On Sums of Locally Testable Affine Invariant Properties

Eli Ben-Sasson*, Elena Grigorescu**, Ghid Maatouk***, Amir Shpilka†, and Madhu Sudan‡

Abstract. Affine-invariant properties are an abstract class of properties that generalize some central algebraic ones, such as linearity and low-degree-ness, that have been studied extensively in the context of property testing. Affine invariant properties consider functions mapping a big field $\mathbb{F}_{q^n}$ to the subfield $\mathbb{F}_q$ and include all properties that form an $\mathbb{F}_q$-vector space and are invariant under affine transformations of the domain. Almost all the known locally testable affine-invariant properties have so-called "single-orbit characterizations" — namely they are specified by a single local constraint on the property, and the "orbit" of this constraint, i.e., translations of this constraint induced by affine-invariance. Single-orbit characterizations by a local constraint are also known to imply local testability. In this work we show that properties with single-orbit characterizations are closed under "summation". To complement this result, we also show that the property of being an n-variate low-degree polynomial over $\mathbb{F}_q$ has a single-orbit characterization (even when the domain is viewed as $\mathbb{F}_{q^n}$ and so has very few affine transformations). As a consequence we find that the sum of any sparse affine-invariant property (properties satisfied by $q^{O(n)}$-functions) with the set of degree d multivariate polynomials over $\mathbb{F}_q$ has a single-orbit characterization (and is hence locally testable) when q is prime. We conclude with some intriguing questions/conjectures attempting to classify all locally testable affine-invariant properties.

Keywords: Property testing, Symmetries, Direct sums, Error-correcting codes

* Department of Computer Science, Technion — Israel Institute of Technology, Haifa, Israel, eli@cs.technion.ac.il. Research funded partially by the European Community's Seventh Framework Programme (FP7/2007-2013) Grant 240258 and the US-Israel Binational Science Foundation Grant 2006104.

** Georgia Tech, Atlanta, GA, elena@cc.gatech.edu. Supported in part by NSF award CCR-0829672 and NSF award 1019343 to the Computing Research Association for the Computing Innovation Fellowship Program.

*** School of Computer and Communications Sciences, EPFL, Switzerland, ghid.maatouk@epfl.ch. Part of this work was conducted while at Microsoft Research. Supported in part by Grant 228021-ECCSciEng of the European Research Council.

† Faculty of Computer Science, Technion — Israel Institute of Technology, Haifa, Israel and Microsoft Research, Cambridge, MA, shpilka@cs.technion.ac.il. Research partially supported by the Israel Science Foundation Grant 339/10.

‡ Microsoft Research New England, Cambridge, Massachusetts, USA, madhu@mit.edu.

1 Introduction

Given finite sets D and R, let $\{D \to R\}$ denote the set of functions mapping D to R. A *property* $\mathcal{F}$ of functions mapping D to R is simply given by a set $\mathcal{F} \subseteq \{D \to R\}$. The goal of property testing [20,12] is to design "query efficient" tests for various properties. Specifically, a (k, ϵ, δ)-tester for $\mathcal{F}$ is a probabilistic oracle algorithm that, given oracle access to a function $f : D \to R$, makes k-queries to f and accepts $f \in \mathcal{F}$ with probability one, while rejecting f that is δ-far from $\mathcal{F}$ with probability at least ϵ. Here, distance is measured by normalized Hamming distance: $\delta(f, g) = |\{x \in D \mid f(x) \neq g(x)\}|/|D|$ denotes the distance between f and g, and $\delta(f, \mathcal{F}) = \min_{g \in \mathcal{F}}\{\delta(f, g)\}$. f is said to be δ-far from $\mathcal{F}$ if $\delta(f, \mathcal{F}) > \delta$ and δ-close otherwise. To minimize notation we say $\mathcal{F}$ is k-locally testable if for every $\delta > 0$ there exists $\epsilon = \epsilon(k, \delta) > 0$ such that $\mathcal{F}$ is (k, ϵ, δ)-locally testable. Our interest is in families of properties that are k-locally testable for some constant k.

In this work we consider testing of "affine-invariant (linear) properties". The domain and range of such properties are fields. Let $\mathbb{F}_q$ denote the field of size q and let $\mathbb{F}_q^*$ denote the non-zero elements of $\mathbb{F}_q$. We consider properties $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ (so q is a prime power and n is a positive integer). $\mathcal{F}$ is *linear* if for every $f, g \in \mathcal{F}$ and $\alpha \in \mathbb{F}_q$, the function $\alpha \cdot f + g$ belongs to $\mathcal{F}$, where $(\alpha \cdot f + g)(x) = \alpha \cdot f(x) + g(x)$. A function $A : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$ is *affine* if there exist $\alpha, \beta \in \mathbb{F}_{q^n}$ such that $A(x) = \alpha x + \beta$. We say A is an *affine permutation* if A is affine and bijective. Note this is equivalent to saying $A(x) = \alpha x + \beta$ for some $\alpha \in \mathbb{F}_{q^n}^*$ and $\beta \in \mathbb{F}_{q^n}$. A property $\mathcal{F}$ is said to be *affine-invariant* if for $f \in \mathcal{F}$ and every affine permutation $A : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$, the function $f \circ A$ is also in $\mathcal{F}$, where $(f \circ A)(x) = f(A(x))$.[1]

The main contribution of this work is to describe a new class of affine-invariant properties that are locally testable. We show that a broad class of locally testable affine-invariant properties (one that includes most known ones) is closed under "sums". But before presenting our results, we motivate the study of affine-invariant properties briefly.

Motivation: The study of affine-invariance was originally motivated in [19] by its connections to locally testable codes and to property testing (cf. the recent survey [21]). Indeed, many "base-constructions" of locally testable codes — crucially used in constructing probabilistically checkable proofs [4,3] — are algebraic in nature and come from families of low-degree polynomials. This motivates the search for the minimal algebraic requirements sufficient to obtain families of locally testable codes, and affine-invariance offers a rich and interesting framework in which to study abstract properties shared by low-degree functions and other algebraic locally testable properties. In this respect, the study of affine-invariant

[1] In all previous works starting with [19], affine-invariance was defined as invariance with respect to all afine functions, and not only with respect to affine permutations. In this paper, we define affine-invariance as invariance with respect to the group of affine-permutations. Fortunately, the class of properties does not change despite the mild change in the definition. We prove this equivalence in the full version [7].

property testing is similar to the study of graph property testing initiated by [12]. Graph-property testing abstracts and unifies properties such as k-colorability and triangle-free-ness, by focussing only on the invariance induced by being a "graph property" (i.e., the property should remain invariant under renaming of the vertices). Affine-invariant testing similarly attempts to abstract and unify algebraic properties such as being linear or of low-degree or a BCH codeword by focussing only on the invariance of the property (and the linearity of the code/property). The study of graph property testing however is much further advanced and has culminated in a complete combinatorial characterization of locally-testable properties in the "dense-graph model"[1,10]. Testing of affine-invariant properties lacks such a characterization and indeed it is not yet clear what shape such a characterization might take.

An additional reason to study affine-invariant properties is because they correspond to *locally correctable codes*. An error correcting code of blocklength n is said to be locally correctable if it has an associated "local corrector". Given an adversarially corrupted codeword $w \in \mathbb{F}_q^n$ and index $i \in \{1, \ldots, n\}$ the (randomized) local corrector makes a *constant* number (hence it is called "local") of queries to entries of w and outputs, with high probability, the ith entry of the "correct" codeword w' — closest in Hamming distance to w. Linear codes that are locally correctable are easily seen to be locally decodable codes as defined by [16] and can be used to construct databases that support private information retrieval [11] (in general though, local correctability is a stronger property than local decodability, see e.g. [6,5]) . It can be verified that affine-invariant locally testable codes are in fact locally correctable [19] hence our results imply new families of locally correctable (and decodable) codes.

Known Testable Properties: Previous works have shown local testability results for two broad categories of affine-invariant properties: (1) Reed-Muller properties, and (2) Sparse properties.

In our language, Reed-Muller properties are obtained by equating the sets $\mathbb{F}_{q^n}$ and $\mathbb{F}_q^n$ with an $\mathbb{F}_q$-linear bijection. This allows us to view $\mathbb{F}_q$-linear subspaces of $\{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ as linear subspaces of $\{\mathbb{F}_q^n \to \mathbb{F}_q\}$ where the latter is the set of n-variate functions over $\mathbb{F}_q$. The *q-ary Reed-Muller property of weight degree w* is given by the set of functions that are n-variate polynomials of degree at most w in this view. The testing result here shows that the Reed-Muller property with parameter w over $\mathbb{F}_q$ is testable with $q^{O(w/q)}$ queries [17] (see also [2,15]), independent of n.

Sparse properties are less structured ones. Roughly, a property is t-sparse if it is of size at most $q^{O(tn)}$. The main theorem here, due to [18] shows that for every prime q and integer t there exists k, such that for every n every t-sparse $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ is k-locally testable.

Aside from the classes above, the known testable properties are less "explicit" and are derived from the concept of single-orbit characterizations, described next.

Single-orbit characterizations: Local tests of linear properties work by picking k query locations $\alpha_1, \ldots, \alpha_k \in \mathbb{F}_{q^n}$ (non-adaptively) and then verifying that

$f(\alpha_1), \ldots, f(\alpha_k)$ satisfy some given constraint (which will restrict this k-tuple to satisfy some linear constraints over $\mathbb{F}_q$). If a property is affine-invariant, it should be equally effective to query $A(\alpha_1), \ldots, A(\alpha_k)$ for some affine permutation A, and then test to see if the function values at these points also satisfy the given constraint. The collection of tests so obtained (by trying out all As) is referred to as the *orbit* of the constraint at $\alpha_1, \ldots, \alpha_k$. If the only functions that satisfy all these constraints are the functions in $\mathcal{F}$, then we say that $\mathcal{F}$ has a *single orbit characterization*.

Single-orbit characterizations seem to be playing a central role in testing of affine-invariant properties. On the one hand, it is known that every k-single-orbit characterized property is k-locally testable [19] and some non-single-orbit characterized properties are known to be not locally-testable even though they can be characterized by a collection of k-local constraints [8]. On the other hand, most known locally testable properties also seem to have some "single-orbit" property. Sparse codes over prime fields were shown to be single-orbit characterized in [18] (see also [14]). The Reed-Muller property has the single orbit property over the (large) group of affine transformations over the vector space $\mathbb{F}_q^n$ by natural considerations. (This will be insufficient for our purposes and so we will strengthen it to get a single-orbit characterization over the field $\mathbb{F}_{q^n}$ in this work.)

Remaining cases of known locally testable codes are obtained in one of two ways: (1) By lifting: This is an operation introduced in [8]. Here we start with a single-orbit property over some field $\mathbb{F}_{q^n}$ and then "lift" this property to one over an extension field $\mathbb{F}_{q^{nm}}$ (in a manner we will describe later). (2) By taking intersections: The intersection of testable properties is always testable. The lifts turn out to be single-orbit characterized by definition, and the intersection of a constant number of single-orbit characterized properties also turns out to be single-orbit characterized essentially by definition.

1.1 Main Result

In this work we extend the class of properties over $\mathbb{F}_{q^n}$ that have single orbit characterizations.

Our first result considers the sum of affine invariant properties. For properties $\mathcal{F}_1, \mathcal{F}_2 \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ their *sum* is $\mathcal{F}_1 + \mathcal{F}_2 = \{f_1 + f_2 \mid f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$. For general linear properties $\mathcal{F}_1 + \mathcal{F}_2$ is also linear, but the testability of $\mathcal{F}_1, \mathcal{F}_2$ does not imply their sum is locally testable. Indeed it may be the case that $\mathcal{F}_1 + \mathcal{F}_2$ satisfies no local constraints. Sums of affine-invariant properties behave more nicely. It is straightforward to see the the sum of affine-invariant properties is affine-invariant. More interestingly, it is also possible to show (relatively easily) that if for every $i \in \{1, 2\}$, $\mathcal{F}_i$ satisfies a k_i-local constraint, then $\mathcal{F}_1 + \mathcal{F}_2$ satisfies a $k_1 \cdot k_2$-local constraint. However this does not seem to imply local-testability. Here we focus on single-orbit characterized properties and prove their sum is single-orbit characterized.

Theorem 1. *For every q, k_1, k_2, there exists $\kappa = \kappa(k_1, k_2, q)$ such that for every n, if $\mathcal{F}_1, \mathcal{F}_2 \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ are affine-invariant properties with $\mathcal{F}_i$ having a k_i-single orbit characterization, then $\mathcal{F}_1 + \mathcal{F}_2$ has a κ-single orbit characterization. Specifically, if $n \geq n_0 = 10k^2 \log k + 10$, where $k = \max\{k_1, k_2\}$, then we can set $\kappa = k_1 \cdot k_2$, else $\kappa = q^{n_0}$ works.*

To apply the theorem above to get new families of single-orbit characterized properties, we need good base properties. However, the two families mentioned earlier, sparse properties and Reed-Muller properties were not known to have the single-orbit property over the same group. Reed-Muller properties were known to have the single-orbit property over the group of affine permutations over $\mathbb{F}_q^n$, while sparse properties are invariant only over $\mathbb{F}_{q^n}$. (And there is no point using the theorem above to prove that the sum of two sparse families is single-orbit — this is already known since the sum of sparse families is also sparse!) To remedy this situation we show that the Reed-Muller property is actually single orbit over the group of affine permutations over $\mathbb{F}_{q^n}$.

Theorem 2 (Reed-Muller codes have local single-orbit property). *Let $q = p^s$ be a prime power. Let w, n be integers such that $w + 1 < \sqrt{\frac{n}{\log_q(3ns)}}$. Denote $w + 1 = r(p-1) + \ell$, where $0 \leq \ell < p - 1$. Then, the q-ary Reed-Muller family of weight degree w, $\mathrm{RM}_q(w, n)$, has a k-single orbit characterization for $k = p^r \cdot (\ell+1)$. In particular, for every w, q there exists a $k = k(w, q)$ such that the q-ary Reed-Muller family of weight degree w has a k-single orbit characterization.*

Indeed an immediate consequence of the two theorems above is that the sum of Reed-Muller and sparse properties over prime fields are locally testable.

Corollary 1. *For integers t, d and prime p, there exists a $k = k(t, d, p)$ such that for every n and every pair of properties $\mathcal{F}_1, \mathcal{F}_2 \in \{\mathbb{F}_{p^n} \to \mathbb{F}_p\}$, where $\mathcal{F}_1$ is the p-ary Reed-Muller property of weight degree d, and $\mathcal{F}_2$ is t-sparse, the property $\mathcal{F}_1 + \mathcal{F}_2$ has a k-single orbit characterization, and is hence k-locally testable.*

The corollary above describes the broadest known class of testable properties when n and q are prime. When n is not prime, the earlier-mentioned notion of lifting leads to other locally testable binary properties, and then intersection also leads to further richness.

Due to space restrictions, we give just a brief hint of the proof of our main theorem. We also describe some of the open questions and conjectures arising from our work. A full version of this work is available as [7].

2 The structure of affine-invariant properties

In what follows $\mathbb{F}_q$ will denote the field of q elements of characteristic p, where $q = p^s$ for some integer s. Let $d = \sum_i d_i p^i$ be the base p representation of an integer d. The **weight** (or p-weight) of d is defined as $\mathrm{wt}(d) = \sum_i d_i$. I.e. it is

the sum of coefficients in the p-ary representation of d. A non-negative integer $e = \sum_i e_i p^i$ is said to be in the p-shadow of d (or simply in the **shadow** of d), denoted $e \leq_p d$, if $e_i \leq d_i$ for all i. We denote $a \equiv_k b$ whenever a is equal to b modulo k. As we will be studying polynomials modulo identities of the form $x^q - x \equiv_p 0$ it will be convenient to define the following variant of the modular operation. Let a and k be integers. We define $a \bmod^* k$ as

$$a \bmod^* k = \begin{cases} 0 & a = 0 \\ b & \text{where } 1 \leq b \leq k \text{ is such that } b \equiv_k a \end{cases}$$

We also say that $a \equiv b \pmod^* k$ if $a \bmod^* k = b \bmod^* k$. Note that the only difference between mod and mod* is that mod* does not send nonzero multiples of k to zero but rather to k. It is now clear that $x^a \equiv_q x^{a \bmod^* q-1}$.

The class of properties that we consider are characterized by their algebraic properties. To describe such properties we need to introduce several notions from the works of [19,13,14,9,8].

We view functions $f : \mathbb{F}_{q^n} \to \mathbb{F}_q$ as functions from $\mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$ whose image just happens to be contained in $\mathbb{F}_q \subseteq \mathbb{F}_{q^n}$. This allows us to view f as (the evaluation of) a univariate polynomial of degree $q^n - 1$.

Let $f(x) = \sum_{d=0}^{q^n-1} c_d x^d$. The *support* of f, denoted $\mathrm{supp}(f)$, is the set $\mathrm{supp}(f) = \{d \mid c_d \neq 0\}$.

The following definition captures an important feature of the structure of affine invariant families.

Definition 1 ($\mathrm{Deg}(\mathcal{F})$)**.** *Let $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ be a family of functions. The **degree set** of $\mathcal{F}$, denoted $\mathrm{Deg}(\mathcal{F})$, is the set of degrees of monomials that appear in some polynomial in $\mathcal{F}$. Formally,*

$$\mathrm{Deg}(\mathcal{F}) = \{d \mid \exists f \in \mathcal{F} \text{ such that } d \in \mathrm{supp}(f)\}.$$

To better understand affine-invariance we need to describe some basic properties of the degree sets (the ones that are known to lead to local testability). We do so in the next two definitions.

Definition 2 ($\mathrm{Shift}(d), \mathrm{Shift}(D)$**, shift-closed, shift-representatives,** $\mathrm{Fam}(D)$)**.** *Let d be an integer in $\{0, \ldots, q^n - 1\}$. The **shift** of d is defined as the set of degrees obtained when taking all q powers of x^d. Formally, $\mathrm{Shift}_{q,n}(d) = \{q^i \cdot d \bmod^* q^n - 1 \mid \forall 0 \leq i \leq n\}$. Recall that $q^i \cdot d \bmod^* q^n - 1$ is the integer d' such that if $d \neq 0$ then $d' \equiv q^i d \mod (q^n - 1)$ and $1 \leq d' \leq q^n - 1$, and if $d = 0$ then $d' = 0$. (In what follows, we will always be considering degrees in the support of functions from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$, so that we drop the subscripts.)*

*We extend the notion to a set of degrees naturally. For a set $D \subseteq \{0, \ldots, q^n - 1\}$, the shift of D is defined as $\mathrm{Shift}(D) = \bigcup_{d \in D} \mathrm{Shift}(d)$. A set D is said to be **shift-closed** if $\mathrm{Shift}(D) = D$. For a shift-closed D, a set $S \subseteq D$ is said to be a set of* shift-representatives *of D if $\mathrm{Shift}(S) = D$ and $\mathrm{Shift}(d) \cap \mathrm{Shift}(d') = \emptyset$ for $d, d' \in S$. (In other words S contains one element from each "shift" class*

in D; by convention we assume each element of S is the smallest amongst its shifts.)[2]

Finally, for a shift-closed D, we define $\mathrm{Fam}(D) = \{\mathrm{Trace}(f) \mid f : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}, \mathrm{supp}(f) \subseteq D\}$.

Another important ingredient that we will use is the *shadow* of a degree.

Definition 3 (Shadow, Shadow-closed set). *For a non-negative integer d, the* ***shadow*** *of d is the set* $\mathrm{Shadow}(d) = \{e \mid e \leq_p d\}$. *The shadow of a set S of non-negative integers is simply the union of the shadows of its elements, i.e.,* $\mathrm{Shadow}(S) = \bigcup_{d \in S} \mathrm{Shadow}(d)$. *A set S of non-negative integers is* ***shadow-closed*** *if* $\mathrm{Shadow}(S) = S$.

For a general (linear) family $\mathcal{F}$, the degree set of $\mathcal{F}$ does not give much useful information about $\mathcal{F}$. However, for affine invariant families, this set completely describes the family. Furthermore, sets of degrees that are closed under shifts and under shadows completely characterize affine-invariant properties.

Our next lemma repeats in different forms in the literature [19,13,14,9]. Specifically, it is Lemma 3.5 in [8].

Lemma 1 (Closed degree sets specify affine-invariant properties). *Let $\mathcal{F}$ be a linear and affine-invariant family. Then* $\mathrm{Deg}(\mathcal{F})$ *is shadow-closed and shift-closed, and* $\mathcal{F} = \mathrm{Fam}(\mathrm{Deg}(\mathcal{F}))$. *Conversely, if D is shadow-closed and shift-closed then D is the degree set of some affine invariant family. More specifically,* $\mathrm{Fam}(D)$ *is affine-invariant and* $D = \mathrm{Deg}(\mathrm{Fam}(D))$.

3 Sums of Affine-Invariant Properties

In this section we prove Theorem 1. The main idea behind the proof is that instead of looking at the sets of degrees of a locally characterizable family $\mathcal{F}$, we look at the *border* set of degrees. These are the integers that do not themselves belong to $\mathrm{Deg}(\mathcal{F})$ but every integer in their shadow is in $\mathrm{Deg}(\mathcal{F})$.

Definition 4 (Border of a family). *Let $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ be a family of functions. The* ***border*** *of $\mathcal{F}$ is the set of degrees given by*

$$\mathrm{Border}(\mathcal{F}) = \{d \notin \mathrm{Deg}(\mathcal{F}) \mid \forall e <_p d,\ e \in \mathrm{Deg}(\mathcal{F})\}.$$

We start by noticing that a *k-single orbit characterization* can be specified by a pair $(\bar{\alpha}; \{\bar{\lambda}_i\}_{i=1}^t)$, where $\bar{\alpha} = (\alpha_1, \ldots, \alpha_k) \in \mathbb{F}_{q^n}^k$ and $\bar{\lambda}_i = (\lambda_{i,1}, \ldots, \lambda_{i,k}) \in \mathbb{F}_q^k$, and $f \in \mathcal{F}$ if and only if it satisfies $\sum_{j=1}^k \lambda_{i,j} f(\pi(\alpha_j)) = 0$ for every $i \in \{1, \ldots, t\}$ and every affine map $\pi : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$. Note further that we can assume $t \leq k$ in the specification above. The following lemma gives several equivalent definitions to being a k-single orbit characterizable family. The lemma can be seen as an extension of Lemma 3.6 in [8].

[2] As $d' \in \mathrm{Shift}(d)$ if and only if $d \in \mathrm{Shift}(d')$, such S always exists.

Lemma 2. *[Equivalent definitions of k-single orbit characterizable family] Let $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ be a linear affine-invariant family. The following are equivalent:*

1. *$(\bar{\alpha}; \{\bar{\lambda}_i\}_{i=1}^t)$ is a k-single orbit characterization of $\mathcal{F}$, where $\bar{\alpha} = (\alpha_1, \ldots, \alpha_k) \in \mathbb{F}_{q^n}^k$ and $\bar{\lambda}_i = (\lambda_{i,1}, \ldots, \lambda_{i,k}) \in \mathbb{F}_q^k$.*
2. *For all d, $d \in \mathrm{Deg}(\mathcal{F}) \Leftrightarrow \forall i \sum_{j=1}^k \lambda_{i,j}(\alpha_j x + y)^d \equiv 0$ (as a formal polynomial in x and y).*
3. *For all d,$d \in \mathrm{Deg}(\mathcal{F}) \Leftrightarrow \forall e \leq_p d, \forall i \sum_{j=1}^k \lambda_{i,j}\alpha_j^e = 0$.*
4. *For all $d \in \mathrm{Deg}(\mathcal{F}) \cup \mathrm{Border}(\mathcal{F})$, $d \in \mathrm{Deg}(\mathcal{F}) \Leftrightarrow \forall i \sum_{j=1}^k \lambda_{i,j}\alpha_j^d = 0$.*

3.1 Proof of Main Theorem

Due to space limitations we give only an idea of the proof of Theorem 1.

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be k-single orbit characterized properties (we assume for simplicity that $k_1 = k_2 = k$, and make several other simplifying assumptions here). Suppose the k-single orbit characterization of $\mathcal{F}_1$ is simply $(\bar{\alpha}, \bar{1})$ (so $\mathcal{F}_1$ satisfies $\sum_{i=1}^k f(\alpha_i) = 0$). Similarly, let the k-single orbit characterization of $\mathcal{F}_2$ be $(\bar{\beta}, \bar{1})$.

A candidate k^2-single orbit characterization of $\mathcal{F}_1 + \mathcal{F}_2$ would be the the "outer product" of the two given constraints, namely the k^2 local constraint given by $((\alpha_i\beta_j)_{i,j}; \bar{1})$.

To analyze this potential constraint, we look at the degree set based descriptions of single-orbit characterizations. First we use the (easily verifiable fact) that $\mathrm{Deg}(\mathcal{F}_1 + \mathcal{F}_2) = \mathrm{Deg}(\mathcal{F}_1) \cup \mathrm{Deg}(\mathcal{F}_2)$. Next we see that for every $d \in \mathrm{Deg}(\mathcal{F}_1) \cup \mathrm{Deg}(\mathcal{F}_2)$, $\sum_{i=1}^k \sum_{j=1}^k \alpha_i^d \beta_j^d = (\sum_{i=1}^k \alpha_i^d) \cdot (\sum_{j=1}^k \beta_j^d) = 0$, so $((\alpha_i\beta_j)_{i,j}; \bar{1})$ is a valid constraint on $\mathcal{F}_1 + \mathcal{F}_2$.

Unfortunately, it is not clear that for every $d \in \mathrm{Border}(\mathcal{F}_1 + \mathcal{F}_2)$ the sum $\sum_{i=1}^k \sum_{j=1}^k \alpha_i^d \beta_j^d$ does not equal 0, which is necessary (by Part 4 of Lemma 2).

To remedy this, we take a random constraint in the orbit of $(\bar{\alpha}; 1)$ and a random constraint in the orbit of $(\bar{\beta}; \bar{1})$ and take their "outer product". Specifically we pick random non-zero $a_1, a_2 \in \mathbb{F}_{q^n}$ and random $b_1, b_2 \in \mathbb{F}_{q^n}$ and consider the potential constraint $(\bar{\gamma}; \bar{1})$ where $\bar{\gamma} = (\gamma_{i,j})_{i,j}$ is given by $\gamma_{i,j} = (a_1\alpha_i + b_1)(a_2\beta_j + b_2)$. It is again easy to verify that $\sum_{i,j} \gamma_{i,j}^d = 0$ for every $d \in \mathrm{Deg}(\mathcal{F}_1) \cup \mathrm{Deg}(\mathcal{F}_2)$.

We then note that for any fixed $d \notin \mathrm{Deg}(\mathcal{F}_1) \cup \mathrm{Deg}(\mathcal{F}_2)$ the formal sum $\sum_{i,j}((x_1\alpha_i + y_1)(x_2\beta_j + y_2))^d \not\equiv 0$ (as a polynomial in x_1, x_2, y_1, y_2 — this uses Part 2 of Lemma 2). Thus when we pick random assignments $x_1 = a_1$, $x_2 = a_2$ etc,, we find that $\sum_{i,j} \gamma_{i,j}^d \neq 0$ with probability at least $1 - O(d/q^n)$, and so this random choice does eliminate any particular "bad" d.

To conclude the argument we need to make sure that every $d \in \mathrm{Border}(\mathcal{F}_1 + \mathcal{F}_2)$ is eliminated (i.e., $\sum_{i,j} \gamma_{i,j}^d \neq 0$). To do so, we use the union bound, with two crucial ingredients: First we use the main theorem from [9] to conclude that all $d \in \mathrm{Border}(\mathcal{F}_1 + \mathcal{F}_2)$ have p-weight at most k and there are only $(q+n)^{O(k)}$ such d's. Next we use the fact that we need to consider only one d from every

"shift" class, to take the smallest one. This allows us to work with $d \leq q^{n(1-1/k)}$. Combining these ingredients, we can take the union bound over all possible bad events and conclude that a random choice a_1, a_2, b_1, b_2 eliminates every $d \in \text{Border}(\mathcal{F}_1 + \mathcal{F}_2)$ with positive probability.

4 Consequences, Questions and Conjectures

Our work further highlights the role played by single-orbit characterizations in the testing of affine-invariant properties. This feature is common (e.g. Reed-Muller property is single-orbit over the smaller group) and also useful (sums of single-orbit characterized properties also have this feature). In this section we describe some of the questions surrounding this concept that emerge from this (and related) research.

At the moment almost all known locally-testable affine-invariant properties are known to be single-orbit characterized. The only exception is the case of sparse properties where the range is not a prime field. This leads to the following question, which we hope can be resolved affirmatively (soon).

Question 1. For every q and t, does there exists a constant $k = k(q,t)$ such that every t-sparse property $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ is k-single orbit characterized?

Assuming an affirmative answer to the questions above, we get a "concise" description of all known testable properties.

4.1 Known locally testable properties

As mentioned earlier, the known "basic" single-orbit characterized affine-invariant families are the Reed-Muller families and sparse families. Three "operations" are also now known that preserve "single-orbit characterizations" and hence local testability of these basic families: (1) Sums of two families, (2) Intersections of two families, and (3) Lift of a single family. Below we define this lifting operator.

Definition 5 (Lifted code [8]). *Let $\mathbb{K} \supsetneq \mathbb{L} \supsetneq \mathbb{F}_q$ be finite fields with $q = p^s$. For $D \subseteq \{0, \ldots, |\mathbb{L}| - 1\}$ we define the* lift *of D from $\mathbb{L}$ to $\mathbb{K}$ to be the set of integers* $\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(D) = \{d' \in \{0, \ldots, |\mathbb{K}| - 1\} \mid (\text{shadow}_p(d') \, (mod^* \, |\mathbb{L}| - 1)) \subseteq D\}$.

For an affine-invariant family $\mathcal{F} \subseteq \{\mathbb{L} \to \mathbb{F}_q\}$ with degree set $D = \text{Deg}(\mathcal{F})$, let $\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(\mathcal{F})$ be the affine-invariant family with degree set $\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(D)$, i.e., $\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(\mathcal{F}) = \{f : \mathbb{K} \to \mathbb{F}_q \mid \text{supp}(f) \subseteq \text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(D)\} = \text{Fam}(\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(D))$.

The following proposition follows easily from the definitions

Proposition 1 ([8]). *Lifts of single orbit characterized families are also single-orbit characterized. Specifically, if $\mathbb{F}_q \subseteq \mathbb{L} \subseteq \mathbb{K}$ and $(\bar{\alpha}, \{\bar{\lambda}_i\}_{i=1}^t)$ is a k-single orbit characterization of $\mathcal{F} \subseteq \{\mathbb{L} \to \mathbb{F}_q\}$ then $(\bar{\alpha}, \{\bar{\lambda}_i\}_{i=1}^t)$ is also k-single orbit characterization of* $\text{lift}_{\mathbb{L} \nearrow \mathbb{K}}(\mathcal{F})$.

Given the operations above, it is easy to see that one can compose a finite number of basic single-orbit characterized families using a "formula" whose operations are sum, intersection and lifts. We define this concept below.

Definition 6 (Formula, size). *A formula Φ of size s, degree d, sparsity t producing a family $\mathcal{F} \subseteq \{\mathbb{K} \to \mathbb{F}_q\}$, denoted $(s, d, t, \mathbb{K}, \mathbb{F})$-formula, is given by the following inductive definition:*

1. *A formula Φ of size 1, is given by $\mathcal{F} \subseteq \{\mathbb{K} \to \mathbb{F}_q\}$ where $\mathcal{F}$ is either a Reed-Muller family of order d, or a t-sparse family.*
2. *A formula of size s is obtained by one of the following operations:*
   (a) *Picking $\mathbb{L}$ such that $\mathbb{F}_q \subseteq \mathbb{L} \subseteq \mathbb{K}$ and letting $\Phi = \mathrm{lift}_{\mathbb{L} \nearrow \mathbb{K}}(\Phi_1)$ where Φ_1 is a $(s-1, t, d, \mathbb{L}, \mathbb{F})$ formula.*
   (b) *Picking s_1, s_2 such that $s_1 + s_2 + 1 = s$ and letting $\Phi = \Phi_1 \cap \Phi_2$ where Φ_i is an $(s_i, t, d, \mathbb{K}, \mathbb{F})$ formula.*
   (c) *Picking s_1, s_2 such that $s_1 + s_2 + 1 = s$ and letting $\Phi = \Phi_1 + \Phi_2$ where Φ_i is an $(s_i, t, d, \mathbb{K}, \mathbb{F})$ formula.*

The following theorem summarizes the state of knowledge of single-orbit characterized families.

Theorem 3. *For every s, t, d, q there exists a $k = k(s, t, d, q)$ such that for every n, every $(s, t, d, \mathbb{F}_{q^n}, \mathbb{F}_q)$-formula produces a k-single orbit characterized family, for prime q.*

Note that the caveat that q is prime can be dropped if we have an affirmative answer to Question 1.

4.2 Conjectures and questions

We start with the most obvious question.

Question 2. Is the following statement true? For every k, q there exist s, t, d such that for every n, if $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ is a k-locally testable affine-invariant family then $\mathcal{F}$ is given by an $(s, t, d, \mathbb{F}_{q^n}, \mathbb{F}_q)$-formula.

At the moment our understanding of affine-invariance with respect to its local testability is so far that it is too optimistic to conjecture an affirmative answer to this question. All we can say is that an affirmative answer is not yet ruled out.

The nature of the question seems to become much simpler if we disallow lifts, by insisting that n is prime (then we get no fields $\mathbb{L}$ strictly between $\mathbb{F}_q$ and $\mathbb{F}_{q^n}$). In this setting, intersections become uninteresting and lead to a much tamer question.

Question 3. Is the following statement true? For every k, q there exist t, d such that for every prime n, if $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ is a k-locally testable affine-invariant family then $\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2$ where $\mathcal{F}_1 = \mathrm{RM}_q(d', n)$ and $\mathcal{F}_2$ is t'-sparse, for some $d' \leq d$ and $t' \leq t$.

This question remains quite challenging even when we restrict to the case where $q = 2$ (where our state of understanding does seem somewhat better), and even when we restrict our families to be contained in $\mathrm{RM}_2(2, n)$.

Conjecture 1. For every k there exists a t such that the following holds for every prime n: If $\mathcal{F} \subsetneq \mathrm{RM}_2(2, n)$ is k-locally testable then $\mathcal{F}$ is t-sparse.

Attempting to prove the conjecture above leads to some interesting questions about the rank of certain Vandermonde like matrices that seem interesting in their own right. We state the conjecture below. We don't prove the connection to the conjecture above, but claim that an affirmative answer to the following implies an affirmative answer to the above.

Conjecture 2. For every k, there exists a t such that for every prime n and every sequence $\alpha_1, \ldots, \alpha_k \in \mathbb{F}_{2^n}$ of elements that are $\mathbb{F}_2$-linearly independent, and every sequence of t distinct elements $e_1, \ldots, e_t \in \{0, \ldots, n-1\}$, the $k \times t$ matrix $M = [M_{ij}]_{ij}$ with $M_{ij} = \alpha_i^{2^{e_j}}$ has rank exactly k.

Finally a couple of questions which relate to the structure of locally-testable codes (an affirmative answer to both is implied by an affirmative answer to Question 2).

Question 4. For every k, q does there exist a $\tilde{k}$ such that for every n, if $\mathcal{F} \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ is k-locally testable, then $\mathcal{F}$ has a $\tilde{k}$-single orbit characterization?

Question 5. For every k, q does there exist a $\tilde{k}$ such that for every n, if $\mathcal{F}_1, \mathcal{F}_2 \subseteq \{\mathbb{F}_{q^n} \to \mathbb{F}_q\}$ are k-locally testable, then $\mathcal{F}_1 + \mathcal{F}_2$ is $\tilde{k}$-locally testable?

Acknowledgments

We would like to thank Shripad Garge, Neeraj Kayal and Dipendra Prasad for valuable pointers.

References

1. Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 251–260. ACM, 2006.
2. Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.
3. Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, May 1998.
4. Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

5. Boaz Barak, Zeev Dvir, Avi Wigderson, and Amir Yehudayoff. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:149, 2010. To appear in STOC 2011.
6. Omer Barkol, Yuval Ishai, and Enav Weinreb. On locally decodable codes, self-correctable codes, and t-private PIR. *Algorithmica*, 58:831–859, 2010.
7. Eli Ben-Sasson, Elena Grigorescu, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. On sums of locally testable affine invariant properties. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:79, 2011.
8. Eli Ben-Sasson, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. Symmetric LDPC codes are not necessarily locally testable. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:199, 2010. To appear in CCC 2011.
9. Eli Ben-Sasson and Madhu Sudan. Limits on the rate of locally testable affine-invariant codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:108, 2010.
10. Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 261–270. ACM, 2006.
11. Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, 1998.
12. Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.
13. Elena Grigorescu, Tali Kaufman, and Madhu Sudan. 2-transitivity is insufficient for local testability. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC 2008, 23-26 June 2008, College Park, Maryland, USA*, pages 259–267. IEEE Computer Society, 2008.
14. Elena Grigorescu, Tali Kaufman, and Madhu Sudan. Succinct representation of codes with applications to testing. In Irit Dinur, Klaus Jansen, Joseph Naor, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 5687 of *Lecture Notes in Computer Science*, pages 534–547. Springer, 2009.
15. Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. *Random Struct. Algorithms*, 35(2):163–193, 2009.
16. Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 80–86, New York, NY, USA, 2000. ACM.
17. T. Kaufman and D. Ron. Testing polynomials over general fields. *SIAM J. on Computing*, 36(3):779–802, 2006.
18. Tali Kaufman and Shachar Lovett. Testing of exponentially large codes, by a new extension to Weil bound for character sums. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:65, 2010.
19. Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 403–412. ACM, 2008.
20. Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Computing*, 25(2):252–271, 1996.
21. Madhu Sudan. Invariance in property testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:51, 2010.