Electronic Colloquium on Computational Complexity, Report No. 23 (2018)

# On coset leader graphs of structured linear codes

Eran Iceland and Alex Samorodnitsky [*]

**Abstract**

We suggest a new approach to obtain bounds on locally correctable and some locally testable binary linear codes, by arguing that their coset leader graphs have high discrete Ricci curvature.

The bounds we obtain for locally correctable codes are worse than the best known bounds obtained using quantum information theory, but are better than those obtained using other methods, such as the "usual" information theory. (We remark that our methods are completely elementary.)

The bounds we obtain for a family of locally testable codes improve the best known bounds.

## 1 Introduction

We are interested in upper bounds on the cardinality of locally structured linear subspaces of the Hamming space $\{0,1\}^n$.

To fix notions and some notation, let $C$ be a linear subspace of $\{0,1\}^n$, and let $C^{\perp} = \big\{y : \langle x, y\rangle = 0, \ \forall x \in C\big\}$ be the dual space. We will assume that $C^{\perp}$ contains a rich family of vectors of constant length, and try to deduce that it is large (alternatively, that $C$ is small).

Specifically, we consider two families of locally constrained linear binary codes. Such codes have numerous applications in theoretical computer science (see [DSW14] and the references therein). With that, essentially in all the cases, there is a significant gap between the best known examples of such codes and upper bounds on their cardinality.

Let $\mathcal{F}$ be the family of local constraints on $C$ (that is, constant length vectors in $C^{\perp}$). We will consider:

- *Locally correctable codes*

  These codes come with two parameters, an integer $q \geq 1$ and a density parameter $0 < \delta < 1/q$. For each $1 \leq i \leq n$, the family $\mathcal{F}$ contains either a unit vector supported at $i$, or at least $\delta n$ vectors of length $q+1$ whose supports contain $i$ and are disjoint otherwise.

[*] School of Engineering and Computer Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Research partially supported by ISF grants 1241/11 and 1724/15, and by BSF grant 2010451.

ISSN 1433-8092

- *Locally testable codes with high* 3*-density*

  Here we have one integer parameter $\sigma$ which tends to infinity with $n$, and we assume that $\mathcal{F}$ contains, for each $1 \leq i \leq n$, at least $\sigma$ vectors of length 3 whose supports contain $i$.

## 1.1 Known bounds

### 1.1.1 Locally correctable codes

- For $q = 1$, there are no locally correctable codes when $n$ is bigger than a constant [KT00].
- For $q = 2$, the answer is $\Theta(\log n)$ [GKST06]. [1]
- For a constant $q > 2$, there is a significant gap between upper and lower bounds. The best known locally correctable codes are the Reed-Muller codes of dimension $\Theta\left((\log n)^{q-1}\right)$ [MS77]. The best upper bound on the dimension is $O\left(n^{\frac{\lceil q/2\rceil - 1}{\lceil q/2\rceil}}\right)$, up to a polylogarithmic factor [KdW04, Woo07, Woo12].

**Remark 1.1:**

In fact, all these bounds hold also for *locally decodable codes*, which is yet another version of locally constrained codes. For a locally decodable code $C$ of (unknown) dimension $D$, we assume that the first $D$ coordinates of a vector in $C$ determine the vector (in other words the projection of a non-zero vector in $C$ on the first $D$ coordinates is non-zero). The parameters of $C$ are $q$ and $\delta$, and the family of local constraints $\mathcal{F}$ contains, for each $1 \leq i \leq D$, at least $\delta n$ vectors of length $q + 1$ whose supports contain $i$ and are disjoint otherwise.

Clearly, any locally correctable code is also locally decodable. The reverse implication does not hold [KV10], but we are not aware of any upper bounds separating these two families of binary codes.

■

### 1.1.2 Locally testable codes

Locally testable codes with high 3-density were considered in [BSV12]. Let us call a linear code $C$ *regular* if each column in its generating matrix appears with the same multiplicity. Then the following claim holds.

- The dimension of a regular locally testable code with 3-density $\sigma$ is at most $O\left(\frac{\log \sigma}{\sqrt{\sigma}} \cdot n\right)$ [BSV12].

[1] The constant hidden in the asymptotic notation here and below is allowed to depend on $\delta$.

### 1.2 Our results

We follow the idea of Friedman and Tillich [FT05] and consider the *coset leader graph* of a binary linear code (see Definition 1.7). At this point it suffices to say that this is a Cayley graph whose cardinality is that of the corresponding code. Following a line of thought in [FT05] we view this graph as a homogeneous space and apply suitably modified tools from Riemannian geometry to upperbound its cardinality. Specifically, we will show this graph to have positive discrete Ricci curvature in the sense of [Oll09]. This will provide an upper bound on its diameter and hence on its size.

**Bounds for locally correctable codes:** In the statement of the next claim, and from now on, we will refer to a locally correctable code with parameters $q$ and $\delta$ as $q$-locally correctable (this in particular emphasizes the fact that $q$ is the more important parameter for the purpose of this discussion). Recall that we allow constants hidden in the asymptotic notation to depend on $\delta$.

**Theorem 1.2:** *Let $C$ be a $q$-locally correctable code with $q \geq 2$. Then the covering radius of $C^{\perp}$ is $O\left(n^{\frac{q-2}{q-1}}\right)$, and $\dim(C) \leq O\left(n^{\frac{q-2}{q-1}}(\log n)^{\frac{1}{q-1}}\right)$.*

While this is weaker than the best known bounds, we observe that for $q > 3$ the exponent of $n$ in our bound on $\dim(C)$ lies strictly between that in [KT00] obtained via Shannon's entropy, and the best known bound [KdW04, Woo07], which uses highly non-trivial facts, such as subadditivity of quantum entropy.

For $q = 3$ we do somewhat better. Theorem 1.2 bounds the dimension of a 3-locally correctable code by $\sqrt{n \log n}$, which is only logarithmically weaker than the best known bound of $\sqrt{n}$ [Woo12] (but stronger than $n^{2/3}$ of [KT00]). In fact, we can recover the $\sqrt{n}$ bound in one special case.

**Definition 1.3:** We say that a $q$-locally correctable code is *perfect* if $\delta = \frac{n-1}{qn}$ for every $i \in [n]$ (that is, the density parameter is as large as possible). ▮

**Theorem 1.4:** *Let $C$ be a perfect 3-locally correctable code. Then $\dim(C) \leq O(\sqrt{n})$.*

**Bounds for locally testable codes with high 3-density:** We improve on the bounds for locally testable codes with high 3-density.

**Theorem 1.5:** *The dimension of a regular locally testable code with 3-density $\sigma$ is at most $\frac{2}{\sqrt{\sigma}} \cdot n$.*

This bound is tight, up to a constant factor [DK11].

We also consider a more general case in which the multiplicity of the columns is allowed to vary.

**Theorem 1.6:** *Let $v_1, v_2, \ldots, v_n$ be the columns of a generating matrix $G$ of a code $C$. Let $p$ be the maximal multiplicity of a column in $G$. Assume that each coordinate participates in at least $\sigma$ linear dependencies of length three, and that $\sigma > p$. Then*

$$dim(C) \leq O\left(\frac{\log\left(\lceil \sigma/p \rceil\right)}{\lceil \sigma/p \rceil} \cdot n\right)$$

This is tight, up to the $\log\left(\lceil \sigma/p \rceil\right)$-factor, see Example 4.2.

## 1.3 Our approach in more detail

Our starting point is the elegant proof of [FT05] for the first linear programming bound for binary linear codes. We start with the definition of coset leader graphs.

**Definition 1.7:** The *coset leader graph* $\mathbb{T}$ of a linear code $C \subseteq \{0,1\}^n$ is the Cayley graph of the quotient group $\mathbb{F}_2^n / C^\perp$ with respect to the set of generators given by the standard basis $e_1 + C^\perp, \ldots, e_n + C^\perp$. ▮

Note that $\mathbb{T}$ has $|C|$ vertices. Note also that $\mathbb{T}$ may have loops or parallel edges (if $C^\perp$ contains non-zero vectors of Hamming weight less than 3).

[FT05] employs discrete versions of comparison theorems in Riemannian geometry comparing, on one hand, the growth of neighborhoods in $\mathbb{T}$ with the growth of neighborhoods in $\{0,1\}^n$ and, on the other hand, the spectral behaviour of the Laplacian of $\mathbb{T}$ with the Laplacian of $\{0,1\}^n$.

Following [FT05], we view $\mathbb{T}$ as a "discrete manifold", and try to estimate the cardinality of $\mathbb{T}$ by employing insights and tools borrowed from Riemannian geometry. The main technical notion we use is that of *discrete Ricci curvature*, due to Ollivier [Oll09]. We show that in the cases we consider, $\mathbb{T}$ has 'positive curvature' which is bounded away from zero (as opposed, say, to the Hamming cube whose curvature is $\frac{2}{n+1}$ and hence goes to zero with dimension).

This allows us to upper bound the diameter of $\mathbb{T}$ (equivalently, the covering radius of $C^\perp$), using a discrete version of the Bonnet-Myers theorem from Riemannian geometry [Oll09]. Since $\mathbb{T}$ is a regular graph of degree $n$, an upper bound on its diameter implies a bound on $|\mathbb{T}|$, and hence on $|C|$.

**Remark 1.8:** While our approach is "curvature based", most of the bounds on local codes in the literature are based on isoperimetric inequalities or their information theoretic versions [KT00, KdW04, Woo07]. In Riemannian geometry the notions of curvature and isoperimetry are closely related. Better isoperimetric inequalities for graphs with "positive discrete curvature" are known in the discrete setting as well [LY10, BHL$^+$15, KKRT15]. It seems natural to ask whether this connection might be exploited in order to improve coding bounds. ▮

## 2 The Main Technical Lemma

Our bounds are based on the following key lemma.

**Lemma 2.1** *Let $u_1, ..., u_m$ and $v_1, ..., v_n$ be correspondingly the rows and the columns of an $m \times n$ matrix over $\mathbb{F}_2$. Let $V = \mathrm{span}(u_1, ..., u_m)$. Suppose that for each $i = 1, ..., n$ with $v_i \neq 0$ there are at least $K > 0$ disjoint pairs of indices $\{j, l\}$ such that $i \notin \{j, l\}$ and $v_i = v_j + v_l$. Then, for the coset leader graph $\mathbb{T} = \{0,1\}^n / V^\perp$ of $V$ holds*

$$\mathrm{diam}(\mathbb{T}) \leq \frac{n}{K+1}.$$

We observe that this implies a bound on the dimension of $V$.

**Corollary 2.2:**

$$\dim V \leq \log_2 \left( \sum_{i=0}^{\mathrm{diam}(\mathbb{T})} \binom{n}{i} \right) \leq \frac{n \log(K+1) + n/\ln 2}{K+1}$$

.

**Proof:** (of Corollary 2.2)

Recall that $\mathbb{T}$ is an $n$-regular Cayley graph of an Abelian group. Hence

$$|V| = |\mathbb{T}| \leq \sum_{i=0}^{\mathrm{diam}(\mathbb{T})} \binom{n}{i} \leq 2^{nH\left(\frac{1}{K+1}\right)} \leq 2^{\frac{n \log(K+1) + n/\ln 2}{K+1}}.$$

Here $H(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$ is the binary entropy function. For the second inequality recall that for any $0 \leq k \leq r$ holds $\sum_{i=0}^{k} \binom{r}{i} \leq 2^{rH\left(\frac{k}{r}\right)}$ (Theorem 1.4.5. in [vL99]). For the third inequality, note that $(1-x) \ln \frac{1}{1-x} \leq x$, for $0 \leq x < 1$. ∎

**Corollary 2.3:** *The case $q = 2$ of Theorem 1.2 holds.*

**Proof:** (of Corollary 2.3)

Let $\mathbb{T}$ be the coset leader graph of a locally correctable code $C$ with parameters 2 and $\delta$. By definition, a generating matrix of $C$ satisfies the assumptions of Lemma 2.1 with $K = \delta n$, and hence $\mathrm{diam}(\mathbb{T}) \leq 1/\delta$ and $\dim(C) \leq O\left(\frac{\log n}{\delta}\right)$. Since it is easy to see that the diameter of $\mathbb{T}$ is precisely the covering radius of $C^\perp$, this proves the case $q = 2$ of Theorem 1.2. ∎

## 2.1 Examples

**Example 2.4:** Let $m$ be a positive integer and let $n = 2^m - 1$. The generating matrix of the *Hadamard code* $C$ of length $n$ [MS77] is the $m \times n$ matrix whose columns are all the non-zero vectors in $\mathbb{F}_2^m$. Let $\mathbb{T}$ be the coset leader graph of $C$. Since the columns of the generating matrix are non-zero and distinct, $\mathbb{T}$ is a simple $n$-regular graph with $2^m = n + 1$ vertices, namely it is the complete graph on $n + 1$ vertices. In this case the assumptions of Lemma 2.1 hold with $K = (n - 1)/2$, and it gives the tight bounds $\mathrm{diam}(\mathbb{T}) \leq \lfloor \frac{2n}{n+1} \rfloor = 1$ and $|\mathbb{T}| \leq n + 1$. ▮

**Example 2.5:** Let $C$ be the direct product of two Hadamard codes. That is, assume $n = 2 \cdot (2^m - 1)$, and let the generating matrix of $C$ be a $2m \times n$ block-diagonal matrix with two $m \times (n/2)$ blocks whose columns are all the non-zero vectors in $\mathbb{F}_2^m$. In this case $\mathbb{T}$ is the Cartesian product of two complete graphs on $n/2 + 1$ vertices. That is, $|\mathbb{T}| = \frac{(n+2)^2}{4}$ and $\mathrm{diam}(\mathbb{T}) = 2$. The conditions of the lemma hold with $K = (n - 2)/4$, leading to an upper bound of 3 on the diameter and of $\binom{n}{3} + \binom{n}{2} + n + 1$ on the cardinality of $\mathbb{T}$. ▮

In the remainder of this section we proceed as follows. We start with comparing Lemma 2.1 to related results in the literature. Next, we describe the key notion of discrete curvature on graphs. Finally, we prove the lemma in Section 2.4.

## 2.2 Lemma 2.1 and related results

In this subsection we expand on Remark 1.8. Upper bounds on locally correctable codes in the literature follow from bounds on locally decodable codes. (Our approach applies directly to locally correctable codes, which might explain its relative simplicity.) A typical approach uses isoperimetric inequalities. [GKST06, Woo07, Woo12] use a weighted version of the edge-isoperimetric inequality on the boolean cube ([GKST06]). Another version of the edge-isoperimetric inequality is proved and used in [BSV12]. We compare Lemma 2.1 and Corollary 2.2 with these two results, which we restate in our language.

**Lemma 2.6** ([GKST06, Lemma 3.3])**:** *Let $u_1, ..., u_m$ and $v_1, ..., v_n$ be, correspondingly, the rows and the columns of an $m \times n$ matrix over $\mathbb{F}_2$. Let $V = \mathrm{span}(u_1, ..., u_m)$ and assume that $v_1, ..., v_{\mathrm{dim} V}$ span the column space. Suppose that for each $i = 1, ..., \dim V$ there are $K_i > 0$ disjoint pairs of indices $\{j, l\}$ such that $i \notin \{j, l\}$ and $v_i = v_j + v_l$. Let $K = \left(\sum_{i=1}^{\dim V} K_i\right) / \dim V$. Then*

$$\dim V \leq \frac{n \log n}{2K}.$$

**Lemma 2.7** ([BSV12, Lemma 3.15])**:** *Let $u_1, ..., u_m$ and $v_1, ..., v_n$ be, correspondingly, the rows and the columns of an $m \times n$ matrix over $\mathbb{F}_2$. Let $V = \mathrm{span}(u_1, ..., u_m)$ and assume that $v_1, ..., v_{\mathrm{dim} V}$ span the column space. Suppose that for each $i = 1, ..., \dim V$ there are $K > 0$ disjoint pairs of indices $\{j, l\}$ such that $i \notin \{j, l\}$ and $v_i = v_j + v_l$. Then*

$$\dim V \leq \frac{n \log K + n}{K}.$$

We collect the assumptions and the conclusions of the three claims in the following table (omitting constants for readability):

| | Lemma 2.1 + Cor. 2.2 | Lemma 2.6 | Lemma 2.7 |
|---|---|---|---|
| Disjoint rep's for: | all columns | basis | basis |
| At least $K$ rep's for: | all columns | average basis column | all basis columns |
| Dimension at most | $\frac{n \log K}{K}$ | $\frac{n \log n}{K}$ | $\frac{n \log K}{K}$ |
| Diameter at most | $\frac{n}{K}$ | $\frac{n \log n}{K}$ | $\frac{n \log K}{K}$ |

This table requires some reading help, which we provide here. The first two rows present the assumptions, and the last two the bounds. In this context, a family of disjoint representations of a column $i$ is a collection of disjoint pairs of indices $\{j, l\}$ such that $i \notin \{j, l\}$ and $v_i = v_j + v_l$.

The first row specifies whether such family is assumed to exist for all column vectors $v_1, \ldots, v_n$ or only for the basis $v_1, \ldots, v_{\dim V}$. The second row indicates whether the lower bound $K$ is on the minimal or the average size of a family (over the relevant coordinates).

The third row bounds the dimension of $V$. The fourth row bounds the diameter of the coset leader graph $\mathbb{T} = \{0,1\}^n / V^{\perp}$, which is the same as the covering radius of $V$. Since lemmas 2.6 and 2.7 do not consider the covering radius, we have filled out the corresponding entries using the fact that the covering radius of a linear code is upper bounded by its dimension [CHLL97, Theorem 2.1.9].

The next example shows that in Lemma 2.1 the assumption on minimal family size cannot be replaced by that on average family size, without affecting the bounds.

**Example 2.8:** For an integer $m$, let $n = 2^m - 1$, and let $A$ be the generating matrix of the Hadamard code (Example 2.4) with $m$ rows and $n$ columns. Let $I_m$ be the $m \times m$ identity matrix. Consider a linear code defined by the following $(2m) \times (n + m)$ generating matrix:

$$\left(\begin{array}{c|c} A & 0 \\ \hline 0 & I_m \end{array}\right)$$

In this case, the average family size is linear in $n$ (but the minimal family size is zero). The coset leader graph is the Cartesian product of the complete graph on $n + 1$ vertices with the $m$-dimensional discrete cube. Hence its diameter is logarithmic in $n$ (as opposed to constant). ▮

## 2.3 Discrete Curvature on Graphs

There are several possible ways to extend the notion of Ricci curvature from Riemannian geometry to the general setting of metric spaces and, in particular, graphs [Cha96, Oll09, Pet11, BJL12]. We use the approach of Ollivier [Oll09, OV12]. In the following discussion $G$ is a finite multigraph with a probability measure $m_x$ on its vertex set $V = V(G)$ assigned to each vertex $x$. We denote by $d$ the graph (shortest path) metric defined by $G$ on $V$.

Recall that the *transportation distance* between two probability measures $\mu$ and $\nu$ on $V$ is defined as

$$W_1(\mu, \nu) = \min_q \sum_{(x', y') \in G \times G} q(x', y') d(x', y') \tag{1}$$

where the minimum is taken over all probability measures $q$ on the product space $V \times V$ whose marginals are $\mu$ and $\nu$.

**Definition 2.9:** Let $x \neq y \in V$. The *coarse Ricci curvature* $\kappa(x, y)$ along $(x, y)$ is

$$\kappa(x, y) = 1 - \frac{W_1(m_x, m_y)}{d(x, y)} \tag{2}$$

▮

A canonical choice for the measure $m_x$ is the uniform probability measure on the metric ball of radius 1 around $x$. In this case the coarse Ricci curvature $\kappa(x, y)$ along $(x, y)$ is positive if and only if the mean distance between the metric balls around $x$ and $y$ (as measured by $W_1(m_x, m_y)$) is smaller than the distance between $x$ and $y$. This conforms to the intuition that in spaces with positive curvature metric balls are closer on average than their centers (and vice versa for spaces with negative curvature).

The *curvature* $\kappa(G)$ of the *graph* $G$ is defined as the minimum of $\kappa(x, y)$ over all pairs of vertices. This minimum is attained on a pair of adjacent vertices [Oll09, Proposition 19].[2] In particular, curvature is a local property.

The key claim we need is the following discrete version of the Bonnet-Myers theorem for Riemannian manifolds[3] [Oll09, Proposition 23]. For $x \in V$, let $\delta_x$ be the probability measure concentrated on $x$, and let $J(x) = W_1(\delta_x, m_x)$. (E.g., if $G$ is a simple $n$-regular graph, and $m_x$ is the uniform measure on the metric ball of radius 1 around $x$, then $J(x) = \frac{n}{n+1}$.)

**Proposition 2.10:** ([Oll09, Proposition 23]) *For any graph $G$ and a family of probability measures $\{\mu_x\}_{x \in V(G)}$ holds*

$$\operatorname{diam}(G) \leq \frac{2 \cdot \max_{x \in G} J(x)}{\kappa(G)}.$$

**Remark 2.11:** Many of these ideas appear also in the theory of random walks on graphs, see [LPW09], especially chapter 14. ▮

[2] This, and Proposition 2.10 below are simple consequences of the triangle inequality for the transportation distance.

[3] The classical Bonnet-Myers theorem for Riemannian manifolds states that if the Ricci curvature of an $n$-dimensional complete Riemannian manifold $M$ is at least $(n-1)\kappa > 0$, then the manifold is compact and its diameter is at most $\pi/\sqrt{\kappa}$.

### 2.4 Proof of Lemma 2.1

In this section we prove Lemma 2.1. We choose a family of probability measures $\{\mu_x\}_{x\in U}$ on the vertex set $U$ of $\mathbb{T}$ which enables us to bound the coarse Ricci curvature on $\mathbb{T}$ from below, and then apply Proposition 2.10.

Recall that $\mathbb{T}$ is an $n$-regular multigraph. For $x \in U$, we define $m_x$ to be the measure induced by the uniform measure on the edges incident to $x$. That is, for $y$ adjacent to $x$ we set $m_x(y)$ to be the number of edges between $x$ and $y$, divided by $n+1$; and we let $m_x(x)$ be the number of loops at $x$ *plus one*, divided by $n+1$. The measure $m_x$ is supported on the metric ball of radius 1 around $x$ and, if $\mathbb{T}$ is a simple graph, then $m_x$ is uniform on this set.

We observe that the local structure of $\mathbb{T}$ at (any) vertex $x$, and hence the measure $m_x$, can be described in terms of the column vectors $v_1, \dots, v_n$. In fact, the number of loops at any vertex of $\mathbb{T}$ equals to the number of zero vectors among $v_1, \dots, v_n$. Similarly, the number of edges between two distinct vertices $x$ and $x + e_i$ (the addition is in the factor group $\mathbb{F}_2^n/V^\perp$) is the number of times $v_i$ appears as a column vector.

Next, we upper bound the transportation distance between measures $m_x$ and $m_y$, for distinct adjacent vertices $x$ and $y$. Let $y = x + e_i$ for some $1 \le i \le n$. By the assumption of the lemma, there are some $M \ge K$ disjoint pairs of indices $\{j,l\} \subseteq [n] \setminus \{i\}$ such that $v_j + v_l = v_i$. Equivalently, $e_i + e_j + e_l \in V^\perp$ and hence

$$x + e_j = y + e_l \qquad \text{and} \qquad x + e_l = y + e_j$$

This means that the points $x + e_j$ and $x + e_l$ belong to the supports of both $m_x$ and $m_y$, and we have identified an overlap between the two measures, of weight $\frac{2}{n+1}$ in each measure. Going over all the $M$ representations $v_j + v_l = v_i$ produces an overlap of weight $\frac{2M}{n+1}$ in each of the measures.

The identity $y = x + e_i$ gives an additional overlap of at least $\frac{2}{n+1}$ between the measures. This brings the total overlap to at least $\frac{2M+2}{n+1}$.

We now transport $m_x$ to $m_y$ as follows. The points in the joint support stay in place. All the remaining mass in $m_x$ is moved by a unit distance in parallel. That is, we move a point $z$ in the unit ball around $x$ to the point $z + e_i$ in the unit ball around $y$.

Computing the total amount of work gives

$$W_1(m_x, m_y) \le \frac{(n+1) - (2M+2)}{n+1} \le \frac{(n+1) - (2K+2)}{n+1}.$$

Hence, by (2), the coarse Ricci curvature along $(x, y)$ is at least $\frac{2(K+1)}{n+1}$. Since this holds for any adjacent pair of vertices, we have $\kappa(\mathbb{T}) \ge \frac{2(K+1)}{n+1}$.

Applying Proposition 2.10 (note that $J(x) \le \frac{n}{n+1}$ for all $x \in U$) gives

$$\operatorname{diam}(\mathbb{T}) \le \frac{\frac{2n}{n+1}}{\kappa(\mathbb{T})} \le \frac{n}{K+1},$$

concluding the proof of the lemma. ▮

# 3 Bounds on Locally Correctable Codes

## 3.1 Proof of Theorem 1.2

The case $q = 2$ of the theorem is treated in Corollary 2.3. In this section we deal with larger values of $q$.

Let $C$ be a locally correctable code with parameters $q > 2$ and $\delta$. Fix a generating matrix of $C$ and let its columns be $v_1, \ldots, v_n$. Let $N = \{i \mid v_i \neq 0\}$. By definition, for each coordinate $i \in N$, there is a family $M_i$ of at least $\delta n$ disjoint $q$-subsets of $[n] \setminus \{i\}$ such that the vectors indexed by each subset sum to $v_i$.

Our argument works (essentially) by reduction to the base case $q = 2$. Let us start with a quick overview. We will show that there is a subset $B$ of $[n]$ such that for any $i \in N \setminus B$ there are many $q$-tuples $\alpha \in M_i$ with $|\alpha \setminus B| \leq 2$. Dividing out by the vector space spanned by the columns in $B$ will produce a code whose generating matrix satisfies the conditions of Lemma 2.1, with a parameter $K$ related to the parameters of the original code. Applying Lemma 2.1 and Corollary 2.2 will complete the proof. Let us mention that this approach is similar to that in [Woo07] and [DK11].

**Lemma 3.1:** *Let $q > 2$. For each $1 \leq a \leq (\log n)^{1/(q-1)}$ there exists a subset $B \subseteq [n]$, such that:*

- $|B| \quad \leq \quad \left(a + \frac{4}{\delta a^{q-2}}\right) \cdot n^{\frac{q-2}{q-1}}$
- *For every $i \in N \setminus B$ holds* $\quad \left|\left\{\alpha \in M_i \ : \ |\alpha \setminus B| \leq 2\right\}\right| \quad \geq \quad \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}}$

**Proof:**

Set $\theta = a \cdot n^{-\frac{1}{q-1}}$ and observe that $0 < \theta < 1$. We construct a random subset $B \subseteq [n]$ satisfying the assertions of the lemma in two steps. In the first step we add to $B$ elements in $[n]$ chosen independently at random with probability $\theta$. With high probability, this will produce a set of cardinality about $n\theta = a \cdot n^{\frac{q-2}{q-1}}$, satisfying the second claim of the lemma for all but a small number of indices $i \in N$. In the second step we will add to $B$ all these exceptional indices and in this way ensure that both claims of the lemma hold.

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables, $X_j = \begin{cases} 1, & \text{w.p.} \quad \theta; \\ 0, & \text{w.p.} \quad 1 - \theta. \end{cases}$

Let $B_0 = \{j : X_j = 1\}$. For a $q$-subset $\alpha \subseteq [n]$, let $W_\alpha$ be indicator of the event $|B_0 \cap \alpha| \geq q - 2$.

For $i \in N$, let $Y_i = \sum_{\alpha \in M_i} W_\alpha$. Since the $q$-tuples in $M_i$ are disjoint, the random variables $\{W_\alpha\}_{\alpha \in M_i}$ are independent, and hence $Y_i$ is a binomial random variable with parameters $|M_i|$ and $\eta = Pr\,(W_\alpha = 1) > \theta^{q-2}$. In particular, $\mathbb{E}(Y_i) = |M_i| \cdot \eta > \delta n \theta^{q-2} = \delta a^{q-2} \cdot n^{\frac{1}{q-1}}$.

Hence, by Chebyshev's inequality,

$$\Pr\left(Y_i < \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}}\right) \leq \Pr\left(Y_i < \frac{\mathbb{E}(Y_i)}{2}\right) \leq \frac{\mathrm{var}(Y_i)}{(\mathbb{E}(Y_i)/2)^2} \leq \frac{4}{\mathbb{E}(Y_i)} \leq \frac{4}{\delta a^{q-2} \cdot n^{\frac{1}{q-1}}}$$

Let $B = B_0 \bigcup \left\{ i \in N \ : \ Y_i < \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}} \right\}$.

By the definition of $B$, for all $i \in N \setminus B$ holds

$$\left| \left\{ \alpha \in M_i \ : \ |\alpha \setminus B| \leq 2 \right\} \right| \geq \left| \left\{ \alpha \in M_i \ : \ |\alpha \setminus B_0| \leq 2 \right\} \right| = Y_i \geq \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}}$$

Therefore, $B$ satisfies the second claim of the lemma. To verify that for some choice of $B$ the first claim holds as well, we upperbound the expectation of $|B|$ appropriately.

$$\mathbb{E}(|B|) = \mathbb{E}(|B_0|) + \mathbb{E}(|B \setminus B_0|) \leq n\theta + \sum_{i \in N} \Pr\left( Y_i < \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}} \right) \leq \left( a + \frac{4}{\delta a^{q-2}} \right) \cdot n^{\frac{q-2}{q-1}}$$

∎

We proceed with the proof of Theorem 1.2. Let $a$ be a parameter in the interval $\left[1, (\log n)^{1/(q-1)}\right]$ (we will optimize over the value of $a$ later on). Let $B = B(a) \subseteq [n]$ be the subset of indices given by Lemma 3.1. Let $U = U(B) = \mathrm{Span}(\{v_i \ : \ i \in B\})$. Let $C_B$ be the subcode of $C$ containing the vectors in $C$ which vanish on $B$. Let $\mathbb{T} = \mathbb{F}_2^n / C^{\perp}$ and $\mathbb{T}_B = \mathbb{F}_2^n / C_B^{\perp}$ be the coset leader graphs of $C$ and $C_B$ respectively. Then the following holds.

**Lemma 3.2:**

- *Any generating matrix of $C_B$ satisfies the conditions of Lemma 2.1 with $K = \frac{\delta}{2} a^{q-2} \cdot n^{\frac{1}{q-1}}$.*
- $\dim C = \dim C_B + \dim U$.
- $\mathrm{diam}(\mathbb{T}) \leq \mathrm{diam}(\mathbb{T}_B) + \dim U$.

**Proof:** We start with the first claim. Let $G_B$ be a generating matrix of $C_B$. Note that the preceding discussion, and in particular the choice of the set $B$, has been independent of the generating matrix of $C$ we have chosen, and hence we may assume that $G_B$ is a row submatrix of this generating matrix, which we will denote by $G$. Let $u_1, ..., u_n$ be the columns of $G_B$. Then $u_i$ is a restriction of $v_i$ to a (fixed) subset of coordinates for all $1 \leq i \leq n$.

Let $u_i$ be a non-zero column of $G_B$. We need to show that there are at least $K$ disjoint pairs of indices $\{j, l\}$ with $u_i = u_j + u_l$. First, note that $i \in N \setminus B$. Indeed, $u_i$ is zero for $i \in B$, by the definition of $C_B$, and $v_i$ (and hence $u_i$) is zero for $i \notin N$, by the definition of $N$.

Since $i \in N \setminus B$, there are at least $K$ disjoint $q$-tuples $\alpha \in M_i$ with $|\alpha \setminus B| \leq 2$. We will find a coordinate pair $\{j, l\}$ with $u_i = u_j + u_l$ contained in each of these tuples, and this will complete the argument. Fix $\alpha$. By definition, $\sum_{k \in \alpha} v_k = v_i$, implying $\sum_{k \in \alpha} u_k = u_i$. Since $u_s = 0$ for $s \in B$, this means $\sum_{k \in \alpha \setminus B} u_k = u_i$. Since $u_i \neq 0$, the set $\alpha \setminus B$ is not empty. If $|\alpha \setminus B| = 2$, take $\{j, l\} = \alpha \setminus B$. If $|\alpha \setminus B| = 1$, take $j$ to be the unique element of $\alpha \setminus B$, and $l$ any element of $\alpha \cap B$.

The second claim is a well-known fact in linear algebra. We provide a brief argument for completeness. Right multiplication by $G$ defines an isomorphism between $\mathbb{F}_2^{\dim C}$ and $C$. The claim is implied by the observation that the pre-image of $C_B$ under this isomorphism is precisely $U^\perp$.

We pass to the third claim. Since the diameter of the coset graph of a code equals to the covering radius of the dual code, the claim is that the covering radius of $C^\perp$ is upper bounded by the covering radius of $C_B^\perp$ plus the dimension of $U$. We will show this by finding, for each vector $x \in C_B^\perp$, a vector $y \in C^\perp$ such that $|x - y| \leq \dim U$. Observe that $C_B^\perp = \Big\{x \in \{0,1\}^n, \ \sum_{i=1}^n x_i v_i \in U\Big\}$. Let $x \in C_B^\perp$, and let $\sum_{i=1}^n x_i v_i = u \in U$. The vector $u$ can be written as a linear combination of columns in $B$, of length at most $\dim U$. Let $z \in \{0,1\}^n$ be the characteristic vector of this linear combination. Then $|z| \leq \dim U$ and $y = x + z \in C^\perp$, completing the proof.

▮

Now we are ready to complete the proof of Theorem 1.2. To bound the covering radius of $C^\perp$, which is the same as the diameter of $\mathbb{T}$, take $a = 1$. This gives $|B| = \left(1 + \frac{4}{\delta}\right) \cdot n^{\frac{q-2}{q-1}}$ in Lemma 3.1 and $K = \frac{\delta}{2} \cdot n^{\frac{1}{q-1}}$ in Lemma 3.2. By Lemma 2.1, $\mathrm{diam}(\mathbb{T}_B) \leq \frac{n}{K+1}$, and hence

$$\mathrm{diam}(\mathbb{T}) \leq \mathrm{diam}\mathbb{T}_B + \dim U \leq \frac{n}{K+1} + |B| \leq O\left(n^{\frac{q-2}{q-1}}\right).$$

To bound the dimension of $C$, take $a = (\log n)^{\frac{1}{q-1}}$. This gives $|B| \approx n^{\frac{q-2}{q-1}} (\log n)^{\frac{1}{q-1}}$ and $K = \frac{\delta}{2} n^{\frac{1}{q-1}} (\log n)^{\frac{q-2}{q-1}}$. By Corollary 2.2, $\dim C_B \leq O\left(n^{\frac{q-2}{q-1}} (\log n)^{\frac{1}{q-1}}\right)$, and hence,

$$\dim C \leq \dim C_B + \dim U \leq \dim C_B + |B| \leq O\left(n^{\frac{q-2}{q-1}} (\log n)^{\frac{1}{q-1}}\right).$$

### 3.2 Proof of Theorem 1.4

Let $\mathbb{T} = \{0,1\}^n / C^\perp$ be the coset leader graph of $C$. We will show that the neighborhoods of (any) vertex in $\mathbb{T}$ grow rather slowly, which will imply that $\mathbb{T}$, and hence $C$, are not too large. For $r \geq 0$, let $S_r^{\mathbb{T}}$ be the sphere of radius $r$ around $C^\perp$ in $\mathbb{T}$. The key observation is that there are many edges in $\mathbb{T}$ between the consecutive spheres $S_{r-1}^{\mathbb{T}}$ and $S_r^{\mathbb{T}}$.

**Lemma 3.3:** *Let $r \geq 2$. Assume that $S_r^{\mathbb{T}}$ is not empty. Then there are at least $(\lfloor r/2 \rfloor)^2$ edges between any vertex $x + C^\perp \in S_r^{\mathbb{T}}$ and $S_{r-1}^{\mathbb{T}}$.*

**Remark 3.4:** This should be compared to the situation in the discrete cube $\{0,1\}^n$, also an $n$-regular graph, in which a vertex at distance $r$ from zero is connected to the sphere of radius $r - 1$ around zero by exactly $r$ edges. ▮

Before proving the lemma, let us show that it implies the claim of the theorem. By the lemma, there are at least $(\lfloor r/2 \rfloor)^2 \cdot |S_r^{\mathbb{T}}|$ edges between $S_{r-1}^{\mathbb{T}}$ and $S_r^{\mathbb{T}}$. On the other hand, $\mathbb{T}$ is an $n$-regular graph, which means that there are at most $n \cdot |S_{r-1}^{\mathbb{T}}|$ such edges. Hence $(\lfloor r/2 \rfloor)^2 \cdot |S_r^{\mathbb{T}}| \leq n \cdot |S_{r-1}^{\mathbb{T}}|$, and this holds for any $r \geq 2$.

The sphere of radius 1 is of cardinality at most $n$. Multiplying consecutive inequalities provides an upper bound on the cardinality of a sphere of radius $r \geq 2$:

$$\left|S_r^{\mathbb{T}}\right| \quad \leq \quad n \cdot \prod_{t=2}^{r} \frac{n}{(\lfloor t/2 \rfloor)^2} \quad \leq \quad \left(\frac{cn}{r^2}\right)^r$$

for an appropriate constant $c > 0$. The second inequality can be deduced e.g., from Stirling's formula. It is easy to see that this implies $|\mathbb{T}| = \sum_r \left|S_r^{\mathbb{T}}\right| \leq c^{\sqrt{n}}$, for a (possibly different) constant $c$, completing the proof of the theorem.

**Proof:** (of Lemma 3.3).

By assumption, $C$ is a perfect 3-locally correctable code. This means that $n$ is 1 modulo 3, and that for all $1 \leq i \leq n$ there is a family $M_i$ of $\frac{n-1}{3}$ disjoint 3-tuples partitioning $[n] \setminus \{i\}$, so that for any such 3-tuple $\alpha$ holds $e_i + \sum_{j \in \alpha} e_j \in C^{\perp}$. That is, for any two indices $i < j$ there is a unique pair of indices $k \neq l$ such that $(j, k, l) \in M_i$. In particular, $\{i, j\} \cap \{k, l\} = \emptyset$, and $e_i + e_j + e_k + e_l \in C^{\perp}$.

Let now $x + C^{\perp} \in S_r^{\mathbb{T}}$. We may assume that $x$ is of minimal weight in its coset, meaning that the Hamming weight of $x$ is $r$. We will also assume, for simplicity, that $x_i = 1$ for $1 \leq i \leq r$ (and $x_i = 0$ for $i > r$).

For two indices $i, j$ with $1 \leq i < j \leq r$, let $k, l$ be such that $(j, k, l) \in M_i$. Note that this necessarily means that $x_k = x_l = 0$ (that is $k, l > r$). Indeed, otherwise $x' = x + e_i + e_j + e_k + e_l$ would be a vector in $x + C^{\perp}$ of weight smaller than $r$. The key point for us is that the edges from $x + C^{\perp}$ in the directions $k, l$ lead down to $S_{r-1}^{\mathbb{T}}$. In fact, the vector $x + e_k = x + e_i + e_j + e_l$ is of weight $r - 1$ (and similarly for $x + e_l$).

Let $V \subseteq [n]$ contain all directions leading from $x$ down to $S_{r-1}^{\mathbb{T}}$. As we have seen, each pair of indices $i, j$ with $1 \leq i < j \leq r$ defines a pair $(k, l) \in V \times V$, which we interpret as an edge with vertices in $V$. From now on we assume, for simplicity, that $r$ is even. Going over $i, j$ with $1 \leq i \leq r/2 < j \leq r$ defines a multigraph $G$ on $V$ with $r^2/4$ edges. In fact, we claim that $G$ is a simple graph, that is distinct pairs $i, j$ and $i_1, j_1$ define distinct edges $(k, l)$ and $(k_1, l_1)$. Indeed, otherwise $e_i + e_j + e_{i_1} + e_{j_1} \in C^{\perp}$, which means that $x' = x + e_i + e_j + e_{i_1} + e_{j_1}$ is a vector in $x + C^{\perp}$ of weight smaller than $r$.

Next, we claim that $G$ is a disjoint union of stars. This would mean that the number of vertices of $G$ is larger than its number of edges, i.e., $|V| > r^2/4$, proving Lemma 3.3. This claim is a simple corollary of the following auxiliary lemma.

**Lemma 3.5:** *Any edge of $G$ contains a vertex of degree* 1.

**Proof:** Assume to the contrary that there exists an edge $(k, l)$ in $G$ such that both $k$ and $l$ have degree at least 2. There are two possible cases. Either $G$ contains a simple path $k_1 \to k \to l \to l_1$ of length 4, or $G$ contains a triangle with vertices $k, l, m$. Consider the first case. Let $(i, j)$, $1 \leq i \leq r/2 < j \leq r$, be the pair of indices defining the first edge of the path, let $(i_1, j_1)$ define the second edge, and $(i_2, j_2)$ the third edge. We claim that $i \neq i_1$. Indeed, otherwise both

$(j, k_1, k)$ and $(j_1, k, l)$ would be in $M_i$, contradicting the fact that $M_i$ is a family of disjoint triples. Next, we claim that $j = j_1$. If not, we would have

$$\Big(e_i + e_{i_1} + e_j + e_{j_1}\Big) + \Big(e_{k_1} + e_l\Big) \quad = \quad \Big(e_i + e_j + e_{k_1} + e_k\Big) + \Big(e_{i_1} + e_{j_1} + e_k + e_l\Big) \in C^{\perp},$$

which would give us a vector $x' = x + \Big(e_i + e_{i_1} + e_j + e_{j_1}\Big) + \Big(e_{k_1} + e_l\Big)$ in $x + C^{\perp}$ of weight smaller than $r$.

A similar argument shows that $i_1 \neq i_2$ and $j_1 = j_2$ (and therefore also $j = j_2$). We now observe that $i$ and $i_2$ also have to be distinct. Indeed, otherwise we would have both $(j, k_1, k)$ and $(j, l, l_1)$ in $M_i$.

Taking everything into account, this means that

$$\Big(e_i + e_{i_1} + e_{i_2} + e_j\Big) + \Big(e_{k_1} + e_{l_1}\Big) \quad = \quad \Big(e_i + e_j + e_{k_1} + e_k\Big) + ... + \Big(e_{i_2} + e_{j_2} + e_l + e_{l_1}\Big) \in C^{\perp},$$

giving a vector $x' = x + \Big(e_i + e_{i_1} + e_{i_2} + e_j\Big) + \Big(e_{k_1} + e_{l_1}\Big)$ in $x + C^{\perp}$ of weight smaller than $r$, and in this way reaching a contradiction.

The second case of the lemma is similar (but simpler). We omit the analysis. This completes the proof of Lemma 3.5 and of Lemma 3.3.

∎

## 4 Bounds on Locally Testable Codes

In this section we prove Theorems 1.5 and 1.6. The proofs of both theorems are based on the following lemma.

**Lemma 4.1:** *Let $G$ be a matrix satisfying the assumptions of Theorem 1.6. Then $G$ satisfies the assumptions of Lemma 2.1 with $K = \lceil \sigma/p \rceil$.*

**Proof:** Let $t$ be the number of distinct columns of $G$ and assume, without loss of generality, that $v_1, \ldots, v_t$ are pairwise distinct. That is, the first $t$ columns represent all the distinct columns in $G$. For $1 \leq i \leq t$, let $w_i$ denote the multiplicity of $v_i$ in $G$. Note that $\sum_{i=1}^{t} w_i = n$. We may, and will, assume that $w_1 \leq \ldots \leq w_t = p$. For $1 \leq i \leq t$ with $v_i \neq 0$, let $N_i = \{(j,k) : \ 1 \leq j < k \leq t, \ v_i = v_j + v_k\}$.

Fix an index $1 \leq i \leq n$ with $v_i \neq 0$. We need to show that there are at least $K = \sigma/p$ disjoint pairs of indices $\{r, s\}$ such that $i \notin \{r, s\}$ and $v_i = v_r + v_s$. It suffices to show this for any of the copies of $v_i$ in $G$, and so we may assume $1 \leq i \leq t$.

Assume first that $G$ has no zero columns. In this case we claim that $v_i$ participates in exactly $\sum_{(j,k) \in N_i} w_j w_k$ dependencies of length three. Indeed, each pair $(j,k) \in N_i$ contributes $w_j w_k$ dependencies, obtained by taking $v_i$ together with any copy of $v_j$ and any copy of $v_k$. On the other hand, every dependency is of this form. Hence, by assumption, $\sum_{(j,k) \in N_i} w_j w_k \geq \sigma$.

Next, we note that any pair $(j,k)$ in $N_i$ contributes $w_j$ disjoint pairs of indices $\{r,s\}$ such that $i \notin \{r,s\}$ and $v_r + v_s = v_i$, obtained by making $v_r$ go over all the copies of $v_j$ in $G$ and matching each $v_r$ with a distinct copy of $v_k$. Here we use the fact that $w_j \leq w_k$. Moreover, these collections of indices are disjoint for different choices of $(j,k) \in N_i$. Altogether this gives

$$\sum_{(j,k)\in N_i} w_j \geq \frac{1}{p} \cdot \sum_{(j,k)\in N_i} w_j w_k \geq \frac{\sigma}{p}$$

such pairs, proving the lemma in this case. For the first inequality, recall that all $w_k$ are bounded from above by $p$.

If $G$ has zero columns, let $1 \leq z \leq t$ be the index with $v_z = 0$. Compared to the previous case, we have $(w_i - 1) \cdot w_z$ additional dependencies of length 3 for $v_i$, obtained by choosing any of the extra copies of $v_i$ together with $v_i$ itself and with any copy of $v_z$. So, in this case the total number of dependencies is $(w_i - 1) \cdot w_z + \sum_{(j,k)\in N_i} w_j w_k$, and this, by assumption, is at least $\sigma$.

On the other hand, we get $\min\{w_i - 1, w_z\}$ additional disjoint pairs of indices $\{r,s\}$ such that $i \notin \{r,s\}$ and $v_r + v_s = v_i$, by matching as many distinct copies of $v_i$ as possible (not counting $v_i$ itself) with distinct copies of $v_z$. Altogether, we get

$$\min\{w_i - 1, w_z\} + \sum_{(j,k)\in N_i} w_j \quad \geq \quad \frac{1}{p} \cdot \left( (w_i - 1) \cdot w_z + \sum_{(j,k)\in N_i} w_j w_k \right) \quad \geq \quad \frac{\sigma}{p}$$

such pairs, proving the lemma in this case as well.

■

The claim of Theorem 1.6 now follows directly by substituting $K = \lceil \sigma/p \rceil$ in Corollary 2.2.

We proceed with the proof of Theorem 1.5, using the notation of Lemma 4.1. We first note that since $C$ is a regular code, each column of $G$ has the same multiplicity $p$, implying $t = n/p$. In particular, the dimension of $C$ is at most $n/p$. Hence we may and will assume $\sigma > 4p^2$, since otherwise we are done.

Next, consider the coset leader graph $\mathbb{T} = \{0,1\}^n / V^\perp$, where $V$ is the row space of $G$. By Lemmas 2.1 and 4.1, the radius of $\mathbb{T}$ is at most $\frac{n}{\sigma/p+1} < \frac{np}{\sigma}$. The key point to observe is that while $\mathbb{T}$ is an $n$-regular multigraph, the edges of $\mathbb{T}$ corresponding to identical columns of $G$ are parallel to each other, and hence each vertex of $\mathbb{T}$ has precisely $t$ distinct neighbors. Proceeding as in the proof of Corollary 2.2, we have

$$|C| \ = \ |\mathbb{T}| \ \leq \ \sum_{i=0}^{\lfloor \frac{np}{\sigma} \rfloor} \binom{t}{i} \ \leq \ 2^{tH\left(\frac{np}{\sigma t}\right)}.$$

Substituting $t = n/p$, and setting $\alpha = \frac{\sigma}{p^2}$, we get

$$\frac{1}{n} \cdot \log_2 |C| \leq \frac{1}{p} H\left(\frac{p^2}{\sigma}\right) = \frac{1}{\sqrt{\sigma}} \cdot \sqrt{\alpha} H\left(\frac{1}{\alpha}\right)$$

To complete the proof, we will show that $\sqrt{\alpha} \cdot H(\frac{1}{\alpha}) < 2$, for all $\alpha \geq 1$. In fact,

$$\alpha \cdot H\left(\frac{1}{\alpha}\right) = \log_2(\alpha) + (\alpha - 1)\log_2\left(1 + \frac{1}{\alpha - 1}\right) \leq \frac{1}{\ln 2} \cdot \Big(\ln(\alpha) + 1\Big).$$

Hence $\sqrt{\alpha} \cdot H(\frac{1}{\alpha}) \leq \frac{1}{\ln 2} \cdot \frac{\ln \alpha + 1}{\sqrt{\alpha}}$. It remains to observe that the function $\frac{\ln \alpha + 1}{\sqrt{\alpha}}$ attains its maximum of $\frac{2}{\sqrt{e}} < 2\ln 2$ at $\alpha = e$. ▮

The next example shows that Theorem 1.6 is tight, up to the $\log\left(\lceil \sigma/p \rceil\right)$-factor.

**Example 4.2:** Let $m$ be a power of 2, and let $k \geq \log_2 m$ be integer. Let $U$ be a linear subspace of $\{0,1\}^k$ of dimension $\log_2 m$ with minimal distance at least 3. Let $u_1, ..., u_m$ be the vectors of $U$. Let $\mathbf{1}$ be the all-1 vector of length $k$, and let $B_i$ be the $k \times k$ matrix given by the outer product $u_i \otimes \mathbf{1}$. Finally, let $I$ be the $k \times k$ identity matrix.

Let $G$ be the following $k \times n$ matrix with $n = 2km$. The first $km$ columns of $G$ are formed by $m$ square blocks $I + B_i$, for $i = 1, ..., m$. The remaining $km$ columns are formed by the blocks $B_1, ..., B_m$.

Clearly the rows of $G$ are linearly independent, and therefore the dimension of the code $C$ it generates is $k$. By construction, for $G$ holds $p = k$ and $\sigma = km$ (since $U$ is a subspace). Hence we have

$$dim(C) = k = \frac{n}{2m} = \frac{n}{2\sigma/p}.$$

▮

## References

[BHL$^+$15] Frank Bauer, Paul Horn, Yong Lin, Gabor Lippner, Dan Mangoubi, and Shing-Tung Yau. Li-Yau inequality on graphs. *J. Differential Geom.*, 99(3):359–405, 2015.

[BJL12] Frank Bauer, Jürgen Jost, and Shiping Liu. Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator. *Math. Res. Lett.*, 19(6):1185–1205, 2012.

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

[Cha96] Ruth Charney. Metric geometry: connections with combinatorics. In *Formal power series and algebraic combinatorics (New Brunswick, NJ, 1994)*, volume 24 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 55–69. Amer. Math. Soc., Providence, RI, 1996.

[CHLL97] Gérard Cohen, Iiro Honkala, Simon Litsyn, and Antoine Lobstein. *Covering codes*, volume 54 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, 1997.

[DK11] Irit Dinur and Tali Kaufman. Dense locally testable codes cannot have constant rate and distance. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 507–518. Springer, Heidelberg, 2011.

[DSW14] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Breaking the quadratic barrier for 3-lcc's over the reals. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 784–793. ACM, 2014.

[FT05] Joel Friedman and Jean-Pierre Tillich. Generalized Alon-Boppana theorems and error-correcting codes. *SIAM J. Discrete Math.*, 19(3):700–718 (electronic), 2005.

[GKST06] Oded Goldreich, Howard Karloff, Leonard J. Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Comput. Complexity*, 15(3):263–296, 2006.

[KdW04] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. Comput. System Sci.*, 69(3):395–420, 2004.

[KKRT15] Bo'az Klartag, Gady Kozma, Peter Ralli, and Prasad Tetali. Discrete curvature and abelian groups. *arXiv preprint arXiv:1501.00516*, 2015.

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 80–86 (electronic), New York, 2000. ACM.

[KV10] Tali Kaufman and Michael Viderman. Locally testable vs. locally decodable codes. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 670–682. Springer, Berlin, 2010.

[LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[LY10] Yong Lin and Shing-Tung Yau. Ricci curvature and eigenvalue estimate on locally finite graphs. *Math. Res. Lett.*, 17(2):343–356, 2010.

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977. North-Holland Mathematical Library, Vol. 16.

[Oll09] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[OV12] Y. Ollivier and C. Villani. A curved Brunn-Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube? *SIAM J. Discrete Math.*, 26(3):983–996, 2012.

[Pet11] Anton Petrunin. Alexandrov meets Lott-Villani-Sturm. *Münster J. Math.*, 4:53–64, 2011.

[vL99] J. H. van Lint. *Introduction to coding theory*, volume 86 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 1999.

[Woo07] D. Woodruff. New lower bounds for general locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

[Woo12] David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *J. Comput. Sci. Tech.*, 27(4):678–686, 2012.