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
Massachussettes 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 ($= n \cdot 2^{\tilde{O}(\sqrt{\log n})}$) size. Such objects were recently shown to exist (nonconstructively) by Goldreich and Sudan [17]. Previous explicit constructions required size $n^{1+\Omega(\epsilon)}$ with $1/\epsilon$ 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.

*Supported by NSF grants CCR-0133096, CCR-9877049, CCR 0205390, and NTT Award MIT 2001-04.

†Supported in part by NSF Awards CCR 0205390, and NTT Award MIT 2001-04.

‡URL: `http://www.eecs.harvard.edu/~salil`. Supported by NSF grant CCR-0133096 and a Sloan Research Fellowship.

Categories and Subject Descriptors

F.2.2 [**Theory of Computation**]: Aanlysis 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 oracle for a small number q of bits. If the theorem is valid, there must exist a proof that is accepted by the verifier with probability one, while if the claimed theorem is not valid then no proof should be accepted with probability greater than, say, $1/2$.

Strikingly sharp PCP constructions are known by now for verifying membership in $\mathcal{NP}$-complete languages. For example, the optimal PCP theorem of Håstad [19] shows that there exists a PCP verifier verifying satisfiability of a given CNF formula (SAT) that queries only 3 of bits of the proof, while maintaining the property that the *size* of the proof is only polynomially longer than a traditional proof (namely a satisfying assignment). However, the blowup in the proof size is by a polynomial of huge degree.

Such blow-ups raise the natural question: What is the smallest penalty in proof size that one has to pay for the benefits of probabilistic verifiability with a constant number of queries. This question was first addressed by Babai et al. [9]. Soon after, Polishchuk and Spielman [30] showed how to construct PCP verifiers making $O(1/\epsilon)$ queries to a proof oracle of size just $O(n^{1+\epsilon})$ for SAT (where n denotes the size of the formula being claimed satisfiable).

The results of Polishchuk and Spielman seem near optimal. Yet a recent result of Goldreich and Sudan [17] suggests something better may be possible. They show that there exists a verifier for SAT that accesses proofs of size $n \cdot 2^{O(\sqrt{\log n})} = n^{1+o(1)}$ while querying only a fixed constant number (19) of bits from the proof. Their verifier is obtained by a nonconstructive argument using the Probabilistic Method. Due to the randomness in the construction of the verifier, the applicability to actual proof checking is unclear. We remedy this situation by derandomizing the main part of their construction. Formally we show:

THEOREM 1.1. *SAT has a PCP verifier that on input of length n: (i) accesses proof oracles of size $n \cdot 2^{\tilde{O}(\sqrt{\log n})}$; (ii) makes a constant number of queries into the proof; (iii) accepts valid proofs with probability one; and (iv) rejects invalid theorems with probability at least* $1/2$.

Above and throughout the paper, $\tilde{O}(f)$ means $O(f \cdot (\log f)^c)$ for a fixed but unspecified constant c. We note that while we roughly maintain the size complexity of [17], our constants (in the query complexity) are significantly larger.

1.2 Locally testable codes

A code $C : \Sigma^n \to \Sigma^m$ is *locally testable* (with q queries), if there is a (probabilistic) testing procedure that on every vector $v \in \Sigma^m$, reads q coordinates of v, and rejects it with probability proportional to its distance from the code (i.e., the image of C). That is, elements of the code should always be accepted, and vectors far away from the code should be rejected with constant probability.[1]

Locally testable codes, first defined by Friedl and Sudan [15], may be viewed as the combinatorial centerparts of PCPs. They bear strong intuitive relations to PCPs, with an evident similarity between the testing procedure and the verifier in a PCP system. Indeed, it has been a common belief that such codes lie at the heart of most PCP constructions, though, as pointed out by Goldreich and Sudan [17], this belief is not too well founded. Nevertheless, existence of a locally testable code with a given choice of parameters strongly suggests the existence of a PCP with related parameters, and vice versa. This motivates the study of locally testable codes.

To focus on the main parameter of interest, let us restrict ourselves to the case of binary codes, i.e. $C : \{0,1\}^n \to \{0,1\}^m$. We are interested in minimizing the blocklength $m = m(n)$. We require that the code has linear (i.e., $\Omega(m)$) minimum distance. Under these conditions, classical results of coding theory show that linear blocklength (i.e., $m = O(n)$) can be achieved, explicitly and efficiently. However, when we are interested in *local testability* in addition, it is an intriguing open problem. No lower bounds prevent the possibility of achieving linear length $m = O(n)$ with LTCs of constant query complexity.[2] The best previously known explicit construction of locally testable codes (as follow from [30, 17]) have size $m = n^{1+1/O(q)}$ for a q-query locally testable code. Goldreich and Sudan [17] proved the existence of almost-linear sized codes (i.e, with $m = n^{1+o(1)}$) by a probabilistic argument. Specifically, they show that there exists a constant q and a GF(2)-linear family of LTCs $C_n : \{0,1\}^n \to \{0,1\}^m$ with $m = n \cdot 2^{O(\sqrt{\log n})}$) which can be tested with q queries. We derandomize their construction and restore explicitness, while maintaining the size of the codes.

THEOREM 1.2. *There exist a constant q and an explicitly constructible* GF(2)*-linear family of LTCs $C_n : \{0,1\}^n \to \{0,1\}^m$ with $m = n \cdot 2^{\tilde{O}(\sqrt{\log n})}$ which can be tested with q queries (for infinitely many n).*

1.3 Derandomized Low Degree Tests

Our constructions of PCPs and locally testable codes come via a derandomization of the "low degree test", so we begin by recalling the low degree testing problem along with the related linearity testing problem. In both cases, the testing algorithm is given oracle access to a function $f : \mathbb{F}^m \to \mathbb{F}$, and aims to determine whether f has a certain property (in our case, either linearity or small total degree). The testing algorithm is a randomized procedure that accesses this oracle in a few (ideally constant number of) positions, and then accepts or rejects. We want the tester to accept every f with the property, and reject every f that is far (in Hamming distance) from having the property. This is actually the original setting from which the general area of Property Testing [32, 16] has evolved.

In both the Blum-Luby-Rubinfeld linearity test [13] and the original low degree test [32], the testing algorithm chooses two random points $\vec{x}, \vec{y} \in \mathbb{F}^m$. In the linearity test, it then tests whether $f(\vec{x}) + f(\vec{y}) = f(\vec{x} + \vec{y})$. In the low degree test, it tests whether $f(\vec{x})$ agrees with the univariate polynomial obtained by interpolating f at $d+1$ points on the line $\{\vec{x} + t\vec{y} : t \in \mathbb{F}\}$.

Note that these tests use a sample space of size quadratic in the size of the domain, i.e. one of size $|\mathbb{F}|^{2m}$, due to choosing two random points. Goldreich and Sudan [17] use the Probabilistic Method to show that the size of the sample space can be reduced to nearly linear, namely $O(|\mathbb{F}|^m \cdot$

[1]Note that this is a different requirement than the one for *locally decodable* codes, which asks that if $v \in \Sigma^m$ is close to a codeword $C(u)$, then any coordinate of u can be probabilistically recovered reading only q coordinates of v.

[2]This is in contrast to locally decodable codes, where there is a super-linear lower bound of $m \geq n^{1+\Omega(q)}$ [24], and the best known upper bounds are exponential in n.

$\log|\mathbb{F}|$), but their proof is nonconstructive. They apply their idea several times in the paper, and achieve significant improvements (albeit nonconstructive) over existing PCPs and LTCs. 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 $\vec{x}$ uniformly from $\mathbb{F}^m$, but $\vec{y}$ is chosen from an *ε-biased set S*. Then we apply the same acceptance criterion as the original test.

Thus, we use a sample space of size $|\mathbb{F}|^m \cdot |S|$, which is nearly linear in $|\mathbb{F}|^m$ if we use the best known explicit constructions of ε-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.[3] This of course relies on particular properties of ε-biased sets, so we turn to discuss those now.

2. EPSILON-BIASED SETS

We now present general background on ε-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 $\mathbb{F}^m$ of dimension m over a finite field $\mathbb{F}$.

We recall basic notation about characters and Fourier representations (For more details see e.g. [10, 21]). For $\mathbb{F}$ a field of characteristic p, a *character* of $\mathbb{F}^m$ is a homomorphism $\chi : \mathbb{F}^m \to \mu_p$, where μ_p is the (multiplicative) group of complex pth roots of unity. The *trivial* character maps $\mathbb{F}^m$ to 1. The set of characters form a basis for the vector space of functions mapping $\mathbb{F}^m$ to $\mathbb{C}$. I.e. every function $f : \mathbb{F}^m \to \mathbb{C}$ can be written as $\sum_\chi \hat{f}_\chi \cdot \chi$ where the sum is over all characters χ and $\hat{f}_\chi$ is the *Fourier coefficient* of f corresponding to character χ. The Fourier coefficient corresponding to χ is defined by $\hat{f}_\chi = \langle f, \chi \rangle \stackrel{\text{def}}{=} \frac{1}{|\mathbb{F}^m|} \sum_{x \in \mathbb{F}^m} f(x)\overline{\chi(x)}$. When $\mathbb{F} = \mathbb{Z}_p$ each character can be written as χ_α for $\alpha \in \mathbb{F}^m$, where $\chi_\alpha(x) \stackrel{\text{def}}{=} \omega^{\sum_{i=1}^m \alpha_i x_i}$ (here ω is a primitive complex pth root of unity). In this case we abuse notation and write $\hat{f}_\alpha$ for $\hat{f}_{\chi_\alpha}$.

A set $S \subseteq \mathbb{F}^m$ is called *ε-biased* if all nontrivial Fourier coefficients of the characteristic function of S are bounded by $\varepsilon|S|/|\mathbb{F}^m|$, in absolute value. That is, for every character $\chi \neq 1$, $|\sum_{y \in S} \chi(y)| \leq \varepsilon|S|$. These sets are interesting when $|S| \ll |\mathbb{F}|^m$ (as $S = \mathbb{F}^m$ is 0-biased). A prime example of an ε-biased set is a random set, for which S can be taken to be extremely small (namely, $|S| = O(m \log|\mathbb{F}|/\varepsilon^2)$). However, most applications of ε-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!

[3]Actually, our analysis of the tests give slightly weaker guarantees. In the case of linearity testing, our test is only guaranteed to reject functions that are far from being *affine*, so (after a slight modification) we obtain an affineness test. In the case of testing for total degree d, the test is only guaranteed to reject functions that are far from total degree md, but we can fix this with a slight augmentation to the test.

The seminal paper of Naor and Naor [29] defined these sets, gave the first explicit constructions of such sets of size poly(m/ε) for $\mathbb{F} = \mathrm{GF}(2)$, 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 $O(m^a \cdot \log^b |F|/\varepsilon^c)$ for various triples of constants $a, b, c \leq 3$.

Since the introduction of explicit ε-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 ε-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 ε-biased sets to the derandomization of low degree tests. To analyze our derandomizations, we rely on two alternative, but equivalent, viewpoints of ε-biased sets, which we describe below.

2.1 Epsilon-Biased Sets as Expanders

Any set $S \subseteq \mathbb{F}^m$ naturally gives rise to a Cayley graph G_S whose vertices are the elements of $\mathbb{F}^m$, 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 λ_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 $S \subseteq \mathbb{F}^m$, S is ε-biased iff the normalized second eigenvalue of G_S is at most ε.*

The derandomized sample spaces used in our tests can be viewed as selecting a random edge $(\vec{x}, \vec{x} + \vec{y})$ 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 λ and any two sets $A, B \subseteq V$ of densities $a = |A|/|V|$, $b = |B|/|V|$, let $\mathbf{e}(A, B)$ be the number of ordered pairs $(u, v), u \in A, v \in B$ such that $(u, v) \in E$. Then $\left|\frac{\mathbf{e}(A,B)}{dn} - ab\right| \leq \lambda\sqrt{ab}$.*

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 2×2 $\mathbb{F}$-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 $\mathbb{F}$ of characteristic p, any set $S \subset \mathbb{F}^m$ defines a linear code $\mathcal{C}_S \subseteq \mu_p^S$ over the alphabet $\mu_p \cong \mathbb{Z}_p$, by restricting all characters $\chi : \mathbb{F}^m \to \mu_p$ to S. In fact, every $\mathbb{Z}_p$-linear code can be obtained in this way.

It was observed in [29] that the ε-bias property of S is closely related to the error-correcting properties of this linear code. Specifically, suppose $\mathbb{F}$ has characteristic 2. Then S is ε-biased iff every pair of distinct codewords in $\mathcal{C}_S$ have relative Hamming distance $(1 \pm \varepsilon)/2$. For fields of characteristic > 2, ε-bias does not correspond to Hamming distance, but rather to the (related) following measure of distance. A code $\mathcal{C} \subseteq \mu_p^S$ over the alphabet μ_p is *ε-orthogonal* if every pair of distinct codewords $x, y \in \mathcal{C}$ are nearly orthogonal, i.e. $|\langle x, y\rangle_S| \leq \varepsilon$, where $\langle x, y\rangle_S \stackrel{\text{def}}{=} \frac{1}{|S|}\sum_{z\in S} x(z)\overline{y(z)}$. (We normalize by $|S|$ so all elements of μ_p^S are unit vectors, i.e. $||x||_S \stackrel{\text{def}}{=} \sqrt{\langle x, x\rangle_S} = 1$.) By definition, $S \subset \mathbb{F}^m$ is ε-biased iff $\mathcal{C}_S$ is ε-orthogonal. It is not hard to show that if $\mathcal{C}$ is ε-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 derandomized linearity test. In particular, we will apply the following variants of two standard coding theory facts to $\mathcal{C}_S$ and thereby show that the nearly orthogonal elements of $\mathcal{C}_S$, 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 $\mathcal{C} \subseteq \mu_p^S$ be an ε-orthogonal code of blocklength $|S|$ over the alphabet μ_p. Let $u \in \mathbb{C}^S$ be a vector with $||u||_S \leq 1$. Then,*

[Unique Decoding] If $|\langle u, w\rangle_S| \geq 1-\delta$ for some $w \in \mathcal{C}$, then for any $v \in \mathcal{C}, v \neq w$,

$$|\langle u, v\rangle_S| \leq \varepsilon + \sqrt{2\delta}$$

[List Decoding] For any distribution D on $\mathcal{C}$,

$$\left|\sum_{v\in\mathcal{C}} D_v \cdot \langle v, u\rangle_S\right| \leq \sqrt{||D||^2 + \varepsilon} \leq \sqrt{||D||_\infty + \varepsilon}$$

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 ε-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 ε-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 (multipilicative), an *affine function* from G to H is a function f such that $\forall x, y \in G, f(x)f(y) = f(x+y)f(0)$. 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^f, on function $f : G \to H$

1. Select $x, y \in G$ uniformly at random.
2. Accept if $f(x)f(y) = f(x + y)f(0)$

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 $f' \stackrel{\text{def}}{=} f^{-1}(0) \cdot f$ 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* **AffTest**^{f} *accpets with probablity $\geq 1 - \frac{2}{9}\delta$, 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 pth roots of unity μ_p. The homomorphisms from $\mathbb{Z}_p^m$ to μ_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 χ_α is $(1 + \hat{f}_\alpha)/2$ and hence the agreement with the affine function $-\chi_\alpha$ is $(1 - \hat{f}_\alpha)/2$. That is, the maximal agreement of f with an affine function equals $(1+\max|\hat{f}_\alpha|)/2$. As to the acceptance probability, they proved the following theorem.

Theorem 3.2. *[10] For $f : \mathbb{Z}_2^m \to \{-1, 1\}$, if* **AffTest**^{f} *accepts with probability $\frac{1+\delta}{2}$ then $|\hat{f}_\alpha| \geq \delta$ for some $\alpha \in \mathbb{Z}_2^m$. In particular, f has agreement $\geq \frac{1+\delta}{2}$ 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. *$f : \mathbb{Z}_p^m \to \mu_p$ 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 $f : \mathbb{Z}_2^m \to \{-1, 1\}$ preserves scalar multiplication. Even when $p > 2$, in some applications of affineness 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 access 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 χ_α is $\frac{1}{p} + (1 - \frac{1}{p}) \cdot \hat{f}_\alpha$. Similarly, for any $\zeta \in \mu_p, \zeta \neq 1$, the agreement of f with the affine function $\zeta \cdot \chi_\alpha$ is $\frac{1}{p}(1 - \hat{f}_\alpha)$. That is, when $\hat{f}_\alpha$ is positive, f has some non-trivial agreement with χ_α. When $\hat{f}_\alpha$ is negative, f has non-trivial agreement with each of the non-zero affine shifts of f. Håstad and Wigderson present the following straightforward generalization of theorem 3.2 to arbitrary prime p.

THEOREM 3.4. *[21] For $f : \mathbb{Z}_p^m \to \mu_p$, if $\mathbf{AffTest}^f$ accepts with probability $\frac{1}{p} + (1 - \frac{1}{p}) \cdot \delta$, then $|\hat{f}_\alpha| \geq \delta$ for some $\alpha \in \mathbb{Z}_p^m$. In particular, f has agreement $\geq \frac{1}{p}(1 + \delta)$ with some affine function.*

3.2 Derandomized Affineness 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 affineness test, that depends on the choice of a set $S \subseteq \mathbb{Z}_p^m$ (which we will take to be ε-biased).

Derandomized $\mathbf{AffTest}_S^f$:

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 $f(x)f(y) = f(x+y)f(0)$

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 $S \subset \mathbb{Z}_p^m$ be an ε-biased set for $\varepsilon \geq 0$, and $f : \mathbb{Z}_p^m \to \mu_p$ be an arbitrary function that preserves scalar multiplication. Assume $\mathbf{AffTest}_S^f$ accepts with probability $\frac{1}{p} + (1 - \frac{1}{p}) \cdot \delta$. Then*

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

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

3.3 Analysis

Previous analysis of the linearity test for $\mathbb{Z}_p^m$ proceed 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 $f|_S$ on the ε-orthogonal code $\mathcal{C}_S$ (here $f|_S$ is the restriction of f to the input set S). Applying the ε-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 $f' \stackrel{\text{def}}{=} f \cdot (f(0))^{-1}$ 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 $\mathbf{AffTest}_S^f$ on f preserving scalar multiplication is*

$$\frac{1}{p}\left(1 + \sum_{a=1}^{p-1} \sum_{\alpha \in \mathbb{Z}_p^m} \hat{f}_\alpha^2 \cdot \langle f^a, \chi_\alpha^a \rangle_S\right) = \sum_{\alpha \in \mathbb{Z}_p^m} \hat{f}_\alpha^2 \cdot \Pr_{y \in S}[f(y) = \chi_\alpha(y)]$$

Where $g^a(x) \stackrel{\text{def}}{=} (g(x))^a$ and $\langle h, g \rangle_S \stackrel{\text{def}}{=} |S|^{-1} \sum_{y \in S} h(y)\overline{g(y)}$.

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

Proof (of Theorem 3.5): The acceptance probability of $\mathbf{AffTest}_S^f$ is $\frac{1}{p} + (1 - \frac{1}{p})\delta$. So by lemma 3.6 and the triangle inequality there must be some $a \neq 0$ such that

$$\delta \leq \left| \sum_{\alpha \in \mathbb{Z}_p^m} \hat{f}_\alpha^2 \cdot \langle f^a, \chi_\alpha^a \rangle_S \right|$$

Fix such an a 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^a to the set S, i.e. u is an element of μ_p^S. Similarly, for $\alpha \in \mathbb{Z}_p^m$ let v_α be the restriction of $\chi_\alpha^a = \chi_{(a \cdot \alpha)}$ to S (v_α is an element of $\mathcal{C}_S \subseteq \mu_p^S$). Using this notation we get $\langle f^a, \chi_\alpha^a \rangle_S = \langle u, v_\alpha \rangle_S$ Recall that $\mathcal{C}_S = \{v_\alpha : \alpha \in \mathbb{Z}_p^m\}$ is an ε-orthogonal code over the alphabet μ_p. View $\{\hat{f}_\alpha^2\}_{\alpha \in \mathbb{Z}_p^m}$ as a distribution on $\mathcal{C}_S$. Indeed, $\hat{f}_\alpha$ is real because f respects scalar multiplication, and $\sum_\alpha \hat{f}_\alpha^2 = 1$ 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 $\delta \leq \sqrt{\max \hat{f}_\alpha^2 + \varepsilon}$. We conclude $\max |\hat{f}_\alpha| \geq \sqrt{\delta^2 - \varepsilon}$. As noted earlier, f respecting scalar multiplication has agreement at least $\frac{1}{p}(1 + \max |\hat{f}_\alpha|)$ 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 α such that $\langle u, v_\alpha \rangle_S \geq 1 - \gamma$. So by the unique decoding part of lemma 2.3, $\langle u, v_\beta \rangle_S \leq \varepsilon + \sqrt{2\gamma}$ for all $\beta \in \mathbb{Z}_p^m, \beta \neq \alpha$. Using Parseval's identity again we get $\sum_{\beta \neq \alpha} \hat{f}_\beta^2 \cdot \langle u, v_\beta \rangle_S \leq \varepsilon + \sqrt{2\gamma}$, so $1 - \gamma \leq (1-\gamma)\hat{f}_\alpha^2 + \varepsilon + \sqrt{2\gamma}$. 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 $f : \mathbb{F}^m \to \mathbb{F}$ is close to an m-variate polynomial of total degree d. For $\vec{x}, \vec{y} \in \mathbb{F}^m$ the *line crossing* $\vec{x}$ in *direction* $\vec{y}$ is the set

$$\ell_{\vec{x},\vec{y}} \stackrel{\text{def}}{=} \{\vec{x} + t \cdot \vec{y} : t \in \mathbb{F}\} \tag{1}$$

Let $\mathbb{L}$ be the set of all possible lines in $\mathbb{F}^m$. Each line ℓ is represented in roughly $|\mathbb{F}|^2$ ways above, so we fix some canonical parameterization of each line, i.e. a pair $(\vec{x}, \vec{y})$ such that $\ell = \ell_{\vec{x},\vec{y}}$. Let $\mathbb{F}[t]^d$ 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 $f : \mathbb{F}^m \to \mathbb{F}$ and $g : \mathbb{L} \to \mathbb{F}[t]^d$. The function g sends each line ℓ to a degree d univariate polynomial called its *line polynomial*, which we interpret as a function from $\ell \to \mathbb{F}$ (under the canonical parameterization of ℓ). Thus for a line ℓ and a point $\vec{z} \in \ell$, $g(\ell)(\vec{z})$ is well defined.[4] We say f *agrees* with the line polynomial $g(\ell)$ on $\vec{z}$ if $f(\vec{z}) = g(\ell)(\vec{z})$.

If f is a degree d polynomial then one can set $g(\ell) = f|_\ell$, 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 Rubinfeld and Sudan [32].

RS Low Degree Test $\mathbf{LDTest}^{f,g}$, on function $f : \mathbb{F}^m \to \mathbb{F}$ and line oracle $g : \mathbb{L} \to \mathbb{F}[t]^d$:

1. Select $\vec{x}, \vec{y} \in \mathbb{F}^m$ uniformly and independently at random.

2. Accept if f agrees with the line polynomial $g(\ell_{\vec{x},\vec{y}})$ on $\vec{x}$, i.e. $f(\vec{x}) = g(\ell_{\vec{x},\vec{y}})(\vec{x})$. Otherwise reject.

We stress that when $g(\ell_{\vec{x},\vec{y}})$ is queried, it is done by specifying the canonical representation of the line $\ell = \ell_{\vec{x},\vec{y}}$ (rather than the points $\vec{x}, \vec{y}$ 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 ℓ. 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 $f : \mathbb{F}^m \to \mathbb{F}$, there is an *optimal line oracle* $g = f_{\mathbb{L}}$ which maximizes the acceptance probability of the test. This line oracle assigns to each line ℓ the degree d polynomial that has the maximal agreement with f on ℓ. If there are several polynomials that have maximal agreement with f then the first one (according to a fixed ordering of $\mathbb{F}[t]^d$) is selected.

It is not hard to show that f has degree d if and only if the test accepts on *all* pairs $\vec{x}, \vec{y} \in \mathbb{F}^m$ (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 size $|\mathbb{F}^m|^2$, 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, $\mathbf{LDTest}_S^{f,g}$:

1. Select $\vec{x} \in \mathbb{F}^m, \vec{y} \in S$ uniformly and independently at random.

2. As in the RS Test, accept if f agrees with the line polynomial $g(\ell_{\vec{x},\vec{y}})$ at $\vec{x}$ (otherwise reject).

We analyze the test when S is λ-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 $d \leq |\mathbb{F}|/3$, $m \leq \alpha|\mathbb{F}|/\log|\mathbb{F}|$, $S \subseteq \mathbb{F}^m$ be a λ-biased set for $\lambda \leq \alpha/(m \log |\mathbb{F}|)$, and $\delta \leq \alpha$. If $\Pr[\mathbf{LDTest}_S^{f,g}$ accepts$] \geq 1 - \delta$, then f has distance at most 4δ from some polynomial of total degree md.*

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

We note that the above Theorem only concludes that f is close to a polynomial of total degree at most md. 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 $\vec{x}$ at random in $\mathbb{F}^m$ and accept if f agrees with the line polynomial $g(\ell_{\vec{0},\vec{x}})$ at $\vec{x}$. (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 $md \leq \alpha|\mathbb{F}|$.

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 $f : \mathbb{F}^m \to \mathbb{F}$ 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 ℓ gets labelled by $f|_\ell$. These right-hand

[4] Specifically, let $\vec{x}, \vec{y}$ give the canonical parameterization of ℓ, let t' be such that $\vec{z} = \vec{x} + t' \cdot \vec{y}$, let $p(t) = g(\ell) \in \mathbb{F}[t]^d$ and set $g(\ell)(\vec{z}) = p(t')$.

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 $g : \mathbb{L} \to \mathbb{F}[t]^d$ 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 $\vec{x}$ to be the majority vote of the value specified by the labelling of its neighbors (i.e. $\mathrm{maj}\{g(\ell)(\vec{x})\}$, where the majority is over all lines ℓ crossing $\vec{x}$). 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 derandomized 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 λ-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 λ-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 λ-biased Lines are good Samplers

As mentioned above, a key component in our analysis 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 $\ell_{\vec{x},\vec{y}}$ with direction from $\vec{y} \in S$ (as used in our $\mathbf{LDTest}_S^{f,g}$) has good sampling properties. Specifically, suppose there is some small "bad set" $B \subseteq \mathbb{F}^m$ of density μ. (Think of B as corresponding to points at which the low-degree test fails.) Then we will show that a random line ℓ is unlikely to have large intersection with B.

We start by showing that the probability that any two particular points $\vec{x}+a\cdot\vec{y}$ and $\vec{x}+b\cdot\vec{y}$ on ℓ land in B is roughly μ^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 $G_{(b-a)S}$). Thus, we have:

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

$$\Pr_{\vec{x}\in\mathbb{F}^m,\vec{y}\in S}[(\vec{x}+a\vec{y}\in B)\wedge(\vec{x}+b\vec{y}\in B)] \le \mu^2+\lambda\mu.$$

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 λ-biased. Then, for any $B \subseteq \mathbb{F}^m$ of density $\mu = |B|/|\mathbb{F}^m|$ and any $\varepsilon > 0$,*

$$\Pr_{\vec{x}\in\mathbb{F}^m,\vec{y}\in S}\left[\left|\frac{|\ell_{\vec{x},\vec{y}}\cap B|}{|\ell_{\vec{x},\vec{y}}|}-\mu\right|>\varepsilon\right] \le \left(\frac{1}{|\mathbb{F}|}+\lambda\right)\cdot\frac{\mu}{\varepsilon^2}$$

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 $\frac{\mu}{|\mathbb{F}|\cdot\varepsilon^2}$. The above lemma says that using a possibly much smaller λ-biased set for S, we obtain almost the same bound. Note that the tail probability will be substantially smaller than μ (if $1/|\mathbb{F}|$ and λ are small relative to ε^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 $\vec{x}+a\vec{y}\in B$. We are interested in bounding the deviation of $X = \sum_a X_a$ from its expectation. We can compute the variance as

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

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

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

Thus,

$$\mathbf{Var}[X] \le |\mathbb{F}|\cdot\mu + |\mathbb{F}|^2\cdot\lambda\mu$$

.

By Chebychev's Inequality we conclude

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

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 λ-biased then one can remove a small set of "bad" directions from S and obtain a function that is both 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 $d \le |\mathbb{F}|/3$, S is a λ-biased subset of $\mathbb{F}^m$, and $T \subseteq S$ s.t. $|T| \ge |S|/2$. If $\Pr[\mathbf{LDTest}_T^{f,f_{\mathbb{L}}}$ accepts$] \ge 1 - \delta$, then for any $\gamma, \delta' > 0$ such that $\gamma\delta' = \delta$ and $\delta' \le 1/60$, there exists $f' : \mathbb{F}^m \to \mathbb{F}$ and $T' \subset T$ with the properties:*

1. *$|T'| \ge (1-\gamma)|T|$. (T' remains large).*
2. *$\Delta(f', f) \le 2\delta$. ($f'$ is close to f).*
3. *$\Pr[\mathbf{LDTest}_{T'}^{f',f_{\mathbb{L}}}$ accepts$] \ge 1 - c(\lambda + \frac{1}{|\mathbb{F}|}) \cdot \delta'$. ($(f', f_{\mathbb{L}})$ passes the new test with much higher probability than $(f, f_{\mathbb{L}})$ did).*

PROOF SKETCH. A full proof will appear in the full version. Let T' be the set of directions $\vec{y} \in T$ such that for at least $1 - \delta'$ fraction of $\vec{x} \in \mathbb{F}^m$, f agrees with the line polynomial going through $\vec{x}$ in direction $\vec{y}$. From the fact that the acceptance probability is at least $1 - \gamma\delta'$, it follows that $|T'| \ge |T| \cdot (1-\gamma)$. For each $\vec{x} \in \mathbb{F}^m$, we define the corrected function $f'(\vec{x})$ to be the most common value of $f_{\mathbb{L}}(\ell_{\vec{x},\vec{y}})(\vec{x})$ over $\vec{y} \in T'$ (breaking ties arbitrarily). It is not hard to see that $\Delta(f', f) \le 2\delta$. In order to prove the third part of the theorem we notice that $\Pr[\mathbf{LDTest}_{T'}^{f',f_{\mathbb{L}}}$ accepts$] \ge \Pr_{\vec{x}\in\mathbb{F}^m,\vec{y},\vec{z}\in T'}[f_{\mathbb{L}}(\ell_{\vec{x},\vec{y}})(\vec{x}) = f_{\mathbb{L}}(\ell_{\vec{x},\vec{z}})(\vec{x})]$. Thus we only need to show that on two random lines $\ell_{\vec{x},\vec{y}}$ and $\ell_{\vec{x},\vec{z}}$ with directions $\vec{y}, \vec{z} \in S$ and common origin $\vec{x}$, the polynomials given by $f_{\mathbb{L}}$ agree at their origin with overwhelming probability (higher 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 $f_{\mathbb{L}}(\ell_{\vec{w},\vec{y}})$ and $f_{\mathbb{L}}(\ell_{\vec{w},\vec{z}})$ at most points $\vec{w}$ on the lines $\ell_{\vec{x},\vec{y}}$ and $\ell_{\vec{x},\vec{z}}$) and at most points $\vec{w}$ on the affine plane $\{\vec{x} + a\vec{y} + b\vec{z} : a, b \in \mathbb{F}\}$. 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. $f_{\mathbb{L}}(\ell_{\vec{x},\vec{y}})(\vec{x}) = f_{\mathbb{L}}(\ell_{\vec{x},\vec{z}})(\vec{x})$, 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 neccessary steps leading them, leaving the details for the final version.

Our starting point is the standard Reed-Muller code. Thus, we view a k-bit message M as describing the coefficients of a degree d, m-variate polynomial P_M over a finite field $\mathbb{F}$ of size q. We code M by giving the evaluation of P_M on all points in $\mathbb{F}^m$. For the right setting of parameters, we get constant rate and linear distance (where distance is measured over the alphabet $\mathbb{F}$). 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 $\mathbb{F}$, i.e. each symbol gives the restriction of P_M to a line in $\mathbb{F}^m$ (described by a degree d univariate polynomial). The low degree test analysis of Rubinfeld and Sudan [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 $\mathbb{F}^m$ 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 $|\mathbb{F}|^{2m}$ 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 $(1 + o(1))|\mathbb{F}^m|$ 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 $\mathrm{GF}(2)$-linear locally testable code mapping k bits to $n = k \cdot 2^{O(\sqrt{\log k})}$ symbols over alphabet $\{0,1\}^\ell$, where $\ell = 2^{\tilde{O}(\sqrt{\log k})}$. 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 $d = m^m$ and $|\mathbb{F}| = q = cm \cdot d$ for a large constant c. A degree d, m-variate polynomial is defined by $\binom{m+d}{m}$ coefficients, so the message length in bits is $k = \log q \cdot \binom{m+d}{m} \ge \log q \cdot d^m \ge m^{m^2}$, and thus $m \le \sqrt{\log k}$. 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 $(d+1)\log q = 2^{\tilde{O}(\sqrt{\log k})}$, as claimed. For the block length n, recall that our low degree tests use λ-biased sets where $\lambda = O(1/(m \log q))$. The best constructions of such spaces have size $\mathrm{poly}(m, \log q, 1/\lambda) = \mathrm{poly}(m)$. So the number of lines used by our test is $q^{m-1} \cdot \mathrm{poly}(m)$ (we save a factor of q because each line passes through q points) and this equals our block length n. Thus,

$$\begin{aligned} n &= q^{m-1} \cdot \mathrm{poly}(m) \\ &= \frac{(cmd)^m}{cmd} \cdot \mathrm{poly}(m) \\ &\le k \cdot c^{m-1} \cdot \mathrm{poly}(m) \\ &= k \cdot 2^{O(\sqrt{\log k})} \end{aligned}$$

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 $\mathbb{F}$) of the same parameters. By the Schwartz-Zippel lemma, this distance is at least $1 - d/q = 1 - o(1)$. Finally, if we select $\mathbb{F} = \mathrm{GF}(2^r)$, then each linear constraint over $\mathbb{F}$ translates to a system of linear constraints over $\mathrm{GF}(2)$, so we get $\mathrm{GF}(2)$-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 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

$$P_1(e) = P_2(e') \quad (2)$$

The first solution given by [17] is to encode $\{P_i(e_j)\}_{e_j \in \mathbb{F}}$ by a locally decodable code over the alphabet $\mathbb{F}$. This results in a reduction of the query complexity, because the value $P_i(e)$ 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 $n = k \cdot 2^{\tilde{O}(\sqrt{\log k})}$ 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 [35] 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 ϕ be a 3-CNF of length n. In the standard PCP construction, a (randomized, poly-time, outer) verifier V wishes to check whether ϕ is satisfiable, while reading only a small number of symbols from the proof π provided by a prover P.[5] V assumes P knows a satisfying assignment $A \in \{0,1\}^n$, and views A as a function $A : H^m \to \{0,1\}$ where H is a subest 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 $\mathbb{F}^m$ where $|\mathbb{F}^m| = n^{1+o(1)}$. This extension is the function $f : \mathbb{F}^m \to \mathbb{F}$ and is viewed by V (and us) as an encoding of the assignment A. V checks that *(i)* f is indeed a low degree polynomial and *(ii)* $f|_{H^m}$ satisfies ϕ. (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 $(1+o(1))m\log|\mathbb{F}| = (1+o(1))\log n$, compared with $2m\log|\mathbb{F}| \approx 2\log n$ required by the original low degree test.

As for part *(ii)*, let C be a clause of ϕ 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 $y_1^C, y_2^C, y_3^C \in H^m$). The standard verifier (e.g. the verifier of [18]) selects a random point $z \in \mathbb{F}^m$ and queries π for $f|_s$, 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 $f|_s$ is also a low degree polynomial. Given the description of a low degree polynomial (that is supposed to be $f|_s$), 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 $z \in \mathbb{F}^m$ 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 $A(y_1^C), A(y_2^C), A(y_3^C)$ (without 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 $\mathbb{F}^m$. 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 $L \in \mathrm{MIP}_{c,s}[p,r,a]$ to say the language L has a p-prover one round proof system with randomness r, answer size a, soundness s and completeness c. Similarly $L \in \mathrm{PCP}_s[r,q]$ 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 $\mathrm{MIP}_{1,1-\gamma}[3,r,a]$.*

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 $SAT \in \mathrm{PCP}_{1,\beta}[r,q]$. In particular the proof oracles have size $2^{r(n)} = n \cdot 2^{\tilde{O}(\sqrt{\log n})}$ on instances of length n.*

[5]For clarity of exposition, we assume the constraints V verifies are precisely the clauses of ϕ. The real verifier we use checks more complicated algebraic local constraints, derived from the De-Bruin 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

[1] N. Alon and O. Goldreich and J. Hastad and R. Peralta. Simple Constructions of Almost $k-$wise Independent Random Variables, *Journal of Random Structures and Algorithms*,3:3 (1992), pp 289–304

[2] M. Ajtai, H. Iwaniec, J. Komlos, J. Pintz, E. Szemeredi, Construction of a Thin Set with Small Fourier Coefficients, *Bull. London Math. Soc.* 22, pp. 583-590, 1990.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy. Proof Verification and Intractability of Approximation Problems. *JACM*, Vol. 45, pages 501–555, 1998. Preliminary version in *33rd FOCS*, 1992.

[4] N. Alon, Y. Mansour. ε-Discrepancy sets and their applications for interpolation of sparse polynomials. *Information Processing Letters*, 54:337-342 (1995).

[5] N. Alon, Y. Roichman. Random Cayley Graphs and Expanders. *Rand. Str. Alg.* vol. 5 (1994), 271–284

[6] S. Arora and S. Safra. Probabilistic Checkable Proofs: A New Characterization of NP. *JACM*, Vol. 45, pages 70–122, 1998. Preliminary version in *33rd FOCS*, 1992.

[7] N. Alon, J. Spencer. The Probabilistic Method, published by John Wiley and Sons, Inc. (1992).

[8] S. Arora and M. Sudan. Improved low degree testing and its applications. In *29th STOC*, pages 485–495, 1997.

[9] L. Babai, L. Fortnow, L. Levin, M. Szegedy. Checking Polynomial Computations in Polylogarithmic Time. *23rd STOC*, pages 21–31, 1991.

[10] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, M. Sudan. Linearity testing over characteristic two, *IEEE Transactions on Information Theory* vol. 42(6), pp 1781–1795, November 1996.

[11] M. Bellare, O. Goldreich, M. Sudan. Free bits, PCP and non-approximability - towards tight results. In *SIAM Journal on Computing*, 27(3): 804-915, June 1998. Preliminary version in *Proceedings of the 36th FOCS*, pages 422-431, Milwaukee, Wisconsin, 23-25 October 1995.

[12] E. Ben-Sasson, P. Harsha, M. Sudan. Algebraic Constraint Satisfaction Problems - A Survey. In preparation.

[13] M. Blum, M. Luby and R. Rubinfeld. Self-Testing/Correcting with Applications to Numerical Problems. *JCSS*, Vol. 47, No. 3, pages 549–595, 1993.

[14] G. Even, O. Goldreich, M. Luby, N. Nisan, B. Velickovic Approximations of General Independent Distributions, *STOC'92*, pages 10-16, 1992.

[15] K. Friedl, M. Sudan. Some Improvements to Low Degree Tests. *3rd Israel symp. on Theory and Computing Systems* 1995.

[16] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *JACM*, pages 653–750, July 1998.

[17] O. Goldreich, M. Sudan. Locally Testable Codes and PCPs of Almost-Linear Length *Electronic Colloquium on Computational Complexity*, Report TR02-050.

[18] P. Harsha, M. Sudan. Small PCPs with low query complexity. In *Computational Complexity*, 9(3-4):157-201, 2000.

[19] Johan Håstad. Some Optimal Inapproximability Results. *Journal of the ACM*, pages 798–859, volume 48, 2001.

[20] J. Håstad, S. Phillips and S. Safra. A Well Characterized Approximation Problem. In *Information processing letters*, Vol 47:6, 1993 pp. 301-305.

[21] J. Håstad, A. Wigderson. Simple Analysis of Graph Tests for Linearity and PCP. To appear in *Random Structures and Algorithms*.

[22] S. M. Johnson. A new upper bound for error-correcting codes. *IEEE Trans. on Info. Theory*, 9(1963) pages 198-205.

[23] N. M. Katz. An Estimate for Character Sums. *J. AMS* 2, (1963) pages 197-200.

[24] J. Katz, L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proc. of 32nd STOC*, pages 80-86, ACM, 2000.

[25] M. Kiwi. Probabilistically Checkable Proofs and the testing of Hadammard-like codes. Ph. D. Thesis, MIT.

[26] R. Meshulam, A. Wigderson. Expanders in Group Algebras, *Proc. of the 34th STOC*, pp. 669-677, 2002

[27] R. Motwani, J. Naor, M. Naor. The Probabilistic Method Yields Deterministic Parallel Algorithms. *JCSS* 49(3): 478-516 (1994)

[28] M. Naor. Constructing Ramsey graphs from small probability spaces. *IBM Research Report* RJ 8810, 1992.

[29] J. Naor, M. Naor. Small Bias Probability Spaces: Efficient Constructions and Applications. *22nd STOC* 1990, pages 213-223.

[30] A. Polishchuk and D.A. Spielman. Nearly-linear size holographic proofs. In *26th STOC*, pages 194–203, 1994.

[31] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *29th STOC*, 1997.

[32] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, Vol. 25 (2), pages 252–271, 1996.

[33] A. A. Razborov, A. Wigderson, E. Szemeredi. Constructing Small Sets that Are Uniform in Arithmetic Progressions, *Combinatorics, Probability and Computing* 2:513-518, 1993.

[34] M. Sipser, D. Speilman Expander Codes. In *IEEE Transactions on Information Theory*, 1996, Vol 42, No 6, pp. 1710-1722.

[35] M. Sudan, Scribed by V. Guruswami. Probabilistically Checkable Proofs. Lecture notes, Graduate Summer School on Computational Complexity. `http://theory.lcs.mit.edu/ madhu/pcp/course.html`