Electronic Colloquium on Computational Complexity, Report No. 151 (2021)

Locally Testable Codes with constant rate, distance, and locality

Irit Dinur*[1] Shai Evra†[2] Ron Livne[2] Alexander Lubotzky‡[1] Shahar Mozes§[2]

[1]Weizmann Institute, Rehovot, Israel
[2]Hebrew University, Jerusalem, Israel

November 8, 2021

Abstract

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 proportional to their distance from the code. The parameter q is called the locality of the tester.

LTCs were initially studied as important components of PCPs, and since then the topic has evolved on its own. High rate LTCs could be useful in practice: before attempting to decode a received word, one can save time by first quickly testing if it is close to the code.

An outstanding open question has been whether there exist "c^3-LTCs", namely LTCs with **c**onstant rate, **c**onstant distance, and **c**onstant locality.

In this work we construct such codes based on a new two-dimensional complex which we call a left-right Cayley complex. This is essentially a graph which, in addition to vertices and edges, also has squares. Our codes can be viewed as a two-dimensional version of (the one-dimensional) expander codes, where the codewords are functions on the squares rather than on the edges.

1 Introduction

A locally testable code (LTC) is an error correcting code that has a property-tester. The tester reads q bits (randomly - but not necessarily uniformly - chosen) from a given word, and rejects words with probability proportional to their distance from the code. The parameter q is called the locality of the tester.

A random code has, with high probability, constant rate and distance, but locality that is proportional to the length. This is true even for random LDPC codes [BHR05], and a priori the mere existence of codes with constant locality is not obvious. The first LTCs appear implicitly in works on program checking [BLR90] and on probabilistically

*I.D. acknowledges support by ERC grant 772839 and ISF grant 2073/21.

†S.E. is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

‡A.L. acknowledges support by ERC grant 882751 and a grant from the Institute for Advanced Study at Princeton.

§S.M. acknowledges support by ISF-Moked grant 2019/19.

ISSN 1433-8092

checkable proofs (PCPs) [BFL91, LFKN92, BFLS91, AS98, ALM+98]. A formal definition of an LTC appeared simultaneously in several places [BFLS91, RS96, FS13, Aro94] (see [Gol10] for a detailed history).

Spielman, in his PhD thesis [Spi96], discusses the possibility of having an error-correcting code that is locally testable (he uses the term 'checkable code') and explains its potential applicability: *"A checker would be able to read only a constant number of bits of a received signal and then estimate the chance that a decoder will be able to correct the errors, then the checker can instantly request a retransmission of that block, before the decoder has wasted its time trying to decode the message. Unfortunately all known codes with local-checkers have rate approaching zero."*

Goldreich and Sudan [GS06] initiated a systematic study of LTCs as objects of interest in their own right. Over the years better and better LTCs were constructed [PS94, GS06, BSVW03, BGH+06, BS05, Din07, KMRS17, GKdO+18], but, nevertheless, experts went back and forth on whether "c^3-LTCs" (namely, LTCs with **c**onstant rate, **c**onstant distance, and **c**onstant locality) are likely to exist, compare [Gol05] with [Gol10, Section 3.3.2].

We construct the first such family of LTCs,

Theorem 1.1. *For all $0 < r < 1$, there exist $\delta, \kappa > 0$ and $q \in \mathbb{N}$ and a polynomial-time construction of an infinite family of error-correcting codes $\{C_n\}$ with rate r and distance δ, such that for all n, C_n is κ-locally testable with q queries.*

Namely, every code C_n comes with a randomized local tester that reads at most q bits from a given word w and then accepts or rejects, such that

- *For all $w \in C_n$, $\mathbb{P}[accept] = 1$.*
- *For all $w \notin C_n$, $\mathbb{P}[reject] \geqslant \kappa \cdot \operatorname{dist}(w, C_n)$.*

We remark that [KMRS17, GKdO+18] have shown (see [GKdO+18, Section 1.2]) how to take an LTC with rate arbitrarily close to 1 and with constant distance, and construct a new LTC with rate and distance approaching the Gilbert-Varshamov bound, and only a constant overhead in the locality q. So the theorem above holds for all $r, \delta > 0$ that satisfy $r + h(\delta) < 1$ where $h(\cdot)$ is the binary entropy function.

Expander codes, one dimension up

The celebrated expander-codes of Sipser and Spielman [SS96] are a family of error-correcting codes constructed from a single base code $C_0 \subseteq \mathbb{F}_2^d$ and a family of d-regular expander graphs $G_n = (V_n, E_n)$ such that the code corresponding to G_n consists of functions on E_n such that for every vertex in V_n, the local view from the neighboring edges (assuming some arbitrary fixed ordering) is itself in the base code C_0,

$$C = \left\{ f : E_n \to \mathbb{F}_2 \,\middle|\, \forall v \in V_n, f|_{edges(v)} \in C_0 \right\}.$$

Similarly, our codes will also be defined via a fixed base-code and an infinite family of expander graphs. Our graphs will have, in addition to vertices and edges, also two-dimensional faces, called squares, where each square touches four edges and four vertices.

Our codewords are functions *on the squares* such that for every edge, the bits on the neighboring squares form a codeword in the base code. It is natural to view our code as a Tanner code [Tan81] with bits on the squares and constraints on the edges; whereas the expander-codes have bits on the edges and constraints on the vertices.

Inspecting our code on the set of squares neighboring a fixed vertex, we see an intermediate code, whose constraints come from the edges neighboring that vertex.

We thus have three codes for the three dimensions of links: the base code C_1 at the link of an edge, the intermediate code C_0 at the link of a vertex, and the global code C at the link of the empty face which is the set of all squares.

Left-Right Cayley Complex. Let us describe our construction of a graph-with-squares, namely a square complex (for a more formal description see Definition 3.1). Let G be a finite group with two sets of generators A, B. We define the left-right Cayley complex $X = Cay^2(A, G, B)$ as follows

- The vertices are $X(0) = G$.
- The edges are $X(1) = X^A(1) \sqcup X^B(1)$ where

$$X^A(1) = \{\{g, ag\} \mid \ g \in G, a \in A\}\,, \qquad X^B(1) = \{\{g, gb\} \mid \ g \in G, b \in B\}\,.$$

The fact that with A we multiply on the left, and with B we multiply on the right, gives a local commutativity which generates many four-cycles, namely, squares. Indeed for every a, g, b the graph has a cycle of length 4 with alternating A and B edges, given by the walk g, gb, agb, ag, g. We place a square for each of these four-cycles.

- The squares are a set of the following four-cycles in the graph,

$$X(2) = \{(g, gb, agb, ag, g) \mid \ g \in G, a \in A, b \in B\}\,.$$

  We denote by $[a, g, b]$ the square containing the edges $\{g, ag\}$ and $\{g, gb\}$. By changing the 'root' of the square we get $[a, g, b] = [a^{-1}, ag, b] = [a^{-1}, ab, b^{-1}] = [a, gb, b^{-1}]$.

The Code. Fix a left-right Cayley complex $X = Cay^2(A, G, B)$, and fix a pair of base codes $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$ (assuming $|A| = |B| = d$ we can take both to be isomorphic to some $C_1 \subseteq \mathbb{F}_2^d$). Our code is defined to be

$$C[A, G, B, C_A, C_B] = \{f : X(2) \to \mathbb{F}_2 \mid \ \forall a, g, b,\ f([\cdot, g, b]) \in C_A,\ \text{and}\ f([a, g, \cdot]) \in C_B\}\,.$$

Observe that for a codeword $f \in C$ and a fixed vertex $g \in G$, the restriction of f to the squares touching g is $f([\cdot, g, \cdot])$. It is not difficult to check that this word necessarily belongs to the tensor code $C_A \otimes C_B$, see Lemma 4.1. Thus, by putting the constraints around each edge, we get an intermediate code on the squares touching a vertex, which turns out to be a tensor code! Tensor codes have non-trivial dependencies among the constraints defining them. This often implies local testability of tensor codes [BS06, DSW06, BV09], and turns out important for showing that our code can be locally tested by the following simple test:

Local test: Choose a random vertex g, and accept iff $f([\cdot, g, \cdot]) \in C_A \otimes C_B$.

We discuss below the type of local to global propagation that goes into proving that this test works.

Finally, to complete our construction of locally testable codes and to prove Theorem 1.1, we describe in Section 6 an explicit construction of a family of groups and pairs of generating sets which give good left-right Cayley complexes, and in Section 5 a matching choice of base codes C_A, C_B.

Propagation from local to global. Sipser and Spielman proved distance of their expander codes [SS96] through propagation: expansion of the underlying graph is used to "lift" the distance of the base code to the distance of the global code. In our codes distance is shown via a similar argument.

More interestingly, local testability of our codes is also shown through expansion. We show that if a received word violates only a small amount of constraints, then locally it can be corrected, as long as the intermediate code $C_A \otimes C_B$ is itself *robustly locally testable*. We describe an iterative decoding algorithm (Algorithm 1) and prove that it converges thanks to sufficient expansion of certain edge-to-edge random walks on our square complex. Conceptually, local local-testability (of the intermediate code $C_A \otimes C_B$), implies global local-testability (of the entire code), through expansion.

The existence of many dependencies among the constraints defining our codes is the point where our codes most clearly differ from expander codes: in expander codes one can have a single violated constraint that does not propagate, and leads to a word that is far from the code but no tester can detect it, as proven in [BHR05].

Locally Testable Codes: historical background and techniques

As mentioned earlier, the study of LTCs arose naturally in works on program checking and PCPs. The Hadamard code was the first code proven to be locally testable in the work of Blum, Luby, and Rubinfeld on linearity testing [BLR90]. The low (logarithmic) rate of this code was quickly improved to polynomial rate by moving from linear functions (codewords of the Hadamard code) to low degree polynomial functions (codewords of the Reed Muller code). Subsequent works studied "low degree tests" which are in fact proofs that the Reed-Muller code is locally testable. These works were crucial for progress leading up to the proof of the PCP theorem. More on the relation between PCPs and LTCs, as well as the historical development, can be found in Goldreich's survey [Gol10].

A systematic study of LTCs was initiated by Goldreich and Sudan in [GS06], and a sequence of works constructed both LTCs and PCPs with improved parameters [GS06, BSVW03, BGH$^+$06, BS08, Din07], achieving constant locality and distance, but rate $1/\operatorname{poly}\log n$. Some experts believed that low rate is inherently needed and some attempts to prove upper bounds on the rate have been made [BGK$^+$10, DK11, BSV12, BSS05], although these lower bounds are in rather restrictive models.

This, perhaps, has triggered works from the other end of the spectrum [KMRS17, GKdO$^+$18] which focused on constructing error correcting codes with constant rate and distance, that are locally testable with smallest possible locality. These works achieve constant rate and quasi-poly-logarithmic distance and locality.

In terms of techniques, many of the earlier constructions of LTCs have two notable features. Firstly, they are based on the properties of low degree polynomials, and secondly, they come hand in hand with a PCP constructions, so that both share the same composition-recursion structure.

The gap amplification technique [Din07] of the first author is a construction of both a PCP and an LTC that relies on expander graphs and concatenation and departs from the domain of low degree polynomials. Meir [Mei08] gave a tensor-code-based construction of LTCs that is neither related to low degree functions nor to PCPs altogether. Further works [KMRS17, GKdO$^+$18] also construct LTCs without any PCP counterpart.

A feature shared by all previous constructions of LTCs with mildly high rate is their recursive nature. One first constructs codes with weaker properties and then enhances them by concatenation, possibly with different iterations. The overall composed structure of the code is somewhat complicated and begs for a more direct "one-shot" construction.

A path leading towards a one-shot construction seemed to open up with the connection to high dimensional expanders.

High Dimensional Expansion

The current paper is mainly elementary and almost self-contained (with the exception of Section 6 which uses the existence of some Ramanujan Cayley graphs with specific properties and can be taken as a black box). But it came up as a result of a much longer and intensive journey. Some interesting open problems were left aside along the way. It is, therefore, worthwhile to give the story here.

The journey started by the first and fourth authors during a year-long program at the IIAS (Israeli Institute of Advanced Studies) on high dimensional expanders in 2017: the hope was to use the Ramanujan complexes (a la [LSV05b, LSV05a]) to construct LTCs as high-dimensional versions of expander-codes over Ramanujan graphs as explained above. Although expander codes are typically not locally testable [BHR05] the hope was that higher dimensional versions would be.

This optimistic belief was inspired by local to global behavior of certain high dimensional complexes that was uncovered already by Garland in his seminal work [Gar73].

In that paper, Garland proved a conjecture of Serre, that the cohomology of cocompact lattices in high-rank simple p-adic groups vanishes. Equivalently, if X is a finite simplicial quotient of a Bruhat-Tits building of dimension at least two, its cohomology vanishes. The proof of Garland is “local-to-global”: he showed that if the links of relevant cells have a spectral gap, then so does the global Laplacian of X. Namely, if X is locally an expander, then it is also globally so. (For a purely combinatorial treatment and generalizations - see [Opp18]). The global spectral gap implies the vanishing of the cohomology.

This “local to global” approach is a high-dimensional phenomenon that does not hold for graphs! In graphs, the local structure does not reveal any information about the global expansion. To illustrate this, the reader may recall the LPS-Ramanujan graphs [LPS88] which are (p+1)-regular expander graphs with large girth. One can easily get (p+1)-regular graphs with large girth (and hence locally isomorphic to the LPS ones) which are far from being expanders. In contrast, the Garland method shows that local expansion implies global expansion in the high dimensional case.

The local to global approach was also the key ingredient, in [KKL14, EK16] where Gromov’s overlapping problem was solved using the Ramanjaun complexes.

At this point there was already some interest from the theoretical computer science community. The fact that high dimensional expansion is related to property testing in computer science was observed for the first time in [KL14]. The first author and Kaufman proved that high dimensional expansion implies an efficient agreement-test [DK17], which is related to both PCPs and LTCs. Anari et al [ALOV19] resolved a conjecture regarding convergence of certain Markov chains by analyzing the global random walk through local analysis at the links.

Inspired by all this, the idea was to construct LTC codes by using the local-to-global behavior of the Ramanujan complexes in an analog to the way [SS96] used Ramanujan graphs for LDPC codes. For simplicity, we will describe it from now on only in dimension 2, but one can do the same in higher dimensions.

The original idea was as follows: fix a large prime p and take an infinite family of Ramanujan complexes X, quotients of the Bruhat-Tits building of $G = SL(3, \mathbb{Q}_p)$. The complex X is a 2-dimensional complex, the link of every edge of it is in one-to-one correspondence with the projective line $\mathbb{P}^1$ over $\mathbb{F}_p$ and the link of every vertex is the graph of lines versus points of the projective plane over $\mathbb{F}_p$. One can define a base code (“the small code”) C_1 on $\mathbb{P}^1$ to be a "projective" variant of the Reed-Solomon code. This code induces a "big code" C as a subspace of the $\mathbb{F}_p$ functions on $X(2)$- the 2-dimensional cells of X- whose local views at every edge are in the base code of the edge. The goal was then to propagate the rate, distance, and local-testability of Reed-Solomon codes from the small code C_1 to the big code C.

This turned out to be easier to say than to do. At some point, we were hoping to use p-adic uniformization. Recall the work of Mumford [Mum79] who used the combinatorial structure of one such Ramanujan complex to prove a result on algebraic surfaces appearing as locally symmetric quotients of $SU(2,1)$. We were hoping to go in the opposite direction and to use the theory of algebraic surfaces to study our combinatorial objects. The theory of p-adic uniformization was developed in depth by Varshavsky in his thesis [Var98] (written under the supervision of the 3rd author of the current paper). This is an opportunity to thank Yakov Varshavsky who gave upon our request a semester-long course describing this work. While we eventually are not using this, we were fortunate to be exposed to an amazing chapter of deep mathematics.

Propagating local testability from the small code to the big code when these are defined over a high dimensional expander is possible. This was proved in [DDFH18] with the hope that it would serve our original plan. For our codes to fit, the intermediate code, C_0 - the one that is defined on the link of a vertex through the small Reed-Solomon codes C_1 on the edges - needed to be itself locally-testable. Unfortunately we failed to prove that C_0 is locally testable. Here the problem is very concrete: Find C_1 inside $\mathbb{F}_p^{\mathbb{P}^1}$ such that the induced intermediate code C_0 on the link of a vertex is locally testable. Here, the link of a vertex is nothing but the lines versus points graph of the projective plane.

One can generalize this challenge to get such a code also on higher dimensional spherical buildings. This is interesting also in higher dimensions: are such spherical codes locally testable?

We, therefore, changed direction and replaced $G = SL(3, \mathbb{Q}_p)$ by a product $G = SL(2, \mathbb{Q}_p) \times SL(2, \mathbb{Q}_q)$. This time the quotients obtained from congruence lattices in G give rise to square complexes. These complexes were shown long ago to be Ramanujan cubical complexes [JL99] and the dynamic of walks along them was studied in [Moz91]. This time the local intermediate code look like tensor codes (since the link of every vertex is the *complete* bipartite graph) and there are plenty of tensor codes that are locally testable as mentioned above. A subtle obstacle arose at this point which does not exist in the graph codes of [SS96]: one needs to name the squares in such a way that the function defined on the link of an edge $\{u, v\}$ will be in or out the code independently if we look at it from the vertex u or the vertex v. It might be that this challenge can be overcome, but at that point, we realized that by changing from these square complexes to the left-right Cayley complexes as defined above, this problem is easily fixed. Moreover, it became also easier to argue about the rate- making the whole paper much simpler than we expected!

As explained, our long journey left a number of unsettled issues. We believe they are interesting in their own right (and in all dimensions) even if not needed anymore for the concrete goal of locally testable codes.

The left-right Cayley complexes seem objects that are worth studying for their own sake. It is actually somewhat surprising that in spite of over 100 years of studying Cayley graphs, these objects, as far as we know, have never been studied before. An immediate curiosity is whether there are higher-dimensional analogs or whether a group "has only two sides" and hence these exist only in dimension 2. Anyway, it seems that this paper solves one problem but opens many others.

Acknowledgements

We wish to thank Prahladh Harsha and Avi Wigderson for many interesting discussions along the way of this project. We also wish to thank Tali Kaufman for her influential role in connecting LTCs and high dimensional expansion.

This work was presented by the first author on October 6, 2021 at the Simon's Institute for the Theory of Computing [Din21a] as part of the lecture series on breakthroughs in computer science, and at the Institute for Advanced Study in Princeton on October 25-26, 2021 [Din21b]. It was also presented by the 4th author on October 27, 2021 at the Simon's HDX21 workshop [Lub21]. The authors are very grateful to these institutions and for the remarks of the audience which improved the exposition of the paper.

2 Preliminaries

2.1 Expander Graphs

A d-regular graph G is said to be a λ-one-sided expander if it has eigenvalues $d = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant -d$ which satisfy $\lambda_i \leqslant \lambda \cdot d$ for all $i > 1$.

The following is a standard lemma by Alon and Chung,

Lemma 2.1 ([AC88])**.** *Let $G = (V, E)$ be a d-regular λ one-sided expander. Let $T \subseteq V$ be such that the graph induced on T, denoted $G(T)$, has average degree at least δd. Then $|T| \geqslant (\delta - \lambda) \cdot |V|$, and the number of edges in $G(T)$ is at least $(\delta - \lambda)\delta \cdot |E|$.*

This lemma holds in more general situations where instead of a d-regular graph we have a weighted Markov operator as long as it has a basis of eigenvectors. Let $\mathcal{D}$ be any probability distribution over a finite set V, and define an inner product by

$$\langle \cdot, \cdot \rangle_{\mathcal{D}} : \mathbb{R}^V \times \mathbb{R}^V \to \mathbb{R}, \qquad \langle f, f' \rangle_{\mathcal{D}} = \mathop{\mathbb{E}}_{x \sim \mathcal{D}}[f(x) f'(x)].$$

Denote by $\mathbf{1} \in \mathbb{R}^V$ the constant 1 function.

Lemma 2.2. *Let $M : \mathbb{R}^V \to \mathbb{R}^V$ be a symmetric Markov operator such that $M\mathbf{1} = \mathbf{1}$, and such that for all f with $\langle f, \mathbf{1} \rangle_{\mathcal{D}} = 0$, $\langle f, Mf \rangle_{\mathcal{D}} \leqslant \lambda \langle f, f \rangle_{\mathcal{D}}$. Let $f = \mathbb{1}_T$ be the indicator of a set $T \subseteq V$, so that $\langle f, f \rangle_{\mathcal{D}} = \mathbb{P}_{\mathcal{D}}[T]$. If $\langle f, Mf \rangle_{\mathcal{D}} \geqslant \delta \cdot \langle f, f \rangle_{\mathcal{D}}$ then $\mathbb{P}_{\mathcal{D}}[T] \geqslant \delta - \lambda$, and $\langle f, Mf \rangle_{\mathcal{D}} \geqslant \delta(\delta - \lambda)$.*

Proof. Denote $p = \mathbb{P}_{\mathcal{D}}[T]$. We can write $f = p\mathbf{1} + f_\perp$ with $\langle f_\perp, \mathbf{1} \rangle_{\mathcal{D}} = 0$. We get

$$\delta \cdot p \leqslant \langle f, Mf \rangle_{\mathcal{D}} = \langle p\mathbf{1} + f_\perp, M(p\mathbf{1} + f_\perp) \rangle_{\mathcal{D}} = p^2 + \lambda \langle f_\perp, f_\perp \rangle_{\mathcal{D}} \leqslant p^2 + \lambda p.$$

which, when rearranging, gives the lemma. □

2.2 Error Correcting Codes

A linear code $C \subset \mathbb{F}_2^n$ is an $\mathbb{F}_2$-linear subspace of $\mathbb{F}_2^n$. The block-length of the code is n. The rate and distance of the code are the relative dimension of the code and relative

Hamming weight of the smallest weight non-zero codeword, respectively, namely,

$$\mathrm{Rate}(C) = \frac{1}{n}\dim(C) \qquad \text{and} \qquad \mathrm{dist}(C) = \frac{1}{n}\min_{w\in C-\{0\}} |\{i \in [n] \mid \ w_i \neq 0\}|.$$

We recall the definition of locally testable codes from [GS06]. The definition given here is that of a "strong" LTC, and implies all other definitions of locally testable codes. See [Gol17, Chapter 13].

Definition 2.3 (Locally Testable Code (LTC))**.** For $\kappa > 0$ and $q \in \mathbb{N}$ we say that an error-correcting code $C \subseteq \mathbb{F}_2^n$ is *κ-locally testable with q queries* if there is a distribution over a collection of q-element subsets $S \subset [n]$ such that each subset S is associated with a set $V_S \subset \mathbb{F}_2^S$ of allowed local views, and such that, denoting by $f|_S$ the restriction of f to the set S, the following hold.

– If $f \in C$ then for every S, $f|_S \in V_S$.

– For every $f \in \mathbb{F}_2^n$,
$$\kappa \cdot \mathrm{dist}(f, C) \leqslant \mathop{\mathbb{P}}_S[f|_S \notin V_S].$$

Definition 2.4 (Tensor Code)**.** Let $n_1, n_2 \in \mathbb{N}$ and let $C_i \subset \{f : [n_i] \to \mathbb{F}_2\}$ for $i = 1, 2$ be two linear codes. Define the tensor code $C = C_1 \otimes C_2$ by

$$C = \{M : [n_1] \times [n_2] \to \mathbb{F}_2 \mid \ \forall i \in [n_1], j \in [n_2], M(i, \cdot) \in C_2, M(\cdot, j) \in C_1\}.$$

It is easy to check that $\dim(C_1 \otimes C_2) = \dim(C_1) \cdot \dim(C_2)$, and that $\mathrm{dist}(C_1 \otimes C_2) = \mathrm{dist}(C_1)\,\mathrm{dist}(C_2)$. We view the elements of C as n_1-by-n_2 matrices and write $w(i, \cdot) \in \mathbb{F}_2^{n_2}$ for the i-th row of w, and similarly $w(\cdot, j) \in \mathbb{F}_2^{n_1}$ is the j-th column of w.

A natural *test* for whether a given matrix $f \in \mathbb{F}_2^{n_1\times n_2}$ is in $C_1 \otimes C_2$ is as follows:

Randomly choose a row or a column, and check whether the restriction of f to that column (or row) is in C_1 (or C_2).

The quality of the test is measured by the relation between the rejection probability and the distance of f from the tensor code. Formally, this is captured by the notion of robust-testability.

Definition 2.5 (Robust testability of tensor codes)**.** Fix $C_i \subseteq \mathbb{F}_2^{n_i}$ linear error correcting codes for $i = 1, 2$. For $f : [n_1] \times [n_2] \to \mathbb{F}_2$, let

$$\delta^{\mathrm{col}}(f) = \mathrm{dist}(f, C_1 \otimes \mathbb{F}_2^{n_2}), \qquad \delta^{\mathrm{row}}(f) = \mathrm{dist}(f, \mathbb{F}_2^{n_1} \otimes C_2).$$

and
$$\rho(f) = (\delta^{\mathrm{col}}(f) + \delta^{\mathrm{row}}(f))/2.$$

The robust testability of $C_1 \otimes C_2$ is defined to be

$$\rho = \min_{f\notin C_1\otimes C_2} \frac{\rho(f)}{\mathrm{dist}(f, C_1 \otimes C_2)},$$

and we say that $C_1 \otimes C_2$ is ρ-robustly testable.

The robust testability of tensor codes was first studied in [BS06], where it was shown that for any code C with sufficiently high distance, the d-dimensional tensor code $C^{\otimes d}$ is robustly testable for all $d \geqslant 3$. The requirement $d \geqslant 3$ was puzzling because the tensor of Reed Solomon codes is known [PS94] to be robustly testable even for $d = 2$ and this was considered the prototype for locally testable codes. Surprisingly, Paul Valiant discovered

[Val05] that there are codes C for which $C \otimes C$ is *not* robustly testable, see also [GM12]. Quickly after that [DSW06] formulated a notion of smooth codes, broadened later to 'weakly smooth' in [BV09], and showed that the tensor product of a smooth code and any other code is in fact robustly testable. To define smooth codes recall the definition of LDPC codes,

Definition 2.6 (LDPC code)**.** Let $c, d, n \in \mathbb{N}$. A (c, d, n)-LDPC code is given by a (c, d)-regular bipartite graph $([n], [m], E)$ (called a factor graph) with n left vertices and $m = nc/d$ right vertices, called parity checks, such that all right vertices have degree d and all left vertices have degree c. The code is defined to be

$$C = \left\{ w : [n] \to \mathbb{F}_2 \;\middle|\; \forall j \in [m], \sum_{i:ij \in E} w(i) = 0 \mod 2 \right\}.$$

Definition 2.7 (Smooth code)**.** Let $c, d, n \in \mathbb{N}, \alpha, \beta, \delta > 0$. A (c, d, n) LDPC code $C \subset \mathbb{F}_2^n$ is (α, β, δ)-smooth if for every $Y \subseteq [n]$ with $|Y| \leqslant \alpha \cdot m$ there is some $X \subseteq [n]$ with $|X| \leqslant \beta \cdot n$ such that the code $C(\bar{Y})|_{\bar{X}}$ has distance at least δ. Here the code $C(\bar{Y})|_{\bar{X}}$ is the code obtained by removing the constraints in Y and then removing the coordinates in X.

In Section 5 we show that random low density parity check codes (LDPC) are smooth.

Agreement-Testability. A related testing notion focuses on the agreement between pairs of overlapping local views. We think of the following situation,

- For each column we are given a codeword of C_1, and these are aggregated into $w_1 \in C_1 \otimes \mathbb{F}_2^{n_2}$.
- For each row we are given a codeword of C_2, and these are aggregated into $w_2 \in \mathbb{F}_2^{n_1} \otimes C_2$.
- We check "agreement", namely, pick a random pair of row i and column j, and check whether they agree on their intersection, i.e. whether

$$w_1(i, j) \overset{?}{=} w_2(i, j).$$

Testability is defined to be the ratio between the amount of pairwise disagreement to the distance from the code. Formally,

Definition 2.8 (agreement-testability)**.** Let $\kappa > 0$. Let $C_i \subset \{f : [n_i] \to \mathbb{F}_2\}$ for $i = 1, 2$. We say that $C_1 \otimes C_2$ is κ-agreement-testable if for every $w_1 \in C_1 \otimes \mathbb{F}_2^{n_2}$, $w_2 \in \mathbb{F}_2^{n_1} \otimes C_2$, there exists $w \in C_1 \otimes C_2$ such that

$$\kappa \cdot (\mathop{\mathbb{P}}_i[w_1(i, \cdot) \neq w(i, \cdot)] + \mathop{\mathbb{P}}_j[w_2(\cdot, j) \neq w(\cdot, j)]) \leqslant \mathop{\mathbb{P}}_{i \in [n_1], j \in [n_2]}[w_1(i, j) \neq w_2(i, j)].$$

In words, given a word w_1 whose rows are in C_1, and given a word w_2 whose columns are in C_2, we say that $C_1 \otimes C_2$ is κ-agreement-testable if the amount of disagreement between w_1 and w_2 is an upper-bound for the fraction of rows or columns one needs to change in order to get to the closest word $w \in C_1 \otimes C_2$, times κ.

It is well known (see for example [DH09]) that agreement testability is equivalent to robust testability:

Lemma 2.9. *Let $C_i \subseteq \mathbb{F}_2^{n_i}$, and assume $\delta_i = \mathrm{dist}(C_i)$ for $i = 1, 2$.*

- *If $C_1 \otimes C_2$ is ρ-robustly testable then it is κ-agreement testable, for $\kappa^{-1} = \frac{1}{2\delta_1\rho} + \frac{1+1/(2\rho)}{\delta_2}$.*
- *If $C_1 \otimes C_2$ is κ-agreement testable, then it is ρ-robustly testable for $\rho = \frac{\kappa}{2(\kappa+1)}$.*

We prove this lemma in Appendix A.

3 The Left-Right Cayley Complex

We describe a new construction of a Cayley graph that in addition to vertices and edges also has two-dimensional faces, called squares. Each square contains four edges that constitute a four-cycle.

Definition 3.1 (Left-Right Cayley Complex)**.** Let G be a finite group with two symmetric sets of generators A, B, namely, each is closed under taking inverses. We assume that the identity element of G is neither in A nor in B. Define the *Left-Right Cayley Complex* $X = Cay^2(A, G, B)$ as follows

- The vertices are $X(0) = G$.
- The edges are $X(1) = X^A(1) \sqcup X^B(1)$ where
$$X^A(1) = \{\{g, ag\} \mid \; g \in G, a \in A\}, \qquad X^B(1) = \{\{g, gb\} \mid \; g \in G, b \in B\}.$$
- The squares are $X(2) = A \times G \times B/\sim$ where for every $a \in A, b \in B, g \in G$,
$$(a, g, b) \sim (a^{-1}, ag, b) \sim (a^{-1}, agb, b^{-1}) \sim (a, gb, b^{-1}),$$
and denote the equivalence class of (a, g, b) by $[a, g, b]$, so
$$[a, g, b] = \{(a, g, b), (a^{-1}, ag, b), (a^{-1}, agb, b^{-1}), (a, gb, b^{-1})\}.$$

The graph $(X(0), X^A(1))$ is none other than the Cayley graph $Cay(G, A)$. Similary $(X(0), X^B(1))$ is the Cayley graph $Cay(G, B)$. The fact that with A we multiply on the left, and with B we multiply on the right, gives a local commutativity which generates many four-cycles, namely, squares.

Remark 3.2. Given a group G and a set of generators A, the Cayley graph $Cay^{left}(G, A)$ with left-multiplication edges is isomorphic to the Cayley graph $Cay^{right}(G, A)$ with right multiplication edges via the map $g \mapsto g^{-1}$. The left-multiplication edge $\{g, ag\}$ maps to the right multiplication edge $\{g^{-1}, g^{-1}a^{-1}\}$. This justifies talking about a Cayley graph without specifying left or right multiplication.

Remark 3.3. The product of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is a square complex $X = G_1 \times G_2$ defined as follows.

- The vertices are $X(0) = V_1 \times V_2$.
- The edges are $X(1) = E_1 \times V_2 \;\sqcup\; V_1 \times E_2$, where an edge $(\{u, u'\}, v) \in E_1 \times V_2$ connects (u, v) with (u', v), and similarly an edge $(u, \{v, v'\}) \in V_1 \times E_2$ connects (u, v) with (u, v').
- The squares $X(2)$ are identified with $E_1 \times E_2$, so that the square corresponding to the pair of edges $e_1 = \{u, u'\} \in E_1$ and $e_2 = \{v, v'\} \in E_2$ is the four-cycle $(u, v) \to (u, v') \to (u', v') \to (u', v) \to (u, v)$.

The left-right Cayley complex is the quotient of the Cartesian product of $G_A = (G, X^A(1))$ and $G_B = (G, X^B(1))$ obtained by identifying the vertex (g, g') with $(gh, h^{-1}g')$ for all $h \in G$. One can check that the map $(g, g') \mapsto gg'$ gives a homomorphism from $G_A \times G_B$ to $Cay^2(A, G, B)$.

Remark 3.4. Left-right Cayley complexes are examples of two-dimensional cubical complexes. Cubical complexes are well-studied, and in particular there are constructions of Ramanujan cubical complexes [JL99] with bounded degree and any dimension, whose walk dynamics was studied in [Moz91]. The left-right Cayley complexes have an additional matching labels feature that other complexes are not known to have.

Definition 3.5 (Links)**.** For each $g \in G$, the link of g is $X_g = \{[a, g, b] \mid \ a \in A, b \in B\}$. There is a natural map $(a, b) \mapsto [a, g, b]$.

For every edge $e = \{g, ag\}$, the link of e is denoted $X_e = \{[a, g, b] \mid \ b \in B\}$. Similarly if $e = \{g, gb\}$ we let $X_e = \{[a, g, b] \mid \ a \in A\}$.

Definition 3.6. A left-right Cayley complex satisfies the total no-conjugacy condition if

$$\forall a \in A, b \in B, g \in G, \qquad g^{-1}ag \neq b. \tag{TNC}$$

Here are a few easy properties of left-right Cayley complexes.

Claim 3.7. Assuming (TNC), each vertex has exactly $|A| + |B|$ distinct neighbors; and each square contains exactly four distinct vertices; and the map $(a, b) \mapsto [a, g, b]$ is a bijection from $A \times B$ to X_g for each $g \in G$.

Proof. Let $a \neq a' \in A$ and $b \neq b' \in B$. Clearly $ag \neq a'g$ and $gb \neq gb'$. If $ag = gb$ then $g^{-1}ag = b$ which contradicts (TNC). So g has $|A| + |B|$ distinct neighbors. Next we show that each square $[a, g, b] \in X(2)$ must have four distinct vertices. Clearly $g \neq ag$ and $g \neq gb$, and we already saw that $ag \neq gb$. Now, if $g = agb$ we would contradict (TNC) because it implies $g^{-1}a^{-1}g = b$ making $a^{-1} \in A$ and $b \in B$ conjugates.

Finally, let us see that the map $(a, b) \mapsto [a, g, b]$ is a bijection between $A \times B$ and X_g for all g. Assume that $[a, g, b] = [a', g, b']$ for some $(a, b), (a', b') \in A \times B$. This implies that $(a', g, b') \in \{(a, g, b), (a^{-1}, ag, b), (a^{-1}, agb, b^{-1}), (a, gb, b^{-1})\}$. We have seen that $g \neq ag, gb, agb$ so this implies that $(a, g, b) = (a', g, b')$ which means that $(a, b) = (a', b')$. □

Remark 3.8. It follows that assuming (TNC)

$$|X(1)| = \frac{|A| + |B|}{2} \cdot |G| \quad \text{and} \quad |X(2)| = \frac{|A||B|}{4} \cdot |G|.$$

It will be natural to consider a weighted version of the 1-skeleton of X, where the weight is distributed evenly between the A and the B edges. When $|A| = |B|$ this is the usual unweighted graph.

Definition 3.9. Let $\mathcal{D}_1$ be the distribution over $X(1)$ given by selecting with probability half a uniform edge in $X^A(1)$, and with probability half a uniform edge in $X^B(1)$. (In case A, B have equal size $\mathcal{D}_1$ is the uniform distribution over $X(1)$.)

We define an inner product on functions over $X(1)$ $f, f' : X(1) \to \mathbb{R}$ by

$$\langle f, f' \rangle_{\mathcal{D}_1} = \mathop{\mathbb{E}}_{e \sim \mathcal{D}_1}[f(e)f'(e)] = \frac{1}{2} \mathop{\mathbb{E}}_{e \in X^A(1)}[f(e)f'(e)] + \frac{1}{2} \mathop{\mathbb{E}}_{e \in X^B(1)}[f(e)f'(e)]. \tag{3.1}$$

This will be the only inner product we consider for functions over $X(1)$ so we sometimes omit the subscript and simply write $\langle f, f' \rangle = \langle f, f' \rangle_{\mathcal{D}_1}$. As usual we let $\|f\| = \langle f, f \rangle$.

Parallel Random Walk. In addition to the standard random walk on the 1-skeleton of X, we will be interested in a random walk on the edges called the parallel walk, which takes an edge e to a random edge e' that is "parallel" to it.

Definition 3.10 (Labels)**.** For each $s \in A \cup B$ let $[s] = \{s, s^{-1}\}$. Let $\tilde{A} = \{[a] \mid \ a \in A\}$ and let $\tilde{B} = \{[b] \mid \ b \in B\}$. The label of an edge $\{g, ag\}$ is defined to be $\{a, a^{-1}\}$, and this is independent of the presentation of the edge as $\{g, ag\}$ or $\{(ag), a^{-1}(ag)\}$. Similarly, the label of an edge $\{g, gb\}$ is defined to be $\{b, b^{-1}\}$.

Let $\tilde{A} \cup \tilde{B}$ denote the set of labels of the edges in the complex. For any $\sigma \in \tilde{A} \cup \tilde{B}$, denote by $X^{\sigma}(1)$ the set of edges labelled σ.

Claim 3.11*.* If $\sigma = \{c, c^{-1}\} \in \tilde{A} \cup \tilde{B}$ and $c \neq c^{-1}$, then $X^{\sigma}(1)$ has size $|G|$, otherwise it has size $|G|/2$.

Proof. We shall prove the claim for $\sigma = \{a, a^{-1}\} \in \tilde{A}$, the claim for $\sigma \in \tilde{B}$ is proven analogously. Observe that every vertex g participates in two edges labelled $\sigma = \{a, a^{-1}\}$, namely $\{g, ag\}$ and $\{g, a^{-1}g\}$. Since every edge is counted twice, from each of its two endpoints, we get that $|G| = |X^{\sigma}(1)|$. [1]

In case $a = a^{-1}$each vertex participates in only a single edge labelled $[a]$, but still every edge has two endpoints so after accounting for the double counting we get $|X^{\sigma}(1)| = |G|/2$. □

Define a Markov operator $M^{||}_{\sigma} : \mathbb{R}^{X^{\sigma}(1)} \to \mathbb{R}^{X^{\sigma}(1)}$ on the space of functions on $X^{\sigma}(1)$ by setting, for any $f : X^{\sigma}(1) \to \mathbb{R}$,

$$M^{||}_{[a]} f(\{g, ag\}) = \mathop{\mathbb{E}}_{b} f(\{gb, agb\}), \qquad M^{||}_{[b]} f(\{g, gb\}) = \mathop{\mathbb{E}}_{a} f(\{ag, agb\}).$$

We define a Markov operator $M^{||} : \mathbb{R}^{X^{\sigma}(1)} \to \mathbb{R}^{X^{\sigma}(1)}$ on the space of functions on the entire set of edges $X(1)$ by letting, for any $f : X(1) \to \mathbb{R}$,

$$M^{||} f = \sum_{\sigma} M^{||}_{\sigma} (f|_{X^{\sigma}(1)}). \tag{3.2}$$

Definition 3.12 (Parallel Random Walk)**.** We define a random walk on the set of edges $X(1)$ as follows. Starting from an edge e, choose uniformly a square containing e and then move to the unique edge $e' \neq e$ with the same label as e. (If (TNC) doesn't hold the square might not contain an edge $e' \neq e$ with the same label, in which case the walk will stay in place).

The Markov operator corresponding to this walk is exactly $M^{||}$, because starting at an edge $e = \{g, ag\}$, a random square containing e is $[a, g, b]$ for a uniformly chosen $b \in B$, and then the only other $[a]$-labeled edge in this square is the edge $e' = \{gb, agb\}$.

Lemma 3.13. *Assume both $Cay(G, A)$ and $Cay(G, B)$ are λ-expanders. Suppose $R \subseteq X(1)$ and assume $f = \mathbf{1}_R : X(1) \to \mathbb{R}$ satisfies $\langle f, M^{||} f\rangle \geqslant c \cdot \langle f, f\rangle$. Then there exists some $\sigma \in \tilde{A} \cup \tilde{B}$ such that*

$$|R \cap X^{\sigma}(1)| \geqslant (c - \lambda)|G|/2.$$

[1] Note also that when $a \neq a^{-1}$ the operator $M^{||}_{[a]}$ on $X^{\sigma}(1)$ is isomorphic to the standard random walk on $Cay(G, B)$, and similarly if $b \neq b^{-1}$ then $M^{||}_{[b]}$ is isomorphic to $Cay(G, A)$.

Proof. We expand $\langle f, M^{||} f\rangle$ according to (3.2), and get

$$\langle f, M^{||} f\rangle = \mathop{\mathbb{E}}_{\sigma} \mathop{\mathbb{E}}_{e \in X^{\sigma}(1)} [f(e) M^{||}_{\sigma} f(e)],$$

where the expectation over σ is obtained by choosing, with probability half, a random label in $\tilde{A}$, and with probability half, a random label in $\tilde{B}$. Clearly also

$$\langle f, f\rangle = \mathop{\mathbb{E}}_{e \sim \mathcal{D}_1} [f(e)^2] = \mathop{\mathbb{E}}_{\sigma} \mathop{\mathbb{E}}_{e \in X^{\sigma}(1)} [f(e)^2].$$

Plugging these into the inequality $\langle f, M^{||} f\rangle - c \cdot \langle f, f\rangle \geqslant 0$ we get

$$\mathop{\mathbb{E}}_{\sigma} \mathop{\mathbb{E}}_{e \in X^{\sigma}(1)} \left[f(e) M^{||}_{\sigma} f(e) - c \cdot f(e)^2 \right] \geqslant 0$$

so there must be at least one σ for which

$$\mathop{\mathbb{E}}_{e \in X^{\sigma}(1)} [f(e) M^{||}_{\sigma} f(e)] \geqslant c \cdot \mathop{\mathbb{E}}_{e \in X^{\sigma}(1)} [f(e)^2]. \tag{3.3}$$

Fix, say, $\sigma = [a]$ and define $h_a : G \to \mathbb{R}$ by $h_a(g) = f(\{g, ag\})$. (The case $\sigma = [b]$ is analogous and omitted). Now,

$$c \cdot \langle h_a, h_a\rangle = c \cdot \mathop{\mathbb{E}}_{g} [h_a(g)^2] = c \cdot \mathop{\mathbb{E}}_{g} [f(\{g, ag\})^2] = c \cdot \mathop{\mathbb{E}}_{e \in X^{[a]}(1)} [f(e)^2] \leqslant \mathop{\mathbb{E}}_{e \in X^{[a]}(1)} [f(e) M^{||}_{\sigma} f(e)]$$

$$= \mathop{\mathbb{E}}_{g \in G} \left[f(\{g, ag\}) \mathop{\mathbb{E}}_{b \in B} [f(\{gb, agb\})] \right] = \mathop{\mathbb{E}}_{g \in G} \left[h_a(g) \mathop{\mathbb{E}}_{b \in B} h_a(gb) \right] = \langle h_a, M_B h_a\rangle, \tag{3.4}$$

where M_B is the random walk operator on $Cay(G, B)$. We relied here on the fact that choosing a uniform edge in $X^{[a]}(1)$ can be done by choosing a uniform $g \in G$ and looking at $\{g, ag\}$. Observe now that h_a indicates the set $T = \{g \in G \mid f(\{g, ag\}) \neq 0\}$, so by Lemma 2.2 applied on the graph $Cay(G, B)$ with the operator M_B we deduce that $|T| \geqslant (c - \lambda)|G|$. Since every non-zero value for f can cause at most two non-zero values in h_a, we get that $|R \cap X^{\sigma}(1)| = |f^{-1}(1) \cap X^{\sigma}(1)| \geqslant |h_a^{-1}(1)|/2 = |T|/2 \geqslant (c - \lambda) \cdot |G|/2$. □

4 Error Correcting Code on a Left-Right Cayley Complex

Let G, A, B and $X = Cay^2(G, A, B)$ as in the previous section. Recall that for any vertex $g \in X(0)$ (resp. any edge $e \in X(1)$) we denote by $X_g \subset X(2)$ (resp. $X_e \subset X(2)$) the set of squares in X containing the vertex g (resp. the edge e). Let $C_A \subset \mathbb{F}_2^A$ and let $C_B \subset \mathbb{F}_2^B$ be two fixed linear error correcting codes with rates $\rho_A = \text{Rate}(C_A), \rho_B = \text{Rate}(C_B)$ and distances $\delta_A = \text{dist}(C_A), \delta_B = \text{dist}(C_B)$, respectively.

Define the code $C = C[G, A, B, C_A, C_B]$ as follows. For an edge $e = \{g, ag\} \in X^A(1)$ we define a local code

$$C_e = \{f : X_e \to \mathbb{F}_2 \mid f([a, g, \cdot]) \in C_B\}.$$

Similarly, for an edge $e = \{g, gb\} \in X^B(1)$ we define a local code

$$C_e = \{f : X_e \to \mathbb{F}_2 \mid f([\cdot, g, b]) \in C_A\}.$$

Note that this definition appears to depend on the choice of $g \in e$ but it does not. Finally, we define a global code

$$C = \{f : X(2) \to \mathbb{F}_2 \mid \forall e \in X(1), f|_{X_e} \in C_e\}.$$

For each vertex $g \in X(0)$, define the local tensor code around the vertex g to be

$$C_g = \{f : X_g \to \mathbb{F}_2 \mid f([\cdot, g, \cdot]) \in C_A \otimes C_B\}.$$

Lemma 4.1 (C is a lifted tensor-code)**.**

$$C = \left\{ f : X(2) \to \mathbb{F}_2 \,\middle|\, \forall g \in X(2), f|_{X_g} \in C_g \right\}.$$

Proof. Immediate from the fact that $f([\cdot, g, \cdot]) \in C_A \otimes C_B$ for any $g \in X(0)$ if and only if $f([a, g, \cdot]) \in C_B$ and $f([\cdot, g, b]) \in C_A$ for any $g \in X(0)$, $a \in A$ and $b \in B$.

□

Observe that for the local code at each vertex to be a tensor code, we must make sure that around every A edge we have *the same* code C_A, and similarly for B. If we choose different base codes at different edges we might still get a code with rate and distance, but local testability will probably fail, because we lose the local tensor structure. This is in contrast to the case of expander codes where the local base code can be chosen arbitrarily and differently at each vertex.

4.1 Properties of The Code

We now look at the rate, distance and local testability of the code $C = C[G, A, B, C_A, C_B]$. Recall $\rho_A = \text{Rate}(C_A)$, $\rho_B = \text{Rate}(C_B)$ and $\delta_A = \text{dist}(C_A)$, $\delta_B = \text{dist}(C_B)$

Lemma 4.2 (Rate)**.** *The rate of the code C is bounded from below by*

$$Rate(C) \geqslant 2(\rho_A + \rho_B) - 3.$$

Proof. For each $e \in X^A(1)$, $codim(C_e) = codim(C_B) = |B| \cdot (1 - \rho_B)$. Similarly for each $e \in X^B(1)$, $codim(C_e) = codim(C_A) = |A| \cdot (1 - \rho_A)$. The number of linearly independent constraints on $f \in C$ is at most

$$|X^A(1)| \cdot |B|(1 - \rho_B) + |X^B(1)| \cdot |A|(1 - \rho_A) = |G||A||B|(1 - \frac{\rho_A + \rho_B}{2})$$

On the other hand, the dimension of the ambient space is the number of squares $|X(2)| = |G||A||B|/4$, see Remark 3.8. Subtracting the number of constraints from the number of bits we get a lower bound on the dimension of the code,

$$dim(C) \geqslant \frac{1}{4}|G||A||B|(1 - (4 - 2(\rho_A + \rho_B))) = \frac{1}{4}|G||A||B|(2(\rho_A + \rho_B) - 3).$$

□

In fact, we can do a little better. Recall that a *vertex cover* of a graph is a set of vertices that touch all of the edges. For example, if the graph is bipartite, then it has a vertex cover whose size is half the size of the graph.

Lemma 4.3 (Rate - better bound)**.** *Suppose the underlying graph of X has a vertex cover of size $\nu|G|$. Then the rate of the code is at least $4\nu\rho_A\rho_B + 1 - 4\nu$. In particular, if the graph is bipartite, then $\nu = \frac{1}{2}$ and we get that*

$$Rate(C) \geqslant 2\rho_A\rho_B - 1.$$

It is interesting to mention that in the expander codes of Tanner [Tan81], (whose distance and decoding was later analyzed in [SS96]), if the local code C_0 has rate ρ_0 then the global rate is shown to be at least $2\rho_0 - 1$. In our code the rate of the local code is $\text{Rate}(C_g) = \text{Rate}(C_A \otimes C_B) = \rho_A\rho_B$, and in case the graph is bipartite, we get the same bound of $2(\rho_A\rho_B) - 1$ on the rate of the global code.

Proof. Let $V^* \subset G$ be a vertex cover, namely, a set of vertices that touches every edge. Then $f \in C$ if and only if for every $g \in V^*$, $f|_{X_g} \in C_g$. The reason is that every edge e touches some $g \in V^*$ and the constraint $f|_{X_e} \in C_e$ is implied by $f|_{X_g} \in C_g$.

Since C_g is isomorphic to $C_A \otimes C_B$ it has $|A| \cdot |B|(1 - \rho_A \rho_A)$ linearly independent constraints. The dimension of the code is at least

$$\begin{aligned} \dim(C) &\geqslant |G||A||B|\frac{1}{4} - |V^*| \cdot |A| \cdot |B|(1 - \rho_A\rho_A) \\ &\geqslant \frac{1}{4}|G||A||B| \cdot (1 - 4\nu(1 - \rho_A\rho_B)) = \frac{1}{4}|G||A||B| \cdot (4\nu\rho_A\rho_B + 1 - 4\nu). \quad (4.1) \end{aligned}$$

□

Lemma 4.4 (Distance)**.** *Suppose that both $Cay(G, A)$ and $Cay(G, B)$ are λ-expanders for $\lambda < 1$. Then the distance of the code C is bounded from below by*

$$\operatorname{dist}(C) \geqslant \delta_A\delta_B \cdot (\min(\delta_A, \delta_B) - \lambda)\,.$$

Proof. Let $0 \neq f \in C$. Let $g_0 \in X(0)$ be some vertex such that $w_{g_0} = f|_{X_{g_0}} \neq 0$ (if they are all zero then $f = 0$). Observe that since $0 \neq w_{g_0} \in C_A \otimes C_B$ then w_{g_0} has at least $\delta_A|A|$ non zero columns and at least $\delta_B|B|$ non-zero rows. Let $A_1 \subset A$ be the labels of these columns, and fix $a_1 \in A_1$. We first show that

$$\mathop{\mathbb{P}}_{g,b}[f([a_1, g, b]) \neq 0] \geqslant \delta_B(\delta_B - \lambda). \quad (4.2)$$

To prove (4.2) consider the graph $Cay(G, B)$ whose vertices are $X(0)$ and the edges are $X^B(1)$, and define a function $f_{a_1} : X^B(1) \to \mathbb{F}_2$ by $f_{a_1}(\{g, gb\}) = f([a_1, g, b])$, observing that f_{a_1} is well defined because for $g' = gb$,

$$f_{a_1}(\{g, g'\}) = f_{a_1}(\{g, gb\}) = f([a_1, g, b]) = f([a_1, g', b^{-1}]) = f_{a_1}(\{g', g'b^{-1}\}) = f_{a_1}(\{g', g\}).$$

Since $f_{a_1} \neq 0$, it must have large weight because it belongs to the expander code defined on $Cay(G, B)$ with local code C_B. More elaborately, for every vertex g that touches an edge where $f_{a_1} \neq 0$, there must be at least $\delta_B|B|$ non-zero edges touching g. By Lemma 2.1 we get at least $\delta_B(\delta_B - \lambda)|X^B(1)|$ edges on which $f_{a_1} \neq 0$, which proves (4.2).

For every $a \in A_1$, the weight of f_a is at least $\delta_B(\delta_B - \lambda)$, so if we choose a random $a \in A$ and then a random edge in $X^B(1)$, the probability that $a \in A_1$ is at least δ_A, and conditioned on this, the probability that $f_a(e) \neq 0$ is at least $\delta_B(\delta_B - \lambda)$, so altogether

$$\mathop{\mathbb{P}}_{a,g,b}[f([a, g, b]) \neq 0] \geqslant \mathop{\mathbb{P}}_{a}[a \in A_1] \cdot \mathop{\mathbb{P}}_{g,b}[f_a(\{g, gb\}) \neq 0 \mid a \in A_1] \geqslant \delta_A\delta_B(\delta_B - \lambda).$$

Symmetrically, the weight of f is also at least $\delta_B\delta_A(\delta_A - \lambda)$, and the lemma follows. □

Theorem 4.5 (Local Testability)**.** *Suppose $X = Cay^2(A, G, B)$ is a left-right Cayley complex such that both $Cay(G, A)$ and $Cay(G, B)$ are λ-expanders, and such that* (TNC) *holds. Assume $C_A \subset \mathbb{F}_2^A$ and $C_B \subset \mathbb{F}_2^B$ are error correcting codes with relative distances $\delta_A, \delta_B > 0$ respectively and such that $C_A \otimes C_B$ is κ_0-agreement-testable. If*

$$c = \frac{\kappa_0}{8 + \kappa_0} \cdot \min(\delta_A, \delta_B) > \lambda \quad (4.3)$$

then $C = C[G, A, B, C_A, C_B]$ is κ-locally testable with $|A| \cdot |B|$ queries, for $\kappa^{-1} = \max(4(1 + |A| + |B|), \frac{2(|A|+|B|)}{c-\lambda})$. Namely, for every $f : X(2) \to \mathbb{F}_2$

$$\kappa \cdot \operatorname{dist}(f, C) \leqslant \mathop{\mathbb{P}}_{g \in X(0)}[f|_{X_g} \notin C_g].$$

In words, given some potential codeword f, each vertex g is associated with a local test that reads f at all of the $|A| \cdot |B|$ squares touching g and checks that these vaues form a codeword in the base code C_g. The theorem says that the distance of f to the code is upper bounded by a constant multiple of the fraction of violated local tests.

We prove the theorem in the next section, by describing an iterative correction algorithm that finds a codeword close to f if the probability that the test rejects is not too large.

4.2 Local Self-Correction Algorithm

In this section we describe a local self-correction algorithm that starts with a given string $f : X(2) \to \mathbb{F}_2$ and either finds a codeword $f_0 \in C$ or gives up. We denote

$$\rho(f) = \mathop{\mathbb{P}}_{g}(f|_{X_g} \notin C_g),$$

the fraction of rejecting local tests. We will show that if $\rho(f) \leqslant \rho_0$ for some constant $\rho_0 > 0$, then the algorithm finds $f_0 \in C$ such that $\mathrm{dist}(f_0, f) \leqslant O(\rho(f))$.

For each vertex g, let $\mathrm{w}_g \in C_g$ be a closest codeword to $f|_{X_g}$ (breaking ties arbitrarily). We focus on the collection of local views $\{\mathrm{w}_g\}$ and whether the local views of neighboring vertices agree on the common squares.

Definition 4.6. Given a collection of local views $\mathrm{w} = \{\mathrm{w}_g \in C_g \mid g \in G\}$, we define the disagreement of the collection to be

$$\mathcal{E}(\mathrm{w}) = \mathop{\mathbb{P}}_{e=\{g,g'\} \in X(1)}[\mathrm{w}_g|_{X_e} \neq \mathrm{w}_{g'}|_{X_e}]. \tag{4.4}$$

Algorithm 1: Iterative decoding algorithm. (input: $f : X(2) \to \mathbb{F}_2$)

```
1. (Initialization:) For each vertex $g$, let $\mathrm{w}_g \in C_g$ be a closest codeword to $f|_{X_g}$
   (breaking ties arbitrarily).
       $\mathrm{w}_g = argmin_{w \in C_g} \mathrm{dist}(w, f|_{X_g})$.
2. (Main loop:) If there is a vertex $g$ and a choice $w \in C_g$ that reduces $\mathcal{E}(\mathrm{w})$ then
   replace $\mathrm{w}_g$ by $w$ and repeat.
3. (End:) If $\mathcal{E}(\mathrm{w}) > 0$ output "far". Otherwise, $\mathcal{E}(\mathrm{w}) = 0$, define $f_0 : X(2) \to \mathbb{F}_2$ by
   choosing for each square $s \in X(2)$ an arbitrary vertex $g \in s$ and setting
   $f_0(s) = \mathrm{w}_g(s)$. Output $f_0$.
```

Observe that $\mathcal{E}(\mathrm{w})|X(1)|$ is a non-negative integer, and this value decreases by at least 1 every step of the algorithm, so the algorithm must halt.

Proposition 4.7. *If the algorithm outputs f_0 then $f_0 \in C$ and*

$$\mathrm{dist}(f, C) \leqslant \mathrm{dist}(f, f_0) \leqslant 4(1 + |A| + |B|) \cdot \rho(f).$$

Let $\mathrm{w}^1 = \{\mathrm{w}_g^1\}$ be the collection of local views initially defined in step 1 of the algorithm, and let $\mathrm{w} = \{\mathrm{w}_g\}$ be the final collection, at the end of the algorithm.

Proposition 4.8. *If the algorithm outputs "far" then* $\mathcal{E}(\mathrm{w}) \geqslant \varepsilon_0 = \frac{c-\lambda}{|A|+|B|}$*, where* $c = \frac{\kappa_0}{8+\kappa_0} \cdot \min(\delta_A, \delta_B)$ *is defined in* (4.3)*.*

We will show that this immediately means that $\rho(f) \geqslant \mathcal{E}(\mathrm{w})/2 \geqslant \frac{c-\lambda}{2(|A|+|B|)}$ and this in turn means that $\mathrm{dist}(f, C) \leqslant 1 \leqslant \frac{2(|A|+|B|)}{c-\lambda} \cdot \rho(f)$, which will prove the theorem.

Proof of Theorem 4.5. Given $f : X(2) \to \mathbb{F}_2$, run the algorithm above. The output is either a function f_0, which by Proposition 4.7, satisfies $\mathrm{dist}(f, C) \leqslant \mathrm{dist}(f, f_0) \leqslant 4(1+|A|+|B|) \cdot \rho(f)$; or the output is "far", in which case $\mathcal{E}(\mathrm{w}) \geqslant \varepsilon_0$ by Proposition 4.8. We observe that

$$\mathcal{E}(\mathrm{w}^1) \leqslant 2\rho(f). \tag{4.5}$$

The reason is that for each edge $\{g, g'\}$ that contributes to $\mathcal{E}$ either $f|_{X_g} \neq \mathrm{w}^1_g$ or $f|_{X_{g'}} \neq \mathrm{w}^1_{g'}$, otherwise

$$\mathrm{w}^1_g|_{X_{gg'}} = (f|_{X_g})|_{X_{gg'}} = f|_{X_{gg'}} = (f|_{X_{g'}})|_{X_{gg'}} = \mathrm{w}^1_{g'}|_{X_{gg'}}.$$

So the process of selecting an edge uniformly and then a random endpoint of it will lead to a rejecting vertex with probability at least $\mathcal{E}(\mathrm{w}^1)/2$, proving (4.5). Now $\rho(f) \geqslant \mathcal{E}(\mathrm{w}^1)/2 \geqslant \mathcal{E}(\mathrm{w})/2 \geqslant \varepsilon_0/2 = \frac{c-\lambda}{2(|A|+|B|)}$, so we can write

$$\mathrm{dist}(f, C) \leqslant 1 \leqslant \frac{2(|A|+|B|)}{c-\lambda} \cdot \rho(f).$$

All in all we get,

$$\mathrm{dist}(f, C) \leqslant \max(4(1+|A|+|B|), \frac{2(|A|+|B|)}{(c-\lambda)}) \cdot \rho(f) = \kappa \cdot \mathop{\mathbb{P}}_{g}(f|_{X_g} \notin C_g)$$

as needed. □

Remark 4.9. Algorithm 1 is clearly also a decoding algorithm in the standard sense: if we know that the given word f is close enough to the code, then the regular structure of the tester implies that it will be rejected with probability proportional to $\mathrm{dist}(f, C)$. The analysis herein shows that for small enough (constant) distance, the algorithm will then find the nearest codeword.

The difficulty in our analysis is to show the same without any a priori guarantee on the distance of f from the code.

We now turn to prove the two propositions.

Proof of Proposition 4.7. By assumption, $\mathcal{E}(\mathrm{w}) = 0$. We first observe that the value of $f_0(s)$ does not depend on the choice of $g \in s$ because $\mathcal{E}(\mathrm{w}) = 0$ implies that $\mathrm{w}_g(s) = \mathrm{w}_{g'}(s)$ for any $g, g' \in s$. (Suppose $g_1, g_2 \in s$ disagree. If they are adjacent this means that w_{g_1} disagrees with w_{g_2} contradicting $\mathcal{E}(\mathrm{w}) = 0$. If they are non-adjacent, they have a common neighbor which cannot agree with both of them). It follows that $f_0 \in C$, because for each g, $f_0|_{X_g} = \mathrm{w}_g \in C_g$. To bound $\mathrm{dist}(f, f_0)$, let

$$V_0 = \left\{g \in X(0) \,\middle|\, f|_{X_g} \neq \mathrm{w}^1_g\right\}, \qquad V_1 = \left\{g \in X(0) \,\middle|\, \mathrm{w}^1_g \neq \mathrm{w}_g\right\}.$$

So V_0 is the set of vertices whose local view doesn't perfectly satisfy the constraints of the code, and V_1 is the set of vertices g for which w_g at the end of the algorithm differs from its initial value.

Observe that $g \in V_0$ iff $f|_{X_g} \notin C_g$, so by definition,

$$|V_0| = \rho(f) \cdot |X(0)|. \tag{4.6}$$

Any square s that does not touch $V_0 \cup V_1$ must have for every $g \in s$

$$f_0(s) = \mathrm{w}_g(s) = \mathrm{w}_g^0(s) = f(s),$$

where the second equality is because $g \notin V_1$ and the third is because $g \notin V_0$. We bound $|V_1|$ by the number of iterations of the algorithm, which is at most $|V_1| \leqslant \mathcal{E}(\mathrm{w}^1) \cdot |X(1)|$. We recall from (4.5) that $\mathcal{E}(\mathrm{w}^1) \leqslant 2\rho(f)$. Thus, we have,

$$|V_1| \leqslant \mathcal{E}(\mathrm{w}^1) \cdot |X(1)| \leqslant 2\rho(f) \cdot \frac{|A| + |B|}{2}|X(0)|. \tag{4.7}$$

Altogether, since every vertex touches $|A||B|$ squares, and since $|X(2)| = |A||B||X(0)|/4$, and using (4.6) and (4.7), we get

$$\mathrm{dist}(f, f_0) \leqslant \frac{|A||B| \cdot |V_0 \cup V_1|}{|X(2)|} = \frac{4 \cdot |V_0 \cup V_1|}{|X(0)|} \leqslant 4(1 + |A| + |B|)\rho(f).$$

□

The interesting part of the proof is to show that if $\mathcal{E}(\mathrm{w}) > 0$ after the algorithm ends, then $\mathcal{E}(\mathrm{w}) > \varepsilon_0 = \frac{c-\lambda}{|A|+|B|}$.

Proof of Proposition 4.8. Let

$$R = \{e = \{g, g'\} \in X(1) \mid \ \mathrm{w}_g|_{X_e} \neq \mathrm{w}_{g'}|_{X_e}\}$$

be the set of "dispute" edges. The rest of the proof is aimed towards showing $\mathcal{E}(\mathrm{w}) \geqslant \varepsilon_0$ or equivalently, since $\mathcal{E}(\mathrm{w}) = |R|/|X(1)|$, that

$$|R| \geqslant \frac{c - \lambda}{|A| + |B|} \cdot |X(1)| = \frac{c - \lambda}{2} \cdot |G|. \tag{4.8}$$

First some more notations. For an edge $\{g, ag\} \in X^A(1)$ let

$$E^{||}(\{g, ag\}) = \left\{\{gb, agb\} \in X^A(1) \,\middle|\, \ b \in B\right\}$$

and similarly for an edge $\{g, gb\} \in X^B(1)$,

$$E^{||}(\{g, gb\}) = \left\{\{ag, agb\} \in X^B(1) \,\middle|\, \ a \in A\right\}.$$

For a vertex g, let

$$E^A(g) = \{\{g, ag\} \mid \ a \in A\}, \qquad E^B(g) = \{\{g, gb\} \mid \ b \in B\}.$$

We now make two claims on the local structure of R. The first is due to the local distance, and the second is due to the local testability of tensor codes.

Claim 4.10. Suppose $\{g, ag\} \in R$, then

$$|R \cap E^B(g)| + |R \cap E^B(ag)| + |R \cap E^{||}\{g, ag\}| \geqslant \delta_B|B|.$$

Similarly, suppose $\{g, gb\} \in R$, then

$$|R \cap E^A(g)| + |R \cap E^A(gb)| + |R \cap E^{||}\{g, gb\}| \geqslant \delta_A|A|.$$

Proof. Let $e = \{g, ag\} \in R$, so $\mathrm{w}_g|_{X_e} \neq \mathrm{w}_{ag}|_{X_e}$. Since $\mathrm{w}_g|_{X_e}, \mathrm{w}_{ag}|_{X_e} \in C_e$, these are two distinct codewords of C_e, and must disagree on at least $\delta_B|B|$ squares. Let $[a, g, b]$ be such a square, and look at the three edges of the square that are not e: $\{g, gb\}, \{gb, agb\}$ and $\{agb, ag\}$. At least one of the three edges must be in R, because $\mathrm{w}_g, \mathrm{w}_{gb}, \mathrm{w}_{agb}, \mathrm{w}_{ag}$ cannot all agree on the value of $[a, g, b]$ without contradicting $\mathrm{w}_g([a, g, b]) \neq \mathrm{w}_{ag}([a, g, b])$. This implies the first part of the claim, and the second part is proven similarly. □

Recall that we assume $C_A \otimes C_B$ is agreement-testable, as per Definition 2.8.

Claim 4.11. Assume $C_A \otimes C_B$ is κ_0-agreement-testable. For every $g \in G$,

$$\underset{a}{\mathbb{P}}[\{g, ag\} \in R] + \underset{b}{\mathbb{P}}[\{g, gb\} \in R] \leqslant \kappa_0^{-1} \cdot \underset{a\in A, b\in B}{\mathbb{P}}[\{ag, agb\} \in R \text{ or } \{gb, agb\} \in R]. \quad (4.9)$$

Proof. Define $w_0, w_1, w_2 : A \times B \to \mathbb{F}_2$ as follows. First, let $w_0(a, b) = \mathrm{w}_g([a, g, b])$. Next, let $w_1(a, b) = \mathrm{w}_{ag}([a^{-1}, ag, b])$. Similarly let $w_2(a, b) = \mathrm{w}_{gb}([a, gb, b^{-1}])$. In words, the ath row of w_1 comes from the "opinion" of w_{ag}, and the bth column of w_2 comes from the "opinion" of w_{gb}. Observe that $w_0 \in C_A \otimes C_B$, $w_1 \in \mathbb{F}_2^A \otimes C_B$, and $w_2 \in C_A \otimes \mathbb{F}_2^B$. Now observe that $w_1(a, \cdot) \neq w_0(a, \cdot)$ iff $\{g, ag\} \in R$, and $w_2(\cdot, b) \neq w_0(\cdot, b)$ iff $\{g, gb\} \in R$. Finally, $w_1(a, b) \neq w_2(a, b)$ implies that the event on the RHS of (4.9) holds, namely, $\{ag, agb\} \in R$ or $\{gb, agb\} \in R$.

By the κ_0-agreement-testability of $C_A \otimes C_B$, there is a word $w \in C_A \otimes C_B$ such that

$$\underset{a}{\mathbb{P}}[w(a, \cdot) \neq w_1(a, \cdot)] + \underset{b}{\mathbb{P}}[w(\cdot, b) \neq w_2(\cdot, b)] \leqslant \kappa_0^{-1} \cdot \underset{a,b}{\mathbb{P}}[w_1(a, b) \neq w_2(a, b)].$$

Since the iterative algorithm has terminated, we know that

$$\underset{a}{\mathbb{P}}[w_0(a, \cdot) \neq w_1(a, \cdot)] + \underset{b}{\mathbb{P}}[w_0(\cdot, b) \neq w_2(\cdot, b)] \leqslant \underset{a}{\mathbb{P}}[w(a, \cdot) \neq w_1(a, \cdot)] + \underset{b}{\mathbb{P}}[w(\cdot, b) \neq w_2(\cdot, b)]$$

otherwise the algorithm would have flipped from $\mathrm{w}_g = w_0$ to $\mathrm{w}_g = w$. Combining the inequalities the claim follows,

$$\begin{aligned}\underset{a}{\mathbb{P}}[\{g, ag\} \in R] + \underset{b}{\mathbb{P}}[\{g, gb\} \in R] &= \underset{a}{\mathbb{P}}[w_0(a, \cdot) \neq w_1(a, \cdot)] + \underset{b}{\mathbb{P}}[w_0(\cdot, b) \neq w_2(\cdot, b)] \\ &\leqslant \underset{a}{\mathbb{P}}[w(a, \cdot) \neq w_1(a, \cdot)] + \underset{b}{\mathbb{P}}[w(\cdot, b) \neq w_2(\cdot, b)] \\ &\leqslant \kappa_0^{-1} \cdot \underset{a,b}{\mathbb{P}}[w_1(a, b) \neq w_2(a, b)] \\ &\leqslant \kappa_0^{-1} \cdot \underset{a\in A, b\in B}{\mathbb{P}}[\{ag, agb\} \in R \text{ or } \{gb, agb\} \in R].\end{aligned}$$

□

Let $M_0 = \frac{1}{2}M_A + \frac{1}{2}M_B$. Clearly for any $f : X(0) \to \mathbb{R}$ such that $E[f] = 0$, $\langle f, M_0 f\rangle = \frac{1}{2}\langle f, M_A f\rangle + \frac{1}{2}\langle f, M_B f\rangle \leqslant \lambda\langle f, f\rangle$. Recall the distribution $\mathcal{D}_1$ over $X(1)$ from Definition 3.9 and the corresponding inner product $\langle \cdot, \cdot\rangle_{\mathcal{D}_1}$. Define $\mathsf{D} : \mathbb{R}^{X(1)} \to \mathbb{R}^{X(0)}$, $\mathsf{U} : \mathbb{R}^{X(0)} \to \mathbb{R}^{X(1)}$ to be down and up operators, moving us from functions on edges to functions on vertices and vice versa. Namely,

$$\forall f_1 \in \mathbb{R}^{X(1)}, \qquad \mathsf{D}f_1(g) = \underset{e\sim\mathcal{D}_1|g}{\mathbb{E}}[f_1(e)] = \frac{1}{2}\underset{a\in A}{\mathbb{E}}[f_1(\{g, ag\})] + \frac{1}{2}\underset{b\in B}{\mathbb{E}}[f_1(\{g, gb\})]$$

and

$$\forall f_0 \in \mathbb{R}^{X(0)}, \qquad \mathsf{U}f_0(\{g_1, g_2\}) = \underset{g\in\{g_1,g_2\}}{\mathbb{E}}[f_0(g)] = \frac{1}{2}(f_0(g_1) + f_0(g_2)).$$

Note that these are averaging operators so they never increase norms, e.g. $\|\mathsf{D}f\| \leqslant \|f\|$ for all f.

Claim 4.12. Let $M = \mathsf{U}M_0\mathsf{D} : \mathbb{R}^{X(1)} \to \mathbb{R}^{X(1)}$. Then M has second largest eigenvalue at most λ.

Proof. We rely on the fact that $\mathcal{D}_1$ can be described by first choosing a uniform vertex g and then a random edge containing g such that with probability half we choose an A edge and with probability half a B edge. For any $f_1 : X(1) \to \mathbb{R}$ and $f_0 : X(0) \to \mathbb{R}$ we have

$$\langle \mathsf{D}f_1, f_0\rangle = \mathop{\mathbb{E}}_{g}[\mathop{\mathbb{E}}_{e\sim\mathcal{D}_1|g}[f_1(e)] \cdot f_0(g)] = \mathop{\mathbb{E}}_{e\sim\mathcal{D}_1}[f_1(e) \mathop{\mathbb{E}}_{g\in e}[f_0(g)]] = \langle f_1, \mathsf{U}f_0\rangle_{\mathcal{D}_1}.$$

Now, if $\langle f_1, \mathbf{1}\rangle = 0$ then $\langle \mathsf{D}f_1, \mathbf{1}\rangle = 0$, so

$$\langle f_1, Mf_1\rangle = \langle f_1, \mathsf{U}M_0\mathsf{D}f_1\rangle = \langle \mathsf{D}f_1, M_0\mathsf{D}f_1\rangle \leqslant \lambda\langle \mathsf{D}f_1, \mathsf{D}f_1\rangle \leqslant \lambda\langle f_1, f_1\rangle.$$

□

The following lemma is based on Claims 4.10 and 4.11.

Lemma 4.13. *Fix $\gamma = \frac{\kappa_0}{8+\kappa_0}$. Let $M = \mathsf{U}M_0\mathsf{D}$ and let $f = \mathbb{1}_R : X(1) \to \mathbb{R}$ be the indiator function of the edge set R. Then*

$$\langle f, (\gamma M^{||} + (1-\gamma)M)f\rangle_{\mathcal{D}_1} \geqslant \gamma \cdot \min(\delta_A, \delta_B) \cdot \langle f, f\rangle_{\mathcal{D}_1}.$$

Proof. We give a combinatorial interpretation to $\gamma M^{||} + (1-\gamma)M$ by observing that for a fixed $e \in X(1)$, $(\gamma M^{||} + (1-\gamma)M)f(e)$ is the probability that $e' \in R$ in the following random process.

1. Start from an edge $e \in X(1)$.
2. With probability γ, output a uniformly random edge $e' \in E^{||}(e)$ and halt. With probability $1-\gamma$ continue.
3. Choose at random one of the endpoints of the edge, $g_1 \in e$.
4. With probability $\frac{1}{2}$ let $g_2 = a_1g_1$ for a random $a_1 \in A$, and with probability $\frac{1}{2}$ let $g_2 = g_1b_1$ for a random $b_1 \in B$.
5. With probability $\frac{1}{2}$ let $e' = \{g_2, a_2g_2\}$ for a random $a_2 \in A$, and with probability $\frac{1}{2}$ let $g_2 = g_2b_2$ for a random $b_2 \in B$. Output e'.

We will prove the lemma by showing that for every $e \in R$,

$$(\gamma M^{||} + (1-\gamma)M)f(e) \geqslant \gamma \cdot \min(\delta_A, \delta_B). \tag{4.10}$$

So fix some $e \in R$, and for convenience assume $e = \{g, ag\}$ for some $g \in G, a \in A$ (if $e = \{g, gb\}$ the argument is symmetric). Let

$$r_0 = |R \cap E^{||}(e)|, \quad r_1 = |R \cap E^B(g)|, \quad r_2 = |R \cap E^B(ag)|.$$

By Claim 4.10, $r_0 + r_1 + r_2 \geqslant \delta_B|B|$. With probability γ step 2 outputs a random $e' \in E^{||}(e)$, and the probability it is in R is $r_0/|B|$.

$$\mathbb{P}[e' \in R] = \gamma \cdot r_0/|B| + (1-\gamma) \cdot \mathbb{P}[e' \in R \,|\, \text{the process entered step 3}] \tag{4.11}$$

Assume we entered step 3. Due to Claim 4.11,

$$\mathop{\mathbb{P}}_{a,b}[\{ag_1, ag_1b\} \in R \text{ or } \{g_1b, ag_1b\} \in R] \geqslant \kappa_0^{-1} \cdot r_i/|B| \tag{4.12}$$

where $i \in \{1,2\}$ depending on whether $g_1 = g$ or $g_1 = ag$ as chosen in step 3. What is the probability that e' is one of the edges $\{ag_1, ag_1b\}$ and $\{g_1b, ag_1b\}$ considered in the LHS of (4.12)? This happens exactly if in steps 4 and 5 we will walk in alternating colors (A, B or B, A). Let E_{AB} be the event that in step 4 we choose an A-edge, i.e. $g_2 = a_1g_1$ for some $a_1 \in A$ and then in step 5 we set e' to be a B-edge, i.e. $e' = \{a_1g_1, a_1g_1b_2\}$ for some $b_2 \in B$. Similarly let E_{BA} be the event that $g_2 = g_1b_1$ and $e' = \{g_1b_1, a_2g_1b_1\}$. Clearly

$$\mathbb{P}[\mathsf{E}_{AB}] = \mathbb{P}[\mathsf{E}_{BA}] = \frac{1}{4}.$$

Now,

$$\mathbb{P}[e' \in R \text{ and } \mathsf{E}_{AB}] = \frac{1}{4} \cdot \mathbb{P}_{a_1,b_2}[\{a_1g_1, a_1g_1b_2\} \in R], \tag{4.13}$$

and

$$\mathbb{P}[e' \in R \text{ and } \mathsf{E}_{BA}] = \frac{1}{4} \cdot \mathbb{P}_{a_2,b_1}[\{g_1b_1, a_2g_1b_1\} \in R]. \tag{4.14}$$

where the probability is taken over the randomness of the random process above conditioned on having entered step 3. Since E_{AB} and E_{BA} are disjoint events,

$$\begin{aligned}
\mathbb{P}[e' \in R] &\geqslant \mathbb{P}[e' \in R \text{ and } \mathsf{E}_{AB}] + \mathbb{P}[e' \in R \text{ and } \mathsf{E}_{BA}] \\
&\geqslant \frac{1}{4} \cdot (\mathbb{P}_{a,b}[\{ag_1, ag_1b\} \in R] + \mathbb{P}_{a,b}[\{g_1b, ag_1b\} \in R]) \\
&\geqslant \frac{1}{4} \cdot \mathbb{P}_{a,b}[\{ag_1, ag_1b\} \in R \text{ or } \{g_1b, ag_1b\} \in R] \\
&\geqslant \frac{1}{4}\kappa_0 \cdot r_i/|B| = \frac{r_i\kappa_0}{4|B|}
\end{aligned}$$

where in the last inequality we have used (4.12). We conclude that if in step 3 we choose $g_1 = g$, then $\mathbb{P}[e' \in R] \geqslant \frac{r_1\kappa_0}{4|B|}$, whereas if in step 3 we choose $g_1 = ag$, then $\mathbb{P}[e' \in R] \geqslant \frac{r_2\kappa_0}{4|B|}$.

Altogether, recalling (4.11),

$$\mathbb{P}[e' \in R] \geqslant \gamma \cdot \frac{r_0}{|B|} + (1-\gamma) \cdot \frac{\kappa_0}{4|B|}(r_1 + r_2)/2.$$

Plugging in $\gamma = \frac{\kappa_0}{8+\kappa_0}$ we get $1 - \gamma = 8\gamma/\kappa_0$, and recalling that $r_0 + r_1 + r_2 \geqslant \delta_B|B|$,

$$\mathbb{P}[e' \in R] \geqslant \gamma(r_0 + r_1 + r_2)/|B| \geqslant \gamma\delta_B.$$

We have seen that if $e = \{g, ag\}$ for some a, g is in R, then $e' \in R$ with probability at least $\gamma\delta_B$. Symmetrically, if $e = \{g, gb\}$ for some g, b is in R then we would get that $e' \in R$ with probability at least $\gamma\delta_A$. Together this proves (4.10) and completes the proof of the lemma. □

Recall from (4.3) that $c = \frac{\kappa_0}{8+\kappa_0} \cdot \min(\delta_A, \delta_B)$. By Lemma 4.13, $\langle f, (\gamma M^{||} + (1-\gamma)M)f\rangle \geqslant c \cdot \langle f, f\rangle$ so either

$$\langle f, Mf\rangle \geqslant c\langle f, f\rangle \tag{4.15}$$

or

$$\langle f, M^{||}f\rangle \geqslant c\langle f, f\rangle. \tag{4.16}$$

If (4.15) holds, then by Lemma 2.2, applied with the operator M whose vertex set is $X(1)$ is endowed with the distribution $\mathcal{D}_1$, we get $\mathbb{P}_{\mathcal{D}_1}[R] \geqslant c - \lambda$ which means that $|R| \geqslant \frac{|G|}{2} \cdot \min(|A|, |B|)(c - \lambda)$.

Otherwise, assume that (4.16) holds. By Lemma 3.13 there exists some $\sigma \in \tilde{A} \cup \tilde{B}$ such that, $|R \cap X^\sigma(1)| \geqslant |G|(c-\lambda)/2$.

This completes the proof of Proposition 4.8 showing that if $\mathcal{E}(\mathrm{w}) > 0$ then $\mathcal{E}(\mathrm{w}) > \frac{2(c-\lambda)}{|A|+|B|}$. □

5 A Concrete Construction

In the previous section we have described a code scheme: Given a left-right Cayley complex $Cay^2(A, G, B)$ together with two base codes $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$, we get an error-correcting code $C[G, A, B, C_A, C_B]$.

In this section we prove our main theorem by showing how to find an infinite family of left-right Cayley complexes and base codes that yield locally testable codes.

Theorem (Restatement of Theorem 1.1)**.** *For all $0 < r < 1$, there exist $\delta, \kappa > 0$, $q \in \mathbb{N}$ and an explicit construction of an infinite family of error-correcting codes $\{C_n\}_n$, such that for each n, $Rate(C_n) \geqslant r$, $\mathrm{dist}(C_n) \geqslant \delta$ and C_n is κ-locally testable with q queries.*

The proof of the theorem relies on the following two lemmas.

Lemma 5.1 (Good base code)**.** *For all $0 < r_0 < 1$, there exist $\delta_0, \kappa_0 > 0$ and $d_0, D_0 \in \mathbb{N}$, such that for every integer $D > D_0$ that is divisible by d_0, there exists a linear error correcting code $C_0 \subseteq \mathbb{F}_2^D$ with rate at least r_0, distance at least δ_0, and such that the tensor code $C_0 \otimes C_0$ is κ_0-agreement testable.*

Lemma 5.2 (Good left-right Cayley complexes)**.** *Let $d_0, D_0 \in \mathbb{N}$. Let q be any odd prime power such that $q \geqslant \max\{2d_0^2, D_0\}$ and define $D = d_0 \cdot \lfloor \frac{q+1}{d_0} \rfloor$. Then there exist an explicit construction of an infinite family of finite groups $G_i = PSL_2(q^i)$, with two symmetric generating subsets $A_i, B_i \subset G_i$, such that for each i, both A_i and B_i are of size D hence divisible by d_0, A_i and B_i satisfy condition (TNC), and the Cayley graphs $Cay(G_i, A_i)$ and $Cay(G_i, B_i)$ are λ-expanders where $\lambda \leqslant 4D^{-1/2}$.*

We prove Lemma 5.1 in Subsection 5.1 by showing that random LDPC codes are smooth. We prove Lemma 5.2 in Section 6 using the known constructions of Ramanujan graphs by Lubotzky, Samuels and Vishne [LSV05a] and Morgenstern [Mor94].

Let us now deduce Theorem 1.1 from Lemmas 5.1 and 5.2.

Proof of Theorem 1.1. Fix $0 < r < 1$ and set $r_0 = \frac{r+3}{4}$ so that $r = 4r_0 - 3$. By Lemma 5.1, given r_0, there exist $\delta_0, \kappa_0 > 0$ and $d_0, D_0 \in \mathbb{N}$, such that for any $D > D_0$ divisible by d_0, there exists a code $C_0 \subset \mathbb{F}_2^D$ with $\mathrm{Rate}(C_0) \geqslant r_0$, $\mathrm{dist}(C_0) \geqslant \delta_0$ and such that $C_0 \otimes C_0$ is κ_0-agreement-testable.

Define $q_0 = \max\{2D_0,\ 2d_0^2,\ 32\left(\frac{\kappa_0+8}{\kappa_0\delta_0}\right)^2\}$. For any $q \geqslant q_0$ odd prime power denote $D = d_0 \cdot \lfloor \frac{q+1}{d_0} \rfloor$. Note that $q+1 \geqslant D \geqslant q+1-d_0 > q - \sqrt{q} > \frac{1}{2}q$. In particular $D > \frac{1}{2}q_0$, hence $4D^{-1/2} < \sqrt{\frac{32}{q_0}} \leqslant \frac{\kappa_0\delta_0}{8+\kappa_0}$, which also implies $4D^{-1/2} < \delta_0$.

By Lemma 5.2 there exists an explicit construction of an infinite family of groups $G_i = PSL_2(q^i)$ together with generating sets A_i, B_i such that for each $i \in \mathbb{N}$, $|A_i| = |B_i| = D$, conditions (TNC) holds, and both $Cay(G_i, A_i)$ and $Cay(G_i, B_i)$ are $\lambda = 4D^{-1/2}$ expanders. In particular, from our choice of D, equation (4.3) holds and $\lambda < \delta_0$.

By Lemma 5.1 there exists a code C_0 of length D, with rate at least r_0, distance at least δ_0 and such that the tensor code $C_0 \otimes C_0$ is κ_0-agreement testable. Since D is a constant we can, theoretically, enumerate over all possible codes in search of a good one.

Define our family of global codes to be $C_i = C[G_i, A_i, B_i, C_0, C_0]$, $i \in \mathbb{N}$, and by the above choices it has the following parameters:

- Block-length $\frac{1}{4}|G_i|D^2$, where $|G_i| = \frac{1}{2}(q^{3i} - q^i)$.
- Distance at least $\delta = \delta_0^2(\delta_0 - 4D^{-1/2}) > 0$, by Lemma 4.4,
- Rate at least $r = 4r_0 - 3 > 0$, by Lemma 4.2,
- It is κ-locally-testable with D^2 queries, by Theorem 4.5, for

$$\kappa = \min \left\{ \frac{1}{4+8D} , \frac{1}{4D} \left(\frac{\delta_0 \kappa_0}{8+\kappa_0} - 4D^{-1/2} \right) \right\} . \tag{5.1}$$

□

5.1 Good Base Codes

In this section we prove Lemma 5.1 by relying on the notion of smooth codes from [DSW06], which was consequently broadened to weakly-smooth codes in [BV09]. These works showed that the tensor product of a smooth code and any other code is robustly testable and therefore, by Lemma 2.9, also agreement-testable.

Definition 5.3 (Smooth Code)**.** Let $c, d, n \in \mathbb{N}, \alpha, \beta, \delta > 0$. A (c, d, n) LDPC code $C \subset \mathbb{F}_2^n$ is (α, β, δ)-smooth if for every $Y_0 \subseteq Y$ with $|Y_0| \leqslant \alpha|Y|$ there is some $X_0 \subseteq X$ with $|X_0| \leqslant \beta|X|$ such that the code $C(\bar{Y}_0)|_{\bar{X}_0}$ has distance at least δ. Here the code $C(\bar{Y}_0)|_{\bar{X}_0}$ is the code obtained by removing the constraints in Y_0 and then removing the coordinates in X_0.

5.1.1 Random LDPC Codes

We will next show that random LDPC codes are smooth. Random LDPC codes, see Definition 2.6, were famously introduced by Gallager in his PhD thesis [Gal63].

Given a (c, d, n)-LDPC code, let $m = nc/d$. By counting constraints it is easy to see that the dimension of an LDPC code is at least $n - m = n(1 - \frac{c}{d})$, so the rate is at least $1 - \frac{c}{d}$. Spielman described in his thesis [Spi96] the following expansion property,

Definition 5.4. A (c, d)-regular bipartite graph $([n], [m], E)$ is a (δ, γ)-expander if every set of left vertices $A \subset [n]$ whose size is at most δn, has at least $c|A|(1 - \gamma)$ neighbors.

An LDPC code whose factor graph is a (δ, γ) expander immediately has distance at least δ, as long as $\gamma < \frac{1}{2}$ [Spi96].

A random (c, d, n)-code is given by selecting a random (c, d)-regular bipartite graph, which in turn is done by taking a random matching between the nc "half-edges" on the left and the md "half-edges" on the right, where we assume that nc/d is an integer.

Claim 5.5 (Claim 6.4 in [BHR05][2]). For every $c > 2$, d, any constant $\gamma > \frac{1}{c}$, and sufficiently large D such that Dc/d is an integer, a random (c, d)-regular bipartite graph with D left vertices is with high probability a (δ, γ)-expander for any δ satisfying

$$\delta \leqslant \left(2e^{1+c(1-\gamma)}(d - d\gamma)^{c\gamma} \right)^{-\frac{1}{c\gamma - 1}} . \quad \square$$

[2] A (δ, γ)-expander here is called a $(c(1 - \gamma), \delta)$-left-expander in [BHR05].

Remark 5.6. It can be extracted from the proof of Claim 6.4 in [BHR05] that the first D for which such a (c, d)-regular (δ, γ)-expander exists, denote it by D_0, is upper bounded by a function of c, d and γ. More explicitly, if $c = 4$ and $\gamma = \frac{5}{12}$, then

$$D_0 \leqslant 2^{42} d^{15}.$$

The proof is similar Gallager's proof [Gal63] that a random LDPC code has constant distance with high probability. Tensors of these codes are robustly testable,

Theorem 5.7 (Robust testability of expander codes)**.** *Let C be a (c, d, D)-code whose factor graph is a (c, d)-regular (δ, γ)-expander. Let C' be any linear code with distance δ'. Then $C \otimes C'$ is ρ-robustly testable for*

- $\rho \geqslant \frac{\delta\delta' \cdot (\frac{1}{6} - \gamma)}{2d}$ *when $\gamma < 1/6$ [DSW06], and*
- $\rho \geqslant \frac{\delta\delta' \cdot}{d^{\log_{0.5+\gamma} 0.05}}$ *for all $\gamma < 1/2$ [BV09].*

Finally, we can prove Lemma 5.1, which we restate for convenience,

Lemma (Restatement of Lemma 5.1)**.** *For all $0 < r_0 < 1$, there exist $\delta_0, \kappa_0 > 0$ and $d_0, D_0 \in \mathbb{N}$, such that for every integer $D > D_0$ that is divisible by d_0, there exists a linear error correcting code $C_0 \subseteq \mathbb{F}_2^D$ with rate at least r_0, distance at least δ_0, and such that the tensor code $C_0 \otimes C_0$ is κ_0-agreement testable.*

Proof. We fix $\gamma = 0.15 < 1/6$ and set $c_0 = 7$ so that $\gamma > 1/c_0$. We choose $d_0 = \lceil \frac{7}{1-r_0} \rceil$ such that $\frac{c_0}{d_0} \leqslant 1 - r_0$. Claim 5.5 guarantees existence of $\delta_0 > 0$ and D_0 such that for all $D > D_0$ divisible by d_0, a random (c_0, d_0)-regular bipartite graph with D left vertices is with high probability a (δ_0, γ)-expander.

For each such bipartite graph, we take C_0 to be the corresponding (c_0, d_0, D)-LDPC code. This code has rate at least r_0, distance at least δ_0, and by taking $C' = C_0$ in Theorem 5.7, we get that $C_0 \otimes C_0$ is robustly testable with $\rho = \Omega(\delta_0^2/d_0) = \Omega(\delta_0^2(1 - r_0))$. By Claim A.1 these codes are κ_0-agreement-testable for $\kappa_0 = \Omega(\delta_0^3(1 - r_0))$. □

We remark that the divisibility condition on D is not really necessary. For D not divisible by d_0 one can redistribute at most d_0 extra edges so that the graph is slightly irregular. The resulting graph is still a (δ, γ)-expander, and one can also prove smoothness, mutatis mutandis, with a negligible change to the parameters.

6 Good Left-Right Cayley Complexes

In the previous section we showed how to construct good locally testable codes on good left-right Cayley complexes provided the latter exists. To finish the proof of the main result of the paper, we should show that such complexes indeed exist and to give explicit construction. Namely, in this section we prove Lemma 5.2.

More generally, we show that for every $\lambda > 0$, there exist $k_1, k_2 \in \mathbb{N}$ and an infinite family of finite groups G_i, with two symmetric subsets of generators A_i, B_i, such that for each i, $|A_i| = k_1$ and $|B_i| = k_2$, the two sets A_i and B_i satisfies (TNC), and the second largest eigenvalues of the adjecancy matrices of $\mathrm{Cay}(G_i, A_i)$ and $\mathrm{Cay}(G_i, B_i)$, denoted $\lambda(\mathrm{Cay}(G_i, A_i))$ and $\lambda(\mathrm{Cay}(G_i, B_i))$, are bounded from above by λ. Moreover, we can take $\lambda = \Theta(k_1^{-1/2}) = \Theta(k_2^{-1/2})$, making both Cayley graphs quasi-Ramanujan.

There are a number of ways in the literature to find Cayley graphs with small $\lambda(\mathrm{Cay}(G, S))$. There are even various methods to give different sets of generators for the same group (see [Lub94], [LSV05a]). The difficulty is to ensure that condition (TNC) is satisfied. We will show two (actually three) ways to do so. In all of our constructions, the elements in the sets B_i will be of order 2, while all the elements in A_i will be of order greater then 2. This ensures that (TNC) is automatically satisfied.

6.1 The Morgenstern Generators, $q = 2^\ell$

In [Mor94], Morgenstern presented for every prime power q, infinitely many groups $G_i = PGL_2(q^i)$ or $G_i = PSL_2(q^i)$ each with a symmetric set B_i of $q+1$ generators such that $\mathrm{Cay}(G_i, B_i)$ are Ramanujan, i.e., $\lambda(\mathrm{Cay}(G_i, B_i)) \leqslant \frac{2\sqrt{q}}{q+1}$.

The case of q even, i.e., $q = 2^\ell$, is special in two ways. First of all, here $PGL_2(q^i) = PSL_2(q^i)$, so this is always a simple group. But more importantly, in this case all the elements of B_i are of order 2 (see Remark 6.3). Assume q is even from now on.

Morgenstern constructed an explicit arithmetic lattice Γ in the group $PSL_2(\mathbb{F}_q((t)))$ which is isomorphic to the free product $\langle b_0\rangle * \ldots * \langle b_q\rangle$, where $B = \{b_0, \ldots, b_q\}$ is a set of elements of order 2 (see [Mor94, Section 5]). The above mentioned Cayley graphs $\mathrm{Cay}(G_i, B_i)$ are identified as quotients of this Γ by normal congruence subgroups, where $B_i = \phi_i(B)$ is the image of B under an epimorphism $\phi_i : \Gamma \to G_i$. Note that by [Mor94] these Cayley graphs are all Ramanujan.

Let us now show how to get another symmetric set of generators A_i for $G_i = PSL_2(q^i)$ with $\lambda(\mathrm{Cay}(G_i, A_i))$ small, and such that A_i and B_i satisfy (TNC).

Let Λ be the index 2 subgroup of Γ - the kernel of the homomorphism $\phi : \Gamma \to C_2$ (= the cyclic group of order 2) where ϕ sends each b_j to the unique non-trivial element of C_2. One can see easily that Λ is exactly the subgroup of all elements of Γ of even length w.r.t. B. It is generated by the set $A = \{b_t b_s \mid b_t, b_s \in B,\ t \neq s\}$ which is of size $k_1 = q^2 + q$. We claim

Claim 6.1. (i) For $i > 1$, the image $A_i = \phi_i(A)$ of A in G_i generates $G_i = PSL_2(q^i)$.

(ii) $\lambda(\mathrm{Cay}(G_i, A_i)) < \frac{3q-1}{q^2+q} < \frac{3\sqrt{k_1 - 1}}{k_1}$.

(iii) For $i > 4$, the images of the elements of A in G_i are distinct from one another, and each element in A_i has order > 2.

Proof. (i) Since $\Lambda = \langle A\rangle$ is of index two in Γ then $\langle A_i\rangle$ is of index at most two in G_i. But $G_i = PSL_2(q^i)$ is simple, hence it has no index 2 subgroup (a subgroup of index 2 must be normal), which implies $\langle A_i\rangle = G_i$.

(ii) Let T_B and T_A be the (non-normalized) adjacency matrices of $\mathrm{Cay}(G_i, B_i)$ and $\mathrm{Cay}(G_i, A_i)$, respectively. Note that $T_B^2 = T_A + (q+1)I$. Hence if μ is an eigenvalue of T_A, then $\mu = \lambda^2 - (q+1)$ for some eigenvalue λ of T_B. Since $\mathrm{Cay}(G_i, B_i)$ is Ramanujan, $|\lambda| = q+1$ or $|\lambda| \leqslant 2\sqrt{q}$. Therefore $\mu = q^2 + q$ or $\mu \leqslant (2\sqrt{q})^2 - (q+1) = 3q - 1$.

(iii) It suffices to show that each reduced word which is a product of length at most 4 in B is not in the kernel of ϕ_i, which is equivalent to the girth of $\mathrm{Cay}(G_i, B_i)$ being greater than 4. By [Mor94, Theorem 5.13 (3)] the girth of $\mathrm{Cay}(G_i, B_i)$ is at least $\frac{2}{3}\log_q |G_i| \geqslant i$, which completes the proof. □

Thus, given $\lambda > 0$ by taking q large enough so that $\frac{3\sqrt{q^2+q-1}}{q^2+q} < \lambda$, we get the desired λ-expanding left-right Cayley complexes with $k_1 = q^2 + q$ and $k_2 = q + 1$.

We can do slightly better. Note that Λ above, being a normal subgroup of a free product of finite groups, with trivial intersection with each factor is a free group (see Section 34 in [Kur55]). In fact, by the Reidemeister-Schreier algorithm applied to the transversal set $\{1, b_0\}$ of Λ in Γ (or by inspection) one can see that Λ is a free group on the q generators $\{b_0 b_j \ : \ j = 1, \ldots, q\}$. As $(b_0 b_j)^{-1} = b_j b_0$ we deduce that $A' = \{b_0 b_j,\, b_j b_0 \ : \ j = 1, \ldots, q\}$ is a symmetric set of generators of Λ.

We can now look at the image $A'_i = \phi_i(A')$ under the epimorphism $\phi_i : \Gamma \to G_i$. Arguing similarly to the proof of Claim 6.1 (i), A'_i generates G_i, and by the proof of (iii) above, the images are all different. Finally:

Claim 6.2. $\lambda(\mathrm{Cay}(G_i, A'_i)) < \frac{3\sqrt{2q-1}}{2q}$.

Proof. Let $V_i = \{f \ : \ G_i \to \mathbb{C}\}$ and for any element $s \in G_i$, define the s-adjecancy $T_s : V_i \to V_i$, $T_s f(g) = f(gs)$, and for any multiset S of G_i, define the S-adjecancy operator $T_S : V_i \to V_i$, $T_S = \sum_{s \in S} T_s$. Note that for any two multisets S, S' of G_i, $T_{S \cup S'} = T_S + T_{S'} - T_{S \cap S'}$ and $T_S T_{S'} = T_{SS'}$, where $SS' = \{ss' \ : \ s \in S, s' \in S'\}$ counted with multiplicities. Therefore $T_{A'_i} = T_b T_{B_i} + T_{B_i} T_b - 2I$, where $b = \phi_i(b_0)$. Let $f \in V_i$ be such that $f \perp 1_{G_i}$, i.e. $\sum_{g \in G_i} f(g) = 0$. Note that for any $s \in G_i$, then $T_s f \perp 1_{G_i}$ and $\|T_s f\| = \|f\|$. By [Mor94, Theorem 5.11], we have $\|T_{B_i} f\| \leqslant 2\sqrt{q}\|f\|$ for any $f \perp 1_{G_i}$. Then

$$\|T_{A'_i} f\| \leqslant \|T_b T_{B_i} f\| + \|T_{B_i} T_b f\| + 2\|f\| \leqslant \|T_{B_i} f\| + \|T_{B_i}(T_b f)\| + 2\|f\|$$

$$\leqslant (2\sqrt{q} - 1)\,\|f\| + (2\sqrt{q} - 1)\,\|f\| + 2\|f\| = 4\sqrt{q}\|f\| \leqslant 3\sqrt{2q-1}\|f\|$$

which completes the proof. □

So this time we have a family of λ-expanders left-right Cayley complexes with $k_1 = 2q$ and $k_2 = q+1$, for any $\lambda > \frac{3\sqrt{2q-1}}{2q}$.

Remark 6.3. Everything said above is explicit. In fact the generator set B_i of $PSL_2(q^i)$ are given explicitly in [Mor94, equation (21)]. Assume i is even. Let $\mathbf{i} \in \mathbb{F}_{q^i}$ be such that $\mathbf{i} \notin \mathbb{F}_q$ and $\varepsilon = \mathbf{i}^2 + \mathbf{i} \in \mathbb{F}_q$. Let $x \in \mathbb{F}_{q^i}$ be such that $1, x, \ldots, x^{e_i - 1}$ form a basis for $\mathbb{F}_{q^i}$ over $\mathbb{F}_q$. Then the $q+1$ elements of B_i are

$$\phi_i(b_j) = \begin{pmatrix} 1 & \gamma_j + \delta_j \mathbf{i} \\ x(\gamma_j + \delta_j + \delta_j \mathbf{i}) & 1 \end{pmatrix}, \qquad j = 0, \ldots, q, \tag{6.1}$$

where $(\gamma_j, \delta_j) \in \mathbb{F}_q^2$ are the $q+1$ solutions in $\mathbb{F}_q$ for $\gamma^2 + \gamma\delta + \delta^2\varepsilon = 1$. One indeed sees that each of the elements of B_i is of order 2.

We will pass now to a different construction, which will give us Cayley graphs of G_i w.r.t. A_i and B_i of the same size: $|A_i| = |B_i| = q+1$, and both are Ramanujan.

6.2 The LSV Generators, q odd

In [LSV05a], Lubotzky, Samuels and Vishne constructed Ramanujan complexes, based on an arithmetic lattice Γ, discovered by Cartwright and Steger [CS98], which acts simply transitively on the Bruhat-Tits building of $PGL_d(\mathbb{F}_q((t)))$. The special case $d = 2$ gave some new Ramanujan graphs. These Ramanujan graphs were highlighted in [KL12], as edge-transitive Ramanujan graphs which have been used there to construct symmetric LDPC codes.

The arithmetic group Γ, acting simply transitive on the Bruhat-Tits tree of $PGL_2(\mathbb{F}_q((t)))$ (q any odd prime power) is obtained there as a the group generated by the $q+1$ conjugates of a specific element b, conjugated by the non-split torus T of order $q+1$ in $PGL_2(\mathbb{F}_q)$. This is a symmetric set of generators A for Γ which generates a free group on $\frac{q+1}{2}$ generators. We will present below a different choice for b, this time b' - an element of order 2, whose conjugation under T forms a symmetric set B of size

$q+1$ and generate a group Γ' which also acts simply transitive on the Bruhat-Tits tree. Moreover, Γ and Γ' are both finite index subgroups of an arithmetic group $G(R)$ - to be defined below.

In [LSV05a] (see also [KL12]) it was shown that $G(R)$ has infinitely many finite congruence quotients G_i, under the maps $\phi_i : G(R) \to G_i$, for which $\mathrm{Cay}(G_i, \phi_i(A))$ are Ramanujan $(q+1)$-regular graphs. We will observe below that the same holds for $\mathrm{Cay}(G_i, \phi_i(B))$. For i large enough (see Claim 6.5) the elements of $\phi_i(A)$ are of order > 2 while $\phi_i(B)$ contains only elements of order 2. Hence we will get two-sided Cayley square complexes with $k_1 = k_2 = q+1$ and $\lambda \leqslant \frac{2\sqrt{q}}{q+1}$. By choosing q large enough, they will be λ-expanders for arbitrarily small $\lambda > 0$.

Now, in more details: Let $0 \neq \varepsilon \in \mathbb{F}_q$ be a non-square element, let $R = \mathbb{F}_q[y, \frac{1}{y}, \frac{1}{1+y}]$ be the subring of $\mathbb{F}_q(y)$, generated by y, $\frac{1}{y}$ and $\frac{1}{1+y}$, and let $A(R)$ be the quaternion R-algebra,

$$A(R) = R + R\alpha + Rz + R\alpha z \qquad : \qquad \alpha^2 = \varepsilon, \quad z^2 = 1+y, \quad z\alpha = -\alpha z. \tag{6.2}$$

Remark 6.4. We note that our choice of basis for $A(R)$, $\{1, \alpha, z, \alpha z\}$, is based on [KL12], while [LSV05a] used a different basis for $A(R)$, $\{\xi, \xi^q, \xi z, \xi^q z\}$, where $\{\xi, \xi^q\}$ forms an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^2} = \mathbb{F}_q[\alpha]$. The change of bases does not affect any of the following constructions.

For any ring D, denote by D^* its group of units. Note that an element of $r(y) \in R$ belongs to R^* if and only if it is of the form $r(y) = cy^n(1+y)^m$, $c \in \mathbb{F}_q^*$, $n, m \in \mathbb{Z}$, and that an element $a = a_1 + a_2\alpha + a_3 z + a_4 \alpha z \in A(R)$ belongs to $A(R)^*$ if and only if its norm $N(a) := a_1^2 - \varepsilon a_2^2 - (1+y)a_3^2 - \varepsilon(1+y)a_4^2 \in R$ belongs to R^*. Note also that R is the center of $A(R)$ and R^* is the center of $A(R)^*$. Then the principal arithmetic group $G(R)$ is defined to be

$$G(R) = A(R)^*/R^* = \{a \in A(R) \,:\, N(a) \in R^*\}/R^*.$$

The Cartwright–Steger arithmetic lattice Γ, and the second arithmetic lattice Γ', are defined to be the subgroups of $G(R)$, generated by the symmetric sets of size $q+1$, A and B, which are the sets of T conjugates of the elemenets, b and b', respectively, where $T = \mathbb{F}_q[\alpha]^*/\mathbb{F}_q^* \leqslant G(R)$ is a non-split torus of order $q+1$, $b = \left(1 - \frac{1}{1+y}z\right)R^* \in G(R)$ and $b' = \alpha b = \left(\alpha - \frac{1}{1+y}\alpha z\right)R^* \in G(R)$, namely,

$$\Gamma = \langle A \rangle \leqslant G(R), \quad A = \left\{tbt^{-1} \,:\, t \in T\right\}, \quad \Gamma' = \langle B \rangle \leqslant G(R), \quad B = \left\{tb't^{-1} \,:\, t \in T\right\},$$

$$T = \mathbb{F}_q[\alpha]^*/\mathbb{F}_q^*, \qquad b = \left(1 - \frac{1}{1+y}z\right)R^*, \qquad b' = \alpha b = \left(\alpha - \frac{1}{1+y}\alpha z\right)R^*.$$

Note that b and b' belongs to $G(R)$, since $N\left(1 - \frac{1}{1+y}z\right) = 1 - (1+y)\frac{1}{(1+y)^2} = \frac{y}{1+y} \in R^*$ and $N\left(\alpha - \frac{1}{1+y}\alpha z\right) = N(\alpha) \cdot N\left(1 - \frac{1}{1+y}z\right) = -\varepsilon \cdot \frac{y}{1+y} \in R^*$.

Claim 6.5. (i) Every element of A is of infinite order, while every element of B is of order 2.

(ii) For $i > 2$, every element of $A_i = \phi_i(A)$ is of order > 2, while every element of $B_i = \phi_i(B)$ is of order 2.

Proof. (i) The claim about the elements of A follows from [LSV05a, Corollary 5.4]. For the claim about the elements of B, since they are all conjugate of one another, it suffice to show $b'^2 = 1$, or equivalently, $\left(\alpha - \frac{1}{1+y}\alpha z\right)^2 \in R^*$. This follows from the following computations,

$$\left(\alpha - \frac{1}{1+y}\alpha z\right)^2 = \alpha^2 - \frac{1}{1+y}\alpha\alpha z - \frac{1}{1+y}\alpha z\alpha + \frac{1}{(1+y)^2}\alpha z\alpha z = *,$$

and by equation 6.2, as $\alpha z = -z\alpha$, $\alpha^2 = \varepsilon$ and $z^2 = 1+y$, we get

$$* = \alpha^2 - \frac{1}{(1+y)^2}\alpha^2 z^2 = \varepsilon - \frac{\varepsilon}{1+y} = \varepsilon\frac{y}{1+y} \in R^*.$$

(ii) This follows from an injectivity radius argument for congruence subgroups, see for instance [LM07]. □

Let $\mathcal{B}$ be the Bruhat-Tits tree of $PGL_2(\mathbb{F}_q((t)))$, which is a $(q+1)$-regular infinite tree. By [LSV05a, Section 3], Γ, Γ' and $G(R)$ are subgroups of $PGL_2(\mathbb{F}_q((t)))$, hence acts on $\mathcal{B}$. In the notation of [LSV05a], let $v_0 = [L_0]$ be the fundamental vertex in $\mathcal{B}$, and let Ω be the set of its neighbors.

Claim 6.6. (i) For each set $X = A$ or $X = B$, the map $g \leftrightarrow g.v_0$ is a bijection between X and Ω.

(ii) The subgroups, Γ and Γ', acts simply transitively on the Bruhat-Tits tree.

(iii) Both subgroups, Γ and Γ', are normal in $G(R)$ and of index $2(q+1)$.

(iv) If $\phi : G(R) \to PSL_2(q^e)$ is an epimorphism, then both subsets, $\phi(A)$ and $\phi(B)$, are symmetric set of generators for $PSL_2(q^e)$.

(v) If $\phi : G(R) \to PSL_2(q^e)$ is an epimorphism whose kernel is a congruence subgroup $G(R,\phi)$ of $G(R)$, then both Cayley graphs, $\mathrm{Cay}(PSL_2(q^e), \phi(A))$ and $\mathrm{Cay}(PSL_2(q^e), \phi(B))$, are Ramanujan $(q+1)$-regular graphs.

Proof. (i) The claim for A is [LSV05a, Proposition 4.3]. The claim for B follows from the claim for A and the identity $tb't^{-1}.v_0 = t\alpha bt^{-1}.v_0 = (t\alpha)b(t\alpha)^{-1}(t\alpha t^{-1}).v_0$. Now, $\alpha \in T$ and T fixes v_0, so $\{tb't^{-1}.v_0 | t \in T\} = \{tbt^{-1}.v_0 | t \in T\}$.

(ii) The transitivity claim for Γ is [LSV05a, Proposition 4.5], which relies solely on the validity of claim (i) for the generating set A of Γ, hence the same proof works also for Γ'. Moreover, the same proof can actually show that for any $n \in \mathbb{N}$, for any vertex v of distance n from v_0, there exists a reduced word $g = s_1 \cdots s_n \in \Gamma$ (resp. Γ'), $s_1, \ldots, s_n \in A$ (resp. B), such that $g.v_0 = v$. This proves that the action is also simple since the number of vertices of distance n is equal the number of reduced words of length n, for any $n \in \mathbb{N}$.

(iii) The claim for Γ follows from [LSV05a, Propositions 4.9 and 3.5], and the same proof also works for Γ'. The fact that the index is $2(q+1)$ follows also from the fact that Γ' acts simply transitively on the Bruhat-Tits tree by (ii). Hence the index of Γ' in $G(R)$ is equal to the order of the stabilizer of v_0 in $G(R)$, which by [LSV05a, Proposition 3.5], is of size $2(q+1)$.

(iv) By (iii) both images, $\phi(\Gamma)$ and $\phi(\Gamma')$, are normal subgroups of index $\leqslant 2(q+1)$ in $PSL_2(q^e)$, and since $PSL_2(q^e)$ is a simple group of size $\geqslant \frac{1}{2}(q+1)q(q-1) > 2(q+1)$, we get that $\phi(\Gamma) = PSL_2(q^e) = \phi(\Gamma')$.

(v) The claim for $\mathrm{Cay}(PSL_2(q^e), \phi(A))$ is [LSV05a, Theorem 7.1], and the same proof holds also for $\mathrm{Cay}(PSL_2(q^e), \phi(B))$. Another way to see this is to observe that both graphs are isomorphic to $G(R,\phi)\backslash\mathcal{B}$ and in particular they are isomorphic, so if one is Ramanujan so is the other. □

6.3 Proof of Lemma 5.2 and Degree Reduction

First we use the LSV generators constructed in the previous subsection to prove the following Lemma.

Claim 6.7. For any odd prime power q there exist an explicit construction of an infinite family of finite groups $G_i = PSL_2(q^i)$, with two symmetric generating subsets A_i, B_i of G_i, such that for each i, $|A_i| = |B_i| = q+1$, condition (TNC) holds for A_i and B_i, and the Cayley graphs $\text{Cay}(G_i, A_i)$ and $\text{Cay}(G_i, B_i)$ are Ramanujan, in particular they are λ-expanders with $\lambda \leqslant 2(q+1)^{-1/2}$.

Proof. From Claim 6.6 we get that for any i, there exists two symmetric generating subsets A_i and B_i of the finite group $G_i = PSL_2(q^i)$, both sets are of size $q+1$, and the Cayley graphs $\text{Cay}(G_i, A_i)$ and $\text{Cay}(G_i, B_i)$ are both Ramanujan. By Claim 6.5, for any $i > 2$, the two sets A_i and B_i satisfy condition (TNC). □

Next we prove the following degree reduction trick, which allows us to start with a λ-expander Cayley graph, and to remove a few elements from the generating set with only negligible effect on λ.

Claim 6.8. (i) Let G be a finite group, let $S' \subset S$ be two symmetric subset of G, and denote $\lambda = \lambda(\text{Cay}(G,S))$ and $\lambda' = \lambda(\text{Cay}(G,S'))$ the normalized second largest eigenvalues of the corresponding Cayley graphs. Then

$$\lambda' \leqslant \lambda + \frac{|S \setminus S'|}{|S'|}.$$

(ii) In particular, if $c \leqslant \lambda \cdot |S|^{1/2} \leqslant C$, where $0 < c < C$, and $|S \setminus S'| \leqslant \frac{1}{2} \cdot c \cdot |S|^{1/2}$, then

$$\lambda' \leqslant 2\lambda \leqslant 2C \cdot |S'|^{1/2}.$$

Proof. (i) Let $M = M_A$ and $M' = M_{A'}$ be the adjacency matrices of $\text{Cay}(G,S)$ and $\text{Cay}(G,S')$, respectively. Since S (resp. S') generates G, the largest eigenvalue of M (resp. M'), which is $|S|$ (resp. $|S'|$), has multiplicity one, and its eigenvector is the constant function 1_G. By the Courant-Fischer Formula we get that

$$\lambda \cdot |S| = \max_{0 \neq v \perp 1_G} \frac{v^t M v}{v^t v} \qquad \text{and} \qquad \lambda' \cdot |S'| = \max_{0 \neq v \perp 1_G} \frac{v^t M' v}{v^t v}.$$

Now the matrix $M - M'$ can be considered as the adjacency matrix of the set $S \setminus S'$, which by the Perron-Frobenius Theorem, all of its eigenvalues are bounded in absolute value by $|S \setminus S'|$, and by the Courant-Fischer Formula $|S \setminus S'| = \max_{0 \neq v} \frac{v^t (M' - M) v}{v^t v}$. Therefore we get that

$$\lambda' \cdot |S'| = \max_{0 \neq v \perp 1_G} \frac{v^t M' v}{v^t v} \leqslant \max_{0 \neq v \perp 1_G} \frac{v^t M v}{v^t v} + \max_{0 \neq v \perp 1_G} \frac{v^t (M' - M) v}{v^t v} \leqslant \lambda \cdot |S| + |S \setminus S'|,$$

and after dividing by $|S'|$ we get the claim.

(ii) follows from (i) together with the fact that

$$|S \setminus S'| \leqslant \frac{1}{2} \cdot c \cdot |S|^{1/2} \leqslant \frac{\lambda |S|}{1+\lambda} \quad \Rightarrow \quad \frac{|S \setminus S'|}{|S|} \leqslant \lambda.$$

□

Finally we combine the above two Claims to prove Lemma 5.2.

Lemma (Restatement of Lemma 5.2)**.** *Let $d_0, D_0 \in \mathbb{N}$. Let q be any odd prime power such that $q \geqslant \max\{2d_0^2, D_0\}$ and define $D = d_0 \cdot \lfloor \frac{q+1}{d_0} \rfloor$. Then there exist an explicit construction of an infinite family of finite groups $G_i = PSL_2(q^i)$, with two symmetric generating subsets $A_i, B_i \subset G_i$, such that for each i, both A_i and B_i are of size D hence divisible by d_0, A_i and B_i satisfy condition* (TNC)*, and the Cayley graphs $Cay(G_i, A_i)$ and $Cay(G_i, B_i)$ are λ-expanders where $\lambda \leqslant 4D^{-1/2}$.*

Proof of Lemma 5.2. First note that D is by definition the largest integer $\leqslant q + 1$ which is divisible by d_0, and that $q + 1 - D \leqslant d_0 \leqslant \frac{1}{2}\sqrt{D}$.

By Claim 6.7, for each i, there exist $\tilde{A}_i, \tilde{B}_i$ two symmetric generating subset of $G_i = PSL_2(q^i)$, such that $\tilde{A}_i, \tilde{B}_i$ are both of size $q + 1$, they satisfy (TNC) and such that the corresponding Cayley graphs are Ramanujan, i.e. λ-expandrs for $\lambda \leqslant \frac{2\sqrt{q}}{q+1} \leqslant 2(q+1)^{-1/2}$.

Let $A_i \subset \tilde{A}_i$ and $B_i \subset \tilde{B}_i$ be any two symmetric subsets of size D. Since $\tilde{A}_i$ and $\tilde{B}_i$ satisfy (TNC), any subsets of them must also satisfy (TNC).

By Claim 6.8, we get that for $G = G_i$, $S = \tilde{A}_i$ or $\tilde{B}_i$, and $S' = A_i$ or B_i, respectively, we get that

$$\lambda(\mathrm{Cay}(G, S')) \leqslant 2\lambda(\mathrm{Cay}(G, S)) \leqslant 4D^{-1/2},$$

which completes the proof of the Lemma. □

References

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988. 7

[ALM$^+$98] 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. 2

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. 5

[Aro94] S. Arora. *Probabilistic checking of proofs and the hardness of approximation problems*. PhD thesis, U.C. Berkeley, 1994. Available via anonymous ftp as Princeton TR94-476. 2

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. 2

[BFL91] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991. 2

[BFLS91] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd ACM Symp. on Theory of Computing*, pages 21–31, 1991. 2

[BGH$^+$06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. Robust PCPs of proximity, shorter PCPs and applications to coding. *SIAM Journal on Computing*, 36(4):889–974, 2006. In special issue on Randomness and Computation. 2, 4

[BGK$^+$10] Eli Ben-Sasson, Venkatesan Guruswami, Tali Kaufman, Madhu Sudan, and Michael Viderman. Locally testable codes require redundant testers. *SIAM J. Comput.*, 39(7):3230–3247, 2010. 4

[BHR05] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3CNF properties are hard to test. *SIAM J. Comput.*, 35(1):1–21, 2005. 1, 4, 5, 23, 24

[BLR90] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proc. 22nd ACM Symp. on Theory of Computing*, pages 73–83, 1990. 1, 4

[BS05] Eli Ben-Sasson and Madhu Sudan. Simple PCPs with poly-log rate and query complexity. In *Proc. 37th ACM Symp. on Theory of Computing*, pages 266–275, 2005. 2

[BS06] Eli Ben-Sasson and Madhu Sudan. Robust locally testable codes and products of codes. *Random Structures & Algorithms*, 28(4):387–402, 2006. 3, 8

[BS08] Eli Ben-Sasson and Madhu Sudan. Short PCPs with polylog query complexity. *SIAM J. Comput.*, 38(2):551–607, 2008. 4

[BSS05] László Babai, Amir Shpilka, and Daniel Stefankovic. Locally testable cyclic codes. *IEEE Trans. Inf. Theory*, 51(8):2849–2858, 2005. 4

[BSV12] Eli Ben-Sasson and Michael Viderman. Towards lower bounds on locally testable codes via density arguments. *computational complexity*, 21(2):267–309, 2012. 4

[BSVW03] Eli Ben-Sasson, Madhu Sudan, Salil P. Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proc. 35th ACM Symp. on Theory of Computing*, pages 612–621, 2003. 2, 4

[BV09] Eli Ben-Sasson and Michael Viderman. Tensor products of weakly smooth codes are robust. *Theory of Computing*, 5(12):239–255, 2009. 3, 9, 23, 24

[CS98] D.I. Cartwright and T. Steger. A family of $\tilde{A}_n$-groups. *Israel J. Math.*, 103(1):125–140, 1998. 26

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, volume 116 of *LIPIcs*, pages 38:1–38:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. 6

[DH09] Irit Dinur and Prahladh Harsha. Composition of low-error 2-query PCPs using decodable PCPs. In *Proc. 50th IEEE Symp. on Foundations of Computer Science*, 2009. 9

[Din07] Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3), 2007. 2, 4

[Din21a] Irit Dinur. Breakthroughs - locally testable codes with constant rate, distance, and locality. https://simons.berkeley.edu/events/breakthroughs-locally-testable-codes-constant-rate 2021. 7

[Din21b] Irit Dinur. Locally testable codes with constant rate, distance, and locality. Part I: https://youtu.be/pz2-bEopa-c, Part II: https://youtu.be/Ydb2OPQ7eqI, 2021. 7

[DK11] Irit Dinur and Tali Kaufman. Dense locally testable codes cannot have constant rate and distance. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, August 17-19, 2011. Proceedings*, pages 507–518, 2011. 4

[DK17] Irit Dinur and Tali Kaufman. Agreement expansion. Work in progress, 2017. 5

[DSW06] Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust local testability of tensor products of LDPC codes. In *Proc. 10th International Workshop on Randomization and Computation (RANDOM)*, 2006. 3, 9, 23, 24

[EK16] Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 36–48, 2016. 5

[FS13] Katalin Friedl and Madhu Sudan. Some improvements to total degree tests. *CoRR*, abs/1307.3975, 2013. 2

[Gal63] R.G̃. Gallager. *Low density parity check codes.* MIT Press, Cambridge, Massachusetts, 1963. 23, 24

[Gar73] H. Garland. p-adic curvature and the cohomology of discrete subgroups of p-adic groups. *Annals of Mathematics*, 97:375, 1973. 5

[GKdO$^+$18] Sivakanth Gopi, Swastik Kopparty, Rafael Mendes de Oliveira, Noga Ron-Zewi, and Shubhangi Saraf. Locally testable and locally correctable codes approaching the gilbert-varshamov bound. *IEEE Trans. Inf. Theory*, 64(8):5813–5831, 2018. 2, 4

[GM12] Oded Goldreich and Or Meir. The tensor product of two good codes is not necessarily robustly testable. *Information Processing Letters*, 112(8-9), 2012. 9

[Gol05] Oded Goldreich. Short locally testable codes and proofs (survey). ECCC Technical Report TR05-014, 2005. 2

[Gol10] Oded Goldreich. *Short Locally Testable Codes and Proofs: A Survey in Two Parts*, pages 65–104. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. 2, 4

[Gol17] Oded Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017. 8

[GS06] Oded Goldreich and Madhu Sudan. Locally testable codes and PCPs of almost-linear length. *J. of the ACM*, 53(4):558–655, 2006. 2, 4, 8

[JL99] B. Jordan and R. Livne. The Ramanujan property for regular cubical complexes. *Duke Mathematical Journal*, 105:85–103, 1999. 6, 11

[KKL14] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Ramanujan complexes and bounded degree topological expanders. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 484–493, 2014. 5

[KL12] T. Kaufman and A. Lubotzky. Edge transitive Ramanujan graphs and symmetric LDPC good codes. In *Proceedings of the 44^{th} symposium on Theory of Computing*, pages 359–366. ACM, 2012. 26, 27

[KL14] Tali Kaufman and Alexander Lubotzky. High dimensional expanders and property testing. In *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 501–506, 2014. 5

[KMRS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):11:1–11:42, 2017. 2, 4

[Kur55] A.G. Kurosh. *The Theory of Groups, vol. 2.* Chelsea publishing company, New York, 1955. 25

[LFKN92] 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. 2

[LM07] A. Lubotzky and R. Meshulam. A Moore bound for simplicial complexes. *Bulletin of the London Mathematical Society*, 39(3):353–358, 2007. 28

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988. 5

[LSV05a] 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. 5, 22, 25, 26, 27, 28

[LSV05b] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Ramanujan complexes of type $\tilde{A}_d$. *Israel J. Math.*, 149(1):267–299, 2005. 5

[Lub94] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures.* Modern Birkhäuser Classics. Birkhäuser Verlag, Basel, 1994. With an appendix by Jonathan D. Rogawski. 25

[Lub21] Alexander Lubotzky. The c^3 problem: Locally testable codes with constant rate and constant distance. MPS Conference on High-Dimensional Expanders, `https://www.simonsfoundation.org/event/2021-mps-conference-on-high-dimensional-expanc` 2021. 7

[Mei08] Or Meir. Combinatorial construction of locally testable codes. In *Proc. 40th ACM Symp. on Theory of Computing*, pages 285–294, 2008. 4

[Mor94] M. Morgenstern. Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *Journal of Combinatorial Theory, Series B*, 62(1):44–62, 1994. 22, 25, 26

[Moz91] S. Mozes. A zero entropy, mixing of all orders tiling system, symbolic dynamics and its applications. *Contemp. Math*, 135:319–325, 1991. 6, 11

[Mum79] David Mumford. An algebraic surface with k ample, $(K^2) = 9, p_g = q = 0$. *American Journal of Mathematics*, 101, 02 1979. 6

[Opp18] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part I: descent of spectral gaps. *Discret. Comput. Geom.*, 59(2):293–330, 2018. 5

[PS94] A. Polishchuk and D. Spielman. Nearly linear size holographic proofs. In *Proc. 26th ACM Symp. on Theory of Computing*, pages 194–203, 1994. 2, 8

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. 2

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1723–1731, 1996. Codes and complexity. 2, 23

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1710–1722, 1996. Codes and complexity. 2, 4, 5, 6, 14

[Tan81] R.M̃. Tanner. A recursive approach to low complexity codes. *IEEE Trans. Inform. Theory*, Vol. IT-27,(5):533–547, 1981. 2, 14

[Val05] Paul Valiant. The tensor product of two codes is not necessarily robustly testable. In *APPROX-RANDOM*, pages 472–481, 2005. 9

[Var98] Yakov Varshavsky. p-adic uniformization of unitary shimura varieties. *Publications Math ’e matiques of the Institut des Hautes É tudes Scientalités*, 87(1):57–119, 1998. 6

A Robust Testability and Agreement Testability

In this section we show the equivalence between the two notions, proving Lemma 2.9

Claim A.1 (Robust-testability implies agreement-testability). Assume $\delta_i = \mathrm{dist}(C_i)$ for $i = 1,2$. If $C_1 \otimes C_2$ is ρ-robustly testable then $C_1 \otimes C_2$ is κ-agreement testable, for $\kappa^{-1} = \frac{1}{2\delta_1\rho} + \frac{1+1/(2\rho)}{\delta_2}$.

Proof. Suppose $w_1 \in C_1 \otimes \mathbb{F}_2^{n_2}$, and $w_2 \in \mathbb{F}_2^{n_1} \otimes C_2$. Let $f = w_1$, so $\delta^{\mathrm{col}}(f) = 0$, and observe that since $w_2(i,\cdot) \in C_2$ for each j,

$$\delta^{\mathrm{row}}(f) = \mathop{\mathbb{E}}_{i\in[n_1]} \mathrm{dist}(f(i,\cdot), C_2) \leqslant \mathop{\mathbb{E}}_{i\in[n_1]} \mathrm{dist}(f(i,\cdot), w_2(i,\cdot)) = \mathrm{dist}(w_1, w_2).$$

By the robust-testability of $C_1 \otimes C_2$ there is some $w \in C_1 \otimes C_2$ such that

$$\mathrm{dist}(w, w_1) = \mathrm{dist}(w, f) \leqslant \frac{1}{\rho} \cdot \frac{\delta^{\mathrm{row}}(f) + \delta^{\mathrm{col}}(f)}{2} \leqslant \frac{1}{2\rho} \cdot (\mathrm{dist}(w_1, w_2) + 0).$$

By the triangle inequality $\operatorname{dist}(w,w_2) \leqslant \operatorname{dist}(w,w_1)+\operatorname{dist}(w_1,w_2) \leqslant (1+\frac{1}{2\rho})\operatorname{dist}(w_1,w_2)$.

Next, observe that $\mathbb{P}_j[w(\cdot,j) \neq w_1(\cdot,j)] \cdot \delta_1 \leqslant \operatorname{dist}(w,w_1)$, and similarly $\mathbb{P}_i[w(i,\cdot) \neq w_2(i,\cdot)] \cdot \delta_2 \leqslant \operatorname{dist}(w,w_2)$. Altogether,

$$\begin{aligned}\mathbb{P}_j[w(\cdot,j) \neq w_1(\cdot,j)] + \mathbb{P}_i[w(i,\cdot) \neq w_2(i,\cdot)] &\leqslant \frac{1}{\delta_1}\operatorname{dist}(w,w_1) + \frac{1}{\delta_2}\operatorname{dist}(w,w_2) \\ &\leqslant (\frac{1}{2\rho\delta_1} + \frac{1+1/(2\rho)}{\delta_2}) \cdot \operatorname{dist}(w_1,w_2)\end{aligned}$$

proving the claim with $\kappa^{-1} = \frac{1}{2\rho\delta_1} + \frac{1+1/(2\rho)}{\delta_2}$, or $\kappa = \frac{2\rho\delta_1\delta_2}{\delta_2+\delta_1(1+2\rho)}$. □

Note that in case $\delta_1 = \delta_2 = \delta$ the statement simplifies slightly to $\kappa = \frac{\rho\delta}{\rho+1}$. The other direction, that we don't need here, is even simpler,

Claim A.2 (Agreement-testability implies robust-testability). If $C_1 \otimes C_2$ is κ-agreement testable, then $C_1 \otimes C_2$ is ρ-robustly testable for $\rho = \frac{\kappa}{2(\kappa+1)}$.

Proof. Assume $C_1 \otimes C_2$ is κ-agreement-testable. Let $w \in \mathbb{F}_2^{n_1 \times n_2}$ satisfy $\rho(w) = \delta$. Let $w_1 \in C_1 \otimes \mathbb{F}_2^{n_2}$ be such that $\delta^{col}(w) = \operatorname{dist}(w,w_1)$. Let $w_2 \in \mathbb{F}_2^{n_1} \otimes C_2$ be such that $\delta^{row}(w) = \operatorname{dist}(w,w_2)$. By the triangle inequality,

$$\operatorname{dist}(w_1,w_2) \leqslant \operatorname{dist}(w_1,w) + \operatorname{dist}(w,w_2) = \delta^{col}(w) + \delta^{row}(w) = 2\rho(w).$$

By the κ-agreement testability there is some $w' \in C_1 \otimes C_2$ such that

$$\kappa \cdot (\mathbb{P}_i[w_1(i,\cdot) \neq w'(i,\cdot)] + \mathbb{P}_j[w_2(\cdot,j) \neq w'(\cdot,j)]) \leqslant \mathbb{P}_{i,j}[w_1(i,j) \neq w_2(i,j)]) = \operatorname{dist}(w_1,w_2) \leqslant 2\rho(w).$$

But clearly

$$\operatorname{dist}(w_1,w') + \operatorname{dist}(w',w_2) \leqslant \mathbb{P}_i[w_1(i,\cdot) \neq w'(i,\cdot)] + \mathbb{P}_j[w_2(\cdot,j) \neq w'(\cdot,j)] \tag{A.1}$$

so again by the triangle inequality,

$$\begin{aligned}\operatorname{dist}(w,w') &\leqslant \frac{1}{2}(\operatorname{dist}(w,w_1) + \operatorname{dist}(w_1,w') + \operatorname{dist}(w,w_2) + \operatorname{dist}(w_2,w')) \\ &= \frac{1}{2}(\operatorname{dist}(w,w_1) + \operatorname{dist}(w,w_2) + \operatorname{dist}(w_1,w') + \operatorname{dist}(w_2,w')) \\ &\leqslant \rho(w) + \kappa^{-1} \cdot \rho(w) = \frac{\kappa+1}{\kappa} \cdot \rho(w).\end{aligned}$$

□