Randomness-efficient Low Degree Tests and Short PCPs via Epsilon-Biased Sets

Eli Ben-Sasson
DEAS, Harvard University and LCS, MIT
Cambridge, MA
eli@eecs.harvard.edu

Madhu Sudan
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, MA
madhu@mit.edu.

Salil Vadhan
Harvard University,
Division of Engineering and Applied Sciences,
Maxwell Dworkin 337, 33 Oxford Street
Cambridge, MA 02138, USA.
salil@eecs.harvard.edu.

Avi Wigderson
Institute for Advanced Study, Princeton and the
Hebrew University, Jerusalem.
Address: Institute for Advanced Study, School of
Math., Einstein Drive, Princeton, NJ 08540.
avi@ias.edu

ABSTRACT
We present the first explicit construction of Probabilistically Checkable Proofs (PCPs) and Locally Testable Codes (LTCs) of fixed constant query complexity which have almost-linear (= n · 2O(√n)) size. Such objects were recently shown to exist (nonconstructively) by Goldreich and Sudan [17]. Previous explicit constructions required size n^4+o(e) with 1/e queries.

The key to these constructions is a nearly optimal randomness-efficient version of the low degree test [32]. In a similar way we give a randomness-efficient version of the BLR linearity test [13] (which is used, for instance, in locally testing the Hadamard code).

The derandomizations are obtained through ε-biased sets for vector spaces over finite fields. The analysis of the derandomized tests rely on alternative views of ε-biased sets — as generating sets of Cayley expander graphs for the low degree test, and as defining linear error-correcting codes for the linearity test.

Categories and Subject Descriptors
F.2.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity—Nonnumerical Algorithms and Problems, Complexity of proof procedures; E.4 [Data]: Coding and Information Theory

General Terms
Theory

Keywords
Probabilistically Checkable Proofs, Locally Testable Codes, Property Testing, Linearity Testing, Low Degree Testing

1. INTRODUCTION
Low degree testing, the problem of testing the proximity of a function to a family of low-degree functions has been a subject of intense examination in the past decade. On the one hand, this task opens up a wide range of intriguing mathematical questions. On the other hand, success in designing and analyzing efficient tests has led to great strides in the design of probabilistically checkable proofs (PCPs) and more recently, in new families of error-correcting codes called locally testable codes (LTCs). In this paper we explore the randomness requirement of such tests and reduce them significantly. Our results translate to explicit constructions of PCPs and LTCs of almost-linear size. We start with some background material.

1.1 PCPs
Probabilistically Checkable Proofs (PCPs) are by now a fundamental object of study in theoretical computer science. The essence of a PCP system is the PCP verifier — a probabilistic algorithm that is given a claimed theorem statement as input and is given oracle access to a purported proof of the theorem. The PCP verifier is allowed to query the proof
12. Locally testable codes

A code \( C \subseteq \Sigma^n \) is locally testable (with \( q \) queries) if there is \((s, \Sigma, \Sigma^*)\)-local explicitly decodable code of rate \( R \) and length \( n \) that can be found in \( \text{poly}(n) \) time. This means that the code \( C \) can be efficiently verified by checking a small number of its coordinates. Specifically, the code \( C \) is locally testable if it can be verified by checking \( q \) randomly chosen bits of the input, and if the probability of accepting a random codeword is \( \geq 1/2 \), and the probability of accepting a random non-codeword is \( \leq 1/3 \).

We say that a code \( C \subseteq \Sigma^n \) is \((s, \Sigma, \Sigma^*)\)-local if there is a polynomial-time algorithm that, given a word \( w \in \Sigma^n \), can determine in \( \text{poly}(n) \) time whether \( w \) is a codeword of \( C \) and, if so, output \( s \) indices \( i_1, \ldots, i_s \) such that \( w[i_1], \ldots, w[i_s] \) is a codeword of \( C \subseteq \Sigma^s \).

The code \( C \subseteq \Sigma^n \) is explicitly decodable if there is a polynomial-time algorithm that, given a word \( w \in \Sigma^n \), can recover the actual codeword from \( w \).

13. Derandomized Low Degree Tests

The construction of \( \gamma \)-constant error \((\gamma, 1\gammaC)\)-testers for \((\gamma, \Sigma, \Sigma^*)\)-local codes is due to Goldreich and Sudan [17].

In [17], Goldreich and Sudan constructed an \((\epsilon, \gamma, \gammaC)\)-tester for \((\gamma, \Sigma, \Sigma^*)\)-local codes, where \( \gammaC \) is a constant that depends only on \( \gamma, \epsilon, \gamma \), and \( \gammaC \) depends only on \( \gamma \). This tester can be used to verify that a given \((\gamma, \Sigma, \Sigma^*)\)-local code is indeed \((\gamma, \Sigma, \Sigma^*)\)-local, and it can also be used to verify that a given \((\gamma, \Sigma, \Sigma^*)\)-local code is \((\gamma, \Sigma, \Sigma^*)\)-local, up to some small error factor.

The derandomized low degree tester for \((\gamma, \Sigma, \Sigma^*)\)-local codes is based on the fact that any \((\gamma, \Sigma, \Sigma^*)\)-local code can be encoded as a polynomial of low degree over \( \Sigma \), and the degree of the polynomial is \( \gamma \).

The low degree tester works by computing the low degree polynomial of the code and checking that it is indeed \((\gamma, \Sigma, \Sigma^*)\)-local. This can be done efficiently, and the tester is polynomial-time.

The derandomized low degree tester is a more efficient version of the low degree tester, which is randomized. The derandomized version uses a small number of random bits to generate the low degree polynomial, and it is deterministic.

The derandomized low degree tester is useful in many applications, including the construction of locally decodable codes and the verification of \((\gamma, \Sigma, \Sigma^*)\)-local codes.
log |F|), but their proofs are nonconstructive. They apply their ideas several times in the paper, and achieve significant improvements (albeit nonconstructive) over existing PCPs and LTFs. At the end, they raise the natural question of derandomizing their technique, which would obviously lead to explicit constructions.

We provide an explicit sample space of nearly linear size for the low degree and linearity tests. The sample space is of the following form: as above, we choose $\tilde{x}$ uniformly from $F^n$, but $\tilde{y}$ is chosen from an $\epsilon$-biased set $S$. Then we apply the same acceptance criterion as the original test.

Thus, we use a sample space of size $|F|^n \cdot |S|$, which is nearly linear in $|F|^n$ if we use the best known explicit constructions of $\epsilon$-biased sets. Of course, the challenge is to show that the test still works correctly, i.e., still rejects functions that are far from being linear (resp., low degree) with high probability. This of course relies on particular properties of $\epsilon$-biased sets, so we turn to discuss those now.

2. Epsilon-Biased Sets

We now present general background on $\epsilon$-biased sets and discuss their equivalent formulations in terms of Cayley expander graphs and linear error-correcting codes, which play a central role in our analyses. Formal definitions and precise statements will appear in the full version. While some of the discussion here generalizes to arbitrary Abelian groups, we restrict ourselves (for simplicity, and since all our applications are such) to vector spaces $F^n$ of dimension $m$ over a finite field $F$.

We recall basic notation about characters and Fourier representations (for more details see e.g. [10, 21]). For a field of characteristic $p$, a character of $F^n$ is a homomorphism $\chi: F^n \to \mu_p$, where $\mu_p$ is the (multiplicative) group of complex $p$th roots of unity. The trivial character maps $F^n$ to 1.

The set of characters forms a basis for the vector space of functions mapping $F^n$ to $\mathbb{C}$. I.e., every function $f: F^n \to \mathbb{C}$ can be written as $\sum_{\alpha} \hat{f}_\alpha \cdot \chi_\alpha$ where the sum is over all characters $\chi$ and $\hat{f}_\alpha$ is the Fourier coefficient of $f$ corresponding to character $\chi$. The Fourier coefficient corresponding to $\chi$ is defined by $\hat{f}_\alpha = \langle f, \chi \rangle \defeq \frac{1}{|F^n|} \sum_{x \in F^n} f(x) \overline{\chi(x)}$. When $F = \mathbb{Z}_p$, each character can be written as $\chi_\alpha$ for $\alpha \in F^n$, where $\chi_\alpha(x) \defeq \omega^{\sum_{i=1}^n \alpha_i x_i}$ (here $\omega$ is a primitive complex $p$th root of unity). In this case we abuse notation and write $f$ for $f_\alpha$.

A set $S \subseteq F^n$ is called $\epsilon$-biased if all nontrivial Fourier coefficients of the characteristic function of $S$ are bounded by $\epsilon |S|/|F^n|$, in absolute value. That is, for every character $\chi \neq 1$, $|\sum_{x \in S} \chi(x)| \leq \epsilon |S|$. These sets are interesting when $|S| \ll |F|^n$ (as $S = 0$-biased). A prime example of an $\epsilon$-biased set is a random set, for which $S$ can be taken to be extremely small (namely, $|S| = O(m \log |F|/\epsilon^2)$). However, most applications of $\epsilon$-biased sets need explicit constructions, i.e., ones which are deterministic and efficient. Clearly, for a derandomization application such as ours, choosing the set at random defeats the whole purpose.

The seminal paper of Naor and Naor [29] defined these sets, gave the first explicit constructions of such sets of size $O(m^c/\epsilon^2)$ for $F = \mathbb{GF}(2)$, and demonstrated the power of these constructions for several applications. Many other constructions followed [1, 23, 2, 33, 14, 4], the best of which have size $O(m^{c} \cdot \log^k |F|/\epsilon^2)$ for various triples of constants $a, b, c \leq 3$.

Since the introduction of explicit $\epsilon$-biased sets by [29], the set and diversity of applications of these objects grew quickly, establishing their fundamental role in theoretical computer science. The settings where $\epsilon$-biased sets are used include: the direct derandomization of algorithms such as fast verification of matrix multiplication and communication protocols for equality [29]; the construction of almost $k$-wise independent random variables, which in turn have many applications [29, 27]; inapproximability results for quadratic equation over $GF(2)$ [20]; learning theory [4]; explicit constructions of Ramsey graphs [28]; and elementary constructions of Cayley expanders [5, 26].

In this paper we add to this long list by applying $\epsilon$-biased sets to the derandomization of low degree tests. To analyze our derandomizations, we rely on two alternative, but equivalent, viewpoints of $\epsilon$-biased sets, which we describe below.

2.1 Epsilon-Biased Sets as Expanders

Any set $S \subseteq F^n$ naturally gives rise to a Cayley graph $G_S$ whose vertices are the elements of $F^n$, and whose edges connect pairs of vectors whose difference is in $S$. (Here we assume $S$ is symmetric ($S = -S$) so as to yield an undirected graph.)

For a $d$-regular graph $G$, the normalized second eigenvalue of $G$ is defined to be $\lambda_2/d \in [0, 1]$, where $\lambda_2$ is the second largest eigenvalue of the adjacency matrix of $G$ in absolute value. It is well-known that this is a good measure of a graph’s expansion (smaller second eigenvalue = better expansion).

The following is implicit in [5, 29].

**Lemma 2.1.** For any $S \subseteq F^n$, $S$ is $\epsilon$-biased if and only if the normalized second eigenvalue of $G_S$ is at most $\epsilon$.

The derandomized sample spaces used in our tests can be viewed as selecting a random edge $(\tilde{x}, \tilde{x} + \tilde{y})$ in $G_S$. This point of view is critical in our analysis of the low degree test. We will use the standard combinatorial implication of the second eigenvalue bound on $G_S$, called the expander mixing lemma (cf. [7]). This lemma says that between every two sets of vertices, the fraction of edges between them is roughly the fraction of edges between them in the complete graph.

**Lemma 2.2 (Expander Mixing Lemma).** (cf. [7]) For $G = (V, E)$ a connected $d$-regular graph over $n$ vertices with normalized second eigenvalue $\lambda$ and any two sets $A, B \subseteq V$ of densities $a = |A|/|V|$, $b = |B|/|V|$, let $\epsilon(a, b)$ be the number of ordered pairs $(u, v) \in A \times B$ such that $(u, v) \in E$. Then $\epsilon(a, b) \leq \lambda \sqrt{ab}$.

In addition to the mixing properties of a Cayley expander, we will also make crucial use of the algebraic structure of the graph, namely that applying $2 \times 2$ $F$-linear transformations to the set of edges yields an isomorphic graph and hence preserves the expansion.
2.2 Epsilon-Biased Sets as Linear Codes

For a field $F$ of characteristic $p$, any set $S \subseteq F^m$ defines a linear code $C_S \subseteq \mu_p^m$ over the alphabet $\mu_p \cong \mathbb{Z}_p$, by restricting all characters $\chi : F^m \to \mu_p$ to $S$. In fact, every $\mathbb{Z}_p$-linear code can be obtained in this way.

It was observed in [29] that the $\varepsilon$-bias property of $S$ is closely related to the error-correcting properties of this linear code. Specifically, suppose $F$ has characteristic 2. Then $S$ is $\varepsilon$-biased iff every pair of distinct codewords in $C_S$ have relative Hamming distance $(1 - \varepsilon)/2$. For fields of characteristic $> 2$, $\varepsilon$-bias does not correspond to Hamming distance, but rather to the (related) following measure of distance. A code $C \subseteq \mu_p^m$ over the alphabet $\mu_p$ is $\varepsilon$-orthogonal if every pair of distinct codewords $x, y \in C$ are nearly orthogonal, i.e., $|(x, y)_S| \leq \varepsilon$, where $(x, y)_S \defeq \frac{1}{|S|} \sum_{z \in S} x(z)y(z)$. (We normalize by $|S|$ so all elements of $\mu_p$ are unit vectors, i.e. $|x|_S^2 \defeq \sqrt{(x, x)_S} = 1$.) By definition, $S \subseteq F^m$ is $\varepsilon$-biased iff $C_S$ is $\varepsilon$-orthogonal. It is not hard to show that if $C$ is $\varepsilon$-orthogonal then every pair of distinct codewords have distance at least $(1 - \varepsilon)/2$, but for $p > 2$ the converse is not necessarily true.

This coding-theoretic point of view will be critical in our analysis of the derandomization linearity test. In particular, we will apply the following variants of two standard coding theory facts to $C_S$ and thereby show that the nearly orthogonal elements of $C_S$, are essentially as good as the orthogonal vectors implicitly arising in the analysis of the original BLR test.

**Lemma 2.3 (ε-Orthogonality Lemma).** For $\varepsilon \geq 0$, let $C \subseteq \mu_p^m$ be an $\varepsilon$-orthogonal code of blocklength $|S|$ over the alphabet $\mu_p$. Let $u \in C^S$ be a vector with $|u|_S \leq 1$. Then,

**[Unique Decoding]** If $|(u, w)_S| \geq 1 - \delta$ for some $w \in C$, then for any $v \in C, v \neq w$,

$|(u, v)_S| \leq \varepsilon + \sqrt{2\delta}$

**[List Decoding]** For any distribution $D$ on $C$,

$\left| \sum_{v \in C} D_v \cdot (u, v)_S \right| \leq \sqrt{|D|} + \varepsilon \leq \sqrt{|D|} + |\varepsilon| + |\delta|$

Let us compare the previous lemma to its “standard” coding theory cousins. Suppose a code has large minimal distance. The cousin of the Unique Decoding part is the simple observation that if $u$ is close to a codeword $v$ then it must be far from any other codeword $w$. The cousin of the List Decoding part is the Johnson Bound. This bound says that only a few codewords can be “somewhat close” to a fixed word $u$. Our List Decoding bound says any large set of codewords in an $\varepsilon$-orthogonal code must have small average inner product with a fixed vector, i.e. most codewords are far from it (on average). Actually, the standard Johnson Bound for binary codes can be deduced from the List Decoding part of the $\varepsilon$-orthogonality lemma.

3. Affineness Testing

In this section we show how to test whether a function $f : \mathbb{Z}_p^m \to \mathbb{Z}_p$ is close to being affine, where the randomness needed is only $(1 + o(1))m \log p$, compared with $2m \log p$ in the previous tests.

3.1 The BLR Affineness Test

For Abelian groups $G$ (additive) and $H$ (multiplicative), an affine function from $G$ to $H$ is a function $f$ such that

$\forall x, y \in G, f(x + y) = f(x) + f(y)$. In the affineness testing problem we are given oracle access to a function $f : G \to H$, and wish to test whether it is close in Hamming distance to some affine function. For $f, g : G \to H$ two functions, we define the agreement of $f$ and $g$ as $Pr_{x \in G}[f(x) = g(x)]$. We are interested in measuring the maximal agreement of $f$ with some affine function. Blum, Luby, and Rubinfeld [13] suggested the following test:

**BLR AffTest**, on function $f : G \to H$

1. Select $x, y \in G$ uniformly at random.
2. Accept if $f(x + y) = f(x) + f(y)$

The original test suggested by [13] was only for homomorphisms, for which $f(0) = 0$, but their techniques generalize to affine functions as well. Indeed, notice that $f$ is affine iff the function $f' \defeq f^{-1}(0) \cdot f$ is a homomorphism. Moreover the acceptance probability of the previous test on the two functions is the same. By definition $f$ is affine if and only if the test accepts for all pairs $x, y \in G$. What is more surprising is that the acceptance probability of the test is a good estimate on the maximal agreement of $f$ with an affine function. This has been shown in [13] by the following theorem.

**Theorem 3.1.** [13] For any finite Abelian groups $G, H$ and any function $f : G \to H$, if $\text{AffTest}$ accepts with probability $\geq 1 - \frac{\varepsilon}{\delta}$, then $f$ has agreement $\geq 1 - \delta$ with some affine function.

For the special case $f : \mathbb{Z}_p^m \to \mathbb{Z}_p$ tighter results are obtained. For the rest of this section we fix $G = \mathbb{Z}_p^m$ and for ease of analysis associate $\mathbb{Z}_p$ with the multiplicative group of complex $p$th roots of unity $\mu_p$. The homomorphisms from $\mathbb{Z}_p^m$ to $\mu_p$ are precisely the characters of $\mathbb{Z}_p^m$ and the affine functions are the characters and their affine shifts, i.e. the set of functions $\{ \zeta \cdot \chi_\alpha : \zeta \in \mu_p, \alpha \in \mathbb{Z}_p^m \}$. Bellare, Coppersmith, Håstad, Kiwi, and Sudan [10] studied the case of $p = 2$ and showed that both the acceptance probability and agreement are naturally expressed using the Fourier representation of $f$. In this case ($p = 2$) the agreement of $f$ with the homomorphism $\chi_\alpha$ is $|1 - f_\alpha|/2$ and hence the agreement with the affine function $-\chi_\alpha$ is $(1 - f_\alpha)/2$. That is, the maximal agreement of $f$ with an affine function equals $(1 + \max|f_\alpha|)/2$. As to the acceptance probability, they proved the following theorem.

**Theorem 3.2.** [10] For $f : \mathbb{Z}_p^m \to \{-1, 1\}$, if $\text{AffTest}$ accepts with probability $\geq \frac{1}{4} + \varepsilon$ then $|f_\alpha| \geq \varepsilon$ for some $\alpha \in \mathbb{Z}_p^m$. In particular, $f$ has agreement $\geq \frac{1}{4} + \varepsilon$ with some affine function.

The extension of the analysis of [10] to $p > 2$ was first considered by Kiwi [25] and later on by Håstad and Wigderson [21]. Although it is easy to express both the acceptance probability and the agreement in terms of Fourier coefficients, now these coefficients might be complex numbers and their interpretation is not as straightforward. Following [21], we add the following assumption on $f$ in order to make the analysis simpler.
Definition 3.3: \( f : \mathbb{Z}_p^m \rightarrow \mu_p \) is said to preserve scalar multiplication if for all \( x \in \mathbb{Z}_p^m \), \( a \in \{1, \ldots, p-1\} \), \( f(a \cdot x) = (f(x))^a \).

Notice that for \( p = 2 \) every function \( f : \mathbb{Z}_p^m \rightarrow \{-1, 1\} \) preserves scalar multiplication. Even when \( p > 2 \), in some applications of affiness testing (e.g. PCPs) this assumption can be effectively achieved by folding [11, 19]: From every class of \( p-1 \) inputs of the type \( \{x, 2x, \ldots, (p-1)x\} \), pick (arbitrarily) a unique representative, and accept it whenever the value of \( f \) is needed on any of these inputs (answering in a way that preserves scalar multiplication).

It is not hard to see that all Fourier coefficients of a multiplication preserving \( f \) are real numbers. Furthermore, the agreement of such an \( f \) with the homomorphism \( \chi_0 \) is \( \frac{1}{p} + \left(1 - \frac{1}{p}\right) \cdot \tilde{f}_0 \). Similarly, for any \( \zeta \in \mu_p \), \( \zeta \neq 1 \), the agreement of \( f \) with the affine function \( \zeta \cdot \chi_0 \) is \( \frac{1}{p} \left(1 - \tilde{f}_0 \right) \). That is, when \( \tilde{f}_0 \) is positive, \( f \) has some non-trivial agreement with \( \chi_0 \). When \( \tilde{f}_0 \) is negative, \( f \) has non-trivial agreement with each of the non-zero affine shifts of \( f \). Hastad and Wigderson present the following straightforward generalization of theorem 3.2 to arbitrary prime \( p \).

Theorem 3.4. [21] For \( f : \mathbb{Z}_p^m \rightarrow \mu_p \), if \( \text{AffTest}^f \) accepts with probability \( \frac{1}{p} + \left(1 - \frac{1}{p}\right) \cdot \delta \), then \( \tilde{f}_0 \geq \delta \) for some \( \alpha \in \mathbb{Z}_p^m \). In particular, \( f \) has agreement \( \geq \frac{1}{p} \left(1 + \delta\right) \) with some affine function.

3.2 Derandomized Affiness Testing

The original testing procedure of [13] uses a sample space of size \( |G|^2 \). We wish to reduce the size of this space (i.e. reduce the amount of randomness required). Notice that we cannot hope to decrease it to less than \( |G|/3 \), because doing so would mean our test does not even query the function on all values. Goldreich and Sudan [17] observed that non-constructively, one can reduce the size of the sample space to \( |G| \cdot \log |H| \). We now present explicit sets which are almost as small as the random construction used by [17] for \( G = \mathbb{Z}_p^m \) and \( H = \mu_p \).

We propose the following derandomized affiness test, that depends on the choice of a set \( S \subseteq \mathbb{Z}_p^m \) (which we will take to be \( \varepsilon \)-biased).

**Derandomized AffTest**

1. Choose \( x \in \mathbb{Z}_p^m \) uniformly at random, and \( y \in S \) uniformly at random.
2. As in the BLR Test, accept if \( f(x) \cdot f(y) = f(x+y) \cdot f(0) \).

It is easy to see that if \( f \) is affine then for any \( S \), the acceptance probability of the derandomized test is 1. The main theorem of this section shows that the converse also holds: if the test accepts with high probability, then \( f \) is close to affine.

Theorem 3.5 (Derandomized Linearity Analysis). Let \( S \subseteq \mathbb{Z}_p^m \) be an \( \varepsilon \)-biased set for \( \varepsilon \geq 0 \), and \( f : \mathbb{Z}_p^m \rightarrow \mu_p \) be an arbitrary function that preserves scalar multiplication. Assume \( \text{AffTest}^f \) accepts with probability \( \frac{1}{p} + \left(1 - \frac{1}{p}\right) \cdot \delta \). Then

1. There exists \( \alpha \in \mathbb{Z}_p^m \) such that \( \tilde{f}_0 \geq \sqrt{2 \gamma - \varepsilon} \). Hence \( f \) has agreement \( \geq \frac{1}{p} \left(1 + \sqrt{2 \gamma - \varepsilon}\right) \) with some affine function.
2. If \( \delta = 1 - \gamma \) then there exists \( \alpha \in \mathbb{Z}_p^m \) such that \( \tilde{f}_0 \geq \sqrt{1 - \sqrt{2 \gamma - \varepsilon}} \). Hence \( f \) has agreement \( \geq \frac{1}{p} \left(1 - \sqrt{2 \gamma - \varepsilon}\right) \) with some affine function.

Since \( S = \mathbb{Z}_p^m \) is 0-biased, for this part 1 of our theorem gives Theorems 3.4, 3.2 as special cases. The best constructions of \( \varepsilon \)-biased sets are of polynomial size in \( \log p \), \( m \) and \( \varepsilon \). Thus we have reduced the randomness required by the linearity test from \( 2m \log p \) bits, down to \( m \log p \cdot \log(\frac{1}{\varepsilon} + \log \log p) = (1 + o(1)) \cdot m \log p \) (the last equality is for fixed \( p \) and \( \varepsilon \), while \( m \) grows to infinity).

3.3 Analysis

Previous analysis of the linearity test for \( \mathbb{Z}_p^m \) proceeds by representing both the agreement of \( f \) with affine functions and the acceptance probability of the test on \( f \), in terms of the Fourier coefficients of \( f \) [10, 21]. We follow a similar path, with a slight twist in the end. As we will shortly see, the acceptance probability of the derandomized test using \( S \), will be expressed in terms of the projection of \( f \) on the \( \varepsilon \)-orthogonal code \( C_S \) (here \( f_S \) is the restriction of \( f \) to the input set \( S \)). Applying the \( \varepsilon \)-orthogonality lemma 2.3 will complete the proof. Now for the details (more of which will appear in the full version).

We assume wlog \( f(0) = 1 \), because the probability of acceptance of our test on the function \( f' \) \( \equiv f \cdot (f(0))^{-1} \) is equal to the acceptance probability on \( f \). Thus, our test checks whether \( f(x)f(y)(f(x+y))^{-1} = 1 \) for random \( x \in \mathbb{Z}_p^m \), \( y \in S \). We express the acceptance probability of the derandomized test on \( f \) in terms of Fourier coefficients of \( f \).

**Lemma 3.6.** The acceptance probability of \( \text{AffTest}^f_S \) on \( f \) preserving scalar multiplication is

\[
\frac{1}{p} \left(1 + \sum_{\alpha \in \mathbb{Z}_p^m} \tilde{f}_0 \cdot \langle f', \chi_\alpha \rangle_S \right) = \sum_{\alpha \in \mathbb{Z}_p^m} \tilde{f}_0 \cdot \Pr[f(y) = \chi_\alpha(y)]
\]

Where \( g^*(x) \equiv (g(x))^a \) and \( \langle h, g \rangle_S \equiv \lvert S \rvert^{-1} \sum_{y \in S} h(y)g(y) \).

Although the second expression above is cleaner, we use the first to complete our analysis.

**Proof (of Theorem 3.5):** The acceptance probability of \( \text{AffTest}^f_S \) is \( \frac{1}{p} + \left(1 - \frac{1}{p}\right) \delta \). So by lemma 3.6 and the triangle inequality there must be some \( \alpha \neq 0 \) such that

\[
\delta \leq \left| \sum_{\alpha \in \mathbb{Z}_p^m} \tilde{f}_0 \cdot \langle f', \chi_\alpha \rangle_S \right|
\]

Fix such an \( \alpha \) and employ the coding-theory ideas and notation of section 2.2. To stress this point of view and to simplify notation, let \( u \) be the restriction of \( f^* \) to the set \( S \), i.e. \( u \) is an element of \( \mu_S^S \). Similarly, for \( \alpha \in \mathbb{Z}_p^m \) let \( v_\alpha \) be the restriction of \( \chi_\alpha = \chi_{(x,0)} \) to \( S \) (\( v_\alpha \) is an element of \( C_S \subseteq \mu_S \)). Using this notation we get \( \langle f^*, \chi_\alpha \rangle_S = \langle u, v_\alpha \rangle_S \). Recall that \( C_S = \{v_\alpha : \alpha \in \mathbb{Z}_p^m\} \) is an \( \varepsilon \)-orthogonal code over the alphabet \( \mu_p \). View \( \{\tilde{f}_0\}_{\alpha \in \mathbb{Z}_p^m} \) as a distribution on \( C_S \).

Indeed, \( \tilde{f}_0 \) is real because \( f \) respects scalar multiplication, and \( \sum_{\alpha} \tilde{f}_0 = 1 \) by Parseval’s identity, so this distribution is well defined.
In order to prove part 1, apply the list decoding part of lemma 2.3 obtaining \( \delta \leq \sqrt{\max f^2} + \varepsilon \). We conclude \( \max |f_{x,\alpha}| \geq \sqrt{N} \varepsilon \). As noted earlier, \( f \) respecting scalar multiplication has agreement at least \( \frac{1}{2}(1 + \max |f_{x,\alpha}|) \) with some affine function, completing the proof of part 1. As to part 2 of the theorem, assume \( \delta > 1 - \gamma \). By averaging there must exist some \( \alpha \) such that \( (u, v_\alpha) \varepsilon \geq 1 - \gamma \). So by the unique decoding part of lemma 2.3. \( (u, v_\alpha) \varepsilon \leq \varepsilon + \sqrt{N} \gamma \), for all \( \beta \in \mathbb{F}^n \). Using Parseval’s identity again we get \( \sum_{\beta \in \mathbb{F}^n} |f_{x,\alpha}(u, v)\beta| \leq \varepsilon + \sqrt{N} \gamma \), so \( 1 - \gamma \leq (1 - \gamma)\sqrt{N} + \varepsilon + \sqrt{N} \gamma \). Moving factors from the right to left hand side completes the proof of the theorem.

4. Low Degree Testing

We start by recalling the low degree testing problem. Let \( \mathbb{F} \) be a finite field of size \( q \). Our aim is to test whether \( f : \mathbb{F}^m \rightarrow \mathbb{F} \) is close to an \( m \)-variate polynomial of total degree \( d \). For \( \bar{x}, \bar{y} \in \mathbb{F}^m \) the line crossing \( \bar{x} \) direction \( \bar{y} \) is the set

\[
\ell_{\bar{x}, \bar{y}} \triangleq \{ \bar{x} + \bar{t} \cdot \bar{y} : t \in \mathbb{F} \}
\]

Let \( L \) be the set of all possible lines in \( \mathbb{F}^m \). Each line \( \ell \) is represented in roughly \( |\mathbb{F}|^2 \) ways above, so we fix some canonical parameterization of each line, i.e. a pair \( (\bar{x}, \bar{y}) \) such that \( \ell = \ell_{\bar{x}, \bar{y}} \). Let \( \mathbb{F}[\ell] \) be the set of univariate polynomials of degree at most \( d \) polynomials over \( \mathbb{F} (t \) is the formal variable). In the version of the low degree testing problem we consider, we are given oracle access to a pair of functions \( f, g \), where \( f : \mathbb{F}^m \rightarrow \mathbb{F} \) and \( g : L \rightarrow \mathbb{F}[\ell] \). The function \( g \) sends each line \( t \) to a degree \( d \) univariate polynomial called its line polynomial, which we interpret as a function from \( t \rightarrow \mathbb{F} \) (under the canonical parameterization of \( t \)). Thus for a line \( \ell \) and a point \( \bar{z} \in \ell \), \( g(\ell)(\bar{z}) \) is well defined. We say \( f \) agrees with the line polynomial \( g(\ell)(\bar{z}) \) if \( f(\bar{z}) = g(\ell)(\bar{z}) \).

If \( f \) is a degree \( d \) polynomial then one can set \( g(\ell) = f \), so \( f \) agrees with all line polynomials on all points in \( \mathbb{F}^m \). This gives rise to the low degree test originally suggested by Rubinfield and Sudan [32].

RS Low Degree Test \( LDTest^{f,g} \), on function \( f : \mathbb{F}^m \rightarrow \mathbb{F} \) and line oracle \( g : L \rightarrow \mathbb{F}[\ell] \):

1. Select \( \bar{x}, \bar{y} \in \mathbb{F}^m \) uniformly and independently at random.
2. Accept if \( f \) agrees with the line polynomial \( g(\ell_{\bar{x}, \bar{y}})(\bar{x}) \) on \( \bar{x} \), i.e. \( f(\bar{x}) = g(\ell_{\bar{x}, \bar{y}})(\bar{x}) \). Otherwise reject.

We stress that when \( g(\ell_{\bar{x}, \bar{y}}) \) is queried, it is done by specifying the canonical representation of the line \( \ell = \ell_{\bar{x}, \bar{y}} \) (rather than the points \( \bar{x}, \bar{y} \) themselves). We also note that the original low degree testing problem involves just a single oracle \( f \), in which case \( g(\ell) \) can be simulated by interpolating \( f \) at \( d+1 \) points on \( t \). But the formulation above, in terms of a separate line oracle, is important in PCP constructions (where one “large” query is preferable to several “small” queries).

For any function \( f : \mathbb{F}^m \rightarrow \mathbb{F} \), there is an optimal line oracle \( g = f \), which maximizes the acceptance probability of the test. This line oracle assigns to each line \( t \) the degree \( d \) polynomial that has the maximal agreement with \( f \) on \( t \). If there are several polynomials that have maximal agreement with \( f \) then the first one (according to a fixed ordering of \( \mathbb{F}[\ell] \)) is selected.

It is not hard to show that \( f \) has degree \( d \) if and only if the test accepts on all pairs \( \bar{x}, \bar{y} \in \mathbb{F}^m \) (when the optimal line oracle is used). The fundamental low degree testing theorem of [32, 6, 3] shows that the rejection probability of the low degree test is a good measure of the Hamming distance of \( f \) from the set of low degree polynomials.

4.1 Derandomized Low Degree Test

The low degree test described above has a sample space of \( \mathbb{F}^m \), which is quadratic in the domain size. We derandomize this test by a method similar to that done for the linearity test:

Derandomized Low Degree Test, \( LDTest^{f,g} \):

1. Select \( \bar{x} \in \mathbb{F}^m , \bar{y} \in \mathbb{S} \) uniformly and independently at random.
2. As in the RS Test, accept if \( f \) agrees with the line polynomial \( g(\ell_{\bar{x}, \bar{y}}) \) at \( \bar{x} \) (otherwise reject).

We analyze the test when \( S \) is \( \lambda \)-biased, and our main theorem is the following. In all theorems presented in this section, we do not try to optimize constants.

**Theorem 4.1 (Low Degree Analysis). There exists a universal constant \( \alpha > 0 \) such that the following holds. Let \( d \leq \mathbb{F}/2, m \leq \mathbb{F}/\log [\mathbb{F}], S \subseteq \mathbb{F}^m \) be a \( \lambda \)-biased set for \( \lambda \geq \mathbb{F}/(m \log |S|) \), and \( \delta \leq \alpha \). If \( Pr[LDTest^{f,g} \text{ accepts}] \geq 1 - \delta \), then \( f \) has distance at most \( \lambda d \) from some polynomial of total degree \( md \).

Using the best constructions of \( \varepsilon \)-biased sets, we get a sample space of nearly linear size, namely \( \mathbb{F}^m \)-polylog\( (|\mathbb{F}|^m) \).

We note that the above Theorem only concludes that \( f \) is close to a polynomial of total degree at most \( \lambda d \). However, the following augmentation to the above test can be used to verify that \( f \) is actually close to degree \( d \). With probability \( 1/2 \), instead of executing the above test, we choose \( \bar{x} \) at random in \( \mathbb{F}^m \) and accept if \( f \) agrees with the line polynomial \( g(\ell_{\bar{x}, \bar{y}}) \) at \( \bar{x} \). (Note that here the direction is completely uniform, instead of being restricted to \( S \).)

In the full version of the paper, we show that if \( f, g \) pass the augmented test with high probability, then \( f \) is actually close to a polynomial of total degree \( d \). The parameters are identical to those in Theorem 4.1, except that we also require \( \lambda d \leq \mathbb{F}/\mathbb{F} \).

4.2 Overview of the Analysis

Our analysis of the derandomized low-degree test can be thought of as a blend of the original low-degree test analysis of [32, 6, 3, 30] and the iterative decoding methods that are used for expander/low-density parity-check (LDPC) codes. To draw this analogy, consider a bipartite graph where the left-hand vertices correspond to points in \( \mathbb{F}^m \) and the right-hand vertices correspond to lines in \( \mathbb{F}^m \). Thus, a degree \( d \) polynomial \( f : \mathbb{F}^m \rightarrow \mathbb{F} \) can be thought of as labelling the left-hand vertices with field elements. \( f \) also induces a labelling of the right-hand vertices by univariate polynomials of degree \( d \) — vertex \( t \) gets labelled by \( f(t) \). These right-hand
vertices can be thought of as giving “consistency” checks on the left-hand vertices, analogous to the parity-check vertices in an LDPC code.

Suppose we have a pair of functions $f : \mathbb{F}^m \to \mathbb{F}$ and $g : L \to \mathbb{F}[d]$ that passes the (original) low degree test with high probability. The original low-degree test analysis in [32, 3] considers a “corrected” function $f' : \mathbb{F}^m \to \mathbb{F}$ obtained by setting the value at each left vertex $x$ to be the majority vote of the value specified by the labelling of its neighbors (i.e. $\text{maj}\{g(f(x))\}$, where the majority is over all lines $t$ crossing $x$). This is analogous to one round of error-correction in the iterative decoding of LDPC codes. Remarkably, with the standard low degree test, it is possible to show that one round of this decoding eliminates all errors (i.e. inconsistent edges), and $f'$ passes the low degree test with probability $1$ (and hence is a degree $d$ polynomial). Since it can also be shown that $f'$ is close to $f$, this completes the analysis. How is it possible that one round of decoding eliminates all the errors? This relies on two main ideas. First, this point-line bipartite graph is extremely well-connected. Every two left-hand vertices have a common neighbor. Thus, one round of error-correction can conceivably yield global consistency. Second, consistency among the lines is forced by their intersections. In particular, lines can be grouped into planes, and the consistency among them can be deduced from the bivariate low-degree analyses of [6, 30].

The analysis of our randomization test begins the same way as for the standard low degree test: we perform majority voting to obtain a corrected function $f'$. However, our graph is not as well-connected by far, since we only consider lines whose direction is in our $\lambda$-biased set. It is much sparser than the full point-line graph, and in particular has nonconstant diameter. So we cannot hope for one round of error-correction to eliminate all errors. However, we can gain hope from the fact that decoding for LDPC codes is done on sparse graphs, and it is known that several rounds of majority voting successfully corrects all errors if the underlying graph is a sufficiently good expander (cf. [34]). In our case, the point-line bipartite graph is very related to the Cayley expander generated by the $\lambda$-biased set. (Recall Section 2.1.) Using this relationship (and the compatibility of our restricted graph structure with the reduction to the bivariate low-degree analysis), we are able to show that after one round of majority voting, the fraction of errors decreases dramatically. That is, we obtain a function $f'$ that passes the low-degree test with much higher probability (roughly speaking). We repeat this process several times, and ultimately obtain a function $F$ passing the low-degree test with probability $1$ (and hence is a low-degree polynomial). It is easy to show that distance accumulated at each step decreases geometrically, so $F$ is indeed close $f$, as desired.

### 4.3 $\lambda$-biased Lines are good Samplers

As mentioned above, a key component in our approach is the relationship between lines used in our test and the expansion properties of the Cayley graph $G_S$. Specifically, we use this connection to show that a random line $\ell_{x,y}$ with direction from $y \in S$ (as used in our LDTest$(\epsilon, \delta)$) has good sampling properties. Specifically, suppose there is some small “bad set” $B \subseteq \mathbb{F}^m$ of density $\mu$. (Think of $B$ as corresponding to points at which the low-degree test fails.) Then we will show that a random line $t$ is unlikely to have large intersection with $B$.

We start by showing that the probability that any two particular points $\hat{x} + a \hat{y}$ and $\hat{x} + b \hat{y}$ on $t$ land in $B$ is roughly $\mu^2$ (even though these points are far from independent). The reason is that any two such points are like a random edge in the expander $G_S$, and thus the Expander Mixing Lemma 2.2 applies. When $a = 0$ and $b = 1$, the points are in fact a random edge in $G_S$. For other values of $a, b$, the algebraic structure of $G_S$ allows us to still relate them to a random edge (they are a random edge in the isomorphic expander $G_{\lambda(-a\epsilon)}$). Thus, we have:

**Lemma 4.2.** Suppose $S \subseteq \mathbb{F}^m$ is $\lambda$-biased. Then, for any $B \subseteq \mathbb{F}^m$ of density $\mu = |B|/|\mathbb{F}^m|$, and any two distinct $a, b \in \mathbb{F}$,

$$\Pr_{\hat{x} \in \mathbb{F}^m \cap (\hat{x} + a \hat{y} \in B) \land (\hat{x} + b \hat{y} \in B)} \leq \mu^2 + \lambda \mu.$$

Once we have analyzed pairs, the sampling property of an entire line follows by a variance computation:

**Lemma 4.3.** *(Sampling Lemma).* Suppose $S \subseteq \mathbb{F}^m$ is $\lambda$-biased. Then, for any $B \subseteq \mathbb{F}^m$ of density $\mu = |B|/|\mathbb{F}^m|$ and any $\varepsilon > 0$,

$$\Pr_{\hat{x} \in \mathbb{F}^m \cap (\hat{x} + \epsilon \hat{y} \in B)} \left( \frac{|E[S \cap B]|}{|E[S]|} - \mu \right) > \varepsilon \leq \left( \frac{1}{|\mathbb{F}|} + \lambda \right) \cdot \frac{\mu}{\varepsilon^2}.$$

In the case $S = \mathbb{F}^m$, this lemma is a standard consequence of pairwise independence of random lines and Chebychev’s inequality. In such a case the error probability is bounded by $\frac{1}{|\mathbb{F}|^2}$. The above lemma says that using a possibly much smaller $\lambda$-biased set for $S$, we obtain almost the same bound. Note that the tail probability will be substantially smaller than $\mu$ (if $1/|\mathbb{F}|$ and $\lambda$ are small relative to $\varepsilon^2$); this will correspond to the error-reduction in one round of correction in our analysis of the low-degree test.

**Proof.** For each $a \in \mathbb{F}$, let $X_a$ be an indicator random variable for $\hat{x} + a \hat{y} \in B$. We are interested in bounding the deviation of $X = \sum X_a$ from its expectation. We can compute the variance as

$$\text{Var}[X] = \sum_{a \in \mathbb{F}} \text{Var}[X_a] + \sum_{a \neq b \in \mathbb{F}} \text{Cov}[X_a, X_b],$$

where $\text{Cov}[X_a, X_b] = E[X_a X_b] - E[X_a] E[X_b]$.

$x$ is picked uniformly at random, so for each $a$, $E[X_a] = \mu$, and since $X_a$ is $\{0, 1\}$-valued we get $\text{Var}[X_a] = \mu - \mu^2 \leq \mu$. For the covariance, we use the previous lemma 4.2 to obtain $E[X_a X_b] \leq \mu^2 + \lambda \mu$, so $\text{Cov}[X_a, X_b] \leq \lambda \mu$.

Thus,

$$\text{Var}[X] \leq |\mathbb{F}| \cdot \mu + |\mathbb{F}|^2 \cdot \lambda \mu.$$

By Chebychev’s Inequality we conclude

$$\Pr[|X - \mu|/|\mathbb{F}| > \varepsilon |\mathbb{F}|] \leq \frac{\text{Var}[X]}{(\varepsilon |\mathbb{F}|)^2} \leq \left( \frac{1}{|\mathbb{F}|} + \lambda \right) \cdot \frac{\mu}{\varepsilon^2}.$$

The proof is complete.

### 4.4 One Round of Correction

The following lemma states what happens in one round of majority-vote error correction. Repeatedly applying this lemma (as in the outline above) yields Theorem 4.1. This
Lemma states that if $S$ is $\lambda$-biased then one can remove a small set of “bad” directions from $S$ and obtain a function that is close to $f$ and passes the new test (using the smaller set of good directions) with significantly higher acceptance probability.

**Lemma 4.4.** There exists a universal constant $c > 0$ such that the following holds. Suppose $d \leq |F|/3$, $S$ is a $\lambda$-biased subset of $F^n$, and $T \subseteq S$ s.t. $|T| \geq |S|/2$. If $Pr[\text{LDTTest}^t_{f, f'} \text{ accepts}] \geq 1 - \delta$, then for any $\gamma, \delta' > 0$ such that $\gamma \delta' = \delta$ and $\delta' \leq 1/60$, there exists $f': F^n \rightarrow F$ and $T' \subseteq T$ with the properties:

1. $|T'| \geq (1 - \gamma)|T|$ (i.e. $T'$ remains large).

2. $\Delta(f', f) \leq 2\delta$ (i.e. $f'$ is close to $f$).

3. $Pr[\text{LDTTest}^t_{f', f} \text{ accepts}] \geq 1 - c(\lambda + m^{-1}) \delta'$. (($f, f_1$) passes the new test with much higher probability than $(f, f_2)$ did).

**Proof Sketch.** A full proof will appear in the full version. Let $T'$ be the set of directions $\hat{y} \in T$ such that for at least $1 - \delta'$ fraction of $x \in F^n$, $f$ agrees with the line polynomial going through $x$ in direction $\hat{y}$. From the fact that the acceptance probability is at least $1 - \delta'$, it follows that $|T'| \geq |T'|(1 - \gamma)$. For each $x \in F^n$ we define the corrected function $f'(x)$ to be the most common value of $f_k(x, \hat{y})$ over $\hat{y} \in T'$ (breaking ties arbitrarily). It is not hard to see that $\Delta(f', f) \leq 2\delta$. In order to prove the third part of the theorem we notice that $Pr[\text{LDTTest}^t_{f', f} \text{ accepts}] \geq Pr_{\hat{y} \in T', \eta \in T'}[f_k(x, \hat{y}) = f_k(x, \hat{z}, \eta)]$. Thus we only need to show that on two random lines $\hat{z}, \hat{z}'$ with directions $\hat{y}, \hat{y}' \in S$ and common origin $\hat{x}$, the polynomials given by $f_k$ agree at their origin with overwhelming probability (greater than the original acceptance probability). As in [3], we prove this by a reduction to bivariate low degree testing. We use Lemma 4.3 to argue that with overwhelming probability, $f$ agrees with $f_k(x, \hat{y})$ and $f_k(x, \hat{z})$ at most points $\hat{w}$ on the lines $\hat{z}$ and $\hat{z}'$ and at most points $\hat{w}'$ on the affine plane $\{x + a \hat{z} + b \hat{z}' : a, b \in F\}$. This enables us to then apply the bivariate testing theorem of [6, 30] to deduce that the two line polynomials agree at the origin of this plane, i.e. $f_k(x, \hat{y}) = f_k(x, \hat{z}, \eta)(\hat{x})$, as needed.

5. **Locally Testable Codes.**

Given the work in [17], it is relatively straightforward to go from a randomness efficient low-degree test to a locally testable code. Here we give a concise statement of the final results and an outline of the necessary steps leading them, leaving the details for the full version.

Our starting point is the standard Reed-Muller code. Thus, we view a bit-class message $M$ as describing the coefficients of a degree $d$, $m$-variate polynomial $P_M$ over a finite field $F$ of size $q$. We code $M$ by giving the evaluation of $P_M$ on all points in $F^m$. For the right setting of parameters, we get constant rate and linear distance (where distance is measured over the alphabet $F$). The next step in the encoding is to give the restriction of $P_M$ to all lines. This was formally defined as a code by Friedl and Sudan [15]. (In [17], it is called the FS-code and we will use their terminology). It is convenient to think of the alphabet of this code as $(d + 1)$-tuples from $F$, i.e. each symbol gives the restriction of $P_M$ to a line in $F^m$ (described by a degree $d$ univariate polynomial). The low degree test analysis of [32] implies that this code is locally testable with query complexity 2, where the test we perform is the natural one, i.e. select a random point in $F^m$ and two random lines going through it, and test if the two polynomials agree on the intersecting point. There are two problems with this code. The first is the quadratic blowup in the encoding size, which translates to inverse quadratic rate (at best). The reason is that there are $|F|^m$ lines. The second problem is the large alphabet size. We deal with these problems one at a time.

Goldreich and Sudan showed that the truncation of the FS-code to a random set of $(1 + o(1))|F|^m$ lines suffices for the low degree test to succeed (i.e. reject any word far from a low degree polynomial with constant probability), resulting in nearly linear size locally testable codes [17]. Plugging in our explicit constructions from the previous section gives explicit codes, summarized by the following lemma.

**Lemma 5.1.** For infinitely many $k$, there exists a polynomial-time constructible family of $GF(2)$-linear locally testable code mapping $k$ bits to $n = k \cdot 2^{O(\sqrt{\log k})}$ symbols over alphabet $\{0, 1\}^k$, where $k = 2^{O(\sqrt{\log k})}$. Furthermore the codes have distance $\Omega(n)$.

**Proof Sketch.** For simplicity we assume all numbers are integers and use the following parameters for our truncated FS-code. Pick $d, m$ such that $d > m^n$ and $|F| = q = cm \cdot d$ for a large constant $c$. A degree $d$, $m$-variate polynomial is defined by $(m^d)$ coefficients, so the message length in bits is $k = \log q \cdot (m^d) \geq \log q \cdot d^m \geq m^2$, and thus $m \leq \sqrt{\log k}$. A codeword in our truncated FS-code consists of the restriction of such a polynomial to all lines used in our low degree test. Each symbol in the alphabet is a degree $d$ univariate polynomial, and hence has bit-length $(d + 1) \log q = 2^{O(\sqrt{\log k})}$, as claimed. For the block length $n$, recall that our low degree tests use $\lambda$-biased sets where $\lambda = O(1/(m \log q))$. The best constructions of such spaces have size $\text{poly}(m, \log q, 1/\lambda) \equiv \text{poly}(m)$. So the number of lines used by our test is $2^{m-1} \cdot \text{poly}(m)$ (we save a factor of $q$ because each line passes through $q$ points) and this equals our block length $n$. Thus,

$$n = q^{m-1} \cdot \text{poly}(m) = \frac{(cmd)^m}{cmd} = k \cdot e^{m-1} \cdot \text{poly}(m) = k \cdot 2^{O(\sqrt{\log k})}$$

as desired. Our low degree test analysis (augmented to test for total degree $d$, rather than $md$) implies that this code is locally testable with 2 queries. The relative distance of this code is at least that of the standard Reed-Muller code (over alphabet $F$) of the same parameters. By the Schwartz-Zippel lemma, this distance is at least $1 - d/q = 1 - o(1)$. Finally, if we select $F = GF(2^t)$, then each linear constraint over $F$ translates to a system of linear constraints over $GF(2)$, so we get $GF(2)$-linear codes. This completes the proof of the lemma.

The remaining problem in the above construction is the large alphabet size $(d + 1) \log q$. This same problem was faced by [17] and they showed how to solve it in two dif-
different ways. For the sake of completeness, we survey their techniques.

Look at a specific test of our testing procedure. We query (the coefficients of) two degree $d$ univariate polynomials $P_1$, $P_2$, and check for some predetermined values $e, e' \in \mathbb{F}$ if the following equation holds
\[ P_1(e) = P_2(e') \tag{2} \]
The first solution given by [17] is to encode \( \{ P_i(e_j) \}_{j \in \mathbb{F}} \) by a locally decodable code over the alphabet $\mathbb{F}$. This results in a reduction of the query complexity, because the value $P_1(e)$ can be decoded by reading only a constant number of values in the locally decodable code (instead of reading $d + 1$ symbols). This reduction in query complexity comes at a dear price in the blocklength, because the only locally decodable codes we know of are very inefficient in terms of their rate.

The second (and better) solution offered by [17] uses PCPs. For each possible test we wish to perform, our code gives a PCP proof of Eq. (2). In other words, our code is the truncated FS code, appended by PCP proofs of the statements in Eq. (2) for all possible tests our randomness-efficient test performs. Now, instead of performing the “line vs. line” low degree test, we test the PCP proof of the corresponding statement Eq. (2). The query complexity of this test is constant, and the blowup to the code size is not too large, because we only need to encode statements of size $2(d + 1) \log q \ll n$, and hence even a polynomial size PCP can be tolerated. The main problem is to show that appending the PCPs results in a code and does not reduce the distance of the truncated FS-code. This problem was once again solved by [17], so plugging in our explicit constructions we get the following theorem. (A proof can be reconstructed using [17] and details will be given in the final version).

**Theorem 5.2.** For infinitely many $k$, there exists a polynomial-time constructible family of linear locally testable codes mapping $k$ bits to $n = k \cdot 2^O(\sqrt{\log n})$ bits. Furthermore, the codes have distance $\Omega(n)$.

## 6. SHORT PCPS

In this section we sketch the short PCPs derived from the efficient low degree test of section 4. We assume the reader has some familiarity with PCP constructions (see [33] for introduction and pointers) and rely heavily on the constructions given in [18].

The PCP construction is built from an outer and inner verifier. The randomness used by the outer verifier dictates to a large extent the total length of the proof. It is this verifier that we make efficient and so focus on it.

Let $\phi$ be a 3-CNF of length $n$. In the standard PCP construction, a (randomized, poly-time, outer) verifier $V$ wishes to check whether $\phi$ is satisfiable, while reading only a small number of symbols from the proof $\pi$ provided by a prover $P$. $V$ assumes $P$ knows a satisfying assignment $A \in \{0, 1\}^n$, and views $A$ as a function $A : H^m \rightarrow \{0, 1\}$ where $H$ is a subset of a small finite field $\mathbb{F}$ ($|H| = h < |\mathbb{F}|$ and $h^m = n$). $V$ requests from $P$ the low degree extension of $A$ to the whole domain $\mathbb{F}^n$ where $|\mathbb{F}^n| = n^{1+o(1)}$. This extension is the function $f : \mathbb{F}^n \rightarrow \mathbb{F}$ and is viewed by $V$ (and us) as an encoding of the assignment $A$. $V$ checks that $f$ is indeed a low degree polynomial and $f[H^m]$ satisfies $\phi$. (Actually, the verifier chooses one of these two tests to perform at random, to save on random bits.)

Our new verifier performs part (i) via the randomness efficient low degree test of section 4. The number of random bits required by this test is $(1 + o(1))\log |\mathbb{F}| = (1 + o(1))\log n$, compared with $2\log |\mathbb{F}| \approx 2\log n$ required by the original low degree test.

As for part (ii), let $C$ be a clause of $\phi$ that depends on the assignment to the three variables given by the triple $A(y_1^C), A(y_2^C), A(y_3^C)$ (where $y_1^C, y_2^C, y_3^C \in H^n$). The standard verifier (e.g., the verifier of [18]) selects a random point $z \in \mathbb{F}^n$ and queries $\pi$ for $f_z$, the restriction of $f$ to a low degree surface $s$ that runs through the four points $(y_1^C, y_2^C, y_3^C, z)$. Since $f$ is assumed to be low degree and $s$ has low degree, then $f_z$ is also a low degree polynomial. Given the description of a low degree polynomial (that is supposed to be $f_z$), $V$ can check if the constraint $C$ is satisfied (because the surface passes through all inputs to $C$).

The problem with this standard verification process is the price we pay in randomness: selecting a constraint $C$ costs $(1 + o(1))\log n$ bits and selecting a random $z \in \mathbb{F}^n$ costs an additional $(1 + o(1))\log n$ bits, summing up to more than $2\log n$ bits.

In order to save in randomness we delve deeper into the efficient PCP constructions of [18] and show that the set of surfaces defined only by the points $A(y_1^C), A(y_2^C), A(y_3^C)$ (with the additional randomness of $z$!) is random enough for our purposes. Namely, a random point on such a random surface is with high probability uniformly distributed over $\mathbb{F}^n$. Using this observation, we show that both tests performed by $V$ can be done with a randomness pricetag of $(1 + o(1))\log n$ bits. Composing this with standard efficient inner verifiers yields the short PCP. Details of the proofs are deferred to the full version.

We state our main theorems using the standard definitions of Multi-Prover Interactive proof systems (MIPs) and Probabilistically Checkable Proofs (PCPs). Namely, we use the notation $L \in \text{MIP}_s[r, a]$ to say the language $L$ has a $r$-prover one round proof system with randomness $r$, answer size $a$, soundness $s$ and completeness $c$. Similarly $L \in \text{PCP}_s[r, q]$ means $L$ has PCP proofs with perfect completeness, soundness $s$, randomness $r$ and query complexity $q$.

The construction sketched above yields the following:

**Theorem 6.1 (Randomness-efficient MIPs).** There exists $\gamma > 0$ and functions $r(n) = \log n + O(\sqrt{\log n \log \log n})$ and $a(n) = 2^{O(\sqrt{\log n})}$ such that SAT is contained in $\text{MIP}_{1,1-\gamma}[3, r, a]$.

Composing this with standard inner verifiers (as in [6, 3]), we obtain:

**Theorem 6.2 (Short PCPs).** There exist constants $\beta < 1$, $q < \infty$, and a function $r(n) = \log n + O(\sqrt{\log n \log \log n})$ such that $\text{SAT} \in \text{PCP}_{1,\beta}[r, q]$. In particular the proof oracles have size $2^{o(n)} = n \cdot 2^{O(\sqrt{\log n})}$ on instances of length $n$. 

---

5 For clarity of exposition, we assume the constraints $V$ verifies are precisely the clauses of $\phi$. The real verifier we use checks more complicated algebraic local constraints, derived from the De-Brun Graph Coloring problem of [30].
7. ACKNOWLEDGEMENTS

We are grateful to Oded Goldreich for getting us started on this research, for contributing to the initial results, and for his active participation in all discussions since.

8. REFERENCES


