A Strange Square Complex and c^3-Locally Testable Codes

Matthew T. Regehr

April 15, 2022

This project is an attempt to communicate as concisely the details I found most relevant[1] in the result of [DEL+21] on the existence of locally testable codes with constant rate, constant distance, and constant locality (c^3). I did some refactoring of the proofs and tried to suppress details I found less enlightening where possible. Feedback is welcome :)

The project is organized as follows:

- Section 1 on preliminaries discusses codes, agreement and local testability, as well as spectral expansion. Feel free to skip this but refer back to it later if something is unclear (I use some slightly unusual notation for walks that I find convenient).
- Section 2 describes the left-right Cayley complex, a half combinatorial half algebraic object that provides the scaffolding on which we build good locally testable codes.
- Section 3 discusses this left-right Cayley code and its properties. Here, we prove local testability under appropriate assumptions. This is where most of the magic happens.
- Finally, Section 4 instantiates everything and proves the main result. It is good to skip to this section now and briefly peak at the main result to get a sense of the destination.

1 Preliminaries

1.1 Codes

For the purposes of this project, we use "code" as a shorthand for binary linear block codes.

Definition 1. *A binary linear block code is a linear subspace C of the vector space $\mathbb{F}_2^L$ where L is some finite index set. We call $|L|$ the "block length" of C and $\dim C$ its "message length". The "rate" of C is*

$$\rho(C) := \frac{\dim C}{|L|}$$

and the (normalized) "distance" of C is

$$\delta(C) := \frac{\min\{\|c - c'\|_1 : c, c' \in C, c \neq c'\}}{|L|} = \frac{\min\{\|c\|_1 : c \in C, c \neq 0\}}{|L|}.$$

Lastly, the distance of a string $w \in \mathbb{F}_2^L$ to the code C is

$$d(w, C) := \frac{\min\{\|w - c\|_1 : c \in C\}}{|L|}.$$

[1] Relevant to CS 860 at UWaterloo in W2022 (`https://cs.uwaterloo.ca/~lapchi/cs860/`).

We will also make frequent use of tensor codes, which consist of matrices whose rows belong to one given code and whose columns belong to another given code.

Definition 2. *Let $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$ be codes. The tensor code of C_A and C_B is*

$$C_A \otimes C_B := \{c \in \mathbb{F}_2^{A \times B} : \forall a \in A \ \forall b \in B \ c(\cdot, b) \in C_A, c(a, \cdot) \in C_B\}$$

where $c(\cdot, b)$ denotes the string $a' \mapsto c(a', b)$ in $\mathbb{F}_2^A$ and likewise for $c(a, \cdot)$.

1.2 Agreement Testable Codes

Agreement testability is a weaker property of a code than local testability (which we will define soon). However, agreement testability is a less rare property that can be refined to local testability once we have the appropriate tools.

Definition 3. *Let $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$ be codes. We say that C_A and C_B are ϕ-agreement testable when, for all $f^A \in C_A \otimes \mathbb{F}_2^B$ and $f^B \in \mathbb{F}_2^A \otimes C_B$, there is $w \in C_A \otimes C_B$ such that*

$$\mathbb{P}_{(a,b) \sim \mathcal{U}(A \times B)}(f^A(a, b) \neq f^B(a, b)) \geq \phi \max \left\{ \mathbb{P}_{a \sim \mathcal{U}(A)}(f^A(a, \cdot) \neq w(a, \cdot)), \mathbb{P}_{b \sim \mathcal{U}(B)}(f^B(\cdot, b) \neq w(\cdot, b)) \right\},$$

i.e. the fraction of entries in which f^A and f^B disagree is proportional to the fraction of rows in which f^A and w differ as well as the fraction of columns in which f^B and w differ.

Lemma 1. *For every $\rho \in (0, 1)$, there are $\delta, \phi > 0$ and $D \in \mathbb{Z}^+$ such that, for every sufficiently large multiple d of D, there is a code $C_0 \subseteq \mathbb{F}_2^d$ for which $\rho(C_0) \geq \rho$, $\delta(C_0) \geq \delta$, and $C_0 \otimes C_0$ is ϕ-agreement testable.*

We do not prove the lemma here to avoid a lengthy (but very interesting) tangent. The idea is to consider random low-density parity check (LDPC) codes, introduced by [Gal63]. It is a straightforward combinatorial calculation to to show that the factor graph of a random LDPC code is a bipartite expander with high probability. Then, one can argue that the tensor product of LDPC codes on expanding factor graphs is agreement testable [DSW06].

1.3 Locally Testable Codes

Informally, a q-local tester for a code randomly chooses a set I of q bits from a string c to read and then accepts or rejects c depending on whether the read bits $c|_I$ belong to some set of valid local views. More precisely:

Definition 4. *A q-local tester for a code $C \subseteq \mathbb{F}_2^L$ is a pair $(Q, \mathcal{V})$ where Q is a probability distribution over $\binom{L}{q}$ and $\mathcal{V} \subseteq \bigcup_{I \in \binom{L}{q}} \mathbb{F}_2^I$.*

We say that $(Q, \mathcal{V})$ is complete when, for any $c \in C$,

$$\mathbb{P}_{I \sim Q}(c|_I \in \mathcal{V}) = 1.$$

Finally, $(Q, \mathcal{V})$ is said to be κ-sound ($\kappa \in (0, 1)$) when, for any $w \in \mathbb{F}_2^L$, we have that

$$\mathbb{P}_{I \sim Q}(w|_I \notin \mathcal{V}) \geq \kappa d(w, C).$$

1.4 Expansion

We consider a slightly unusual notion of expansion that allows us to speak more easily of walks on non-regular graphs. It will be notationally convenient to consider spaces of signed measures on a set but it is fine to think of the more usual spaces $\mathbb{R}^{|V|}$ instead.

Definition 5. *Let V be any finite set. We associate to V the Hilbert space $(\mathcal{M}(V), \langle \cdot, \cdot \rangle)$ where $\mathcal{M}(V)$ denotes the set of signed measures on V and $\langle f, g \rangle := \sum_{v \in V} f(\{v\}) g(\{v\})$. We identify a point $v \in V$ with the Dirac measure δ_v (these form a basis). In the same spirit, we identify 1 with the uniform distribution $\mathcal{U}(V)$. A Markov operator on this space is a positive semi-definite linear map $M : \mathcal{M}(V) \to \mathcal{M}(V)$ s.t. $M1 = 1$. We say that M is a λ-expander (for $\lambda \in [0, 1)$) if for every $f \in \mathcal{M}(V)$ such that $\langle f, 1 \rangle = 0$, we have*

$$\langle Mf, f \rangle \leq \lambda \langle f, f \rangle .$$

An important random walk on graphs is the neighbor walk in which we choose a neighbor of the current vertex uniformly at random at every step.

Definition 6. *The neighbor walk operator of a graph $G = (V, E)$, denoted N_G, is the Markov operator $\mathcal{M}(V) \to \mathcal{M}(V)$ s.t. $N_G v = \mathcal{U}(\{v' \in V : \{v, v'\} \in E\})$ for all $v \in V$. We say that G is a λ expander if N_G is a λ-expander.*

The crucial property of expander walks is that they remain in small regions with low probability. We will exploit this fact later to show that a certain region must be large.

Theorem 1 (Alon–Chung [AC88])**.** *Let M be a λ-expander on V and let $\varnothing \neq T \subseteq V$ be such that $\mathbb{E}_{v \sim \mathcal{U}(T)}[\mathbb{P}_{v' \sim Mv}(v' \in T)] \geq p$. Then $|T| \geq (p - \lambda) |V|$.*

The proof of this key result is actually quite simple. The main idea is to decompose $\mathcal{U}(T)$ as $\frac{|T|}{|V|} 1 + h$ for some $h \perp 1$ and notice that $\langle M(\mathcal{U}(T)), \mathcal{U}(T) \rangle = \mathbb{E}_{v \sim \mathcal{U}(T)}[\mathbb{P}_{v' \sim Mv}(v' \in T)]$.

2 The Left-Right Cayley Complex

In order to lift agreement testable codes to locally testable codes, we require some algebraic apparatus. Recall that a group is a set G equipped with an associative binary operator $\cdot : G^2 \to G$, an identity $e \in G$, and inverses $(\cdot)^{-1} : G \to G$. We say that $T \subseteq G$ is symmetric when it is closed under inversion, i.e. $g^{-1} \in T$ for all $g \in T$. We call such a symmetric set T a generating set of G if any element of G can be expressed as a product of elements in T.

Definition 7. *Let G be a group and $A, B \subseteq G$ symmetric generating sets not containing the identity of G such that $|A| = |B|$ and $ag \neq gb$ for all $g \in G$, $a \in A$, and $b \in B$. The left-right Cayley (LRC) complex on G with left edges in A and right edges in B is $\mathrm{Cay}(A, G, B) := X$ where $X(0) := G$ is the set of "vertices" of $\mathrm{Cay}(A, G, B)$,*

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

forms its "edges", and

$$X(2) := \{[a, g, b] : a \in A, g \in G, b \in B\}$$

makes up its "squares" where $[a, g, b] := \{g, ag, agb, gb\}$ *denotes the square obtained by beginning at* g*, travelling along the* a*-edge* $\{g, ag\}$*, the* b*-edge* $\{ag, agb\}$*, the* a^{-1}*-edge* $\{agb, gb\}$*, and finally the* b^{-1}*-edge* $\{gb, g\}$*.*

For convenience, we denote by $X^A := (X(0), X^A(1))$ *and* $X^B := (X(0), X^B(1))$ *the* A *and* B *1-skeletons of* X*. We also write* $X_g(2)$ *for the set of squares containing a vertex* $g \in X(0)$ *and* $X_e(2)$ *for the set of squares containing an edge* $e \in X(1)$*.*

A few quick remarks on the pre-conditions for the curious: Symmetry lets us "complete the square" by following a^{-1} and b^{-1} edges. Using generating sets makes X^A and X^B connected. We eliminate self-loops by excluding the identity from A and B. Lastly, $ag \neq gb$ prevents the squares of the complex from "collapsing" into edges.

Definition 8. *An LRC complex* X *is a* λ*-expander if* X^A *and* X^B *are both* λ*-expanders.*

2.1 The Parallel Walk

In this subsection, we introduce a random walk on the edges of an LRC complex given by jumping across squares to "parallel" edges. This walk may not seem interesting now but we will see that it captures an important aspect of the structure of the complex once we analyze the LRC code in Section 3. For this reason, it may be easier to skip this subsection for now and come back later once the role of the parallel walk becomes clear.

Definition 9. *Let* $X := \mathrm{Cay}(A, G, B)$ *be an LRC complex. The parallel walk graph of* X*, denoted* $X^{||}$*, is the neighbor walk operator of the graph with vertices* $X(1)$ *and an edge from* $\{g, ag\}$ *to* $\{agb, gb\}$ *as well as an edge from* $\{ag, agb\}$ *to* $\{gb, g\}$ *for every* $g \in G$*,* $a \in A$*,* $b \in B$*. See Figure 2.1. The parallel walk operator of* X*, denoted* P_X*, is just the neighbor walk operator on* $X^{||}$*.*

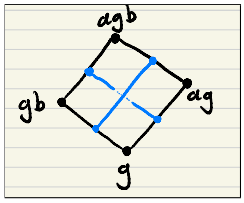

Figure 1: A snapshot of the parallel walk graph $X^{||}$ (shown in blue) over a single square of X (shown in black).

Lemma 2. *If* $X = \mathrm{Cay}(A, G, B)$ *is a* λ*-expanding LRC complex, then every connected component of* $X^{||}$ *is a* λ*-expander under* P_X *of size at least* $\frac{|X(1)|}{|A|+|B|}$*.*

The idea of the proof is to show that each component of $X^{||}$ is of the form $X^{||}|_{X^\sigma(1)}$ for a "label" $\sigma \in \tilde{A} \cup \tilde{B}$ where $\tilde{A} := \{\{a, a^{-1}\} : a \in A\}$, $\tilde{B} := \{\{b, b^{-1}\} : b \in B\}$, and $X^\sigma(1) := \{\{g, cg\} : c \in \sigma, g \in G\}$ denotes the set of edges in $X(1)$ "labelled" by σ.

Then, for $\sigma = \{a, a^{-1}\} \in \tilde{A}$ such that $a \neq a^{-1}$, $X^{||}|_{X^\sigma(1)}$ is isomorphic to X^B via $\{g, ag\} \mapsto g$ and is thus a λ-expander of size $|X(0)| = 2\frac{|X(1)|}{|A|+|B|}$. The case $a = a^{-1}$ is slightly trickier (see the proof of Lemma 3.13 in [DEL+21]). The case $\sigma \in \tilde{B}$ is analogous.

2.2 The Down-Neighbor-Up Walk

In this subsection, we discuss a walk whose role is complementary to that of the parallel walk in characterizing the structure of an LRC complex. In this walk, one steps from an edge "down" to a random incident vertex, walks to a random neighbor of this vertex, and then steps back "up" to a random incident edge. Once again, it is fine to just skip ahead to Section 3 and come back once the role of this walk become clearer.

Definition 10. *Let X be an LRC complex. The down operator on X is the unique linear map $D_X : \mathcal{M}(X(1)) \to \mathcal{M}(X(0))$ s.t. $D_X e = \mathcal{U}(e)$ for all $e \in X(1)$. The up operator for X is the unique linear map $U_X : \mathcal{M}(X(0)) \to \mathcal{M}(X(1))$ s.t. $U_X g = \mathcal{U}(\{e \in X(1) : g \in e\})$ for $g \in X(0)$. Finally, the down-neighbor-up operator for X is $\nabla_X := U_X N_X D_X$.*

Lemma 3. *If X is a λ-expander, then so is ∇_X.*

The main idea of this proof is to notice that U_X and D_X are adjoint operators. That is, $\langle D_X f, g\rangle = \langle f, U_X g\rangle$ for all $f \in \mathcal{M}(X(1))$ and $g \in \mathcal{M}(X(0))$. Moreover, since X^A and X^B are λ-expanders, it follows immediately that $N_X = \frac{N_{X^A} + N_{X^B}}{2}$ is as well. Then, for any $f \in \mathcal{M}(X(1)$ with $\langle f, 1\rangle = 0$, we have $\langle D_X f, 1\rangle = \langle f, U_X 1\rangle = \langle f, 1\rangle = 0$ and thus

$$\langle \nabla_X f, f\rangle = \langle U_X N_X D_X f, f\rangle = \langle N_X D_X f, D_X f\rangle \leq \lambda \langle D_X f, D_X f\rangle \leq \lambda \langle f, f\rangle .$$

3 The Left-Right Cayley Code

We now see how such a complex can be used to construct locally testable codes with good rate and distance from an agreement testable base code.

Definition 11. *Let $X := \mathrm{Cay}(A, G, B)$ be a LRC complex and let $C_0 \subseteq \mathbb{F}_2^{A\times B}$ be a code. We say that X lifts C to the code*

$$C_0^X := \{c \in \mathbb{F}_2^{X(2)} : \forall g \in X(0)\ c([\cdot, g, \cdot]) \in C_0\}$$

where $c([\cdot, g, \cdot])$ denotes the string $(a, b) \mapsto c([a, g, b])$ in $\mathbb{F}_2^{A\times B}$.

For convenience, we denote by $C_g := \{c|_{X_g(2)} : c \in C\}$ the set of "local views" of the code when restricted to the squares containing $g \in X(0)$.

It is straightforward to check that C_0^X is in fact a linear code. Moreover, we note that C_0 will always be taken to be a tensor code $C_A \otimes C_B$, in which case we have

$$C_0^X = \{c \in \mathbb{F}_2^{X(2)} : \forall g \in X(0)\ \forall a \in A\ \forall b \in B\ c([\cdot, g, b]) \in C_A, c([a, g, \cdot]) \in C_B\}. \quad (1)$$

3.1 Rate

Lemma 4. *Let $X := \mathrm{Cay}(A, G, B)$ be a LRC complex and let $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$ be any codes with rates at least ρ. Then*

$$\rho((C_A \otimes C_B)^X) \geq 4\rho - 3.$$

The details of this proof are not so interesting and thus we omit them. Nonetheless, the main idea is just to count number of linear constraints in Equation 1 in terms of the dimensions of C_A and C_B and then use the fact that the number of squares in X is $\frac{|G||A||B|}{4}$ (the 4 reflects the fact that each square has four representations as $[a, g, b]$).

3.2 Distance

Lemma 5. *Let $X := \mathrm{Cay}(A, G, B)$ be a λ-expanding LRC complex and let $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$ be any codes with distances at least δ. Then*

$$\delta((C_A \otimes C_B)^X) \geq \delta^2(\delta - \lambda).$$

Again, we skip some details. The idea is to consider a non-zero $c \in (C_A \otimes C_B)^X$ and then find $g_0 \in G$ such that $c|_{X_g(2)} \neq 0$. Then, for $w(a, b) := c([a, g_0, b])$, we have $0 \neq w \in C_A \otimes C_B$ and hence there is a δ fraction of rows $A^* \subseteq A$ and a δ-fraction of columns $B^* \subseteq B$ such that $w(a, \cdot) \neq 0$ and $w(\cdot, b) \neq 0$ for all $a \in A^*$ and $b \in A^*$. Then, for any $a \in A^*$, apply the Alon–Chung lemma to the support of

$$f_a : X^B(1) \to \mathbb{F}_2, \{g, gb\} \mapsto c([a, g, b])$$

to show that it has size at least $\delta(\delta - \lambda)\left|X^B(1)\right|$. Then conclude by noting that a δ-fraction of these f_a satisfy this property.

3.3 Local Testability

Now we come to the meat of the paper. We need to argue that, under the appropriate assumptions, our LRC codes are locally testable with locality and soundness constant w.r.t. their block size. Our test is simple: choose $g \in X(0)$ uniformly at random and check that a candidate agrees with a true codeword on the $|X_g(2)| = |\{[a, g, b] : a \in A, b \in B\}| = |A|\,|B|$ squares containing g.

Definition 12. *Let $C = C_0^X$ be an LRC code. The canonical local test for C is $(Q, \mathcal{V})$ where Q is the distribution on $\binom{X(2)}{|A||B|}$ such that $g \sim \mathcal{U}(X(0)) \implies X_g(2) \sim Q$ and where $\mathcal{V} := \bigcup_{g \in G} C_g$. By construction of C, this test rejects a candidate $w \in \mathbb{F}_2^{X(2)}$ with probability*

$$rej_C(w) := \mathop{\mathbb{P}}_{g \sim \mathcal{U}(X(0))}(w|_{X_g(2)} \notin C_0).$$

This local test is clearly complete, but we claim that it is also sound as long as the underlying complex is an expander and the base codes are agreement testable.

Theorem 2. *Let X be a λ-expanding LRC complex and $C = (C_A \otimes C_B)^X$ an LRC code such that C_A and C_B have distance at least δ and are ϕ-agreement testable. As long as δ and ϕ are large enough[2] w.r.t. λ, then*

$$rej_C(w) \geq \kappa d(w, C)$$

for all $w \in \mathbb{F}_2^{X(2)}$, where κ is some constant[3] independent of $|X(0)|$. In particular, C is locally testable with locality $|A|\,|B|$ and soundness κ.

---

[2] $\lambda < \frac{\phi\delta}{\phi+8}$

[3] $\kappa := \min\left\{\frac{\frac{\phi\delta}{\phi+8} - \lambda}{2(|A|+|B|)}, \frac{1}{4(1+|A|+|B|)}\right\}$.

Algorithm 1 A decoder for a left-right Cayley code $C = C_0^X$

procedure $\text{DECODE}_C(w)$
  for $g \leftarrow X(0)$ **do**
    $c_g \leftarrow \arg\min_{c \in C_g} \left\| c - w|_{X_g(2)} \right\|_1$
  end for
  repeat
    for $g \leftarrow X(0)$ **do**
      if $\exists c'_g \in C_g$ that reduces $R := \{e = \{g_1, g_2\} \in X(1) : c_{g_1}|_{X_e(2)} \neq c_{g_2}|_{X_e(2)}\}$ **then**
        $c_g \leftarrow c'_g$
      end if
    end for
  until all ifs fail
  if $R \neq \varnothing$ **then**
    return "stuck"
  else
    return $[a, g, b] \mapsto c_g([a, g, b])$
  end if
end procedure

The key to proving soundness lies in the analysis of Algorithm 1, about which we now make a few useful remarks: First, if you forgot what C_g means, refer to Definition 11. Note that the algorithm always terminates since R decreases in size by at least one on every iteration of the repeat loop. Finally, note that the output of the algorithm $[a, g, b] \mapsto c_g([a, g, b])$ is well-defined since $R = \varnothing$ implies that c_{g_1} and c_{g_2} must agree on all squares to which both g_1 and g_2 belong (it is a good exercise to convince yourself of this before proceeding to the analysis).

The following notation is helpful for the analysis.

Definition 13. *We denote by $c_g^0(w)$ the value of c_g before the repeat loop executes and by $c_g^{-1}(w)$ its value after the repeat loop terminates when* DECODE_C *is run on w. We give $R^0(w)$ and $R^{-1}(w)$ analogous meanings.*

Proposition 1. *For any $w \in \mathbb{F}_2^{X(2)}$, $\frac{|R^0(w)|}{|X(1)|} \leq 2rej_C(w)$.*

This is a short proof and I really encourage the reader to try writing it out formally just to check their basic understanding of the objects at play here. The idea is just to apply a kind of triangle inequality: for any edge $\{g_1, g_2\}$ that contributes to $R^0(w)$, $c_{g_1}^0(w)$ and $c_{g_2}^0(w)$ disagree, so they can't both agree with w (when restricted to the appropriate squares).

Now, we arrive at what I believe is the most important (and interesting) result in the paper. Roughly, it says if a string is far from an LRC code, then it must violate a constant fraction of the linear constraints defining the code. This is in stark contrast to earlier attempts at locally testable codes such as the LDPC codes in which a string may be very far from the code yet only violate one constraint [Gol10]. The stars of this show are the expansion of the underlying complex, the agreement testability of the base code, and their interplay with the Alon–Chung lemma (Theorem 1).

Lemma 6. *Let X be a λ-expander and $C = (C_A \otimes C_B)^X$ be an LRC code such that C_A and C_B have distance at least δ and are ϕ-agreement testable. Then there is a constant[4] α independent of $|X(0)|$ such that, for any $w \in \mathbb{F}_2^{X(2)}$ on which* DECODE$_C$ *gets stuck, we have*

$$\left|R^{-1}(w)\right| \geq \alpha \left|X(1)\right|.$$

Before we dive into the proof, now is a good time to quickly review Subsections 2.1 and 2.2 in case you don't remember our good friends P_X and ∇_X (the parallel and down-neighbor-up walks respectively).

Proof of Lemma 6. To ease the notational burden throughout this proof, we supress (w) from the $R^{-1}(w)$ and $c_g^{-1}(w)$.

Now, since DECODE$_C$ gets stuck on w, we have $R^{-1} \neq \varnothing$. Our goal is to show that R^{-1} contains a constant fraction of the edges $X(1)$. We will achieve this by applying an averaging argument to the random walks P_X and ∇_X to show that at least one of them remains in R^{-1} with high probability and then applying Alon–Chung (Theorem 1). In particular, we claim that, for some carefully chosen[5] $\gamma \in (0,1)$,

$$\gamma \mathop{\mathbb{P}}_{e' \sim P_X e}(e' \in R^{-1}) + (1-\gamma) \mathop{\mathbb{P}}_{e' \sim \nabla_X e}(e' \in R^{-1}) \geq \gamma\delta \tag{2}$$

holds for all $e \in R^{-1}$. To this end, consider any edge $e \in R^{-1}$. Without loss of generality, assume $e = \{g_0, a_0 g_0\}$ (the argument is symmetric for a b-edge).

We first claim that a constant fraction of the edges sharing a square with e must also belong to R^{-1}. Let's give these edges names first. Define

$$\begin{aligned} E^L &:= \{\{g_0, g_0 b\} : b \in B\}; \\ E^{||} &:= \{\{a_0 g_0 b, g_0 b\} : b \in B\}; \text{and} \\ E^R &:= \{\{a_0 g_0, a_0 g_0 b\} : b \in B\}. \end{aligned}$$

See Figure 3.3 for a picture of these sets. Now, since $e \in R^{-1}$, we have $c_{g_0}|_{X_e(2)} \neq c_{a_0 g_0}|_{X_e(2)}$. But $\{c|_{X_e(2):c\in C}\}$ is isomorphic to C_B (via $c \mapsto b \mapsto c([a_0, g_0, b])$) and thus has distance at least δ, so $c_{g_0}|_{X_e(2)}$ and $c_{a_0 g_0}|_{X_e(2)}$ must disagree on at least $\delta |B|$ squares. On any such square $[a_0, g_0, b]$, by transitivity either c_{g_0} must disagree with $c_{g_0 b}$, $c_{g_0 b}$ must disagree with $c_{a_0 g_0 b}$, or $c_{a_0 g_0 b}$ must disagree with $c_{a_0 g_0}$. Thus

$$\underbrace{\left|R^{-1} \cap E^L\right|}_{=:m^L} + \underbrace{\left|R^{-1} \cap E^{||}\right|}_{=:m^{||}} + \underbrace{\left|R^{-1} \cap E^R\right|}_{=:m^R} \geq \delta |B|. \tag{3}$$

With this in the back of our minds, let's start by computing the first term of the LHS of Equation 2. Since $E^{||}$ are exactly the neighbors of e in the parallel walk graph $X^{||}$ and since P_X is the neighbor walk operator on this graph,

$$\mathop{\mathbb{P}}_{e' \sim P_X e}(e' \in R^{-1}) = \frac{\left|R^{-1} \cap E^{||}\right|}{|E^{||}|} = \frac{m^{||}}{|B|}.$$

[4] $\alpha := \frac{\frac{\phi\delta}{\phi+8} - \lambda}{|A|+|B|}$

[5] The choice wouldn't make sense now, so we will define it later at the right moment.

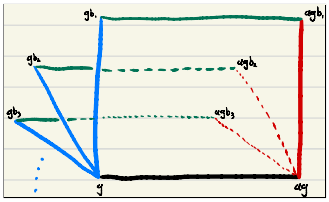

Figure 2: A few squares containing an edge $\{g, ag\}$ with E^L shown on the left in blue, E^R on the right in red, and $E^{||}$ on the top in green.

Now, we come to the most technically involved part of the proof: we need to lower bound the second term of Equation 2. To this end, consider the strings $f^A_G(a,b) := c^{-1}_{gb}([a,g,b])$ and $f^B_g(a,b) := c^{-1}_{ag}([a,g,b])$. Since $c^{-1}_{gb} \in C_{gb}$ and $c^{-1}_{ag} \in C_{ag}$, it follows that $f^A_g \in C_A \otimes \mathbb{F}_2^B$ and $f^B_g \in \mathbb{F}_2^A \otimes C_B$. By agreement testability (recall Definition 3), there is $w_g \in C_A \otimes C_B$ such that the the fraction of entries in which f^A_g and f^B_g differ is at least ϕ times the fraction of columns in which f^A_g and w_g differ. Set $c'_g([a,g,b]) := w_g(a,b)$ so that $c' \in C_g$. Now, writing $\mathbb{P}_a$ as shorthand for $\mathbb{P}_{a\sim\mathcal{U}(A)}$ (and similarly for $\mathbb{P}_b$), we have

$$
\begin{aligned}
\mathop{\mathbb{P}}_{e'\sim\nabla_X e}(e' \in R^{-1}) &= \mathop{\mathbb{E}}_{g\sim D_X e}\left[\mathop{\mathbb{P}}_{e'\sim U_X N_X g}(e' \in R^{-1})\right] \\
&\geq \mathop{\mathbb{E}}_{g\sim D_X e}\left[\frac{1}{2^2}\mathop{\mathbb{P}}_{a,b}(\{ag, agb\} \in R^{-1}) + \frac{1}{2^2}\mathop{\mathbb{P}}_{a,b}(\{gb, agb\} \in R^{-1})\right] \\
&\qquad\qquad (N_X \text{ takes } a\text{-neighbor and } U_X \text{ takes } b\text{-edge or vice-versa}) \\
&\geq \frac{1}{4}\mathop{\mathbb{E}}_{g\sim D_X e}\left[\mathop{\mathbb{P}}_{a,b}(\{ag, agb\} \in R^{-1} \text{ or } \{gb, agb\} \in R^{-1})\right] \\
&\geq \frac{1}{4}\mathop{\mathbb{E}}_{g\sim D_X e}\left[\mathop{\mathbb{P}}_{a,b}(c^{-1}_{ag}([a,g,b]) \neq c^{-1}_{gb}([a,g,b]))\right] \\
&\qquad\qquad (\text{Transitivity along path } ag \to agb \to gb) \\
&= \frac{1}{4}\mathop{\mathbb{E}}_{g\sim D_X e}\left[\mathop{\mathbb{P}}_{a,b}(f^A_g(a,b) \neq f^B_g(a,b))\right] \\
&= \frac{\phi}{4}\mathop{\mathbb{E}}_{g\sim D_X e}\left[\mathop{\mathbb{P}}_{b}(f^A_g(\cdot,b) \neq w_g(\cdot,b))\right] \qquad (\text{Agreement testability}) \\
&= \frac{\phi}{4\,|B|}\mathop{\mathbb{E}}_{g\sim D_X e}\left[|\{b \in B : c^{-1}_{gb}|_{X_{\{g,gb\}}(2)} \neq c'_g|_{X_{\{g,gb\}}(2)}\}|\right] \\
&\geq \frac{\phi}{4\,|B|}\mathop{\mathbb{E}}_{g\sim D_X e}\left[|\{b \in B : c^{-1}_{gb}|_{X_{\{g,gb\}}(2)} \neq c^{-1}_{g}|_{X_{\{g,gb\}}(2)}\}|\right] \\
&\qquad\qquad (\text{Were this not true, } \textsc{decode}_C \text{ would have set } c_g \leftarrow c'_g) \\
&= \frac{\phi(m^L + m^R)}{8\,|B|}. \qquad (D_X e = \mathcal{U}(\{g_0, a_0 g_0\}))
\end{aligned}
$$

Finally, choose $\gamma := \frac{\phi}{\phi+8}$ (so that $1-\gamma = 8\gamma/\phi$). As $e \in R^{-1}$ was arbitrary, we have

$$\begin{aligned}
\gamma \mathop{\mathbb{E}}_{e\sim R^{-1}}\left[\mathop{\mathbb{P}}_{e'\sim P_X e}(e' \in R^{-1})\right] + (1-\gamma)\mathop{\mathbb{E}}_{e\sim R^{-1}}\left[\mathop{\mathbb{P}}_{e'\sim \nabla_X e}(e' \in R^{-1})\right] &\geq \frac{\gamma m^{||}}{|B|} + \frac{(1-\gamma)\phi(m^L+m^R)}{8\,|B|} \\
&= \frac{\gamma(m^L+m^{||}+m^R)}{|B|} \\
&\geq \gamma\delta \qquad \text{(Equation 3)}
\end{aligned}$$

In particular, either $\mathbb{E}_{e\sim R^{-1}}[\mathbb{P}_{e'\sim P_X e}(e' \in R^{-1})] \geq \gamma\delta$ or $\mathbb{E}_{e\sim R^{-1}}[\mathbb{P}_{e'\sim \nabla_X e}(e' \in R^{-1})] \geq \gamma\delta$. In the former case, a simple averaging argument implies that there is some connected component K of $X^{||}$ for which $\mathbb{E}_{e\sim R^{-1}\cap K}[\mathbb{P}_{e'\sim P_X e}(e' \in R^{-1}\cap K)] \geq \gamma\delta$, whereupon Lemma 2 together with Theorem 1 yield that

$$\left|R^{-1}\right| \geq \left|R^{-1}\cap K\right| \geq (\gamma\delta-\lambda)\,|K| \geq \frac{\gamma\delta-\lambda}{|A|+|B|}\,|X(1)|\,.$$

Likewise, if $\mathbb{E}_{e\sim R^{-1}}[\mathbb{P}_{e'\sim \nabla_X e}(e' \in R^{-1})] \geq \gamma\delta$, then Lemma 3 and Alon–Chung directly yield

$$\left|R^{-1}\right| \geq (\gamma\delta-\lambda)\,|X(1)|\,.$$

□

Proof of Theorem 2. Let $w \in \mathbb{F}_2^{X(2)}$. We need to show that $\mathrm{rej}_C(w)$ is bounded below by a constant times $d(w,C)$.

First, assume $\textsc{decode}_C$ gets stuck on w. Then, by Proposition 1 as well as Lemma 6,

$$2\mathrm{rej}_C(w) \geq \frac{|R^0(w)|}{|X(1)|} \geq \frac{|R^{-1}(w)|}{|X(1)|} \geq \alpha \geq \alpha d(w,C).$$

On the other hand, assume that $\textsc{decode}_C(w)$ does not get stuck and returns $c \in C$. Let $V_1 := \{g \in X(0) : w|_{X_g(2)} \neq c_g^0(w)\}$ be the vertices at which the local view does not initially agree with w and let $V_2 := \{g \in X(0) : c_g^0(w) \neq c_g^{-1}(w)\}$ be the vertices at which the local view changes during the execution of the algorithm. Since each $c_g^0(w)$ is the closest point in C_g to w, we have

$$|V_1| = \left|\{g \in X(0) : w|_{X_g(2)} \notin C_g\}\right| = \mathrm{rej}_C(w)\,|X(0)|\,.$$

On the other hand, since R decreases by at least one at every step of the repeat loop,

$$|V_2| \leq \left|R^0(w)\right| \leq 2\mathrm{rej}_C(w)\,|X(1)| = \mathrm{rej}_C(w)(|A|+|B|)\,|X(0)|$$

where the second inequality follows from Proposition 1 and the equality follows from the handshake lemma and the fact that every vertex of X has degree $|A|+|B|$. Finally, notice that if a square $s \in X(2)$ does not meet V_1 or V_2, then we must have $w(s) = c(s)$ (since all local views in s match w initially and do not change), so, since every vertex meets $|A|\,|B|$ squares, it follows that

$$d(w,C) \leq \frac{\|w-c\|_1}{|X(2)|} \leq \frac{|V_1\cup V_2|\,|A|\,|B|}{\frac{|X(0)||A||B|}{4}} \leq 4(1+|A|+|B|)\mathrm{rej}_C(w).$$

□

4 Locally Testable Codes of Constant Rate, Distance, and Locality

We've now seen most of the main insights of [DEL$^+$21]. The machinery of the LRC complexes and codes is now in place. It just remains to furnish this machinery with the right group structure and instantiate it with the agreement testable base codes that we have seen.

Theorem 3. *For every rate $\rho \in (0,1)$, there exist $\delta > 0$, $q \in \mathbb{N}$, $\kappa > 0$, and a family of locally testable codes $\{C_n\}_{n\in\mathbb{N}}$ such that C_n has rate at least ρ, distance at least δ, locality q, and soundness κ and such that the block size of C_n tends to infinity as $n \to \infty$.*

To prove the theorem, we require good LRC complexes, whose existence can be proved by considering the projective special linear groups over large finite fields.

Lemma 7. *For every $d \in \mathbb{Z}^+$ and every sufficiently large odd prime power q, there exists a multiple d' of d such that, for all $n \in \mathbb{Z}^+$, there is a group G_n of size much larger than n and a pair $A_n, B_n \subseteq G_n$ of symmetric generating sets with size d' not containing the identity of G_n such that $ag \neq gb$ for all $a \in A_n$, $b \in B_n$, and $g \in G_n$ and such that $\mathrm{Cay}(A_n, G_n, B_n)$ is an $8/\sqrt{d}$-expander.*

Proof of Theorem 3. Let $\rho \in (0,1)$. By Lemma 1, there are $\delta, \phi > 0$ and $D \in \mathbb{Z}^+$ such that, for every sufficiently large multiple d of D, there is a code $C_d \subseteq \mathbb{F}_2^d$ for which $\rho(C_d) \geq \rho$, $\delta(C_d) \geq \delta$, and $C_d \otimes C_d$ is ϕ-agreement testable. By Lemma 7, there is some multiple d of D for which "good" A_n, B_n, and G_n exist and, making d larger as needed, so that $8/\sqrt{d} < \frac{\delta\phi}{\phi+8}$ and so that there is $C_0 \subseteq \mathbb{F}_2^d$ as in Lemma 1. Then, for every $n \in \mathbb{Z}^+$, let $C_{A_n} \subseteq \mathbb{F}_2^{A_n}$ and $C_{B_n} \subseteq \mathbb{F}_2^{B_n}$ be isomorphic to C_0 and put $C_n := (C_{A_n} \otimes C_{B_n})^{\mathrm{Cay}(A_n,G_n,B_n)}$ so that C_n has block size at least $|G_n| \gg n$ and so that

$$\rho(C_n) \geq 4\rho - 3 \qquad \text{(Lemma 4)}$$

$$\delta(C_n) \geq \delta^2(\delta - 8/\sqrt{d}) \qquad \text{(Lemma 5)}$$

and so that C_n is d^2-agreement testable with soundness

$$\min\left\{\frac{\frac{\phi\delta}{\phi+8} - 8/\sqrt{d}}{4d}, \frac{1}{4(1+2d)}\right\}. \qquad \text{(Theorem 2)}$$

□

References

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

[DEL+21] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. *arXiv preprint arXiv:2111.04808*, 2021.

[DSW06] Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust local testability of tensor products of ldpc codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 304–315. Springer, 2006.

[Gal63] Robert G Gallager. Low density parity check codes, no. 21 in research monograph series, 1963.

[Gol10] Oded Goldreich. Short locally testable codes and proofs: A survey in two parts. In *Property testing*, pages 65–104. Springer, 2010.