Electronic Colloquium on Computational Complexity, Report No. 72 (2020)

Locally testable codes via high-dimensional expanders*

Yotam Dikstein[†] Irit Dinur[‡] Prahladh Harsha[§] Noga Ron-Zewi[¶]

May 5, 2020

Abstract

Locally testable codes (LTC) are error-correcting codes that have a local tester which can distinguish valid codewords from words that are far from all codewords, by probing a given word only at a very small (sublinear, typically constant) number of locations. Such codes form the combinatorial backbone of PCPs. A major open problem is whether there exist LTCs with positive rate, constant relative distance and testable with a constant number of queries.

In this paper, we present a new approach towards constructing such LTCs using the machinery of high-dimensional expanders. To this end, we consider the Tanner representation of a code, which is specified by a graph and a base code. Informally, our result states that if this graph is part of an *agreement expander* then the local testability of the code follows from the local testability of the base code. Agreement expanders allow one to stitch together many mostly-consistent local functions into a single global function. High-dimensional expanders are known to yield agreement expanders with constant degree.

This work unifies and generalizes the known results on testability of the Hadamard, Reed-Muller and lifted codes, all of which are proved via a single round of local self-correction: the corrected value at a vertex v depends on the values of all vertices that share a constraint with v. In the above codes this set includes all of the vertices. In contrast, in our setting the degree of a vertex might be a constant, so we cannot hope for one-round self-correction. We overcome this technical hurdle by performing iterative self correction with logarithmically many rounds and tightly controlling the error in each iteration using properties of the agreement expander.

Given this result, the missing ingredient towards constructing a constant-query LTC with positive rate and constant relative distance is an instantiation of a base code and a constant-degree agreement expander that interact well with each other.

1 Introduction

In this work, we study an approach to constructing locally testable codes (LTCs) based on high-dimensional expansion. LTCs are error-correcting codes that have a local tester which can test if a given word is a valid

[*] Part of this work was done when the authors were visiting the Simons Institute of Theory of Computing, Berkeley for the summer cluster on "Error-Correcting Codes and High-Dimensional Expansion".

[†] Weizmann Institute, yotam.dikstein@weizmann.ac.il

[‡] Weizmann and IAS, irit.dinur@weizmann.ac.il. Research supported by ERC CoG grant 772839 and the National Science Foundation under agreement No. CCF-1900460.

[§] Tata Institute of Fundamental Research, Mumbai, India. prahladh@tifr.res.in. Research supported by the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0500 and in part by the Swarnajayanti fellowship.

[¶] Haifa University, noga@cs.haifa.ac.il

ISSN 1433-8092

codeword or far (in Hamming distance) from all codewords, by probing the given word only at a very small (sublinear, typically constant) number of locations. Reed-Muller codes were the first codes shown to be locally-testable [FS95, RS96]. These codes are based on low degree polynomial functions, and have inverse polynomial rate. Later on, LTCs with inverse poly-logarithmic rate were constructed by [BS08, Din07]. Obtaining an LTC family with rate that is not vanishing is a major open question in this area. Such codes are known as "good" LTCs or c^3-LTCs since they have **c**onstant rate, **c**onstant relative distance, and testable with a **c**onstant number of queries [Gol10]. This question is interesting in its own right, and also could potentially lead towards constructing linear-length PCPs (as LTCs are the combinatorial backbone of all PCP constructions). The problem of constructing c^3-LTCs is particularly difficult as we do not know if such good codes exist, even non-explicitly (say using a probabilistic argument). The difficulty stems from the fact that local testability requires redundancy in the constraints. In known LTCs, the constraints are highly overlapping, a property that in the past went hand in hand with relatively dense families of constraints. Alas this density seems to significantly limit the rate. In contrast, high-dimensional expanders give sparse families of subsets that are heavily overlapping. Perhaps if we manage to find appropriate constraints on these subsets we may find higher rate LTCs.

In this work, the vague notion of "overlapping constraints" is captured through so-called agreement-expansion (which will be formally defined below).

Informally speaking, we show that if an error-correcting code is defined through a collection of local constraints that *sit on an agreement expander*, then to prove local testability of the entire code it suffices to prove local testability of the local components (which are of merely constant size in the case of constant-degree agreement expanders). This is similar in spirit to recent applications of high-dimensional expanders towards proving other local-to-global results. This passing from local to global is particularly important because known constructions of high-dimensional expanders are very difficult to analyze on a global level. So far, successful analyses focused on the local structure (in neighborhoods, or so-called links) of these objects. Through this work, the task of constructing global LTCs is reduced to the task of constructing LTCs on the local structure, which appears to be a much more reasonable task.

This work can be viewed as providing a generic scheme for constructing an LTC on a high-dimensional expander (or an agreement expander), and the (big) missing ingredient is an appropriate instantiation. We comment that the flagship example of an LTC, namely Reed-Muller codes, can be viewed as an instantiation of this scheme, with the underlying agreement expander being the Grassmannian complex and the base code being the Reed-Solomon code (see Section 5). The hope is that replacing the "dense" Grassmannian complex by a bounded-degree complex, together with finding an appropriate base code, could potentially lead to a c^3 LTC.

Tanner Codes. To elucidate the main result, we begin by recalling a well-studied family of codes, the *Tanner codes* [Gal60, Tan81]. A Tanner code $C \subseteq \{0,1\}^n$ is given by a family of (small, often constant-sized) subsets $t_1, \dots, t_m \subset [n]$ and for each subset a base code $C_{t_i} \subset \{0,1\}^{t_i}$. A string $w \in \{0,1\}^n$ is in the code C if for each i, $w|_{t_i} \in C_{t_i}$.[1] Many known codes, including Reed-Muller codes, lifted codes, tensor codes, and expander codes, are in fact Tanner codes. In all of these cases, there is a single base code C_0 such that $C_{t_i} = C_0$ for all i, but this need not be the case.

The Tanner representation of a code also gives a natural candidate for a local test for checking whether a given word $w \in \{0,1\}^n$ is in the code.

Natural Tanner Test: Choose a random $i \in [m]$ and accept iff $w|_{t_i} \in C_{t_i}$.

[1] A Tanner code is equivalently described on a bipartite graph (called the Tanner graph) with n right vertices corresponding to the coordinates of the code and m left vertices corresponding to the sets t_i, with an edge between v and t_i if $v \in t_i$.

We say that C is ρ-locally-testable with the natural tester if

$$\rho \cdot \mathrm{dist}(w, C) \leqslant \mathbb{P}\left[\text{Test fails}\right].$$

A family of codes is a locally testable code (LTC) if it satisfies the above inequality for some test (not necessarily the natural Tanner test) with a constant ρ (that does not decrease with the block length of the code).

Many Tanner codes, including expander codes and random LDPC codes, that are very good in terms of rate and distance, *and* can be characterized by "low density" constraints (that look at only a constant number of bits in the codeword) fail quite miserably at being LTCs [BHR05].

Imagine that in addition to $T = \{t_1, \dots, t_m\}$ we also have a family S of subsets of $[n]$, such that each $s \in S$ has constant size, but slightly larger than the size of the t_i's. For each such $s \in S$ we consider the 'local' Tanner code

$$C_s = \{w \in \{0,1\}^s \mid \; w|_t \in C_t, \forall t \in T, \; t \subset s\}.$$

(Of course, C_s is non-trivial only if there are some $t \in T$ contained in s.)

In this work, we show that if for each $s \in S$, the code C_s itself is locally testable with the natural Tanner test, then the code C too must be locally testable with respect to the natural Tanner test. This holds as long as we assume some nice structure on the families S and T, namely that they are part of a "multi-layered agreement sampler", MAS for short, which is described below.

Let us change point of view and look at the codes $\{C_s\}$ as a collection of base codes, giving rise to the Tanner code C. Our main result is that local testability of the base codes C_s *lifts* to local testability of the entire code C, assuming an expander-like MAS condition on the underlying Tanner graph. This is analogous to the celebrated expander codes [SS96] in which distance of the base codes gets lifted to distance of the entire code, assuming expansion of the underlying Tanner graph. Whereas expansion alone does not suffice for local testability, the MAS structure does.

High-dimensional expanders and Agreement Expanders. There are several interesting and non-equivalent definitions for high-dimensional expanders, the two main ones being topological definitions of coboundary or cosystolic expansion [LM06, Gro10, DKW18], and, more relevant to this work, random walk definitions either locally at the link level [KM17, DK17] or globally [DDFH18, KO18]. Without going into details, high-dimensional expansion has already been shown to imply some surprising local to global theorems. For example the trickling down theorem of [Opp18] proves global spectral expansion using local spectral expansion in the links (which are the neighborhoods of individual vertices). Another example is the list decoding of [DHKNT19] which deduces global list decoding from list-decoding on the local pieces.

Yet another example, which is crucial for this work, is that high-dimensional expanders give rise to agreement expanders [DK17, DD19]. An agreement expander allows one to stitch together many mostly-consistent local functions into a single global function. We elaborate a little more on this notion. Let V be a ground set of n elements, and let S be a collection of subsets of V of some fixed size. Let $\mathcal{A}$ be a graph whose vertices are the subsets in S, and each edge $\{s, s'\}$ is labeled by a subset $k \subset s \cap s'$. Let K be a collection of subsets labelling the edges.

$(V, K, S, \mathcal{A})$ is an α-agreement expander if whenever an ensemble has agreement value $1 - \varepsilon$ there exists a global function $F \colon V \to \{0,1\}$ such that $f_s = F|_s$ for all but at most ε/α of $s \in S$. (See Section 2.5 for the full definition). An agreement expander is given by V, K, S and the edge-labelled graph $\mathcal{A}$. Suppose that for each $s \in S$ we are given a local function $f_s \in \{0,1\}^s$. The *agreement value* of the ensemble $\{f_s\}$ is the probability of $f_s|_k = f_{s'}|_k$ for a randomly chosen edge $\{s, s'\}_k$ (this is notation for an edge between s, s'

labeled by k) in the graph $\mathcal{A}$. Whenever there is a global function $F\colon V \to \{0,1\}$ such that $f_s = F|_s$ for all $s \in S$, the agreement value of $\{f_s\}$ is clearly 1. We say that

Agreement expanders have been studied and used in the LTC and PCP literature for years (under different names such as direct product tests or sometimes low degree tests). However, prior to the recent connection with high-dimensional expanders, the only known agreement expanders were relatively dense. The existence of sparse such objects seems promising and could potentially lead to LTCs with positive rate. This work shows how agreement expansion can be useful for constructing LTCs.

Multilayered Agreement Samplers (MAS). We now describe the MAS combinatorial structure needed for our LTC scheme. Let V be a ground set of n elements, and let T, K, S be three families of subsets of V of sizes $q_0 < q_1 < q_2$. The system (V, T, K, S) is said to be a (λ, α) *-MAS* if the following two conditions are met.

- V, K, S are part of an α-agreement-expander.
- The bipartite containment graph of T vs. K is a λ-sampler.

The above definition is stricter than what we actually need, see the formal more refined definition in Definition 3.1. We are now ready to state our main result.

Main Result. Let V, T, K, S be a (λ, α) *-MAS*. Suppose that for each $t \in T$ we have a local code $C_t \subset \{0,1\}^t$. Let $C \subset \{0,1\}^n$ be the Tanner code defined by $\{C_t\}$ for all $t \in T$. Namely,

$$C := \left\{ w \in \{0,1\}^V \;\middle|\; w|_t \in C_t \text{ for every } t \right\}.$$

Similarly, for each $s \in S$, let C_s be the Tanner code defined by $\{C_t \mid \; t \subset s\}$, namely,

$$C_s = \{ w \in \{0,1\}^s \mid \; w|_t \in C_t \text{ for every } t \subset s \},$$

and similarly define for each $k \in K$, $C_k = \left\{ w \in \{0,1\}^k \mid \; w|_t \in C_t \text{ for every } t \subset k \right\}$.

Theorem 1.1. *Let V, T, S be layers in a (λ, α)-MAS satisfying $\lambda \leqslant \rho\delta\alpha/64$. Suppose $C_k \subset \{0,1\}^k$ has relative distance δ for all $k \in K$ and suppose that C_s is ρ-locally testable with the natural Tanner tester. Then C is $\rho\delta\alpha/16$ locally testable (with the natural Tanner tester).*

We state our full main theorem in Theorem 4.1.

Overview of proof. Our proof of local testability, like previous proofs of testability, goes via self correction. The main difficulty in our setting is that a single round of self-correction is insufficient to correct the word.

Let w be a word that satisfies a $(1-\varepsilon)$-fraction of the constraints in the Tanner graph. We would like to show that there exists a $w^* \in C$ such that $\operatorname{dist}(w, w^*) = O(\varepsilon)$. For specific codes, one could use the properties of the code to perform this self-correction (cf. Reed-Muller testing, one could use the properties of polynomials). However, we cannot resort to such properties since we are working in an abstract setting. Instead, we rely on simple majority decoding: each vertex takes a value that satisfies the majority of the constraints it participates in. The main engine driving our proof is agreement expansion. Our proof strategy is as follows:

Construct a word w' from the received word w via self correction (or otherwise) and show

(a) w is close to w', and

(b) w' is a valid codeword.

Property (a) is easy to show if w' is constructed via self correction using majority decoding. Property (b) is not very hard in the context of Hadamard testing and Reed-Muller testing: every vertex participates in a constraint with every other vertex (indeed the diameter of the Tanner graph is a constant), hence one round of self-correction results in a valid codeword w'. However, since our proof is general enough to work even for constant-degree Tanner graphs wherein the diameter can be as large as logarithmic, one does not expect a single step of self correction via majority decoding to yield a codeword in a single step.

Our proof instead relies on a novel iterative self correction procedure that slowly corrects a given word in logarithmically many iterations. A standard problem that arises when using iterative procedures is that the error grows linearly in the number of iterations, which is prohibitively expensive in our setting. We use the properties of MAS to show that the number of unsatisfied constraints by the self-corrected word w' reduces by a constant factor in each iteration. This allows us to perform an arbitrary number of rounds in the iterative self-correction procedure till we reach a perfect codeword $w^* \in C$ (actually a logarithmic number of rounds will suffice). This type of argument is new in the context of locally testable codes.

Given this we can proceed with the proof overview as follows. Since w satisfies $(1-\varepsilon)$-fraction of the constraints, an averaging argument shows that a $(1-O(\varepsilon))$-fraction of the s's satisfy most of the constraints within them. Hence, by the local testability of the code C_s we get that for most s's, $w|_s$ is close to a local codeword, say $w_s \in C_s$. Furthermore, it is not hard to show that these local codewords satisfy that for a typical $k \in K$ and $s, s' \in S$ such that $k \subset s \cap s'$, we have $w_s|_k \equiv w_{s'}|_k$. In other words, the w_s's satisfy the hypothesis of the agreement test. From the agreement expansion of the MAS, there exists a "global" word w' that explains most of the w_s's. Furthermore, it is not hard to show that w' is close to the original word w. We then use the sampler property of the MAS to show that w' violates *significantly fewer* constraints than w (in particular, w' violates at most $\varepsilon/2$-fraction of constraints).

We iteratively apply the above self-correction procedure to get a sequence of words such that $w^{(0)} := w, w^{(1)}, w^{(2)}, \ldots$ such that $w^{(i)}$ violates at most $\varepsilon/2^i$-fraction of constraints and $\mathrm{dist}(w^{(i)}, w^{(i+1)}) = O(\varepsilon)/2^i$. Since the fraction of violated constraints cannot infinitely decrease, we have that eventually for a large enough i, $w^* := w^{(i)} \in C$ and $\mathrm{dist}(w, w^*) \leqslant \sum_{j=0}^{i-1} dist(w^{(j)}, w^{(j+1)}) = O(\varepsilon)$.

Relation to previous work. We begin by recalling the history of LTCs and the close connection between PCP and LTC constructions. LTCs were first studied in the context of program checking by Blum, Luby and Rubinfeld [BLR93] and Gemmell *et al.* [GLRSW91]. The notion of LTCs is implicit in the work on locally checkable proofs by Babai et al. [BFLS91] and subsequent works on PCPs. The explicit definition appeared independently in the works of Rubinfeld and Sudan [RS96], Friedl and Sudan [FS95], Arora's PhD thesis [Aro94] and Spielman's PhD thesis [Spi95]. A formal study of LTCs was initiated by Goldreich and Sudan [GS06]. Most known constructions of PCPs yield LTCs with similar parameters. In fact, there is a generic transformation to convert a PCP of proximity (which is a PCP with more requirements) into an LTC with comparable parameters [BGHSV06, Tre04]. See a survey by Goldreich [Gol10] for the interplay between PCP and LTC constructions. In fact, the current best construction of LTCs (constant-query, constant fractional distance and inverse polylogarithmic rate) is obtained from the PCP constructions of Ben-Sasson and Sudan [BS08] and Dinur [Din07]. PCP-based constructions are unlikely to yield LTCs with constant rate since PCP constructions typically involve at least a logarithmic overhead. Nevertheless LTC constructions that aren't derived from PCPs perhaps have a better chance at achieving the coding-theory gold-standard of positive rate and distance.

Agreement expansion and the multilayered set system structure play a central role in our proof of local testability. Another application of agreement expansion towards local testability was studied in [DHKRZ19], where it was used to enhance the local testability of a code in the context of the subspaces (Grassmannian) complex. We remark that use of such multilayered agreement samplers in the context of locally-testable codes is actually implicit in many previous constructions of locally testable codes. The Raz-Safra [RS97] proof of the local testability of the Reed-Muller codes works with points-lines-planes structure, a subgraph of the Grassmannian complex which is an excellent agreement expander as explained in detail in Section 5. The original proof due to Blum, Luby and Rubinfeld [BLR93] (as well as subsequent improvements due to Coppersmith) of the local testability of the Hadamard codes as well as Kaufman and Sudan's proof of testability of affine-invariant codes [KS08], relies on the three-layered structure comprising of the points, the three-point tests and certain nine-point sets, sometimes referred to as "magic squares" [KS08].

Our proof makes explicit this use of MAS to construct LTCs and shows that four-layered MAS are sufficient to transform “local” local testability to “global” local testability. In this sense, our proof can be viewed as bringing together these seemingly different proofs of local-testability under a common umbrella.

We already remarked that our construction has a similar paradigm as the Sipser-Spielman construction of expander codes [SS96] which demonstrates that if the base code has good distance then the Tanner code also has good distance provided the graph is an expander. Another construction of the same flavor is the result of Dinur et al. [DHKNT19] that demonstrates that if the local code is efficiently list-decodable then so is the global code defined by ABNNR distance amplification property via an expander [ABNNR92], provided the expander is part of a large high-dimensional expander.

Further Discussion and Future Work. This work gives a general scheme for constructing an LTC. It needs to be instanciated with an appropriate MAS and base codes. As mentioned earlier, and explained in detail in Section 5, one such instanciation is to choose the Grassmannian complex as the MAS, and the Reed Solomon code as the base codes. This gives the well-studied locally testable codes called Reed-Muller codes, as well as the more recent so-called lifted codes.

The most interesting direction is to instantiate this scheme with an MAS that comes from some bounded-degree high-dimensional expander, and to combine it with appropriate choice of locally testable base code. The main hurdle in choosing the base codes is to be able to certify that the resulting Tanner code maintains positive rate. In some similar situations this is done by a simple counting of the number of constraints. However, such an argument cannot work in the setting of LTCs, and we leave it as an open question.

2 Preliminaries

2.1 Error Correcting Codes

Let Σ be some finite set. A code is some $C \subseteq \Sigma^n$. Let p be a prime power and $\Sigma = \mathbb{F}_p^n$ be an n-dimensional vector space over a field with q elements. We say that C is a *linear code* when C is a subspace of $\mathbb{F}_p^n$. The rate of the code is $rate(C) = \frac{\log_q |C|}{n}$.

It is convenient to think about $\mathbb{F}_p^n$ as functions $f : [n] \to \mathbb{F}_p$. The distance between two functions $f, g : [n] \to \Sigma$, denoted by $\operatorname{dist}(f, g)$, is the fraction of $x \in [n]$ so that $f(x) \neq g(x)$. The distance of a code is defined to be $\operatorname{dist}(C) = \min_{f,g\in C, f\neq g} \operatorname{dist}(f, g)$. When C is linear, this is the same as $\min_{0\neq f\in C} \operatorname{dist}(f, 0)$.

2.2 Tanner Codes

A Tanner code [Gal60, Tan81] over an alphabet Σ (also called a lifted code) is defined through two objects: a family T of q-element subsets of $[n]$, and with each subset $t \in T$ a base code $C_t \subset \Sigma^t$. The code $C \subseteq \Sigma^n$ is given by

$$C = \{w \in \Sigma^n \mid \ w|_t \in C_t, t \in T\}.$$

The family T is often described through a bipartite graph on vertex sets $[n]$ and T connecting $t \in T$ to $i \in [n]$ whenever $i \in t$. Several well-known families of codes can be constructed as Tanner codes, including tensor codes, Reed-Muller codes, and the codes considered by Sipser and Spielman [SS96]. A family of Tanner codes that is especially related to our context is the family of so-called lifted codes. Lifted codes were first introduced by Ben-Sasson, Maatouk, Shpilka and Sudan [BMSS11] and their local testability was studied by Guo, Kopparty and Sudan [GKS13]. These codes can be described as Tanner codes where $[n]$ is identified with points of a vector space and the family T contains all possible affine subspaces of a prescribed dimension m. The base code C_0 is taken to be affine invariant. A prime example for such codes is the Reed-Muller code.

2.3 Locally Testable Codes

A (Q, ρ)-local tester for the code C is a probabilistic oracle algorithm that determines whether a word is in the code. It does the following: given oracle access to a function $f : [n] \to \Sigma$, it queries f at Q input locations. Then if $f \in C$ it accepts with probability 1. If $f \notin C$ it rejects with probability at least $\rho \cdot dist(f, C)$. Here $\rho \in (0, 1)$ is some constant parameter, and $dist(f, C)$ is the distance between f and the closest codeword to it in C.

For linear codes C, [BHR05] showed that without loss of generality, we can assume that the local testers picks a random subset $t \subset [n]$ according to some distribution, and accept if and only if $w|_t \in C_t$ (that is, that there exists some codeword $w' \in C$ so that $w|_t = w'|_t$). Thus we formally define the locally testable codes as following:

Definition 2.1 (Locally Testable Codes)**.** Let V some finite set, and C be some linear code on V. Let D be some distribution on subsets of V, and suppose every set $t \sim D$ of of size at most Q. Let $\rho > 0$. We say C is (Q, ρ)-testable with respect to D if

$$\rho \cdot \mathrm{dist}(f, C) \leqslant \mathop{\mathbb{P}}_{t \sim D}\left[f|_t \notin C|_t\right].$$

An alternate way of describing a locally testable code is using the Tanner graph $G = ([n], T, E)$ representation of a code. In this representation, $[n]$ corresponds to the n input locations in the codeword. T corresponds to the subsets of indexes that are queried by the local tester. We connect $i \in [n]$ and $t \in T$ if $i \in t$.

A local tester that corresponds to this representation picks a random constraint $t \in T$ and checks if the corresponding constraint is satisfied.

2.4 Sampler Graphs

Let $G = (U, V, E)$ be a bipartite graph, and assume that each edge carries a non-negative weight p_{uv} such that $\sum_{uv \in E} p_{uv} = 1$. This probability distribution induces a marginal probability distribution on U and similarly on V given by $p_u = \sum_{uv \in E} p_{uv}$. For every set $B \subseteq U$ (and V respectively) we denote by $\mathbb{P}[B] = \mathbb{P}_{u \in U}[u \in B]$. As a slight abuse of notation, for a set $B \subseteq U$ and a vertex $v_0 \in V$ we denote by

$$\mathbb{P}[B \mid \ v_0] = \mathop{\mathbb{P}}_{uv \in E}[u \in B \mid \ v = v_0].$$

A sampler graph is a graph where for all $B \subseteq U$, most of the vertices $v_0 \in V$ have that $\mathbb{P}[B] \approx \mathbb{P}[B \mid \ v_0]$.

Definition 2.2 (λ-sampler)**.** Let $G = (V, U, E)$ be a bipartite graph. For any $B \subset U$, we define $N = N(B, \delta) = \{v \in V \mid \mathbb{P}_{u \in U}[u \in B \mid u \sim v] > \mathbb{P}[B] + \delta\}$. For $\lambda \in (0, 1)$ we say G is a λ-sampler if for every $B \subseteq U$ and every $\delta > 0$,

$$\mathbb{P}[N] \leqslant \frac{\lambda}{\delta^2} \mathbb{P}[B].$$

There is a tight connection between expander bipartite graphs and sampler graphs. For more on this, see [Gol11].

2.5 Agreement Expanders

Let V be a finite universe, S a collection of subsets of V, and for each subset $s \in S$, a local function $f_s \in \Sigma^s$. An ensemble $\{f_s\}$ is *perfectly global* if it comes from a single global function $w : V \to \Sigma$, namely, $f_s = w|_s$ for all s. We denote by $\mathcal{G}$ the collection of all perfectly global ensembles. An agreement tester is given by a non-negatively weighted graph $\mathcal{A}$ with vertex set S, and such that each edge $\{s, s'\}$ is labelled by some $k \subseteq s \cap s'$. Without loss of generality we require that the weights sum to 1, so that the edges form a distribution over pairs s, s'. Given a collection $\{f_s\}$ of local functions, the tester selects an edge s, s' at random, and accepts if $f_s(v) = f_{s'}(v)$ for each $v \in k$. We call this *the value of* $\{f_s\}$ *under* $\mathcal{A}$ and denote it by $\mathcal{A}(\{f_s\})$,

$$\mathcal{A}(\{f_s\}) := \mathop{\mathbb{P}}_{s,s' \sim \mathcal{A}}[f_s(v) = f_{s'}(v),\ \forall v \in k].$$

It is clear that a perfectly global ensemble has value 1. Indeed for any pair s, s' and any $v' \in s \cap s'$, $f_s(v) = g(v) = f_{s'}(v)$ assuming that $g : V \to \Sigma$ is the global function that agrees with $\{f_s\}$. The graph $\mathcal{A}$ is an agreement expander if a robust converse holds, namely any ensemble with $A(\{f_s\}) \approx 1$ has to be close to a perfectly global ensemble. Formally,

Definition 2.3. Let $V, K, S, \mathcal{A}$ be as above. We call $\mathcal{A}$ an α-agreement expander if for every ensemble of local functions $\{f_s\}$

$$\alpha \cdot \operatorname{dist}(\{f_s\}, \mathcal{G}) \leqslant 1 - \mathcal{A}(\{f_s\}). \tag{2.1}$$

where the distance $\operatorname{dist}(\{f_s\}, \mathcal{G})$ is the distance between $\{f_s\}$ and the closest perfectly global ensemble; where distance between two ensembles $\{f_s\}, \{g_s\}$ is defined as probability $f_s \neq g_s$ when s is chosen from the marginal distribution of $\mathcal{A}$.

More refined notions of agreement-expansion. We also say that $\mathcal{A}$ *is an agreement expander with respect to* δ*-ensembles* if (2.1) holds for every ensemble $\{f_s\}$ that is a δ-ensemble, namely, such that for every edge $\{s, s'\}_k$ in the graph $\mathcal{A}$, we have either $f_s|_k = f_{s'}|_k$ or else the Hamming distance between $f_s|_k$ and $f_{s'}|_k$ is at least $\delta|k|$.

Furthermore, we allow a slightly weaker notion of distance from being perfectly global. We say that $\mathcal{A}$ has (K, α) soundness wrt δ ensembles if the following holds. Suppose that for every $s \in S$ there is a distribution $k \sim D_s$ that samples $k \in K$ that are subsets of s. We say that $\mathcal{A}$ is (K, α)-sound wrt $\{f_s\}$ when

$$\alpha \cdot \min_{G \in \mathcal{G}} \mathop{\mathbb{P}}_{s \in S, k \sim D_s}[f_s|_k \neq G|_k] \leqslant 1 - \mathcal{A}(\{f_s\}). \tag{2.2}$$

We say that $\mathcal{A}$ is (K, α) sound with respect to δ ensembles if (2.2) holds for all δ ensembles.

3 Multilayer Agreement Samplers

The structure we use to construct locally testable codes has a sampler component and an agreement expander component, that sit together in four layers. We call these structures Multilayer Agreement Samplers.

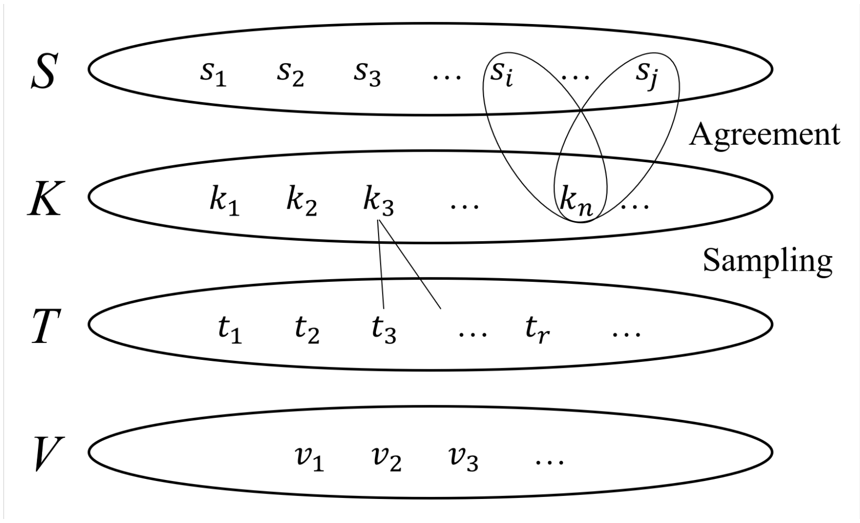
S
s_1 s_2 s_3 … s_i … s_j
Agreement
K
k_1 k_2 k_3 … k_n …
Sampling
T
t_1 t_2 t_3 … t_r …
V
v_1 v_2 v_3 …

Figure 1: Multilayer Agreement Sampler

Definition 3.1 (Multilayer Agreement Samplers (MAS))**.** Let $\delta, \lambda, \alpha \geqslant 0$. Let V be a set of elements, and $T, K, S \subset 2^V$ be families of subsets so that there is a non-degenerate Markov chain that samples (v, t, k, s) from V, T, K, S respectively, so that $v \in t \subset k \subset s$. (Spelling out the Markov chain requirement we have a distribution over (v, t, k, s) such that the choice of t is conditioned only on v, the choice of k is conditioned only on t, and finally the choice of s is conditioned only on k.)

We say that (V, T, K, S) are a $(\lambda, \delta, \alpha)$ *-MAS* if

1. There is an agreement expander $\mathcal{A}$ with vertex set S and edge labels K so that:
   - The marginal distribution of sampling a labeled edge $\{s, s'\}_k$ in $\mathcal{A}$, and returning s, k, is the same as the marginal distribution s, k of the Markov chain.
   - $\mathcal{A}$ is (K, α)-sound for δ-ensembles.
2. The bipartite containment graph of K vs. T, is a λ-sampler. Here the probability of sampling an edge (k, t) is the probability of sampling (k, t) together in the Markov chain.

A natural example for an MAS is the Grassmannian complex, that is, a four layer structure where $V = \mathbb{F}_p^n$ and T, K, S are affine spaces of $\mathbb{F}_p^n$ of fixed dimensions. We elaborate on this example in the subsection below. The Grassmannian complex is dense, that is, the number of subspaces grows exponentially with the dimension. No known codes on the Grassmannian complex have good rate.

Currently known constant degree MASs arise from high-dimensional expanders which are simplicial complexes. However, we cannot use MASs that are directly simplicial complexes to construct any code with non-trivial rate and distance. It is conceivable that high-dimensional expanders that are not simplicial complexes[2] may yield good LTCs.

3.1 MASs coming from the Grassmannian Complex

The set system for the Grassmannian MAS is corresponds to points and affine subspaces of a vector space. Formally, let p be some prime power, and $q_0 < q_1 < q_2 < n$ be some integers greater than 0. Our ground

[2] These can still arise from high dimensional expanders. For example an MAS whose subsets are *links* of a high dimensional expander.

set is $V = \mathbb{F}_p^n$, and over it we define the following set system $X = (V, T, K, S)$ where T, K, S consist of all affine subspaces of dimensions q_0, q_1 and q_2 respectively. The Markov process of this set system, is sampling $(v \in t \subseteq k \subseteq s)$ uniformly.

The edge distribution of the test graph $\{s, s'\}_k \sim \mathcal{A}$, is to sample a subspace $k \in K$, and then two subspaces $s, s' \in S$ independently, given that $s, s' \supset k$. We call this the q_2, q_1-agreement test.

We claim that this set system is an MAS:

Lemma 3.2. *There is a universal constant $\alpha > 0$ so that the following holds. Let $q_0 < q_1 < q_2 < n$ be as above, and assume $q_2 \geqslant 3q_1 + 2$. Let p be any prime power. Let $X = (V, T, K, S)$ be as above. Then X is a $(p^{q_0 - q_1}, \delta, \delta\alpha)$-MAS for every $\delta > 0$.*

the constant above does not depend on p, nor on q_0, q_1, q_2, n.

Proof. The sampling properties of the layers of a Grassmannian complex are folklore:

Fact 3.3. *Let $G = (K, T, E)$ be the graph where K are subspaces $\mathbb{F}_p^n$ of dimension q_1 and T are subspaces of dimension q_0, and $(t, k) \in E$ if $t \subset k$ with uniform weights. This graph is a $p^{-|q_0 - q_1|}$-sampler.*

Agreement of the q_2, q_1-agreement test graph was proven by [DD19] (Theorem 6.2).

Theorem 3.4 (Agreement for Grassmannian)**.** *There exists a constant $\alpha > 0$ such that for every prime power p, $\delta > 0$, and integers q_1, q_2, n such that $3q_1 + 2 < q_2 \leqslant n$ the following holds. The q_2, q_1-Grassmannian agreement test is $(K, \delta\alpha)$-sound for δ-ensembles.*

Combining these two statements together we get that there exists some $\alpha > 0$ so that for every $\delta > 0$, (V, T, K, S) defined above are a $(p^{q_0 - q_1}, \delta, \delta\alpha)$-MAS. □

In Section 5 use our main theorem, Theorem 4.1, to show that local testability of lifted on the Grassmannian complex, is implied by the local testability of the base code.

4 Main Theorem - Locally Testable Codes on MASs

Given an *MAS* (V, T, K, S) and a set of base codes $\{C_t \mid \ t \in T\}$, the *lifted code* to V is

$$C = \{w : V \to \Sigma \mid \ w|_t \in C_t, \forall t \in T\}.$$

Similarly, for every $s \in S$ or $k \in K$, the local lifts to s or k are

$$C_s = \{w : s \to \Sigma \mid \ w|_t \in C_t, \forall t \subseteq s\}, \ C_k = \{w : k \to \Sigma \mid \ w|_t \in C_t, \forall t \subseteq k\}.$$

The next theorem is a reformulation of Theorem 1.1.

Theorem 4.1 (Main)**.** *Let V be a finite set and $\rho, \delta, \lambda, \alpha \geqslant 0$ so that $\lambda \leqslant \frac{\rho\delta\alpha}{64}$. Let $X = (V, T, K, S)$ be a $(\delta, \lambda, \alpha)$-MAS. Let $\{C_t \mid \ t \in T\}$ be a set of base codes, and let C be the lifted code. Suppose that*

1. Local Distance: *C_k has distance δ for every $k \in K$.*
2. Local local testability: *For every $s \in S$, the code C_s is ρ-locally testable with respect to sampling $t \in T$ given that $t \subset s$.*

Then C is $\frac{\rho\delta\alpha}{16}$-locally testable with respect to the distribution of choosing $t \in T$.

We encourage the readers to think of λ, α as some fixed constants of the set system. Then the theorem states that if $\{C_k\}$ have large relative distance $\delta = \Omega(1)$, and $\{C_s\}$ are ρ-locally testable for a large enough ρ, then the lifted code is $\Omega(\rho)$-locally testable.

4.1 Proof of the Main Theorem

Proof of Theorem 4.1. Let $w_0 : V \to \Sigma$ be some word so that

$$Fail(w_0) \overset{\text{def}}{=} \underset{t\in T}{\mathbb{P}}[w_0|_t \notin C_t] = \varepsilon.$$

We need to find a word w^* so that $\text{dist}(w_0, w^*) \leqslant \frac{16\varepsilon}{\rho\delta\alpha}$. We will find a word $w_1 : V \to \Sigma$ so that $\text{dist}(w_0, w_1) = \frac{8\varepsilon}{\rho\delta\alpha}$, and

$$Fail(w_1) = \underset{t\in T}{\mathbb{P}}[w_1|_t \notin C_t] \leqslant \frac{1}{2}\varepsilon.$$

As a first step we define a function ensemble $\{f_s \mid\ s \in S\}$ so that $f_s \in C_s$ is the closest code word to $w_0|_s$ (ties broken arbitrarily). For each $k \subset s, s'$ both $f_s|_k \in C_k$ and $f_{s'}|_k \in C_k$, and since C_k is a code with relative distance δ, we get that $\{f_s\}$ is a δ-ensemble.

We claim that the ensemble passes the agreement test with high probability.

Claim 4.2.

$$\underset{\{s_1,s_2\}_k\sim\mathcal{A}}{\mathbb{P}}[f_{s_1}|_k = f_{s_2}|_k] = 1 - \frac{4\varepsilon}{\rho\delta}.$$

As there is an agreement expander $\mathcal{A}$ that is (K, α)-sound with respect to δ-ensembles, there exists some function $w_1 : V \to \Sigma$ so that

$$\underset{k\subset s}{\mathbb{P}}[w_1|_k = f_s|_k] = 1 - \frac{4\varepsilon}{\rho\delta\alpha}. \tag{4.1}$$

We claim that w_0 is close to w_1, and that w_1 fails the test with probability $\leqslant \frac{\varepsilon}{2}$.

Claim 4.3. $\text{dist}(w_0, w_1) \leqslant \frac{8\varepsilon}{\rho\delta\alpha}$.

Claim 4.4. $Fail(w_1) \leqslant \frac{1}{2}\varepsilon$.

Modulo Claim 4.3 and Claim 4.4, we repeat the correction process $poly(\log(\min_{t\in T} \mathbb{P}[t]))$ times. In the beginning of the i-th iteration we start with w_i that fails the test with probability $\leqslant \varepsilon/2^i$. In the end of the iteration, we find w_{i+1} that fails the test with probability $\leqslant \varepsilon/2^{i+1}$, and so that $\text{dist}(w_i, w_{i+1}) \leqslant \frac{8\varepsilon}{\rho\delta\alpha 2^i}$. Thus we obtain a sequence of functions $w_0, w_1, w_2, ..., w_r$ that ends with $w_r = w^*$ that always passes the test. The distance we accumulate from w_0 is

$$\text{dist}(w_0, w_r) \leqslant \sum_{i=0}^{r-1} \text{dist}(w_i, w_{i+1}) \leqslant \frac{8}{\rho\delta\alpha} \sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{16}{\rho\delta\alpha}.$$

□

Proof of Claim 4.2. By the local testability of the base code C_s,

$$\underset{s}{\mathbb{E}}[\text{dist}(w_0|_s, C_s)] \leqslant \rho^{-1} \underset{s}{\mathbb{E}}\left[\underset{t\subset s}{\mathbb{P}}[w_0|_t \notin C_t]\right] = \rho^{-1} \underset{t}{\mathbb{P}}[w_0|_t \notin C_t] \leqslant \frac{\varepsilon}{\rho}. \tag{4.2}$$

As f_s is closest code word to $w_0|_s$,

$$\frac{\varepsilon}{\rho} \geqslant \underset{s}{\mathbb{E}}[\text{dist}(w_0|_s, f_s)] = \underset{s}{\mathbb{E}}\left[\underset{k\subset s}{\mathbb{E}}[\text{dist}(w_0|_k, f_s|_k)]\right].$$

By Markov's inequality, with probability $1 - \frac{4\varepsilon}{\rho\delta}$ of sampling $\{s_1, s_2\}_k \sim \mathcal{A}$, it holds that $\text{dist}(w_0|_k, f_{s_i}|_k) < \frac{\delta}{2}$ where f_{s_i} is the closest codeword in C_{s_i} to $w_0|_{s_i}$.

By the local distance assumption, C_k has distance δ, and if $\text{dist}(f_{s_1}|_k, f_{s_2}|_k) < \delta$, then

$$f_{s_1}|_k = f_{s_2}|_k.$$

□

Proof of Claim 4.3. We note that

$$\mathrm{dist}(w_0, w_1) = \mathop{\mathbb{E}}_{s}\left[\mathrm{dist}(w_0|_s, w_1|_s)\right].$$

We show closeness by the triangle inequality. Fix $s \in S$, then

$$\mathrm{dist}(w_0|_s, w_1|_s) \leqslant \mathrm{dist}(w_0|_s, f_s) + \mathrm{dist}(f_s, w_1|_s).$$

By (4.2),

$$\mathop{\mathbb{E}}_{s}\left[\mathrm{dist}(w_0|_s, f_s)\right] \leqslant \frac{\varepsilon}{\rho}.$$

By the (K, α)-soundness of the agreement expander $\mathcal{A}$,

$$\mathrm{dist}(w_1|_s, f_s) = \mathop{\mathbb{E}}_{k\subset s}\left[\mathrm{dist}(w_1|_k, f_s|_k)\right] \leqslant \mathop{\mathbb{P}}_{k\subset s}\left[w_1|_k \neq f_s|_k\right] = \frac{4\varepsilon}{\rho\delta\alpha}.$$

By the triangle inequality, and using the fact that both $\delta, \alpha < 1$

$$\mathrm{dist}(w_0, w_1) \leqslant \frac{8\varepsilon}{\rho\delta\alpha}.$$

□

The proof of Claim 4.4 relies on the λ-sampling property of the MAS.

Proof of Claim 4.4. By assumption the containment graph between T and K is has the λ-sampling property. Let $B = \{k \in K \mid \ \forall s \supset k,\ f_s|_k \neq w_1|_k\}$. We observe the following:

1. By the agreement property, $\mathbb{P}\left[B\right] \leqslant \frac{8\varepsilon}{\rho\delta\alpha}$, and without loss of generality $\mathbb{P}\left[B\right] \leqslant \frac{1}{2}$ (if we want to show that the code is $\frac{\rho\delta\alpha}{16}$-locally testable, it is enough to consider ε so that $\frac{16\varepsilon}{\rho\delta\alpha} \leqslant 1$).
2. If $t \in T$ contributes to the failure (i.e $w_1|_t \notin C_t$), then $w_1|_k \neq f_s|_k$ for all $k \supset t$ and $s \supset k$. Thus *all* its neighbours are in B.

Denote by N the set of $t \in T$ so that all of t's neighbours are in B. By item 2 above we have that $\mathbb{P}_{t\in T}\left[w_1|_t \notin C_t\right] \leqslant \mathbb{P}_{t\in T}\left[N\right]$. We note that if we sample a neighbour of t, we get some $k \in B$ with probability $1 \geqslant \mathbb{P}\left[B\right] + \frac{1}{2}$. Thus by the λ-sampling property, we have that

$$\mathop{\mathbb{P}}_{t\in T}\left[N\right] \leqslant 4\lambda\frac{8\varepsilon}{\rho\delta\alpha}.$$

We chose $\lambda \leqslant \frac{\rho\delta\alpha}{64}$, hence $Fail(w_1) \leqslant \frac{1}{2}\varepsilon$.

□

Remark 4.5. The MAS has four layers. The vertex layer V and the layer T are required to define the lifted code itself. It is also natural to introduce a higher layer S, since without any other requirements we can't expect any lifted code to be locally testable.

However, the intermediate layer K is possibly unneeded. While it is a crucial part of the *proof*, it is not needed for lifting the code, nor for the local tests. We believe it is interesting to understand whether it is enough to study a three-layered set system, namely (V, T, S). Are there similar properties, in terms of agreement, sampling and expansion, that also give us a similar result?

5 Local Testability in Vector Spaces

In this section we demonstrate how the main theorem fits in with, and generalizes, the known results on testability of Reed-Muller codes. In this case the MAS is the Grassmannian complex MAS described in Lemma 3.2 for $V = \mathbb{F}_p^n$ and T, K, S being the collections of all affine subspaces of dimension q_0, q_1, q_2 respectively.

We define the code on V by lifting base codes $\{C_t \mid \ t \in T\}$. Namely

$$C = \left\{ w : \mathbb{F}_p^n \to \mathbb{F}_p \mid \ w|_t \in C_t, \ \forall t \in T \right\}.$$

One example of such a code, is the (n, r)-Reed-Muller code on $\mathbb{F}_p^n$. This code consists of all polynomials of degree $\leqslant r$. When $n = 1$ we call this the Reed-Solomon code. Take T to be the set of all affine lines (i.e. $q_0 = 1$), and let C_t be the r-Reed-Solomon code on every line. Lifting this code to V results in all functions $w : \mathbb{F}_p^n \to \mathbb{F}_p$ so that for every line $t \in T$, $f|_t$ is a function of degree at most r. For some parameters n, r, p this results in the (n, r)-Reed-Muller code. Surprisingly, [GKS13] showed that for some other parameters r, n, p the code lifted from the r-Reed-Solomon code, contains more than the (n, r)-Reed-Muller code. Nevertheless, these codes are locally testable as well [GHS15, HRS15].

Our main theorem states that to prove local testability of C it is enough to prove that $C_s = \{w : s \to \mathbb{F}_p \mid \ w|_t \in C_t, \ \forall t \in T, t \subset s\}$ is locally testable, for to each subspace $s \in S$.

This gives rise to testability results for Reed-Muller codes (which are well studied, see [RS96, RS97, AS03]) as well as to lifted codes as were studied in [GKS13] (given of course, that we check their local testability in a some small fixed space). Moreover, this statement continues to hold for more general sets of base codes $\{C_t \mid \ t \in T\}$: If the lifts of $\{C_t \mid \ t \in T\}$ to dimension q_2 subspaces are locally testable (with good enough parameters), then the lifted code to dimension n is also locally testable. This is particularly useful in the regime where q_0, q_2 are fixed, and n tends to infinity. This includes the examples above, but is a more general statement.

Theorem 5.1. *There is a universal constant $\alpha > 0$ so that the following holds. Let $q_0 < q_1 < q_2 < n$ be as above, and assume $q_2 \geqslant 3q_1 + 2$. Let p be any prime power. Let $X = (V, T, K, S)$ be as above. Let $\{C_t \mid \ t \in T\}$ be a set of base codes, and suppose that there exists some $\delta > 0$ and $\rho \geqslant \frac{64p^{q_1 - q_0}}{\alpha\delta^2}$ so that:*

1. *For any q_1 dimensional space $k \in K$, C_k has distance $\geqslant \delta$.*[3]
2. *For every q_2 dimensional space $s \in S$, C_s is ρ-locally testable.*

Then for any $n > q_2$, the lift of $\{C_t \mid \ t \in T\}$ to $\mathbb{F}_p^n$ is $\frac{\rho\delta^2\alpha}{16}$-locally testable.

The constant α doesn't depend on any of the other parameters, nor on the field size.

We encourage the readers to think of $\delta = \Omega(1)$. Then for every fixed dimensions q_0, q_1 and field size p there is some ρ, so that for every lifted code that is ρ-locally testable on spaces of dimension q_2, the code is also $\Omega(\rho)$-locally testable on all spaces of dimension $n > d$ (for a large enough ρ). Note that this theorem applies both to the regime where the field size is small (e.g. $p = 2, 3$), and where the field size goes to infinity. When p grows, the conditions of the theorem become easier to satisfy, that is, that the lower bound on ρ becomes smaller as well.

Proof of Theorem 5.1. Let α be the constant stated in Lemma 3.2. The system $X = (V, T, K, S)$ defined above is a $(p^{q_0 - q_1}, \delta, \delta\alpha)$-MAS, by Lemma 3.2, for that α.

Denote by C the lift of $\{C_t \mid \ t \in T\}$ to $\mathbb{F}_p^n$. This code satisfies the distance and local local testability properties:

[3] [GKS13] showed this holds, for example, whenever the base codes C_t themselves have distance $\geqslant \delta + \frac{1}{p^{q_0}}$.

1. The lift of $\{C_t \mid t \in T\}$ to an q_1 dimensional space $k \in K$ has distance $\geqslant \delta$.
2. The lift of $\{C_t \mid t \in T\}$ to a q_2-dimensional space $s \in S$ is ρ-locally testable.

Hence by Theorem 4.1, this code is $\frac{\rho\delta^2\alpha}{16}$-locally testable. □

References

[ABNNR92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. *Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs*. IEEE Trans. Inform. Theory, 38(2):509–516, 1992.

[Aro94] Sanjeev Arora. *Probabilistic checking of proofs and the hardness of approximation problems*. Ph.D. thesis, University of California, Berkeley, 1994.

[AS03] Sanjeev Arora and Madhu Sudan. *Improved low-degree testing and its applications*. Combinatorica, 23(3):365–426, 2003. (Preliminary version in *29th STOC*, 1997). eccc:1997/TR97-003.

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. *Checking computations in polylogarithmic time*. In *Proc. 23rd ACM Symp. on Theory of Computing (STOC)*, pages 21–31. 1991.

[BGHSV06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. *Robust PCPs of proximity, shorter PCPs and applications to coding*. SIAM J. Comput., 36(4):889–974, 2006. (Preliminary version in *36th STOC*, 2004). eccc:2004/TR04-021.

[BHR05] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. *Some 3CNF properties are hard to test*. SIAM J. Comput., 35(1):1–21, 2005. (Preliminary version in *35th STOC*, 2003).

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. *Self-testing/correcting with applications to numerical problems*. J. Comput. Syst. Sci., 47(3):549–595, December 1993. (Preliminary version in *22nd STOC*, 1990).

[BMSS11] Eli Ben-Sasson, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. *Symmetric LDPC codes are not necessarily locally testable*. In *Proc. 26th IEEE Conf. on Comput. Complexity*, pages 55–65. 2011. eccc:2010/TR10-199.

[BS08] Eli Ben-Sasson and Madhu Sudan. *Short PCPs with polylog query complexity*. SIAM J. Comput., 38(2):551–607, 2008. (Preliminary version in *37th STOC*, 2005). eccc:2004/TR04-060.

[DD19] Yotam Dikstein and Irit Dinur. *Agreement testing theorems on layered set systems*. In *Proc. 60th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 1495–1524. 2019. arXiv:1909.00638, eccc:2019/TR19-112.

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. *Boolean function analysis on high-dimensional expanders*. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, eds., *Proc. 20th International Workshop on Randomization and Computation (RANDOM)*, volume 116 of *LIPIcs*, pages 38:1–38:20. Schloss Dagstuhl, 2018. arXiv:1804.08155, eccc:2018/TR18-075.

[DHKNT19] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon TaShma. *List decoding with double samplers*. In *Proc. 30th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 2134–2153. 2019. eccc:2018/TR18-198.

[DHKRZ19] Irit Dinur, Prahladh Harsha, Tali Kaufman, and Noga Ron-Zewi. *From local testing to robust testing via agreement testing*. In Avrim Blum, ed., *Proc. 10th Innovations in Theor. Comput. Sci. (ITCS)*, volume 124 of *LIPIcs*, pages 29:1–29:18. Schloss Dagstuhl, 2019. eccc:2016/TR16-160.

[Din07] Irit Dinur. *The PCP theorem by gap amplification*. J. ACM, 54(3):12, 2007. (Preliminary version in *38th STOC*, 2006). eccc:2005/TR05-046.

[DK17] Irit Dinur and Tali Kaufman. *High dimensional expanders imply agreement expanders*. In *Proc. 58th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 974–985. 2017. eccc:2017/TR17-089.

[DKW18] Dominic Dotterrer, Tali Kaufman, and Uli Wagner. *On expansion and topological overlap*. Geometriae Dedicata, 195:307—317, 2018. (Preliminary version in *32nd Symp. Comput. Geom.*, 2016). arXiv:1506.04558.

[FS95] Katalin Friedl and Madhu Sudan. *Some improvements to total degree tests*. In *Proc. 3rd Israel Symp. on Theoretical and Computing Systems*, pages 190–198. 1995. (See arXiv for corrected version). arXiv:1307.3975.

[Gal60] Robert G. Gallager. *Low Density Parity Check Codes*. Ph.D. thesis, Massachusetts Institute of Technology, 1960.

[GHS15] Alan Guo, Elad Haramaty, and Madhu Sudan. *Robust testing of lifted codes with applications to low-degree testing*. In *Proc. 56th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 825–844. 2015. eccc:2015/TR15-043.

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. *New affine-invariant codes from lifting*. In Robert D. Kleinberg, ed., *Proc. 4th Innovations in Theor. Comput. Sci. (ITCS)*, pages 529–540. ACM, 2013. eccc:2012/TR12-149.

[GLRSW91] Peter Gemmell, Richard J. Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. *Self-testing/correcting for polynomials and for approximate functions*. In *Proc. 23rd ACM Symp. on Theory of Computing (STOC)*, pages 32–42. 1991.

[Gol10] Oded Goldreich. *Short locally testable codes and proofs: A survey in two parts*. In Oded Goldreich, ed., *Property Testing*, volume 6390 of *LNCS*, pages 65–104. Springer, 2010. eccc:2005/TR05-014.

[Gol11] ———. *A sample of samplers: A computational perspective on sampling*. In Oded Goldreich, ed., *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, volume 6650 of *LNCS*, pages 302–332. Springer, 2011. eccc:1997/TR97-020.

[Gro10] Mikhail Gromov. *Singularities, expanders and topology of maps. Part 2: from combinatorics to topology via algebraic isoperimetry*. Geom. Funct. Anal., 20:416—526, 2010.

[GS06] Oded Goldreich and Madhu Sudan. *Locally testable codes and PCPs of almost linear length*. J. ACM, 53(4):558–655, 2006. (Preliminary version in *43rd FOCS*, 2002). eccc:2002/TR02-050.

[HRS15] Elad Haramaty, Noga Ron-Zewi, and Madhu Sudan. *Absolutely sound testing of lifted codes*. Theory Comput., 11:299–338, 2015. (Preliminary version in *17th RANDOM*, 2013). eccc:2013/TR13-030.

[KM17] Tali Kaufman and David Mass. *High dimensional random walks and colorful expansion*. In Christos Papadimitriou, ed., *Proc. 8th Innovations in Theor. Comput. Sci. (ITCS)*, volume 67 of *LIPIcs*, pages 4:1–4:27. Schloss Dagstuhl, 2017. arXiv:1604.02947.

[KO18] Tali Kaufman and Izhar Oppenheim. *High order random walks: Beyond spectral gap*. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, eds., *Proc. 20th International Workshop on Randomization and Computation (RANDOM)*, volume 116 of *LIPIcs*. Schloss Dagstuhl, 2018. arXiv:1707.02799.

[KS08] Tali Kaufman and Madhu Sudan. *Algebraic property testing: the role of invariance*. In *Proc. 40th ACM Symp. on Theory of Computing (STOC)*, pages 403–412. 2008. eccc:2007/TR07-111.

[LM06] Nathan Linial and Roy Meshulam. *Homological connectivity of random 2-complexes*. Combinatorica, 26(4):475–487, 2006.

[Opp18] Izhar Oppenheim. *Local spectral expansion approach to high dimensional expanders part I: Descent of spectral gaps*. Discrete Comput. Geom., 59(2):293–330, 2018. arXiv:1709.04431.

[RS96] Ronitt Rubinfeld and Madhu Sudan. *Robust characterizations of polynomials with applications to program testing*. SIAM J. Comput., 25(2):252–271, April 1996. (Preliminary version in *23rd STOC*, 1991 and *3rd SODA*, 1992).

[RS97] Ran Raz and Shmuel Safra. *A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP*. In *Proc. 29th ACM Symp. on Theory of Computing (STOC)*, pages 475–484. 1997.

[Spi95] Daniel A. Spielman. *Computationally Efficient Error-Correcting Codes and Holographic Proofs*. Ph.D. thesis, Massachusetts Institute of Technology, June 1995.

[SS96] Michael Sipser and Daniel A. Spielman. *Expander codes*. IEEE Trans. Inform. Theory, 42(6):1710–1722, November 1996. (Preliminary version in *35th FOCS*, 1994).

[Tan81] Michael R. Tanner. *A recursive approach to low complexity codes*. IEEE Trans. Inform. Theory, 27(5):533–547, 1981.

[Tre04] Luca Trevisan. *Some applications of coding theory in computational complexity*. Quaderni di Matematica, 13:347–424, 2004. arXiv:cs/0409044, eccc:2004/TR04-043.