


New Codes on High Dimensional Expanders

Irit Dinur 

Weizmann Institute of Science, Rehovot, Israel

Siqi Liu 

UC Berkeley, CA, USA

Rachel Yun Zhang 

Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We describe a new parameterized family of symmetric error-correcting codes with low-density parity-check matrices (LDPC).

Our codes can be described in two seemingly different ways. First, in relation to Reed-Muller codes: our codes are functions on a subset of the points in \mathbb{F}^n whose restrictions to a prescribed set of affine lines has low degree. Alternatively, they are Tanner codes on high dimensional expanders, where the coordinates of the codeword correspond to triangles of a 2-dimensional expander, such that around every edge the local view forms a Reed-Solomon codeword.

For some range of parameters our codes are provably locally testable, and their dimension is some fixed power of the block length. For another range of parameters our codes have distance and dimension that are both linear in the block length, but we do not know if they are locally testable. The codes also have the *multiplication property*: the coordinate-wise product of two codewords is a codeword in a related code.

The definition of the codes relies on the construction of a specific family of simplicial complexes which is a slight variant on the coset complexes of Kaufman and Oppenheim. We show a novel way to embed the triangles of these complexes into \mathbb{F}^n , with the property that links of edges embed as affine lines in \mathbb{F}^n .

We rely on this embedding to lower bound the rate of these codes in a way that avoids constraint-counting and thereby achieves non-trivial rate even when the local codes themselves have arbitrarily small rate, and in particular below $1/2$.

2012 ACM Subject Classification Theory of computation \rightarrow Error-correcting codes; Theory of computation \rightarrow Expander graphs and randomness extractors

Keywords and phrases error correcting codes, high dimensional expanders, multiplication property

Digital Object Identifier 10.4230/LIPIcs.CCC.2025.27

Related Version Full Version: <https://arxiv.org/abs/2308.15563>

Funding Irit Dinur: Supported by ERC grant 772839, and ISF grant 2073/21. Part of the work was done while visiting the Simons Institute for the Theory of Computing.

Siqi Liu: Supported in part by the Berkeley Haas Blockchain Initiative and a donation from the Ethereum Foundation.

Rachel Yun Zhang: This research was supported in part by DARPA under Agreement No. HR00112020023, an NSF grant CNS-2154149, and NSF Graduate Research Fellowship 2141064.

Acknowledgements We wish to thank Gilles Zemor for pointing us to the work of [16]. We thank Amnon Ta-Shma for early comments on this work.

1 Introduction

A locally testable code (LTC) is an error correcting code that has a property-tester. The tester reads q bits that are randomly chosen, and rejects words with probability that grows with their distance from the code.



© Irit Dinur, Siqi Liu, and Rachel Yun Zhang;
licensed under Creative Commons License CC-BY 4.0

40th Computational Complexity Conference (CCC 2025).

Editor: Srikanth Srinivasan; Article No. 27; pp. 27:1–27:42



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



LTCs were initially studied and constructed side by side with PCPs (probabilistically checkable proofs). It was only recently that the question of existence of LTCs with the c^3 property was resolved in the affirmative [6, 31]. A code has the c^3 property if it has constant rate, constant distance and testable with constant locality.

It was initially hoped that high dimensional expanders, a la [25, 24], can serve as the combinatorial structure underlying the code, and all one needs to do is to find an appropriate collection of local codes (at the links) to match up with the combinatorics, see [8, 7, 4].

This approach has not borne fruit up until now, essentially due to the stringent requirements on the local codes, which turned out difficult to fulfill. Nevertheless, c^3 codes were eventually constructed by circumventing the problem and switching from simplicial to square complexes. The main benefit of square complexes is that their local views support tensor codes, which satisfy the requirements for local testability.

In this work we return to the simplicial setting and construct a new parameterized family of locally testable codes on simplicial (bounded-degree) high dimensional expanders. In addition to serving as a new and potentially interesting family of LDPC error-correcting codes, these codes satisfy further properties that could be potentially useful for other applications such as PCPs and beyond. In particular, the codes are symmetric in the sense of [22, 18], meaning that there is a group acting on the coordinates of the codeword, that takes every coordinate to every coordinate. In addition, the fact that local views of our codes are Reed-Solomon, immediately implies that the codes satisfy the *multiplication property*: the coordinate-wise product of two codewords is a codeword in a related code.

These two properties (symmetry and multiplicativity) do not hold for the known constructions of c^3 LTCs due to their using random linear codes as local codes.

Our codes can be described in two seemingly different ways. First, in relation to Reed-Muller codes: our codes are functions on a subset $\bar{X}_n \subset \mathbb{F}^n$ whose restrictions to a prescribed set of affine lines has low degree. Alternatively, they are Tanner codes on high dimensional expanders, where the coordinates of the codeword correspond to triangles of a 2-dimensional expander X , such that around every edge the local view forms a Reed-Solomon codeword. The definition of the codes relies on the construction of a specific family of simplicial complexes whose triangles embed naturally into \mathbb{F}^n , with the property that links of edges embed as affine lines¹ in \mathbb{F}^n .

► **Theorem 1.** *Let $\mathbb{F} = \mathbb{F}_q$ be a fixed finite field. For every n divisible by 9 there exists a connected 2-dimensional 3-partite simplicial complex X_n , such that*

- (a) *For each vertex $v \in X_n(0)$, the link of v is a bipartite q -regular graph on $2q^2$ vertices whose normalized adjacency matrix has second largest eigenvalue $\lambda_2 = 1/\sqrt{q}$.*
- (b) *There is a set of points $\bar{X}_n \subset \mathbb{F}^n$ and an injective map $\iota : X_n(2) \rightarrow \mathbb{F}^n$, such that $\bar{X}_n = \text{Im}(\iota) \subset \mathbb{F}^n$, and*

$$|X_n(2)| = |\bar{X}_n| \geq q^{cn}$$

for some absolute constant $c > 0$.

- (c) *There is a set \mathcal{L}_n of affine lines in \mathbb{F}^n with one line per edge $e \in X_n(1)$. The line corresponding to an edge e is given by a bijection $\ell_e : \mathbb{F} \rightarrow X_{+e}$ such that $\iota \circ \ell_e : \mathbb{F} \rightarrow \mathbb{F}^n$ is an affine line in \mathbb{F}^n , where we denote $X_{+e} = \{t \in X_n(2) \mid t \supset e\}$. Let $\mathcal{L}_n = \{\iota \circ \ell_e \mid e \in X(1)\}$.*

¹ An affine line in \mathbb{F}^n is given by $\mathbf{a}_0 \in \mathbb{F}^n$ and $\mathbf{a}_1 \in \mathbb{F}^n \setminus \{0\}$ so that $\ell_e = \{\mathbf{a}_0 + t\mathbf{a}_1 \mid t \in \mathbb{F}\}$

(d) X_n is 3-partite, so the edges have 3 distinct types and we denote by \mathcal{L}_n^i the set of lines corresponding to type- i edges.

Furthermore, X_n and the maps $\iota, \{\ell_e\}_e$ are constructible in polynomial time in the size of the complex X_n .

The complexes $\{X_n\}$ are a slight variant on the coset complexes of [19]. The embedding of $X_n(2)$ into \mathbb{F}^n , so that the links of edges map into affine lines is new, and allows us to define a new family of codes on X_n ,

► **Definition 2** (HDX local Reed-Solomon Codes). Let $\mathbb{F} = \mathbb{F}_q$ and $\{X_n\}_n$ be as above, and let $d_1, d_2, d_3 < q$ be positive integers. We define a family of codes as $n \rightarrow \infty$,

$$C_{n,d_1,d_2,d_3} = \{f : \bar{X}_n \rightarrow \mathbb{F} \mid \forall i = 1, 2, 3, \forall \ell \in \mathcal{L}_n^i, f \circ \ell \in RS(q, d_i)\} \quad (1)$$

where $RS(q, d) \subset \mathbb{F}^{\mathbb{F}}$ is the Reed Solomon code of degree d over \mathbb{F} . Note that we have three distinct degree parameters since the edges of the complex have three distinct types.

The code C_{n,d_1,d_2,d_3} can alternatively be described as

$$C_{n,d_1,d_2,d_3} = \{f : X_n(2) \rightarrow \mathbb{F} \mid \forall e \in X(1), f|_{X_{+e}} \in C_e\}. \quad (2)$$

for an appropriate choice of codes $C_e \subset \mathbb{F}^{X_{+e}}$ such that C_e isomorphic to $RS(q, d_i)$ whenever e is an edge of type i .

We also denote $C_{n,d} = C_{n,d,d,d}$. In the sequel we will often focus on $C_{n,d}$ as it contains all of the ideas, but we wanted to include the slightly more general definition of C_{n,d_1,d_2,d_3} .

The isomorphism between definitions (1) and (2) is essentially given by ι from Theorem 1. The map ι identifies $X(2)$ and $\bar{X}_n \subset \mathbb{F}^n$, and further identifies the set of triangles containing an edge $e \in X(1)$ with an affine line in \mathbb{F}^n (see also Section 3.4 and Claim 25 for a full proof).

By definition (1) we can see that every n -variate polynomial of total degree at most $d = \min(d_1, d_2, d_3)$ gives rise to a codeword in $C_{n,d}$. This allows us to give a non-trivial lower bound on the dimension of $C_{n,d}$ even when d is small and constraint-counting fails. On the other hand, definition (2) allows us to prove distance and local testability through the high dimensional expansion machinery.

► **Theorem 3.** Fix a prime power q and let $d < q$. The family $\{C_{n,d}\}_n$ has the following properties

1. **Rate:** The dimension of the code is at least the dimension of the Reed-Muller code with $n/3$ variables and degree d . Furthermore, for $d > \frac{2}{3}q$ the code has dimension $\Omega(|X_n(2)|)$ (namely, linear rate as $n \rightarrow \infty$).
2. **Distance:** If $d < q - \Omega(\sqrt{q})$ the code has constant relative distance $\Omega_q(1)$.
3. **LDPC and Local Testability:** For all d and n , the code is defined by parity checks of length $d + 2$. When q is prime, if $d < q/4$ the code is locally testable with $d + 2$ queries.
4. **Multiplication:** For any d_1, d_2 , and for any $w_1 \in C_{n,d_1}$ and $w_2 \in C_{n,d_2}$, we have $w_1 \odot w_2 \in C_{n,d_1+d_2}$ where $w_1 \odot w_2$ is the coordinate-wise product of w_1 and w_2 .
5. **Symmetric:** There is a transitive group action on the coordinates of the code that preserves the code. Transitivity means that for every pair of coordinates t, t' there is a group element that takes t to t' .

► **Remark 4.** Here is a table that summarizes the rate, relative distance, and local testability of $\{C_{n,d}\}_n$ in different regimes of d .

d	Rate	Distance	Local testability
$(0, q/4)$	$\geq \binom{n/3}{d}$	$\Omega_q(1)$	$d + 2$
$[q/4, \frac{2}{3}q]$	$\Omega(X_n(2))$?
$[\frac{2}{3}q, q - \Omega(\sqrt{q})]$			

We fail to give a family of c^3 codes (with constant relative rate, constant relative distance, and local testability with constant number of queries). Specifically, in the low degree regime $d < q/4$, the lower bound on the relative rate is subconstant $> \binom{n/3}{d}/|X_n(2)| = q^{-\Omega(n)}$. In the high degree regime $d > 2q/3$, our approach for showing local testability fails, and we don't know whether or not local testability occurs.

There is a natural way to generalize our complexes and codes to higher dimensions. In Section 6 we describe this generalization and show that distance as well as local testability “trickles down”. We also show that the codes have a homological description as the space of cycles on the top dimension.

1.1 Background

Recent works [6, 31] give locally testable codes that have the c^3 property, namely they have constant relative rate, constant relative distance and locally testable with a constant number of queries. These codes are constructed on an especially designed squares complex. Similar schemes have been hypothesized to exist on simplicial complexes (see discussion in [6], and see [10]).

The simplicial complex codes are defined as

$$C = \left\{ w \in \mathbb{F}_q^{X(2)} \mid w|_{X_{+e}} \in C_e \right\} \quad (3)$$

where we fix, for each edge e , a local code $C_e \subseteq \mathbb{F}_q^{X_{+e}}$ where X_{+e} is the set of triangles containing the edge e . In other words, C is defined by aggregating the local codes C_e into one big parity check matrix, using the structure of the HDX. The definition is simple, but the challenge is in analysing properties of C such as rate, distance, or local testability.

1. Distance follows quite directly from the distance of each C_e together with expansion of X .
2. Local testability can be shown if we manage to show a local version of local testability, per each vertex v . Namely, we look at $C_v := C|_{X_{+v}} \subseteq \mathbb{F}_q^{X_{+v}}$, the restriction of the code to the set X_{+v} of triangles containing a vertex v . This itself is a linear subspace of $\mathbb{F}_q^{X_{+v}}$.

We say that C_v itself is locally testable if a good test for $z \in C_v$ is to select a random $e \ni v$ and check if $z|_{X_{+e}} \in C_e$.

If each C_v is locally testable, then the high dimensional expansion of X would allow us to deduce local testability of the entire code C .

This is the content of Theorem 41, whose proof is based on the same ideas underlying the proof of local testability in the the squares complex [6, 31] and also earlier the proof of cosystolic expansion in high dimensional expanders [17, 9], and is similar to [10, Proposition 6.4], although there are technical differences.²

² They show that if the links are coboundary expanders, then the global sheaf is a cosystolic expander. This is morally equivalent to saying that local robustness implies global local-testability. However, our proof makes use of a weaker local robustness condition, which is what we manage to prove for the local codes.

3. The dimension (or rate) of C could a priori be 0, so one needs to show that C is non-trivial somehow. So far the only successful method has been through constraint counting. One shows that the total number of parity checks is smaller than the number of degrees of freedom. Of course the constraints are often highly dependent so this argument is not tight, and one would like new techniques for lower bounding the dimension.

A major difficulty in construction of LTCs is in coming up with a collection of local codes C_e that simultaneously allow proving that each C_v is locally testable, and at the same time maintain non-trivial dimension for C . Indeed, several prior works gave a general framework for HDX codes but did not know how to instantiate them non-trivially. In [5], the authors gave a definition based on double samplers, and the proof of local testability relies on agreement testing as does ours. In [6], the c^3 problem has been resolved via a similar approach, but where the underlying complex is a squares complex that has a simpler link structure, which, in turn, implies more flexibility in the choice of local codes. A similar framework is introduced and studied in [10], where codes are described as sections in sheaved high dimensional expanders. All of these works, including the current paper, follow one unifying local to global principle, and are descendants of the work of [17, 9] that showed coboundary expansion based on the same property in the links, plus some form of spectral expansion.

In this work we give a first family of codes on high dimensional expanders that are defined by instantiating (3). As mentioned before, our local edge codes C_e are Reed-Solomon codes. The codes C_v at each vertex turn out to be the following

$$C_v \cong C_{d_x, d_y} = \{f : \mathbb{F}_q^3 \rightarrow \mathbb{F}_q \mid \forall a, b, c, \deg_x(f(x, b, c)) \leq d_x; \deg_y(f(a, y, ay + c)) \leq d_y\}. \quad (4)$$

Quite surprisingly we have learned, after completing this work, that these (local) codes have already been studied in [16], and their rate and distance is known (for the case $d_x = d_y$). We compute the rate of these codes in the case $d_x \neq d_y$ for $d_x + d_y < q/2$ (see Theorem 36), and prove that C_{d_x, d_y} is agreement-testable when $d_x, d_y < q/4$ and when q is a prime (see Theorem 38). Namely, given two functions $X, Y : \mathbb{F}_q^3 \rightarrow \mathbb{F}_q$ such that for all $b, c \in \mathbb{F}_q$, $X(x, b, c)$ is a low degree function of x ; and such that for all $a, c \in \mathbb{F}_q$, $Y(a, y, ay + c)$ is a low degree function of y , if

$$\mathbb{P}_{a, b, c} [X(a, b, c) \neq Y(a, b, c)] < \epsilon$$

then there is some $f \in C_{d_x, d_y}$ that is close to both X and Y . To show that this is a good test we adapt the analysis of Polychuk and Spielman [32] of the local testability of bivariate polynomial functions. In order to make the analysis go through, we need a characterization of the codewords of the local code, namely that every codeword is described by a polynomial that is low degree in all three coordinate directions. This is given in Lemma 34. We remark that the local testability we obtain is slightly weaker: there is a quadratic, rather than linear, relation between the distance of the word to a code and the probability of failing the test. Nevertheless we derive local-testability of C from this slightly weaker local testability of the C_v 's.

Finally, as for the dimension of the code C . For degree parameters that allow local testability, constraint counting is out of the question, because there are more constraints than degrees of freedom. Instead we show that the code contains many codewords by showing that every low degree multi-variate polynomial is a valid codeword in of our code. We leave for future research to try and get tighter bounds on the rate of these codes.

Multiplication Property

A triple C_1, C_2, C_3 of codes are so-called *multiplication codes* if for any $w_1 \in C_1$ and $w_2 \in C_2$ the word $w_1 \odot w_2$ obtained by coordinate-wise product satisfies $w_1 \odot w_2 \in C_3$. Codes with the multiplication property enable computation “under the hood”, namely, without decoding. Roughly speaking, the product of two codewords is an encoding of the product of the two messages. This can clearly be useful for secure computation. For example, if the two messages represent two subsets, then their product is the intersection of the sets, and we can compute set disjointness without decoding.

This property was introduced in [27] who used it towards a proof for $IP = PSPACE$ involving abstract codes, not necessarily Reed-Muller. The original proof of [26, 34] relied on Reed-Muller codes which are indeed multiplication codes. The multiplication property is used for example in the proof of the PCP theorem [26, 2, 1], and could potentially be useful for constructing PCPs from LTCs.

Since our codes are defined as lifts of Reed-Solomon codes they are automatically multiplication codes for appropriate degree choices (a Tanner codes inherits the multiplication property from the local-view codes). Alternatively, they are also Tanner codes with Reed-Solomon codes as local codes. So these Tanner codes inherit their multiplication property from the local codes. By choosing the parameters appropriately, we can get C_1, C_3 an LDPC code with linear rate, and C_2 a locally testable (LDPC) code with polynomial rate, such that this triple has the multiplication property (see Lemma 27).

Connection to Reed-Muller codes, lifted codes, and Tanner Codes

Recall that a Tanner code is given by two pieces of data:

- A bipartite graph $\mathcal{G} = (\mathcal{B}, \mathcal{P}, E)$ with left vertices \mathcal{B} (for bits), right vertices \mathcal{P} (for parity checks), and edges E .
 - For each $p \in \mathcal{P}$ there is a local code $C_p \subset \mathbb{F}^{\Gamma(p)}$ where $\Gamma(p) \subset \mathcal{B}$ are the neighbors of p .
- Given \mathcal{G} and $\{C_p\}$, the Tanner code is

$$\mathcal{T}(\mathcal{G}, \{C_p\}) = \{w \in \mathbb{F}^{\mathcal{B}} \mid \forall p \in \mathcal{P}, w|_{\Gamma(p)} \in C_p\}.$$

The Reed-Muller code $RM_{n,d}$ is the space of all n -variate polynomials of total degree at most d . For a broad set of parameters, Reed-Muller codes are themselves known to be Tanner codes, with local parity check codes being Reed Solomon codes, and with a $\mathcal{B} = \mathbb{F}^n$ and \mathcal{P} being the set of all affine lines in \mathbb{F}^n , so that the bipartite graph \mathcal{G}_{RM} is the point-vs.-line graph which connects a point in \mathbb{F}^n to all lines passing through it. For some parameter regimes (when the degree is high) such Tanner codes form a larger space than the Reed-Muller codes, as was discovered and studied in [14], where such codes are termed “lifted codes”. Later in [11], partial lifts are considered where some of the vertices of \mathcal{P} are erased and one considers the remaining punctured code. Clearly, different choices of where to puncture will have different properties. The codes studied in [11] are locally correctable and each point in \mathbb{F}^n is contained in $\text{poly}(n)$ affine lines in \mathcal{P} . This type of codes resembles the situation in our codes, as we explain next, but a key difference is that in our construction each point in \mathbb{F}^n only participates in constant number of parity checks.

The code C_{n,d_1,d_2,d_3} , as per (1), can be viewed a Tanner code whose graph is a subgraph of \mathcal{G}_{RM} obtained by keeping a subset \mathcal{L}_n of the lines, and a subset $\tilde{X}_n \subset \mathbb{F}^n$ of the points. We also allow different degree restrictions on different types of lines. If we set $d = d_1 = d_2 = d_3$ the parity check matrix of $C_{n,d}$ is thus a sub-matrix of the parity check matrix of the Reed-Muller code matrix. As for (2), it can be seen as a Tanner code on the bipartite graph $\mathcal{G} = (X(2), X(1), E)$ connecting each edge $e \in X(1)$ on the right to the set $X_{+e} \subset X(2)$ of triangles containing it, and putting a Reed-Solomon code on every X_{+e} in a specified way.

Two-query testability

One can often convert a code that is testable with q queries to another code that is testable with two queries, while increasing the alphabet. This is done by converting each codeword to the list of its local views. For example, in the case of Reed-Muller codes, instead of representing a polynomial by its evaluation on points, we represent it by providing its restriction to each plane. This is sometimes called the planes table, and is two-query testable by the plane-vs.-plane test, see [33]. In the case of c^3 LTCs of [6, 31], this conversion is immediate, although it is not described there explicitly. For our codes it also holds, and we include details in Claim 17. This property was also highlighted in [10].

1.2 Further Work

This work raises many questions for further investigation.

- **Better bounds.** What is the exact rate of our codes? We give a lower bound based on Reed-Muller codes, but we don't know how close this bound is to the truth. Are there parameter regimes where our codes are c^3 ?
- **Efficient encoding.** Our codes are specified by their parity-check matrices. A generator matrix can be computed in $O(N^3)$ time, where N is the block length. Given the generator matrix, computing an encoding takes time $O(N^2)$. Can encoding be done more efficiently, ideally even in linear-time?
- **Quantum codes.** It has recently become clear that locally testable HDX codes are related to quantum LDPC codes of the CSS type. Our construction gives a 2-chain (from vertices to edges to triangles) which can automatically be viewed as a quantum LDPC code. Indeed, we show in Section 2.6 how any HDX code gives rise to a sheaf which, by [10, Section 7.4], gives rise to a quantum code. Our local testability analysis implies distance in one direction, whereas for a good quantum LDPC code, one needs to prove both distance as well as co-distance. This is an interesting challenge. Moreover, in Section 6 we show how our construction generalizes to k dimensions, which, by the same recipe as above, can be converted to a sheaf and a $k + 1$ -chain. Studying the resulting quantum codes at intermediate levels $0 < i < k$ seems like an interesting direction.
- **Sparsified Grassmannian.** We have constructed a collection of affine lines with the property that each point has exactly three lines passing through it. This is a very sparse collection of lines, that never-the-less expands quite well when we walk from point to line to point. How does this generalize to higher dimensional complexes? The question of finding sparse Grassmannians has been studied in [29, 3, 21, 13] and could provide new PCP gadgets with properties similar to [23].
- **Other coset complexes.** We have considered a variant of the coset complexes of Kaufman and Oppenheim, by choosing subgroups based on a sub-ring of the ring chosen in [19]. Many high dimensional expanders are based on matrix groups [25, 24, 19, 30]. Thus similar embeddings of the complexes into \mathbb{F}^n could potentially be useful for coming up with new codes, whose properties can be studied through the high dimensional expansion framework.
- **Generalized sumcheck.** Many interactive proof systems reduce proving certain relations to checking that a multi-variate polynomial $p(x_1, \dots, x_m)$ over \mathbb{F}^m sums up to zero over some subset $\mathbb{H}^m \subseteq \mathbb{F}^m$. The sumcheck protocol is designed to check this relation by recursively restricting variables of p and checking the condition over smaller sets of random variables. Can one design sumcheck-like protocols for codes other than Reed-Muller? For

example, in the case of our code, we would need to find subsets of points that play the role of H^m , so that one can test whether some partial sums of a codeword sum up to zero.

Perhaps a first step would be to find a sumcheck protocol for our local codes C_{d_x, d_y} .

1.3 Organization

We define our coset complex in Section 3 and prove its properties (Theorem 1) in Sections 3.1, 3.2, and 3.3. We then define our code on this coset complex in Section 3.4. Its properties as detailed in Theorem 3 are proved in the following sections:

- In Section 3.5 we show that our code has the multiplication property, and in Section 3.6, we show that our code has constant distance in appropriate parameter regimes (items 2 and 4 of Theorem 3). We also show that our code has the transivity property (item 5).
- We discuss the rate (item 1) of the global and local codes in Sections 4.1 and 4.2 respectively.
- In Sections 5.1 and 5.3, we prove that the local and global codes are agreement testable. Finally, in Section 6 we describe the generalization to higher dimensions.

2 Preliminaries

2.1 Expander Graphs

A d -regular graph G is said to be a γ -one-sided expander if it has eigenvalues $d = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -d$ which satisfy $\lambda_i \leq \gamma \cdot d$ for all $i > 1$.

► **Lemma 5** (Alon-Chung). *Let $G = (V, E)$ be a d -regular γ one-sided expander. Let $T \subseteq V$ be such that the graph induced on T has average degree at least δd . Then $|T| \geq (\delta - \gamma) \cdot |V|$.*

Proof. Let A be the normalized adjacency matrix of G and let f be the indicator function of T . Using the spectral decomposition $f = \frac{|T|}{|V|} \mathbf{1} + f^\perp$ we get

$$\delta d|T| \leq 2E(T) = f^\top A f \leq |T|^2 d / |V| + \gamma d|T|$$

where $E(T)$ denotes the number of edges in the induced graph on T . Dividing both sides by $d|T|$ and rearranging gives the lemma. ◀

2.2 High Dimensional Expanders

A pure k -dimensional simplicial complex X is a set system (or hypergraph) consisting of a set of vertices $X(0)$ and an arbitrary collection of subsets of size $k + 1$ together with all their subsets. The sets of size $i + 1$ in X are denoted by $X(i)$. We will sometimes omit set brackets and write for example $uvw \in X(2)$ instead of $\{u, v, w\} \in X(2)$. As convention $X(-1) = \{\emptyset\}$. Unless it is otherwise stated, we always assume that X is finite.

Let $i < k$ and $s \in X(i)$. It is standard to define the link of s to be a $k - i - 1$ -dimensional simplicial complex defined by $X_s = \{t \setminus s \mid t \in X, t \supseteq s\}$. We also define the less-standard but useful notation

► **Definition 6** (Star). *For a k -dimensional complex X and a face $s \in X(i)$ for some $i < k$, the star of s is the k -dimensional complex containing all faces that contain s .*

$$X_{+s}(j) = \{t \in X(j) \mid t \supseteq s\}.$$

For a face $s \in X(i)$, there is a natural bijection $X_{+s}(j) \rightarrow X_s(j-i-1)$ mapping $t \in X_{+s}$ to $t \setminus s \in X_s$.

► **Definition 7** (High dimensional link expander). *Let X be a k -dimensional simplicial complex. Let $\lambda \geq 0$. We say that X is a one-sided λ -link expander if for every $s \in X^{\leq k-2}$, the graph $(X_s(0), X_s(1))$ is a λ -one sided spectral expansion.*

► **Definition 8** (Coset complex). *A k -dimensional coset complex is given by a group G and subgroups K_1, \dots, K_{k+1} . The vertices are all cosets of the $k+1$ subgroups, and the i -faces are all $i+1$ -tuples of cosets that have a non-empty intersection. The complex is denoted $X[G; K_1, \dots, K_{k+1}]$.*

The degree of a complex is defined as $\max_{v \in X(0)} |\{e \in X(1) \mid v \subset e\}|$, the maximum edge degree of a vertex. A beautiful construction of a constant-degree coset complex that is a high dimensional expander was given in [19], see also [15, 30].

It is not hard to see that links in a coset complex are themselves coset complexes.

2.3 Random Walks on High Dimensional Expanders

Let X be a regular two-dimensional complex, so that every vertex touches the same number edges, and every edge touches the same number of triangles. This assumption is not needed, but it is satisfied by our coset complexes and it slightly simplifies the definitions below.

Let $V = X(0)$, $E = X(1)$, $X_+ = X(2)$, and also denote $X(-1) = \{\phi\}$.

The down operator $D^i : \mathbb{R}^{X(i)} \rightarrow \mathbb{R}^{X(i-1)}$ (for $0 \leq i \leq 2$) and the up operator $U^i : \mathbb{R}^{X(i)} \rightarrow \mathbb{R}^{X(i+1)}$ (for $-1 \leq i \leq 1$) are defined by

$$\begin{aligned} \forall a \in X(i-1), \quad D^i f(a) &= \mathbb{E}_{b \supset a} [f(b)], \\ \forall b \in X(i+1), \quad U^i g(b) &= \mathbb{E}_{a \subset b} [g(a)]. \end{aligned}$$

We will often drop the superscript i in U^i and D^i when it is clear what i is.

Let $e \sim e'$ denote the lower random walk on the edges (choose a random e , then $v \in e$, then $e' \ni v$). It is easy to see that the Markov operator corresponding to this walk is just UD . Let $e \frown e'$ denote the non-lazy upper random walk on the edges (choose a random e , then $t \supset e$, then $e' \in t$ such that $e' \neq e$). It is not hard to see that the Markov operator corresponding to this walk, denoted M^+ , satisfies $DU = \frac{2}{3}M^+ + \frac{1}{3}I$.

We define inner products on the spaces $\mathbb{R}^V, \mathbb{R}^E$ by expectation according to the uniform distribution (here we are using the regularity assumption).

For example, for $f, g : V \rightarrow \mathbb{R}$

$$\langle f, g \rangle = \mathbb{E}_{x \in V} [f(x)g(x)], \quad \|f\| = (\langle f, f \rangle)^{1/2}.$$

The following is by now well known, see [8, 20], and we include a proof for completeness.

► **Lemma 9.** *Let X be a one-sided γ -link expander. Every $g \in \mathbb{R}^E$ satisfies*

$$\langle g, M^+ g \rangle \leq \langle g, (UD + \gamma I) g \rangle.$$

In particular, for a set $R \subset X(1)$ such that $\beta = \mathbb{P}_{e \frown e'} [e \in R | e' \in R]$ it must be that

$$\mathbb{P}_{e \frown e'} [e \in R | e' \in R] \geq \beta - \gamma$$

by applying the lemma on $g = \mathbf{1}_R$ and observing that $\mathbb{P}_{t \in X(2), e \neq e' \subset t} [e, e' \in R] = \beta \mathbb{P}[R]$.

27:10 New Codes on High Dimensional Expanders

Proof.

$$\begin{aligned}
\langle g, M^+g \rangle &= \mathbb{E}_{abc \in X_+} g(ab)g(ac) \\
&= \mathbb{E}_a \mathbb{E}_{bc \in X_a(1)} [g(ab)g(ac)] \\
&\leq \mathbb{E}_a \mathbb{E}_{b,c \in X_a(0)} [g(ab)g(ac)] + \gamma \mathbb{E}_a \mathbb{E}_{b \in X_a(0)} [g(ab)^2] \\
&= \langle g, UDg \rangle + \gamma \|g\|^2
\end{aligned}$$

where the inequality follows since X_a is a γ -one-sided expander for each a , and this means that for any $f : X_a(0) \rightarrow \mathbb{R}$ (and in particular setting $f(b) := g(ab)$),

$$\mathbb{E}_{bc \in X_a(1)} f(b)f(c) \leq \mathbb{E}_{b,c \in X_a(0)} f(b)f(c) + \gamma \cdot \mathbb{E}_{b \in X_a(0)} f(b)^2.$$

Note that the distribution of choosing a and then two edges ab, ac independently, is equivalent to choosing a random edge e' , then a random vertex $a \in e'$ and then another edge $e \ni a$. ◀

The following swap walk $S_{0,1}$ that starts from a random edge in $X(1)$ and ends at some vertex in $X(0)$ will be used in the analysis of local testability of the global code. Starting with a random edge e , choose a random triangle $t \ni e$ and output $v = t \setminus e$.

► **Lemma 10.** *Let X be a two-dimensional γ -link-expander. The random walk $\tilde{M} = S_{0,1}D$ on the edges has second largest eigenvalue bounded by 3γ .*

Proof. We claim that

$$M^+UD = \frac{1}{2}S_{0,1}D + \frac{1}{2}UD.$$

Let us analyze the random walk corresponding to the operator M^+UD . We note that when viewed as an operator for functions in \mathbb{R}^E ,

$$M^+UDf(e) = \mathbb{E}_{e_1 \stackrel{M^+}{\sim} e} [UDf(e_1)] = \mathbb{E}_{e' \stackrel{UD}{\sim} e_1 \stackrel{M^+}{\sim} e} [f(e')].$$

The underlying random walk from e to e' is as follows. We start from an edge e , go (via M^+) to a random edge e_1 such that $e \cup e_1 \in X(2)$. We then go from e_1 , via UD , to an edge e' . In this step two things can happen, each with probability $1/2$. Either e' doesn't contain $\{v\} = e_1 \cap e$, in which case we end up with the distribution of $S_{0,1}D$; or $e' \ni v$, in which case we end up with UD . This proves the required equality. We can thus write $S_{0,1}D = 2M^+UD - UD$. Furthermore, by Lemma 9 we have that for any $f \in \mathbb{R}^E$,

$$\langle f, (2M^+UD - UD)f \rangle \leq \langle f, (2(UDUD + \gamma UD) - UD)f \rangle.$$

So for any $f \in \mathbb{R}^E$ with $\mathbb{E}[f] = 0$ we get,

$$\langle f, S_{0,1}Df \rangle \leq \langle f, (2UDUD - UD)f \rangle + 2\gamma \|f\|^2 = \langle f, U(2DU - I)Df \rangle + 2\gamma \|f\|^2 \leq 3\gamma \|f\|^2$$

where in the last inequality we have used the fact that $2DU - I$ is nothing other than the random walk from vertex to vertex in the graph $(X(0), X(1))$, so by assumption on X , $\langle (2DU - I)g, g \rangle \leq \gamma \|g\|^2$ for every function $g \in \mathbb{R}^V$ such that $\mathbb{E}[g] = 0$. ◀

2.4 HDX Codes

An HDX code is defined by two objects

- A k -dimensional simplicial complex X , and a dimension $0 < i \leq k$.
- A collection $\{C_s\}$ of local codes $C_s \subseteq \mathbb{F}_q^{X_{+s}(k)}$, one per face $s \in X(k-i)$.

The HDX code at dimension k is defined as

$$C^k[X, \{C_s\}] = \left\{ f \in \mathbb{F}_q^{X(k)} \mid \forall s \in X(k-i), (f(s \cup v) : v \in X_s(i-1)) \in C_s \right\} \quad (5)$$

When X is one dimensional these are the expander codes of [35].

Example 1: expander code

An expander code is given by an expander graph $X = (V, E)$ and a local code C_v for every vertex $v \in V$ such that $C_v \subseteq \mathbb{F}_2^{\{e \ni v\}}$. A word $f \in \mathbb{F}_2^E$ is in the code if for every vertex v , the bits on the edges touching v form a codeword in a small code C_v . Formally, if $(f(e) : e \ni v) \in C_v$. Often the graph (V, E) is d -regular and the local codes are taken to all be copies of some $C_0 \subset \mathbb{F}_2^d$.

This is a special case of (5) for dimension $k = 1$, where $X(0) = V$ and $X(1) = E$ and $C_v \subseteq \mathbb{F}_2^{X_v(0)}$ via the identification $X_v(0) \leftrightarrow \{e \ni v\}$.

Example 2: cocycle codes

Fix some simplicial complex X , and some dimension $0 < k < \dim(X)$. Suppose for every $s \in X(k-1)$, the local code C_s is taken to be the parity code consisting of all even length words, namely the local code at $s \in X(k-1)$ is,

$$Z_s = \left\{ f \in \mathbb{F}_2^{X_s(0)} \mid \sum_{v \in X_s(0)} f(s \cup v) = 0 \right\}.$$

Then the k dimensional HDX code $C[X, \{Z_s\}]$ coincides with the space of k -cocycles of X . For example, when $k = 1$, the code is spanned by all closed walks.

2.5 Local-Testability and Agreement-Testability

For a function $f : A \rightarrow B$ and a subset $A' \subset A$ we denote by $f|_{A'} : A' \rightarrow B$ the restriction of f to A' .

Let X be a k -dimensional simplicial complex and assume that for every $v \in X(0)$ we are given a local code $C_v \subseteq \mathbb{F}_q^{X_{+v}}$. Let $C = C^k[X, \{C_v\}] \subseteq \mathbb{F}_q^{X(k)}$ be an HDX code.

► **Definition 11 (Locally testable code).** Let $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a strictly increasing function with $\rho(0) = 0$, and let $\epsilon > 0$. A code $C \subseteq \mathbb{F}_q^n$ is a $(Q, \epsilon, \rho(\cdot))$ -locally testable code if there is a randomized tester that, upon receiving a given word $f \in \mathbb{F}_q^n$, queries f in at most Q locations and then accepts or rejects, such that if $p = \mathbb{P}[\text{Tester rejects } f] \leq \epsilon$ then $\text{dist}(f, C) \leq \rho(p)$.

We specialize this definition to the case of HDX codes by considering a “canonical” local tester that selects a random vertex v and checks if the restriction of the codeword to the star of v is in the local code C_v . Namely, if $f|_{X_{+v}} \in C_v$. The number of queries made by this tester is equal to the maximal number of k -faces containing a vertex v .

27:12 New Codes on High Dimensional Expanders

► **Definition 12** (Local testability of HDX Codes). *Let $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a strictly increasing function with $\rho(0) = 0$, and let $\epsilon > 0$. The HDX code $C = \mathcal{C}^k[X, \{C_v\}]$ is $(\epsilon, \rho(\cdot))$ -locally testable if for any $f \in \mathbb{F}_q^{X(k)}$, denoting $p = \mathbb{P}_v[f|_{X_{+v}} \notin C_v]$, the following holds. If $p \leq \epsilon$ then*

$$\text{dist}(f, C) \leq \rho(p).$$

► **Remark 13.** If an HDX code is locally testable as per Definition 12, then the number of queries made by the tester is at most $\max_v |X_{+v}(k)|$ which is the maximal number of k -faces containing a vertex v . In a bounded-degree complex X this is a constant number. Moreover, if each C_v is an LDPC, we can reduce the query complexity further (without changing the code), as in the following Claim.

▷ **Claim 14.** Suppose each local code C_v is defined by at most m_0 parity checks, each looking at most q_0 bits. If an HDX code $\mathcal{C}^k[X, \{C_v\}]$ is $(\epsilon, \rho(\cdot))$ -locally testable per Definition 12, then it is $(q_0, \frac{\epsilon}{m_0}, \rho'(\cdot))$ -locally testable as per Definition 11, where $\rho'(x) := \rho(m_0 x)$.

Proof. For any vertex v , $f|_{X_{+v}} \in C_v$ iff none of the m_0 parity checks fail. By union bound, the fraction of vertices v for which at least one of the m_0 parity checks fail is at most $m_0 \cdot \frac{\epsilon}{m_0} = \epsilon$. This means that $\mathbb{P}_v[f|_{X_{+v}} \notin C_v] \leq m_0 p < \epsilon$, and by the local testability per Definition 12 we deduce that $\text{dist}(f, C) \leq \rho(m_0 p)$. ◁

► **Definition 15** (Agreement testability of HDX Codes). *Let $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a strictly increasing function, and let $\epsilon > 0$. The HDX code $\mathcal{C}^k[X, \{C_v\}]$ is called $(\epsilon, \rho(\cdot))$ -agreement testable if, for any given collection $\{z_v \in C_v \mid v \in X(0)\}$, if*

$$\alpha := \mathbb{P}_{uv \in X(1)}[z_v(X_{+uv}(k)) \neq z_u(X_{+uv}(k))] < \epsilon$$

then there exists some $h \in C$ such that

$$\mathbb{P}_{v \in X(0)}(z_v \neq h|_{X_{+v}(k)}) \leq \rho(\alpha).$$

▷ **Claim 16** (Agreement testability implies local testability). Let X be a k -dimensional simplicial complex, and assume that every vertex is contained in the same number of k faces. Let $C = \mathcal{C}^k[X, \{C_v\}]$ be an HDX code. If C is $(\epsilon_0, \rho_0(\cdot))$ -agreement testable then it is $(\epsilon_0/2, \rho_1(\cdot))$ -locally testable, where $\rho_1(p) := \rho_0(2p) + p$.

Proof. Suppose $f \in \mathbb{F}_q^{X(k)}$. Assume that $p = \mathbb{P}_v[f|_{X_{+v}} \notin C_v]$. Let $V^* = \{v \in X(0) \mid f|_{X_{+v}} \in C_v\}$. Set $z_v = \begin{cases} f|_{X_{+v}} & v \in V^* \\ 0 & v \notin V^* \end{cases}$. Choose a random edge $uv \in X(1)$.

The probability that either u or v are not in V^* is at most $2p$. In the remaining probability they surely agree, so the disagreement is at most $\alpha \leq 2p$. If $2p \leq \epsilon_0$, then by the assumption on agreement testability, there must be some codeword $h \in C$ such that $\mathbb{P}_v[z_v \neq h|_{X_{+v}}] \leq \rho_0(\alpha) \leq \rho_0(2p)$. The codeword h disagrees with f on some X_{+v} either when $v \notin V^*$ or when $z_v \neq h|_{X_{+v}}$. This event is upper bounded by $p + \rho_0(2p) = \rho_1(p)$. By assumption every vertex is contained in the same number of k -faces. So choosing a random vertex and then a random k -face containing it, is the same as choosing a random k -face. Therefore,

$$\text{dist}(f, C) \leq \text{dist}(f, h) = \mathbb{P}_{s \in X(k)}[f(s) \neq h(s)] \leq \mathbb{P}_{v \in X(0)}[v \notin V^*] + \mathbb{P}_{v \in X(0)}[z_v \neq h|_{X_{+v}}] \leq \rho_1(p).$$

◁

Since there has been some discussion of this in the literature [10], we spell out how any HDX code that is agreement testable automatically gives rise to the “local-view” code that is two-query testable. The idea is to move from a codeword $f \in C$ to the collection $\{z_v\}_{v \in X(0)}$ of local views where $z_v = f|_{X_{+v}}$.

▷ **Claim 17 (Agreement testability implies 2-query LTCs).** Suppose C is an HDX code. Let Σ be a finite alphabet such that $|\Sigma| = \max_v |C_v|$, and fix an injection $\sigma_v : C_v \rightarrow \Sigma$ for each $v \in X(0)$. Define

$$LC = \{z : X(0) \rightarrow \Sigma \mid \exists f \in C, \text{ s.t. } z(v) = \sigma_v(f|_{X_{+v}}) \forall v \in X(0)\}. \quad (6)$$

If C is agreement testable, then LC is locally testable with two queries. If C is $(\epsilon, \rho(\cdot))$ -agreement testable, then LC is $(2, \epsilon, \rho'(\cdot))$ -locally testable per Definition 11, where $\rho'(p) = p + \rho(p)$.

Proof. We only give a sketch. Given $z \in LC$, the tester will select a random edge uv and read $z(u), z(v) \in \Sigma$. It will interpret $z(u), z(v)$ as words in C_u, C_v respectively. This is done by computing $z_v = \sigma_v^{-1}(z(v))$ and similarly $z_u = \sigma_u^{-1}(z(u))$. This inversion may fail since σ_v is an injection but not necessarily a bijection, in which case the tester rejects. Otherwise, the tester will accept iff $z_u|_{X_{+uv}} = z_v|_{X_{+uv}}$. The analysis follows from the definition of agreement testability: define $\hat{z}_v = \sigma_v^{-1}(z(v))$ for each $v \in X(0)$, where if the inversion fails we define $\hat{z}_v = 0 \in C_v$, then the probability $p \leq \epsilon$ the tester fails is at least $\mathbb{P}_{uv \in X(0)} [\hat{z}_u|_{X_{+uv}} \neq \hat{z}_v|_{X_{+uv}}] =: \alpha$, which by Definition 15 means that there is some $x \in C$ such that $\mathbb{P}_{v \in X(0)} [\hat{z}_v \neq x|_{X_{+v}(k)}] \leq \rho(\alpha) \leq \rho(p)$. Define $y : v \rightarrow \sigma_v(x|_{X_{+v}}) \in LC$, so that $\mathbb{P}_{v \in X(0)} [\sigma_v(\hat{z}_v) \neq y(v)] \leq \rho(p)$.

Now note that

$$\begin{aligned} \text{dist}(z, LC) &= \min_{lc \in LC} \mathbb{P}_{v \in X(0)} [z(v) \neq lc(v)] \\ &\leq \mathbb{P}_{v \in X(0)} [z(v) \neq y(v)] \\ &\leq \mathbb{P}_{v \in X(0)} [z(v) \notin \sigma_v(C_v)] + \mathbb{P}_{v \in X(0)} [\sigma_v(\hat{z}_v) \neq y(v)] \\ &\leq p + \rho(p), \end{aligned}$$

where in the last line we used that $\mathbb{P}_{v \in X(0)} [z(v) \notin \sigma_v(C_v)] \leq \mathbb{P}_{uv \in X(1)} [z(u) \notin \sigma_u(C_u) \vee z(v) \notin \sigma_v(C_v)] \leq p$. \triangleleft

2.6 Sheaves and HDX codes

Meshulam shows in [28] how to view the expander codes of [35] as a twisted homology of a graph with certain local coefficients. He also describes a higher dimensional generalization of systems of local coefficients attached to higher dimensional simplicial complexes. First and Kaufman [10] focus on the cohomological (as opposed to homological) variant and give a framework for studying codes as sheaves. The HDX codes we have defined can be placed in this framework, as we briefly explain next.

An \mathbb{F} -sheaf $\mathcal{F}X$ over a simplicial complex X is a collection of \mathbb{F} -vector spaces $F(x)$ for every $x \in X$, together with linear maps $\text{Res}_{t \leftarrow s} : F(s) \rightarrow F(t)$ for every pair of faces $s \subset t$. These maps are called *restriction maps*, and are required to satisfy certain transitive consistency, see more details in [10].

A k -dimensional HDX code naturally gives rise to a sheaf as follows.

► **Definition 18.** Let \mathbb{F} be a field, let X be a k -dimensional simplicial complex, and let $\{C_s \subseteq \mathbb{F}^{X_{+s}}\}_{s \in X(k-1)}$ be a collection of \mathbb{F} -linear local codes. Define a sheaf over X with respect to $\{C_s\}$ to be

- $F(t) = \mathbb{F}$ for every $t \in X(k)$,
- $F(s) = C_s \subseteq \mathbb{F}^{X_{+s}(k)}$ for every $s \in X(k-1)$,
- $F(r) = \{f \in \mathbb{F}^{X_{+r}(k)} \mid \forall s \in X_{+r}(k-1), f|_{X_{+s}(k)} \in C_s\}$ for every $r \in X(i)$ and every i .
In other words, $F(r)$ is isomorphic to the HDX code defined over X_r with a local code appropriately isomorphic to C_s at a face $s \setminus r$,
- Since $X_{+s} \subset X_{+r}$ whenever $s \supset r$, we can define the restriction maps by actual restriction.

The coboundary operator from vertices to edges is $\delta : \oplus_{v \in X(0)} F(v) \rightarrow \oplus_{e \in X(1)} F(e)$ is defined by $\delta f(e) = \sum_{v \in e} \text{Res}_{e \leftarrow v} f(v)$ (and the boundary operator is defined as the dual). For full definitions of the coboundary and boundary operators, see [10] (the coboundary operators are defined in Section 4 and the boundary operators in Section 7.4).

▷ **Claim 19.** Let \mathbb{F} be a field, let X be a k -dimensional simplicial complex, and let $\{C_s \subseteq \mathbb{F}^{X_{+s}}\}_{s \in X(k-1)}$ be a collection of \mathbb{F} -linear local codes. Let $C_s^\perp = \{f \in \mathbb{F}^{X_{+s}} \mid f \perp C_s\}$ be the code dual to C_s , and let $\mathcal{F}X^\perp$ be the sheaf over X with respect to $\{C_s^\perp\}$. Then, letting $\partial_k : C^k(X, \mathcal{F}X^\perp) \rightarrow C^{k-1}(X, \mathcal{F}X^\perp)$ denote the k -th boundary operator, $Z^k = \text{Ker} \partial_k$ is the HDX code $\mathcal{C}^k[X, \{C_s\}]$. ◀

Moreover, for the sheaf $\mathcal{F}X$ with respect to $\{C_s\}$, the cocycle code $Z_0 = \text{Ker} \delta_0$ is the related “local-views” code defined in (6), given by replacing each codeword with its ensemble of local views (see the definition in (6) and the ensuing discussion). Agreement testability of our code is equivalent to cosystolic expansion of the sheaf $\mathcal{F}X$ in dimension 0, as proven in [10, Proposition 7.6].

Finally, First and Kaufman describe in [10, Section 7.4] how a sheaf gives rise to a quantum CSS code.

3 Coset Complex and Code

3.1 The Triangle Complex

We describe a family of simplicial complexes $\{X_n\}_n$. The construction is a variant of the coset complexes constructed by Kaufman and Oppenheim [19].

Let $\mathbb{F} = \mathbb{F}_q$ be a fixed finite field. Let $\varphi \in \mathbb{F}_q[t]$ be a primitive (and irreducible) polynomial of degree n and let $R_n = \mathbb{F}_q[t]/\langle \varphi \rangle \cong \mathbb{F}_{q^n}$ (i.e. the ring of univariate polynomials of degree $\leq n-1$ where multiplication is done modulo φ). Further assume that we choose n so that $3 \nmid q^n - 1$ (this is easy as long as $q \not\equiv 1 \pmod{3}$).

We define three matrix groups

$$\begin{aligned} K_1 &= \left\{ \begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} \in M_3(R_n) \right\}, \\ K_2 &= \left\{ \begin{pmatrix} 1 & 0 & 0 \\ ct^2 & 1 & at \\ bt & 0 & 1 \end{pmatrix} \in M_3(R_n) \right\}, \\ K_3 &= \left\{ \begin{pmatrix} 1 & at & 0 \\ 0 & 1 & 0 \\ bt & ct^2 & 1 \end{pmatrix} \in M_3(R_n) \right\}. \end{aligned} \tag{7}$$

and let $G = G_n$ be the group generated by K_1, K_2, K_3 . Clearly $G_n \subseteq SL_3(R_n)$. We will show that for ϕ and n as we specified above, it will hold that $G_n = SL_3(R_n)$.

We define three additional (smaller) subgroups,

$$\begin{aligned} H_1 &= K_2 \cap K_3 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha t & 0 & 1 \end{pmatrix} \mid \alpha \in \mathbb{F}_q \right\}, \\ H_2 &= K_1 \cap K_3 = \left\{ \begin{pmatrix} 1 & \alpha t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mid \alpha \in \mathbb{F}_q \right\}, \\ H_3 &= K_1 \cap K_2 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha t \\ 0 & 0 & 1 \end{pmatrix} \mid \alpha \in \mathbb{F}_q \right\}. \end{aligned} \quad (8)$$

▷ **Claim 20.** Each subgroup H_1, H_2, H_3 is isomorphic to the abelian group $(\mathbb{F}_q, +)$ via the isomorphism $\alpha \leftrightarrow h_i(\alpha)$ for $h_i(\alpha) \in H_i$ the matrix with αt in the appropriate location. ◀

The coset complex considered in this paper is

$$X = X[G; K_1, K_2, K_3] \quad (9)$$

as per Definition 8. The group $G = G_n$ depends on the underlying ring $R = R_n$. When we let the size of the ring R_n grow by increasing the degree n , we get a family of complexes X_n as required.

By definition X is a 3-partite simplicial complex satisfying the following,

- ▷ **Claim 21.** Let $G, K_1, K_2, K_3, H_1, H_2, H_3$ be as above. Let $X = X[G; K_1, K_2, K_3]$. Then
1. The vertices correspond to cosets of K_i : $X(0) \cong G/K_1 \sqcup G/K_2 \sqcup G/K_3$. Each vertex is contained in $q^3 = |K_i|$ triangles.
 2. The edges of X connect a vertex $u = g_i K_i$ to a vertex $v = g_j K_j$ iff $i \neq j$ and the cosets intersect. In this case their intersection corresponds to a coset of H_k for $k \neq i, j$. The elements of this coset are in 1-1 correspondence with the set T_{uv} of triangles containing the edge uv . There are exactly $|H_k| = q$ such triangles.
 3. $X(2) \cong G$. Moreover, given three vertices $u = g_1 K_1, v = g_2 K_2, w = g_3 K_3$, the triangle uvw belongs to $X(2)$ iff the three cosets have a nonempty intersection $g_1 K_1 \cap g_2 K_2 \cap g_3 K_3 = \{g\}$.
 4. Assuming that φ is primitive and $3 \nmid q^n - 1$, then $G = \text{SL}_3(R_n)$.

Proof. The first three items are properties of coset complexes, as in [19]. We prove the last item:

Let $e_{ij}(\alpha)$ denote the matrix with 1's along the diagonal and α in the (i, j) 'th position. (Note that $h_1(\alpha) = e_{12}(\alpha t)$, $h_2(\alpha) = e_{23}(\alpha t)$, and $h_3(\alpha) = e_{31}(\alpha t)$.) The following so-called Steinberg relations hold for all $i \neq j \neq k$,

$$[e_{ij}(\alpha), e_{jk}(\beta)] = e_{ik}(\alpha\beta),$$

where $[g, h] = ghg^{-1}h^{-1}$ denotes the commutator (this can be verified by direct calculation).

We claim that using $h_1(1) = e_{12}(t)$, $h_2(1) = e_{23}(t)$, and $h_3(1) = e_{31}(t)$ it's possible to generate all matrices $e_{ij}(t^\beta)$ where $(i, j) \in \{(1, 2), (2, 3), (3, 1)\}$ and $\beta \equiv 1 \pmod{3}$, as well as where $(i, j) \in \{(1, 3), (2, 1), (3, 2)\}$ and $\beta \equiv 2 \pmod{3}$. We will prove this via induction.

Suppose that for some $B \in \mathbb{N}_{\geq 0}$ all $e_{ij}(t^\beta)$ with $\beta \leq B, \beta \equiv j - i \pmod 3$ are generatable. Then, we will show the claim for $B + 1$. If $B + 1$ is a multiple of 3, then there's nothing to show. Otherwise, if $B + 1$ is 1 mod 3, then by assumption we have that $e_{ik}(t^2)$ and $e_{kj}(t^{B-1})$ are both generatable where $k - i \equiv j - k \equiv 2 \pmod 3$, so $e_{ij}(t^{B+1}) = [e_{ik}(t^2), e_{kj}(t^{B-1})]$ is also generatable. If $B + 1$ is 2 mod 3, then we have that $e_{ik}(t)$ and $e_{kj}(t^B)$ are both generatable where $k - i \equiv j - k \equiv 1 \pmod 3$ by inductive hypothesis, so $e_{ij}(t^{B+1}) = [e_{ik}(t), e_{kj}(t^B)]$ are also both generatable.

Now, since φ is primitive, it holds that the elements t, \dots, t^{q^n-1} are all distinct and thus must cover all of $R_n \setminus \{0\}$. If $3 \nmid q^n - 1$, then the elements $t^{3\beta+1}$ where $0 \leq \beta < q^n - 1$ are also distinct and range over $R_n \setminus \{0\}$, as do the elements $t^{3\beta+2}$. The reason is that if t generates the multiplicative group of $R_n = \mathbb{F}_q[t]/\langle \varphi \rangle$, whose order is $q^n - 1$, then as long as $3 \nmid q^n - 1$ then t^3 also generates the group as well. In other words, from $h_1(1), h_2(1), h_3(1)$ we can generate all $e_{ij}(\gamma)$ for any $i \neq j$ and $\gamma \in R_n \setminus \{0\}$ (that is, we can generate all elementary matrices).

To finish, it is well known that if we can generate all elementary matrices then we can generate all of $\text{SL}_3(R_n)$. \triangleleft

This claim proves item (b) in Theorem 1. Item (a) is proven in the next subsection.

3.2 Structure of the Links

In this section we describe the links. That is, we understand the structure of $X(K_i; H_{i+1}, H_{i-1})$. Without loss of generality we focus on the link $G_1 = (K_1/H_2 \sqcup K_1/H_3, E_1)$, a bipartite graph described in more details below.

Recall that the vertices of G_1 corresponds to cosets of the form kH_2 and $k'H_3$ while the edges correpsond to pairs of cosets $\{kH_2, k'H_3\}$ that have non-empty intersection $k_1H_1 \cap k_2H_2$.

For any coset kH_2 with representative $k = \begin{pmatrix} 1 & 0 & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix}$, we use $(*, b, c)$ to denote the vertice

in G_1 . Similarly for any coset $k'H_3$ with representation $k' = \begin{pmatrix} 1 & \alpha t & \gamma t^2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ we use $(\alpha, *, \gamma)$

to denote the corresponding vertice in G_1 . So an edge connects $(*, b, c)$ and $(\alpha, *, \gamma)$ iff the following equation has a solution in \mathbb{F}_q^2 :

$$\begin{pmatrix} 1 & xt & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \alpha t & (\alpha y + \gamma)t^2 \\ 0 & 1 & yt \\ 0 & 0 & 1 \end{pmatrix}. \quad (10)$$

This system of equations is solvable iff $c = \alpha b + \gamma$. Therefore we can deduce the following statement about the degree of G_1 .

\triangleright **Claim 22.** Every vertex in G_1 has degree q .

Furthermore let A be the normalized adjacency matrix of G_1 . By the edge characterization Equation (10), we derive the following spectral gap for A .

\triangleright **Claim 23.** $|\lambda_2(A)| = \frac{1}{\sqrt{q}}$.

Proof. Let B be the $|K_1/H_2| \times |K_1/H_3|$ unnormalized biadjacency matrix of G_1 such that

$$A = \begin{pmatrix} \mathbf{0} & \frac{1}{q} \cdot B \\ \frac{1}{q} \cdot B^T & \mathbf{0} \end{pmatrix}.$$

Then the matrix BB^T is the $|K_1/H_2| \times |K_1/H_2|$ matrix whose value in its $((*, b, c), (*, b', c'))$ -th entry is the number of 2-step walks from $(*, b, c)$ to $(*, b', c')$. Equivalently, by Equation (10), the value is also the number of solutions to the equations $c = xb + y$ and $c' = xb' + y$ over \mathbb{F}_q^2 . Therefore

$$BB^T[(*, b, c), (*, b', c')] = \begin{cases} 1 & \text{when } b \neq b' \\ q & \text{when } b = b' \wedge c = c' \\ 0 & \text{o.w.} \end{cases}$$

We can thus explicitly write $BB^T = (J_q - I_q) \otimes J_q + q \cdot I_q \otimes I_q$ where $J_q \in \mathbb{R}^{q \times q}$ is the all-ones matrix.

So its top eigenvalue satisfies $\lambda_1(BB^T) = q^2$ and the second-largest eigenvalue is $\lambda_2(BB^T) = q$. From this we can deduced that $|\lambda_2(A)| = \frac{1}{\sqrt{q}}$. \triangleleft

3.3 Embedding the Complex into a Vector Space

Recall that there is a natural isomorphism between the group G and the triangles of the complex, $X(2)$. We describe a natural way to biject G to a set of points $S \subset \mathbb{F}_q^{9n}$,

$$X(2) \cong G \xhookrightarrow{\iota} S \subset R^9 \cong \mathbb{F}_q^{9n}.$$

Every $g \in G$ is a 3×3 matrix $(r_{ij})_{1 \leq i, j \leq 3}$ such that $r_{ij} \in R$ for each i, j . An element in R is a univariate polynomial of degree at most $n-1$ so it is specified by n coefficients $r_{ij}(t) = \sum_{\ell=0}^{n-1} r_{ij}^{(\ell)} t^\ell$. We simply map each of the nine matrix entries into a vector of coefficients in \mathbb{F}_q^n and concatenate them all:

$$(r_{ij}) \xhookrightarrow{\iota} \left(r_{11}^{(0)}, r_{11}^{(1)}, \dots, r_{11}^{(n-1)}, r_{12}^{(0)}, r_{12}^{(1)}, \dots, r_{12}^{(n-1)}, \dots, r_{33}^{(0)}, r_{33}^{(1)}, \dots, r_{33}^{(n-1)} \right) \quad (11)$$

This embedding is clearly injective and it is linear in the coefficients of the matrix entries, namely $\iota(\alpha g + \beta g') = \alpha \iota(g) + \beta \iota(g')$ for any $\alpha, \beta \in \mathbb{F}_q$ and $g, g' \in G$.

\triangleright **Claim 24.** For $g \in G$ and $i = 1, 2, 3$ let $\ell_{g,i} : \mathbb{F}_q \rightarrow gH_i$ be defined by $\ell_{g,i}(\alpha) = gh_i(\alpha)$. Then $\iota \circ \ell_{g,i} : \mathbb{F}_q \rightarrow \mathbb{F}_q^n$ is an affine line that can be written as

$$\iota(\ell_{g,i}(\alpha)) = v_0 + \alpha v_i$$

where $v_0 = \iota(g)$ and for $g = \begin{pmatrix} | & | & | \\ g_1 & g_2 & g_3 \\ | & | & | \end{pmatrix}$, we have

$$v_1 = \iota \left(\begin{pmatrix} | & 0 & 0 \\ tg_3 & 0 & 0 \\ | & 0 & 0 \end{pmatrix} \right); \quad v_2 = \iota \left(\begin{pmatrix} 0 & | & 0 \\ 0 & tg_1 & 0 \\ 0 & | & 0 \end{pmatrix} \right); \quad v_3 = \iota \left(\begin{pmatrix} 0 & 0 & | \\ 0 & 0 & tg_2 \\ 0 & 0 & | \end{pmatrix} \right).$$

Here the entries in tg_i are taken modulo $\varphi(t)$, and $v_i \neq 0$ for $i = 1, 2, 3$.

Moreover, for any $g' \in gH_i$, the line $\iota \circ \ell_{g',i}$ is a re-parameterization of $\iota \circ \ell_{g,i}$, satisfying

$$\iota \circ \ell_{g',i}(\alpha) = \iota \circ \ell_{g,i}(\alpha + \alpha')$$

for $\alpha' \in \mathbb{F}$ such that $g' = gh_i(\alpha')$.

Proof. Fix $g \in G$. We prove the first part for $i = 1$, the other cases are similar. The matrix $gh_1(\alpha)$ is obtained from g by adding αt times the third column of g to the first column of g , namely,

$$gh_1(\alpha) = (g_1, g_2, g_3) + (\alpha t \cdot g_3, 0, 0)$$

Since the embedding ι is linear in the coefficients, we get that

$$\iota(gh_1(\alpha)) = \iota((g_1, g_2, g_3)) + \iota((\alpha t \cdot g_3, 0, 0)) = v_0 + \alpha v_1$$

as in the claim.

Regarding the moreover part, by definition $\ell_{g',i}(\alpha) = g'h_i(\alpha) = gh_i(\alpha')h_i(\alpha) = gh_i(\alpha + \alpha') = \ell_{g,i}(\alpha + \alpha')$. Since the expression is linear in α , and since ι is additive, $\iota \circ \ell_{g,i}$ is clearly the same affine line as $\iota \circ \ell_{g',i}$, reparameterized. \triangleleft

We define, for $i = 1, 2, 3$,

$$\mathcal{L}_n^i = \{ \iota \circ \ell_{g,i} : \mathbb{F}_q \rightarrow \mathbb{F}_q^n \mid g \in G \}, \quad (12)$$

and $\mathcal{L}_n = \mathcal{L}_n^1 \cup \mathcal{L}_n^2 \cup \mathcal{L}_n^3$. This establishes item (c) in Theorem 1.

3.4 The Global Code

Let $RS(q, d)$ be the Reed Solomon code of degree d over \mathbb{F}_q . Namely, $RS(q, d)$ is the set of length- q tuples $(p(\alpha) : \alpha \in \mathbb{F}_q)$ where p is a univariate polynomial of degree at most d .

For any three parameters $0 \leq d_1, d_2, d_3 \leq q$, we defined the code C_{d_1, d_2, d_3} in two ways, see Definition 2. First, we defined it as a (punctured) lifted Reed-Solomon code,

$$C^1 = C_{n, d_1, d_2, d_3} = \{ f : \bar{X}_n \rightarrow \mathbb{F}_q \mid \forall i = 1, 2, 3, \ell \in \mathcal{L}_n^i, f \circ \ell \in RS(q, d_i) \} \quad (1)$$

where \mathcal{L}_n^i is the set of affine lines defined in (12), and then we defined it as an HDX code,

$$C^2 = C_{n, d_1, d_2, d_3} = \{ f : X_n(2) \rightarrow \mathbb{F}_q \mid \forall e \in X(1), f|_{X_{+e}} \in C_e \} \quad (2)$$

where C_e is isomorphic to $RS(q, d_i)$ for edges of type i . We now define C_e more explicitly. Recall from Claim 21 (item 2) that every edge e of type i corresponds to a coset gH_i , in the sense that the triangles containing e , which we denote X_{+e} , correspond to the elements of gH_i . Choose, for each edge, one group element g to be a coset representative. We write $X_{+e} = \{gh_i(\alpha)\}_{\alpha \in \mathbb{F}}$ and define

$$C_e = \{ f \in \mathbb{F}_q^{X_{+e}} \mid f(gh_i(\alpha)) \text{ is a degree } d_i \text{ function of } \alpha \}. \quad (13)$$

The definition of C_e appears to depend on the choice of coset representative but it does not, because the degree of $f(gh_i(\alpha))$ as a function of α is the same as the degree of $f(gh_i(\alpha + \alpha')) = f(g'h_i(\alpha))$ as a function of α .

\triangleright **Claim 25.** The codes C^1, C^2 are isomorphic, and $\iota : G \rightarrow \bar{X}_n$ gives the isomorphism.

Proof. Fix $f \in C^1$. We define $\tilde{f} : G \rightarrow \mathbb{F}_q$ by $\tilde{f} = f \circ \iota$, and show $\tilde{f} \in C^2$. We need to check that $\tilde{f}|_{X_{+e}} \in C_e$ for each e . By definition this is true iff $(\tilde{f}(gh_i(\alpha)))_\alpha \in RS(q, d_i)$ for e a type i edge. But $\tilde{f}(gh_i(\cdot)) = \tilde{f} \circ \ell_{g,i} = f \circ \iota \circ \ell_{g,i} \in RS(q, d_i)$ where the last inclusion is because $f \in C^1$ and $\iota \circ \ell_{g,i} \in \mathcal{L}_n^i$.

The other direction is easy as well. Given $\tilde{f} \in C^2$, we let $f = \tilde{f} \circ \iota^{-1} : \bar{X}_n \rightarrow \mathbb{F}_q$ the unique function such that $f \circ \iota = \tilde{f}$. To check that $f \in C^1$ consider any line $\ell = \iota \circ \ell_{g,i} \in \mathcal{L}_n$ and observe that $f \circ \ell = \tilde{f} \circ (\iota \circ \ell_{g,i}) = \tilde{f} \circ \ell_{g,i} \in RS(q, d_i)$. \triangleleft

Since there is a 1 – 1 correspondence between $X(2)$ and G , we may also write codewords of C^2 as $f : G \rightarrow \mathbb{F}_q$. A group always acts transitively on itself by left multiplication. Moreover,

▷ **Claim 26.** If $w : G \rightarrow \mathbb{F}_q$ is a codeword, then w^g is a codeword, where $w^g(g') = w(gg')$

Proof. This is clear since for any $g' \in G$ and any $i = 1, 2, 3$, $(w^g(g'h_i(\alpha)))_\alpha = (w(gg'h_i(\alpha)))_\alpha \in RS_q^{d_i}$. ◀

This establishes the transitivity of the code, as claimed in last item of Theorem 3. It implies that the restriction of our code to K_i is isomorphic to the restriction of our code to any coset gK_i . To study the local view of the code at a link of a vertex it suffices to study its restriction to K_i for $i = 1, 2, 3$.

3.5 Multiplication Property

It is immediate that the codes have the multiplication property by inheritance from the Reed-Solomon code. Recall that for $w, w' \in \mathbb{F}_q^N$ we define $w'' = w \odot w' \in \mathbb{F}_q^N$ by coordinate-wise product: $w''[i] = w[i] \cdot w'[i]$.

▶ **Lemma 27.** Suppose $C = C_{n,d_1,d_2,d_3}$ and $C' = C_{n,d'_1,d'_2,d'_3}$, then for every $w \in C$ and $w' \in C'$, we have $w \odot w' \in C'' = C_{n,d_1+d'_1,d_2+d'_2,d_3+d'_3}$.

Proof. For every $e \in X(1)$ the local views of $w|_{X_{+e}}, w'|_{X_{+e}}$ are Reed-Solomon codewords of degrees d_i, d'_i (relying on the definition in (2)). So the coordinate-wise product, $w''|_{X_{+e}}$, is a Reed-Solomon codeword of degree at most $d_i + d'_i$, as needed. ◀

3.6 Distance

The global code C can also easily be shown to have constant relative distance,

▶ **Lemma 28 (Distance).** If the relative distance of C_e is at least $\delta > 0$ for every $e \in X(1)$ then C has relative distance at least $(\delta - 2\gamma)(\delta - \gamma)\delta$.

Proof. Let $0 \neq x \in C$. Let $V^* = \{v \in X(0) \mid x|_{X_{+v}} \neq 0\}$ and let $v \in V^*$. We will first show that at least $(\delta - \gamma)$ of the edges touching v are non-zero. (An edge e is non zero iff $x(e) \neq 0$.)

Let $A_v = \{u \in X_v(0) \mid x|_{X_{+uv}} \neq 0\}$. Each $u \in A_v$ has $0 \neq x|_{X_{+uv}} \in C_{uv}$ so $x|_{X_{+uv}}$ must have at least δ fraction of non-zero entries. Each of these non-zero entries corresponds to some vertex w such that $uvw \in X_+$ with $x(uvw) \neq 0$ so $w \in A_v$. We found that for each $u \in A_v \subset X_v(0)$, at least δ of its neighbors (inside X_v) are in A_v . Since the graph X_v is a γ -expander, the Alon-Chung lemma (Lemma 5) implies that $|A_v| \geq (\delta - \gamma)|X_v(0)|$.

Observe that every $u \in A_v$ must itself be in V^* , so each $v \in V^*$ has at least $\delta - \gamma$ fraction of its neighbors in V^* . We can again apply Lemma 5, (now using the fact that the graph $(X(0), X(1))$ is a γ -expander, to deduce $|V^*| \geq (\delta - 2\gamma)|X(0)|$. We have seen that each $v \in V^*$ has at least $\delta - \gamma$ fraction of non-zero edges touching it, so the total fraction of non-zero edges is at least $(\delta - 2\gamma)(\delta - \gamma)$. Each such edge touches at least δ non-zero triangles, so the total fraction of non-zero triangles is at least $(\delta - 2\gamma)(\delta - \gamma)\delta$ as claimed. ◀

This along with the fact that q is constant and $\gamma = 1/\sqrt{q}$ from Claim 23 establishes the second item in Theorem 3.

We now prove a generalization of the above distance lemma to higher dimensional HDX codes.

► **Lemma 29** (Distance of k -Dimensional HDX Code). *Let X be a k -dimensional γ -one-sided local expander, and assume that for every $t \in X(k-1)$ we have a code $C_t \subset \{f : X_{+t}(k) \rightarrow \mathbb{F}_q\}$ with minimum relative distance $\geq \delta > 0$. Define, for every $-1 \leq i \leq k-2$ and every face $t \in X(i)$, the code*

$$C_t = \{f : X_{+t}(k) \rightarrow \mathbb{F}_q \mid f|_{X_{+t}} \in C_t \forall t \in X_{+t}(k-1)\}.$$

Then for every $-1 \leq i < k-2$ and every $t \in X(i)$ the code C_t has relative distance $\prod_{j=0}^{k-1-i}(\delta - j\gamma)$. In particular, the code $C = C_\emptyset$ on the entire complex has relative distance $\prod_{j=0}^k(\delta - j\gamma)$.

Before giving the proof we remark that the distance in the lemma decays exponentially with the dimension, assuming that $\gamma \ll \delta$. This is necessary, as can be seen from the example the $(k+1)$ -dimensional tensor code $C^{\otimes k+1}$, for C any code with distance δ . This code has distance δ^{k+1} and it can be viewed as an HDX code on the complete $k+1$ -partite k -dimensional complex³.

Proof. Let $0 \neq x \in C_\emptyset$. For all $-1 \leq i \leq k-1$ and $s \in X(i)$ we define $A_s = \{u \in X_s(0) \mid x|_{X_{+(s \cup \{u\})}} \neq 0\}$. We claim that if $x|_{X_{+s}} \neq 0$, then $|A_s| \geq (\delta - (k-1-i)\gamma) |X_s(0)|$, and furthermore that C_s has relative distance $\prod_{j=0}^{k-1-i}(\delta - j\gamma)$.

To see this, we proceed by (downwards) induction on i . This is clearly the case for $i = k-1$. Now for $i < k-1$, let $t \in X(i)$. For $v \in X_t(0)$ such that $t \cup \{v\} \in A_t$, we have that each $u \in A_{t \cup \{v\}}$ has $0 \neq x|_{X_{+(t \cup \{u, v\})}} \in C_{t \cup \{u, v\}}$ so $t \cup \{u\} \in A_t$ as well. By the inductive hypothesis, we have that $|A_{t \cup \{v\}}| \geq (\delta - (k-2-i)\gamma) |X_{t \cup \{v\}}(0)|$ for all $v \in V_t^*$. Since X_t is a γ -expander, the Alon-Chung lemma (Lemma 5) implies that $|A_t| \geq (\delta - (k-2-i)\gamma - \gamma) |X_t(0)| = (\delta - (k-1-i)\gamma) |X_t(0)|$.

Now, to compute the distance, we have that each $t \in A_v$ has a $\delta - (k-1-i)\gamma$ fraction of non-zero $(i+1)$ -faces touching it, each of which has a $\prod_{j=0}^{k-2-i}(\delta - j\gamma)$ fraction of k -faces touching it, so the total fraction of non-zero k -faces in X_{+t} is at least $\prod_{j=0}^{k-1-i}(\delta - j\gamma)$. ◀

3.7 Local Code at a Vertex

For each $v \in X(0)$, let

$$C_v = \{f \in \mathbb{F}_q^{X_{+v}} \mid \forall e \ni v, f|_{X_{+e}} \in C_e\}. \quad (14)$$

It is easy to see that our code can be written as

$$C = \left\{ f \in \mathbb{F}^{X(2)} \mid \forall v, f|_{X_{+v}} \in C_v \right\}$$

because we are simply aggregating the constraints differently than in (2).

What does C_v look like when moving to \bar{X}_n ?

► **Lemma 30.** *Fix $v \in X(0)$ a vertex of type i .*

- $\iota(X_{+v})$ is a 3 dimensional affine subspace in \bar{X}_n .
- The code $C_v \subset \mathbb{F}_q^{X_{+v}}$ is isomorphic to C_{d_x, d_y} for $d_x = d_{i+1}$, $d_y = d_{i-1}$, where we define C_{d_x, d_y} by

$$C_{d_x, d_y} = \left\{ f : \mathbb{F}_q^3 \rightarrow \mathbb{F}_q \mid \forall a, b, c, \deg_x(f(x, b, c)) \leq d_x; \deg_y(f(a, y, ay + c)) \leq d_y \right\}. \quad (15)$$

³ There is a natural identification of $[n]^{k+1}$ with the faces of this complex. The link of every $k-1$ face is identified with a row in the appropriate direction.

Proof. Fix first $v = gK_1 = K_1$ given by $g = id$. The elements of $\iota(K_1)$ are

$$\left\{ \iota \left(\begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} \right) \mid a, b, c \in \mathbb{F}_q \right\}$$

and when we range over all possible choices of a, b, c we get an \mathbb{F}_q -linear subspace. If $v = gK_1$ for some $g \notin K_1$, every element becomes

$$\begin{pmatrix} | & | & | \\ g_1 & g_2 & g_3 \\ | & | & | \end{pmatrix} \cdot \begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} | & | & | \\ g_1 & atg_1 + g_2 & ct^2g_1 + btg_2 + g_3 \\ | & | & | \end{pmatrix}$$

which after embedding into \mathbb{F}_q^n becomes $\iota(g_1, g_2, g_3) + a\iota(0, tg_1, 0) + b\iota(0, 0, tg_2) + c\iota(0, 0, t^2g_1)$. This is a 3 dimensional affine subspace. A similar proof applies to vertices of type 2, 3.

For the second item, we focus again on $v = K_1$. The code C_v consists of all $f \in \mathbb{F}_q^{X+v}$ that, for each $e \ni v$, satisfy $f|_{X+e} \in C_e$. We identify functions on $X+v$ with functions on \mathbb{F}_q^3 through

the isomorphism $\mathbb{F}_q^3 \rightarrow K_1$ given by $(a, b, c) \mapsto \begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix}$. An edge $e \ni v$ corresponds to

a coset of H_2 or H_3 in K_1 , say gH_2 , given by a coset representative $g = \begin{pmatrix} 1 & 0 & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix}$. The

group elements of the coset are $gh_2(x)$ for all $x \in \mathbb{F}_q$,

$$\left\{ \begin{pmatrix} 1 & 0 & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & xt & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & xt & ct^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} \mid x \in \mathbb{F}_q \right\},$$

each corresponding to an triangle of $X+e$. So, by definition of C_e (see (13)) the constraint $f|_{X+e} \in C_e$ translates to $f(gh_2(x))$ having degree d_2 in x . In other words, $f(x, b, c)$ must have degree at most d_2 in x .

Similarly, suppose the edge e is gH_3 for $g = \begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The group elements of the

coset are $gh_3(y)$ for all $y \in \mathbb{F}_q$, where

$$\left\{ \begin{pmatrix} 1 & at & ct^2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & yt \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & at & (ay+c)t^2 \\ 0 & 1 & yt \\ 0 & 0 & 1 \end{pmatrix} \mid y \in \mathbb{F}_q \right\}.$$

The constraint $f|_{X+e} \in C_e$ translates to requiring that $f(a, y, ay+c)$ have degree is at most d_3 in y .

It is now clear that C_v is isomorphic to C_{d_2, d_3} when $v = K_1$. The same also holds for any $v = gK_1$ since by Claim 26 the code is invariant under the action of G . This implies that $C_v \cong C_{v'}$, for any v, v' of the same color (since the group action moves any K_i coset to any other K_i coset, it is thus transitive on each color class).

To complete the proof we check that for $v = K_2$ we have $C_v \cong C_{d_3, d_1}$ and for $v = K_3$ $C_v \cong C_{d_1, d_2}$.

Let us start with $v = K_2$. Our requirement is that $f : K_2 \rightarrow \mathbb{F}_q$ evaluates to a degree d_1 polynomial on cosets of $H_1 < K_2$ and a degree d_3 polynomial on cosets of $H_3 < K_2$.

Fix some $g = \begin{pmatrix} 1 & 0 & 0 \\ ct^2 & 1 & at \\ bt & 0 & 1 \end{pmatrix} \in K_2$. For any element $g_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & xt \\ 0 & 0 & 1 \end{pmatrix} \in H_3$, and

$g_y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ yt & 0 & 1 \end{pmatrix} \in H_1$, we have

$$gg_x = \begin{pmatrix} 1 & 0 & 0 \\ ct^2 & 1 & (a+x)t \\ bt & 0 & 1 \end{pmatrix}, \quad gg_y = \begin{pmatrix} 1 & 0 & 0 \\ (c+ay)t^2 & 1 & at \\ (b+y)t & 0 & 1 \end{pmatrix}.$$

Writing now $f(a, b, c) = f\left(\begin{pmatrix} 1 & 0 & 0 \\ ct^2 & 1 & at \\ bt & 0 & 1 \end{pmatrix}\right)$, we require that for all a, b, c ,

- $f(a+x, b, c)$ must have degree d_3 in x ; and
- $f(a, b+y, c+ay)$ must have degree at most d_1 in y .

This is clearly equivalent to requiring

- $f(x, b, c)$ must have degree d_3 in x for all b, c ; and
- $f(a, y, c+ay)$ must have degree at most d_1 in y for all a, c .

Namely, it is equivalent to requiring $f \in C_{d_3, d_1}$

Finally, for $v = K_3$, we fix some $g = \begin{pmatrix} 1 & at & 0 \\ 0 & 1 & 0 \\ bt & ct^2 & 1 \end{pmatrix} \in K_3$. For any element $g_x =$

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ xt & 0 & 1 \end{pmatrix} \in H_1$, and $g_y = \begin{pmatrix} 1 & yt & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in H_2$, we have

$$gg_x = \begin{pmatrix} 1 & at & 0 \\ 0 & 1 & 0 \\ (b+x)t & ct^2 & 1 \end{pmatrix}, \quad gg_y = \begin{pmatrix} 1 & (y+a)t & 0 \\ 0 & 1 & 0 \\ bt & (yb+c)t^2 & 1 \end{pmatrix}.$$

Writing now $f(a, b, c) = f\left(\begin{pmatrix} 1 & at & 0 \\ 0 & 1 & 0 \\ bt & ct^2 & 1 \end{pmatrix}\right)$, we require that for all a, b, c ,

- $f(a, b+x, c)$ must have degree d_1 in x ; and
- $f(a+y, b, c+by)$ must have degree at most d_2 in y .

This is clearly equivalent to requiring that $f \in C_{d_1, d_2}$.

Moving to cosets gK_i , by the fact that the code is invariant under the group action, for any coset gK_i , the local code is isomorphic to $C_{d_{i+1}, d_{i-1}}$ ◀

4 Rate

4.1 Rate of Global Code

We analyze the rate of our code in two regimes. The first, is when the relative rate of the local codes C_e is at least $2/3$. In this case a standard constraint counting implies that the global code has constant relative rate.

► **Lemma 31.** *Suppose $\dim(C_e) > (2/3 + \epsilon)q$ for each $e \in X(1)$. Then $\dim(C) > 3\epsilon$.*

The second parameter regime is when the relative rate of the local codes C_e is arbitrarily small. In this case we give a non-trivial lower bound by demonstrating a collection of linearly independent codewords. These are

$$C' = \{f|_S \mid f : \mathbb{F}_q^{9m} \rightarrow \mathbb{F}_q \text{ has degree} \leq d\}$$

where $d = \min(d_1, d_2, d_3)$ and $S \subset \mathbb{F}_q^{9m}$ is the image of G when embedded into the vector space, see (11).

► **Lemma 32.** $\dim(C') \geq \dim(RM_d^{3m})$. If we choose $|\mathbb{F}_q| \approx \text{poly}(m)$ we get polynomial rate.

Proof. Consider the upper triangular matrices with 1 on the diagonal. This set of matrices belongs to S and is isomorphic to \mathbb{F}^{3m} , so any polynomial in $9m$ variables that depends only on these $3m$ variables and has total degree at most d gives rise to a distinct codeword in C' . ◀

An alternative way to bound the rate is as follows. If the ring R is a field (which by Claim 21 is true whenever φ is a primitive polynomial), then the fraction of matrices with determinant 0 is about $1/|R|$, which is tiny. Moreover, let $S_r = \{m \in M_{3 \times 3}(R) \mid \det(m) = r\}$. Since every $0 \neq r \in R$ has an inverse, there is a bijection between S_{r_1} and S_{r_2} for every $r_1, r_2 \neq 0$,

given by $m_1 \leftrightarrow m_2 = \begin{pmatrix} r_2 r_1^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot m_1$. In particular, our set $S = S_1$ bijects to each S_r .

Each S_r for $r \neq 0$ supports a copy of our complex, obtained by a linear transformation of the entire vector space which moves shifts the lines. We leave open the question of obtaining better bounds on the global rate.

4.2 Rate of Local Code at a Vertex

In this section we analyse the rate of the code $C_{q,d_x,d_y} \subseteq \{f : \mathbb{F}_q^3 \rightarrow \mathbb{F}_q\}$ defined in (15). In this section, we will only consider the case where q is a prime p . We remark that a proof of the rate in the case of $d_x = d_y$ was given in an earlier work [16]. However, in our case, we will need a stronger statement about the nonexistence of monomials above a certain degree in order to prove local testability in Section 5.1, which is given in Lemma 34.

Since $f(x, b, c)$ lies on a degree d_x polynomial in x for any b, c , we can write $f(x, y, z)$ as a polynomial that is degree d_x in x and degree $p - 1$ in y and z :

$$f(x, y, z) = \sum_{\substack{0 \leq i \leq d_x \\ 0 \leq j, k \leq p-1}} c_{ijk} x^i y^j z^k.$$

Plugging in $(x, y, xy + z)$, we get that

$$f(x, y, xy + z) = \sum_{\substack{0 \leq i \leq d_x \\ 0 \leq j, k \leq p-1}} c_{ijk} x^i y^j (xy + z)^k.$$

We can reduce $f(x, y, xy + z)$ modulo $y^p - y$ to get a polynomial $g(x, y, z)$ that is degree $\leq p - 1$ in y and z , and $\leq d_x + p - 1$ in x . We let $g_j(x, z)$ and $g_{jk}(x)$ denote the coefficient of y^j and $y^j z^k$ respectively, so that

$$g(x, y, z) = \sum_{0 \leq j \leq p-1} g_j(x, z) y^j = \sum_{0 \leq j, k \leq p-1} g_{jk}(x) y^j z^k.$$

Note that we can also write

$$g_j(x, z) = \begin{cases} h_j(x, z) & j = 0 \\ h_j(x, z) + h_{j+p-1}(x, z) & j \neq 0 \end{cases}$$

$$g_{jk}(x) = \begin{cases} h_{j,k}(x) & j = 0 \\ h_{j,k}(x) + h_{j+p-1,k}(x) & j \neq 0 \end{cases}.$$

The condition that $f(a, y, ay + c)$ lies on a degree d_y polynomial in y for all a and c means that for all $d_y < j \leq p-1$, it holds that $g_j(x, z) = 0$ for all x, z . In particular, for any $0 \leq k \leq p-1$, it must hold that $g_{jk}(x) = 0$ for all x . So, if we know that $g_{jk}(x)$ is degree $\leq p-1$ in x , then in fact all coefficients in $g_{jk}(x)$ must be 0. We will use this fact extensively in the analysis of the rate.

First, we will need the following lemma.

► **Lemma 33** ([12]). *Assuming that $r \leq k \leq m < p$, the following matrix has full rank in \mathbb{F}_p :*

$$\begin{pmatrix} \binom{m}{k} & \binom{m}{k-1} & \cdots & \binom{m}{k-r} \\ \binom{m-1}{k} & \binom{m-1}{k-1} & \cdots & \binom{m-1}{k-r} \\ \vdots & \vdots & & \vdots \\ \binom{m-r}{k} & \binom{m-r}{k-1} & \cdots & \binom{m-r}{k-r} \end{pmatrix}$$

Proof. We divide the entries of row i by $(m-i)!$ (where the top row is row 0), and multiply the entries of column j by $(m-k+j)!(k-j)!$ (where the leftmost column is column 0). Since $p > m$ and $m \geq k \geq r \geq i, j$, both $(m-i)!$ and $(m-k+j)!(k-j)!$ are nonzero in \mathbb{F}_p . We obtain that the rank of the above matrix is equivalent to the rank of the below matrix:

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ h_1(m-k) & h_1(m-k+1) & h_1(m-k+2) & \cdots & h_1(m-k+r) \\ h_2(m-k) & h_2(m-k+1) & h_2(m-k+2) & \cdots & h_2(m-k+r) \\ \vdots & \vdots & \vdots & & \vdots \\ h_r(m-k) & h_r(m-k+1) & h_r(m-k+2) & \cdots & h_r(m-k+r) \end{pmatrix},$$

where $h_i(x) = x(x-1)(x-2) \cdots (x-i+1)$ is a degree i polynomial.

This matrix is invertible. To see this, let $\alpha_j = m - k - j$. The above matrix is rewritten as

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ h_1(\alpha_0) & h_1(\alpha_1) & h_1(\alpha_2) & \cdots & h_1(\alpha_r) \\ h_2(\alpha_0) & h_2(\alpha_1) & h_2(\alpha_2) & \cdots & h_2(\alpha_r) \\ \vdots & \vdots & \vdots & & \vdots \\ h_r(\alpha_0) & h_r(\alpha_1) & h_r(\alpha_2) & \cdots & h_r(\alpha_r) \end{pmatrix}.$$

Using the smaller degree monomials in the rows above it, the polynomial h_i in each row can iteratively be made into simply the monomial x^i . In particular, the rank of the above matrix is the same as the rank of the Vandermonde matrix, which has full rank. ◀

► **Lemma 34.** *Assuming that $p \geq d_x + d_y + 2$, then for $j + k > d_x + d_y$ and for all $0 \leq i \leq d_x$, it holds that $c_{ijk} = 0$. In other words, the combined degree of y and z in $f(x, y, z)$ is at most $d_x + d_y$.*

Proof. Recall that we've written

$$f(x, y, z) = \sum_{0 \leq i \leq d_x} c_{ijk} x^i y^j z^k.$$

We will backwards induct on the value of $j + k$ to show that $c_{ijk} = 0 \forall i \in [0, d_x]$. The base case is $j + k = 2p - 1$. In this case, there are no terms with $j + k = 2p - 1$, so $c_{ijk} = 0$ for all $i \in [0, d_x]$.

Now, assume that for $s > d_x + d_y$ and $j + k > s$, it holds that $c_{ijk} = 0 \forall i \leq d_x$. We will prove that for all $j + k = s$, it holds that $c_{ijk} = 0$ for $i \in [0, d_x]$.

To begin, we express the polynomial $g_{jk}(x)$ in terms of the nonzero coefficients c_{ijk} .

▷ **Claim 35.** Assume that $c_{ij'k'} = 0$ for $j' + k' > s$ and $i \in [0, d_x]$. Then for $j + k = s$, it holds that

$$g_{jk}(x) = \sum_{\substack{0 \leq i \leq d_x \\ k \leq k' \leq p-1}} c_{i, s-k', k'} \cdot \binom{k'}{k} x^{k'-k+i}.$$

Proof. Consider a term $c_{ij'k'} x^i y^{j'} (xy + z)^{k'}$ that appears in $f(x, y, xy + z)$. This expands to

$$c_{ij'k'} \cdot \sum_{\substack{0 \leq i \leq d_x \\ 0 \leq k \leq k'}} \binom{k'}{k} x^{k'-k+i} y^{j'+k'-k} z^k.$$

Thus, the coefficient $c_{ij'k'}$ appears in the expression of $g_{jk}(x)$ when $(j' + k' - k \bmod^* p) = j$. Note that all the sum of the exponents of y and z in every term of the above expression is $(j' + k' - k) + k = j' + k'$. Then, the coefficients that appear in the expression of $g_{jk}(x)$ must satisfy the property that $j' + k' \in \{s, s + p - 1\}$. Since we've assumed that $c_{ij'k'} = 0$ for $j' + k' > s$, it holds that the $c_{ij'k'}$ that appear are precisely those where $j' + k' = s$. We thus obtain the formula in the claim. ◁

We now continue to prove the inductive step for $j + k = s$. We split into three cases.

Case 1: $s > p - 1 + d_y$. In this case, we proceed by backwards induction on the value of k to show that $c_{ijk} = 0$. For $k > p - 1$, there are no such terms, so $c_{i, s-k, k} = 0$ for all $i \in [0, d_x]$.

Now assume the inductive hypothesis that for all $k' > k$ it holds that $c_{i, s-k', k'} = 0$. Using Claim 35, we have that

$$\begin{aligned} g_{s-k, k}(x) &= \sum_{\substack{0 \leq i \leq d_x \\ k \leq k' \leq p-1}} c_{i, s-k', k'} \cdot \binom{k'}{k} x^{k'-k+i} \\ &= \sum_{0 \leq i \leq d_x} c_{i, s-k, k} \cdot x^i, \end{aligned}$$

where in the second line we've used the inductive hypothesis. But $s - k \geq s - (p - 1) > d_y$, so $g_{s-k, k}(x) = 0 \forall x \in \mathbb{F}_p$. Since $g_{s-k, k}(x)$ is a polynomial of degree $d_x < p$, this implies that all the coefficients $c_{i, s-k, k}$ are equal to 0. This completes the inductive step.

Case 2: $p - 1 < s \leq p - 1 + d_y$. In this case, to prove that $c_{i,s-k,k} = 0$ for all $i \in [0, d_x]$ and $k \in [0, p-1]$, our strategy is to find independent linear relations between these coefficients. It will be convenient to view a matrix which represents the data given by Claim 35, as follows:

$$\begin{array}{l}
 c_{i,s-p+1,p-1} \\
 c_{i,s-p+2,p-2} \\
 c_{i,s-p+3,p-3} \\
 \vdots \\
 c_{i,d_y+1,s-d_y-1} \\
 c_{i,d_y+2,s-d_y-2} \\
 \vdots \\
 c_{i,p-1,s-p+1}
 \end{array}
 \begin{bmatrix}
 g_{d_y+1,s-d_y-1} & g_{d_y+2,s-d_y-2} & \cdots & g_{p-1,s-p+1} \\
 \binom{p-1}{s-d_y-1} x^{p-s+d_y+i} & \binom{p-1}{s-d_y-2} x^{p-s+d_y+i+1} & \cdots & \binom{p-1}{s-p+1} x^{2p-s+i-2} \\
 \binom{p-2}{s-d_y-1} x^{p-s+d_y+i-1} & \binom{p-2}{s-d_y-2} x^{p-s+d_y+i} & \cdots & \binom{p-2}{s-p+1} x^{2p-s+i-3} \\
 \binom{p-3}{s-d_y-1} x^{p-s+d_y+i-2} & \binom{p-3}{s-d_y-2} x^{p-s+d_y+i-1} & \cdots & \binom{p-3}{s-p+1} x^{2p-s+i-4} \\
 \vdots & \vdots & & \vdots \\
 \binom{s-d_y-1}{s-d_y-1} x^i & \binom{s-d_y-1}{s-d_y-2} x^{i+1} & \cdots & \binom{s-d_y-1}{s-p+1} x^{p-d_y+i-2} \\
 0 & \binom{s-d_y-2}{s-d_y-2} x^i & \cdots & \binom{s-d_y-2}{s-p+1} x^{p-d_y+i-3} \\
 \vdots & \vdots & & \vdots \\
 0 & 0 & \cdots & \binom{s-p-1}{s-p-1} x^i
 \end{bmatrix}$$

The rows correspond to the coefficients $c_{i,s-k,k}$ for $k = p-1, \dots, s-p+1$ as labeled on the left. Each row actually corresponds to $d_x + 1$ coefficients, for $i \in [0, d_x]$, but for sake of space we condense these into a single row.

The columns correspond to $g_{d',s-d'}(x)$ for $d' \in [d_y + 1, p-1]$. Because $d' > d_y$, these $g_{d',s-d'}(x)$ all should be 0. The way to read off the value of $g_{d',s-d'}(x)$ is by looking at the corresponding column, and summing the product of each entry with the corresponding row coefficient $c_{i,s-k',k'}$, remembering that each row actually corresponds to $d_x + 1$ rows for $i \in [0, d_x]$. The entries of the matrix were chosen according to the expansion of $g_{d',s-d'}(x)$ given in Claim 35.

Now, to prove that $c_{i,s-k,k} = 0$ for all $i \in [0, d_x]$ and $k \in [0, p-1]$, we will backwards induct on the value of $i + k$, which we will denote by t . So assume that for $i' + k' > t$ that $c_{i',s-k',k'} = 0$. This is true for $t = d_x + (p-1)$, since there do not exist coefficients with larger values of $i' + k'$ (as $i' \leq d_x$ and $k' \leq p-1$).

Consider all the coefficients with $i + k = t$, meaning coefficients $c_{i,s-t+i,t-i}$ where $i \in [0, d_x]$ and $t-i, s-t+i \in [0, p-1]$. These coefficients correspond to $\leq d_x + 1$ consecutive rows of the matrix with increasing values of i . Note that in each column, these coefficients only contribute to a single monomial $x^{i'}$: namely, in the column for $g_{d',s-d'}$, the exponents of x that appear for $i + k = t$ are all equal to $i' = i + (k - (s - d')) = t - s + d'$. Thus, for any $g_{d',s-d'}(x)$ which has degree $< p$, we obtain a linear constraint on $c_{i,s-t+i,t-i}$ by restricting the corresponding column corresponding to the rows corresponding to the coefficients (we may also drop the powers of x , so that the coefficients of the linear constraint are simply binomial coefficients). For instance, for $t = p$, the column $g_{d',s-d'}$ gives us the constraint

$$\begin{pmatrix} c_{1,s-p+1,p-1} & c_{2,s-p+2,p-2} & \cdots & c_{d_x,s-p+d_x,p-d_x} \end{pmatrix} \begin{pmatrix} \binom{p-1}{s-d'} \\ \binom{p-2}{s-d'} \\ \vdots \\ \binom{p-d_x}{s-d'} \end{pmatrix} = 0.$$

Our strategy may therefore be summarized as follows: for each $t \leq p-1 + d_x$, assuming that there are m coefficients of the form $c_{i,s-t+i,t-i}$, we will find m consecutive columns such that the degree of x in $g_{d',s-d'}(x)$ is $< p$. These columns will also satisfy that if we restrict the matrix to these columns and to the rows corresponding to the coefficients, the resulting $m \times m$ matrix has diagonal that lies above the lower left triangle of 0's. Then, by Lemma 33, we know these m linear constraints are independent, telling us that $c_{i,s-t+i,t-i} = 0$ for each i .

The columns will be chosen as follows.

- For $t \geq p - 1$, we are considering the $m = p - t + d_x$ coefficients

$$c_{t-p+1, s-p+1, p-1}, c_{t-p+2, s-p+2, p-2}, \dots, c_{d_x, s-t+d_x, t-d_x}.$$

We pick the first $p - t + d_x$ columns of the matrix, corresponding to $g_{d_y+1, s-d_y-1}, g_{d_y+2, s-d_y-2}, \dots, g_{p+d_x+d_y-t, s+t-p-d_x-d_y}$. Each of $g_{d', s-d'}(x)$ for $d' \in [d_y + 1, p + d_x + d_y - t]$ is a polynomial in x . Using the inductive hypothesis that $c_{i', s-k', k'} = 0$ for $i' + k' > t$, we see that the maximum degree of x in any of these polynomials is $\leq i + k - \min(s - d') = t - (s + t - p - d_x - d_y) = p + d_x + d_y - s < d_x + d_y < p$.

- For $s - d_y - 1 \leq t < p - 1$, we are looking at the $d_x + 1$ coefficients

$$c_{0, s-t, t}, c_{1, s-t+1, t-1}, \dots, c_{d_x, s-t+d_x, t-d_x}.$$

Consider the $d_x + 1$ columns corresponding to $g_{d_y+1, s-d_y-1}, \dots, g_{d_x+d_y+1, s-d_x-d_y-1}$. Because $t \geq s - d_y - 1$, the submatrix has nonzero diagonal. The largest exponent of x in any of these columns is $\leq i + k - (s - d_x - d_y - 1) \leq t - s + d_x + d_y - 1 < d_x + d_y - 1 < p$.

- For $t < s - d_y - 1$, let $\hat{d} = \min(d_x, p - 1 - s + t)$. We consider the $\hat{d} + 1$ coefficients

$$c_{0, s-t, t}, c_{1, s-t+1, t-1}, \dots, c_{\hat{d}, p-1, t-\hat{d}}.$$

Look at the $\hat{d} + 1$ columns corresponding to $g_{s-t, t}, \dots, g_{s-t+\hat{d}, t-\hat{d}}$. This matrix has nonzero diagonal. The largest degree of x in any of these columns is $\leq i + k - (t - \hat{d}) \leq \hat{d} < p$.

Case 3: $d_x + d_y < s \leq p - 1$. Similar to the previous case, we consider the following matrix, interpreted the same way as before.

$$\begin{array}{c} c_{i, 0, s} \\ c_{i, 1, s-1} \\ c_{i, 2, s-2} \\ \vdots \\ c_{i, d_y+1, s-d_y-1} \\ c_{i, d_y+2, s-d_y-2} \\ c_{i, d_y+3, s-d_y-3} \\ \vdots \\ c_{i, s, 0} \end{array} \begin{bmatrix} g_{d_y+1, s-d_y-1} & g_{d_y+2, s-d_y-2} & \cdots & g_{s, 0} \\ \binom{s}{s-d_y-1} x^{i+d_y+1} & \binom{s}{s-d_y-2} x^{i+d_y+2} & \cdots & \binom{s}{0} x^{i+s} \\ \binom{s-1}{s-d_y-1} x^{i+d_y} & \binom{s-1}{s-d_y-2} x^{i+d_y+1} & \cdots & \binom{s-1}{0} x^{i+s-1} \\ \binom{s-2}{s-d_y-1} x^{i+d_y-1} & \binom{s-2}{s-d_y-2} x^{i+d_y} & \cdots & \binom{s-2}{0} x^{i+s-2} \\ \vdots & \vdots & \vdots & \vdots \\ \binom{s-d_y-1}{s-d_y-1} x^i & \binom{s-d_y-1}{s-d_y-2} x^{i+1} & \cdots & \binom{s-d_y-1}{0} x^{i+s-d_y-1} \\ 0 & \binom{s-d_y-2}{s-d_y-2} x^i & \cdots & \binom{s-d_y-2}{0} x^{i+s-d_y-2} \\ 0 & 0 & \cdots & \binom{s-d_y-3}{0} x^{i+s-d_y-3} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \binom{0}{0} x^i \end{bmatrix}.$$

As in the previous case, we will downwards induct on the value of $i + k = t$. Assume that for $i' + k' > t$ it holds that $c_{i', s-k', k'} = 0$. This is true for $t = d_x + s$ (the maximal possible value of $i + k$) since there do not exist coefficients with larger values of $i' + k'$.

We again consider all m coefficients satisfying $i + k = t$. These occupy a number of consecutive rows of the matrix. Furthermore, for any column with degree in x less than p , restricting that column to those rows gives a linear constraint on the coefficients. Thus, we again demonstrate, for each t , m consecutive columns that each have exponent in x less than p . Restricting these columns to the coefficient rows will result in a $m \times m$ submatrix lying above the lower left triangle of 0's, which is full rank by Lemma 33, implying that all m coefficients must in fact be 0.

27:28 New Codes on High Dimensional Expanders

- For $t \geq s$, we are interested in the $s - t + d_x + 1$ coefficients

$$c_{t-s,0,s}, c_{t-s+1,1,s-1}, \dots, c_{d_x,s-t+d_x,t-d_x}.$$

Look at the first $s - t + d_x + 1$ columns, which correspond to $g_{d',s-d'}$ for $d' \in [d_y + 1, s - t + d_x + d_y + 1]$. The maximum degree of any of these polynomials is $i + k - (t - d_x - d_y - 1) = d_x + d_y + 1 < p$.

- For $s - d_y - 1 \leq t < s$, we consider the $d_x + 1$ coefficients

$$c_{0,s-t,t}, c_{1,s-t+1,t-1}, \dots, c_{d_x,s-t+d_x,t-d_x}.$$

The columns we are interested in are $g_{d_y+1,s-d_y-1}, \dots, g_{d_x+d_y+1,s-d_x-d_y-1}$. Since $t \geq s - d_y - 1$, the submatrix has nonzero diagonal. The maximum degree of any of these polynomials is $\leq i + k - (s - d_x - d_y - 1) = t - s + d_x + d_y + 1 < d_x + d_y + 1 < p$.

- For $t < s - d_y - 1$, let $\hat{d} = \min(d_x, t)$. Then we are considering the following $\hat{d} + 1$ coefficients.

$$c_{0,s-t,t}, c_{1,s-t+1,t-1}, \dots, c_{\hat{d},s-t+\hat{d},t-\hat{d}}.$$

For these, we will look at the $\hat{d} + 1$ columns corresponding to $g_{s-t,t}, \dots, g_{s-t+\hat{d},t-\hat{d}}$. This clearly has nonzero diagonal, and the maximum degree of x in any of these polynomials is $i + k - (t - \hat{d}) = \hat{d} < p$. \blacktriangleleft

► **Theorem 36.** For $p \geq d_x + d_y + 2$, the dimension of the local code is $\frac{1}{2} \cdot (d_x + 1)(d_y + 1)(d_x + d_y + 2)$.

Proof. From Lemma 34, we know that $c_{ijk} = 0$ for all $i \in [0, d_x]$ and $j + k > d_x + d_y$. Thus we can write

$$f(x, y, z) = \sum_{\substack{0 \leq i \leq d_x \\ 0 \leq j+k \leq d_x+d_y}} c_{ijk} x^i y^j z^k.$$

Note that $c_{i'j'k'} x^{i'} y^{j'} (xy + z)^{k'} = \sum_{0 \leq k \leq k'} \binom{k'}{k} x^{k'-k+i'} y^{k'-k+j'} z^k$ and in particular the sum of the exponents of y and z is always $j' + k'$, so in particular the value of $g_{jk}(x)$ only depends on coefficients $c_{i'j'k'}$ where $j' + k' = j + k$ (note also that we're now in the setting where $j + k \leq d_x + d_y < p$, so $g(x, y, z) = f(x, y, xy + z)$ without having to reduce modulo $y^p - y$).

Again, we will use the fact that for all $j > d_y$ and $k \in [0, p - 1]$, it holds that $g_{jk}(x) = 0 \forall x \in \mathbb{F}_p$. Note that $g_{jk}(x)$ could have degree as large as $2d_x + d_y$ which could be larger than p . We thus let $\hat{g}_{jk}(x)$ denote $g_{jk}(x)$ reduced modulo $x^p - x$. The condition that $g_{jk}(x) = 0 \forall x \in \mathbb{F}_p$ then is equivalent to $\hat{g}_{jk}(x) \equiv 0$.

We will work through the polynomials $\hat{g}_{jk}(x)$ and set each to 0 by choosing coefficients c_{ijk} appropriately. We will consider all the polynomials \hat{g}_{jk} with a fixed value of $j + k$, denoted by s , simultaneously.

As such, let $d_y < s \leq d_x + d_y$. The relevant polynomials $g_{jk}(x)$ that evaluate to 0 everywhere are $g_{d_y+1,s-d_y-1}, \dots, g_{s,0}$, and the relevant coefficients are $c_{i,0,s}, \dots, c_{i,s,0}$, where $i \in [0, d_x]$. The dependency of $g_{d',s-d'}$ on the coefficients is given in the following matrix, interpreted the same as before.

$$\begin{array}{c}
c_{i,0,s} \\
c_{i,1,s-1} \\
c_{i,2,s-2} \\
\vdots \\
c_{i,d_y+1,s-d_y-1} \\
c_{i,d_y+2,s-d_y-2} \\
c_{i,d_y+3,s-d_y-3} \\
\vdots \\
c_{i,s,0}
\end{array}
\begin{bmatrix}
g_{d_y+1,s-d_y-1} & g_{d_y+2,s-d_y-2} & \cdots & g_{s,0} \\
\binom{s}{s-d_y-1}x^{i+d_y+1} & \binom{s}{s-d_y-2}x^{i+d_y+2} & \cdots & \binom{s}{0}x^{i+s} \\
\binom{s-1}{s-d_y-1}x^{i+d_y} & \binom{s-1}{s-d_y-2}x^{i+d_y+1} & \cdots & \binom{s-1}{0}x^{i+s-1} \\
\binom{s-2}{s-d_y-1}x^{i+d_y-1} & \binom{s-2}{s-d_y-2}x^{i+d_y} & \cdots & \binom{s-2}{0}x^{i+s-2} \\
\vdots & \vdots & \vdots & \vdots \\
\binom{s-d_y-1}{s-d_y-1}x^i & \binom{s-d_y-1}{s-d_y-2}x^{i+1} & \cdots & \binom{s-d_y-1}{0}x^{i+s-d_y-1} \\
0 & \binom{s-d_y-2}{s-d_y-2}x^i & \cdots & \binom{s-d_y-2}{0}x^{i+s-d_y-2} \\
0 & 0 & \cdots & \binom{s-d_y-3}{0}x^{i+s-d_y-3} \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \binom{0}{0}x^i
\end{bmatrix}.$$

We will set the coefficients $c_{i,s-k,k}$ starting with those with the largest value of $t := i + k$, which is $d_x + s$, and proceeding downwards. We will show that if we've already set all coefficients $c_{i,s-k,k}$ with $i + k > t$, then the degree of $\hat{g}_{d',s-d'}(x)$ is $\leq d' + t - s$ (that is, all larger monomials have been set to 0 by the previous choices of coefficient assignments). This is certainly satisfied in the base case: if $t = d_x + s$, then since $i \leq d_x$ we have that the largest power of x in any $g_{d',s-d'}(x)$ is $d_x + d' = d' + t - s$. This largest exponent can only decrease when we pass to $\hat{g}_{d',s-d'}(x)$.

Now, suppose that we're part way through this process and have already set the values for all $c_{i,s-k,k}$ where $i + k > t$. The number of ways we have to set the values $c_{t-k,s-k,k}$ is as follows.

- For $t = d_x + s, d_x + s - 1, \dots, d_x + d_y + 1$, the coefficients of interest are $c_{t-s,0,s}, c_{t-s+1,1,s-1}, \dots, c_{d_x,s+d_x-t,t-d_x}$. We will show that these must all be set to 0. We look at the columns corresponding to $g_{d_y+1,s-d_y-1}, \dots, g_{d_x+d_y-t+s+1,t-d_x-d_y-1}$. Note that there are as many columns as coefficients. By the inductive assumption, the maximal degree of column $g_{d',s-d'}$ is $\leq d' + t - s \leq d_x + d_y + 1 < p$. Thus the coefficient of $x^{d'+t-s}$ in $g_{d',s-d'}(x)$, which is equal to $\sum_{t-d_x \leq k \leq s} \binom{k}{s-d'} c_{t-k,s-k,k}$, must also be 0. This gives us the following linear system of equations:

$$\begin{pmatrix} c_{t-s,0,s} \\ c_{t-s+1,1,s-1} \\ \vdots \\ c_{d_x,s-t+d_x,t-d_x} \end{pmatrix}^T \cdot \begin{pmatrix} \binom{s}{s-d_y-1} & \binom{s}{s-d_y-2} & \cdots & \binom{s}{t-d_x-d_y-1} \\ \binom{s-1}{s-d_y-1} & \binom{s-1}{s-d_y-2} & \cdots & \binom{s-1}{t-d_x-d_y-1} \\ \vdots & \vdots & \vdots & \vdots \\ \binom{t-d_x}{s-d_y-1} & \binom{t-d_x}{s-d_y-2} & \cdots & \binom{t-d_x}{t-d_x-d_y-1} \end{pmatrix} = 0,$$

where the matrix is also a top left submatrix of the aforementioned matrix, ignoring the powers of x . Since $s > s - d_y - 1$, this matrix has nonzero diagonal and by Lemma 33 is invertible. This implies that $c_{t-k,s-k,k} = 0$ for all $t - d_x \leq k \leq s$.

Now, we've set all c_{ijk} with $j + k \geq t$. It remains to show that $\hat{g}_{d',s-d'}(x)$ is now degree $\leq d' + t - s - 1$. We already have that $\hat{g}_{d',s-d'}(x)$ was degree $\leq d' + t - s$ from the inductive assumption. For any $\hat{g}_{d',s-d'}(x)$, the coefficient of the $x^{d'+t-s}$ term is the sum of the coefficients of $x^{d'+t-s}$ and $x^{d'+t-s+(p-1)}$ in $g_{d',s-d'}(x)$. But recall that we've set all c_{ijk} with $j + k \geq t$ to 0, so both these coefficients in $g_{d',s-d'}$ must be 0. Therefore, $\hat{g}_{d',s-d'}(x)$ is degree $\leq d' + t - s - 1$.

- Next, for $t = d_x + d_y, \dots, s + 1$, we are looking at the $s + d_x - t + 1 > s - d_y$ coefficients $c_{t-s,0,s}, \dots, c_{d_x, s-t+d_x, t-d_x}$. We have so far that all $\hat{g}_{d', s-d'}(x)$ have degree $\leq d' + t - s$. The coefficient of $x^{d'+t-s}$ in $\hat{g}_{d', s-d'}(x)$ is equal to the sum of the coefficients of $x^{d'+t-s}$ and $x^{d'+t-s+(p-1)}$ in $g_{d', s-d'}(x)$, which in turn is equal to

$$\sum_{t-d_x \leq k \leq s} \binom{k}{s-d'} \cdot c_{t-k, s-k, k} + \sum_{0 \leq k \leq s} \binom{k}{s-d'} \cdot c_{t-k+(p-1), s-k, k}, \quad (16)$$

where in the second summation the terms $c_{t-k+(p-1), s-k, k}$ that don't exist are understood to be 0. Note that we've already set the values of $c_{t-k+(p-1), s-k, k}$. Thus, in order to set (16) to 0, we need to choose $c_{t-k, s-k, k}$, $t - d_x \leq k \leq s$ so that

$$\sum_{t-d_x \leq k \leq s} \binom{k}{s-d'} \cdot c_{t-k, s-k, k} = - \sum_{0 \leq k \leq s} \binom{k}{s-d'} \cdot c_{t-k+(p-1), s-k, k}.$$

There are $s - d_y$ such equations for the $s - d_y$ polynomials $\hat{g}_{d', s-d'}(x)$, which are all independent by Lemma 33 since $s \geq s - d_y - 1$. Then, there are $p^{s+d_x-t+1-(s-d_y)} = p^{d_x+d_y+1-t}$ ways to choose the coefficients $c_{t-k, s-k, k}$. We remark also that once $c_{t-s,0,s}, \dots, c_{d_x, s-t+d_x, t-d_x}$ are set, all $\hat{g}_{d', s-d'}(x)$ must have degree $\leq d' + t - s - 1$ since we chose the values so that the coefficient of $x^{d'+t-s}$ was 0 for all columns.

- The next case is $t = s, s-1, \dots, d_x$. In this case, we are looking at the $d_x + 1$ coefficients $c_{0, s-t, t}, \dots, c_{d_x, s-t+d_x, t-d_x}$. We have so far that all $\hat{g}_{d', s-d'}(x)$ have degree $\leq d' + t - s$, and the coefficient of $x^{d'+t-s}$ in $\hat{g}_{d', s-d'}(x)$ is equal to the sum of the coefficients of $x^{d'+t-s}$ and $x^{d'+t-s+(p-1)}$ in $g_{d', s-d'}(x)$. Similar to the previous case, this results in $s - d_y$ independent linear equations (by Lemma 33, using the fact that $t \geq s - d_y - 1$). Thus, there are $p^{d_x+d_y+1-s}$ ways to set $c_{0, s-t, t}, \dots, c_{d_x, s-t+d_x, t-d_x}$. Note that after we've set these coefficients, the degree of $\hat{g}_{d', s-d'}(x)$ necessarily must be $\leq d' + t - s - 1$ by choice of these coefficients.
- If $t = d_x - 1, \dots, s - d_y$, then the coefficients we care about are $c_{0, s-t, t}, \dots, c_{t, s, 0}$. We have that $\hat{g}_{d', s-d'}(x)$ has degree $\leq d' + t - s$ and wish to set the $c_{t-k, s-k, k}$ so that the coefficient of $x^{d'+t-s}$ is 0. As before, this results in $s - d_y$ linearly independent equations, which are independent because $t \geq s - d_y - 1$. Thus, there are $p^{t+1-s+d_y}$ ways to set the coefficients $c_{t-k, s-k, k}$. The degrees of all $s - d_y$ polynomials $\hat{g}_{t-k, s-k, k}(x)$ are now $\leq d' + t - s - 1$.
- For $t = s - d_y - 1, \dots, 0$, we are considering the $t + 1$ coefficients $c_{0, s-t, t}, \dots, c_{t, s, 0}$. Note also that for $d' < s - t$, we must already have that $\hat{g}_{d', s-d'}(x) \equiv 0$ since the maximum degree, if it exists, is already $\leq d' + t - s$. So, we look at the $t + 1$ polynomials $\hat{g}_{s-t, t}(x), \dots, \hat{g}_{s, 0}(x)$. Since we want to set the coefficient of $x^{d'+t-s}$ to be 0, this results in $t + 1$ linearly independent equations in $c_{0, s-t, t}, \dots, c_{t, s, 0}$. So there is exactly one way to set $c_{0, s-t, t}, \dots, c_{t, s, 0}$ to make all these coefficients 0.

In total, for $d_y < s \leq d_x + d_y$, if C_s is the number of ways to assign all the coefficients $c_{i, s-k, k}$, then

$$\begin{aligned} \log_p C_s &= 0 + \sum_{t=s+1}^{d_x+d_y} (d_x + d_y + 1 - t) + \sum_{t=d_x}^s (d_x + d_y + 1 - s) \\ &\quad + \sum_{t=s-d_y}^{d_x-1} (t + 1 - s + d_y) + 0 \\ &= (d_y + 1)(d_x + d_y + 1 - s). \end{aligned}$$

Furthermore, if $j + k = s \leq d_y$, then $f(x, y, xy + z)$ is always degree $\leq d_y$ in y , so all such coefficients c_{ijk} are permissible. There are $(d_x + 1) \cdot \binom{d_y+2}{2}$ such coefficients. In total, this gives that the dimension of the local code is

$$\begin{aligned} (d_x + 1) \cdot \binom{d_y + 2}{2} + \sum_{s=d_y+1}^{d_x+d_y} ((d_y + 1)(d_x + d_y + 1 - s)) \\ = \frac{1}{2} \cdot (d_x + 1)(d_y + 1)(d_x + d_y + 2). \end{aligned} \quad \blacktriangleleft$$

5 Code Testability

In this section, we will work with the field size $q = p$ being a prime. The main reason is that we will need results from Section 4.2 about the degree of codewords in C_{d_x, d_y} , viewed as low degree polynomials.

5.1 Testability of the Local Code at a Vertex

In this section we prove that $C_v \cong C_{d_x, d_y}$ (as defined in Lemma 30) is agreement-testable whenever $d_x, d_y < p/4$. Our definition of agreement testability (see Definition 15) applies to HDX codes. Let us restate it in a form that is specialized for C_v :

► **Definition 37.** *The local code C_v defined in (15) is $(\epsilon, \rho(\cdot))$ -agreement testable if whenever we are given a collection of $z_e \in C_e$ for each $e \ni v$, such that*

$$\alpha := \mathbb{P}_{uv \in X_v(1)} [z_{uv}(T_{uvw}) \neq z_{vw}(T_{uvw})] < \epsilon$$

then there exists some $x \in C_v$ such that

$$\mathbb{P}_{u \in X(0)} (z_{uv} \neq x|_{T_{uv}}) \leq \rho(\alpha).$$

Recall $C_v \cong C_{d_x, d_y}$. The local views have two kinds. Let us see the case where $v = K_1$: Cosets of H_2 correspond to lines (x, b, c) for all $b, c \in \mathbb{F}_p$. Cosets of H_3 correspond to lines $(a, y, ay + c)$ for all $a, c \in \mathbb{F}$. Therefore, the local views can be packaged through X, Y , which will be collections of degree d_2 and d_3 polynomials on the lines corresponding to cosets of H_2 and H_3 , respectively.

The following theorem will immediately imply local testability.

► **Theorem 38.** *Let $X, Y : \mathbb{F}_p^3 \rightarrow \mathbb{F}_p$ be such that*

- *For each b, c , the x degree of $X(x, b, c)$ is at most d_x .*
- *For each a, c , the y degree of $Y(a, y, ay + c)$ is at most d_y .*

Then if $\mathbb{P}[X(x, y, z) \neq Y(x, y, z)] = \delta^3$ so that $p \geq d_x + d_y + 2 \max(d_x, d_y) + 4\delta p$, there exists codeword $Q(x, y, z) \in C_{d_x, d_y}$ such that

$$\mathbb{P}_{b,c} [X(x, b, c) \neq Q(x, b, c)] + \mathbb{P}_{a,c} [Y(a, y, ay + c) \neq Q(a, y, ay + c)] \leq 4\delta.$$

Before proving Theorem 38, let us see that it immediately implies agreement testability.

► **Corollary 39.** *If $d_x, d_y < \frac{p}{4}$, the local code $C_v = C_{d_x, d_y}$ is $\left(\left(\frac{p/2 - (d_x + d_y)}{4p} \right)^3, 4(\cdot)^{1/3} \right)$ -agreement testable.*

Proof. Without loss of generality, let v be the coset K_1 in $X(G; K_1, K_2, K_3)$ (the argument applies to other choices of v since G acts transitively on the code C_v , see Claim 26). Then every edge $e \ni v$ is of type 2 or type 3. Recall from Theorem 1 that type 2 edges e are cosets $\{gh_2(x)\}_x$ while type 3 edges e correspond to cosets of the form $\{gh_3(y)\}_y$. Given a collection of local views $z_e \in C_e$, we define functions $X, Y : \mathbb{F}_p^3 \rightarrow \mathbb{F}_p$ in the following way:

$$\begin{aligned} X(x, b, c) &= z_{gh_2}(gh_2(x)) \text{ where } g = \iota^{-1}(0, b, c), \\ Y(a, y, ay + c) &= z_{gh_3}(gh_3(y)) \text{ where } g = \iota^{-1}(a, 0, c). \end{aligned}$$

Here ι is the embedding of elements in K_1 to \mathbb{F}_p^3 as used in the proof of Lemma 30. Since $z_{gh_2}(gh_2(x))$ has degree at most d_2 and $z_{gh_3}(gh_3(y))$ has degree at most d_3 , X and Y satisfy the conditions in the theorem statement for $d_x = d_2, d_y = d_3$. Also by construction of X and Y

$$\mathbb{P}_{uv \in X_v(1)}[z_{uv}(T_{uvw}) \neq z_{uv}(T_{uvw})] = \mathbb{P}_{x,y,z}[X(x, y, z) \neq Y(x, y, z)]$$

Similarly, given a codeword $Q \in C_{d_x, d_y}$, by Lemma 30 we can define its corresponding codeword $f \in C_v$ as

$$f(g) = Q(\iota(g)).$$

Then the disagreement probability between z_e and f satisfies:

$$2 \mathbb{P}_{u \in X(0)}(z_{uv} \neq f|_{T_{uv}}) = \mathbb{P}_{b,c}[X(x, b, c) \neq Q(x, b, c)] + \mathbb{P}_{a,c}[Y(a, y, ay + c) \neq Q(a, y, ay + c)].$$

Now applying Theorem 38 to z_e and f we have that the local code C_v is $\left(\left(\frac{p - (d_x + d_y) - 2 \max(d_x, d_y)}{4p} \right)^3 \geq \left(\frac{p/2 - (d_x + d_y)}{4p} \right)^3, 4(\cdot)^{1/3} \right)$ -agreement testable. \blacktriangleleft

We now move to prove Theorem 38. Let S denote the set of all points (x, y, z) on which $X(x, y, z) \neq Y(x, y, z)$. Consider all polynomials $e(x, y, z)$ that are degree δp in x , and degree δp in y in $e(x, y, xy + z)$. By Theorem 36, the dimension of the space of such polynomials is $(\delta p + 1)^3$. Thus, by dimension counting, there is a nonzero polynomial $E(x, y, z)$ that evaluates to 0 on all points of S . Then, we have that for all x, y, z , $X(x, y, z)E(x, y, z) = Y(x, y, z)E(x, y, z)$.

We have that $X(x, y, z)E(x, y, z)$ is degree $d_x + \delta p$ in x , and $Y(x, y, xy + z)E(x, y, xy + z)$ is degree $d_y + \delta p$ in y . Thus there is a polynomial $P(x, y, z)$ that is degree $d_x + \delta p$ in x , and degree $d_y + \delta p$ in y in $P(x, y, xy + z)$, that agrees on all points. (To see this, we can write $P(x, y, z)$ as a polynomial that is degree $q - 1$ in each of x, y, z . Since $P(x, y_0, z_0)$ has the same evaluation as a degree $d_x + \delta p$ polynomial for each y_0, z_0 , it follows that P must be degree $d_x + \delta p$ in x . Next, by Lemma 34, we have that P must be combined degree $d_x + \delta p + d_y + \delta p < q$ in y and z . Thus, $P(x, y, xy + z)$ is degree $< q$ in y . Since $P(x_0, y, x_0y + z_0)$ has the same evaluation as a degree $d_y + \delta p$ polynomial for each x_0, z_0 , it follows that $P(x, y, xy + z)$ must actually be degree $d_y + \delta p$ in y .)

We may write

$$X(x, y, z)E(x, y, z) = Y(x, y, z)E(x, y, z) = P(x, y, z) \quad \forall x, y, z.$$

We'd like to formally divide $P(x, y, z)$ by $E(x, y, z)$ to obtain a polynomial $Q(x, y, z)$ that is degree d_x in x and degree d_y in the skew- y direction.

► **Lemma 40.** *There exists a polynomial $Q(x, y, z)$ that is degree d_x in x and degree d_y in the skew- y direction, such that $E(x, y, z)Q(x, y, z) = P(x, y, z)$.*

Proof. First, suppose $P(x, y, z)$ and $E(x, y, z)$ share a common factor $F(x, y, z) \not\equiv E(x, y, z)$ of degree (e, f) in the x and skew- y directions, and set

$$P(x, y, z) \equiv \bar{P}(x, y, z)F(x, y, z) \text{ and } E(x, y, z) \equiv \bar{E}(x, y, z)F(x, y, z)$$

so that $\bar{P}(x, y, z)$ and $\bar{E}(x, y, z)$ share no common factors.

Note that whenever $F(x, y_0, z_0)$ is not the 0 polynomial (in x), then we have that $\bar{E}(x, y_0, z_0) | \bar{P}(x, y_0, z_0)$. Since the combined degree of y and z in $F(x, y, z)$ is $\leq e + f < |\mathbb{F}|$ by Lemma 34 and $F(x, y, xy + z)$ is degree f in y , there can be at most f values of $y_0 \in \mathbb{F}$ for which $F(x, y_0, z) \equiv 0$ and also at most f values of $z_0 \in \mathbb{F}$ for which $F(x, y, z_0) \equiv 0$. Let Y_x denote the values of y_0 for which $F(x, y_0, z) \not\equiv 0$ and Z_x denote the values z_0 for which $F(x, y, z_0) \not\equiv 0$, so that $|Y_x|, |Z_x| \geq |\mathbb{F}| - f$. For any $y_0 \in Y_x$, there can be at most $e + f$ values of z_0 for which $f(x, y_0, z_0) \equiv 0$, and for any $z_0 \in Z_x$, there can be at most $e + f$ values of y_0 for which $f(x, y_0, z_0) \equiv 0$. Thus, there are at least $\max(|Y_x|, |Z_x|)(|\mathbb{F}| - e - f)$ pairs $(y_0, z_0) \in \mathbb{F} \times \mathbb{F}$ for which $\bar{E}(x, y_0, z_0) | \bar{P}(x, y_0, z_0)$. Let the set of such pairs (y_0, z_0) be denoted by M_x . We may show a similar statement for the skew- y direction, that there is a set M_y of size at least $\max(|Z_x|, |Z_y|)(|\mathbb{F}| - e - f)$ consisting of pairs (x_0, z_0) for which $\bar{E}(x_0, y, x_0y + z_0) | \bar{P}(x_0, y, x_0y + z_0)$ as polynomials in y .

Our goal is to show that $\bar{E}(x, y, z)$ divides $\bar{P}(x, y, z)$. Assume for the sake of contradiction that $\bar{P}(x, y, z)$ and $\bar{E}(x, y, z)$ have no common factors. Let the degree of $\bar{E}(x, y, z)$ be $(a' = a - e, b' = b - f)$ in the x and skew- y directions, so that $\bar{P}(x, y, z)$ has degree $(d_x + a', d_y + b')$ in the x and skew- y directions. Assume without loss of generality that $a' \geq b'$, otherwise we can consider the change of basis $P'(x, y, z) = \bar{P}(y, x, xy - z)$ and $E'(x, y, z) = \bar{E}(y, x, xy - z)$.

Write

$$\begin{aligned} \bar{P}(x, y, z) &\equiv P_0(y, z) + P_1(y, z)x + \cdots + P_{a'+d_x}(y, z)x^{a'+d_x} \\ \bar{E}(x, y, z) &\equiv E_0(y, z) + E_1(y, z)x + \cdots + E_{a'}(y, z)x^{a'}. \end{aligned}$$

Then, $\bar{P}(x, y, z)$ and $\bar{E}(x, y, z)$ have a common factor if there exists $A(x, y, z)$ that is degree $\leq a' - 1$ in x and $B(x, y, z)$ that is degree $\leq a' + d_x - 1$ in x such that

$$\bar{P}(x, y, z)A(x, y, z) = \bar{E}(x, y, z)B(x, y, z),$$

or

$$\begin{aligned} P_{a'+d_x}A_{a'-1} &= E_{a'}B_{a'+d_x-1} \\ P_{a'+d_x-1}A_{a'-1} + P_{a'+d_x}A_{a'-2} &= E_{a'-1}B_{a'+d_x-1} + E_{a'}B_{a'+d_x-2} \\ &\vdots \\ P_0A_0 &= E_0B_0. \end{aligned}$$

These equations are linear in A_i and $-B_i$ and can be summarized by the following $(2a' + d_x) \times (2a' + d_x)$ matrix:

$$M(\bar{P}, \bar{E})(y, z) = \begin{pmatrix} P_{a'+d_x} & P_{a'+d_x-1} & \cdots & \cdots & \cdots & P_0 & 0 & \cdots & 0 \\ 0 & P_{a'+d_x} & & \cdots & \cdots & P_1 & P_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & P_{a'+d_x} & \cdots & \cdots & \cdots & P_1 & P_0 \\ E_{a'} & E_{a'-1} & \cdots & E_1 & E_0 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & & & & & \ddots & & \vdots \\ \vdots & & \ddots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & E_{a'} & E_{a'-1} & \cdots & E_0 \end{pmatrix}$$

consisting of a' rows with \bar{P} entries and $a' + d_x$ rows of \bar{E} entries.

We define $R(\bar{P}, \bar{E})(y, z)$, the resultant of \bar{P} and \bar{E} , to be the polynomial in the coefficients of \bar{P} and \bar{E} (viewing both as a polynomial in x , with coefficients that are polynomials in y and z) obtained by taking the determinant of $M(\bar{P}, \bar{E})$. Solutions A and B exist if $R(\bar{P}, \bar{E}) = 0$, which would give our contradiction.

Viewing $R(\bar{P}, \bar{E})$ as a polynomial in y and z , and recalling by Lemma 34 that \bar{P} (resp. \bar{E}) is total degree $\leq d_x + d_y + a' + b'$ (resp. $\leq a' + b'$) in y and z , we see that $R(\bar{P}, \bar{E})$ has degree $\leq a' \cdot (d_x + d_y + a' + b') + (a' + d_x) \cdot (a' + b')$.

Meanwhile, for each $(y_0, z_0) \in M_x$, we have that $E(x, y_0, z_0) | P(x, y_0, z_0)$, so each of the top a' rows are linear combinations of the bottom $a' + d_x$ rows. Thus, the polynomial $R(P, E)$ has a zero at each $(y_0, z_0) \in M_x$ of multiplicity a' . Then, because

$$\begin{aligned} a' \cdot |M_x| &\geq a' \cdot \max(|Y_x|, |Z_x|)(|\mathbb{F}| - e - f) \\ &\geq a' \cdot \max(|Y_x|, |Z_x|)(3d_x + d_y + 3a' + b') \\ &> (a' \cdot (d_x + d_y + a' + b') + (a' + d_x) \cdot (a' + b')) \cdot \max(|Y_x|, |Z_x|) \\ &\geq \deg R \cdot \max(|Y_x|, |Z_x|), \end{aligned}$$

$R(\bar{P}, \bar{E})(y, z)$ must be the zero polynomial by Schwarz-Zippel. This implies that $\bar{P}(x, y, z)$ and $\bar{E}(x, y, z)$ must have non-trivial common factor when considered as polynomials in x , which implies (since we assumed that they're coprime) that $\bar{E}(x, y, z) | \bar{P}(x, y, z)$ in $\mathbb{F}(y, z)$. Then, by Gauss' Lemma, this implies that $E(x, y, z) | P(x, y, z)$ when considered as polynomials in $\mathbb{F}[y, z]$, the ring of polynomials in y and z .

Thus, polynomial $Q(x, y, z)$ that is degree d_x in x and degree d_y in the skew- y directions exists. \blacktriangleleft

The final step in the proof of Theorem 38 is to analyze the probability of the local views X and Y disagreeing with Q .

We have that

$$X(x, y, z)E(x, y, z) = Y(x, y, z)E(x, y, z) = Q(x, y, z)E(x, y, z) \quad \forall x, y, z.$$

Thus, for any (b, c) for which $E(x, b, c)$ is nonzero, we have that Q agrees with X on the entire row. Since E can be zero on at most $2\delta p^2$ rows (since the combined degree of y and z in $E(x, y, z)$ is at most $2\delta p$ by Lemma 34), this means that

$$\mathbb{P}_{b,c}[X(x, b, c) \neq Q(x, b, c)] \leq 2\delta.$$

Similarly, we have that $\mathbb{P}_{a,c}[Y(a, y, ay + c) \neq Q(a, y, ay + c)] \leq 2\delta$. This proves Theorem 38.

5.2 Local Testability: From Local to Global

Our main theorem is that if the local codes around an edge have distance δ , and the local codes around a vertex are agreement-testable, then the global code C is agreement-testable, and therefore robustly locally testable.

► **Theorem 41.** *Let $\epsilon_0, \delta > 0$, let $\rho_0(\cdot)$ be a monotone increasing function and assume that $\gamma < \min(\frac{\delta}{8}, \frac{\delta^2}{128} \min(\epsilon_0, \rho_0^{-1}(\frac{\delta}{4})))$. Let X be a two-dimensional bounded-degree γ -one-sided link expander. Suppose we are given codes $C_e \subset \mathbb{F}^{X+e}$ with relative minimum distance at least δ for each edge $e \in X(1)$, and let C_v and C be as defined above. If C_v is $(\epsilon_0, \rho_0(\cdot))$ -agreement-testable for all $v \in X(0)$ then C is $(\epsilon, \rho(\cdot))$ -agreement-testable, where $\epsilon = \frac{\delta^2}{128} \min(\epsilon_0, \rho_0^{-1}(\frac{\delta}{4}))$ and where $\rho(t) = Dt$ for D the maximal degree of a vertex in X . Namely, given any collection of local views $\{z_v \in C_v \mid v \in X(0)\}$, if*

$$\alpha(z) = \mathbb{P}_{uv \in X(1)}[z_u(X_{+uv}) \neq z_v(X_{+uv})] < \epsilon$$

then there is some $x \in C$ such that

$$\mathbb{P}_v[x|_{X_{+v}} \neq z_v] \leq D \cdot \alpha(z).$$

► **Corollary 42.** *Under the assumptions above, and assuming that each local code C_v is defined by at most m_0 parity checks each looking at most q_0 bits, the code C is $(d+2, \epsilon/m_0, \rho'(\cdot))$ locally testable under Definition 11 where $\rho'(t) = \rho(m_0 t)$.*

The corollary follows from a standard conversion from agreement-testability to robust testability, see Section 2.5, particularly Claims 16 and 14.

Proof of Theorem 41. Fix $z^0 = \{z_v^0\}$ a collection of local views, such that $\alpha(z^0) < \epsilon$. Run the following local correction algorithm:

Local Algorithm. If there is a vertex v and a choice $z'_v \in C_v$ that reduces $\alpha(z)$ then replace z_v by z'_v and repeat.

Let $z = \{z_v\}$ be the final collection, after the termination of the local algorithm. The algorithm must halt in at most $\alpha(z^0)|X(1)|$ steps. At this point, either $\alpha(z) = 0$, or $\alpha(z) > 0$. In the first case, we will show a nearby codeword. In the second case, we will show that in fact $\alpha(z) > \epsilon$, a contradiction since $\alpha(z) \leq \alpha(z^0) \leq \epsilon$.

▷ **Claim 43.** If $\alpha(z) = 0$ then z corresponds to a codeword $\hat{x} \in C$ such that $\mathbb{P}_{v \in X(0)}[z_v^0 \neq \hat{x}|_{X_{+v}}] \leq D\alpha(z^0)$, where D bounds the maximal degree of a vertex in X .

Proof. For each triangle $t \in X(2)$ choose arbitrarily a vertex $v \in t$ and set $\hat{x}(t) = z_v(t)$. This choice does not depend on the choice of v because $\alpha(z) = 0$ implies that $z_v(t) = z_{v'}(t)$ for any $v, v' \in t$.

At every step of the local algorithm, one local view changes. So the number of local views where $\hat{x}|_{X_{+v}} \neq z_v^0$ is at most the number of steps of the algorithm, which is at most $\alpha(z^0)|E|$. So

$$\mathbb{P}[\hat{x}|_{X_{+v}} \neq z_v^0] \leq \frac{\alpha(z^0)|E|}{|V|} = D \cdot \alpha(z^0). \quad \blacktriangleleft$$

Assume $\alpha(z) > 0$, and let

$$R = \{uv \in X(1) \mid z_u|_{X_{+uv}} \neq z_v|_{X_{+uv}}\}.$$

The rest of the proof will show that R has large size. First, we claim that

$$\mathbb{P}_{e \sim e'}[e \in R | e' \in R] \geq \delta/2. \quad (17)$$

where $e \sim e'$ is shorthand for the distribution of selecting e, e' from the upper random walk, namely, first choose a random triangle and then choose two distinct edges in it.

Fix an edge $e = uv \in R$. By definition, $(z_u)|_{X_{+uv}} \neq (z_v)|_{X_{+uv}}$. Since both are codewords in C_{uv} , for at least δ fraction of the triangles $uvw \in X_{+uv}$ either $z_u(uvw) \neq z_w(uvw)$ or $z_v(uvw) \neq z_w(uvw)$ and in particular either $uw \in R$ or $vw \in R$. So the random walk from uv to a triangle uvw and then to uw or vw has probability at least $\delta/2$ of staying in R . This establishes (17). By Lemma 9 we further deduce that

$$\mathbb{P}_{e \sim e'}[e \in R | e' \in R] \geq \delta/2 - \gamma. \quad (18)$$

where $e \sim e'$ is shorthand for the distribution of the lower random walk, namely, selecting a two random edges that intersect on a vertex.

The next step is to focus on the neighborhood of a fixed vertex. Fix $v \in V$. For every neighbor u of v let $y_u = z_u|_{X_{+uv}} \in C_{uv}$. This gives us a local view for each neighbor of v , which may or may not agree with $(z_v)|_{X_{+uv}}$. Let $R(v) = R \cap X_v(1) \supset \{uw \in X_v(1) \mid y_u(uvw) \neq y_w(uvw)\}$. We next show that vertices v with small $R(v)$ have few R edges adjacent to them. Here we use the agreement testability of the code C_v .

▷ **Claim 44.** Fix $v \in V$ such that $\epsilon_v \triangleq |R(v)|/|X_v(1)| \leq \epsilon_0$. Then $\mathbb{P}_{u \in X_v(0)}[uv \in R] \leq \rho_0(\epsilon_v)$. Contrapositively, if $\mathbb{P}_{u \in X_v(0)}[uv \in R] \geq \tau$, then $\epsilon_v \geq \min(\epsilon_0, \rho_0^{-1}(\tau))$.

Proof. For every neighbor u of v , $y_u = z_u|_{X_{+uv}} \in C_{uv}$ either agrees or disagrees with z_v . This is measured by the fraction of R edges touching v ,

$$\mathbb{P}_{u \in X_v(0)}[uv \in R] = \mathbb{P}_{u \in X_v(0)}[z_v|_{X_{+uv}} \neq y_u]. \quad (19)$$

Note also that for $uw \in X_v(1)$, if $y_u(uvw) \neq y_w(uvw)$ then $uw \in R(v)$ and so $uw \in R$. The assumption of the claim implies that $\Pr_{uw \in X_v(1)}[y_u(uvw) \neq y_w(uvw)] \leq \epsilon_v \leq \epsilon_0$. The $(\epsilon_0, \rho_0(\cdot))$ agreement-testability of C_v (see Definition 15) guarantees that in this case there exists a codeword $\hat{z}_v \in C_v$ such that

$$\mathbb{P}_{u \in X_v(0)}[\hat{z}_v|_{X_{+uv}} \neq y_u] \leq \rho_0(\epsilon_v) \quad (20)$$

where we have used the monotonicity of $\rho_0(\cdot)$.

Since the local algorithm halted without changing z_v to \hat{z}_v , we conclude, together with (19) and (20), that

$$\mathbb{P}_{u \in X_v(0)}[uv \in R] = \mathbb{P}_{u \in X_v(0)}[z_v|_{X_{+uv}} \neq y_u] \leq \mathbb{P}_{u \in X_v(0)}[\hat{z}_v|_{X_{+uv}} \neq y_u] \leq \rho_0(\epsilon_v).$$

To prove the contrapositive notice that if $\epsilon_v \geq \epsilon_0$ it is immediate, and if not, $\tau \leq \rho_0(\epsilon_v)$ which means that $\rho_0^{-1}(\tau) \leq \epsilon_v$ as needed (ρ_0 is invertible since it is monotone). ◁

Let $f = \mathbf{1}_R$. By (18),

$$\langle Df, Df \rangle = \langle f, U Df \rangle \geq (\delta/2 - \gamma) \|f\|^2. \quad (21)$$

Observe also that

$$\mathbb{E}_v Df(v) = \mathbb{E}_e f(e) = \mathbb{E}_e f(e)^2 = \|f\|^2.$$

▷ **Claim 45.** For any $\tau > 0$ let $V_\tau = \{v \in V \mid Df(v) > \tau\}$. Then

$$\mathbb{P}_v(V_\tau) \geq (\delta/2 - \gamma - \tau)\|f\|^2$$

Proof. Let $\mu(v)$ denote the probability of a vertex.

$$\begin{aligned} (\delta/2 - \gamma)\|f\|^2 &\leq \|Df\|^2 \\ &= \sum_{v \in V_\tau} \mu(v) Df(v)^2 + \sum_{v \notin V_\tau} \mu(v) Df(v)^2 \\ &\leq \sum_{v \in V_\tau} \mu(v) \cdot 1 + \sum_{v \notin V_\tau} \mu(v) Df(v) \cdot \tau \\ &\leq \sum_{v \in V_\tau} \mu(v) \cdot 1 + \sum_{v \in V} \mu(v) Df(v) \cdot \tau \\ &= \mathbb{P}[V_\tau] + \tau\|f\|^2 \end{aligned}$$

where we have used (21) in the first inequality, the definition of V_τ in the next inequality, and $Df(v) \geq 0$ in the last one. \triangleleft

Let $\tilde{M} = S_{1,0}D$ be the Markov operator corresponding to the random walk that starts at an edge e , selects a random vertex such that $e \cup \{v\} \in X(2)$ (we denote this condition by $e \perp v$), and then chooses a random $e' \ni v$. Then,

$$\begin{aligned} \langle f, \tilde{M}f \rangle &= \langle Df, S_{0,1}f \rangle = \sum_v \mu(v) Df(v) \cdot S_{0,1}f(v) \\ &= \sum_v \mu(v) \left(\mathbb{E}_{e \ni v} f(e) \right) \left(\mathbb{E}_{e' \perp v} f(e') \right) \\ &\geq \sum_v \mu(v) \left(\mathbb{E}_{e \ni v} f(e) \right) \cdot \epsilon_v \\ &\geq \sum_{v \in V_\tau} \mu(v) \cdot \tau \cdot \min(\epsilon_0, \rho_0^{-1}(\tau)) \\ &= \tau \min(\epsilon_0, \rho_0^{-1}(\tau)) \mathbb{P}[V_\tau] \\ &\geq \tau \min(\epsilon_0, \rho_0^{-1}(\tau)) (\delta/2 - \gamma - \tau) \|f\|^2. \end{aligned}$$

where the second inequality is because whenever $v \in V_\tau$, Claim 44 implies that $\epsilon_v \geq \min(\epsilon_0, \rho_0^{-1}(\tau))$.

Lemma 10 gives a bound of 3γ on the second largest eigenvalue of \tilde{M} . We finally apply Lemma 5 on the graph corresponding to \tilde{M} to deduce that

$$|R| \geq (\tau(\delta/2 - \gamma - \tau) \min(\epsilon_0, \rho_0^{-1}(\tau)) - 3\gamma) |X(1)|. \quad (22)$$

Choosing for example $\tau = \delta/4$ and as long as $\gamma < \min(\delta/8, \frac{\delta^2}{128} \min(\epsilon_0, \rho_0^{-1}(\frac{\delta}{4})))$ we deduce $|R| > \frac{\delta^2}{128} \min(\epsilon_0, \rho_0^{-1}(\frac{\delta}{4})) \cdot |X(1)| = \epsilon |X(1)|$. \blacktriangleleft

5.3 Testability of HDX codes in Dimensions Above Two

In this section we prove a “trickle-down” statement for agreement-testability. We show that if the local codes at i -links are agreement-testable, then this implies agreement-testability for the codes at $i - 1$ links.

► **Lemma 46.** *Let $\delta, \gamma > 0$, and let X be a k -dimensional γ -one-sided local expander, and assume that for every $t \in X(k-1)$ we have a code $C_t \subset \{f : X_{+t}(k) \rightarrow \mathbb{F}\}$ with minimum relative distance δ . Let D_i denote the maximal number of $i+1$ faces that touch an i face in X . Define, for every $-1 \leq i \leq k-2$ face $s \in X(i)$, the code*

$$C_s = \{f : X_{+s}(k) \rightarrow \mathbb{F} \mid f|_{X_{+t}} \in C_t \ \forall t \in X_{+s}(k-1)\},$$

and assume that for $s \in X(i)$ the code C_s has minimum relative distance δ_i . If there is $\epsilon_{k-2} > 0$ and a monotone increasing function $\rho_{k-2}(\cdot)$ such that for every $t \in X(k-2)$ the code C_t is $(\epsilon_{k-2}, \rho_{k-2}(\cdot))$ -agreement testable, then for every $-1 \leq i < k-2$ and every $s \in X(i)$ the code C_s is $(\epsilon_i, \rho_i(\cdot))$ -agreement testable with

$$\epsilon_i = \frac{\delta_{i+1}^2}{128} \min \left(\epsilon_{i+1}, \rho_{i+1}^{-1} \left(\frac{\delta_{i+1}}{4} \right) \right), \quad \rho_i(t) = D_{i+1}t$$

as long as $\gamma < \min \left(\frac{\delta_{i+1}}{8}, \frac{\delta_{i+1}^2}{128} \min \left(\epsilon_{i+1}, \rho_{i+1}^{-1}(\delta_{i+1}/4) \right) \right)$ for all i . In particular, the code $C_\emptyset \subset \{f : X(k) \rightarrow \mathbb{F}\}$ is $(\epsilon_{-1}, \rho_{-1}(\cdot))$ -agreement testable.

Proof. The proof is very similar to the proof of Theorem 41. Fix $s \in X(i)$. Our goal is to prove that C_s is testable assuming that for all $v \in X_s(0)$, $C_{s \cup \{v\}}$ is testable. By bijecting X_{+s} to X_s (mapping each $t \in X_{+s}$ to $t \setminus s$), we can move to the case where $s = \emptyset$, and our codes sit on the faces of dimension $k-i-1$, namely $C_v \subseteq \mathbb{F}^{X_{+v}(k-i-1)}$ and $C_\emptyset \subseteq \mathbb{F}^{X(k-i-1)}$. The complex X_s is a γ high dimensional expander, and each vertex v touches at most D_{i+1} edges. We also let δ_{i+1} denote the distance of the code $C_{s \cup \{v\}}$ (see Lemma 29 for a calculation).

When $i = k-3$, this was done in the previous section. So the only difference is that now C_v itself is a $(k-i-1)$ -dimensional HDX code for possibly $k-i-1 > 1$. In fact we will see that this makes almost no difference.

We start with $z^0 = \{z_v^0 \in C_v\}_{v \in X(0)}$ a collection of local views, and define $\alpha(z^0) = \mathbb{P}_{uv \in X(1)}[z_u(X_{+uv}) \neq z_v(X_{+uv})]$. Same as in the proof of Theorem 41, we run the local algorithm, replacing z_v with $z'_v \in C_v$ whenever it reduces $\alpha(z)$, until no more changes can be made. Let $z = \{z_v\}_{v \in X(0)}$ be the final collection.

Then, by following the same steps as in the proof of Theorem 41, we have the following:

- If $\alpha(z) = 0$, then just as in Claim 43 we have that z corresponds to a codeword $\hat{x} \in C$ satisfying $\mathbb{P}_{v \in X(0)}[z_v^0 \neq \hat{x}|_{X_{+v}}] \leq D_{i+1}\alpha(z^0)$.
- If $\alpha(z) \neq 0$, then the goal is to show that there must have been many edge disagreements to begin with. Define $R = \{uv \in X(1) \mid z_u|_{X_{+uv}} \neq z_v|_{X_{+uv}}\}$. Then as long as $\gamma < \min \left(\frac{\delta_{i+1}}{8}, \frac{\delta_{i+1}^2}{128} \min \left(\epsilon_{i+1}, \rho_{i+1}^{-1}(\delta_{i+1}/4) \right) \right)$, it holds that

$$|R| \geq \frac{\delta_{i+1}^2}{128} \min \left(\epsilon_{i+1}, \rho_{i+1}^{-1} \left(\frac{\delta_{i+1}}{4} \right) \right) \cdot |X(1)| = \epsilon_i |X(1)|.$$

Therefore, the number of edge disagreements to begin with in the original ensemble z^0 must have been at least $\epsilon_i |X(1)|$ also. ◀

6 Codes in Higher Dimensions

The coset complexes considered here have a higher dimensional version as follows. Fix $k > 2$ and define $H_i = \{h_i(\alpha) \mid \alpha \in \mathbb{F}\}$ where $h_i(\alpha) = e_{i,i+1}(\alpha t) + I_{k+1}$ and let $K_i = \text{span}(H_j : j \neq i)$ and $G = \text{span}(H_1, \dots, H_{k+1})$. Let X be the k -dimensional complex $X[G; K_1, \dots, K_{k+1}]$.

We have the following properties of X :

- X is a γ -expander, where $\gamma = \frac{1}{\sqrt{|\mathbb{F}|-(k-1)}}$. We use the trickle down theorem together with the fact that the link of any $t \in X(d-2)$ is either a $\frac{1}{\sqrt{|\mathbb{F}|}}$ -expander (Claim 23) or a complete bipartite graph (justification in proof of Lemma 48).
- For any $t \in X(i)$, the number of $(i+1)$ -faces that touch t is at most $D_i = (k-i) \cdot |\mathbb{F}|^{k-i-1}$. The reason is that $t \in X(i)$ corresponds to a coset of the group generated by $k-i$ subgroups $H_{j_1}, \dots, H_{j_{k-i}}$. Each $(i+1)$ face that touches t is a coset of the group generated by $k-i-1$ of those subgroups. For each such collection of $k-i-1$ subgroups, there are at most $|\mathbb{F}|^{k-i-1}$ cosets of the resulting group contained within t .

Define like before an embedding of the group elements of X into a vector space

$$G \rightarrow R^{(k+1)^2} \cong \mathbb{F}^{n(k+1)^2}$$

so that the elements in each gH_i embed to an entire affine line. Here we will be working with fields $\mathbb{F} = \mathbb{F}_p$ that have prime order. Fix a degree parameter $k\gamma|\mathbb{F}| < d < |\mathbb{F}|/4$ and define for every $t \in X(k-1)$ the code C_t to be the Reed-Solomon code $RS(|\mathbb{F}|, d)$. Further define, for every $i < k-1$ and every face $w \in X(i)$,

$$C_w = \{f : X_{+w}(k) \rightarrow \mathbb{F} \mid f|_{X_{+t}} \in C_t \forall t \in X_{+w}(k-1)\}.$$

It is immediate that the code $C = \mathcal{C}^k[X, \{C_v\}_{v \in X(0)}] \subseteq \mathbb{F}^{X(k)}$ is an HDX code as defined in Section 2.5. Furthermore, by Lemma 29, for any $w \in X(i)$, the code C_w has distance $\geq \delta_i = \prod_{j=0}^{k-1-i} (\delta - j\gamma)$.

► **Theorem 47.** *Let \mathbb{F} be a field of prime order, and let $X = X[G; K_1, \dots, K_{k+1}]$ be a k -dimensional complex. If $k\gamma|\mathbb{F}| < d < (\frac{1}{4} - \frac{1}{8}\delta_{k-2})|\mathbb{F}|$, then for every $i < k-1$ and every $w \in X(i)$, the code C_w is (ϵ_i, ρ_i) -agreement testable, where*

$$\epsilon_{k-2} = \left(\frac{1}{8} - \frac{d}{2p}\right)^3 \quad \text{and} \quad \rho_{k-2}(\cdot) = 4(\cdot)^{1/3},$$

and for $-1 \leq i < k-2$,

$$\epsilon_i = \frac{\delta_{k-2}^3 \cdot \prod_{j=i+1}^{k-2} \delta_j^2}{2^5 \cdot 2^{7(k-i-1)} \cdot (k+1) \cdot |\mathbb{F}|^k} \quad \text{and} \quad \rho_i = D_{i+1} \cdot (\cdot),$$

where $\delta_i = \prod_{j=0}^{k-1-i} (\delta - j\gamma)$ and $D_i = (k-i) \cdot |\mathbb{F}|^{k-i-1}$.

In particular, the code $C_\phi \subset \{f : X(k) \rightarrow \mathbb{F}\}$ is $(\epsilon_{-1}, \rho_{-1}(\cdot))$ -agreement testable.

The rest of this section is dedicated to proving Theorem 47. Suppose first that $i = k-2$ and $s \in X(k-2)$. Recall that X is $(k+1)$ -partite, and denote $\text{color}(s) \subset [k+1]$ the set of colors of s . For $k > 2$ we observe two kinds of links X_s , and therefore two kinds of codes C_s , depending on the color of s .

► **Lemma 48.** *Let $s \in X(k-2)$. Let $\text{color}(s) = [k+1] \setminus \{i, j\}$. If $|i-j| \equiv 1 \pmod{k+1}$ then C_s is isomorphic to $C_{d,d}$ as defined in (15). Otherwise, $C_s \cong RS(|\mathbb{F}|, r)^{\otimes 2}$.*

In both cases C_s is $\left(\left(\frac{1}{8} - \frac{d}{2p}\right)^3, 4(\cdot)^{1/3}\right)$ -agreement testable.

Proof. When $|i-j| > 1 \pmod{(k+1)}$ the subgroups H_i and H_j commute since $[h_i(\alpha), h_j(\beta)] = 0$. Then $\text{span}(H_i, H_j) \cong \mathbb{F}^2$, and $C_s \cong RS(|\mathbb{F}|, d)^{\otimes 2}$. Also, $\left(\left(\frac{p-2d}{2p}\right)^2, 2(\cdot)\right)$ -agreement testability was proven in [32] as long as $d < |\mathbb{F}|/2$. Note that $\left(\frac{p-2d}{2p}\right)^2 > \left(\frac{1}{8} - \frac{d}{2p}\right)^3$ when $0 \leq d \leq p/4$ and $2(\cdot) \leq 4(\cdot)^{1/3}$.

When $|i-j| = 1 \pmod{(k+1)}$, we can ignore a large part of the matrices (which is identity). For example, if $i = 1$ and $j = 2$ then we can restrict attention to the first 3 rows and columns of the matrices. Thus, we are back in the $k = 2$ case, with $\text{span}(H_i, H_j)$ isomorphic to the group K_{6-i-j} defined in Section 3.1, and the code C_s is isomorphic to $C_{d,d}$ as defined in (15). In this case, by Corollary 39, C_s is $\left(\left(\frac{1}{8} - \frac{d}{2p}\right)^3, 4(\cdot)^{1/3}\right)$ -agreement testable when $d < |\mathbb{F}|/4$. \blacktriangleleft

Moving to $i < k-2$, our proof relies on reverse induction, deducing agreement testability of the level i codes from agreement testability of the level $i+1$ codes.

► **Lemma 49.** *If $d < \left(\frac{1}{4} - \frac{1}{8}\delta_{k-2}\right)|\mathbb{F}|$, then for every $-1 \leq i < k-2$ and every $s \in X(i)$ the code C_s is $(\epsilon_i, \rho_i(\cdot))$ -agreement testable with*

$$\epsilon_i = \frac{\delta_{k-2}^3 \cdot \prod_{j=i+1}^{k-2} \delta_j^2}{2^5 \cdot 2^{7(k-i-1)} \cdot (k+1) \cdot |\mathbb{F}|^k} \quad \text{and} \quad \rho_i(x) = D_{i+1} \cdot x,$$

where $\delta_i = \prod_{j=0}^{k-1-i} (\delta - j\gamma)$ and $D_i = (k-i) \cdot |\mathbb{F}|^{k-i-1}$.

Proof. We have from Lemma 48 that for any $s \in X(k-2)$ C_s is $(\epsilon_{k-2}, \rho_{k-2}(\cdot))$ -agreement testable, where $\epsilon_{k-2} = \left(\frac{1}{2} - \frac{d}{2p}\right)^3$ and $\rho_{k-2}(\cdot) = 4(\cdot)^{1/3}$. Then for $s \in X(k-3)$, we have by Lemma 46 that C_s is $(\epsilon_{k-3}, \rho_{k-3}(\cdot))$ -agreement testable where $\rho_{k-3}(x) = D_{k-2} \cdot x$ and

$$\begin{aligned} \epsilon_{k-3} &= \frac{\delta_{k-2}^2}{128} \min(\epsilon_{k-2}, \rho_{k-2}^{-1}(\delta_{k-2}/4)) \\ &= \frac{\delta_{k-2}^2}{128} \min\left(\left(\frac{1}{8} - \frac{d}{2p}\right)^3, (\delta_{k-2}/16)^3\right) \\ &= \frac{\delta_{k-2}^5}{2^{19}} \\ &\geq \frac{\delta_{k-2}^5}{2^{19}D_{-1}}, \end{aligned}$$

where the third equality holds whenever $\frac{1}{8} - \frac{d}{2p} > \frac{\delta_{k-2}}{16}$, which happens whenever $d < \left(\frac{1}{4} - \frac{1}{8}\delta_{k-2}\right)|\mathbb{F}|$.

In general, for $s \in X(i)$, the code C_s is $(\epsilon_i, \rho_i(\cdot))$ -agreement testable where $\rho_i(x) = D_{i+1} \cdot x$ and

$$\epsilon_i = \frac{\delta_{i+1}^2}{128} \min\left(\epsilon_{i+1}, \frac{\delta_{i+1}}{4D_{i+2}}\right) \geq \frac{\delta_{k-2}^3 \cdot \prod_{j=i+1}^{k-2} \delta_j^2}{2^5 \cdot 2^{7(k-i-1)} \cdot D_{-1}}$$

since

$$\epsilon_{i+1} = \frac{\delta_{k-2}^3 \cdot \prod_{j=i+2}^{k-2} \delta_j^2}{2^5 \cdot 2^{k-i-2} \cdot D_{-1}} < \frac{\delta_{i+1}}{4D_{i+2}}.$$

Plugging in $D_{-1} = (k+1)|\mathbb{F}|^k$ finishes the calculation. \blacktriangleleft

References

- 1 S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. 6
- 2 S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. doi:10.1145/273865.273901. 6
- 3 Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *Proc. 20th International Workshop on Randomization and Computation (RANDOM)*, volume 116, pages 37:1–37:21, Princeton, NJ, 2018. RANDOM/APPROX. doi:10.4230/LIPIcs.APPROX/RANDOM.2018.37. 7
- 4 Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. Locally testable codes via high-dimensional expanders. *Electron. Colloquium Comput. Complex.*, TR20-072, 2020. URL: <https://eccc.weizmann.ac.il/report/2020/072>. 2
- 5 Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. Locally testable codes via high-dimensional expanders. *arXiv preprint*, 2020. arXiv:2005.01045. 5
- 6 Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022. doi:10.1145/3519935.3520024. 2, 4, 5, 7
- 7 Irit Dinur, Prahladh Harsha, Tali Kaufman, and Noga Ron-zewi. From local to robust testing via agreement testing. In *Proceedings of the 11th Innovations in Theoretical Computer Science (ITCS) Conference*, pages 29:1–29:18, 2019. doi:10.4230/LIPIcs.ITCS.2019.29. 2
- 8 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. doi:10.1109/FOCS.2017.94. 2, 9
- 9 Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proc. 48th ACM Symp. on Theory of Computing (STOC)*, pages 36–48, 2016. doi:10.1145/2897518.2897543. 4, 5
- 10 Uriya A First and Tali Kaufman. On good 2-query locally testable codes from sheaves on high dimensional expanders. *arXiv preprint*, 2022. doi:10.48550/arXiv.2208.01778. 4, 5, 7, 13, 14
- 11 S. Luna Frank-Fischer, Venkatesan Guruswami, and Mary Wootters. Locality via partially lifted codes. *CoRR*, abs/1704.08627, 2017. arXiv:1704.08627. 6
- 12 Ira Gessel and Gérard Viennot. Binomial determinants, paths, and hook length formulae. *Advances in Mathematics*, 58(3):300–321, 1985. doi:10.1016/0001-8708(85)90121-5. 24
- 13 Louis Golowich. From grassmannian to simplicial high-dimensional expanders. *CoRR*, abs/2305.02512, 2023. doi:10.48550/arXiv.2305.02512. 7
- 14 Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In Robert D. Kleinberg, editor, *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 529–540. ACM, 2013. doi:10.1145/2422436.2422494. 6
- 15 Prahladh Harsha and Ramprasad Saptharishi. A note on the elementary construction of high-dimensional expanders of kaufman and oppenheim. *arXiv*, December 2019. arXiv:1912.11225. 9
- 16 Tom Hoholdt and Jorn Justesen. Graph codes with Reed-Solomon component codes. In *2006 IEEE International Symposium on Information Theory*, pages 2022–2026, 2006. doi:10.1109/ISIT.2006.261904. 1, 5, 23
- 17 Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Ramanujan complexes and bounded degree topological expanders. In *Proc. 55th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 484–493, 2014. doi:10.1109/FOCS.2014.58. 4, 5

- 18 Tali Kaufman and Alexander Lubotzky. Edge transitive ramanujan graphs and symmetric LDPC good codes. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 359–366. ACM, 2012. doi:10.1145/2213977.2214011. 2
- 19 Tali Kaufman and Izhar Oppenheim. Construction of new local spectral high dimensional expanders. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 773–786, 2018. doi:10.1145/3188745.3188782. 3, 7, 9, 14, 15
- 20 Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Comb.*, 40(2):245–281, 2020. doi:10.1007/s00493-019-3847-0. 9
- 21 Tali Kaufman and Ran J. Tessler. Garland’s technique for posets and high dimensional grassmannian expanders. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA*, volume 251 of *LIPIcs*, pages 78:1–78:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ITCS.2023.78. 7
- 22 Tali Kaufman and Avi Wigderson. Symmetric LDPC codes and local testing. *Comb.*, 36(1):91–120, 2016. doi:10.1007/s00493-014-2715-1. 2
- 23 Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 592–601. IEEE Computer Society, 2018. doi:10.1109/FOCS.2018.00062. 7
- 24 Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of Ramanujan complexes of type \tilde{A}_d . *European J. Combin.*, 26(6):965–993, 2005. doi:10.1016/j.ejc.2004.06.007. 2, 7
- 25 Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Ramanujan complexes of type \tilde{A}_d . *Israel J. Math.*, 149(1):267–299, 2005. doi:10.1007/BF02772543. 2, 7
- 26 C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859–868, October 1992. doi:10.1145/146585.146605. 6
- 27 Or Meir. $IP = PSPACE$ using error-correcting codes. *SIAM J. Comput.*, 42(1):380–403, 2013. doi:10.1137/110829660. 6
- 28 Roy Meshulam. Graph codes and local systems, 2018. arXiv:1803.05643. 13
- 29 Dana Moshkovitz and Ran Raz. Sub-constant error low degree test of almost-linear size. *SIAM J. Computing*, 38(1):140–180, 2008. doi:10.1137/060656838. 7
- 30 Ryan O’Donnell and Kevin Pratt. High-dimensional expanders from chevalley groups. In Shachar Lovett, editor, *37th Computational Complexity Conference, CCC 2022, July 20-23, 2022, Philadelphia, PA, USA*, volume 234 of *LIPIcs*, pages 18:1–18:26. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CCC.2022.18. 7, 9
- 31 Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC ’22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022. doi:10.1145/3519935.3520017. 2, 4, 7
- 32 Alexander Polishchuk and Daniel A. Spielman. Nearly-linear size holographic proofs. In Frank Thomson Leighton and Michael T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 194–203. ACM, 1994. doi:10.1145/195058.195132. 5, 40
- 33 R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *STOC 1997*, pages 475–484, 1997. 7
- 34 A. Shamir. $IP = PSPACE$. *Journal of the ACM*, 39(4):869–877, October 1992. Prelim. version in 1990 FOCS, pages 11–15. doi:10.1145/146585.146609. 6
- 35 Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996. doi:10.1109/18.556667. 11, 13