.

Organization of the paper. The rest of the paper is organized as follows. In Section 3 we set some notation, provide some definitions, and present the sampling properties of subspaces that we use. The transformation from agreement testing to robust testing is given in Section 4, while the transformation from local testing to agreement testing appears in Section 5. We conclude in Section 6 with the full transformation from local testing to robust testing that proves our Main Theorem 1.1 and Main Corollary 1.2.

3 Preliminaries

For a prime power q, let $\mathbb{F}_q$ denote the finite field of q elements. Let $\{\mathbb{F}_q^m \to \mathbb{F}_q\}$ denote the set of functions mapping $\mathbb{F}_q^m$ to $\mathbb{F}_q$. In what follows we focus on codes which are sets of functions, $C \subseteq \{\mathbb{F}_q^m \to \mathbb{F}_q\}$. For a pair of functions $f, g : \mathbb{F}_q^m \to \mathbb{F}_q$, we use $\operatorname{dist}(f,g)$ to denote the fraction of inputs $x \in \mathbb{F}_q^m$ for which $f(x) \neq g(x)$. The minimum distance $\operatorname{mindist}(C)$ of the code C is $\min_{f \neq g \in C}\{\operatorname{dist}(f,g)\}$. For a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$, we use $\operatorname{dist}(f,C)$ to denote $\min_{g \in C}\{\operatorname{dist}(f,g)\}$.

The code C is said to be *linear* if it is an $\mathbb{F}_q$-linear subspace, i. e., $C \neq \emptyset$ and for every $\alpha \in \mathbb{F}_q$ and $f, g \in C$, we have $\alpha f + g \in C$. A function $A : \mathbb{F}_q^m \to \mathbb{F}_q^m$ is said to be an *affine transformation* if there exist a matrix $M \in \mathbb{F}_q^{m \times m}$ and a vector $b \in \mathbb{F}_q^m$ such that $A(x) = Mx + b$. The code C is said to be *affine invariant* if for every affine transformation A and every $f \in C$ we have $f \circ A \in C$ (where $(f \circ A)(x) = f(A(x))$).

3.1 Lifted codes

A subset $L \subseteq \mathbb{F}_q^m$ is said to be an *ℓ-dimensional affine subspace* if there exist $u_0 \in \mathbb{F}_q^m$ and linearly independent $u_1, \ldots, u_\ell \in \mathbb{F}_q^m$ such that $L = \{u_0 + \sum_{i=1}^{\ell} u_i x_i \mid x_1, \ldots, x_\ell \in \mathbb{F}_q\}$. We fix an arbitrary invertible affine map $\gamma_L : \mathbb{F}_q^\ell \to L$ (which we can view as a parameterization of L). For a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$, the restriction $f|_L$ is viewed as a function in $\{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ through $f \circ \gamma_L : \mathbb{F}_q^\ell \to \mathbb{F}_q$. In particular, when we ask if $f|_L \overset{?}{\in} C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ what we are really asking is whether $f \circ \gamma_L \in C$. Note that if C is affine-invariant, whether $f|_L \in C$ does not depend on the choice of the parametrization γ_L.

Definition 3.1 (Lifted codes). Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq \ell$. The *m-dimensional lift $C^{\ell \nearrow m}$* of C is given by

$$C^{\ell \nearrow m} := \left\{ f : \mathbb{F}_q^m \to \mathbb{F}_q \mid f|_L \in C \text{ for every } \ell\text{-dimensional affine subspace } L \subseteq \mathbb{F}_q^m \right\}.$$

Proposition 3.2 (Distance of lifted codes, [24, Theorem 5.1, Part (2)]). *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq \ell$. Then* $\operatorname{mindist}(C^{\ell \nearrow m}) \geq \operatorname{mindist}(C) - q^{-\ell}$.

3.2 Local testing, robust testing, and agreement testing

We now formally define the notions of local testing, robust testing, and agreement testing, specialized to the class of lifted codes and subspace testers. In the case of local testing and robust testing this simply means that the tester samples a uniform random k-dimensional affine subspace and its accepting views are codewords in $C^{\ell \nearrow k}$.

Definition 3.3 (Local testing of lifted codes)**.** Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. The m-dimensional lift $C^{\ell \nearrow m}$ is (k,α)-*testable* if for every $f : \mathbb{F}_q^m \to \mathbb{F}_q$ it holds that

$$\Pr_K \left[f|_K \notin C^{\ell \nearrow k} \right] \geq \alpha \cdot \mathrm{dist}(f, C^{\ell \nearrow m}),$$

where the probability is over a uniform random k-dimensional affine subspace $K \subseteq \mathbb{F}_q^m$.

Recall the definition of the ℓ-*dimensional test* in the paragraph before Main Theorem 1.1.

Theorem 3.4 ([28, Theorem 2.9])**.** *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq \ell$. Then the ℓ-dimensional test rejects a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ with probability at least $\frac{1}{2} \cdot \min\left\{q^{-2\ell}, \mathrm{dist}(f, C^{\ell \nearrow m})\right\}$. In particular, $C^{\ell \nearrow m}$ is $\left(\ell, \frac{q^{-2\ell}}{2}\right)$-testable.*

Definition 3.5 (Robust testing of lifted codes)**.** Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. The m-dimensional lift $C^{\ell \nearrow m}$ is (k,α)-*robust* if for every $f : \mathbb{F}_q^m \to \mathbb{F}_q$ it holds that

$$\mathbb{E}_K \left[\mathrm{dist}(f|_K, C^{\ell \nearrow k}) \right] \geq \alpha \cdot \mathrm{dist}(f, C^{\ell \nearrow m}),$$

where the expectation is over a uniform random k-dimensional affine subspace $K \subseteq \mathbb{F}_q^m$.

We note the following easy implications.

Proposition 3.6. *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. Then the following hold.*

1. *If $C^{\ell \nearrow m}$ is (k,α)-robust then it is (k,α)-testable.*
2. *If $C^{\ell \nearrow m}$ is (k,α)-testable then it is $(k,\alpha \cdot q^{-k})$-robust.*
3. *If $C^{\ell \nearrow m}$ is (k,α)-testable then it is (r,α)-testable for any $k \leq r \leq m$.*
4. *If $C^{\ell \nearrow m}$ is (k,α)-robust then it is (r,α)-robust for any $k \leq r \leq m$.*

Proof. Part (1) follows since $f|_K \notin C^{\ell \nearrow k}$ whenever $\mathrm{dist}(f|_K, C^{\ell \nearrow k}) > 0$, while Part (2) follows since $\mathrm{dist}(f|_K, C^{\ell \nearrow k}) \geq q^{-k}$ whenever $f|_K \notin C^{\ell \nearrow k}$.

Part (3) follows by observing that for a uniform random r-dimensional affine subspace R,

$$\begin{aligned}
\Pr_R \left[f|_R \notin C^{\ell \nearrow r} \right] &= \mathbb{E}_R \left[\mathbb{1}_{f|_R \notin C^{\ell \nearrow r}} \right] \\
&\geq \mathbb{E}_R \left[\Pr_{K \subseteq R} \left[f|_K \notin C^{\ell \nearrow k} \right] \right] \\
&= \Pr_K \left[f|_K \notin C^{\ell \nearrow k} \right],
\end{aligned}$$

where the inequality follows since $f|_R \in C^{\ell \nearrow r}$ implies that $f|_K \in C^{\ell \nearrow k}$ for any K.

Finally, Part (4) follows by letting f_R be the codeword in $C^{\ell\nearrow r}$ that is closest to $f|_R$, and noting that

$$\begin{aligned}\mathbb{E}_R\left[\operatorname{dist}(f|_R, C^{\ell\nearrow r})\right] &= \mathbb{E}_R\left[\operatorname{dist}(f|_R, f_R)\right] \\ &= \mathbb{E}_R\left[\mathbb{E}_{K\subseteq R}\left[\operatorname{dist}(f|_K, f_R|_K)\right]\right] \\ &\geq \mathbb{E}_R\left[\mathbb{E}_{K\subseteq R}\left[\operatorname{dist}(f|_K, C^{\ell\nearrow k})\right]\right] \\ &= \mathbb{E}_K\left[\operatorname{dist}(f|_K, C^{\ell\nearrow k})\right],\end{aligned}$$

where the inequality follows since $f_R|_K \in C^{\ell\nearrow k}$ for any K. □

We now turn to the definition of agreement testing. The agreement testers we consider are testers that for $t < s$, sample a uniform random t-dimensional affine subspace T, and a pair of uniform random s-dimensional affine subspaces S, S' containing T, and accept if and only if $f_S, f_{S'}$ agree on T.

For a code $C \subseteq \{\mathbb{F}_q^m \to \mathbb{F}_q\}$, we let $C(s)$ be the code containing all collections $(f_S)_S$ of partial assignments to s-dimensional affine subspaces that are consistent with some global codeword $g \in C$, formally,

$$C(s) := \left\{(f_S)_S \mid \exists\, g \in C \text{ such that } g|_S = f_S \text{ for any } s\text{-dimensional affine subspace } S\right\}.$$

For a pair of collections $(f_S)_S, (g_S)_S$ of partial assignments to s-dimensional affine subspaces, we denote by $\operatorname{dist}((f_S)_S, (g_S)_S)$ the fraction of s-dimensional affine subspaces S for which $f_S \neq g_S$, and we define $\operatorname{dist}((f_S)_S, C(s))$ accordingly.

Definition 3.7 (Agreement testing of lifted codes)**.** Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq s > t \geq \ell$. The m-dimensional lift $C^{\ell\nearrow m}$ is (s,t,α)-*agreement testable* if for every collection $(f_S)_S$ where $f_S \in C^{\ell\nearrow s}$ for every s-dimensional affine subspace S it holds that

$$\Pr_{T,\, S\supseteq T,\, S'\supseteq T}\left[f_S|_T \neq f_{S'}|_T\right] \geq \alpha \cdot \operatorname{dist}\left((f_S)_S, C^{\ell\nearrow m}(s)\right),$$

where the probability is over a uniform random t-dimensional affine subspace $T \subseteq \mathbb{F}_q^m$, and uniform random s-dimensional affine subspaces S, S' containing T.

3.3 Sampling properties of affine subspaces

Let $\mathcal{T}$ be a collection of affine subspaces $T \subseteq \mathbb{F}_q^m$ of dimension t, and let $\mathcal{S}$ be a collection of affine subspaces $S \subseteq \mathbb{F}_q^m$ of dimension s, where $t < s$. Let $\mathcal{I}(\mathcal{T}, \mathcal{S})$ denote the bipartite *inclusion graph* whose left side is $\mathcal{T}$ and right side is $\mathcal{S}$ and $T \in \mathcal{T}$ and $S \in \mathcal{S}$ are adjacent if and only if $T \subseteq S$. For $\alpha, \beta \in (0,1)$, we say that $\mathcal{I}(\mathcal{T}, \mathcal{S})$ is an (α,β)-*hitting sampler* if for every subset $A \subseteq \mathcal{T}$ with $|A| \geq \alpha|\mathcal{T}|$, it holds that $|N(A)| \geq (1-\beta)|\mathcal{S}|$, where $N(A)$ denotes the set of neighbors of A in $\mathcal{S}$.

First we note that in the case where $\mathcal{T}$ is the collection of all points in $\mathbb{F}_q^m$ (i. e., 0-dimensional affine subspaces), and $\mathcal{S}$ is the collection of all s-dimensional affine subspaces, the hitting sampler property is an immediate consequence of the fact that points in a random affine subspace are uniformly distributed and pairwise independent, and Chebyshev's inequality.

Lemma 3.8. *Let $0 < s < m$, let $\mathcal{T}$ be the collection of all points in $\mathbb{F}_q^m$, and let $\mathcal{S}$ be the collection of all s-dimensional affine subspaces of $\mathbb{F}_q^m$. Then for any $\alpha > 0$, the inclusion graph $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an $(\alpha, \frac{q^{-s}}{\alpha})$-hitting sampler.*

The following lemma from [26] is an extension for the case where for a fixed affine subspace $R \subseteq \mathbb{F}_q^m$ of dimension $r < s$, $\mathcal{T}$ is the collection of all points in $\mathbb{F}_q^m \setminus R$, and $\mathcal{S}$ is the collection of all s-dimensional affine subspaces containing R.

Lemma 3.9 ([26, Lemma 2.12])**.** *Let $0 < r < s < m$, and let $R \subseteq \mathbb{F}_q^m$ be an affine subspace of dimension r. Let $\mathcal{T}$ be the collection of all points in $\mathbb{F}_q^m \setminus R$, and let $\mathcal{S}$ be the collection of all s-dimensional affine subspaces in $\mathbb{F}_q^m$ containing R. Then for any $\alpha > 0$, the inclusion graph $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an $(\alpha, \frac{4q^{-(s-r-2)}}{\alpha})$-hitting sampler.*

Remark 3.10. [26, Lemma 2.11] is only stated for the case where $s = \frac{m}{2}$, while [26, Lemma 2.12] is only stated for the case where $r = \frac{s}{2}$. However, it can easily be seen that the proof implies the bounds we state for general values of $0 < r < s < m$.

Finally, we cite the following lemma, due to Impagliazzo, Kabanets and Wigderson [27], which deals with the case where $\mathcal{T}$ is the collection of all t-dimensional *linear* subspaces, and $\mathcal{S}$ is the collection of all s-dimensional *linear* subspaces.

Lemma 3.11 ([27, Lemma 2.3])**.** *Let $0 < t < s < m$, let $\mathcal{T}$ be the collection of all t-dimensional linear subspaces in $\mathbb{F}_q^m$, and let $\mathcal{S}$ be the collection of all s-dimensional linear subspaces in $\mathbb{F}_q^m$. Then for any $\alpha > 0$, the inclusion graph $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an $(\alpha, \frac{18q^{-(s-t)}}{\alpha})$-hitting sampler.*

Remark 3.12. [27, Lemma 2.3] gives a worse bound for the case that $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an *averaging sampler*.[7] However, the proof first gives the bound we state, and only then shows how to deduce the worse bound for the case of an averaging sampler.

We further observe that Lemma 3.11 gives a similar bound for *affine* subspaces.

Corollary 3.13. *Let $0 < t < s$, let $\mathcal{T}$ be the collection of all t-dimensional affine subspaces in $\mathbb{F}_q^m$, and let $\mathcal{S}$ be the collection of all s-dimensional affine subspaces in $\mathbb{F}_q^m$. Then for any $\alpha > 0$, the inclusion graph $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an $(\alpha, \frac{18q^{-(s-t-1)}}{\alpha})$-hitting sampler.*

Proof. Let $A \subseteq \mathcal{T}$ be a collection of t-dimensional affine subspaces in $\mathbb{F}_q^m$ so that $|A| \geq \alpha|\mathcal{T}|$. Let $S \in \mathcal{S}$ be a uniformly random s-dimensional affine subspace. Then S can be sampled by first picking a random affine shift $v \in \mathbb{F}_q^m$, then picking a random s-dimensional linear subspace $V \subseteq \mathbb{F}_q^m$ so that $v \notin V \setminus \{0\}$, and finally letting $S = v + V$. As for a fixed vector $v \in \mathbb{F}_q^m$, and a random s-dimensional linear subspace $V \subseteq \mathbb{F}_q^m$, we have that $v \in V$ with probability at most $q^{-(m-s)}$, which is negligible in our setting, we may assume that V is a completely random s-dimensional linear subspace. Next observe that $v + V$ contains a t-dimensional affine subspace $T = w + W \in A$ if V contains $\mathrm{Span}\{w - v, W\}$. The proof is completed by noting that, conditioned on choosing v, by Lemma 3.11, V contains $\mathrm{Span}\{w - v, W\}$ for some $T = w + W \in A$ with probability at least $1 - \left(18 \cdot q^{-(s-t-1)}\right)/\alpha$. □

[7] For $\alpha, \beta \in (0,1)$, we say that $\mathcal{I}(\mathcal{T},\mathcal{S})$ is an (α,β)-*averaging sampler* if for every subset $A \subseteq \mathcal{T}$ with $|A| \geq \alpha|\mathcal{T}|$, for at least a $(1-\beta)$-fraction of the vertices $S \in \mathcal{S}$ it holds that $\left|\frac{|A \cap N(S)|}{\deg(S)} - \alpha\right| \leq \frac{\alpha}{2}$. Note that an (α,β)-averaging sampler is in particular an (α,β)-hitting sampler.

4 From agreement testing to robust testing

In this section we prove the following lemma, showing the agreement testing to robust testing implication.

Lemma 4.1 (Agreement testing implies robust testing)**.** *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq s > t \geq \ell$. Suppose that $C^{\ell\nearrow m}$ is (s,t,α)-agreement testable, and let $\delta := \mathrm{mindist}(C^{\ell\nearrow t})$. Then $C^{\ell\nearrow m}$ is $\big(s, \Omega(\alpha\delta)\big)$-robust.*

Proof. For simplicity of notation, in what follows we let T,S denote the random variables obtained by sampling a uniform random affine subspace of dimension t,s respectively. Suppose that $f : \mathbb{F}_q^m \to \mathbb{F}_q$ has $\mathbb{E}_S[\mathrm{dist}(f|_S, C^{\ell\nearrow s})] \leq \varepsilon$. Our goal is to find a codeword $g \in C^{\ell\nearrow m}$ such that $\mathrm{dist}(f,g) \leq O(\varepsilon/(\alpha\delta))$.

The proof proceeds as follows. We would like to apply our assumption on agreement testability, and towards this, we create an instance $(f_S)_S$ for the agreement tester by letting f_S be the codeword in $C^{\ell\nearrow s}$ that is closest to $f|_S$. We then use the fact that f_S is typically close to $f|_S$, together with the fact that t-dimensional affine subspaces sample well inside s-dimensional affine subspaces, and the assumption that C has large distance on t-dimensional affine subspaces of the domain $\mathbb{F}_q^s$, to show that $\Pr_{T,\,S\supseteq T,\,S'\supseteq T}[f_S|_T \neq f_{S'}|_T]$ is small. Agreement testability then gives a codeword $g \in C^{\ell\nearrow m}$ that is consistent with most f_S, and using the fact that s-dimensional affine subspaces sample well inside $\mathbb{F}_q^m$ this implies in turn that $\mathrm{dist}(f,g)$ is small. Details follow.

We begin by showing that $\Pr_{T,\,S\supseteq T,\,S'\supseteq T}[f_S|_T \neq f_{S'}|_T]$ is small. Recall first that $\mathbb{E}_S[\mathrm{dist}(f|_S, f_S)] = \mathbb{E}_S[\mathrm{dist}(f|_S, C^{\ell\nearrow s})] \leq \varepsilon$. Next observe that for a fixed s-dimensional affine subspace S, any point in a uniform random t-dimensional affine subspace contained in S is uniform in S. Thus we also have that $\mathbb{E}_{S,\,T\subseteq S}[\mathrm{dist}(f|_T, f_S|_T)] \leq \varepsilon$, and consequently

$$\begin{aligned}\Pr_{T,\,S\supseteq T,\,S'\supseteq T}\big[\mathrm{dist}(f_S|_T, f_{S'}|_T) \geq \delta\big] &\leq 2\cdot \Pr_{T,\,S\supseteq T}\big[\mathrm{dist}(f|_T, f_S|_T) \geq \delta/2\big]\\ &= 2\cdot \Pr_{S,\,T\subseteq S}\big[\mathrm{dist}(f|_T, f_S|_T) \geq \delta/2\big]\\ &\leq \frac{4\varepsilon}{\delta}.\end{aligned}$$

But since $f_S|_T$, $f_{S'}|_T$ are both codewords of $C^{\ell\nearrow t}$, a code of minimum distance δ, the above implies in turn that $\Pr_{T,\,S\supseteq T,\,S'\supseteq T}[f_S|_T \neq f_{S'}|_T] \leq \frac{4\varepsilon}{\delta}$.

Our assumption on agreement testability now gives a codeword $g \in C^{\ell\nearrow m}$ that has $\Pr_S[g|_S \neq f_S] \leq 4\varepsilon/(\alpha\delta)$. But since any point in a uniform random s-dimensional affine subspace is uniform in $\mathbb{F}_q^m$ this gives in turn that

$$\mathrm{dist}(f,g) = \mathbb{E}_S\big[\mathrm{dist}(f|_S, g|_S)\big] \leq \mathbb{E}_S\big[\mathrm{dist}(f|_S, f_S)\big] + \mathbb{E}_S\big[\mathrm{dist}(f_S, g|_S)\big] \leq \varepsilon + \frac{4\varepsilon}{\alpha\delta} \leq \frac{5\varepsilon}{\alpha\delta}. \qquad \square$$

5 From local testing to agreement testing

In this section we prove the following lemma that gives the local testing to agreement testing implication.

Lemma 5.1 (Local testing implies agreement testing)**.** *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. Suppose that $C^{\ell\nearrow m}$ is (k,α)-testable, and let $\delta := \min_{k\leq r\leq m} \mathrm{mindist}(C^{\ell\nearrow r})$. Then $C^{\ell\nearrow m}$ is $(2k+2+\log_q(4/\delta), k+1, \Omega(\alpha\cdot\delta^2))$-agreement testable.*

Proof outline. For simplicity of notation, in what follows let $s := 2k+2+\log_q(4/\delta)$ and $t := k+1$. We let both S, S' (T, T' and K, K', resp.) denote random variables obtained by sampling a uniform random affine subspace of dimension s (t, k, resp.).

Let $(f_S)_S$ be a collection of partial assignments such that $f_S \in C^{\ell\nearrow s}$ for every S, and

$$\Pr_{T,\, S\supseteq T,\, S'\supseteq T}[f_S|_T \neq f_{S'}|_T] \leq \varepsilon\,. \tag{5.1}$$

Our goal is to find a global codeword $g \in C^{\ell\nearrow m}$ that has

$$\Pr_S[g|_S \neq f_S] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)\,. \tag{5.2}$$

We find the codeword g in three stages.

1. In the initial stage (Section 5.1) we define, for any k-dimensional affine subspace K, a 'most popular function' $\mathrm{Plur}_K : \mathbb{F}_q^k \to \mathbb{F}_q$ by choosing the most common value among $f_S|_K$ going over all $S \supseteq K$. We show that for a typical K, this function is obtained with an overwhelming plurality of $1-O(\varepsilon)$.

2. In the "local structure" stage (Section 5.2) we define, for any k-dimensional affine subspace K, a function $g_K : \mathbb{F}_q^m \to \mathbb{F}_q$ by letting $g_K(x)$ be the most common 'vote' among all f_S that contain both K and x and agree with Plur_K on K. We then show that for a typical K, g_K is close to some codeword $h_K \in C^{\ell\nearrow m}$, and moreover $h_K|_S = f_S$ for most S containing K.

3. In the "global structure" stage (Section 5.3) we show that there exists $\hat{K}$ for which $h_{\hat{K}}|_S = f_S$ for most S (not necessarily containing $\hat{K}$). We can then set our 'global function' g to be equal to $h_{\hat{K}}$.

5.1 Initial stage

For any k-dimensional affine subspace K, we let $\mathrm{Plur}_K : \mathbb{F}_q^k \to \mathbb{F}_q$ denote the most common value among $f_S|_K$ for S containing K. That is,

$$\mathrm{Plur}_K := \mathrm{plurality}_{S\supseteq K}\{f_S|_K\}\,.$$

Next we use our assumption (5.1) to show that for a typical K, the function Plur_K is obtained with overwhelming plurality.

Lemma 5.2.

$$\mathbb{E}_K\left[\Pr_{S\supseteq K}\left[f_S|_K \neq \mathrm{Plur}_K\right]\right] \leq 2\varepsilon\,.$$

Proof. Since the collision probability lower bounds the probability of hitting the most common value, it suffices to show that

$$\Pr_{K,\, S\supseteq K,\, S'\supseteq K}\left[f_S|_K \neq f_{S'}|_K\right] \leq 2\varepsilon\,. \tag{5.3}$$

Clearly if $t=k$ we would be done by (5.1), so the whole point is to show the same for $t>k$. We describe a distribution on triples (S_1, S', S_2) such that (S_1, S_2) are distributed as in (5.3) but the pairs (S_1, S') and (S', S_2) are distributed as in (5.1):

1. Choose a uniform random k-dimensional affine subspace K.

2. Choose a pair of uniform random t-dimensional affine subspaces T_1, T_2 containing K.

3. For $i = 1,2$, choose a uniform random s-dimensional affine subspace S_i containing T_i.

4. Choose a uniform random s-dimensional affine subspace S' containing $T_1 \cup T_2$ (this can be done since $t = k+1$ and $s \geq k+2$).

One can check that indeed K, S_1, S_2 are distributed as in (5.3) while T_i, S_i, S' are distributed as in (5.1). Thus by our assumption (5.1),

$$\Pr_{K,\, S_1 \supseteq K,\, S_2 \supseteq K}\left[f_{S_1}|_K \neq f_{S_2}|_K\right] \leq \Pr_{T_1,\, S_1 \supseteq T_1,\, S' \supseteq T_1}\left[f_{S_1}|_{T_1} \neq f_{S'}|_{T_1}\right] + \Pr_{T_2,\, S' \supseteq T_2,\, S_2 \supseteq T_2}\left[f_{S'}|_{T_2} \neq f_{S_2}|_{T_2}\right] \leq 2\varepsilon\,. \qquad \square$$

5.2 Local structure

Next we define for every k-dimensional affine subspace K the function $g_K : \mathbb{F}_q^m \to \mathbb{F}_q$. As described above, for every $x \in \mathbb{F}_q^m$, we let $g_K(x)$ be the most common value among $f_S(x)$ for S that contain both K and x and agree with Plur_K on K. That is,

$$g_K(x) := \mathrm{plurality}_{S \supseteq K \cup \{x\},\, f_S|_K = \mathrm{Plur}_K}\{f_S(x)\}\,.$$

Next we would like to show that for a typical K, g_K is close to some codeword $h_K \in C^{\ell \nearrow m}$, and additionally $h_K|_S = f_S$ for most S containing K. We show these in three steps:

1. Boosting step (Lemma 5.3): In this step we show that for typical K, x, the plurality in the definition of $g_K(x)$ is obtained with overwhelming probability.

2. LTC step (Lemma 5.4): In this step we use the previous step to show that for a typical g_K, for most K', $g_K|_{K'}$ agrees with some f_S on K', and therefore is a codeword of $C^{\ell \nearrow k}$. By local testability assumption this implies in turn that such g_K is close to being in the code $C^{\ell \nearrow m}$, and we denote by $h_K \in C^{\ell \nearrow m}$ the closest codeword to g_K.

3. Agreement step (Lemma 5.5): In this step we show that a typical h_K agrees with most f_S for $S \supseteq K$.

We start with the boosting step, showing that for typical K, x, the plurality in the definition of $g_K(x)$ is obtained with overwhelming probability. Intuitively, this follows by the assumption that the code has large distance on t-dimensional affine subspaces of the domain $\mathbb{F}_q^s$, together with the fact that k-dimensional affine subspaces sample well inside t-dimensional affine subspaces, which imply that if a pair of f_S agree on K then they must typically also agree on their whole intersection.

Lemma 5.3 (Boosting step)**.**

$$\mathbb{E}_{K,\, x \notin K}\left[\Pr_{S \supseteq K \cup \{x\}}\left[f_S(x) \neq g_K(x) \mid f_S|_K = \mathrm{Plur}_K\right]\right] \leq O\left(q^{-k} \cdot \frac{\varepsilon}{\delta \cdot (1 - 4\varepsilon)}\right).$$

Proof. Since the collision probability lower bounds the probability of hitting the most common value, it suffices to show that

$$\Pr_{K,\,x\notin K,\,S\supseteq K\cup\{x\},\,S'\supseteq K\cup\{x\}}\left[f_S(x)\neq f_{S'}(x)\mid f_S|_K=f_{S'}|_K=\mathrm{Plur}_K\right]\leq O\left(q^{-k}\cdot\frac{\varepsilon}{\delta\cdot(1-4\varepsilon)}\right).$$

Now we have that

$$\begin{aligned}&\Pr_{K,\,x\notin K,\,S\supseteq K\cup\{x\},\,S'\supseteq K\cup\{x\}}\left[f_S(x)\neq f_{S'}(x)\mid f_S|_K=f_{S'}|_K=\mathrm{Plur}_K\right]\\&\leq\Pr_{K,\,T\supseteq K,\,S\supseteq T,\,S'\supseteq T}\left[f_S|_T\neq f_{S'}|_T\mid f_S|_K=f_{S'}|_K=\mathrm{Plur}_K\right]\\&=\frac{\Pr_{T,\,S\supseteq T,\,S'\supseteq T,\,K\subseteq T}\left[f_S|_K=f_{S'}|_K=\mathrm{Plur}_K\mid f_S|_T\neq f_{S'}|_T\right]\cdot\Pr_{T,\,S\supseteq T,\,S'\supseteq T}\left[f_S|_T\neq f_{S'}|_T\right]}{\Pr_{K,\,T\supseteq K,\,S\supseteq T,\,S'\supseteq T}\left[f_S|_K=f_{S'}|_K=\mathrm{Plur}_K\right]},\end{aligned}$$

where the inequality follows recalling that $t=k+1$, and so K and x span a random t-dimensional subspace. Next we bound each of the terms above.

By our assumption (5.1), the right hand term in the numerator is not greater than ε. To bound the denominator note that by Lemma 5.2,

$$\Pr_{K,\,T\supseteq K,\,S\supseteq T,\,S'\supseteq T}[f_S|_K=f_{S'}|_K=\mathrm{Plur}_K]\geq 1-2\cdot\Pr_{K,\,S\supseteq K}[f_S|_K\neq\mathrm{Plur}_K]\geq 1-4\varepsilon\,. \tag{5.4}$$

To bound the left hand term in the numerator note first that since $f_S,f_{S'}$ are both codewords of $C^{\ell\nearrow s}$ then $f_S|_T,f_{S'}|_T$ are distinct codewords in $C^{\ell\nearrow t}$, and so $\mathrm{dist}(f_S|_T,f_{S'}|_T)\geq\delta$. We now apply Theorem 3.8 on the graph $\mathcal{I}(T,\mathcal{K})$, where the ambient space is T, T contains all points in T, and $\mathcal{K}$ contains all k-dimensional subspaces in T. By Theorem 3.8, the graph $\mathcal{I}(T,\mathcal{K})$ is a $(\delta,\frac{q^{-k}}{\delta})$-hitting sampler, and so taking $A'=\{x\in T\mid\ f_S(x)\neq f_{S'}(x)\}$ we deduce that at most a $\frac{q^{-k}}{\delta}$-fraction of K can miss A altogether. Thus,

$$\Pr_{T,\,S\supseteq T,\,S'\supseteq T,\,K\subseteq T}\left[f_S|_K=f_{S'}|_K\;\middle|\;f_S|_T\neq f_{S'}|_T\right]\leq\frac{q^{-k}}{\delta}\,. \tag{5.5}$$

The final bound is obtained by combining the bounds in (5.1), (5.4), and (5.5). □

Next we use the assumption on local testability to show that for a typical K, g_K is close to being a codeword of $C^{\ell\nearrow m}$.

Lemma 5.4 (LTC step)**.**

$$\mathbb{E}_K\left[\mathrm{dist}\left(g_K,C^{\ell\nearrow m}\right)\right]\leq O\left(\frac{\varepsilon}{\alpha\cdot\delta}\right).$$

Proof. To apply our assumption on local testability we first show that $g_K|_{K'}$ is typically a codeword of $C^{\ell\nearrow k}$. For this, first observe that if $g_K|_{K'}$ is not a codeword of $C^{\ell\nearrow k}$ then $g_K|_{K'}\neq f_S|_{K'}$ for all S (since $f_S\in C^{\ell\nearrow s}$ and so $f_S|_{K'}\in C^{\ell\nearrow k}$). Thus we have

$$\begin{aligned}\mathbb{E}_{K,\,K'}\left[\mathbb{1}_{g_K|_{K'}\notin C^{\ell\nearrow k}}\right]&\leq\mathbb{E}_{K,\,K'}\left[\Pr_{S\supseteq K\cup K'}[g_K|_{K'}\neq f_S|_{K'}]\right]\\&\leq\mathbb{E}_{K,\,K'}\left[\Pr_{S\supseteq K\cup K'}\left[g_K|_{K'}\neq f_S|_{K'}\;\middle|\;f_S|_K=\mathrm{Plur}_K\right]\right]+\Pr_{K,\,S\supseteq K}[f_S|_K\neq\mathrm{Plur}_K]\,.\end{aligned}$$

We claim that the above expression is at most $O(\varepsilon/\delta)$. To see this note first that by Lemma 5.2 the right hand term is at most 2ε . To bound the left hand term, note that since each individual point in K' is uniformly distributed in $\mathbb{F}_q^m$ this term is not greater than

$$q^k \cdot \mathbb{E}_{K,\,x}\left[\Pr_{S\supseteq K\cup\{x\}}\left[g_K(x) \neq f_S(x) \;\middle|\; f_S|_K = \mathrm{Plur}_K\right]\right],$$

which is in turn at most $O(\varepsilon/\delta)$ by Lemma 5.3 (noting that the probability in the above expression is zero whenever $x \in K$).

Finally, for any k-dimensional affine subspace K, let $\varepsilon_K := \Pr_{K'}\left[g_K|_{K'} \notin C^{\ell\nearrow k}\right]$. Then on the one hand $\mathbb{E}_K[\varepsilon_K] = \mathbb{E}_{K,\,K'}\left[\mathbb{1}_{g_K|_{K'}\notin C^{\ell\nearrow k}}\right] \leq O(\varepsilon/\delta)$, and on the other hand $\mathrm{dist}(g_K, C^{\ell\nearrow m}) \leq \varepsilon_K/\alpha$ for any K by assumption that $C^{\ell\nearrow m}$ is (k,α)-testable. We conclude that

$$\mathbb{E}_K\left[\mathrm{dist}\left(g_K, C^{\ell\nearrow m}\right)\right] \leq \mathbb{E}_K\left[\frac{\varepsilon_K}{\alpha}\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta}\right). \qquad \square$$

For any k-dimensional affine subspace K, let $h_K \in C^{\ell\nearrow m}$ be the codeword that is closest to g_K. Then by the above lemma,

$$\mathbb{E}_K\left[\mathrm{dist}(g_K, h_K)\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta}\right).$$

The following lemma says that for a typical K, we have that $h_K|_S = f_S$ for most S containing K, which follows by the fact that s-dimensional affine subspaces sample well inside $\mathbb{F}_q^m$ and by assumption that the code has large distance on s-dimensional affine subspaces of the domain $\mathbb{F}_q^m$.

Lemma 5.5 (Agreement step)**.**

$$\mathbb{E}_K\left[\Pr_{S\supseteq K}\left[h_K|_S \neq f_S\right]\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right).$$

Proof. We show that for typical $S \supseteq K$, on the one hand, by Lemma 5.4 and the fact that s-dimensional affine subspaces sample well inside $\mathbb{F}_q^m$, $\mathrm{dist}(h_K|_S, g_K|_S)$ is small, and on the other hand, by Lemma 5.3, $\mathrm{dist}(g_K|_S, f_S)$ is small. We then conclude by triangle inequality that $\mathrm{dist}(h_K|_S, f_S)$ is small, which implies in turn that $h_K|_S = f_S$ by assumption that the code has large distance on s-dimensional subspaces of the domain $\mathbb{F}_q^m$.

We start by showing that $\mathrm{dist}(h_K|_S, g_K|_S)$ is typically small. For this note that for a fixed k-dimensional affine subspace K, and uniform random S containing K, each individual point in $S\setminus K$ is uniformly distributed in $\mathbb{F}_q^m \setminus K$. Thus we have

$$\mathbb{E}_{K,\,S\supseteq K}\left[\mathrm{dist}(h_K|_{S\setminus K}, g_K|_{S\setminus K})\right] \leq \frac{\mathbb{E}_K\left[\mathrm{dist}(h_K, g_K)\right]}{1-q^{-(m-k)}} \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta}\right),$$

where the last inequality follows by Lemma 5.4. Markov's inequality then implies that $\mathrm{dist}(h_K|_{S\setminus K}, g_K|_{S\setminus K}) \leq \frac{\delta}{4}$ with probability at least $1-O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K and $S \supseteq K$. We conclude that with probability at least $1-O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K and $S \supseteq K$, it holds that

$$\mathrm{dist}(h_K|_S, g_K|_S) \leq \frac{\delta}{4} + q^{-(s-k)} \leq \frac{\delta}{2}, \tag{5.6}$$

where the last inequality follows by choice of $s \geq k + \log_q(4/\delta)$.

Next we show that $\mathrm{dist}(g_K|_S, f_S)$ is typically small. For this note that

$$\begin{aligned}
&\mathbb{E}_{K,\, S\supseteq K}\left[\mathrm{dist}(g_K|_S, f_S)\right] \\
&\leq \Pr_{K,\, S\supseteq K}[g_K|_K \neq f_S|_K] + \mathbb{E}_{K,\, S\supseteq K}\left[\mathrm{dist}(g_K|_S, f_S) \mid g_K|_K = f_S|_K\right] \\
&\leq \Pr_{K,\, S\supseteq K}[g_K|_K \neq f_S|_K] + \mathbb{E}_{K,\, S\supseteq K}\left[\mathrm{dist}(g_K|_{S\setminus K}, f_S|_{S\setminus K}) \mid g_K|_K = f_S|_K\right] \\
&= \Pr_{K,\, S\supseteq K}[f_S|_K \neq \mathrm{Plur}_K] + \mathbb{E}_{K,\, x\notin K}\left[\Pr_{S\supseteq K\cup\{x\}}\left[g_K(x) \neq f_S(x) \mid f_S|_K = \mathrm{Plur}_K\right]\right] \\
&\leq O\left(\frac{\varepsilon}{\delta}\right), \qquad (5.7)
\end{aligned}$$

where the last inequality follows by Lemmas 5.2 and 5.3. Markov's inequality then implies that

$$\mathrm{dist}(g_K|_S, f_S) \leq \frac{\delta}{4} \qquad (5.8)$$

with probability at least $1 - O\left(\frac{\varepsilon}{\delta^2}\right)$ over the choice of K and $S \supseteq K$.

Combining (5.6) and (5.7), by a union bound and triangle inequality, we have that $\mathrm{dist}(h_K|_S, f_S) < \delta$ with probability at least $1 - O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K and $S \supseteq K$. Finally, since both $h_K|_S$ and f_S are codewords of $C^{\ell\nearrow s}$ and $\mathrm{mindist}(C^{\ell\nearrow s}) \geq \delta$ we conclude that $h_K|_S \neq f_S$ with probability at most $O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K and $S \supseteq K$. □

5.3 Global structure

We now complete the proof of Lemma 5.1 by showing that there exists a codeword $g \in C^{\ell\nearrow m}$ that agrees with most f_S. We start by showing that most functions h_K are in fact identical, which follows by Lemma 5.5 and the fact that s-dimensional affine subspaces sample well inside $\mathbb{F}_q^m$.

Lemma 5.6. *There exists a k-dimensional affine subspace $\hat{K}$ such that*

$$\Pr_K\left[h_K \neq h_{\hat{K}}\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right).$$

Proof. By Lemma 5.5,

$$\Pr_{K,\, K',\, S\supseteq K\cup K'}\left[h_K|_S \neq h_{K'}|_S\right] \leq 2\cdot \Pr_{K,\, S\supseteq K}\left[h_K|_S \neq f_S\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right),$$

and so by averaging there exists $\hat{K}$ such that

$$\Pr_{K,\, S\supseteq K\cup\hat{K}}\left[h_K|_S \neq h_{\hat{K}}|_S\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right).$$

Markov's inequality then implies that

$$\Pr_{S\supseteq K\cup\hat{K}}\left[h_K|_S \neq h_{\hat{K}}|_S\right] \geq \frac{1}{2}$$

with probability at most $O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K.

Next we claim that when $h_K \neq h_{\hat{K}}$ then $\Pr_{S\supseteq K\cup\hat{K}}\left[h_K|_S \neq h_{\hat{K}}|_S\right] \geq \frac{1}{2}$, and so $h_K \neq h_{\hat{K}}$ with probability at most $O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K. To this end, observe that if $h_K \neq h_{\hat{K}}$ then since h_K, $h_{\hat{K}}$ are both codewords of $C^{\ell\nearrow m}$ and $\mathrm{mindist}(C^{\ell\nearrow m}) \geq \delta$, it must hold that $\mathrm{dist}(h_K, h_{\hat{K}}) \geq \delta$. If h_K and $h_{\hat{K}}$ differ on some point of $K\cup\hat{K}$, then we clearly have that $\Pr_{S\supseteq K\cup\hat{K}}\left[h_K|_S \neq h_{\hat{K}}|_S\right] = 1$, and so we are done. Hence, we may assume that h_K agrees with $h_{\hat{K}}$ on $K\cup\hat{K}$, and so $\mathrm{dist}\left(h_K|_{\mathbb{F}_q^m\setminus(K\cup\hat{K})}, h_{\hat{K}}|_{\mathbb{F}_q^m\setminus(K\cup\hat{K})}\right) \geq \delta$.

We now apply Lemma 3.9 on the graph $\mathcal{I}(\mathbb{F}_q^m\setminus(K\cup\hat{K}), \mathcal{S})$ that connects the points of $\mathbb{F}_q^m\setminus(K\cup\hat{K})$ on the left to the s-dimensional affine subspaces containing $K\cup\hat{K}$ on the right. By Lemma 3.9, the graph $\mathcal{I}(\mathbb{F}_q^m\setminus(K\cup\hat{K}), \mathcal{S})$ is a $(\delta, \frac{4q^{-(s-2k-2)}}{\delta})$-hitting sampler, and so taking

$$A' = \left\{x \in \mathbb{F}_q^m\setminus(K\cup\hat{K}) \,\middle|\, h_K(x) \neq h_{\hat{K}}(x)\right\},$$

we deduce that

$$\Pr_{S\supseteq\hat{K}\cup K}\left[h_{\hat{K}}|_S \neq h_K|_S\right] \geq 1 - \frac{q^{-(s-2k-2)}}{\delta} \geq \frac{1}{2},$$

where the last inequality follows by assumption that $s \geq 2k+2+\log_q(2/\delta)$.

It now follows that $h_K \neq h_{\hat{K}}$ with probability at most $O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right)$ over the choice of K. □

We can now complete the proof of Lemma 5.1.

Proof of Lemma 5.1. Set $g \in C^{\ell\nearrow m}$ to be the function $h_{\hat{K}}$ guaranteed by Lemma 5.6. By Lemmas 5.5 and 5.6,

$$\Pr_S\left[g|_S \neq f_S\right] = \Pr_{K,\,S\supseteq K}\left[g|_S \neq f_S\right] \leq \Pr_K\left[g \neq h_K\right] + \Pr_{K,\,S\supseteq K}\left[h_K|_S \neq f_S\right] \leq O\left(\frac{\varepsilon}{\alpha\cdot\delta^2}\right).$$

So g satisfies (5.2) as required. □

6 From local testing to robust testing

6.1 Proof of Main Theorem 1.1

We can now combine Lemmas 4.1 and 5.1 to prove our Main Theorem 1.1, restated below, showing a transformation from local testing to robust testing.

Main Theorem 1.1. *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq k \geq \ell$. Suppose that $C^{\ell\nearrow m}$ is locally testable using the k-dimensional test with query complexity q^k and soundness α, and let $\delta := \min_{k\leq r\leq m}\mathrm{mindist}(C^{\ell\nearrow r})$. Then $C^{\ell\nearrow m}$ is robustly testable using the $(2k+2+\log_q(4/\delta))$-dimensional test with query complexity $O(q^{2k+2}/\delta)$ and robust soundness $\Omega(\alpha\cdot\delta^3)$.*

Proof of Main Theorem 1.1. By Lemma 5.1 we have that $C^{\ell\nearrow m}$ is $(2k+2+\log_q(4/\delta), k+1, \Omega(\alpha\cdot\delta^2))$-agreement testable, and by Lemma 4.1 this implies in turn that $C^{\ell\nearrow m}$ is $(2k+2+\log_q(4/\delta), \Omega(\alpha\cdot\delta^3))$-robust. □

6.2 Proof of Main Corollary 1.2

We now instantiate our Main Theorem 1.1 with Theorem 3.4 to show that lifted codes are robustly testable. For this, we first observe that one can amplify the soundness of the tester given by Theorem 3.4 to a constant (independent of q and ℓ) at the cost of increasing the testing dimension to $\approx 3\ell$.

Proposition 6.1. *Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code, and $m \geq 3\ell + 1 + \log_q 72$. Then $C^{\ell\nearrow m}$ is $\big(3\ell+1+\log_q 72, \Omega(1)\big)$-testable.*

Proof. If the ℓ-dimensional test rejects with probability at least $\frac{1}{2}\cdot \mathrm{dist}(f, C^{\ell\nearrow m})$ then by Part (3) of Proposition 3.6, the $(3\ell+\log_q 4)$-dimensional test also rejects with the same probability and we are done. Otherwise, by Theorem 3.4, the ℓ-dimensional test rejects with probability at least $\frac{1}{2}\cdot q^{-2\ell}$.

Consider the graph $\mathcal{I}(\mathcal{S},\mathcal{T})$ with left hand side being all ℓ-dimensional affine subspaces of $\mathbb{F}_q^m$ and right hand side being all $(3\ell+1+\log_q 72)$-dimensional affine subspaces of $\mathbb{F}_q^m$. Next we apply Corollary 3.13 on the graph $\mathcal{I}(\mathcal{S},\mathcal{T})$ with A' being the collection of all ℓ-dimensional affine subspaces on which the ℓ-dimensional test rejects. Noting that the $(3\ell+1+\log_q 72)$-dimensional test will reject on any neighbor of A we conclude that the $(3\ell+1+\log_q 72)$-dimensional test rejects with probability at least

$$1-\frac{18q^{-(2\ell+\log_q 72)}}{q^{-2\ell}/2} = 1-\frac{q^{-(2\ell+\log_q 4)}}{q^{-2\ell}/2} = \frac{1}{2}. \qquad \square$$

We now turn to the proof of Main Corollary 1.2, restated below.

Main Corollary 1.2. *There exists an absolute constant $c > 1$ so that the following holds. Let $C \subseteq \{\mathbb{F}_q^\ell \to \mathbb{F}_q\}$ be an affine-invariant linear code of minimum distance δ, and $m \geq \ell$. Then $C^{\ell\nearrow m}$ is robustly testable using the $(6\ell+4+\log_q(c/\delta))$-dimensional test with robust soundness $\Omega(\delta^3)$.*

Proof. Suppose first that $\delta < 2q^{-\ell}$. In this case by Theorem 3.4, $C^{\ell\nearrow m}$ is $(\ell, \Omega(q^{-2\ell}))$-testable, and so by Part (2) of Proposition 3.6, $C^{\ell\nearrow m}$ is also robustly testable using the ℓ-dimensional test with robust soundness $\Omega\big(q^{-3\ell}\big) \geq \Omega(\delta^3)$. By Part (4) of Proposition 3.6 it follows that the $(6\ell+4+\log_q(c/\delta))$-dimensional test also has robust soundness $\Omega(\delta^3)$.

Next assume that $\delta \geq 2q^{-\ell}$. In this case Proposition 3.2 gives that $\mathrm{mindist}(C^{\ell\nearrow r}) \geq \delta/2$ for any $\ell \leq r \leq m$, and so we may apply Proposition 6.1 and Main Theorem 1.1 and conclude that $C^{\ell\nearrow m}$ is $(6\ell+4+\log_q(c/\delta), \Omega(\delta^3))$-robust for an absolute constant $c > 1$. $\square$

References

[1] SANJEEV ARORA: *Probabilistic Checking of Proofs and the Hardness of Approximation Problems*. Ph. D. thesis, UC Berkeley, 1994. ECCC. 8

[2] SANJEEV ARORA, CARSTEN LUND, RAJEEV MOTWANI, MADHU SUDAN, AND MARIO SZEGEDY: Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. Preliminary version in FOCS'92. [doi:10.1145/278298.278306, ECCC:TR98-008] 5, 8

[3] SANJEEV ARORA AND SHMUEL SAFRA: Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. Preliminary version in FOCS'92. [doi:10.1145/273865.273901] 5

[4] SANJEEV ARORA AND MADHU SUDAN: Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003. Preliminary version in STOC'97. [doi:10.1007/s00493-003-0025-0, ECCC:TR97-003] 6, 8

[5] LÁSZLÓ BABAI, LANCE FORTNOW, LEONID A. LEVIN, AND MARIO SZEGEDY: Checking computations in polylogarithmic time. In *Proc. 23rd STOC*, pp. 21–31. ACM Press, 1991. [doi:10.1145/103418.103428] 5

[6] LÁSZLÓ BABAI, LANCE FORTNOW, AND CARSTEN LUND: Non-deterministic exponential time has two-prover interactive protocols. *Comput. Complexity*, 1(1):3–40, 1991. Preliminary version in FOCS'90. [doi:10.1007/BF01200056] 5

[7] ELI BEN-SASSON, ODED GOLDREICH, PRAHLADH HARSHA, MADHU SUDAN, AND SALIL VADHAN: Robust PCPs of proximity, shorter PCPs and applications to coding. *SIAM J. Comput.*, 36(4):889–974, 2006. Preliminary version in STOC'04. [doi:10.1137/S0097539705446810, ECCC:TR04-021] 3

[8] ELI BEN-SASSON, PRAHLADH HARSHA, AND SOFYA RASKHODNIKOVA: Some 3CNF properties are hard to test. *SIAM J. Comput.*, 35(1):1–21, 2005. Preliminary version in STOC'03. [doi:10.1137/S0097539704445445] 2

[9] ELI BEN-SASSON, GHID MAATOUK, AMIR SHPILKA, AND MADHU SUDAN: Symmetric LDPC codes are not necessarily locally testable. In *Proc. 26th IEEE Conf. on Comput. Complexity (CCC'11)*, pp. 55–65. IEEE Comp. Soc., 2011. [doi:10.1109/CCC.2011.14, ECCC:TR10-199] 3

[10] ELI BEN-SASSON AND MADHU SUDAN: Robust locally testable codes and products of codes. *Random Struct. Algor.*, 28(4):387–402, 2006. Preliminary version in RANDOM'04. [doi:10.1002/rsa.20120, arXiv:cs/0408066, ECCC:TR04-046] 3

[11] ELI BEN-SASSON AND MICHAEL VIDERMAN: Composition of semi-LTCs by two-wise tensor products. *Comput. Complexity*, 24(3):601–643, 2015. Preliminary version in RANDOM'09. [doi:10.1007/s00037-013-0074-8, ECCC:TR11-070] 3

[12] MANUEL BLUM, MICHAEL LUBY, AND RONITT RUBINFELD: Self-testing/correcting with applications to numerical problems. *J. Comput. System Sci.*, 47(3):549–595, 1993. Preliminary version in STOC'90. [doi:10.1016/0022-0000(93)90044-W] 6

[13] YOTAM DIKSTEIN AND IRIT DINUR: Agreement testing theorems on layered set systems. In *Proc. 60th FOCS*, pp. 1495–1524. IEEE Comp. Soc., 2019. [doi:10.1109/FOCS.2019.00088, arXiv:1909.00638, ECCC:TR19-112] 6

[14] IRIT DINUR AND ELAZAR GOLDENBERG: Locally testing direct product in the low error range. In *Proc. 49th FOCS*, pp. 613–622. IEEE Comp. Soc., 2008. [doi:10.1109/FOCS.2008.26] 6

[15] IRIT DINUR, PRAHLADH HARSHA, TALI KAUFMAN, AND NOGA RON-ZEWI: From local to robust testing via agreement testing. In *Proc. 10th Innovations in Theoret. Comp. Sci. Conf. (ITCS'19)*, pp. 29:1–18. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.29] 1, 6

[16] IRIT DINUR AND TALI KAUFMAN: High dimensional expanders imply agreement expanders. In *Proc. 58th FOCS*, pp. 974–985. IEEE Comp. Soc., 2017. [doi:10.1109/FOCS.2017.94, ECCC:TR17-089] 6

[17] IRIT DINUR AND INBAL LIVNI NAVON: Exponentially small soundness for the direct product Z-test. In *Proc. 32nd Comput. Complexity Conf. (CCC'17)*, pp. 29:1–29:50. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.CCC.2017.29, ECCC:TR17-096] 6

[18] IRIT DINUR AND OMER REINGOLD: Assignment testers: Towards a combinatorial proof of the PCP Theorem. *SIAM J. Comput.*, 36(4):975–1024, 2006. Preliminary version in FOCS'04. [doi:10.1137/S0097539705446962] 3, 6

[19] IRIT DINUR AND DAVID STEURER: Direct product testing. In *Proc. 29th IEEE Conf. on Comput. Complexity (CCC'14)*, pp. 188–196. IEEE Comp. Soc., 2014. [doi:10.1109/CCC.2014.27, ECCC:TR13-179] 6

[20] URIEL FEIGE, SHAFI GOLDWASSER, LÁSZLÓ LOVÁSZ, SHMUEL SAFRA, AND MARIO SZEGEDY: Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, March 1996. Preliminary version in FOCS'91. [doi:10.1145/226643.226652] 5

[21] PETER GEMMELL, RICHARD J. LIPTON, RONITT RUBINFELD, MADHU SUDAN, AND AVI WIGDERSON: Self-testing/correcting for polynomials and for approximate functions. In *Proc. 23rd STOC*, pp. 32–42. ACM Press, 1991. [doi:10.1145/103418.103429] 6

[22] ODED GOLDREICH AND SHMUEL SAFRA: A combinatorial consistency lemma with application to proving the PCP theorem. *SIAM J. Comput.*, 29(4):1132–1154, 2000. Preliminary version in RANDOM'97. [doi:10.1137/S0097539797315744, ECCC:TR96-047] 6

[23] ALAN GUO, ELAD HARAMATY, AND MADHU SUDAN: Robust testing of lifted codes with applications to low-degree testing. In *Proc. 56th FOCS*, pp. 825–844. IEEE Comp. Soc., 2015. [doi:10.1109/FOCS.2015.56, ECCC:TR15-043] 3, 4, 5, 8

[24] ALAN GUO, SWASTIK KOPPARTY, AND MADHU SUDAN: New affine-invariant codes from lifting. In *Proc. 4th Innovations in Theoret. Comp. Sci. Conf. (ITCS'13)*, pp. 529–540. ACM Press, 2013. [doi:10.1145/2422436.2422494, arXiv:1208.5413, ECCC:TR12-149] 3, 4, 9

[25] ELAD HARAMATY, NOGA RON-ZEWI, AND MADHU SUDAN: Absolutely sound testing of lifted codes. *Theory of Computing*, 11(12):299–338, 2015. Preliminary version in RANDOM'13. [doi:10.4086/toc.2015.v011a012, ECCC:TR13-030] 3, 4

[26] RUSSELL IMPAGLIAZZO, RAGESH JAISWAL, VALENTINE KABANETS, AND AVI WIGDERSON: Uniform direct product theorems: Simplified, optimized, and derandomized. *SIAM J. Comput.*, 39(4):1637–1665, 2010. Preliminary version in STOC'08. [doi:10.1137/080734030, ECCC:TR08-079] 12

[27] RUSSELL IMPAGLIAZZO, VALENTINE KABANETS, AND AVI WIGDERSON: New direct-product testers and 2-query PCPs. *SIAM J. Comput.*, 41(6):1722–1768, 2012. Preliminary version in STOC'09. [doi:10.1137/09077299X, ECCC:TR09-090] 6, 12

[28] TALI KAUFMAN AND MADHU SUDAN: Algebraic property testing: the role of invariance. In *Proc. 40th STOC*, pp. 403–412. ACM Press, 2008. [doi:10.1145/1374376.1374434, ECCC:TR07-111] 4, 5, 10

[29] SWASTIK KOPPARTY, OR MEIR, NOGA RON-ZEWI, AND SHUBHANGI SARAF: High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):11:1–11:42, 2017. Preliminary version in STOC'16. [doi:10.1145/3051093, arXiv:1504.05653, ECCC:TR15-068] 3

[30] RAN RAZ AND SHMUEL SAFRA: A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th STOC*, pp. 475–484. ACM Press, 1997. [doi:10.1145/258533.258641] 6

[31] RONITT RUBINFELD AND MADHU SUDAN: Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. Preliminary versions in STOC'91 and SODA'92. [doi:10.1137/S0097539793255151] 6, 8

[32] MICHAEL VIDERMAN: A combination of testability and decodability by tensor products. *Random Struct. Algor.*, 46(3):572–598, 2015. Preliminary version in RANDOM'12. [doi:10.1002/rsa.20498, arXiv:1105.5806, ECCC:TR11-087] 3, 4, 8

AUTHORS

Irit Dinur
Department of Mathematics and Computer Science
Weizmann Institute
Rehovot, Israel
irit.dinur@weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~dinuri/

Prahladh Harsha
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, India
prahladh@tifr.res.in
http://www.tcs.tifr.res.in/~prahladh/

Tali Kaufman
Department of Computer Science
Bar-Ilan University
Ramat Gan, Israel
kaufmant@mit.edu

Noga Ron-Zewi
Department of Computer Science
University of Haifa
Haifa, Israel
noga@cs.haifa.ac. il
https://cs.haifa.ac.il/~noga

ABOUT THE AUTHORS

IRIT DINUR is a professor at the Weizmann Institute of Science. She received her Ph. D. in 2001 from Tel Aviv University under the supervision of Shmuel Safra. She is interested broadly in theoretical computer science and mathematics, and more specifically in complexity theory, probabilistically checkable proofs, hardness of approximation, and most recently in the growing area of high-dimensional expansion. She has a wife and three kids.

PRAHLADH HARSHA is a theoretical computer scientist at the Tata Institute of Fundamental Research (TIFR). He received his B. Tech. degree in Computer Science and Engineering from the IIT Madras in 1998 and his S. M. and Ph. D. degrees in Computer Science from MIT in 2000 and 2004, respectively; his Ph. D. advisor was Madhu Sudan. Prior to joining TIFR in 2010, he was at Microsoft Research, Silicon Valley and at the Toyota Technological Institute at Chicago. His areas of interest include computational complexity, hardness of approximation, coding theory and information theory. Prahladh credits his mother for his interest in mathematics and dance. He is also deeply indebted to U Koteswara Rao, his high school mentor, for exposing him to the beauty and rigour in mathematics.

TALI KAUFMAN is a professor at Bar Ilan University in Israel. She received her Ph. D. degree from Tel-Aviv University in 2007 under the supervision of Noga Alon. Tali is interested in the theory of computation with a specific interest in locally testable codes, high-dimensional expanders, and their potential implications to CS.

NOGA RON-ZEWI is a professor at the Department of Computer Science at the University of Haifa in Israel. She graduated from the Technion (Israel Institute of Technology) in 2014. Her Ph. D. advisor was Eli Ben-Sasson. Her research interests are in the theory of computation, with a focus on research topics at the interface of communication and computation.