Locally testable codes and cayley graphs

Citation

Published Version
doi:10.1145/2554797.2554807

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:34604533

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Abstract

We give two new characterizations of (\(F_2\)-linear, smooth) locally testable error-correcting codes in terms of Cayley graphs over \(F_2^h\):

1. A locally testable code is equivalent to a Cayley graph over \(F_2^h\) whose set of generators is significantly larger than \(h\) and has no short linear dependencies, but yields a shortest-path metric that embeds into \(\ell_1\) with constant distortion. This extends and gives a converse to a result of Khot and Naor (2006), which showed that codes with large dual distance extend and give a converse to a result of Khot and Naor (2012), which showed that locally testable codes imply Cayley graphs that are small-set expanders but have many large eigenvalues.

2. A locally testable code is equivalent to a Cayley graph over \(F_2^h\) that has significantly more than \(h\) eigenvalues near 1, which have no short linear dependencies among them and which “explain” all of the large eigenvalues. This extends and gives a converse to a recent construction of Barak et al. (2012), which showed that locally testable codes imply Cayley graphs that are small-set expanders but have many large eigenvalues.

1.1 Locally Testable Codes

Informally, a locally testable code (LTC) is an error-correcting code in which one can distinguish received words that are in the code from those that are far from the code by a randomized test that probes only a few coordinates of the received word. Local testing algorithms for algebraic error-correcting codes (like the Hadamard code and the Reed–Muller code) were developed in the literature on program testing [BLR93, KS96], inspired the development of the field of property testing [GGR98], and played a key role in the constructions of multi-prover interactive proofs and the proof of the PCP Theorem [BFL91, FG’98, BFLS91, AS98, ALM+98]. Indeed, they are considered to be the “combinatorial core” of PCPs (cf., [GS06, BG06]), and thus understanding what is possible and impossible with locally testable codes can point the way to a similarly improved understanding of PCPs. See the surveys [Tre04, Gol11, Ben10].

We focus on the commonly studied case of linear codes over \(F_2\). Thus a code is specified by a linear subspace \(C \subset F_2^n\), \(n\) is called the blocklength of the code, and \(k = \dim(C)\) is its rate. The minimum distance is \([d = \min_{x \neq y \in C} d(x, y) = \min_{x \in C \setminus \{0\}} |x|,}\) where \(d(\cdot, \cdot)\) denotes Hamming distance and \(|\cdot|\) denotes Hamming weight. These parameters are typically summarized by referring to \(C\) as an \([n, k, d]\) linear code.

A local tester for \(C\) is a randomized algorithm \(T\) that, when given oracle access to a received word \(r \in F_2^n\), makes at most a small number \(q\) of queries to symbols of \(r\) and accepts or rejects. If \(r \in C\), then \(T^r\) should accept with high probability (completeness), and if \(r\) is “far” from \(C\) in Hamming distance, then \(T^r\) should reject with high probability (soundness). It was shown in [BHR05] that any tester for a linear code can be converted into one with the following structure: the tester \(T\) randomly samples a string \(\alpha \leftarrow \mathcal{D}\) and accepts if \(\alpha \cdot r = 0\), where \(\mathcal{D} = \mathcal{D}_F\) is some distribution on the dual code \(C^\perp = \{\alpha \in F_2^n : \alpha \cdot c = 0, \forall c \in C\}\). In particular, such a tester has perfect completeness (accepts with probability 1 if \(r \in C\)). We say the tester (which is now specified solely by \(\mathcal{D}\)) has (strong) soundness \(\delta\) if for every \(r \in F_2^n\),

\[P_{\alpha \leftarrow \mathcal{D}}[\alpha \cdot r = 1] \geq \delta \cdot d(r, C),\]

where \(d(r, C) = \min_{c \in C} d(r, c)\). This formulation of soundness is often referred to as strong soundness in the literature. Typically, we want \(\delta = \Omega(1/\ell)\), where \(\ell\) is the minimum distance of the code, so that received words at distance \(\Omega(\ell)\) from \(C\) are rejected with constant probability. If there are received words at distance \(\omega(d)\) from \(C\), then it is common to cap the rejection probability at a constant (e.g. require \(P_{\alpha \leftarrow \mathcal{D}}[\alpha \cdot r = 1] \geq \min\{\delta \cdot d(r, C), 1/3\}\)), but we ignore this issue in the introduction for simplicity. An alternate formulation known as weak soundness only requires that the
test reject with probability $\Omega(1)$ when $r$ is at distance $\beta d$ from the code, for some fixed constant $\beta < 1/2$.

We are interested in two paramater regimes for LTCs:

**Asymptotically Good LTCs** Here we seek $r = \Omega(n)$, minimum distance $d = \Omega(n)$, soundness $\delta = \Omega(1/d) = \Omega(1/n)$, and query complexity $q = O(1)$. Unfortunately, we do not know whether such codes exist — this is the major open problem about LTCs first posed by [GS06], and it is closely related to the long-standing open question about whether SAT has constant-query PCPs of linear length (which would enable proving that various approximation problems require time $2^{\Omega(n)}$ under the exponential-time hypothesis). The closest we have is Dinur’s construction [Din07], which has inverse-polynomial (rather than constant) relative rate (i.e. $n = k \cdot \text{polylog}(k)$). A recent result by Viderman additionally achieves strong soundness [Vid13].

**Constant-distance LTCs** Here we are interested in codes where the minimum distance $d$ is a fixed constant, and the traditional coding question is how large the rate $k$ can be as $n \to \infty$. BCH codes have (optimal) rate $k = n - (d/2) \cdot \log n$, but do not have any local testability properties. Reed–Muller codes yield the best known locally testable codes in this regime, with rate $k = n - O((\log n)^{\omega(d)})$ and query complexity $q = O(n/d)$ to achieve soundness $\delta = \Omega(1/d)$ [BKS +10]. This (query complexity is optimal, as $O(n/d)$ queries is needed to detect $d/2$ random corruptions to a codeword with constant probability.) An open question is whether rate $k = n - c_d \cdot \log n$ is possible, for some constant $c_d$ depending only on $d$ (but not on $n$).

In terms of limitations of LTCs, there are a number of results shedding light on the structure of an LTC with good parameters (see the survey [Ben10]), but there are essentially no nontrivial upper-bounds on rate known for arbitrary $F_2$-linear LTCs.

Instead of bounding the query complexity of our LTCs, it is convenient for us to work with smooth LTCs, where we simply require that the tester does not query any one coordinate too often. Formally, a tester, specified by a distribution $D$ on $C^n$, is $\varepsilon$-smooth if for every $i \in [n]$, $\Pr_{x \sim D}[x_i = 1] \leq \varepsilon$. This is analogous to the notion of smooth locally decodable codes (LDCs) defined by Katz and Trevisan [KT00]. Like in the case of LDCs, smoothing soundness is closely related to bounding query complexity, where query complexity $q$ corresponds to smoothness $Q(q/n)$ (as would be the case for testers that make $q$ uniformly distributed queries). In particular, we want the soundness to be $\varepsilon = O(1/n)$ in the asymptotically good regime, and $\varepsilon = O(1/d)$ in the constant-distance regime.

As we show in Appendix A, given a code with a $q$-query tester that gives weak soundness, we can convert it into a code with similar parameters and a tester with smoothness $O(q/d)$ that gives weak soundness. In the reverse direction, we can convert an $\varepsilon$-smooth tester for a code into a $O(\varepsilon n)$ query tester while preserving weak soundness. However, we are unaware of such reductions that preserve strong soundness, and there are examples of codes in the literature (such as those in [Vid13]) which have $q$-query testers that give strong soundness, but which are not known to have smooth testers with strong soundness. Below we focus mainly on smooth

---

1. This is a case where the rejection probabilities need to be capped at a constant, since there may be received words at distance $\omega(d)$ from $C$.

2. We are thankful to Or Meir [Mei13] for clarifying this and pointing out an error in a previous version of this paper.

---

1.2 Cayley Graphs

Cayley graphs are combinatorial structures associated with finite groups and are useful for applications ranging from pure group theory to reasoning about the mixing rates of Markov chains to explicit constructions of expander graphs [Big93, HLW06]. In this paper, we focus on the case that the group is a finite-dimensional vector space $V$ over $\mathbb{F}_2$. (So $V \cong \mathbb{F}_2^n$ for some $h \in \mathbb{N}$.) Given a multiset $S \subseteq V$, the Cayley (multi)graph $\text{Cay}(V, S)$ has vertex set $V$ and edges $(x, x + s)$ for every $s \in S$ (with appropriate multiplicities if $S$ is a multiset). Note that this is an $|S|$-regular undirected graph, since every element of $V$ is its own additive inverse. If we take $S$ to be a basis of $V$, then $\text{Cay}(V, S)$ is simply the $h$-dimensional hypercube, where $h = \dim V$. We will be interested in the properties of such graphs when $|S|$ is larger than $h$.

1.3 LTCs and Metric Embeddings of Cayley Graphs

Our first result shows that locally testable codes are equivalent to Cayley graphs with low-distortion metric embeddings into $\ell_1$. We refer the reader to [Mat02, Chapter 15] for background on metric embeddings. An embedding of a metric space $(X_1, d_1)$ into metric space $(X_2, d_2)$ with distortion $c \geq 1$ is a function $f : X_1 \to X_2$ such that for some $\alpha \in \mathbb{R}^+$ and every $x, y \in X_1$,

$$ \alpha \cdot d_1(x, y) \leq d_2(f(x), f(y)) \leq c \cdot \alpha \cdot d_1(x, y). $$

A commonly studied case is when $(X_1, d_1)$ is the shortest-path metric $d_0$ on a graph $G$, and $(X_2, d_2)$ is an $\ell_2$ metric. Indeed, we will take $(X_1, d_1)$ to be the shortest-path metric on a Cayley graph, and $(X_2, d_2)$ to be an $\ell_2$ metric. We will use the well-known characterization of $\ell_1$ metrics as the cone of the “cut metrics”\(^2\); a finite metric space $(X_2, d_2)$ is an $\ell_1$ metric if and only if there is a constant $\alpha \in \mathbb{R}^+$ and a distribution $F$ on boolean functions on $X_2$ such that for all $x, y \in X_2$, $d_2(x, y) = \alpha \cdot \Pr_{f \sim F}[f(x) \neq f(y)]$.

Note that the shortest-path metric on the hypercube $\text{Cay}(V, \{e_1, \ldots, e_n\})$ is an $\ell_1$ metric. Indeed, this metric is simply the Hamming distance $d$, and $d(x, y) = \Pr_{x \sim \text{Bin}(n, 1/2)}[x \neq y]$. We show that the existence of locally testable codes is equivalent to being able to approximate this property (i.e. have low-distortion embeddings into $\ell_1$) even when the number of generators is noticeably larger than $h$. To avoid trivial ways of increasing the number of generators (like duplicating generators, or taking small linear combinations), we will also require that the generators are $d$-wise linearly independent (i.e. have no linear dependency of length smaller than $d$).

**Theorem 1.** 1. If there is an $\mathbb{F}_2$-linear code of blocklength $n$, rate $r$, and distance $d$ with an $\varepsilon$-smooth local tester of strong soundness $\delta$, then there is a Cayley graph $G = \text{Cay}(\mathbb{F}_2^n, S)$ such that $|S| = n$, $h = n - k$, $S$ is $d$-wise linearly independent, and the shortest path metric on $G$ embeds into $\ell_1$ with distortion at most $\varepsilon/\delta$.

2. If there is a Cayley graph $G = \text{Cay}(\mathbb{F}_2^n, S)$ such that $|S| = n$, $S$ is $d$-wise linearly independent, and the shortest path metric on $G$ embeds into $\ell_1$ with distortion at most $c$, then there is an $\mathbb{F}_2$-linear code of blocklength $n$, rate $k = n - h$, and distance $d$ with an $\varepsilon$-smooth local tester of strong soundness $\delta$, for some $\delta$ and $c$ such that $c/\delta \leq c$.

Note that the theorem provides an exact equivalence between locally testable codes and $\ell_1$ embeddings of Cayley graphs, except
that the equivalence only preserves the ratio $\varepsilon/\delta$ rather than the two quantities separately. It turns out that this ratio is the appropriate parameter to measure when considering strong soundness. (See Section 3.1.) For weaker notions of soundness, we obtain equivalences with weaker notions of low-distortion embeddings, such as “single-scale embeddings” (where we replace the requirement that $d_2(f(x), f(y)) \leq \alpha d_1(x, y)$ with $d_1(x, y) \geq D \Rightarrow d_2(f(x), f(y)) \geq \alpha D$, see [Lee05] and references therein).

The theorem specializes as follows for the two main parameter regimes of interest:

**Corollary 2.** There is an asymptotically good smooth LTC with strong soundness $(k = n^{\Omega(n)}, d = \Omega(n), \varepsilon/\delta = O(1))$ iff there is a Cayley graph $G = \text{Cay}(F_2^n, S)$ with $|S| = (1 + \Omega(1))h$ such that $S$ is $\Omega(h)$-wise linearly independent and the shortest path metric on $G$ embeds into $\ell_1$ with distortion $O(1)$.

**Corollary 3.** For a constant $d$, there is a distance $d$ smooth LTC of blocklength $n$ with rate $k = n - c_d \log n$ and $\varepsilon/\delta = O(1)$ iff there is a Cayley graph $G = \text{Cay}(F_2^n, S)$ with $|S| = 2^{d/4c}$ such that $S$ is $d$-wise linearly independent and the shortest path metric on $G$ embeds into $\ell_1$ with distortion $O(1)$.

To interpret the theorem, let’s consider what the conditions on the Cayley graph $G$ mean. The condition that $|S| = n$ and the elements of $S$ are $d$-wise independent means that locally, in balls of radius $d$, the graph $G$ looks like the $n$-dimensional hypercube (which embeds into $\ell_1$ with no distortion). However, it is squeezed into a hypercube of significantly lower dimension $h$ (which may make even constant distortion impossible).

The canonical example of graphs that do not embed well into $\ell_1$ are expander, we could take $D$ and a distribution $\epsilon \rightarrow \{2\}$ on pairs of vertices in the graph, and deduce a superconstant lower bound on $\epsilon$ with no distortion). However, it is squeezed into a hypercube of significantly lower dimension $h$ (which may make even constant distortion impossible).

The canonical example of graphs that do not embed well into $\ell_1$ are expander, we could take $D$ and a distribution $\epsilon \rightarrow \{2\}$ on pairs of vertices in the graph, and deduce a superconstant lower bound on $\epsilon$ with no distortion). However, it is squeezed into a hypercube of significantly lower dimension $h$ (which may make even constant distortion impossible).

The canonical example of graphs that do not embed well into $\ell_1$ are expander, we could take $D$ and a distribution $\epsilon \rightarrow \{2\}$ on pairs of vertices in the graph, and deduce a superconstant lower bound on $\epsilon$ with no distortion). However, it is squeezed into a hypercube of significantly lower dimension $h$ (which may make even constant distortion impossible).

The canonical example of graphs that do not embed well into $\ell_1$ are expander, we could take $D$ and a distribution $\epsilon \rightarrow \{2\}$ on pairs of vertices in the graph, and deduce a superconstant lower bound on $\epsilon$ with no distortion). However, it is squeezed into a hypercube of significantly lower dimension $h$ (which may make even constant distortion impossible).
Let’s compare these properties with those of the $\varepsilon$-noisy hypercube. Recall that, in the $\varepsilon$-noisy hypercube, the coordinate cuts $S = \{e_1, \ldots, e_k\}$ are linearly independent and all give eigenvalues of $1-2\varepsilon$. And for every $b$, $\lambda(b) = (1-2\varepsilon)^{|b|} = 1 - 2\varepsilon |\text{rank}_S(b)| (\text{provided } |\text{rank}_S(b)| \leq O(1/\varepsilon)).$

Like in our metric embedding result, the main difference here is that we are asking for the set $S$ to be of size larger than $h$ (while retaining $d$-wise independence among the generators), so we need to squeeze many large eigenvalues into a low-dimensional space. One reason that these spectral properties are interesting is that they imply that the graph $\mathcal{G}$ is a small-set expander for sets of size $|V|/\exp(d)$ (see Lemma 18).

One direction of the above theorem (from LTCs to Cayley graphs) is extracted from the work of Barak et al. [BGH+12], who used locally testable codes (in the constant distance regime) to construct small-set expanders that have a large number of large eigenvalues (as a function of the number $H = 2^h$ of vertices). Such graphs provide barriers to improving the analysis of the Arora–Barak–Steurer algorithm for approximating small-set expansion and unique games [ABS10], and were also used by Barak et al. [BGH+12] to construct improved integrality gap instances for semidefinite programming relaxations of the unique games problem. Our contribution is showing that the connection can be reversed, when formulated appropriately (in terms of spectral properties rather than small-set expansion).

We can specialize Theorem 4 to the two parameter regimes of interest to us (see Corollaries 16 and 17 in Section 4.4 for precise statements). The existence of asymptotically good smooth LTCs with strong soundness is equivalent to the existence of asymptotically good LTCs blocklength $n$, rate $c$, and distance $d$ is testable if it has a tester $D$ which has smoothness $\varepsilon$ and soundness $\delta$. By considering received words at distance $1$ from the code, we get $\delta \leq \text{Rej}(e_i, D) \leq \varepsilon$. The upper bound is an easy consequence of the smoothness. Ideally, we want $\delta = O(\varepsilon)$.

Given $v \in V$, let $\bar{v} \in V/C$ denote the coset of $C$ containing it. Let $\bar{C} = \{e_1, \ldots, e_n\}$ denote the coset representatives of the basis vectors $\{e_1, \ldots, e_n\}$. The $e_i$s are not independent over $F_2$, indeed we have

$$\sum_{i \in S} \bar{e}_i = 0 \iff \sum_{i \in S} e_i \in C$$

Hence the shortest non-trivial linear dependence is of weight exactly $d$.

For $\bar{v} \in V/C$, there could be several ways to write it as a linear combination over $\bar{\mathcal{E}}$. We have

$$\bar{v} = \sum_{i \in S} \bar{e}_i \iff v + \sum_{i \in S} e_i \in C$$

Hence if we define $d(v, C) = d(\bar{v}, C)$ for any $v \in \bar{v}$ (the exact choice does not matter), it follows that $d(v, C) = \text{rank}_D(v)$. Similarly, for every $c \in C$, $v \in V$ and $\alpha \in C^*$, $\alpha(v) = \alpha(v + c)$.

Hence

$$\text{Rej}(v, D) = \text{Rej}(v + c, D)$$

This lets us define $\text{Rej}(\bar{v}, D) = \text{Rej}(v, D)$ for any $v \in \bar{v}$.

We can now rephrase smoothness and soundness in terms of coset representatives.

$$\Pr_{\alpha \in D}[\alpha_1 = 1] = \Pr_{\alpha \in D}[\alpha(e_i) = 1] = \text{Rej}(e_i, D)$$

Thus $D$ is $\varepsilon$-smooth if every $\bar{e}_i$ is rejected with probability at most $\varepsilon$.

We say a set $S$ of vectors in an $F_2$-linear space is $d$-wise independent if every $T \subseteq S$ where $|T| < d$ is linearly independent over $F_2$. For a set of vectors $S = \{x_1, \ldots, x_n\}$ which span a space $T$, we use $\text{rank}_S(t)$ for $t \in T$ to denote the smallest $k$ such that $t$ can be expressed as the sum of $k$ vectors from $S$. With this notation, $D$ has soundness $\delta$ if for every $\bar{v} \in V/C$ such that $\text{rank}_D \geq d'$, $\text{Rej}(\bar{v}) \geq \delta d'$.

We summarize these observations in the following lemma:
Lemma 5. Let $C$ be an $[n, k]_2$ code and let $D$ be a tester for $C$.

- $C$ has distance $d$ iff the set $\mathcal{E}$ is $d$-wise independent.

- The tester $D$ is $\varepsilon$-smooth iff $\text{Rej}(\varepsilon_i, D) \leq \varepsilon$ for all $i \in [n]$.

- For $v \in V/C, d(v, C) = \text{rank}_G(v)$. Hence $D$ has soundness $\delta$ iff for every $v \in V/C$,

\[
\text{Rej}(v, D) \geq \delta \cdot \text{rank}_G(v)
\]

3. **Locally Testable Codes and Metric Embeddings**

Let $S = \{s_1, \ldots, s_n\} \subseteq \mathbb{F}_2^n$ be set of $n \geq h$ generators of $\mathbb{F}_2^n$ that are $d$-wise independent. Let $G = \text{Cay}(\mathbb{F}_2^n, S)$ be the Cayley graph whose edges correspond to the set $S$. The graph $G$ can be naturally associated with a code $C_G$ which consists of all vectors $c = (c_1, \ldots, c_n)$ such that $\sum_i c_i s_i = 0$. It is easy to see that $C_G$ is an $[n, n-h, \frac{d}{2}]_2$ linear code.

Similarly, one can start from an $[n, k, \frac{d}{2}]_2$ linear code $C$ and construct a Cayley graph $G_C$ on $\mathbb{F}_2^n$. We take the vertex set to be $\mathbb{F}_2^n$. We add an edge $(x, y)$ if there exist $x \in \bar{x}$ and $y \in \bar{y}$ such that $d(x, y) = 1$. It is easy to see that is equivalent to taking $S = \{x_1, \ldots, x_n\}$, and this set is $d$-wise independent by the distance property of $C$.

It is easy to see that this construction inverts the previous construction. Henceforth we will use a code $C$ and a graph $G$ that can be derived from one another. The vertex set of $G$ is given by $V(G) = \mathbb{F}_2^n/C$ and the edge set $E(G)$ by $(x, x + \varepsilon_i)$ for $x \in \mathbb{F}_2^n/C$ and $i \in [n]$.

**Lemma 6.** Let $d_G$ denote the shortest path metric on $G$. We have

\[
d_G(x, y) = d_G(x + \varepsilon_i, 0) = d(x + y, C)
\]

**Proof:** If $d_G(x, y) \leq d$ then there exists $T \subseteq [n]$ of size $d$ such that

\[
y = x + \sum_{j \in T} \varepsilon_i \Rightarrow x + y = \sum_{j \in T} \varepsilon_i
\]

hence $d(x + y, C) \leq d$. Similarly, if $d(x + y, C) \leq d$, that gives an $x-y$ path of length $d$ in $G$.

An $\ell_1$-embedding of the shortest path metric $d_G$ on the graph $G$ is a distribution $D$ over Boolean functions $f : V(G) = \mathbb{F}_2^n/C \rightarrow \{\pm 1\}$. The distance $d_G(x, y)$ between a pair of vertices $x$ and $y$ under this embedding is

\[
d_G(x, y) = \text{Pr}_{f \in D} [f(x) \neq f(y)]
\]

We define the stretch of an edge $(x, y)$ to be the ratio $d_G(x, y)/d(x, y)$. The distortion $c_D$ of the embedding $D$ is the ratio of the maximum to the minimum stretch of any pair of vertices. It is given by

\[
c_D = \frac{\max_{x, y \in V(G)} d_G(x, y)/d(x, y)}{\min_{x, y} d_G(x, y)/d(x, y)}
\]

It follows by the triangle inequality that the stretch is maximized by some edge. Hence we get

\[
c_D = \frac{\max_{x, y \in V(G), i \in [n]} d(G, x, y)/d(x, y)}{\min_{x, y} d_G(x, y)/d(x, y)}
\]

The minimum $c$ achieved over all $\ell_1$-embeddings of $G$ is denoted $c_1(G)$.

**Definition 7.** An embedding $D$ of $G = \text{Cay}(\mathbb{F}_2^n, S)$ into $\ell_1$ is linear if $D$ is supported on functions $\chi_{\alpha}(x) = (-1)^{\alpha(x)}$ where $\alpha$ is an $\mathbb{F}_2$-linear function on $V(G)$.

The space of linear functions on $\mathbb{F}_2^n/C$ is isomorphic to $C^\perp$. In a linear embedding, we have

\[
\delta(x, y) = \text{Pr}_{\alpha \in \mathbb{F}_2} [\chi_{\alpha}(x) \neq \chi_{\alpha}(y)] = \text{Pr}_{\alpha \in \mathbb{F}_2} [\chi_{\alpha}(x + y) \neq 1] = \delta(x + y, 0)
\]

Thus, the distance $\delta$ is invariant under shifting in linear embeddings, just like the shortest path distance $d_G$. Indeed, the next lemma shows that we can replace any embedding by a linear embedding without increasing the distortion.

**Lemma 8.** There is a linear embedding $D$ of $G$ into $\ell_1$ achieving distortion $c_1(G)$.

To prove this lemma, we set up some machinery. Let $f : \mathbb{F}_2^n/C \rightarrow \{\pm 1\}$ be a Boolean function on $\mathbb{F}_2^n/C$. We can extend $f$ to a function on all of $\mathbb{F}_2^n$ by setting $f(x) = f(x)$. (We will henceforth switch freely between both notions). The resulting function is invariant under cosets of $C$, which implies that its Fourier spectrum is supported on $C^\perp$. Hence we have

\[
f(x) = \sum_{\alpha \in C^\perp} \hat{f}(\alpha)\chi_{\alpha}(x)
\]

Further, we have

\[
\sum_{\alpha \in C^\perp} \hat{f}(\alpha)^2 = 1
\]

We now proceed to the proof of Lemma 8.

**Proof:** Given an arbitrary distribution $D$ on Boolean functions, we define a new distribution $D'$ on functions where we sample $a \in \mathbb{F}_2^n$ uniformly at random, $f \in D$, and return the function $f' : \mathbb{F}_2^n \rightarrow \{\pm 1\}$ defined by $f'(x) = f(x + a)$. We will show that $c_D' \leq c_D$.

Let $a_1$ and $a_2$ be the values of $a$ that minimize and maximize $\delta(x + a, y + a)$ respectively. Then

\[
\delta(x + a_1, y + a_1) \leq \delta(x, y) \leq \delta(x + a_2, y + a_2)
\]

Hence we have

\[
\frac{\delta(x + a_1, y + a_1)}{d(x + a_1, y + a_1)} \leq \frac{\delta(x, y)}{d(x, y)} \leq \frac{\delta(x + a_2, y + a_2)}{d(x + a_2, y + a_2)}
\]

since all the denominators are equal. But this implies that

\[
\min_{x, y} \frac{\delta(x, y)}{d(x, y)} \geq \min_{x, y} \frac{\delta(x, y)}{d(x, y)}
\]

\[
\max_{x, y} \frac{\delta(x, y)}{d(x, y)} \leq \max_{x, y} \frac{\delta(x, y)}{d(x, y)}
\]

and hence

\[
c_D' = \frac{\max_{x, y} \delta(x, y)/d(x, y)}{\min_{x, y} \delta(x, y)/d(x, y)} \leq \frac{\max_{x, y} \delta(x, y)/d(x, y)}{\min_{x, y} \delta(x, y)/d(x, y)} = c_D
\]

Next we show that there is a linear embedding $D''$ with distortion $c_D'' = c_D'$. The embedding is simple to describe: we first sample
We denote this distribution on \(C^\perp\) by \(\hat{f}\). Note that

\[
\delta'(x, y) = \Pr_{a \in \mathbb{F}_2^2, f \in \mathcal{D}} \left[ f(x + a) \neq f(y + a) \right]
\]

We sample \(\chi_\alpha \in C^\perp\) with probability \(f(\alpha)^2\). Then

\[
\delta'(x, y) = \mathbb{E}_{f \in \mathcal{D}} \left[ \frac{1}{2} \mathbb{E}_{a \in \mathbb{F}_2^2} [1 - f(x + a)f(y + a)] \right]
\]

We have

\[
\delta'(x, y) = \mathbb{E}_{f \in \mathcal{D}} \left[ \frac{1}{2} \left( \sum_{\alpha, \beta \in C^\perp} \hat{f}(\alpha)\hat{f}(\beta)\chi_\alpha(x)\chi_\beta(y) \right) \right]
\]

We get

\[
\delta'(x, y) = \mathbb{E}_{f \in \mathcal{D}} \left[ \frac{1}{2} \left( \sum_{\alpha \in C^\perp} \hat{f}(\alpha)^2 \chi_\alpha(x)\chi_\alpha(y) \right) \right]
\]

And we have

\[
\delta'(x, y) = \Pr_{f \in \mathcal{D}, \alpha \in \mathbb{F}_2^2} [\chi_\alpha(x) \neq \chi_\alpha(y)]
\]

From this it follows that \(c_p' = c_p \leq c_D\).

The lemma follows by taking \(\mathcal{D}\) to be the \(\ell_1\) embedding of \(G\) that minimizes distortion.

We now prove the main result of this section.

**Theorem 9.** We have \(c_i(\mathcal{G}) \leq \epsilon \text{ if there exists an } (\varepsilon, \delta)\text{-tester for } \mathcal{C}\) where \(\delta \geq \varepsilon/c\).

**Proof:** For linear embeddings, we can use shift invariance to simplify the expression for distortion. Since \(\mathcal{D}\) is a distribution on \(C^\perp\), we can view it as a tester for \(\mathcal{C}\). Note that \(\Rej(\bar{x}, \mathcal{D}) = \delta(\bar{x}, 0)\). Recall by Equation (3)

\[
c_D = \max_{x \in \mathcal{V}(G)} \frac{\delta(\bar{x}, \bar{x} + \bar{e}_i)}{\min_{x, y} \delta(\bar{x}, \bar{y})/d_G(\bar{x}, \bar{y})}
\]

We can use \(\delta(\bar{x}, \bar{y}) = \delta(\bar{x} + \bar{y}, 0) = \Rej(\bar{x} + \bar{y}, \mathcal{D})\) to rewrite this as

\[
c_D = \max_{x \in \mathcal{V}(G)} \frac{\Rej(\bar{e}_i, \mathcal{D})}{\min_{x \in \mathcal{V}(G)} \Rej(\bar{x}, \mathcal{D})/d(\bar{x}, 0)}
\]

Given a linear embedding specified by a distribution \(\mathcal{D}\) that gives distortion \(c_D\), we view \(\mathcal{D}\) as a tester. By definition, it has smoothness \(\varepsilon\) for

\[
\varepsilon \geq \max_{i \in [n]} \Rej(\bar{e}_i, \mathcal{D})
\]

and has soundness \(\delta\) for

\[
\delta \leq \min_{x \in \mathcal{V}(G)} \Rej(\bar{x}, \mathcal{D})/d(\bar{x}, 0)
\]

since any such \(\delta\) satisfies the condition

\[
\Rej(\bar{x}, \mathcal{D}) \geq \delta d(\bar{x}, 0).
\]
low-weight dual codewords is a necessary condition for local testability. Our results allow for a simple alternative proof of Theorem 11.

Proof of Theorem 11. By Theorem 9, there exists a $(\varepsilon, \delta)$-tester $D$ so that $c_1(D) = \varepsilon / \delta$. We may assume without loss of generality that $D$ is supported on non-zero code-words in $C^\perp$, each of which is of weight at least $d^\delta$. As we have $\varepsilon \geq d^\delta / n$.

As in Section 3.1, we have $\delta \leq 1/t$, where $t$ is the covering radius of $C$. We lower bound $t$ by a standard volume argument:

$$2^{n-h} \cdot \sum_{t' = 0}^t \binom{n}{t'} \geq 2^n \Rightarrow t = \Omega \left( \frac{h}{n \log(n/h)} \right).$$

So we have

$$c_1(D) \geq \varepsilon \delta \geq \frac{d^\delta}{n} = \Omega \left( \frac{d^\delta}{n \log(n/h)} \right).$$

☐

Lower bounds for basis testers:

A basis tester for a code $C$ is a tester $D$ which is supported on a basis for $C$. Ben-Sasson et al. showed a strong lower bound for such testers [BSGK+10]. Their main result when restated in our notation says:

**Theorem 12.** [BSGK+10, Theorem 5] Let $C$ be an $[n, k, d]_2$ code with an $(\varepsilon, \delta)$-basis tester. Then

$$\frac{\varepsilon}{\delta} \geq \frac{kd}{3n}.$$

If $C$ is an $[n, k, d]_2$ code, a basis tester for $C$ yields an embedding into $(n-k)$-dimensional space. Hence their result implies that any linear embedding of $F_2^n / C$ into $(n-k)$-dimensional space requires distortion $\Omega(kd/n)$ (even though $F_2^n / C$ has dimension $n-k$ as a vector space over $F_2$), and hence low-distortion embeddings must have larger support. Note that since our reduction from arbitrary embeddings to linear embeddings could blow up the support, this does not imply a similar bound for arbitrary embeddings.

4. **Locally Testable Codes and Derandomized Hypercubes**

4.1 Derandomized Hypercubes

We consider Cayley graphs over groups of characteristic 2. Let $A = F_2^h$ for some $h > 0$. A distribution $D$ over $A$ gives rise to a weighted graph $Cay(A, D)$ where the weight of edge $(\alpha, \beta)$ equals $D(\alpha + \beta)$.

The symmetry of Cayley graphs makes it easy to compute their eigenvectors and eigenvalues explicitly. Let $A^\perp$ denote the space of all linear functions $b : A \to F_2$.

2. **Linear Independence.** The linear functions $\{e_1, \ldots, e_n\}$ corresponding to the top eigenvectors are linearly independent over $F_2$.

3. **Spectral Decay.** For $a \in A^\perp$, we have $\lambda(a) = \sum_{i \in S} e_i$, then $\lambda(a) \leq (1 - \varepsilon)^{\|a\|}$.

We are interested in Cayley graphs whose threshold rank $n$ is possibly (much) larger than $h$. But this means that the corresponding dual vectors which lie in the space $A^\perp$ of dimension $h < n$ can no longer be linearly independent. So we relax the Linear Independence condition, and only ask that there should be no short linear dependencies between these vectors. The Spectral Decay condition will stay the same, except that we need to modify the notion of rank to account for linear dependencies.

**Definition 13.** Let $Cay(A, D)$ be a Cayley graph on the group $A = F_2^h$. Let $\mu, \nu \in [0, 1]$ and $d \in \{1, \ldots, n\}$. Let $B^* = \{b_1, \ldots, b_h\}$ be a $d$-wise independent set of generators for $A^\perp$ of cardinality $n$. We say that $B^*$ is a $(\mu, \nu)$-spectrum generator for $Cay(A, D)$ if it satisfies the following properties:

- **Large Eigenvalues.** $\lambda(b) \geq 1 - \mu$ for every $b \in B^*$.
- **Spectral Decay.** For $a \in A^\perp$, $\lambda(a) \leq 1 - \nu \cdot \operatorname{rank}(b \cdot a)$.

Note that any set of generators $B^*$ for $A^\perp$ gives us some values of $d$, $\nu$. We would like $n, d, \mu$ and $\nu$. We would like $n, d, \mu$, $\mu$ and $\nu$. Ideally, we would like them to be within a constant factor of each other.

We refer to such graphs as “derandomized hypercubes”. The reason is that if there is a generating set of size $n$ which is significantly larger than the dimension $h$, then the resulting graph has spectral properties that resemble the $n$ dimensional hypercube, although it has only $2^n \ll 2^h$ vertices. Every Cayley graph $Cay(A, D)$ together with a generating set $B^*$ gives us a derandomized hypercube, the parameters $n, d, \mu, \nu$ tell us how good the derandomization is (just like any code $C$ and dual distribution $D$ gives us local tester, whose quality is governed by the parameters it achieves).

4.2 Derandomized hypercubes from Locally Testable Codes

Barak et al. proposed the following construction of Derandomized Hypercubes from any Locally Testable Code [BGH+12]. Given $C$ which is an $[n, k, d]_2$ code with a linear tester $D$, they consider the Cayley graph $Cay(C^\perp, D)$ on $C^\perp \cong F_2^{-k}$ whose edge weights are distributed according to $D$.

**Theorem 14.** Let $C$ be an $[n, k, d]_2$ code for $d \geq 3$, and let $D$ be an $(\varepsilon, \delta)$-tester for $C$. There exists a $d$-wise independent set $E$ of size $n$ which is a $(2\varepsilon, 2\delta)$-spectrum generator for $Cay(C^\perp, D)$.

**Proof.** Observe that $(C^\perp)^* \cong V / C$. This is because each $v \in V$ defines a linear function on $C^\perp$ given by $v(\alpha) = \alpha(v)$, and $v, v'$ define the same function iff the lie in the same coset of $C$.

We take $E = \{e_1, \ldots, e_n\}$ to be the cosets corresponding to the received words $e_1, \ldots, e_n$. By Lemma 5, since $C$ has distance $d$, the set $E$ is $d$-wise independent. We will show that it is a $(2\varepsilon, 2\delta)$-spectrum generator for $Cay(C^\perp, D)$. 
We bound the eigenvalues using the correspondence between the spectrum of \( \text{Cay}(C^+, D) \) and the soundness of the tester \( D \) established by Barak et al.. For \( \bar{v} \in V/C \), let \( \chi_\bar{v} \) and \( \lambda(\bar{v}) \) denote the corresponding eigenvalue. [BGH\textsuperscript{12}, Lemma 4.5] says that

\[
\lambda(\bar{v}) = 1 - 2\text{Rej}(\bar{v}, D).
\]

- **Smoothness implies Large Eigenvalues.** By Lemma 5 the smoothness of \( D \) implies \( \text{Rej}(\bar{e}_i, D) \leq \epsilon \). By Equation 7,

\[
\lambda(\bar{e}_i) = 1 - 2\text{Rej}(\bar{e}_i, D) \geq 1 - 2\epsilon.
\]

- **Soundness implies Spectral decay.** Fix \( \bar{v} \in V/C \) so that \( \text{rank}_{\text{gf}}(\bar{v}) \geq d' \). By Lemma 5, the soundness of \( D \) implies \( \text{Rej}(\bar{v}, D) \geq \delta d' \). By Equation 7,

\[
\lambda(\bar{v}) = 1 - 2\text{Rej}(\bar{v}, D) \leq 1 - 2\delta d'.
\]

\[\square\]

### 4.3 Locally Testable Codes from Derandomized Hypercubes

We show how to start from a Cayley graph on \( A = F_2^n \) and a set of generators for \( A^\ast \) and get a locally testable code from it. Our construction takes a Cayley graph \( \text{Cay}(A, D') \) and a \((d, \mu, \nu)\)-spectrum generator \( B^\ast \).

We define the locally testable code \( C \) by specifying the dual code \( C^\perp \) and the tester \( D \). We view elements \( \alpha \in A \) as messages, and embed them into \( F_2^n \) using the map

\[
f(\alpha) = (b_1(\alpha), \ldots, b_n(\alpha)).
\]

Since \( B^\ast \) generates \( A^\ast \), the mapping \( f \) is injective. Its image is a \( h \)-dimensional subspace of \( F_2^n \) which we denote by \( C^\perp \).

\[\square\]

### 4.4 Some consequences of this equivalence

This equivalence lets us reformulate questions regarding LTCs as questions regarding the existence of certain families of derandomized hypercubes.

**Corollary 16.** There exists an asymptotically good family of codes \( \{C_n\} \) where \( C_n \) has blocklength \( n \) and is \((O(1/n), \Omega(1/n))\)-locally testable iff for infinitely many \( h \) there exists a Cayley graph \( G_n = \text{Cay}(F_2^n, D) \) and a set \( B^\ast \) of generators for \( (F_2^n)^\ast \) such that

- the elements of \( B^\ast \) are \((\rho_0 h)\)-wise independent for \( \rho_0 > 0 \),
- \(|B^\ast| \geq \rho_1 h \) for \( \rho_1 > 1 \),
- \( B^\ast \) is an \((O(1/h), \Omega(1/h))\)-spectrum generator for \( G_n \).

**Corollary 17.** Let \( d \geq 3 \). There exists an asymptotic family of codes \( \{C_n\} \) where \( C_n \) has parameters \([n, n - c_4 \log n, d/2] \) and is \((\epsilon, \Omega(\epsilon))\)-locally testable iff for infinitely many \( h \) there exists a Cayley graph \( G_n = \text{Cay}(F_2^n, D) \) and a set \( B^\ast \) of generators for \( (F_2^n)^\ast \) such that

- the elements of \( B^\ast \) are \( d \)-wise independent,
- \(|B^\ast| \geq 2^{h/c_4} \),
- \( B^\ast \) is an \((\epsilon, \Omega(\epsilon))\)-spectrum generator for \( G_n \).

Next we show that derandomized hypercubes are small-set expanders. We say that a regular graph \( G \) with \( n \) vertices is a \((\tau, \phi)\)-expander if for every set \( S \) of at most \( \tau n \) vertices, at least a fraction \( \phi \) of the edges incident to \( S \) leave \( S \) (i.e. are on the boundary between \( S \) and \( \overline{S} \)).

The following lemma says that if a graph has a \((\mu, \nu)\) spectrum generator, then it is a \((\tau, \phi)\)-expander for appropriately chosen \( \tau \) and \( \phi \). The lemma is proved in [BGH\textsuperscript{12}]. Since our terminology and notation is different, we present a proof of the Lemma in Appendix C.
Lemma 18. \[BGH^{+}12\] Let \(G = \text{Cay}(\mathcal{A}, \mathcal{D})\) be a Cayley graph on the group \(\mathcal{A} = \mathbb{F}_2^n\). Let \(\mathcal{B}^n\) be a \(d\)-wise independent set which is a \((\mu, \tau, \nu)\)-spectrum generator for \(G\). Then \(G\) is a \((\tau, \phi_\tau)\) expander for \(\phi_\tau = \nu d/4 - 3^{d/2}/\tau^{1/4}\).

To interpret the expansion bound, think of \(\nu d/4 - 3^{d/2}/\tau^{1/4}\) (we can assume that the graphs obtained from LTCs have this property, since this is analogous to saying that words at distance \(d/4\) are rejected with constant probability). So if we take \(\tau = \exp(-d)\), then \(\phi_\tau = \Omega(1)\). A particularly interesting instantiation of this bound is obtained by combining Corollary 17 and Lemma 18:

Corollary 19. For \(d \geq 3\), suppose there exists an asymptotically good code \(\{C_n\}\) where \(C_n\) has parameters \([n, n - c_d \log n, d]\) and is \((O(1/d), 1(1/d))-\)locally testable. There for infinitely many \(h\) there exists a Cayley graph \(G_h = \text{Cay}(\mathbb{F}_2^n, \mathcal{D})\) such that \(G_h\) is \((O(9^{-h}), 1(1))-\)expander and has \(2^{h/c_d}\) eigenvalues greater than \(1 - O(1/d)\).

In contrast, Arora et al. \[ABS10\] showed that if \(G\) is an \((\tau, \Omega(1))-\)expander, then there are at most \(n^{O(n)/\tau}\) eigenvalues greater than \(1 - \epsilon\). Their bound implies that the graph \(G_h\) obtained from Corollary 19 cannot have too many \(2^{h/c_d}\) eigenvalues greater than \(1 - O(1/d)\). If there exist LTCs with \(c_d = O(d)\), the resulting graphs \(G_h\) would meet the ABS bound. The only lower bound we know of for \(c_d\) is \(c_d \geq d/2\) by the Hamming bound.

5. ACKNOWLEDGEMENTS

This work was done in part when the second author was on leave as a Visiting Researcher at Microsoft Research Silicon Valley and a Visiting Scholar at Stanford University, and the third author was visiting Microsoft Research Silicon Valley.

The second author was supported in part by NSF grant CCF-1116616 and US-Israel BSF grant 2010196. The third author was supported in part by a grant from the Simons Foundation (Award Number 252545).

We thank Or Meir for pointing out an error in an earlier version of this paper, regarding the relationship between smoothness and bounded query complexity for locally testable codes. We thank Alex Andoni, Anupam Gupta and Kunal Talwar for useful discussions and pointers to the literature on metric embeddings. And we thank Raghu Meka, Prasad Raghavendra and Madhu Sudan for helpful discussions.

6. REFERENCES


[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions (extended abstract). In 29th Annual Symposium on Foundations of Computer Science, pages 68–80,


APPENDIX

A. SMOOTHNESS VERSUS BOUNDED QUERY COMPLEXITY

In this section, we describe an equivalence between constructing LTCs with smooth testers and constructing LTCs with bounded query complexity. The equivalence is only for weak soundness, so we define that more formally now.

**Definition 20.** Let $C \subseteq F_2^n$ be a linear code, and $D$ a distribution on $C^*$ (which we view as a tester for $C$). For $\delta > 0$, we say that $D$ has weak soundness at distance $d'$, if for every received word $r \in F_2^n$ such that $d(r, C) \geq d'$, we have $\Pr_{r \sim D}[\alpha \cdot r = 1] \geq 1/2$.

Thus, a tester with strong soundness $\delta$ has weak soundness at distance $d' = 1/2\delta$, but the converse does not hold. If $d$ is the minimum distance of the code, we typically want strong soundness at distance $d' = \beta d$ for an arbitrarily small constant $\beta$. (If $\beta$ is too large, e.g. $\beta \geq 1/2$, then there may be no received words at distance at least $\beta d$, from the code, and weak soundness becomes vacuous.)

Now we note that it is easy to convert an $\epsilon$-smooth tester into one with bounded query complexity. Since an $\epsilon$-smooth tester for a code queries at most $\epsilon n$ coordinates in expectation, we can get a tester with query complexity $q = O(\epsilon n)$ by discarding tests that query more than $q$ coordinates. However, this transformation costs an additive constant (namely $\epsilon n/q$) in the soundness probability, and hence it only preserves weak soundness. (We can amplify the soundness probability back to $1/2$ by repeating the tester a constant number of times.)

The other direction is more involved and requires a modification to the underlying code.

**Theorem 21.** Let $C$ be an $[n,k,d]_2$ code that has a $q$-query tester $D$ giving weak soundness at distance $d'$. Then for every $d'' \in \mathbb{N}$ there exists an $[n',k,d']_2$ code $C'$ that has a $q/d''$-smooth tester $D'$ giving weak soundness at distance $d' + d''$, with blocklength $n' \leq n + d''$.

Thus taking $d'' = \Theta(d')$, we only increase the distance at which we have soundness by a constant factor, and obtain smoothness $\epsilon = O(q/d'')$. In an asymptotically good LTC, we have $d'' = \Omega(n)$, so smoothness $\epsilon = O(q/n)$ is indeed equivalent to query complexity $O(q)$. When $d'' = o(n)$, the relationship is not tight, and it would be nice to improve Theorem 21 to yield smoothness $O(q/n')$ in general.

**Proof:** Let $C$ be an $[n,k,d]_2$ code that has a $q$-query tester $D$ with weak soundness at distance $d'$. Our goal is to construct from it a smooth LTC with comparable parameters. Let $\epsilon_i$ denote the probability that $D$ queries coordinate $i$. We define a new code $C'$ by replicating coordinate $i$

$$m_i = \left\lceil \frac{\epsilon_i d''}{q} \right\rceil$$

times. This operation cannot decrease the minimum distance of the code.

The tester $D'$ for $C'$ is obtained by first sampling a test $\alpha$ according to $D$ and then for each $i$ in the support of $\alpha$, choosing one of the $m_i$ copies at random. Now each copy of coordinate $i$ is queried with probability $\epsilon_i/m_i \leq q/d''$, so the resulting tester $D'$ is $q/d''$-smooth.

We bound the blocklength of $C'$ as follows:

$$n' = \sum_{i=1}^{n} m_i$$

$$\leq \sum_{i=1}^{n} \left( 1 + \frac{\epsilon_i d''}{q} \right)$$

$$= n + \frac{d''}{q} \sum_{i=1}^{n} \epsilon_i$$

$$\leq n + d''$$,

where the last inequality follows from the fact that $D$ is a $q$-query tester. We now analyze the soundness. Running a test according to $D'$ on a received word $r'$ is the same as running a test from $D$ on a distribution $R(r')$ of received words in $C$, where we sample $r \in R(r')$ by setting the $i$-th coordinate in $r$ randomly to one of the $m_i$ coordinates of $r'$ which are meant to be copies of coordinate $i$. Note that if $r'$ is actually obtained by replicating coordinates of some received word $r$, then $R(r')$ is supported entirely on $r$.

Now consider a received word $r'$ where $d(r', C') \geq d' + d''$. We claim that every $r$ in the support of $R(r')$ satisfies

$$d(r, C) \geq d'$$

Suppose there exists an $r \in R(r')$ and $c \in C$ such that $d(r, c) < d'$; equivalently $r$ and $c$ agree in more than $n - d'$ coordinates. Then
consider the codeword \( c' \in C' \) that is obtained by replicating coordinates of \( c \). Then \( r' \) and \( c' \) agree in at least as many coordinates as \( r \) and \( c \), so
\[
d(r', c') \leq n' - (n - d') \leq (n + d') - (n - d') = d' + d'' ,
\]
which contradicts our assumption that \( d(r', c') \leq d' + d'' \).
Thus every \( r \in R(r') \) is at distance at least \( d' \) from \( C \) and hence is rejected with probability at least \( 1/2 \) by \( D' \). Hence \( D' \) rejects words at distance \( d' + d'' \) from \( C' \) with probability at least \( 1/2 \), and thus has weak soundness at distance \( d' + d'' \).

## B. PROOF OF THEOREM 10

Let \( D^\delta_\ell \) denote the distribution on \( C^\ell \) where we sample \( \ell \)-independent codewords according to \( D \) and add them. We claim that for suitable \( \ell \), both \( \varepsilon \) and \( \delta \) scale by roughly a factor of \( \ell \).

**Lemma 22.** Let \( \ell \) be such that
\[
\ell \leq \frac{1}{4 \text{Rej}(\bar{x}, D)} \quad \text{for all } \bar{x} \in F^n_2 / C.
\]
Then \( D^\delta_\ell \) is an \( (\varepsilon \ell, \delta \ell / 2) \)-tester for \( C \).

**Proof:** The tester \( D^\delta_\ell \) is an \( \varepsilon \ell \)-smooth tester by the union bound. Its soundness can be analyzed by noting that
\[
1 - 2 \text{Rej}(\bar{x}, D^\delta_\ell) = (1 - 2 \text{Rej}(\bar{x}, D))^{\ell}
\]
Using the bound on \( \ell \) to truncate the RHS, we get
\[
\text{Rej}(\bar{x}, D^\delta_\ell) \geq \frac{\ell}{2} \text{Rej}(\bar{x}, D) \geq \frac{\ell}{2} d(\bar{x}, 0).
\]

We use this to prove Theorem 10.

**Proof of Theorem 10.** We start with an \( (\varepsilon, \delta) \)-tester where \( \delta \geq \varepsilon / c_1(G) \). If \( \delta \) exceeds the claimed bound, we are already done. Assume this is not true, so
\[
\delta \leq \frac{1}{16c_1(G) \ell} = \frac{\varepsilon}{c_1(G) \delta} \leq \frac{1}{16 \ell}.
\]
Since the covering radius is \( \ell \), we have that for every \( \bar{x} \in F^n_2 \),
\[
\text{Rej}(\bar{x}, D) \leq \ell \varepsilon \leq 1/16.
\]
Let \( \ell = [1/(4 \ell \varepsilon)] \) so that
\[
\frac{1}{8 \ell \varepsilon} \leq \ell \leq \frac{1}{4 \ell \varepsilon}.
\]
By Lemma 22, \( D^\delta_\ell \) has smoothness \( \varepsilon' \) where
\[
\varepsilon' \leq \ell \varepsilon \leq \frac{1}{4 \ell}
\]
and soundness \( \delta' \) where
\[
\delta' \geq \frac{1}{2} \ell \delta \geq \frac{1}{2} \frac{1}{8 \ell \varepsilon} \quad c_1(G) \geq \frac{1}{16c_1(G) \ell}.
\]
\[
\square
\]

## C. PROOF OF LEMMA 18

We first show the following hypercontractive inequality:

**Claim 23.** Let \( B = \{b_1, \ldots, b_k\} \subseteq F^n_2 \) be \( (4d + 1) \)-wise independent. For every function \( f : F^n_2 \to \mathbb{R} \) defined as
\[
f(x) = \sum_{S \subseteq [n], |S| \leq d} \hat{f}(S) \prod_{i \in S} \chi_{b_i}(x),
\]
we have \( E_x[f(x)^4] \leq 9^d \left( E_x[f(x)^2]^2 \right)^2 \).

**Proof:** Let \( g : F^n_2 \to \mathbb{R} \) be
\[
g(y) = \sum_{S \subseteq [n], |S| \leq d} \hat{f}(S) \prod_{i \in S} y_i.
\]
The statement is proved by the standard \( (2, 4) \)-hypercontractive inequality (applied to \( g \) function) and the observation that
\[
E_x[f(x)^4] = E_y [g(y)^4]
\]
and
\[
E_y[g(y)^4] \leq \sum_{S_1 \cup S_2 \cup S_3 \cup S_4 = \emptyset} \hat{f}(S_1) \hat{f}(S_2) \hat{f}(S_3) \hat{f}(S_4).
\]

We now proceed to prove Lemma 18.

**Proof of Lemma 18.** For any two functions \( f, g : F^n_2 \to \mathbb{R} \), define their inner-product as
\[
\langle f, g \rangle = E_{x \in F^n_2} [f(x) g(x)]
\]
and the \( p \)-norm of \( f \) to be
\[
||f||_p = \left( E_{x \in F^n_2} [f(x)^p] \right)^{1/p}.
\]
For every function \( f : F^n_2 \to \mathbb{R} \) with Fourier expansion
\[
f(x) = \sum_{a \in F^n_2} \hat{f}_a \chi_a(x),
\]
let
\[
f^{<d/4}(x) = \sum_{a : \text{rank}_G(a) < d/4} \hat{f}_a \chi_a(x),
\]
\[
f^{\geq d/4}(x) = \sum_{a : \text{rank}_G(a) \geq d/4} \hat{f}_a \chi_a(x).
\]

Fix a set \( S \subseteq F^n_2 \). Let \( \tau = \mu(S) \) be the volume of \( S \). Let \( \mathbf{1}_S(x) = \mathbf{1}_{x \in S} \) be the indicator function \( S \). Note that \( ||\mathbf{1}_S||_2 = \tau \) for every \( p \geq 1 \). We will lower bound the expansion \( \Phi(S) = 1 - \langle \mathbf{1}_S, G \mathbf{1}_S \rangle / \tau \), which is the fraction of the edges incident to \( S \) leaving \( S \). Observe that
\[
\langle \mathbf{1}_S, G \mathbf{1}_S \rangle = \langle \mathbf{1}_S, G \mathbf{1}_S^{<d/4} \rangle + \langle \mathbf{1}_S, G \mathbf{1}_S^{\geq d/4} \rangle.
\]

The first term in the RHS of (11) is upper bounded as
\[
\langle \mathbf{1}_S, G \mathbf{1}_S^{<d/4} \rangle \leq \|\mathbf{1}_S\|_4 / \|G \mathbf{1}_S^{<d/4}\|_4 \leq \|\mathbf{1}_S\|_4 / \sqrt{3} \|\mathbf{1}_S\|_2 = \sqrt{3} \tau^{5/4}
\]
by Hölder’s inequality and Claim 23. The second term in the RHS of (11) is upper bounded as
\[
\langle \mathbf{1}_S, G \mathbf{1}_S^{\geq d/4} \rangle \leq (1 - \nu d/4) \|\mathbf{1}_S\|_2^2 = (1 - \nu d/4) \tau.
\]
In all, we have
\[
\Phi(S) = 1 - \langle \mathbf{1}_S, G \mathbf{1}_S \rangle \geq 1 - \sqrt{3} \tau^{1/4} - (1 - \nu d/4) \nu d/4 = \nu d/4 - \sqrt{3} \tau^{1/4}.
\]
\[
\square