# Linear Codes and Character Sums

Nathan Linial *  Alex Samorodnitsky †

**Abstract**

Let $V$ be an $rn$-dimensional linear subspace of $Z_2^n$. Suppose the smallest Hamming weight of non-zero vectors in $V$ is $d$. (In coding-theoretic terminology, $V$ is a linear code of length $n$, rate $r$ and distance $d$.) We settle two extremal problems on such spaces.

First, we prove a (weak form) of a conjecture by Kalai and Linial and show that the fraction of vectors in $V$ with weight $d$ is exponentially small. Specifically, in the interesting case of a small $r$, this fraction does not exceed $2^{-\Omega(\frac{r^2}{\log(1/r)+1}n)}$.

We also answer a question of Ben-Or and show that if $r > \frac{1}{2}$, then for every $k$, at most $C_r \cdot \frac{|V|}{\sqrt{n}}$ vectors of $V$ have weight $k$.

Our work draws on a simple connection between extremal properties of linear subspaces of $Z_2^n$ and the distribution of values in short sums of $Z_2^n$-characters.

*Hebrew University, Jerusalem 91904, Israel. E-mail: nati@cs.huji.ac.il. Supported in part by grants from the Israeli Academy of Sciences and the Binational Science Foundation Israel-USA.

†Institute for Advanced Study, Princeton, NJ 08540. E-mail: asamor@ias.edu. This work was done while the author was a student in the Hebrew University of Jerusalem, Israel.

# 1 Introduction

This paper deals with extremal problems on linear codes. Our approach utilizes harmonic analysis on the cube $Z_2^n$. We observe that extremal questions about linear codes readily translate into questions about the possible value distributions of character sums over $Z_2^n$. To fix ideas, we quote a simple but useful lemma which illustrates this connection.

**Lemma 1.1:** *Let $V$ be an $m$-dimensional subspace of $Z_2^n$. Let $A$ be a generating matrix of $V$, namely an $m \times n$ matrix whose rows $v_1, \ldots, v_m$ are a basis of $V$. Let $u_1, \ldots, u_n$ be the columns of $A$, and let $f$ be the characteristic function of the (multi)set $\{u_1, \ldots, u_n\}$ (in the $m$-dimensional cube $Z_2^m$). For $B \subseteq \{1, \ldots, m\}$, the Hamming weight of $v = \bigoplus_{i \in B} v_i$ is $\frac{n - 2^m \hat{f}(B)}{2}$. ($\hat{f}$ being the Fourier Transform on $Z_2^m$, see next section for details.)*

We turn to the two problems that are addressed here, and formulate them in both coding-theoretic and harmonic-analytic terms. In [9] Kalai and Linial conjecture that the number of codewords of minimal weight in a linear code of length $n$ is subexponential in $n$. Recall that the characters on $Z_2^m$ are the Walsh functions $W_v$ with $v \in Z_2^m$, defined as $W_v(u) = (-1)^{<u,v>}$. Here is the translation of this question to the functional terminology: Let $f = \sum_{i=1}^n W_{u_i}$, where $u_i$ span $Z_2^m$. Is it true that the maximum of $f$ on $Z_2^m \setminus \{0\}$ is attained only at $\exp(o(n))$ points?

In Section 3 we prove a weaker statement, namely that the probability of $f$ attaining its maximum is exponentially small. Translating back to subspaces, we conclude that the fraction of non-zero vectors of minimal weight in an $(rn)$-dimensional linear subspace of $Z_2^n$ is exponentially small, and does not exceed $2^{-\Omega(\frac{r^2}{\log(1/r)+1} n)}$.

What is the largest possible number of vectors in an affine subspace that have a given weight? Ben-Or [2] considered affine subspaces $E \subseteq Z_2^n$ of dimension $a \cdot n$ ($a > \frac{1}{2}$ fixed, $n$ even and large) and asked for the largest possible number of vectors in $E$ whose Hamming weight is $n/2$. Consider the expected number of times we draw a vector of weight $n/2$, when we draw uniformly at random $2^{an}$ vectors from the cube. Ben-Or's conjecture states that the number of vectors with weight $n/2$ in an affine subspace never exceeds this expectation by much. Let $L_k = L_{k,n}$ be the set of vectors in $Z_2^n$ whose Hamming weight is $k$ ("the $k$-th level"). The conjecture says:

$$|E \cap L_{n/2}| \leq C_a \frac{|E||L_{\frac{n}{2}}|}{2^n} = \Theta(\frac{|E|}{\sqrt{n}}).$$

We prove more, namely that for every $0 \leq k \leq n$:

$$|E \cap L_k| \leq C_a \frac{|E|}{\sqrt{n}}$$

The harmonic-analytic formulation says that if $f = \sum_{i=1}^n W_{u_i}$ is a real function on $Z_2^m$ with $\{u_i\}_1^n$ a spanning set of $Z_2^m$, and $n/m$ is bounded away from above by 2, then $f$ attains every value at most $O(\frac{2^m}{\sqrt{n}})$ times. Examples show that the bound 2 on $n/m$ is best possible.

The extremal behavior of linear subspaces of $Z_2^n$ is of fundamental importance for coding theory. One of the foremost questions in this area concerns the largest possible size of a linear code with a given distance. Namely, given $d$ and $n$, the question is to find a linear subspace $V$ of

$Z_2^n$ with largest possible dimension, where every non-zero vector of $V$ has weight at least $d$. (See [23] for an introduction to the coding theory.) A more general question (and not as well studied) is to estimate the maximum (minimum) possible number of vectors in $V$ of given weight $k$. We point out the connection between this problem and the uncertainty principle.

Extremal properties of short sums of $Z_2^n$-characters are of interest in probability [7] and also have applications in the Local Theory of Banach Spaces [18] and in combinatorics [6]. We are interested in the possible value distribution of such functions $f : Z_2^n \to \mathbb{R}$. Specifically, how large are the higher values of $f$, and how often are they attained. We are also interested in $f$'s concentration function: $Q_1(f) = max_{x\in\mathbb{R}} Pr(x \leq f \leq x+1)$.

The connection between character sums and subspaces may hopefully contribute to the study of both subjects.

# 2 Preliminaries

This section contains necessary definitions, terminology and references, as well as some of the facts that are required later on. Elements of $Z_2^n$ will be viewed as either vectors or as subsets of $\{1, \dots, n\}$. In this context, we interchange freely between a subset and its characteristic vector.

## 2.1 Harmonic Analysis on $Z_2^n$

$Z_2^n$ is a finite Abelian group, therefore its characters $\{W_T\}_{T\in Z_2^n}$ constitute a group (the *dual group* which is isomorphic to $Z_2^n$.) The character $W_T$ is a function from $Z_2^n$ to $\{-1, 1\}$, defined as: $W_T(S) = (-1)^{|T\cap S|}$. The characters $\{W_T\}_{T\in Z_2^n}$ form an orthonormal basis in the space of real-valued functions on $Z_2^n$, equipped with uniform probability distribution.

For $f : Z_2^n \to \mathbb{R}$, define $\hat{f} : Z_2^n \to \mathbb{R}$ as $\hat{f}(T) = \frac{1}{2^n}\sum_{S\in Z_2^n} f(S)W_T(S)$. The function $\hat{f}$ is the *Fourier Transform* of $f$. See [8] for more on the Fourier Transform in $Z_2^n$.

For $0 \leq k \leq n$ define $K_k : Z_2^n \to \mathbb{R}$ as $K_k(S) = \sum_{|T|=k} W_T(S)$. Clearly, $K_k(S)$ depends only on $|S|$, and can therefore be viewed as a function on integers $0 \leq s \leq n$. With some abuse of notation, we may, therefore, view $K_k$ as $K_k(x) = \sum_{j=0}^{k}(-1)^j\binom{x}{j}\binom{n-x}{k-j}$, the $k$-th *Krawchouk polynomial*, a real polynomial of degree $k$. The polynomials $K_k$ with $0 \leq k \leq n$ form an orthogonal system with respect to the binomial measure on $\{0, \dots, n\}$. See [22] for more on orthogonal polynomials and [12] for a survey of Krawchouk polynomials. We need several more specific properties of Krawchouk polynomials.

**Lemma 2.1:** *If $n = 2d$ is even, then*

$$K_d(x) = \frac{(-1)^d 2^d}{d!}(x-1)(x-3)(x-5)...(x-(n-1)).$$

**Proof:** An easily verified property of Krawchouk polynomials is that $K_s(x+d)$ is an even function of $x$ if $s$ is even, and an odd function for odd $s$. Therefore $K_s(d)$ vanishes for odd $s$. Using the identity $\binom{n}{s}K_t(s) = \binom{n}{t}K_s(t)$, we deduce that $K_d(s) = 0$ for every odd integer $s$ between 0 and $n$. These $d$ roots of $K_d$, a polynomial of degree $d$, determine $K_d$ up to a constant factor. This constant may be obtained using the fact that $K_s(0) = \binom{n}{s}$. ∎

**Lemma 2.2:**
*Let $n = 2d$ be even. Then for any integer $k$ and for an even integer $s$ between 0 and $n$:*

$$|K_k(s)| \leq |K_d(s)|$$

**Proof:** This proof is essentially due to Solé [21]. We are grateful to S. Litsyn for this reference. The generating function of Krawchouk polynomials is $\sum_0^\infty K_k(x)z^k = (1+z)^{n-x}(1-z)^x$. Therefore, by Cauchy's integral formula, for nonnegative integers $x$:

$$K_k(x) = \frac{1}{2\pi i}\oint \frac{(1+z)^{n-x}(1-z)^x}{z^{k+1}}dz =$$

$$2^n \cdot \frac{(-i)^x}{2\pi}\int_0^{2\pi} exp[i(\frac{n}{2}-k)\theta]cos^{n-x}(\frac{\theta}{2})sin^x(\frac{\theta}{2})d\theta.$$

Thus, for $s$ even:

$$|K_k(s)| \leq \frac{2^n}{2\pi}\int_0^{2\pi} |exp[i(\frac{n}{2}-k)\theta]cos^{n-s}(\frac{\theta}{2})sin^s(\frac{\theta}{2})|d\theta \leq \frac{2^n}{2\pi}\int_0^{2\pi} cos^{n-s}(\frac{\theta}{2})sin^s(\frac{\theta}{2})d\theta = |K_d(s)|$$

∎

**Corollary 2.3:** *For even $s$ and $n = 2d$,*

$$|K_k(s)| \leq |K_d(s)| = \frac{2^d}{d!}\frac{s!(n-s)!}{2^d(\frac{s}{2})!(\frac{n-s}{2})!} = \binom{n}{d}\frac{\binom{d}{\frac{s}{2}}}{\binom{n}{s}} = O\left(\frac{2^n}{\sqrt{n}}\frac{\binom{d}{\frac{s}{2}}}{\binom{n}{s}}\right)$$

## 2.2 Coding Theory

We recall some standard notation in this area. As usual, the *rate* of a code $C \subseteq Z_2^n$ is defined as $R(C) = \frac{1}{n}\log_2|C|$. For $\frac{1}{2} \geq \delta \geq 0$ one defines

$$R(\delta) = \limsup_{n\to\infty}\{R(C) \mid C \text{ has length } n \text{ and distance } \geq \delta n\}.$$

For future use we record the essentially strongest known upper bounds on $R(\delta)$ (see [16]). [1]

**Theorem 2.4:** *Let $H(x) = -x\log(x) - (1-x)\log(1-x)$ denote the binary entropy function, and let $\mu(x) = H(\frac{1}{2} - \sqrt{x(1-x)})$. Then*

$$R(\delta) \leq \mu(\delta)$$

*for $0 \leq \delta \leq \frac{1}{2}$.*

Theorem 2.4 is a consequence of the following non-asymptotic result [16]:

[1] Strictly speaking, this is not the best bound in [16], but it matches their best bound for a wide range of $\delta$ and suffices for our purposes.

**Theorem 2.5:** *Let $C$ be a code in $Z_2^n$ with minimal distance $d$. Let $0 \le k \le n/2$ be such that the first root $x_k$ of the Krawchouk polynomial $K_k$ satisfies $x_k \le d$. Then*

$$|C| \le \binom{n}{k}(n+1)^2.$$

In order to apply theorem 2.5 we need an estimate on the first root $x_k$ of the Krawchouk polynomial $K_k$. The following, sufficiently precise estimate is implied [2] by the bounds in [13], section 5.2. For $2 \le k \le n/2$

$$x_k \le n/2 - \sqrt{(n-k+2)(\sqrt{k}-1)^2}.$$

We also record several simple but useful properties of the binary entropy $H$:

**Lemma 2.6:**

- *$H(0) = H(1) = 0$, $H(\frac{1}{2}) = 1$, $H$ is concave on $[0,1]$.*
- *For $|x| \le \frac{1}{2}$, $H(\frac{1}{2} - x) = 1 - O(x^2)$.*
- *$\binom{n}{an} = 2^{n[H(a)+o(1)]}$, and $\sum_{0\le k\le an}\binom{n}{k} = 2^{n[H(a)+o(1)]}$, where $0 \le a \le \frac{1}{2}$ is constant and $n \to \infty$.*
- *For any $0 \le r \le n/2$, $\sum_{0\le k\le r}\binom{n}{k} \le 2^{nH(r/n)}$.*

Combining theorem 2.5, the estimate for $x_k$ and the properties of the entropy function we obtain the following corollary [3]:

**Corollary 2.7:** *Let $C$ be a code in $Z_2^n$ with minimal distance $d$. Then*

$$|C| \le 2^{n\cdot\left[H\left(\frac{1}{2}-\sqrt{\frac{d}{n}(1-\frac{d}{n})}\right)+O\left(\frac{\log n}{\sqrt{n}}\right)\right]}.$$

We will also need estimates from [16] on the largest possible size of constant weight codes. Namely, we need estimates on the largest cardinality of a subset of $L_d$, in which all pairwise distances are at least $d$. Denote by $M(n,d,w)$ the largest cardinality of a subset of $L_w$, in which all pairwise distances are at least $d$ and set

$$R(\delta,\alpha) = \limsup_{n\to\infty}\frac{1}{n}\log_2 M(n,d_n,w_n),$$

where the lim sup is taken over all sequences $d_n = \delta n + o(n)$ and $w_n = \alpha n + o(n)$. We quote a bound on $R(\delta,\delta)$:

**Proposition 2.8:** *([16])*

$$R(\delta,\delta) \le H((1-\sqrt{1-x})/2),$$

*where*

$$x = 2\delta(1-\delta) - 2\delta\sqrt{2\delta - 3\delta^2}.$$

[2] Inequality (5.44), page 82: $x_k \le n/2 - \sqrt{(n-k+2)/2}\cdot h_k$, where $h_k$ is the maximal root of the $k$'th Hermite polynomial. Identity (5.35), page 79: $h_k = max\sqrt{2}\sum_{i=0}^{k-2} x_i x_{i+1}\sqrt{i+1}$, where the maximum is taken over all $\{x_i\}$ satisfying $\sum_{i=0}^{k-1} x_i^2 = 1$. Now choose $x_i = (2k)^{-1/4}$, for $i = k-\sqrt{2k}...k-1$, to obtain the bound we need.

[3] See the Appendix for the proof.

## 2.3 Matroid Theory

We need the following classical result of Edmonds (E.g., [19], which is a good introduction to matroid theory in general).

**Theorem 2.9: The Matroid Union Theorem** *Let $M$ be a matroid on a ground set $E$ with a rank function $\rho$.*

- *$M$ has $k$ disjoint bases if and only if, for every subset $X$ of $E$,*

$$k \cdot \rho(X) + |E - X| \geq k \cdot \rho(E).$$

- *$E$ can be expressed as the union of $k$ independent sets iff, for every subset $X$ of $E$,*

$$k \cdot \rho(X) \geq |X|.$$

## 2.4 An outstanding debt

We still owe the reader a proof of Lemma 1.1:
**Proof:** Let $b$ be the characteristic vector of the set $B$. Then

$$\hat{f}(B) = \frac{1}{2^m} \sum_{S \in Z_2^m} f(S) W_B(S) = \frac{1}{2^m} \sum_{1 \leq i \leq n} (-1)^{<b,u_i>}.$$

Therefore, $2^m \hat{f}(B)$ is the difference between the number of indices $i$ with $< b, u_i >= 0$ and those where $< b, u_i >= 1$. It follows that $| \oplus_{i \in B} v_i| = |\{i| < b, u_i >= 1\}| = \frac{n - 2^m \hat{f}(B)}{2}$. ∎

# 3 Minimal weight

In [9] Kalai and Linial show that the distribution of short distances in a code conveys significant information about the whole code. Specifically, they establish the existence a certain short interval $[d, d']$ near the minimum distance $d$, wherein there is a $k$ such that

$$|V \cap L_k| \geq \left( \frac{|V||L_k|}{2^n} \right)^{1-o(1)}.$$

This, naturally, draws one's attention to the behavior of a linear code near the code's minimal distance. In [9] the following surprising possibility is raised:

**Conjecture 3.1:** For every linear code $V$ of length $n$ and minimal distance $d$, the cardinality $|V \cap L_d|$ is subexponential in $n$. In other words, for every $\epsilon$ there is $N = N(\epsilon)$ such that if $V$ is a code of length $n > N$ and minimal distance $d$, then $|V \cap L_d| \leq (1 + \epsilon)^n$. ∎

Codes are known [1] in which there are $2^{\Theta(\sqrt{n})}$ vectors of the smallest weight. This is currently the best known lower bound.

In this section we prove a weaker version of the conjecture. In particular, we show that if $V$ has an exponential size, then $|V \cap L_d|$ is exponentially smaller than $V$.

**Remark 3.2:** The coding theoretic estimate in theorem 2.4 implies that a very large code has a small minimal distance. In particular, in large enough codes $V$, already the size of the whole level $|L_d|$ is exponentially smaller than $|V|$, making the statement obvious. (This reasoning holds for every code, whether linear or not.) Calculations with the estimate given in the theorem imply the validity of a weak version of the Conjecture 3.1 for all codes (linear or nonlinear) of size $|V| \geq 2^{0.60\ldots n}$. Indeed, the theorem provides a bound on $V$'s rate:

$$R(V) \leq \mu(\delta) + o(1),$$

where $\delta = \frac{d}{n}$ and $\mu(x) = H(\frac{1}{2} - \sqrt{x(1-x)})$. Denote $\phi(x) = \frac{1}{2} - \sqrt{x(1-x)}$. Note that $H(y)$ strictly increases from zero to one throughout the interval $y \in [0, \frac{1}{2}]$, and so is invertible on this interval. The function $\phi(y)$ strictly decreases from $\frac{1}{2}$ to zero on $y \in [0, \frac{1}{2}]$, and $\phi(\phi(y)) = y$. Using these facts, and ignoring asymptotically vanishing factors, we conclude:

$$d \leq n\phi(H^{-1}(R)) = n\left(\frac{1}{2} - \sqrt{H^{-1}(R)(1 - H^{-1}(R))}\right).$$

However (Lemma 2.6), $|L_d| = \binom{n}{d} = 2^{n(H(\delta)+o(1))}$. So if $R$ satisfies:

$$R > H\left(\frac{1}{2} - \sqrt{H^{-1}(R)(1 - H^{-1}(R))}\right),$$

then $|L_d \cap V|$ (in fact even all of $|L_d|$) is exponentially smaller than $|V|$. As $R$ increases from zero to one, the right term decreases from one to zero; so there is a critical $R_0$ where equality holds. By applying $H^{-1}$, it follows that $x_0 = H^{-1}(R_0) \in (0, 1/2)$ satisfies $x_0 = \phi(x_0)$, i.e. $x_0 = \frac{2-\sqrt{2}}{4}$, so $R_0 = H(x_0) = 0.60\ldots$.

Better still, rather than use 2.4, we can resort to estimates in [16] on the largest possible size of constant weight codes. Refer to Proposition 2.8 and define $r(\delta) = H((1 - \sqrt{1-x})/2)$, where $x = 2\delta(1-\delta) - 2\delta\sqrt{2\delta - 3\delta^2}$. The following are easy to check: $r \geq 0$, $r(0) = r(1/2) = 0$, the maximum of $r(\delta)$ for $0 \leq \delta \leq 1/2$ is obtained in $\delta_0 = \frac{1}{6}$, and $r(\frac{1}{6}) = H(\frac{3-2\sqrt{2}}{6}) = x_1 = 0.189\ldots$. Therefore, a code $V$ with rate above $x_1$, is exponentially bigger than $V$'s lowest level. On the other hand, when the rate is positive but $\leq x_1$, we cannot preclude the possibility of a code that is contained in the $k$-th level $L_k$ for $k$ around $n/6$, where all distances are $\geq k$. ∎

## 3.1 A geometric perspective

Almost all work done so far on extremal asymptotic problems in coding theory has taken place within the framework of linear programming. This approach views the distance distribution of $\mathcal{C}$, a code of length $n$, as an $n+1$-dimensional vector. The convex hull of all such vectors, as $\mathcal{C}$ ranges over all codes of length $n$ is a polytope $\mathcal{P}_{\mathrm{n}}$. The extremal problem is then tantamount to optimizing a linear function over the polytope $\mathcal{P}_{\mathrm{n}}$, i.e., finding a solution of a linear program. A successful completion of this plan of research would resolve essentially all extremal asymptotic problems about codes. At this writing, however, neither part of this research program is completed, and we do not even have a complete list of the facets (defining inequalities) of $\mathcal{P}_{\mathrm{n}}$. What we do know is the system of linear inequalities found by Delsarte [3]. This system of inequalities defines a polytope $\mathcal{D}_{\rm n}$ that properly contains $\mathcal{P}_{\rm n}$. One obvious question, then, is to find closer approximations (more facets) of $\mathcal{P}_{\rm n}$. Furthermore, even the current solutions for the relevant optimization problems on $\mathcal{D}_{\rm n}$ are not known to be optimal. The best results so far in this direction were found in [16], but their optimality is still in question (see [12] and [20] for more on this.) Similarly to $\mathcal{P}_{\rm n}$ we also define $\mathcal{L}_{\rm n}$, the convex hull of the weight distributions of all length $n$ linear codes. We know that $\mathcal{D}_{\rm n} \supset \mathcal{P}_{\rm n} \supset \mathcal{L}_{\rm n}$ and the inclusions are proper. We still seem far from a complete description of the latter two polytopes.

In view of these remarks, it is natural to ask whether the analogue of Kalai and Linial's problem holds on Delsarte's polytope $\mathcal{D}_{\rm n}$. The answer is negative. Here is a simple counterexample: Select an even $d$ proportionate with $n$, and set $f(d) = 2^{rn}$. Also set $f(0) = 1$, and $f(i) = 0$ for all $i \neq 0, d$. It is easy to check that $f \in \mathcal{D}_{\rm n}$ for any sufficiently small constant $r$.

These observations suggest, then, an alternative interpretation of results in this paper. Theorem 3.3 expresses a linear inequality on the distribution of values in linear codes, and by the above mentioned example, these linear inequalities hold in $\mathcal{L}_{\rm n}$ but not in $\mathcal{D}_{\rm n}$. A closer look at the proof of Theorem 3.3 shows that it yields, in fact, a whole class of inequalities (for weights near the minimum) which define a proper subpolytope of $\mathcal{D}_{\rm n}$. It is an interesting possibility to try and improve the bounds in [16] by working in this new smaller polytope, but we have not done this yet. Our impression is that an improvement over [16] will require a more far-reaching reduction of $\mathcal{D}_{\rm n}$. As observed in [9], the original Kalai-Linial conjecture, and even weaker versions, will certainly suffice.

## 3.2 Proofs

We turn to prove one of the main results of this paper:

**Theorem 3.3:** *Let $V$ be an $rn$-dimensional linear subspace of $Z_2^n$ in which the minimal weight of a non-zero vector is $d$. Then, for $r \gg \frac{\log n}{\sqrt{n}}$, the fraction of vectors in $V$ whose weight is $d$ does not exceed $2^{-\Omega(\frac{r^2 n}{\log(1/r)+1})}$.*

Lemma 1.1 suggests a translation into the language of character sums:

**Theorem 3.4:** *Let $f$ be a function on the $m$-dimensional cube $Z_2^m$, which is a sum of $n = \frac{m}{r} > m$ characters, $f = \sum_{i=1}^{n} W_{u_i}$ and where $\{u_i\}_{i=1}^{n}$ span $Z_2^m$. Let $M(f)$ be the second largest value of $f$, $M(f) = max_{S \neq 0} f(S)$. Then,*

$$Pr(f = M(f)) \leq 2^{-\Omega(\frac{rm}{\log(1/r)+1})}$$

**Proof:** First, we need several simple facts [4].

**Lemma 3.5:** *Let $B : Z_2^m \to Z_2^m$ be a non-singular linear transformation, and let $f$ be a real function on $Z_2^m$. Define another real function $f_B$ on $Z_2^m$ via $f_B(S) = f(B(S))$. Then the Fourier Transforms of $f$ and $f_B$ are related thus: $\widehat{f_B} = \hat{f}_{(B^{-1})^T}$. In particular, the functions $\hat{f}$ and $\widehat{f_B}$ have the same distribution of values.*

[4]It is convenient for us to state them in the language of character sums. Note, however, that they easily translate back to statements about linear codes. For instance, corollary 3.6 just says that switching to another basis of a code preserves the code, and therefore its weight distribution.

In terms of character sums, this lemma reads:

**Corollary 3.6:** *Let $B : Z_2^m \to Z_2^m$ be a non-singular linear transformation. Then $f = \sum W_{u_i}$ and $f_B = \sum W_{Bu_i}$ have the same distribution.*

The next lemma covers a special case which is very simple to analyze and is crucial for our results.

**Lemma 3.7:** *If the vectors $\{x_i | i \in I\}$ are linearly independent, then the values of the function $\sum_{i \in I} W_{x_i}$ are binomially distributed.*

**Proof:** By the above corollary, the distribution of $\sum_{i \in I} W_{x_i}$ is the same as in the case where $x_i$ is the $i$-th unit vector, which is clearly binomial. ∎

We turn to the proof of Theorem 3.3:

**Proof:** The plan of this proof is as follows: Denote $\mathcal{U} = \{u_i | 1 \le i \le n\}$. Consider a process at each step of which we find a set of $m/2$ linearly independent column vectors in $\mathcal{U}$ and remove them from $\mathcal{U}$. This process can go on as long as the rank of the remaining matrix does not fall below $m/2$. Say that $t$ such sets $\mathcal{B}_j$ are extracted and the set of remaining columns is $\mathcal{N}$. Thus, a decomposition is obtained: $\mathcal{U} = \cup_1^t \mathcal{B}_j \cup \mathcal{N}$, where each $\mathcal{B}_j$ is a linearly independent sequence of length $m/2$, and $k = \text{rank}(\mathcal{N}) < m/2$. Consider a change of basis (achieved by multiplying $A$ on the left by $T$, a nonsingular matrix) following which we may assume that span$(\mathcal{N})$ is a subspace of $Z_2^m$ that is contained in the span of the last $k < m/2$ unit vectors $e_{m-k+1}, \ldots, e_m$. This change of basis clearly does not affect the linear independence of the vectors in $\mathcal{B}_j$. At this stage, we should introduce names for the rows and columns of $TA$, the matrix whose rows are the new basis vectors of $V$. For simplicity's sake, we keep calling $TA$'s rows $v_1, \ldots, v_m$ and the columns $u_1, \ldots, u_n$. It is not hard to see that this is only a notational convention and that nothing is lost by these assumptions. Moreover, we are allowed to assume also that $\mathcal{B}_j$ consists of the columns $u_{\frac{(j-1)m}{2}+1}, \ldots, u_{\frac{jm}{2}}$ and that $\mathcal{N} = \{u_{\frac{tm}{2}+1}, \ldots, u_n\}$. These assumptions imply that in each of the rows $v_1, \ldots, v_{m/2}$ the last $n - 1/2mt$ coordinates vanish. This allows us to view $v_1, \ldots, v_{m/2}$ as vectors in $Z_2^{\frac{mt}{2}}$ and apply the bound from Corollary 2.7 (in $\frac{mt}{2}$ dimensions) to their linear span $V_1$.

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \mathcal{B}_1 & \cdots & & \mathcal{B}_t & \mathcal{N} \end{array} \\
\begin{array}{c} v_1 \\ \vdots \\ v_{m/2} \\ \vdots \\ v_m \end{array} &
\left[\begin{array}{c|c|c|c|c} & & \cdots & & 0 \\ \hline & & & & \end{array}\right]
\end{array}
$$

If $d$ is the minimal weight of a non-zero vector in $V_1$, then:

$$|V_1| = 2^{m/2} \leq 2^{\frac{mt}{2} \cdot \left[ H\left(\frac{1}{2} - \sqrt{\frac{2d}{mt}(1-\frac{2d}{mt})}\right) + O\left(\frac{\log(mt)}{\sqrt{mt}}\right) \right]}$$

Consequently,

$$1 \leq t \cdot \left[ H\left(\frac{1}{2} - \sqrt{\frac{2d}{mt}(1-\frac{2d}{mt})}\right) + O\left(\frac{\log(mt)}{\sqrt{mt}}\right) \right]$$

Recall that $r \gg \frac{\log n}{\sqrt{n}}$, and therefore $m = rn \gg \sqrt{n}\log n$. On the other hand, $mt \leq n$, implying $t = o\left(\frac{\sqrt{n}}{\log n}\right)$. Therefore $t \cdot O\left(\frac{\log(mt)}{\sqrt{mt}}\right) = o(1)$. Assuming $n$ is sufficiently large, we get:

$$\frac{1}{2} \leq t \cdot H\left(\frac{1}{2} - \sqrt{\frac{2d}{mt}(1-\frac{2d}{mt})}\right)$$

As in our discussion in remark 3.2, the properties of $\phi(x) = \frac{1}{2} - \sqrt{x(1-x)}$ imply:

$$d \leq d_0 = \frac{1}{2}mt\phi\left(H^{-1}\left(\frac{1}{2t}\right)\right) = \frac{1}{2}mt\left( \frac{1}{2} - \sqrt{H^{-1}\left(\frac{1}{2t}\right)\left(1 - H^{-1}\left(\frac{1}{2t}\right)\right)} \right)$$

We have thus established an upper bound on the minimal weight in $V_1$ and therefore in all of $V$. The idea now is that, by an averaging argument, every codeword of small weight must have a small weight within some of the "windows" $\mathcal{B}_j$. It is not hard to show that such an event has low probability, since by Lemma 3.7 the weights within such window are binomially distributed. Specifically, let us denote

$$f_j = \sum_{T \in \mathcal{B}_j} W_T$$

Using the functional notation of theorem 3.4, we deduce from the above that $f \geq M(f)$ implies

$$\sum_{j=1}^{t} f_j \geq \frac{1}{2}mt - 2d_0 = mt \cdot \sqrt{H^{-1}\left(\frac{1}{2t}\right)\left(1 - H^{-1}\left(\frac{1}{2t}\right)\right)}$$

and thus there exists an index $1 \leq j \leq t$ such that

$$f_j \geq m \cdot \sqrt{H^{-1}\left(\frac{1}{2t}\right)\left(1 - H^{-1}\left(\frac{1}{2t}\right)\right)} \geq m \cdot \sqrt{H^{-1}(r/4)(1 - H^{-1}(r/4))}.$$

The latter inequality holds, since the middle term decreases with $t$, whence we may assume that $\frac{1}{2}mt = n$, i.e., $t = \frac{2n}{m} = \frac{2}{r}$. Therefore:

$$Pr(f \geq M(f)) \leq t \cdot Pr\left( \; f_j \geq m \cdot \sqrt{H^{-1}(r/4)(1 - H^{-1}(r/4))} \; \right)$$

The standard tail estimates for the binomial distribution imply:

$$Pr(f_j \geq \theta m) \leq e^{-\frac{m\theta^2}{4}}.$$

An easy calculation shows:

$$H^{-1}(r/4)(1-H^{-1}(r/4)) = \Theta(\frac{r}{\log 1/r+1}).$$

Taking all the estimates together, we can now complete the proof of theorem 3.4:

$$Pr(f \geq M(f)) \leq n \cdot 2^{-\Omega(\frac{rm}{\log(1/r)+1})} \leq 2^{-\Omega(\frac{rm}{\log(1/r)+1})}.$$

In the last inequality $n$ vanishes, since the conditions $r \gg \frac{\log n}{\sqrt{n}}$, and $m = rn$ imply $log(n) = o\left(\frac{rm}{\log(1/r)+1}\right)$. ∎

## 3.3 Intersection of subspaces and balls

A simple consequence of the theorem is that there are no constant-weight linear codes $V$ of dimension $rn$, with $r$ bounded away from zero and $n \to \infty$. The analogous statement for affine subspaces is, however, false. In Example 4.2 we will encounter an affine subspace of dimension $n/2$, which is a subset of the middle level $L_{n/2}$.

Still, essentially the same argument yields the following "affine version" of Theorem 3.3:

**Theorem 3.8:** *Let $V$ be an $rn$-dimensional linear subspace of $Z_2^n$ with minimal weight $d$. Then, for any $x \in Z_2^n$, the fraction of vectors in $V + x$ of weight $d$ does not exceed $2^{-\Omega(\frac{r^2 n}{\log(1/r)+1})}$.*

**Proof:** (Sketch) Pick a generating matrix $A$ for $V$ and decompose its columns as in the above proof: $\cup_1^t \mathcal{B}_j \cup \mathcal{N}$, where $|\mathcal{B}_j| = rn/2$, $\text{rank}(\mathcal{N}) < rn/2$. We are interested in those elements of $V$ whose distance from $x$ is $\leq d$. Apply Theorem 2.5 to $V$ to obtain an upper bound $d_0 \geq d$. If $y \in V$ is at distance $\leq d_0$ from $x$, then there is an index $1 \leq j \leq t$, such that the section of $y$ within the $j$-th "window" is at distance $\leq \frac{d_0}{t}$ from the corresponding segment of $x$. As before, the weights of codewords within $\mathcal{B}_j$ are binomially distributed, and the tail estimates on the binomial distribution yield the theorem.

An alternate statement is:

**Theorem 3.9:** *Let $V$ be an $rn$-dimensional linear subspace of $Z_2^n$ with minimal weight $d$. Then the intersection of $V$ with any Hamming ball of radius $d$ contains at most $|V| \cdot 2^{-\Omega(\frac{r^2 n}{\log(1/r)+1})}$ vectors. Consequently, $V$ cannot be covered by fewer than $2^{\Omega(\frac{r^2 n}{\log(1/r)+1})}$ Hamming balls of radius $d$.*

## 3.4 A matroid-theoretic detour

The *girth* $g(M)$ of a matroid $M = (E, \rho)$ is the smallest cardinality of a cycle in $M$. In matroid-theoretic terms Conjecture 3.1 (or rather, its matroid dual) states: "The number of cycles of cardinality $g(M)$ in a binary matroid $M = (E, \rho)$ is subexponential in the size of the ground set $E$."

While this conjecture is open for binary matroids in general, it can be established in certain interesting instances. Consider, for example, cographic matroids - or, equivalently, subspaces $V$

of $Z_n^2$ where each column of the generating matrix $A$ in Lemma 1.1 has weight $\leq 2$. The cycles of the cographic matroid are the edge cutsets of the corresponding graph, so the statement reads:

"The number of minimal (non-empty) edge cutsets in a graph with $e$ edges is subexponential in $e$."

In fact, much more is true, and this number cannot even exceed $\binom{e+1}{2}$. Indeed, it is clearly sufficient to prove this for connected graphs. And it is known [14] that the number of minimal edge cutsets in a connected graph with $v$ vertices cannot exceed $\binom{v}{2}$. (This bound is tight, as demonstrated by $C_v$, the cycle of length $v$.) See also the ingenious probabilistic proof from [10]. We illustrate the methods of the present paper and prove a slightly weaker bound.

**Proposition 3.10:** *Let $G$ be a simple connected graph on $v$ vertices whose edge-connectivity is $t$. Then at most $v^2$ cuts in $G$ have exactly $t$ edges.*

**Proof:** Replace each edge of $G$ by two parallel edges, to obtain a $2t$-edge-connected multigraph $G'$. A simple and well-known consequence of Matroid Union Theorem (Theorem 2.9) is that a $2t$-edge-connected multigraph contains $t$ disjoint spanning trees, say, $\mathcal{T}_1, \ldots, \mathcal{T}_t$. Let $f_i = \sum_{e \in E(\mathcal{T}_i)} W_e$, and $f = \sum_{e \in E} W_e$. (Here we view an edge $e$ as an element in $Z_2^{v-1}$. Recall that for $x \in Z_2^{v-1}$, the character $W_x : Z_2^{v-1} \to \{-1, 1\}$ is defined by $W_x(y) = (-1)^{<x,y>}$.) The assumption that $2t$ is the edge-connectivity of $G'$ translates into $M(f) = |E(G')| - 4t$. This, in turn, implies that when $f(x) = M(f)$, then there is an $1 \leq i \leq t$ for which $f_i \geq v - 3$. By Lemma 3.7, $|\{x | f_i(x) = v - 2j - 1\}| = \binom{v-1}{j}$ (where $i = 1, \ldots, t$ and $j = 0, \ldots, v-1$). Therefore, $|\{x : f(x) = M(f)\}| \leq t |\{x : f_i(x) \geq v - 3\}| \leq tv \leq v^2$. ∎

Similarly, the statement of Conjecture 3.1 for graphic matroids is that:

"If the shortest cycle in a graph $G$ with $e$ edges has length $g$, then $G$ has only subexponentially many in $e$ cycles of length $g$".

Again, much more is true.

**Proposition 3.11:** *Let $G$ be a graph with $e$ edges whose shortest cycle has length $g$, then at most $O(e^2)$ cycles in $G$ have length $g$.*

**Proof:** Pick from every $g$-cycle in $G$ two antipodal edges (at distance $\lfloor \frac{g}{2} \rfloor - 1$). It is not hard to see that every such pair of edges uniquely determines the cycle to which it belongs. Therefore the number of $g$-cycles is at most $O(e^2)$. ∎

**Remark 3.12:** In both the graphical and the cographical problems, the same proofs extend and yield bounds on the number of near-minimal cuts/cycles. ∎

# 4 The middle level

Questions concerning weight distributions come up also outside of coding theory, graph theory and matroid theory. The specific question addressed in this section arose in the field of computational complexity. There is a well-developed theory of decision trees in which internal nodes compute boolean or algebraic functions of the input, and the computation proceeds according to the outcomes. Much less is known when internal nodes are allowed to compute *analytic* functions. These so-called analytic decision trees are easier to understand by considering infinitesimally small inputs. This is the approach taken in a recent paper by Ben-Or [2]. His investigations have led him to state:

**Conjecture 4.1:** Let $E \subseteq Z_2^n$ be an affine subspace of dimension $\lambda \cdot n$, where $\lambda > \frac{1}{2}$ is fixed, and $n$ is large. Then at most $O(\frac{|E|}{\sqrt{n}})$ vectors in $E$ have Hamming weight $n/2$. ∎

As observed by Ben-Or, the condition $\lambda > \frac{1}{2}$ is essential, in view of the following example:

**Example 4.2:** Consider the following affine subspace: $V = \{(x_1, x_2 ... x_n) \in Z_2^n : x_1 + x_2 = 1, x_3 + x_4 = 1, \ldots, x_{n-1} + x_n = 1\}$. Clearly $V$ is an $\frac{n}{2}$-dimensional affine subspace of $Z_2^n$, which is contained in $L_{\frac{n}{2}}$, the middle level. ∎

Ben-Or's problem may be put in the context of the uncertainty principle of harmonic analysis. Informally, this principle posits that it is impossible for both $f$ and $\hat{f}$ to have small support (see [4, 17] for more on this). The obvious quantitative formulation would be an upper bound on the number of zeros a real function $f$ on $Z_2^n$ can have, given $|\text{support}(\hat{f})|$, the cardinality of the support of its transform. It is easy to construct examples where $|\text{support}(\hat{f})| = 2^s$, while $Pr(f = 0) = 1 - \frac{1}{2^s}$, e.g., $f = \prod_1^s (1 + R_i)$, where $R_i = W_{e_i}$ is the $i$-th Rademacher function. These examples are tight (e.g., [17]). Also, these examples are degenerate, in that support$(\hat{f})$ fails to span $Z_2^n$. The large number of zeros in these examples can be accounted to this degeneracy. Henceforth we will require $f$ to be *nondegenerate*, i.e., such that support$(\hat{f})$ spans $Z_2^n$. One nondegenerate example is $f = (1 + R_1) \sum_2^n R_i$. Here $|\text{support}(\hat{f})| = 2n - 2$ and $Pr(f = 0) = \frac{1}{2} + \Theta(n^{-1/2})$. It may be worth noting that this example is essentially the functional counterpart of Ben-Or's example. Below (Corollary 4.6) we show that if $f$ is nondegenerate, and moreover $\hat{f}$ takes a constant non-zero value on a basis of $Z_2^n$, and $|\text{support}(\hat{f})| \leq (\frac{3}{2} - \epsilon)n$, then $Pr(f = 0) \leq O(n^{-1/2})$. We do not know whether this is implied by the weaker assumption that $\hat{f}$ takes a constant non-zero value on a basis of $Z_2^n$, and $|\text{support}(\hat{f})| \leq (2 - \epsilon)n$, which would be tight in view of the above example. This weaker assumption does suffice, if $\hat{f}$ takes only the values $-1, 0, 1$ (Corollary 4.6). Perhaps for a nondegenerate $f$ with $Pr(f = 0) = 1 - \frac{1}{2^s}$, the transform $\hat{f}$ must have support $\geq (n - s)2^s$, as in the obvious extension of the above construction.

Let us mention two papers on related subjects: In [5] it is shown that a linear subspace of dimension $> n/2$ must meet the middle level (here $n$ is required to be divisible by 4.) Also, under the additional condition that the dual distance of $V$ is large, it is shown in [11] that for $k$ in a certain range around $\frac{n}{2}$,

$$|V \cap L_k| \leq O\left(\sqrt{n} \cdot \frac{|V||L_k|}{2^n}\right).$$

The following variation on Ben-Or's example deals with subspaces of even larger dimension. These may, in fact, be the extremal examples for the problems considered in the present section.

**Example 4.3:** Again $n$ is even and $t \leq n$ is even as well. Let $V = \{(x_1, x_2 ... x_n) \in Z_2^n : x_1 + x_2 = 1, x_3 + x_4 = 1, \ldots, x_{t-1} + x_t = 1\}$. Clearly, $\dim(V) = n - \frac{t}{2}$. Maintaining the notation

$\dim(V) = \lambda \cdot n$, this subspace satisfies:

$$|V \cap L_{\frac{n}{2}}| \geq \Omega\left(\frac{1}{\sqrt{2\lambda - 1}} \frac{|V|}{\sqrt{n}}\right).$$

∎

We prove a statement that is stronger that Ben-Or's Conjecture 4.1:

**Theorem 4.4:** *Let $F$ be an affine subspace of $Z_2^n$ with $\dim(F) = \lambda \cdot n$, where $\lambda > 1/2$. Then for each $0 \leq k \leq n$:*

$$|F \cap L_k| \leq O\left(\frac{1}{(2\lambda - 1)^4} \frac{|F|}{\sqrt{n}}\right).$$

*The bound is tight up to the constant 4 in the exponent of $2\lambda - 1$.*

Using the translation in Lemma 1.1, we obtain (with $b = \frac{1-\lambda}{\lambda}$):

**Corollary 4.5:** *Let $b < 1$, and let $g = \sum_{i=1}^{(1+b)m} W_{x_i}$ be a sum of $(1+b)m$ characters of $Z_2^m$, such that $x_i$, $1 \leq i \leq (1+b)m$, span $Z_2^m$. Then the concentration function of $g$ satisfies:*

$$Q_1(g) \leq O\left(\left(\frac{1+b}{1-b}\right)^4 \frac{1}{\sqrt{m}}\right),$$

*where $Q_1(g)$ is defined as $sup_{x \in \mathcal{R}} Pr\left(x \leq g < x+1\right)$.*

**Corollary 4.6:**

1. *Let $R_i = W_{e_i}$ be the $i$-th Rademacher function on $Z_2^m$, and let $R = \sum_{1 \leq i \leq m} R_i$. Consider an approximation of $R$ by a signed sum of $bm$ Walsh functions: $\sum_{1 \leq i \leq bm} \epsilon_i W_{T_i}$, where $b < 1$ and $\epsilon_i \in \{-1, 1\}$. For any choice of the $T_i$ and the $\epsilon_i$*

$$Pr\left(R = \sum_{1 \leq i \leq bm} \epsilon_i W_{T_i}\right) \leq O\left(\left(\frac{1+b}{1-b}\right)^4 \frac{1}{\sqrt{m}}\right).$$

2. *If, moreover, $b < \frac{1}{2}$, then $R$ is hard to approximate by* any *linear combination of $bm$ Walsh functions:*

$$Pr\left(R = \sum_{1 \leq i \leq bm} a_i W_{T_i}\right) \leq O\left(\frac{1}{(1-2b)^4} \frac{1}{\sqrt{m}}\right)$$

*for any choice of characters $W_{T_i}$ and real $a_i$ ($1 \leq i \leq bm$).*

**Proof:**

1. Let $n = (1+b)m$, and consider the $m \times n$ zero-one matrix whose multiset of columns $x_1, \ldots, x_n$ is $e_1, .., e_m, T_1, .., T_{bm}$. Let the rows be denoted $z_1, \ldots, z_m$, and let $V \subset Z_2^n$ be their linear span. Augment this matrix with an additional row $z_{m+1}$, the first $m$ coordinates of which are zero. In later coordinates $z_{m+1}(i) = 0$ or 1, according to $\epsilon_i = -1$ or 1. Let the augmented matrix be called $A$. With the notation of Lemma 1.1, let $f$ be the characteristic function of the (multi)set of $A$'s ($(m+1)$-dimensional) columns. By Corollary 4.5 (or rather a slight modification thereof, where the $x_i$ may span a subspace of codimension one),

$$Pr(f = 0) \leq O\left(\left(\frac{1+b}{1-b}\right)^4 \frac{1}{\sqrt{m}}\right).$$

   Note that $f(x_1, \ldots, x_m, 1) = g(x_1, \ldots, x_m)$, where $g = R - \sum_{1 \leq i \leq bm} \epsilon_i W_{T_i}$. Therefore, $Pr(g = 0) = Pr(f = 0 | x_{m+1} = 1) \leq 2Pr(f = 0)$, and the claim follows.

2. Set $h = \sum_{1 \leq i \leq bm} a_i W_{T_i}$, and let $U$ be a linear subspace of $Z_2^m$ spanned by the $\{T_i\}$. The space $Z_2^m$ is partitioned by the cosets of $U^\perp$, so let the quotient space $X = Z_2^m / U^\perp$ be viewed as a set of distinct representatives for the cosets. Then,

$$Pr(R = \sum_{1 \leq i \leq bm} a_i W_{T_i}) = \frac{1}{|X|} \sum_{x \in X} Pr(h(y) = R(y) | y \in U^\perp \oplus x).$$

   Now, $h$ is constant on each coset of $U^\perp$, for if $y \in (U^\perp \oplus x)$, say $y = u^\perp \oplus x$, then

$$h(y) = \sum_{1 \leq i \leq bm} a_i W_{T_i}(y) = \sum_{1 \leq i \leq bm} a_i (-1)^{<T_i, y>} = \sum_{1 \leq i \leq bm} a_i (-1)^{<T_i, u^\perp \oplus x>} = \sum_{1 \leq i \leq bm} a_i (-1)^{<T_i, x>}.$$

   Therefore, in the relation $h(y) = R(y)$, considered over $y \in U^\perp \oplus x$, the term $h(y)$ stays constant. However, $R$ depends only on Hamming weights, $R(y) = m - 2|y|$. Therefore, $Pr(h(y) = R(y) | y \in U^\perp \oplus x)$ cannot exceed the fraction of elements in $U^\perp \oplus x$ that reside in any single level. But $U^\perp \oplus x$ is a $(1-b)m$-dimensional affine subspace of $Z_2^m$, and $b < 1/2$, so Theorem 4.4 applies, yielding for every $x \in X$,

$$Pr(h(y) = R(y) | y \in U^\perp \oplus x) \leq O\left(\frac{1}{(1-2b)^4} \frac{1}{\sqrt{m}}\right).$$

   The claim follows.

▮

**Proof:** (Of Theorem 4.4) We argue by induction on the dimension $n$ that, for some absolute constant $c > 0$, holds $|F \cap L_k| \leq c \cdot \frac{1}{(2\lambda - 1)^4} \frac{|F|}{\sqrt{n}}$. The theorem is true for $n = 1, 2$. From now on we assume that $n$ is even. Note that this causes no loss of generality. If the theorem is true for all even $n$, it is also true for all odd $n$ (with a constant $c' = 2c$, say). To see this, embed $Z_2^n$ in $Z_2^{n+1}$ by adding a zero last coordinate.

$F$ is an affine subspace, say $F = V \oplus z$, where $V$ is a linear subspace of $Z_2^n$. Pick a basis $v_1, \ldots, v_m$ of $V$, and let $A$ be the matrix with rows $v_1, ..., v_m$. Denote the columns of $A$ by $x_1, .., x_n$. Let $\rho$ be the rank function of the binary matroid defined by $\{x_1, .., x_n\}$.

We distinguish between two cases:
(I) There exists $Y \subseteq \{x_1, .., x_n\}$ with $|Y| \geq 2\rho(Y)$.
(II) No such $Y$ exists.

In the first case, the proof can be completed by induction. In the second case, by the Matroid Union Theorem 2.9, $\{x_1, .., x_n\}$ is a union of two linearly independent sets. In this case, the problem is solved straightforwardly.

We start with case (I). Without loss of generality, $Y = \{x_{n-2t+1}, .., x_n\}$ has rank $\rho(Y) = t$, and the set $\{x_{n-2t+1}, .., x_{n-t}\}$ is a linearly independent spanning set of $Y$. By a proper change of basis, we can further assume $x_i = e_{i-(n-t-m)}$ for $n - 2t + 1 \leq i \leq n - t$:

$$\begin{array}{cc}
 & \begin{array}{ccccccc} x_1 & \cdots & x_{n-2t} & \cdots & x_{n-t} & \cdots & x_n \end{array} \\
\begin{array}{c} v_1 \\ \vdots \\ v_{m-t} \\ \vdots \\ v_m \end{array} & \left[\begin{array}{c|c|c} & 0 & 0 \\ \hline & \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} & \end{array}\right]
\end{array}$$

So, the first $m - t$ coordinates of the vectors $x_i$, for $n - 2t + 1 \leq i \leq n$, are zero. Consider the linear subspace $V_1$ of $V$ spanned by $v_1, .., v_{m-t}$. Since the last $2t$ coordinates of these vectors are zero, $V_1$ may be viewed as a subspace of $Z_2^{n-2t}$. We apply the induction hypothesis to $V_1$ and its translates. To this end, let $n_1 = n - 2t$ and $m_1 = \dim(V_1) = m - t$. Note that $m_1 = \lambda_1 n_1$, where $\lambda_1 = \frac{m-t}{n-2t}$. We wish to estimate $|F \cap L_k|$ for every $0 \leq k \leq n$. Now, $F = V \oplus z$, and we can express $V$ as the disjoint union of cosets of $V_1$, namely:

$$|F \cap L_k| = |(V \oplus z) \cap L_k| = \sum_{w \in V/V_1} |(V_1 \oplus w \oplus z) \cap L_k|.$$

All the vectors in $V_1 \oplus w \oplus z$ have the same last $2t$ coordinates. Say that $l$ of these $2t$ coordinates are 1. Therefore $|(V_1 \oplus w \oplus z) \cap L_k| = |F_1 \cap L_{k-l,n-2t}|$. Here $F_1$ is the restriction of $V_1 \oplus w \oplus z$ to the first $n - 2t$ coordinates, $F_1$ being an $m_1$-dimensional affine subspace of $Z_2^{n-2t}$. The induction hypothesis may be applied to yield

$$|F_1 \cap L_{k-l,n-2t}| \leq c \cdot \frac{1}{(2\lambda_1 - 1)^4} \frac{|F_1|}{\sqrt{n - 2t}}.$$

Expand

$$\frac{1}{(2\lambda_1 - 1)^4 \sqrt{n - 2t}} \leq \frac{1}{(2\lambda - 1)^4} \frac{(n - 2t)^{\frac{7}{2}}}{n^4} \leq \frac{1}{(2\lambda - 1)^4} \frac{1}{\sqrt{n}}$$

to conclude that

$$|F_1 \cap L_{k-l,n-2t}| \le c \cdot \frac{1}{(2\lambda-1)^4} \frac{|F_1|}{\sqrt{n}}.$$

Finally:

$$|F \cap L_k| \le c \cdot \frac{|V/V_1|}{(2\lambda-1)^4} \frac{|V_1|}{\sqrt{n}} \le c \cdot \frac{1}{(2\lambda-1)^4} \frac{|F|}{\sqrt{n}},$$

which completes the proof in case (I).

Case (II):

By the Matroid Union theorem, the columns of $A$ can be divided into two linearly independent sets. It is possible to augment one of these sets to a basis, using vectors from the other set. Therefore, we may assume that $x_1 = e_1, \ldots, x_m = e_m$, and that $\{x_{n-m+1}, .., x_n\}$ are linearly independent. Note that $2^{-n}|F \cap L_k|$ is the inner product of the characteristic functions of $F$ and $L_k$. It follows that

$$\frac{|F \cap L_k|}{|F|} = \frac{2^n}{|F|} < 1_F, 1_{L_k} >= \frac{2^{2n}}{|F|} < \hat{1}_F, \hat{1}_{L_k} >$$

by the Parseval identity. Observe that $\hat{1}_{L_k} = \frac{1}{2^n} K_k$. To evaluate $\hat{1}_F$, recall that $F = V \oplus z$, so that $1_F = 2^n \cdot 1_V * 1_{\{z\}}$ (convolution), whence $\hat{1}_F = 2^n \cdot \hat{1}_V \cdot \hat{1}_{\{z\}}$. But $\hat{1}_V = \frac{|V|}{2^n} 1_{V^\perp}$ and $\hat{1}_{\{z\}} = \frac{1}{2^n} W_z$, and we conclude that

$$\frac{|F \cap L_k|}{|F|} =< 1_{V^\perp} W_z, K_k > \le \frac{1}{2^n} \sum_{s=0}^{n} f_s |K_k(s)|,$$

where $f_s = |\{x \in V^\perp, |x| = s\}|$, is the number of linear dependencies of length $s$ among the $x_i$. This uses the fact that $W_z$ takes only the values $-1, 1$.

To facilitate the following computations, we would like to assume that there are no linear dependencies among the $\{x_i\}$ that have an odd length. That is, $f_s = 0$ for every odd $s$. Let us see why this assumption causes no loss of generality. If $\bar{1}$, the all-one vector, is in $V$, then $V^\perp$ is supported only on the even levels, and this assumption holds. If $\bar{1} \notin V$, then we add it to $V$ to create $V^+$, of dimension larger by one. In other words, we augment the matrix $A$ with the row $\bar{1}$, to obtain the matrix $A^+$. The partition of $A$'s columns into two independent sets still works for $A^+$. Also, upper bounds on the intersection of $V^+$ with various levels certainly apply to $V$.

We turn to prove an upper bound on $\frac{|F \cap L_k|}{|F|}$. Since $\{x_i\}$ are a union of a basis and a linearly independent set of size $(1-\lambda)n$, it is easy to see that

$$f_s \le \sum_{j=0}^{s} \min \left\{ \binom{n-\lambda n}{j}, \binom{\lambda n}{s-j} \right\}.$$

Therefore,

$$\frac{|F \cap L_k|}{|F|} \le \frac{1}{2^n} \sum_{s=0}^{n} f_s |K_k(s)| \le \frac{1}{2^{n-1}} \sum_{s=0}^{\frac{n}{2}} g_s |K_k(s)|,$$

where $g_s = \sum_{j=0}^{s} \min \left\{ \binom{n-\lambda n}{j}, \binom{\lambda n}{s-j} \right\}$. The last step relies on the relation $|K_k(s)| = |K_k(n-s)|$ and the observation that $f_{n-s} \le g_s$ as well as $f_s \le g_s$. As mentioned above, we may assume that

$f_s$ vanishes for odd $s$, so

$$\frac{|F \cap L_k|}{|F|} \leq \frac{1}{2^{n-1}} \sum_{n/2 \geq s \geq 0,\ even} g_s |K_k(s)| \leq \frac{1}{2^{n-1}} \sum_{n/2 \geq s \geq 0,\ even} g_s |K_{\frac{n}{2}}(s)| \leq O\left( \sum_{n/2 \geq s \geq 0,\ even} \frac{g_s}{\sqrt{n}} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \right).$$

The last two inequalities are based on Corollary 2.3.

If we can show that

$$\sum_{n/2 \geq s \geq 0,\ even} g_s \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq O\left(\frac{1}{(2\lambda - 1)^4}\right),$$

then this will conclude the proof of Case (II) and the proof of the theorem. To proceed, we need upper bounds on $g_s$. To this end, we split $g_s = g_s^{(1)} + g_s^{(2)}$, where
$g_s^{(1)} = \sum_{j > \frac{s}{2}}^{s} \min\left\{\binom{n-\lambda n}{j}, \binom{\lambda n}{s-j}\right\}$, $g_s^{(2)} = \sum_{j=0}^{\frac{s}{2}} \min\left\{\binom{n-\lambda n}{j}, \binom{\lambda n}{s-j}\right\}$.
We start with $g_s^{(1)}$. Obviously,

$$\sum_{j > \frac{s}{2}}^{s} \min\left\{\binom{n - \lambda n}{j}, \binom{\lambda n}{s-j}\right\} \leq \sum_{j > \frac{s}{2}}^{s} \sqrt{\binom{n - \lambda n}{j}\binom{\lambda n}{s-j}}.$$

But in the sum $\sum_{j > \frac{s}{2}}^{s} \sqrt{\binom{n-\lambda n}{j}\binom{\lambda n}{s-j}}$ the first term is clearly the largest, i.e.,

$$\sum_{j > \frac{s}{2}}^{s} \min\left\{\binom{n - \lambda n}{j}, \binom{\lambda n}{s-j}\right\} \leq s \cdot \sqrt{\binom{n - \lambda n}{s/2}\binom{\lambda n}{s/2}}.$$

Therefore,

$$\frac{g_s^{(1)} \binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq s \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \sqrt{\binom{n - \lambda n}{s/2}\binom{\lambda n}{s/2}} = s \frac{\binom{\frac{n}{2}}{\frac{s}{2}}^2}{\binom{n}{s}} \sqrt{\frac{\binom{n-\lambda n}{s/2}\binom{\lambda n}{s/2}}{\binom{\frac{n}{2}}{\frac{s}{2}}^2}} \leq O(s \cdot (4\lambda(1-\lambda))^{s/2}).$$

We have to justify the last inequality. Indeed, $\frac{\binom{\frac{n}{2}}{\frac{s}{2}}^2}{\binom{n}{s}} \leq 1$, and to estimate the square root we use the following simple fact: for positive integers $x, y, a$,

$$\frac{\binom{x+y}{a}\binom{x-y}{a}}{\binom{x}{a}^2} \leq \left(1 - \frac{y^2}{x^2}\right)^a.$$

To convince yourself about this inequality, put together the $i$-th terms in the product form of the binomials and observe that $\frac{(x+y-i)(x-y-i)}{(x-i)^2} \leq 1 - \frac{y^2}{x^2}$ for any $0 \leq i \leq x - y$. In our case, $x = n/2$, $y = (\lambda - \frac{1}{2})n$, and $a = s/2$.

To estimate $g_s^{(2)}$ we use

$$g_s^{(2)} = \sum_{j=0}^{\frac{s}{2}} \min\left\{\binom{n - \lambda n}{j}, \binom{\lambda n}{s-j}\right\} \leq \sum_{j=0}^{\frac{s}{2}} \binom{n - \lambda n}{j}.$$

There are two cases to consider, depending on the value of $s$. If $s \leq n - \lambda n$, then

$$\sum_{j=0}^{\frac{s}{2}} \binom{n-\lambda n}{j} \leq s \binom{n-\lambda n}{\frac{s}{2}}$$

and so

$$\frac{g_s^{(2)} \binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq s \frac{\binom{\frac{n}{2}}{\frac{s}{2}} \binom{n-\lambda n}{\frac{s}{2}}}{\binom{n}{s}} \leq s \frac{\binom{(3/2-\lambda)n}{s}}{\binom{n}{s}} \leq s \left(\frac{3}{2} - \lambda\right)^s .$$

Two easy facts used here are: $\binom{a_1}{b_1}\binom{a_2}{b_2} \leq \binom{a_1+a_2}{b_1+b_2}$ and the inequality $\frac{\binom{x}{a}}{\binom{y}{a}} \leq (\frac{x}{y})^a$ for $0 \leq a \leq y \leq x$.

In the complementary range, $n - \lambda n \leq s \leq \frac{n}{2}$, it suffices to use

$$\sum_{j=0}^{\frac{s}{2}} \binom{n-\lambda n}{j} \leq 2^{n-\lambda n},$$

whence

$$\frac{g_s^{(2)} \binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq \frac{2^{n-\lambda n} \binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}}.$$

Now, we sum up all the previous bounds, and show that indeed

$$\sum_{n/2 \geq s \geq 0,\ even} g_s \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq O\left(\frac{1}{(2\lambda-1)^4}\right).$$

First,

$$\sum_{n/2 \geq s \geq 0,\ even} g_s^{(1)} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq \sum_{s\ even} s(4\lambda(1-\lambda))^{s/2} \leq \sum_s s(4\lambda(1-\lambda))^s.$$

We use the identity $\sum_s sx^s = \frac{1}{(1-x)^2}$, with $x = 4\lambda(1-\lambda)$ to conclude:

$$\sum_{n/2 \geq s \geq 0,\ even} g_s^{(1)} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq O\left(\frac{1}{(2\lambda-1)^4}\right).$$

Our estimates on $g_s^{(2)}$ imply:

$$\sum_{n/2 \geq s \geq 0,\ even} g_s^{(2)} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq \sum_s s\left(\frac{3}{2} - \lambda\right)^s + 2^{n-\lambda n} \sum_{n/2 \geq s \geq n-\lambda n,\ even} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}}.$$

Now, $\sum_s s(\frac{3}{2} - \lambda)^s = O\left(\frac{1}{(2\lambda-1)^2}\right)$, and we are left with estimating the second sum, which we call $S$.

We have to distinguish two cases, depending on the value of $\lambda$. If $\lambda$ is bounded away from $1/2$, say $\lambda \geq 0.6$, then the sum of $\frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}}$, taken over even $s$ in the range $n/2 \geq s \geq n - \lambda n$, is dominated by the first term. To see this, use the fact that the first, say $\frac{2\lambda-1}{8}n$, terms decrease geometrically, while later terms continue to decrease, and therefore make a negligible contribution to the sum. That is,

$$\sum_{n/2 \geq s \geq n-\lambda n,\ even} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq O\left(\frac{\binom{\frac{n}{2}}{\frac{n-\lambda n}{2}}}{\binom{n}{n-\lambda n}}\right).$$

Therefore,

$$S \leq O\left(\frac{2^{n-\lambda n}\binom{\frac{n}{2}}{\frac{n-\lambda n}{2}}}{\binom{n}{n-\lambda n}}\right) \leq O(2^{n(1-\lambda-\frac{1}{2}H(1-\lambda))}),$$

where we have employed Stirling's formula to estimate the binomials. But $2^{n(1-\lambda-\frac{1}{2}H(1-\lambda))}$ is bounded by 1, since $H(x) \geq 2x$, for $x \leq 1/2$.

The last range to check is $0.5 \leq \lambda \leq 0.6$. Since $\frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}}$ decreases as $s$ goes from $n - \lambda n$ to $n/2$, we can bound

$$S = 2^{n-\lambda n} \cdot \sum_{n/2 \geq s \geq n-\lambda n,\ even} \frac{\binom{\frac{n}{2}}{\frac{s}{2}}}{\binom{n}{s}} \leq 2^{n-\lambda n} \cdot (\lambda - 1/2)n \cdot \frac{\binom{\frac{n}{2}}{\frac{n-\lambda n}{2}}}{\binom{n}{n-\lambda n}} \leq O\left((\lambda - 1/2)n \cdot 2^{n(1-\lambda-\frac{1}{2}H(1-\lambda))}\right),$$

again using Stirling. Set $x = (\lambda - 1/2)n$. Then the last expression is $x \cdot 2^{n((\frac{1}{2}-\frac{x}{n})-\frac{1}{2}H(\frac{1}{2}-\frac{x}{n}))}$, where $0 \leq \frac{x}{n} \leq 0.1$. The concavity of $H$ (Lemma 2.6) implies that $H(1/2 - \epsilon) \geq 1 - 0.3 \cdot \epsilon$ whenever $0 \leq \epsilon \leq 0.1$. Therefore,

$$\frac{1}{2} - \frac{x}{n} - \frac{1}{2} \cdot H(\frac{1}{2} - \frac{x}{n}) \leq -0.85 \cdot \frac{x}{n},$$

implying $S \leq O(x \cdot 2^{-0.85 \cdot x})$. It is easily verified that the expression $x \cdot 2^{-0.85 \cdot x}$ is bounded for all $x \geq 0$. Consequently, $S \leq O(1)$, where the constant in the $O(1)$ does not depend on $\lambda$. Finally, all cases are checked, and the theorem is established. ∎

# 5 Acknowledgment

Gabor Tardos pointed out the many inaccuracies we had in the paper. Following his suggestions, the statements of Propositions 3.10 and 3.11 are now somewhat stronger, and the proofs are actually simpler. Jaikumar Radhakrishnan made several very helpful remarks: on the proof of Theorem 3.3 and on the girth problem. Simon Litsyn and Ilya Krasikov have generously shared their knowledge in coding theory. We also thank Yuval Peres for several useful conversations.

# References

[1] N. Alon, Packings with large minimum kissing numbers, *Discrete Math.* 175(1997), 249-251.

[2] M. Ben-Or, Randomized analytic decision trees, preprint, Hebrew University, 1996.

[3] P. Delsarte, An algebraic approach to the association schemes of coding theory, *Phillips Research Reports Supplements*, 10(1973).

[4] D. L. Donoho and P. B. Stark, Uncertainty principles and signal recovery, *SIAM J. Appl. Math.* 49(1989), 906-931,

[5] H. Enomoto, P. Frankl, N. Ito and K. Nomura, Codes with given distances, *Graphs and Combinatorics* 3(1987), 25 - 38.

[6] E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another, *Israel Journal of Mathematics*, 105(1998), 251-256.

[7] P. Hitczenko, Domination inequality for martingale transforms of a Rademacher sequence, *Israel Journal of Mathematics*, 84(1993), 161-178.

[8] J. Kahn, G. Kalai and N. Linial, The influence of variables on boolean functions, *29th Symposium on the Foundations of Computer Science*, White Planes, 1988, 68 - 80.

[9] G. Kalai and N. Linial, On the distance distribution of codes, *IEEE Trans. Inf. Th.*, 41(1995), 1467 - 1472.

[10] D. R. Karger and C. Stein, A new approach to the minimum cut problems, Journal of the ACM, 43(1996), 601-640.

[11] I. Krasikov and S. Litsyn, Estimates for the range of binomiality on code's spectra, *IEEE Trans. Inf. Th.*, 43(1997), 987-991.

[12] V. I. Levenshtein, Krawtchouk polynomials and universal bounds for codes and design in Hamming spaces, *IEEE Trans. Inf. Th.*, 41(1995), 1303 - 1321.

[13] V. I. Levenshtein, Universal bounds for codes and designs, *Handbook of Coding Theory* (V. Pless and W. C. Huffman, eds.), vol. 1, Elsevier Science, Amsterdam 1998, 499-648.

[14] M. L. Lomonosov and V. P. Polesskii, Lower bounds of network reliability, *Problems of Information Transmission* 8(1972), 118-123.

[15] J. MacWilliams and N. J. A. Sloane, **The Theory of Error Correcting Codes**, Amsterdam, North-Holland, 1977.

[16] R. J. McEliece, E. R. Rodemich, H. C. Rumsey and L. R. Welch, New upper bounds on the rate of codes via the Delsarte-MacWilliams inequalities, *IEEE Trans. Inf. Th.*, 23(1977), 157-166.

[17] R. Meshulam, An uncertainty inequality for groups of order pq, *Europ. J. Comb.* 13(1992), 401-407.

[18] V. Milman and G. Schechtman, **The asymptotic theory of finite dimensional normed spaces**, Berlin, Springer 1986 (LNM, 1200).

[19] J. G. Oxley, **Matroid Theory**, Oxford University Press, 1992.

[20] A. Samorodnitsky, On the optimum of Delsarte's linear program, *J. Comb. Th., Ser. A*, to appear.

[21] P. Solé, A limit law on the distance distribution of binary codes, *IEEE Trans. Inf. Th.*, 36, 1990, 229 - 232.

[22] G. Szego, **Orthogonal Polynomials**, American Mathematical Society, 1939.

[23] J. H. van Lint, **Introduction to Coding Theory**, Springer-Verlag, 1982.

# 6 Appendix

**Proof:** Of Corollary 2.7

We have to prove that if $C$ is a code in $Z_2^n$ with minimal distance $d$, then

$$|C| \leq 2^{n \cdot \left[H\left(\frac{1}{2} - \sqrt{\frac{d}{n}(1-\frac{d}{n})}\right) + c_0 \cdot \frac{\log n}{\sqrt{n}}\right]}, \tag{1}$$

where $c_0$ is some absolute constant.

Lemma 2.6 implies $n \cdot H\left(\frac{1}{2} - \sqrt{\frac{d}{n}(1-\frac{d}{n})}\right) \geq n - c_1 d(1-\frac{d}{n}) \geq n - c_1 d$, for some absolute constant $c_1 > 0$. Observe that if we choose $c_0 \geq c_1$, then for $d \leq \sqrt{n}\log n$ the claim is trivially true, since the RHS of (1) is greater than $2^n$.

Next, we show that if $c_2$ is a sufficiently small constant, and $2 < d < c_2 n$, then

$$|C| \leq 2^{n \cdot H\left(\frac{1}{2} - \sqrt{\frac{d}{n}(1-\frac{d}{n})}\right)}.$$

Indeed, by the Hamming bound [15], $|C| \leq \frac{2^n}{\sum_{i=0}^{(d-1)/2} \binom{n}{i}}$. The claim now follows since $\binom{n}{\lfloor (d-1)/2 \rfloor} \geq 2^{c_1 d}$ for $2 < d < \frac{n}{2^{3c_1+1}}$.

Next, we deal with $d > n/2$. For this range of $d$, the Plotkin inequality [15] gives

$$|C| \leq \frac{2d}{2d-n},$$

implying (1).

Assume now that $c_2 n \leq d \leq n/2$. We intend to apply theorem 2.5, and so we are looking for a $k$ such that $x_k \leq d$. Using the bound $x_k \leq n/2 - \sqrt{(n-k+2)(\sqrt{k}-1)^2}$, we want to solve for $k$ the inequality

$$(n-k)(3\sqrt{k}-1)^2 \geq (n/2-d)^2. \tag{2}$$

We are going to choose $k$ of the form $n/2-\sqrt{d(n-d)}+c_3\sqrt{n}$, where a sufficiently large constant $c_3$ should do the job. Indeed, (2) transforms to

$$2c_3\sqrt{nd(n-d)} \geq c_3^2 n + (n-k)(2\sqrt{k}-1). \tag{3}$$

Note that for $c_2 n \leq d \leq n/2$, the LHS of (3) is at least $2\sqrt{c_2(1-c_2)}c_3 n^{3/2}$, while the LHS is at most $2n^{3/2}+o(n^{3/2})$. Therefore, for a sufficiently large $c_3$, (3) holds.

It remains to apply theorem 2.5. We get

$$|C| \leq \binom{n}{n/2-\sqrt{d(n-d)}+c_3\sqrt{n}} \cdot (n+1)^2 \leq 2^{n\cdot\left[H\left(\frac{1}{2}-\sqrt{\frac{d}{n}(1-\frac{d}{n})}\right)+O\left(\frac{\log n}{\sqrt{n}}\right)\right]},$$

completing the proof. ∎