Removing The w-Robustness Assumptions.

David Ponarovsky

February 7, 2023

Abstract

We propose a simple alternative construction of good LTC codes. In contrast to previews, constructions made by [Din+22], [LZ22], and [PK21], our construction does not require unique properties of the small codes, such as w-robustness and p-resistance for puncturing.

1 Preambles

Locally Testable Codes, or LTC, are error correction codes such that verifying a uniformly randomly chosen check would be enough to detect any error with probability proportional to its size. Simply put, one can imagine puzzle parts such that any trial to connect pieces in order far from a correct assignment would fail (w.p) at an early step of the process. The analogy for not testability is the case in which the contradiction is observed only in the attempt to putting the last piece.

Besides their clear computational advantage, they are known for their significant roles in the early PCP theorems proofs. And still, the existence of good LTC was considered an open question for decades. Moreover, Sasson proved that codes obtained by the standard randomized constructions could not be LTC [BHR03], which raises the suspicion that maybe codes can not be both good and locally testable. However, recent works by [Din+22], [PK21], and [LZ22] yield a positive answer.

In a nutshell, their sophisticated constructions ensure that no sublinear dependency of restriction exists and yet guarantee that the restrictions are linear far from independent. Namely, no restriction is more important than another, and removing a linear number of constraints would yield the same code.

Their constructions require that the local restrictions, or the local codes, have two properties: the w-robustness and p-resistance for puncturing. Even though they showed probabilistic proof for the existence of an infinite family of such codes, they are more oversized for any practical use. Therefore, we would not formally restate them here; instead, we refer the reader to [LZ22]. Nevertheless, any assumption over the local structure of the code is also an obstacle to encoding a universal computation in the code.

In this work, we propose a new construction for good LTC that demands small codes only to have a large distance. In short, by associating each check with a small code over $2/3$-fraction of the vertex's edges, instead of all of them as in the standard Tanner code, we successfully obtain an LTC with a constant rate. Then by considering graphs, such that both the graph and his subgraph obtained by taking an $\frac{1}{2}$-fraction of the edges of each vertex are good expanders, we also succeed in proving that the codes have linear distance.

Finally, we show how to construct such a graph given a Ramanujan *Cayley* graph. Nevertheless, although we succeeded in simplifying the LTC, we still needed to understand how they can be used to encode a universal computation.

2 Introduction

Coding theory has emerged by the need to transfer information in noisy communication channels. By embedding a message in higher dimension space, one can guarantee robustness against possible faults. The ratio of the original content length to the passed message *length* is the *rate* of the code, and it measures how consuming our communication protocol is. Furthermore, the *distance* of the code quantifies how many faults the scheme can absorb such that the receiver can recover the original message. We could consider the code as all the strings that satisfy a specified restrictions collection.

Non-formally, code is good if its distance and rate are scaled linearly in the encoded message length. In practice, one is also interested in implementing those checks efficiently. We say that a code is an LDPC if any bit is involved in a constant number of restrictions, each of which is a linear equation, and if any restriction contains a fixed number of variables.

Furthermore, finally, another characteristic of the code is its testability, which is the complexity of the number of random checks one should do to negate that a given candidate is in the code. Besides good codes being considered efficient in terms of robustness and overhead, they are also vital components in establishing secure multiparty computation [BGW19] and have a deep connection to probabilistic proofs.

First, we state the notations, definitions, and formal theorem in section 2. Then in sections 3 and 4, we review past results and provide their proofs to make this paper self-contained. Readers familiar with the basic concepts of LDPC, Tanner, and Expanders codes construction should consider skipping directly to section 5, in which we provide our proof.

2.1 Notations, Definitions, And Our Contribution

Here we focus only on linear binary codes, which one could think about as linear subspaces of $\mathbb{F}_2^n$. A common way to measure resilience is to ask how many bits an evil entity needs to flip such that the corrupted vector will be closer to another vector in that space than the original one. Those ideas were formulated by Hamming [Ham50], who presented the following definitions.

Definition 1. *Let $n \in \mathbb{N}$ and $\rho, \delta \in (0,1)$. We say that C is a **binary linear code** with parameters $[n, \rho n, \delta n]$. If C is a subspace of $\mathbb{F}_2^n$, and the dimension of C is at least ρn. In addition, we call the vectors belong to C codewords and define the distance of C to be the minimal number of different bits between any codewords pair of C.*

From now on, we will use the term code to refer to linear binary codes, as we don't deal with any other types of codes. Also, even though it is customary to use the above parameters to analyze codes, we will use their percent forms called the relative distance and the rate of code, matching δ and ρ correspondingly.

Definition 2. *A **family of codes** is an infinite series of codes. Additionally, suppose the rates and relative distances converge into constant values ρ, δ. In that case, we abuse the notation and call that family of codes a code with $[n, \rho n, \delta n]$ for fixed $\rho, \delta \in [0,1)$, and infinite integers $n \in \mathbb{N}$.*

Notice that the above definition contains codes with parameters attending to zero. From a practical view, it means that either we send too many bits, more than a constant amount, on each bit in the original message. Or that for big enough n, adversarial, limited to changing only a constant fraction of the bits, could disrupt the transmission. That distinction raises the definition of good codes.

Definition 3. *We will say that a family of codes is a **good code** if its parameters converge into positive values.*

Apart from distance and rate here, we interest also that the checking process will be robust. In particular, we wish that against significant errors, forgetting to perform a single check will sabotage the computation only with a tiny probability.

Definition 4. *Consider a code C a string x, and denote by $\xi(x)$ the fraction of the checks in which x fails. C will be called a **local-testability** $f(n)$ If there exists $\kappa > 0$ such that*

$$\frac{d(x, C)}{n} \leq \kappa \cdot \xi(x) f(n)$$

Nowadays, we are aware of a wide range of constructions yield good codes, including the expander codes of Sipser and Spilman [SS96] and the LTC codes of Dinur [Din+22], [PK21], [LZ22]. Thus if a decade ago, the main question was the existence of a good code and its construction, now, and particularly in this work, we concentrate on getting a deep understanding of what makes those constructions work. By utilizing those insights, we succussed in achieving significantly simpler construction. Our results:

Theorem (Theorem 1)**.** *There exist a constant $\alpha > 0$ and an infinte family of Tanner Codes $C = \mathcal{T}(G, C_0)$ such that any ireducable codeword x of a coresponding disagreement code $x \in C_\oplus$ at length n, weight at least αn.*

Theorem (Theorem 1+)**.** *There exist a constant $\alpha > 0$ and infinite familiy of codes which satesfies Theroem 1 and also good.*

2.2 Singleton Bound

To get a feeling of the behavior of the distance-rate trade-of, Let us consider the following two codes; each demonstrates a different extreme case. First, define the repetition code $C_r \subset \mathbb{F}_2^{n \cdot r}$, In which, for a fixed integer r, any bit of the original string is duplicated r times. Second, consider the parity check code $C_p \subset \mathbb{F}_2^{n+1}$, in which its codewords are only the vectors with even parity. Let us analyze the repetition code. Clearly, any two n-bits different messages must have at least a single different bit. Therefore their corresponding encoded codewords have to differ in at least r bits. Hence, by scaling r, one could achieve a higher distance as he wishes. Sadly the rate of the code decays as $n/nr = 1/r$. In contrast, the parity check code adds only a single extra bit for the original message. Therefore scaling n gives a family which has a rate attends to $\rho \to 1$. However, flipping any two different bits of a valid codeword is conversing the parity and, as a result, leads to another valid codeword.

To summarize the above, we have that, using a simple construction, one could construct the codes $[r, 1, r]$, $[r, r-1, 2]$. Each has a single perfect parameter, while the other decays to the worst. In the next section, we will review the Singleton bound, which states that for any code (not necessarily good), there must be a zero-sum game between the relative distance and the rate. Now, we are ready to formulate our contribution.

Besides being the first bound, Singleton bound demonstrates how one could get results by using relatively simple elementary arguments. It is also engaging to ask why the proof yields a bound that, empirically, seems far from being tight.

Theorem (Singleton Bound.). *For any linear code with parameter $[n, k, d]$, the following inequality holds:*

$$k + d \leq n + 1$$

Proof. Since any two codewords of C differ by at least d coordinates, we know that by ignoring the first $d - 1$ coordinate of any vector, we obtain a new code with one-to-one corresponding to the original code. In other words, we have found a new code with the same dimension embedded in $\mathbb{F}_2^{n-d+1}$. Combine the fact that dimension is, at most, the dimension of the container space, we get that:

$$\dim C = 2^k \leq 2^{n-d+1} \Rightarrow k + d \leq n + 1$$

□

It is also well known that the only binary codes that reach the bound are: $[n, 1, n]$, $[n, n-1, 2]$,$[n, n, 1]$ [AF22]. In particular, there are no good binary codes that obtain equality. Next, we will review Tanner's construction, that in addition to being a critical element to our proof, also serves as an example of how one can construct a code with arbitrary length and positive rate.

2.3 Tanner Code

The constructions require two main ingredients: a graph Γ, and for simplicity, we will restrict ourselves to a Δ regular graph. Secondly, a small code C_0 at length equals the graph's regularity, namely $C_0 = [\Delta, \rho\Delta, \delta\Delta]$. We can think about any bit string at length E as an assignment over the edges of the graph. Furthermore, for every vertex $v \in \Gamma$, we will call the bit string, which is set on its edges, the local view of v. Then we can define, [Tan81]:

Definition 5. *Let $C = \mathcal{T}(\Gamma, C_0)$ be all the codewords which, for any vertex $v \in \Gamma$, the local view of v is a codeword of C_0. We say that C is a **Tanner code** of Γ, C_0. Notice that if C_0 is a binary linear code, So C is.*

It's also worth mentioning that the first construction of good classical codes, due to Sipser and Shpilman, are Tanner codes over expanders graphs [SS96].

Theorem. *Tanner codes have a rate of at least $2\rho - 1$.*

Proof. The dimension of the subspace is bounded by the dimension of the container minus the number of restrictions. So assuming non-degeneration of the small code restrictions, we have that any vertex count exactly $(1-\rho)\Delta$ restrictions. Hence,

$$\dim C \geq \frac{1}{2}n\Delta - (1-\rho)\Delta n = \frac{1}{2}n\Delta(2\rho - 1)$$

Clearly, any small code with rate $> \frac{1}{2}$ will yield a code with an asymptotically positive rate □

2.4 Expander Codes

We saw how a graph could give us arbitrarily long codes with a positive rate. We will show, Sipser's result that if the graph is also an expander, we can guarantee a positive relative distance. We notice that the name expander codes is coined for a more general version than the one we will present.

Definition 6. *Denote by λ the second eigenvalue of the adjacency matrix of the Δ-regular graph. For our uses, it will be satisfied to define expander as a graph $G = (V, E)$ such that for any two subsets of vertices $T, S \subset V$, the number of edges between S and T is at most:*

$$\mid E(S,T) - \frac{\Delta}{n}|S||T| \mid \leq \lambda\sqrt{|S||T|}$$

This bound is known as the Expander Mixining Lemma. We refer the reader to [HLW06] for more deatilied survery.

Theorem. *Theorem, let C be the Tanner Code defined by the small code $C_0 = [\Delta, \delta\Delta, \rho\Delta]$ such that $\rho \geq \frac{1}{2}$ and the expander graph G such that $\delta\Delta \geq \lambda$. C is a good LDPC code.*

Proof. We have already shown that the graph has a positive rate due to the Tunner construction. So it's left to show also the code has a linear distance. Fix a codeword $x \in C$ and denote By S the support of x over the edges. Namely, a vertex $v \in V$ belongs to S if it connects to nonzero edges regarding the assignment by x, Assume towards contradiction that $|x| = o(n)$. And notice that $|S|$ is at most $2|x|$, Then by The Expander Mixining Lemma we have that:

$$\begin{aligned}\frac{E(S,S)}{|S|} &\leq \frac{\Delta}{n}|S| + \lambda \\ &\leq_{n\to\infty} o(1) + \lambda\end{aligned}$$

Namely, for any such sublinear weight string, x, the average of nontrivial edges for the vertex is less than λ. So there must be at least one vertex $v \in S$ that, on his local view, sets a string at a weight less than λ. By the definition of S, this string cannot be trivial. Combining the fact that any nontrivial codeword of the C_0 is at weight at least $\delta\Delta$, we get a contradiction to the assumption that v is satisfied, videlicet, x can't be a codeword □

2.5 Tanner testability.

This subsection will explain why testability is so hard to achieve. Let C be a good Tanner expander code as defined above. And consider an arbitrary vertex $u \in V$ and arbitrary restriction of C_0, h. Now define $\tilde{C}$ as the code obtained by requiring all the restrictions of C except h on u. That it, u is satisfied if his local view satisfies all the C_0 restrictions apart from h. Also, for convenience, denote the small code u enforces on his local view by C_0^u. Let us assume that the distance of C_0^u is at least $\delta\Delta$. Then, by repeating almost exactly the steps above with caution, one could prove that $\tilde{C}$ also has a linear distance.

Assume that $\tilde{C} \neq C \Rightarrow$ there exists $x \in \tilde{C}/C$. By definition, for any $v \in V/\{u\}$ it holdes that $x|_v \in C_0$. Hence, the assumption that $x \notin C$ implies $x|_u \notin C_0$ So, clearly, x fails at a constant number of C's checks. On the other hand, the closest codeword $y \in C$ to x is also a codeword of $\tilde{C}$ as $y|_v \in C_0$ for every $v \in V$. Hence:

$$d(x,C) = d(x,y) \geq d\left(\tilde{C}\right) = \Theta(n)$$

Even if a linear number of bits needed to be flipped to correct x, only a single check observes that x is indeed an error.

3 Construction

3.1 Almost LTC With Zero Rate

Definition 7 (The Disagreement Code). *Given a Tanner code $C = \mathcal{T}(G, C_0)$, define the code $C_\oplus$ to contain all the words equal to the formal summation $\sum_{v\in V(G)} c_v$ when c_v is an assignment of a codeword $c_v \in C_0$ on the edges of the vertex $v \in V(G)$. We call to such code the **disagreement code** of C, as edges are set to 1 only if their connected vertices contribute to the summation codewords that are different on the corresponding bit to that edge. In addition, we will call to any contribute c_v, the **suggestion** of v. And notice that by linearity, each vertex suggests, at most, a single suggestion.*

Finally, given a bits assessment $x \in \mathbb{F}_2^E$ over the edges of G, we will denote by $x^\oplus \in C_\oplus$ the codeword which obtained by summing up suggestions set such each vertex suggests the closet codeword to his local view. Namely, for each $v \in V$ define:

$$\begin{aligned}c_v &\leftarrow \arg_{\tilde{c}\in C_0}\min d(x|_v, \tilde{c}) \quad \forall v \in V \\ x^\oplus &\leftarrow \sum_{v\in V} c_v\end{aligned}$$

We will think about $x^\oplus$ as the disagreement between the vertices over x.

Lemma 1 (Linearity Of The Disagreement). *Consider the code $C = \mathcal{T}(G, C_0)$. Let $x \in \mathbb{F}_2^E$ then for any $y \in C$ it holds that:*

$$(x+y)^\oplus = (x)^\oplus$$

Proof. Having that $y \in C$ followes $y|_v \in C_0$ and therefore $\arg_{\tilde{c}\in C_0}\min d(z,\tilde{c}) = y|_v + \arg_{\tilde{c}\in C_0}\min d(z, \tilde{c} + y|_v)$, Hence the suggestion made by vertrx v is:

$$\begin{aligned}c_v &\leftarrow \arg_{\tilde{c}\in C_0}\min d((x+y)|_v, \tilde{c}) \\ &\leftarrow y|_v + \arg_{\tilde{c}\in C_0}\min d((x+y)|_v, \tilde{c} + y|_v) \\ &\leftarrow y|_v + \arg_{\tilde{c}\in C_0}\min d(x|_v, \tilde{c})\end{aligned}$$

It follows that:

$$(x+y)^{\oplus} = \sum_{v\in V} c_v = \sum_{v\in V} y|_v + \sum_{v\in V} \arg_{\tilde{c}\in C_0} \min d(x|_v, \tilde{c})$$
$$= y^{\oplus} + x^{\oplus} = x^{\oplus}$$

When the last transition follows immediately by the fact that $y \in C$ and therefore any pair of connected vertices contribute the same value for their associated edge ☐

Definition 8. *Let $C = \mathcal{T}(G, C_0)$. We say that $x \in C_{\oplus}$ is **reducable** if there exists a vertex v and a small codeword c_v, for which, adding the assignment of c_v over the v's edges to x decreases the weight. Namely, $|x + c_v| < |x|$. If $x \in C_{\oplus}$ is not a reducable codeword then we say that x is **ireducable** .*

Theorem (Theorem 1)**.** *There exist a constant $\alpha > 0$ and an infinte family of Tanner Codes $C = \mathcal{T}(G, C_0)$ such that any ireducable [8] codeword x of a coresponding disagreement code $x \in C_{\oplus}$ at length n, weight at least αn.*

Proof. By induction over the number of vertices $V' \subset V$, which suggest a nontrivial codeword to x. Base, assume that a single vertex $v \in V$ suggests a nontrivial codeword $c_v \in C_0$. Then it's clear that $x = c_v$. And therefore, we have that $|x + c_v| = 0 < |x|$.

Assume the correctness of the argument for every codeword defined by at most m nontrivial suggestions made by $V' \subset V$. And consider the graph (V', E') induced by them. If the graph has more than a single connectivity component, then any of them is also a codeword of $C_{\oplus}$ but composed of at most $m-1$ nontrivial suggestions. Therefore, by the assumption, we could find a vertex v and a proper small codeword $c_v \in C_0$, such that the addition of the suggestion will decrease the weight of the codeword defined on that component and therefore decrease the total weight of x.

So, we can assume that the vertices in V' compose a single connectivity component. Let be $x|_v \in \mathbb{F}_2^{\Delta}$ the bits of x on the indices corresponding to v's edges. For any $S \subset E$, define $w_S(x)$ as the weight that x induces over S. Sometimes we will refer to $w_S(x)$ as the **flux** induced by x over S.

The genreal idea of the proof is to show that if the distance of the small code is large ($\geq \frac{2}{3}$) and x is ireducable [8] codeword then there exist an indepandent subset of vertices $U \subset V'$, at linear size, that induce a significant flux over E/E'. If U has linear size than also x has a linear size, And if not, Then we will show that no serious interface has been occurred. claim 1 and claim 2 state that if one is willing to hide an ireducable [8] error then he has to touch at least a linar number of verties. claim 3 and claim 5 quanitify the flux that induced by such errors.

Claim 1. *For any $v \in V'$ and corresponded suggestion c_v it holds that: $w_{E'}(c_v) \geq \frac{1}{2}\delta_0\Delta$.*

Proof. Notice that any edge of E connected only to a single vertex in V' equals the corresponding bit in the original suggestion made by c_v. Hence for every $v \in V'$, it holds that:

$$w_{E/E'}(x|_v) = w_{E/E'}(c_v) \Rightarrow w_{E/E'}(x|_v) \leq |x \cap c_v|$$

Now consider the weight of $x + c_v$, By the assumption that x is ireducable code word of $c_{\oplus}$ we have that:

$$|x + c_v| = |x| + |c_v| - 2|x \cap c_v| > |x|$$
$$\Rightarrow |x \cap c_v| < \frac{1}{2}|c_v|$$
$$w_{E'}(c_v) = |c_v| - w_{E/E'}(c_v) = |c_v| - w_{E/E'}(x|_v)$$
$$\geq |c_v| - |x \cap c_v| \geq \frac{1}{2}|c_v| = \frac{1}{2}\delta_0\Delta$$

☐

Consider an arbitrary vertex $r \in V'$, and consider the DAG obtained by the BFS walk over the subgraph (V', E') starting at r. Denote this directed tree by T.

Claim 2. *The size of T is at least:*

$$|T| \geq \left(\frac{1}{4}\delta_0 - \frac{\lambda}{\Delta}\right) n$$

Proof. By claim 1 any $v \in T$ the degree of v is at least $\frac{1}{2}\delta_0\Delta$ we have that: $E(T,T) \geq \frac{1}{2} \cdot \frac{1}{2}\delta_0\Delta|T|$. Combine the Mixining Expander Lemma we obtain:

$$\begin{aligned}
\frac{1}{4}\delta_0\Delta|T| \leq & \frac{\Delta}{n}|T|^2 + \lambda|T| \\
\Rightarrow & \left(\frac{\Delta}{n}|T| + \lambda - \frac{1}{4}\delta_0\Delta\right)|T| \geq 0 \\
\Rightarrow & |T| \geq \left(\frac{1}{4}\delta_0 - \frac{\lambda}{\Delta}\right) n
\end{aligned}$$

□

Claim 3. *Suppose that G is an expander graph with a second eigenvalue λ, then For any layer U there exist a layer U' such that:*

$$\begin{aligned}
&(1) \quad |U'| \geq |U| \\
&(2) \quad w_{E/E'}\left(x|_{U'}\right) \geq \Delta|U'|\left(\delta_0 - \frac{2}{3} - \frac{2\lambda}{\Delta}\right)
\end{aligned}$$

Proof. Consider layer U and denote by U_{-1} and U_{+1} the preceding and the following layers to U in T. It follows from the expander mixing lemma that:

$$\begin{aligned}
w_{E/E'}\left(x|_{U}\right) \geq \delta_0\Delta|U| - w\left(E(U_{-1}\bigcup U_{+1}, U)\right) \geq \\
\delta_0\Delta|U| - E(U_{-1}\bigcup U_{+1}, U) \\
\delta_0\Delta|U| - \Delta\frac{|U||U_{-1}|}{n} - \Delta\frac{|U||U_{+1}|}{n} \\
- \lambda\sqrt{|U||U_{-1}|} - \lambda\sqrt{|U||U_{+1}|}
\end{aligned}$$

Claim 4. *We can assume that $|U| \geq |U_{-1}|, |U_{+1}|$.*

Proof. Suppose that $|U_{+1}| > |U|$, so we could choose U to be U_{+1}. Continuing stepping deeper till we have that $|U| > |U_{+1}|, |U_{-1}|$. Simiraly, if $|U| > |U_{+1}|$ but $|U_{-1}| > |U|$, the we could take steps upward by replacing U_{-1} with U. At the end of the process, we will be left with U at a size greater than the initial layer and $|U| > |U_{+1}|, |U_{-1}|$ □

Using claim 4, we have that $\left(|U_{+1}| + |U_{-1}|\right)/n < \frac{2}{3}$ and therefore:

$$w_{E/E'}\left(x|_{U}\right) \geq \left(\delta_0 - \frac{2}{3} - \frac{2\lambda}{\Delta}\right)\Delta|U|$$

□

That immediately yields the following: let $U_{\max} = \arg\max_{U \text{ layer in } T} |U|$ then:

$$|x| \geq w_{E/E'}\left(x|_{U_{\max}}\right) \geq \left(\delta_0 - \frac{2}{3} - \frac{2\lambda}{\Delta}\right)\Delta|U_{\max}|$$

Claim 5. *Consider again the maximal layer $U_{\max}$ then:*

$$w_{E/E'}(x) \geq \left(\delta_0 - \frac{|U_{\max}|}{n} - \frac{\lambda}{\Delta}\right)\Delta|T|$$

Proof. Similarly to above, now we will bound the flux that all the nodes in T induce over E/E'. Denote by $U_0, U_1..U_m$ the layers of T ordered corresponded to their height, thus we obtain:

$$\begin{aligned}
w_{E/E'}(x) \geq & \delta_0\Delta|T| - \sum_{i\in[m]} w\left(E\left(U_i, U_{i+1}\right)\right) \\
\geq & \delta_0\Delta|T| - \sum_{i\in[m]} E\left(U_i, U_{i+1}\right) \\
\geq & \delta_0\Delta|T| - \sum_{i\in[m]} \frac{\Delta}{n}|U_i||U_{i+1}| + \lambda\sqrt{|U_i||U_{i+1}|} \\
\geq & \delta_0\Delta|T| - \sum_{i\in[m]} \frac{\Delta}{n}|U_i||U_{i+1}| + \lambda\frac{|U_i| + |U_{i+1}|}{2} \\
\geq & \delta_0\Delta|T| - \frac{\Delta}{n}|T||U_{\max}| - \lambda|T| \\
\geq & \left(\delta_0 - \frac{|U_{\max}|}{n} - \frac{\lambda}{\Delta}\right)\Delta|T|
\end{aligned}$$

□

Claim 6. *Alternate proof of fulx inequality, which doesn't assume that there is no interference inside the layers.* $w(E(U,U)) > 0$.

Proof. Separeate into the following cases, First assume that $U_{\max}/n > \frac{1}{3}$ then we have that the total interference with $U_{\max}$ layers is at most:

$$\frac{\Delta|U_{\max}|(n-|U_{\max}|)}{n} + \lambda\sqrt{U_{\max}n} \leq \left(1 - \frac{U_{\max}}{n} + \sqrt{3}\frac{\lambda}{\Delta}\right)\Delta|U_{\max}|$$

□

Proof of Theorem 1. Consider the size of the maxiaml layer $|U_{\max}|$ and sepearte to the following two cases. First, consider the case that $|U_{\max}| \geq \alpha n$ in that case it follows immedily by claim 3 that if $\delta_0 > \frac{2}{3} - \frac{2\lambda}{\Delta}$ there exists $\alpha' > 0$ such that:

$$|x| \geq \left(\delta_0 - \frac{2}{3} - \frac{2}{\lambda}\Delta\right)\Delta|U_{\max}| \geq \alpha' n$$

So, it is lefts to consider the second case in which $|U_{\max}| < \alpha n$ in that case, we have from claim 5 inequality that:

$$\begin{aligned}|x| \geq w_{E/E'}(x) &\geq \left(\delta_0 - \frac{|U_{\max}|}{n} - \frac{\lambda}{\Delta}\right)\Delta|T| \\ &\geq \left(\delta_0 - \alpha - \frac{\lambda}{\Delta}\right)\Delta|T|\end{aligned}$$

Setting $\alpha \geq \frac{2}{3}$ we complete the proof □

Unfortunately, Singelton bound doesn't allow both $\delta_0 > \frac{2}{3}$ and $\rho_0 \geq \frac{1}{2}$, so in total, we prove the existence of code LDPC code which is good in terms of testability and distance yet has a zero rate. In the following subsection, we will prove that one can overcome this problem, by considering a variton of Tanner code, in which every vertex cheks only a $\frac{2}{3}$ fraction of the edges in his support.

3.2 Overcoming The Vanishing Rate.

Consider the following code; instead of associating each edge with pair of checks, let's define the vertices to be the checks of small codes over $q \in [0,1]$ fraction of their edges. That is, now each vertex defines only $(1-\rho_0)q\Delta$ restrictions. Hence, the rate of the code is at least:

$$\begin{aligned}\rho\frac{1}{2}\Delta n &\geq \frac{1}{2}\Delta n - (1-\rho_0)q\Delta n \\ \Rightarrow \rho &\geq \left(2\rho_0 + \left(\frac{1}{q} - 2\right)\right)q \\ \rho_0 &\geq 1 - \frac{1}{2q}\end{aligned}$$

for example, if $q = 2/3$, then for having constant rate, it is enough to ensure that $\rho_0 \geq 1 - \frac{3}{4} = \frac{1}{4}$.

Intuition For Testability. Before expand the construction let's us justifiy why one should even expects that removing constrainsts preserves testability. Assume that is gurnted that the lower bound of the flux on the trivial vertices remains up to multiplication by the fraction factor q, or put it diffrently, one could just stick q in every inequalitiy without lose correcntess, Then:

$$\begin{aligned}w_{E/E'}(x|_U) &\geq \delta_0 q\Delta|U| - qw\left(E(U_{-1}\bigcup U_{+1}, U)\right) \\ \Rightarrow |x| &\geq \left(\delta_0 - \frac{2}{3} - \frac{2\lambda}{\Delta}\right)q\Delta|U_{\max}|\end{aligned}$$

As you can see, ireducable [8] words of the disagreement have a linear weight, dispite that the orignal code has non-vanish rate.

Yet, We still require more to prove a linear distance. By repeating on the *Singleton Bound 2.2* proof it follows that the small code $\tilde{C}_0$ obtained by ignoring arbitrary $\left(q - \frac{1}{2}\right)\Delta$ coordinates yield a code with distance:

$$\left(\delta_0 - \left(q - \frac{1}{2}\right)\right)\Delta$$

So assume that we could engineer an expander family such that the graphs obtained by removing $\frac{1}{2}$ of the edges connected for each vertex result also expanders, and in addition, regarding $\tilde{C}_0$ each edge is checked by both vertices on its support. Namely, a good Tanners Code could be defined on the restricted graphs; Then, any string that satisfies the original checks also has a linear weight. To achieve this property, we will restrict ourselves to a particular family of Cayley Graphs.

Theorem (Theorem 1+)**.** *There exist a constant $\alpha > 0$ and infinite familiy of codes which satesfies Theroem 1 and also good.*

Definition 9 (Testability Tanner Code)**.** *Let $q > \frac{1}{2}$ and let J be a geneator set for group Γ such that $|J| = \Delta$, $q|\Delta$, J closed for inverse, and there exist subset of J, denote it by, J' such that J' is a generator set of Γ and $|J'| = \frac{1}{2}\Delta$. Let C_0 be a code with parameters $C_0 = q\Delta[1, \rho_0, \delta_0]$. For any vertex assoicate a subset $\bar{J}_v \subset J/J'$ at size:*

$$|\bar{J}_v| = \left(q - \frac{1}{2}\right)\Delta \Rightarrow |\bar{J}_v \cup J'| = q\Delta$$

Define the code $\mathcal{T}(J, q, C_0)$ to be the subspace such that any vertex's local view over the edges defined by $\bar{J}_v \cup J'$ is a codeword of C_0. In addition, let's associate a code $\tilde{C}_v$ obtained for any vertex by ignoring the bits supported on the $\bar{J}_v$ coordinates. Notice that code defined by requiring that the local view of any vertex v of Cayley(Γ, J') *is a codeword of $\tilde{C}_v$ is a TannerCode. Denote it by $\tilde{\mathcal{T}}(J, q, C_0)$.*

Claim 7. *Let J be defined as above such that both* Cayley(Γ, J), Cayley(Γ, J') *are expanders with algebric expansion greater then λ and C_0 with the parameters $\rho_0 > 1 - \frac{1}{2q}$ and $\delta_0 - \left(q - \frac{1}{2}\right) > 2\lambda/\Delta$. Then the code $\mathcal{T}(J, q, C_0)$ is a good code.*

Proof. We already proved that the code has a positive rate, So it left to show a constant relative distance.

Consider a codeword x and denote by x' the restriction of x to *Cayley* (Γ, J') which is a codeword of $\tilde{C} = \tilde{\mathcal{T}}(J, q, C_0)$. But $\tilde{C}$ is a Tanner Code such that any vertex sees at least $\tilde{\delta_0}\Delta := \left(\delta_0 - \left(q - \frac{1}{2}\right)\right)\Delta$ nontrivial bits. Denote by S the vertics subset supports x', and by $E(S, S)$ the edges from S to itself, and by using the fact that *Cayley*(Γ, J') is an expander with seconed eigenvalue at most δ we have that:

$$\frac{|x'|}{|S|} \geq \tilde{\delta}_0\Delta \Rightarrow |S| \geq \left(\tilde{\delta} - \frac{2\lambda}{\Delta}\right)\Delta n$$

By the assumption that $\tilde{\delta} > 2\lambda/\Delta$ we have that S must has liner size, and therefore $|x'|$ also must to be linear in n. Finally as $x' \subset x$ we obtain the correctness of the claim. □

Claim 8 (Existence of such *Cayley's*)**.** *Let S be a generator set such that* Cayley(Γ, S) *has a second largest eigenvalue greater then λ, And consider an arbitray group element $g \in \Gamma$ and denote by S_g the set gSg^{-1}. Then the second eigenvalue of the graph obtained by $(\Gamma, S) \cup (\Gamma, S)$ is at most 2λ.*

Proof. Denote by G, G' the *Cayley* graphs coressponding to S, S_g, for conviniet we will use the notation of $\sum_{v \sim_G u}$ to denote a summation over all the neighboors of v in the graph G. Let $A_{G'}$ be the adjacency matrix of G'. Recall that G' is a Δ regular graph, and therefore the uniform distribution $\mathbf{1}$ is the eigenstate with the maximal eigenvalue, and the second eigenvalue is given by the min-max principle:

$$\begin{aligned}
\max_{f \perp \mathbf{1}} \frac{f^\top A_{G'} f}{f^\top f} &= \max_{f \perp \mathbf{1}} \sum_v \sum_{u \sim_{G'} v} \frac{f(u) f(v)}{f^\top f} \\
&= \max_{f \perp \mathbf{1}} \sum_v \sum_{\tau \in S} \frac{f\left(g\tau g^{-1} v\right) f(v)}{f^\top f} \\
&= \max_{f \perp \mathbf{1}} \sum_{gv} \sum_{\tau \in S} \frac{f\left(g\tau g^{-1} gv\right) f(gv)}{f^\top f} \\
&= \max_{f \perp \mathbf{1}} \sum_{gv} \sum_{\tau \in S} \frac{f(g\tau v) f(gv)}{f^\top f} \\
&= \max_{f \perp \mathbf{1}} \sum_{gv} \sum_{u \sim_G v} \frac{f(gu) f(gv)}{f^\top f}
\end{aligned}$$

As for any function $f : V \to \mathbb{R}$ one could define a function $f' : E \to \mathbb{R}$ such that $f'(v) = f(v)$ and f' preservs the norm:

$$\begin{aligned}
f'^\top f' &= \sum_{v \in V} f'(v) f'(v) = \sum_{v \in V} f^\top(vg) f(vg) = f^\top f \\
\Rightarrow \max_{f \perp \mathbf{1}} \frac{f^\top A_{G'} f}{f^\top f} &= \max_{f \perp \mathbf{1}} \sum_{gv} \sum_{u \sim_G v} \frac{f(gu) f(vg)}{f^\top f}
\end{aligned}$$

By the Interlacing Theorem, [Hae95] the second eigenvalue of any subgraph of G' is less than the λ', In particular, the eigenvalue of the graph obtained by taking the edges that are associated with elements of the S_g/S.

Denote that subgraph by $G'_{/S}$. Because $S_g/S \cap S = \emptyset$, we have that the edges sets of G, G' are disjointness sets. Hence the adjacency matrix of the graphs union equals the sum of their adjacency matrices. So in total, we obtain that:

$$\begin{aligned}\lambda' &= \max_{f \perp \mathbf{1}} \frac{f^\top \left(A_G + A_{G'_{/S}}\right) f}{f^\top f} \\ &\leq \max_{f \perp \mathbf{1}} \frac{f^\top A_G f}{f^\top f} + \max_{f \perp \mathbf{1}} \frac{f^\top A_{G'_{/S}} f}{f^\top f} \\ &\leq \lambda + \lambda = 2\lambda\end{aligned}$$

□

Claim 9. *If Δ is a constant greater than two, and G is a λ-algebric expander with girth at length $\Omega(\log n)$, then there exists a $g \in \Gamma$ such that $S_g \cap S = \emptyset$.*

Proof. As $\Delta > 2$ there must be at least two different elements $s_1, s_2 \in S$ such that $s_1 \neq s_2, s_2^{-1}$. Pick $g = s_1 s_2$. Now assume through contradiction that there exists a pair $s, r \in S$ such that $gsg^{-1} = r \Rightarrow gs = rg$ and notice that the fact that $s_1 \neq s_2^{-1}$ guarantees that both terms are a product of 3 element group.

Therefore either that there is a 6-length cycle in the graph, Or that there is element-wise equivalence, namely $s_1 = r, s_2 = s_1, s = s_2$. The first case contradict the lower bound on the expander girth, which is at least $\Omega(\log_\Delta(n))$, while the other stand in contradiction to the fact that $s_1 \neq s_2$ □

Remark. Regarding Quantum Codes. Notice that any complex designed to hold CSS qLDPC codes must have constant length cycles. Otherwise, the distance of C_x will not be constant, and therefore the condition $H_x H_z^\top = 0$ could be satisfied only if H_z is not a constant row-weight matrix, Put differently C_z is not an LDPC code. Consequently, any trial to generalize the construction for obtaining quantum codes must not rely on claim 9.

Remark. Note On Random Construction. One might wondring if using *Cayley* is necssery. We conjecure that there is a constant $c > 0$ such that sampling pair of $(1+c)\frac{1}{2}\Delta$ regulr random graphs, and than take the anti-symatry union of them might also obtain a good expander such that each of the reseuide part also has good expansion with heigh probability.

Claim 10. *Consider the graph G and the code C as defind in [9] and let S, T be a pair od disjointness vertices subsets. And let x_S and x_T codewords of the $C_\oplus$ such that x_S suggested only by vertices in S, and in similiar manner x_T suggested only by T's vertices. Then the flux of S over T is at most:*

$$w_T(x_S) \leq E_{G'}(S, T) \leq \frac{1}{2}\Delta\frac{|S||T|}{n} + \lambda\sqrt{|S||T|}$$

Proof. The only edges that can interference are the edge defined by J', Namly the edges which belong to *Cayley*(Γ, J'). Therefore it's enough to use the mixing expander lemme on the $\frac{1}{2}\Delta$-regular graph. □

Proof of Theorem 1+. Noitce that $\frac{1}{2} < \frac{2}{3} = q$, Thus reapting exactly over proof above obtains that:

$$\begin{aligned}w_{E/E'}(x|_U) &\geq \delta_0 q\Delta|U| - w\left(E(U_{-1}\bigcup U_{+1}, U)\right) \\ &\geq \delta_0 q\Delta|U| - \frac{\Delta}{2}\frac{|U|\left(|U_{-1}| + |U_{+1}|\right)}{n} - \lambda\sqrt{|U|\left(|U_{-1}| + |U_{+1}|\right)} \\ &\geq \left(\delta_0 - \frac{2}{3} - \frac{1}{q}\frac{\lambda}{\Delta}\right) q\Delta|U|\end{aligned}$$

When the last inequalitiy followes form the fact that the proof of claim 4 doesn't relay on graph structure arguments. The same arguments leads also to anlog inequalitiy for claim 5.

Choseing J such that *Cayley*(Γ, J) is ramnujan provid that $\frac{2\lambda}{\Delta q}$ sacle as $\Theta\left(\frac{1}{\sqrt{\Delta}}\right)$. That close the case in which there is a linear size layer of nontrival suggestions. In other case, in which any such layer is at size less than $\alpha' n$ ($\alpha' = \left(\delta_0 - \left(q - \frac{1}{2}\right)\right)$?) then we obtain the testbility for free □

4 Good Quantum Codes, logaritmic-check-weight.

In the following section we will construct a family of complexes on which we will define a pairs of Tanner Codes, evently, they will used to compose a CSS pairs of good quantum codes.

Inifinte Family Of Tanner Quantum Codes. Let p be a prime and $\delta \in (0,1)$. Consider the Cayly graphs obtained by taking uniformly a $c(\delta)\log n$ generators of the cyclic group at order p, denote that set by S. It was shown by N.Alon **[COMMENT]** cite Noga that with high probability that process yield a Graph with δ-algebric expansion. Now, consider the double cover of that graph and denote it by $G = \left(V = V^{+} \cup V^{-}, E\right)$. And define the folowing graph denoted by $\Gamma^{\pm} = \left(V^{\pm}, E'\right)$:

$$((u,\pm),(v,\pm)) \in E' \Leftrightarrow \exists a \neq b \in S \ s.t \ abu = v$$

5 Decodeing and Testing

For completeness, we show exactly how Theorem 1 implies testability. The following section repeats Leiverar's and Zemor's proof [LZ22]. Consider a binary string x that is not a codeword. The main idea is the observation that the number of bits filliped by (any) decoder, while decoding x, bounds the distance $d(x,C)$ from above. In addition, the number of positive checks in the first iteration is exactly the number of violated restrictions.

Definition 10. *Let $L = \{L_i\}_0^{2|E|}$ be a series of $2|E|$. Such that for each vertex $v \in V$ $\sum_{e=\{u,v\}} L_{e_v} \in C_0$. We will call L a Potential list and refer to the e_v'the element of L as a suggestion made by the vertex $v \in V$ for the edge $e \in E$. Sometimes we will use the notation L_v to denote all the L's coordinates of the form $L_{e_v} \forall e \in Support(v)$. Define the Force of L to be the following sum $F(L) = \sum_{e=\{v,u\}\in E} (L_{e_v} + L_{e_u})$ and notice that $F(L) \in C_{\oplus}$. And define the state $S(L) \subset \mathbb{F}_2^{|E|}$ of L as the vector obtained by choosing an arbitrary value from $\{L_{e_v}, L_{e_u}\}$ for each edge $e \in E$.*

Claim 11. *Let L be the Potential list. If $F(L) = 0$ then $S(L) \in C$.*

Proof. Denote by $\phi(e) \subset \{L_{e_v}, L_{e_u}\}$ the value which was chosen to $e = \{v,u\} \in E$. By $F(L) = 0$, it follows that $L_{e_v} + L_{e_u} = 0 \Rightarrow L_{e_v} = L_{e_u} = \phi(e)$ for any $e \in E$. Hence for every $v \in V$ we have that $S(L)|_v = \sum_{u\sim v} \phi(\{v,u\}) = \sum_{u\sim v} L_{e_v} \in C_0 \Rightarrow S(L) \in C$ □

The decoding goes as follows. First, each vertex suggests the closet C_0's codeword to his local view. Those suggestions define a Potential list, denote it by L, then if $F(L) < \tau$, by Theorem 1, one could find a suggestion of vertex v and a codeword c_v such that updating the value of $L_v \leftarrow L_v + c_v$ yields a Potential list with lower force. Therefore repeating the process till the force vanishes, obtain a Potential list in which its state is a codeword.

Definition 11. *Let $\tau > 0, f : \mathbb{N} \to \mathbb{R}^{+}$, and consider a Tanner Code $C = \mathcal{T}(G, C_0)$. Let us Define the following decoder and denote it by $\mathcal{D}$.*

Algorithm 1: Decoding

Data: $x \in \mathbb{F}_2^n$
Result: $\arg\min\{y \in C : |y + x|\}$ if $d(y,C) < \tau$ and False otherwise.

$L \leftarrow$ Array{}
for $v \in V$ **do**
  $c'_v \leftarrow \arg\min\{y \in C_0 : |y + x|_v|\}$
  $L_v \leftarrow c'_v$
end
$z \leftarrow \sum_{v\in V} c'_v$
if $|z| < \tau\frac{n}{f(n)}$ **then**
  while $|z| > 0$ **do**
    find v and $c \in C_0$ such that $|z + c_v| < |z|$
    $z \leftarrow z + c_v$
    $L_v \leftarrow L_v + c_v$
  end
else
  reject.
end
return $S(L)$

Theorem 1. *Consider a Tanner Code $C = [n, n\rho, n\delta]$ and the disagreement code $C_{\oplus}$ defined by it. Suppose that for every codeword $z \in C_{\oplus}$ in $C_{\oplus}$ such that $|z| < \tau' n / f(n)$, there exists another codeword $y \in C_{\oplus}$ such that $|y| < |z|$, set $\tau \leftarrow \frac{\tau'}{6\Delta}\delta$ then,*

1. *$\mathcal{D}$ corrects any error at a weight less than $\tau n / f(n)$.*
2. *C is $f(n)$ testable code.*

Proof. So it is clear from the claim claim 11 above that if the condition at line (6) is satisfied, then $\mathcal{D}$ will converge into some codeword in C. Hence, to complete the first section, it left to show that $\mathcal{D}$ returns the closest codeword. Denote by e the error, and by simple counting arguments; we have that $\mathcal{D}$ flips at most:

$$d_{\mathcal{D}}(x, C) \leq 2|e|\Delta + \tau \frac{n}{f(n)}\Delta$$

bits. Hence, by the assumption,

$$d_{\mathcal{D}}(x, C) \leq 3\Delta\tau \frac{n}{f(n)} \leq 3\Delta\tau\delta n < \frac{1}{2}\delta n$$

Therefore the code word returned by $\mathcal{D}$ must be the closet. Otherwise, it contradicts the fact that the relative distance of the code is δ. To obtain the correctness of the second section, we will separate when the conditional at the line (5) holds and not. And prove that the testability inequality holds in both cases. Let $x \in \mathbb{F}_2^n$ and consider the running of $\mathcal{D}$ over x. Assume the first case, in which the conditional at line (5) is satisfied. In that case, $\mathcal{D}$ decodes x into its closest codeword in C. Therefore:

$$\begin{aligned} d(x, C) &\leq d_D(x, C) \leq m\xi(x)\Delta + |z|\Delta \\ &\leq m\xi(x)\Delta + m\xi(x)\Delta^2 \\ \frac{d(x, C)}{n} &\leq \kappa_1\xi(x) \end{aligned}$$

Now, consider the other case in which: $|z| \geq \tau\frac{n}{f(n)}$.

$$\begin{aligned} \frac{d(x, C)}{n} &\leq 1 \leq \frac{|z|}{\tau n} f(n) \leq \frac{m}{n}\frac{1}{\tau}\Delta\xi(x) f(n) \\ &\leq \kappa_2\xi(x) f(n) \end{aligned}$$

Picking $\kappa \leftarrow \max\{\kappa_1, \kappa_2\}$ proves $f(n)$-testability $\square$

References

[Ham50] R. W. Hamming. "Error detecting and error correcting codes". In: *The Bell System Technical Journal* 29.2 (1950), pp. 147–160. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

[Tan81] R. Tanner. "A recursive approach to low complexity codes". In: *IEEE Transactions on Information Theory* 27.5 (1981), pp. 533–547. DOI: 10.1109/TIT.1981.1056404.

[Hae95] Willem H. Haemers. "Interlacing eigenvalues and graphs". In: *Linear Algebra and its Applications* 226-228 (1995). Honoring J.J.Seidel, pp. 593–616. ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(95)00199-2. URL: https://www.sciencedirect.com/science/article/pii/0024379595001992.

[SS96] M. Sipser and D.A. Spielman. "Expander codes". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1710–1722. DOI: 10.1109/18.556667.

[BHR03] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. "Some 3CNF Properties Are Hard to Test". In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 345–354. ISBN: 1581136749. DOI: 10.1145/780542.780594. URL: https://doi.org/10.1145/780542.780594.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561.

[BGW19] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. "Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation". In: *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 351–371. ISBN: 9781450372664. URL: https://doi.org/10.1145/3335741.3335756.

[PK21] Pavel Panteleev and Gleb Kalachev. *Asymptotically Good Quantum and Locally Testable Classical LDPC Codes*. 2021. DOI: 10.48550/ARXIV.2111.03654. URL: https://arxiv.org/abs/2111.03654.

[AF22] "Maximum distance separable (MDS) code". In: *The Error Correction Zoo*. Ed. by Victor V. Albert and Philippe Faist. 2022. URL: https://errorcorrectionzoo.org/c/mds.

[Din+22] Irit Dinur et al. *Good Locally Testable Codes*. 2022. DOI: 10.48550/ARXIV.2207.11929. URL: https://arxiv.org/abs/2207.11929.

[LZ22] Anthony Leverrier and Gilles Zémor. *Quantum Tanner codes*. 2022. arXiv: 2202.13641 [quant-ph].