

ONE MORE PROOF OF  
THE FIRST LINEAR PROGRAMMING BOUND  
FOR BINARY CODES AND TWO CONJECTURES

BY

ALEX SAMORODNITSKY

*The Institute of Computer Science, The Hebrew University of Jerusalem  
Givat Ram, Jerusalem 91904, Israel  
e-mail: salex@cs.huji.ac.il*

*Dedicated to Nati Linial on his 70th birthday*

ABSTRACT

We give one more proof of the first linear programming bound for binary codes, following the line of work initiated by Friedman and Tillich [9]. The new argument is somewhat similar to the one given in [24], but we believe it to be both simpler and more intuitive. Moreover, it provides the following ‘geometric’ explanation for the bound. A binary code with minimal distance  $\delta n$  is small because the projections of the characteristic functions of its elements on the subspace spanned by the Walsh–Fourier characters of weight up to  $(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n$  are essentially independent. Hence the cardinality of the code is bounded by the dimension of the subspace.

We present two conjectures, suggested by the new proof, one for linear and one for general binary codes which, if true, would lead to an improvement of the first linear programming bound. The conjecture for linear codes is related to and is influenced by conjectures of Håstad and of Kalai and Linial. We verify the conjectures for the (simple) cases of random linear codes and general random codes.

## 1. Introduction

A binary error-correcting code  $C$  of length  $n$  and minimal distance  $d$  is a subset of the Hamming cube  $\{0, 1\}^n$  in which the distance between any two distinct points is at least  $d$ . Let  $A(n, d)$  be the maximal size of such a code. In this paper we are interested in the case in which the distance  $d$  is linear in the length  $n$  of the code, and we let  $n$  go to infinity. In this case  $A(n, d)$  is known (see, e.g., [18]) to grow exponentially in  $n$ , and we consider the quantity

$$R(\delta) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 A(n, \lfloor \delta n \rfloor),$$

also known as the **asymptotic maximal rate** of the code with relative distance  $\delta$ , for  $0 \leq \delta \leq \frac{1}{2}$ .

The best known upper bounds on  $R(\delta)$  were obtained in [22] using the linear programming relaxation, constructed in [7], of the combinatorial problem of bounding  $A(n, d)$ . While the precise value of the linear program of [7] is still unknown, there is convincing numerical evidence [5] that on the exponential scale the bounds of [22] are the best possible to derive from this program. It follows that in order to improve these bounds we need either to augment the linear program of [7], or to look for a different way to prove the bounds of [22]. The first approach was adopted by [26] (see also, e.g., [3]), who suggested a positive semidefinite relaxation of the problem to bound  $A(n, d)$ , augmenting the linear programming relaxation of [7] by studying the geometry of a code in more detail (see also the discussion before Proposition 1.8 below). The second approach was taken by [9], where the first linear programming bound for linear binary codes was proved by a different and a more direct argument. Given a linear code  $C$ , that is, a linear subspace of the Hamming cube, [9] proved comparison theorems (adapting ideas from Riemannian geometry to the discrete setting) between two metric spaces defined by the two Cayley graphs: The Hamming cube  $\{0, 1\}^n$  and the Cayley graph of the quotient  $\{0, 1\}^n/C^\perp$  with respect to the set of generators given by the standard basis of  $\{0, 1\}^n$ . The key observations were that the metric balls in  $\{0, 1\}^n$  grow faster than their counterparts in  $\{0, 1\}^n/C^\perp$ , while their eigenvalues (the eigenvalue of a set is the maximal eigenvalue of the adjacency matrix of the graph restricted to this set) are bounded from above by these of their counterparts.

Following [9], where the importance of working with Hamming balls and their eigenvalues was established, the expediency of working with the maximal eigenfunctions of Hamming balls was observed in [23, 24]. It was, in effect, shown that, given any nonnegative function  $f$  on  $\{0, 1\}^n$  with a small support, such that the adjacency matrix of the Hamming cube acts on  $f$  by multiplying it pointwise by a large factor, one can obtain an upper bound on the cardinality of error-correcting codes, whose applicability will depend on the cardinality of the support of  $f$  and on the size of the multiplying factor. Using the maximal eigenfunctions of Hamming balls of different radii, with their corresponding parameters, led to a simple proof of the first linear programming bound for linear codes, which was then extended to prove the bound for general binary codes as well. One appealing feature of the argument for linear codes was that it established the following ‘covering’ explanation for the first linear programming bound (stated explicitly in [24] and contained implicitly in [9]). A linear code  $C$  with minimal distance  $d$  is small, because its dual  $C^\perp$  is large, in the following sense: A union of Hamming balls of radius  $r = r(d)$  centered at the points of  $C^\perp$  covers almost the whole space. This implies that, up to negligible errors,  $|C| \leq |B|$ , where  $B$  is a ball of radius  $r$ . Unfortunately, the extension of the argument to general binary codes seemed to allow no such natural geometric interpretation.

In this paper we give another proof of the first linear programming bound for general binary codes. This proof is somewhat similar to the one given in [24], but we believe it to be both simpler and more satisfactory, in that it provides a ‘geometric’ explanation for the bound. We show that a binary code with minimal distance  $\delta n$  is small because the projections of the characteristic functions of its elements on the subspace spanned by the Walsh–Fourier characters of weight up to  $(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n$  are essentially independent. Hence the cardinality of the code is essentially bounded by the dimension of the subspace. Let  $0 \leq r \leq n$ , and let  $\Lambda_r$  be the orthogonal projection on the span of the Walsh–Fourier characters of weight at most  $r$  (see Section 1.3 for background and definitions of relevant notions). For  $x \in \{0, 1\}^n$ , let  $\delta_x$  be the characteristic function of the point  $x$ . Let  $\langle v_1, \dots, v_N \rangle$  denote the linear span of the vectors  $v_1, \dots, v_N$ . We prove the following claim.

**THEOREM 1.1:** *Let  $0 < \delta < \frac{1}{2}$ . There exists a function  $r = r_\delta : \mathbb{N} \rightarrow \mathbb{N}$ , with  $r(n) = (\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n + o(n)$ , such that for any code  $C$  of length  $n$  and minimal distance  $d = \lfloor \delta n \rfloor$  we have*

$$\dim(\langle \{\Lambda_{r(n)}\delta_x\}_{x \in C} \rangle) \geq \frac{1}{2d} \cdot |C|.$$

Let us make several comments about this result.

– It follows that  $|C| \leq 2d \cdot \sum_{k=0}^{r(n)} \binom{n}{k}$ . By the known exponential estimates for binomial coefficients (see (4) below) this implies the first linear programming bound on the asymptotic rate function:

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

where  $H(x)$  is the binary entropy function.

– If  $C$  is a linear code, then it is not hard to see that

$$\dim(\langle \{\Lambda_r\delta_x\}_{x \in C} \rangle) = \frac{|\bigcup_{z \in C^\perp} (z + B(r))|}{|C^\perp|},$$

where  $B(r)$  is the Hamming ball of radius  $r$  around zero. Hence, for linear codes the claim of the theorem reduces to saying that the union of Hamming balls of radius  $r(n)$  centered at the points of  $C^\perp$  covers at least  $\frac{1}{2d}$ -fraction of the space. In this sense, the claim of the theorem is a proper generalization of the covering argument for linear codes given in [24].

– The span  $V_r$  of the Walsh–Fourier characters of weight at most  $r$  is the space spanned by the eigenfunctions of the Laplacian operator on  $\{0, 1\}^n$  corresponding to its smallest eigenvalues  $0, 2, 4, \dots, 2r$ . In some metric spaces  $X$  the spaces  $V_r$  spanned by the eigenfunctions of the Laplacian corresponding to its lowest eigenvalues are the spaces of the  $(r)$ -‘simple’ functions on  $X$ . For instance, if  $X$  is the Hamming cube  $\{0, 1\}^n$ , or the Euclidean sphere  $\mathbb{S}^{n-1}$ , then  $V_r$  is the space of the real multivariate polynomials of degree at most  $r$  on  $\mathbb{R}^n$  restricted to either  $\{0, 1\}^n$  or  $\mathbb{S}^{n-1}$ . One can ask for the value of  $r$  for which the space  $V_r$  becomes ‘complex’ enough to describe a given distance  $d$  in the ambient space, in the sense that the projection of the characteristic function of any metric ball of radius  $d$  in the space on  $V_r$  retains a significant fraction of its  $\ell_2$  norm. For the Hamming cube  $\{0, 1\}^n$  and  $d = \delta n$ , the appropriate value of  $r$  is  $(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n + o(n)$ . (This is closely related to the fact that the Krawchouk polynomial  $K_d$  essentially attains its  $\ell_2$  norm very close to its first root, but not much before that; see, e.g., (11) and Proposition 2.15 in [16].)

In other words, roughly speaking, the cardinality of a binary code of minimal distance  $d$  is upper bounded by the dimension of the space  $V_r$  spanned by the ‘simple’ eigenfunctions of the Laplacian if  $r$  is large enough for the space  $V_r$  to describe distances  $d$  in  $\{0, 1\}^n$ . Let us remark that this phenomenon can also be shown to hold if the ambient space is the Hamming sphere (in effect recovering, via the Bassalygo–Elias inequality, the **second linear programming bound** for binary codes), and we believe that it should be possible to show this, by similar methods, for other symmetric spaces, such as distance regular graphs or the Euclidean sphere. This seems to be rather intriguing, and we wonder whether this could be a special case of a more general principle.

– The proof of Theorem 1.1 relies on the existence of a nonnegative function  $f$  on  $\{0, 1\}^n$  with a small support (in fact it suffices to require that  $\frac{\mathbb{E} f^2}{\mathbb{E} f}$  is large), such that the adjacency matrix  $A$  of the Hamming cube acts on  $f$  by multiplying it pointwise by a large factor. It is not hard to see that any such function can be used to construct a feasible solution to the dual linear program of [7].<sup>1</sup> In particular, if  $f$  is the maximal eigenfunction of a Hamming ball, we (essentially) recover a solution to this program constructed in [22]. In this sense, this line of research is subsumed by that of [7] and [22]. In addition, it can be shown [25] that the best bound one can obtain following this approach is the first linear programming bound. With that, we believe that this approach leads to simpler proofs of this bound which provide additional information (we do not know how to derive the claim of the theorem from the linear program of [7]) and furthermore suggest new possible ways to proceed in order to improve this bound. In fact, we present two conjectures, suggested by the new proof, one for linear and one for general binary codes which, if true, would lead to an improvement of the first linear programming bound. We start with discussing the conjecture for linear codes.

1.1. LINEAR CODES. Before stating the conjecture, let us mention two related conjectures that influenced it. Both conjectures posit, in different ways, that the behavior of a linear code near its minimal weight is highly constrained.

---

<sup>1</sup> Let  $f \geq 0$  with  $\frac{\mathbb{E} f^2}{\mathbb{E} f} \geq \frac{2^n}{s}$ , so that  $Af \geq \lambda f$ , for some  $\lambda \geq 1$ . Let  $G = (Af) * f - (\lambda - 1)(f * f)$ . It is easy to see that  $\hat{G}$  is a feasible solution to the dual program of [7] for codes with minimal distance  $d = \frac{n - \lambda + 1}{2}$ , and the bound we get from the linear program is  $A(n, d) \leq s$ .

The first conjecture, due to Kalai and Linial [13], states that if  $C \subseteq \{0, 1\}^n$  is a linear code with minimal distance  $d$ , then the number of codewords of weight  $d$  in  $C$  is at most subexponential in  $n$ . In fact, they implicitly conjecture more, namely that there is also a strong upper bound on the number of codewords of weight close to  $d$ . They observe that if this is true, then the first linear programming bound for linear codes could be improved. This was elucidated in the subsequent work of [1], where it was shown (among other things) that a linear code attaining the first linear programming bound must have (up to a negligible error) as many codewords of some weight close to  $d$  as a random code of the same cardinality.

A weak version of this conjecture, namely that the number of vectors of weight close to  $d$  is exponentially smaller than the cardinality of  $C$  (assuming  $C$  is exponentially large, which is the interesting case here), was proved in [19]. However, the full conjecture was shown to be false in [2], where a code of minimal distance  $d$  with exponentially many codewords of weight  $d$  was constructed.

The second conjecture is due to Håstad [11]. It states that for any absolute constants  $0 < \alpha < 1$  and  $k \geq 1$  there exists an absolute constant  $K = K(\alpha, k)$ , such that the following is true. Let  $C$  be a nice (pseudorandom in some sense) linear code of length  $n$  with minimal distance  $d \leq n^\alpha$ . Then for any non-zero function  $f$  on  $\{0, 1\}^n$  whose Fourier transform is supported on vectors of weight at most  $kd$  in  $C$  we have

$$(1) \quad \frac{\|f\|_4}{\|f\|_2} \leq K,$$

Some comments:

– As pointed out in [11] some precondition on  $C$  is necessary. To see this, let  $d = \lfloor n^\alpha \rfloor$ , and let  $k = 3$ . Let  $m = 3d$ , and let  $C' \subseteq \{0, 1\}^m$  be a linear code of minimal distance  $d$  and dimension linear in  $m$ . Add  $n - m$  zero coordinates to each vector in  $C'$ , obtaining a code  $C \subseteq \{0, 1\}^n$ . Take

$$f = \sum_{x \in C, |x| \leq kd} W_x$$

(here  $\{W_x\}_x$  are the Walsh–Fourier characters of  $\{0, 1\}^n$ ). Then it is easy to see that  $f$  is essentially proportional to the characteristic function of the dual code  $C^\perp$ , and in particular  $\frac{\|f\|_4}{\|f\|_2}$  is exponential in  $m$ .

– The inequality (1) is a Khintchine-type inequality. Recall that Khintchine-type inequalities establish an upper bound on the ratio of two  $\ell_p$  norms for

functions coming from a certain linear space, typically a space of multivariate polynomials of a specified degree over a given product space. In particular, the prototypical Khintchine inequality [14] states that the ratio of  $\ell_2$  and  $\ell_1$  norms of linear polynomials over the boolean cube  $\{0, 1\}^n$  is bounded by an absolute constant. See [12] for a recent discussion and references.

Our conjecture is in a sense a combination of the two conjectures above in that it considers the set of vectors in a linear code whose weight is close to the minimal distance of the code, but it replaces the ‘hard’ cardinality constraint of [13] by a ‘softer’ analytical constraint of [11].

CONJECTURE 1.2: *Let  $0 < \delta < \frac{1}{2}$ . Then for*

$$c = c(\delta) = \frac{\delta(1 - 2\delta) \cdot \log_2 \left( \frac{\frac{1}{2} + \sqrt{\delta(1-\delta)}}{\frac{1}{2} - \sqrt{\delta(1-\delta)}} \right)}{16\sqrt{\delta(1 - \delta)}},$$

*there is a positive constant  $\epsilon_0 = \epsilon_0(\delta)$  such that for any  $0 \leq \epsilon \leq \epsilon_0$  the following holds. Let  $C \subseteq \{0, 1\}^n$  be a linear code with minimal distance  $d = \lfloor \delta n \rfloor$ , let  $d \leq i \leq (1 + \epsilon)d$ , and let  $A$  be the set of vectors of weight  $i$  in  $C$ . Assume  $A \neq \emptyset$ . Let  $f$  be a non-zero function on  $\{0, 1\}^n$  whose Fourier transform is supported on  $A$ . Then*

$$(2) \quad \frac{\|f\|_4}{\|f\|_2} \leq 2^{c\epsilon n + o(n)}.$$

Some comments:

- If  $f$  is a non-zero function whose Fourier transform is supported on the vectors of minimal weight in a linear code, we conjecture that the ratio  $\frac{\|f\|_4}{\|f\|_2}$  is at most subexponential in  $n$ . Since the ratio of the fourth and the second norm of a function is upper-bounded by the fourth root of the cardinality of its Fourier support (see, e.g., Proposition 1.1 in [15]) this conjecture is weaker than the corresponding conjecture of [13].
- For weights close to minimal, the upper bounds on the ratio of the fourth and the second norm required in (2) are in general smaller than the explicit bounds on cardinality required in [1]. On the other hand, the quantity we want to bound in (2) is smaller, and in general could be much smaller (cf. Proposition 1.4 in which for  $i = (1 + \epsilon)d$  with  $\epsilon > 0$ , the set of vectors of weight  $i$  in the code is exponentially large, but the ratio of the moments is bounded by a constant). In this sense, the two conjectures are incomparable.

– Conjecture 1.2 and the conjecture of [11] are also incomparable. Apart from the fact that the two conjectures speak about codes in different regimes (the conjecture of [11] considers codes with sublinear distance, while we are interested in codes with linear distance), the conclusion of Conjecture 1.2 has to hold for all linear codes, and not only for the ‘nice’ ones. On the other hand, the conclusion itself is much weaker, while the conditions under which it is supposed to hold are stronger.

We claim that if Conjecture 1.2 holds, then the first linear programming bound for binary linear codes can be improved. In fact, it suffices to prove the conjecture only for symmetric functions  $f$ . Here we call function  $f$  **symmetric** if for any point  $x$  in its Fourier support,  $\widehat{f}(x)$  depends only on  $|x|$ . Let  $A_L(n, d)$  be the maximal size of a linear code of length  $n$  and distance  $d$ . Let

$$R_L(\delta) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 A_L(n, \lfloor \delta n \rfloor).$$

PROPOSITION 1.3: *Assume that Conjecture 1.2 holds for symmetric functions  $f$ . Then for all  $0 < \delta < \frac{1}{2}$  we have*

$$R_L(\delta) \leq H\left(\frac{1}{2} - \sqrt{\delta(1 - \delta)}\right) - \theta_L(\delta),$$

where  $\theta_L(\delta) > 0$  for all  $0 < \delta < \frac{1}{2}$ .

Some comments:

– In the notation of Conjecture 1.2, if  $f$  is symmetric, then  $\widehat{f}$  is constant on  $A$  and we may assume, without loss of generality, that  $f = \sum_{a \in A} W_a$ . With that, this special case essentially captures the complexity of the conjecture in full generality. In fact, it is known (see e.g., Proposition 1.1 in [15]) that for any subset  $A \subseteq \{0, 1\}^n$  the maximum of the ratio  $\frac{\|f\|_4}{\|f\|_2^2}$  over non-zero functions  $f$  whose Fourier transform is supported on  $A$  is attained, up to a polylogarithmic in  $|A|$  factor, on the characteristic function of some subset  $B \subseteq A$ . Hence, it would suffice to prove the conjecture for the linear code  $C'$  spanned by the vectors in  $B$  and for the appropriate symmetric function.

– The fourth and the second norm in (2) may be replaced by any two norms  $q > p$ , changing the value of the constant  $c(\delta)$  accordingly.

We show that Conjecture 1.2 holds (in a strong sense) for random linear codes. A random linear code  $C$  of length  $n$  and (prescribed) dimension  $k$  is chosen as follows (see e.g., [4] for this and for properties of random linear codes): choose  $k$  vectors  $v_1, \dots, v_k$  independently at random from  $\{0, 1\}^n$  and take  $C$  to



be the linear span of these vectors. It is convenient to define parameters in the following claim in terms of the dimension rather than the minimal distance of a code. This is justified by the following fact. Let  $0 < R < 1$ . A random linear code of dimension  $k = \lfloor Rn \rfloor$  has minimal distance  $H^{-1}(1 - R) \cdot n \pm o(n)$  with probability tending to 1 with  $n$ . So in case of random linear codes we may speak about the dimension of the code and its minimal distance interchangeably.

PROPOSITION 1.4: *Let  $0 < R < 1$ . There exists a positive constant  $K = K(R)$  and a positive constant  $\epsilon_0 = \epsilon_0(R)$  such that the following holds with probability tending to 1 with  $n$  for a random linear code  $C \subseteq \{0, 1\}^n$  of dimension  $\lfloor Rn \rfloor$ . Let  $d$  be the minimal distance of  $C$ , let  $d \leq i \leq (1 + \epsilon_0)d$ , and let  $A$  be the set of vectors of weight  $i$  in  $C$ . Assume  $A \neq \emptyset$ . Let  $f$  be a function whose Fourier transform is supported on  $A$ . Then*

$$\frac{\|f\|_4}{\|f\|_2} \leq K.$$

1.2. GENERAL CODES. We start with some definitions and a preliminary discussion.

Definition 1.5: For a binary code  $C$  and  $1 \leq i \leq n$  let  $G(C, i)$ , the distance- $i$  graph of  $C$ , be the graph with  $|C|$  vertices indexed by the elements of  $C$ , with two vertices connected by an edge iff the corresponding elements of  $C$  are at distance  $i$  from each other.

If  $C$  is a linear code, then  $G(C, i)$  is a regular graph, for any  $1 \leq i \leq n$ . For a general code  $C$  the graphs  $G(C, i)$  could be highly irregular, which will lead to difficulties (see below) in formulating a conjecture for general codes analogous to Conjecture 1.2, which is not disproved by a simple counterexample. Consequently, we will need to introduce additional constraints in the conjecture below. Specifically, we will require the graphs  $G(C, i)$  for  $i$  close to the minimal distance of the code  $C$  to behave somewhat similarly to regular graphs. In a regular graph  $G$ , the density of edges of any induced subgraph  $G'$  is bounded by that of the whole graph. We will call a graph  $t$ -balanced if it shares this property of regular graphs, up to a multiplicative factor of  $t$ .

*Definition 1.6:* A non-empty graph  $G = (V, E)$  will be called  $t$ -balanced, for some  $t \geq 1$ , if the density of edges in any of its induced subgraphs is at most  $t$  times the density of edges in  $G$ . That is, denoting by  $E(X, X)$  the set of edges from a subset  $X$  of vertices to itself, for any  $X \subseteq V$  we have

$$\frac{|E(X, X)|}{|X|} \leq t \cdot \frac{|E|}{|V|}.$$

We can now state our conjecture for general graphs.

CONJECTURE 1.7: Let  $0 < \delta < \frac{1}{2}$ . Then for

$$c = c(\delta) = \frac{\delta(1 - 2\delta) \cdot \log_2 \left( \frac{\frac{1}{2} + \sqrt{\delta(1-\delta)}}{\frac{1}{2} - \sqrt{\delta(1-\delta)}} \right)}{16\sqrt{\delta(1-\delta)}},$$

there is a positive constant  $\epsilon_0 = \epsilon_0(\delta)$ , such that for any  $0 \leq \epsilon \leq \epsilon_0$  the following holds. Let  $C \subseteq \{0, 1\}^n$  be a code with minimal distance  $d = \lfloor \delta n \rfloor$ , let  $d \leq i \leq (1 + \epsilon)d$ , and let  $G = G(C, i)$  be the distance- $i$  graph of  $G$ . Assume that  $G$  is  $t$ -balanced, for  $t = t(n) = 2^{o(n)}$ . Let  $\lambda \in \mathbb{R}^C$  be the vector of eigenvalues of  $G$ , viewed as a function on  $C$ , where we endow  $C$  with the uniform probability measure. Then

$$(3) \quad \frac{\|\lambda\|_4}{\|\lambda\|_2} \leq 2^{c\epsilon n + o(n)}.$$

Let us make several comments about this conjecture.

- If  $C$  is a linear code, then  $G = G(C, i)$  is a regular graph, and hence it is 1-balanced. Furthermore, if  $A$  is the set of vectors of weight  $i$  in  $C$ , then it is easy to see that the distribution of the eigenvalues of  $G$  on  $C$  is the same as the distribution of the function  $f = \sum_{a \in A} W_a$  on  $\{0, 1\}^n$  (see also the discussion in the proof of Proposition 1.3 below). Hence Conjecture 1.7 generalizes Conjecture 1.2 for symmetric functions.

- Some additional condition on  $G = G(C, i)$  is necessary. Indeed, let  $C$  be an exponentially large code with minimal distance  $d$  which is linear in  $n$ . Let  $k = \lfloor d/3 \rfloor$ . Assume, without loss of generality, that  $k$  is even. Choose a point  $x \in C$ , and choose two points  $y, z \in \{0, 1\}^n$  both at distance  $k$  from  $x$ , such that the distance between  $y$  and  $z$  is also  $k$  (clearly such points exist). Add  $y$  and  $z$  to  $C$ , obtaining a new code  $C'$  with minimal distance  $k$ . It is easy to see that the graph  $G(C, k)$  has only 3 non-zero eigenvalues, and hence

$$\frac{\|\lambda\|_4}{\|\lambda\|_2} \geq \Omega(|C|^{\frac{1}{4}}).$$

– Recall that if  $\lambda$  is the vector of eigenvalues of a graph  $G$ , then  $\sum_i \lambda_i^k$  equals to the number of closed walks of length  $k$  in  $G$  (see, e.g, Lemma 2.5 in [6]). Hence Conjecture 1.7 can be informally restated as follows: Let  $C$  be a binary code of length  $n$  and distance  $d$ . Let  $i$  be close to  $d$  and assume that the graph  $G(C, i)$  is  $2^{o(n)}$ -balanced (which always holds if  $C$  is linear). Then the number of ‘rhombic’ 4-tuples  $(x, y, z, w) \in C^4$  with

$$|y - z| = |z - y| = |w - z| = |x - w| = i$$

is not much larger than  $|C|B_i^2$ , where

$$B_i = \frac{1}{|C|} |\{(x, y) \in C \times C, |x - y| = i\}|$$

is the  $i^{th}$  component of the distance distribution vector of  $C$ .

– Continuing from the preceding comment, we recall that the distance distribution vector  $(B_0, \dots, B_n)$  of a code is a central object of study in the linear programming approach of [7]. The positive semidefinite approach of [26] (see also [3]) studies the geometry of a code in more detail, collecting the statistics of the possible  $\binom{k}{2}$ -tuples of pairwise inner distances in all  $k$ -tuples of elements of a code, for some  $k \geq 2$ . In particular, statistics of inner distances of quadruples of codewords are studied in [10]. In this sense Conjecture 1.7 points out a possible connection between the two above-mentioned approaches whose eventual goal is to improve the linear programming bounds

We claim that if Conjecture 1.7 holds, then the first linear programming bound for binary codes can be improved.

PROPOSITION 1.8: *Assume that Conjecture 1.7 holds. Then for all  $0 < \delta < \frac{1}{2}$  we have*

$$R(\delta) \leq H\left(\frac{1}{2} - \sqrt{\delta(1 - \delta)}\right) - \theta(\delta),$$

where  $\theta(\delta) > 0$  for all  $0 < \delta < \frac{1}{2}$ .

We show that Conjecture 1.7 holds (in a strong sense) for random codes. See Section 1.3 for more details on the (standard) model of random codes that we use. Let  $0 < R < 1$ . A random code of cardinality  $2^{Rn}$  has minimal distance  $H^{-1}(1 - R) \cdot n \pm o(n)$  with probability tending to 1 with  $n$ . So in case of random codes we may speak about the cardinality of the code and its minimal distance interchangeably.

PROPOSITION 1.9: *Let  $0 < R < 1$ . There exists a positive constant  $K = K(R)$  and a positive constant  $\epsilon_0 = \epsilon_0(R)$  such that the following holds with probability tending to 1 with  $n$  for a random code  $C \subseteq \{0, 1\}^n$  of cardinality  $2^{Rn}$ . Let  $d$  be the minimal distance of  $C$ , let  $d \leq i \leq (1 + \epsilon_0)d$ , and let  $G = G(C, i)$  be the distance- $i$  graph of  $G$ . Let  $\lambda \in \mathbb{R}^C$  be the vector of eigenvalues of  $G$ , viewed as a function on  $C$ , where we endow  $C$  with the uniform probability measure. Then*

$$\frac{\|\lambda\|_4}{\|\lambda\|_2} \leq K.$$

ORGANIZATION OF THIS PAPER. The remainder of this paper is organized as follows. We describe the relevant notions and provide some additional background in the next subsection. Theorem 1.1 is proved in Section 2. Propositions 1.3 and 1.8 are proved in Section 3. Propositions 1.4 and 1.9 are proved in Section 4.

1.3. BACKGROUND, DEFINITIONS, AND NOTATION. We view  $\{0, 1\}^n$  as a metric space, with the **Hamming distance** between  $x, y \in \{0, 1\}^n$  given by

$$|x - y| = |\{i : x_i \neq y_i\}|.$$

The **Hamming weight** of  $x \in \{0, 1\}^n$  is

$$|x| = |\{i : x_i = 1\}|.$$

For  $x, y \in \{0, 1\}^n$ , we write  $x + y$  for the modulo 2 sum of  $x$  and  $y$ . Note that the weight  $|x + y|$  of  $x + y$  equals the distance  $|x - y|$  between  $x$  and  $y$  (we will use this simple observation several times below). The **Hamming sphere** of radius  $r$  centered at  $x$  is the set

$$S(x, r) = \{y \in \{0, 1\}^n : |x - y| = r\}.$$

The **Hamming ball** of radius  $r$  centered at  $x$  is the set

$$B(x, r) = \{y \in \{0, 1\}^n : |x - y| \leq r\}.$$

Clearly, for any  $x \in \{0, 1\}^n$  and  $0 \leq r \leq n$  we have

$$|S(x, r)| = \binom{n}{r} \quad \text{and} \quad |B(x, r)| = \sum_{k=0}^r \binom{n}{k}.$$

Let

$$H(t) = t \log_2 \left( \frac{1}{t} \right) + (1 - t) \log_2 \left( \frac{1}{1 - t} \right)$$

be the binary entropy function. We will make use of the following estimate (see, e.g., Theorem 1.4.5. in [20]): For  $x \in \{0, 1\}^n$  and  $0 < r \leq \frac{n}{2}$

$$(4) \quad |B(x, r)| \leq 2^{H(\frac{r}{n}) \cdot n}.$$

The asymptotic notation will always refer to the behavior of a function of an integer argument  $n$  when  $n$  tends to infinity (unless specifically stated otherwise). The  $O$ ,  $\Omega$  and  $\Theta$  asymptotic notation always hides absolute constants.

We write  $a \in b \pm \epsilon$  as a shorthand for  $b - \epsilon \leq a \leq b + \epsilon$ .

1.3.1. *Fourier analysis, Krawtchouk polynomials, and spectral projections.* We recall some basic notions in Fourier analysis on the boolean cube (see, e.g., [8, Section 1.5]). For  $\alpha \in \{0, 1\}^n$ , define the Walsh–Fourier character  $W_\alpha$  on  $\{0, 1\}^n$  by setting

$$W_\alpha(y) = (-1)^{\sum \alpha_i y_i} \quad \text{for all } y \in \{0, 1\}^n.$$

The **weight** of the character  $W_\alpha$  is the Hamming weight  $|\alpha|$  of  $\alpha$ . The characters  $\{W_\alpha\}_{\alpha \in \{0,1\}^n}$  form an orthonormal basis in the space of real-valued functions on  $\{0, 1\}^n$ , under the inner product  $\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x)$ . The expansion  $f = \sum_{\alpha \in \{0,1\}^n} \hat{f}(\alpha)W_\alpha$  defines the Fourier transform  $\hat{f}$  of  $f$ . Explicitly,

$$\hat{f}(\alpha) = \langle f, W_\alpha \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)W_\alpha(x).$$

It will be convenient to denote by

$$\langle \hat{f}, \hat{g} \rangle_{\mathcal{F}} := \sum_{\alpha \in \{0,1\}^n} \hat{f}(\alpha)\hat{g}(\alpha)$$

the inner product of functions in the “Fourier domain”. Note that this inner product is not normalized.

The **Parseval identity** states that

$$\langle f, g \rangle = \sum_{\alpha \in \{0,1\}^n} \hat{f}(\alpha)\hat{g}(\alpha) = \langle \hat{f}, \hat{g} \rangle_{\mathcal{F}}.$$

The **convolution** of  $f$  and  $g$  is defined by

$$(f * g)(x) = \frac{1}{2^n} \sum_{y \in \{0,1\}^n} f(y)g(x + y).$$

The convolution transforms to dot product:  $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$ . The convolution operator is commutative and associative. We will use one additional simple fact. Let  $A$  be the adjacency matrix of  $\{0, 1\}^n$ . Then  $\widehat{Af}(\alpha) = (n - 2|\alpha|)\widehat{f}(\alpha)$  (in fact, note that  $A$  is a symmetric matrix, and it is easy to verify that  $W_\alpha$  is an eigenvector of  $A$  with eigenvalue  $n - 2|\alpha|$ ).

*Krawtchouk polynomials.* (see, e.g., [17]). For  $0 \leq s \leq n$ , let  $F_s$  be the sum of all Walsh–Fourier characters of weight  $s$ , that is  $F_s = \sum_{|\alpha|=s} W_\alpha$ . Note that  $F_s$  is the Fourier transform of  $2^n \cdot L_s$ , where  $L_s$  is the characteristic function of the Hamming sphere of radius  $s$  around 0. It is easy to see that  $F_s(x)$  depends only on the Hamming weight  $|x|$  of  $x$ , and it can be viewed as a univariate function on the integer points  $0, \dots, n$ , given by the restriction to  $\{0, \dots, n\}$  of the univariate polynomial

$$K_s = \sum_{k=0}^s (-1)^k \binom{x}{k} \binom{n-x}{s-k}$$

of degree  $s$ , that is,  $F_s(x) = K_s(|x|)$ . The polynomial  $K_s$  is the  $s^{th}$  **Krawtchouk polynomial**. Abusing notation, we will also call  $F_s$  the  $s^{th}$  Krawtchouk polynomial, and write  $K_s$  for  $F_s$  when the context is clear.

*Spectral projections.* For  $0 \leq r \leq n$  we define  $\Lambda_r$  to be the orthogonal projection to the subspace spanned by Walsh–Fourier characters of weight at most  $r$ . That is, for a function  $f$  on  $\{0, 1\}^n$ , and  $0 \leq r \leq n$ , we have

$$\Lambda_r f = \sum_{|\alpha| \leq r} \widehat{f}(\alpha) W_\alpha.$$

1.3.2. *Bounds on the asymptotic rate function.* The best known lower bound on  $R(\delta)$  is the **Gilbert–Varshamov bound**  $R(\delta) \geq 1 - H(\delta)$  (see, e.g., [21]). The existence of codes asymptotically attaining this bound is demonstrated, e.g., by random codes (see Section 1.3.3). The best known upper bounds on  $R(\delta)$  are the linear programming bounds [22], obtained via the linear programming approach of [7]. To be more specific, [7] suggested a systematic approach to obtaining a linear programming relaxation of the combinatorial problem of bounding the cardinality of an error-correcting code with a given minimal distance (equivalently, of a metric ball packing with a given radius of a ball) in a metric space with a large group of isometries and, more generally, in an association scheme. In [22] tools from the theory of orthogonal polynomials were used to construct

good feasible solutions to the linear programs constructed in [7] for the Hamming cube and the Hamming sphere. This led to two families of bounds. The first linear programming bound is obtained by solving the linear program for the cube. Solving the linear program for the sphere gives bounds for codes in the sphere (also known as constant weight codes). Bounds on codes in the sphere lead to bounds for codes in the cube, via the Bassalygo–Elias inequality, reflecting the fact that the sphere (of an appropriate dimension) is a subset of the cube. Optimizing over the radius of the embedded sphere leads to the second linear programming bound. The two bounds coincide for relative distance  $\delta \geq 0.273\dots$ . In the remaining range the second bound is better.

1.3.3. *Random codes.* There are standard models of random binary codes (see, e.g., [4]). A **random linear code**  $C$  of length  $n$  and (prescribed) dimension  $k$  is chosen as follows: choose  $k$  vectors  $v_1, \dots, v_k$  independently at random from  $\{0, 1\}^n$  and take  $C$  to be the linear span of these vectors. If  $k = \lfloor Rn \rfloor$  for some  $0 < R < 1$ , then the following two events hold with probability tending to 1 with  $n$ . The code  $C$  has dimension  $k$  and minimal distance  $d \in H^{-1}(1 - R) \cdot n \pm o(n)$ . Shorthand: Here and below we will write ‘with high probability’ (w.h.p.) as a shorthand for ‘with probability tending to 1 with  $n$ ’.

We pass to nonlinear codes. Let  $0 < R < 1$ . A **random code**  $C$  of (prescribed) cardinality  $N = \lfloor 2^{Rn} \rfloor$  is chosen in two steps. First, we choose  $N$  points  $x_1, \dots, x_N$  independently at random from  $\{0, 1\}^n$ . Next, we erase from this list pairs of points whose distance from each other lies below a certain threshold. There are several essentially equivalent ways to choose this threshold. The model we use is as follows: Fix a sufficiently small (see the discussion before the proof of Proposition 1.9) constant  $\tau = \tau(R)$ , and define  $d_0$  to be the maximal integer between 1 and  $N$  so that

$$\frac{N}{2^n} \sum_{\ell=0}^{d_0-1} \binom{n}{\ell} \leq \tau.$$

For all  $1 \leq i < j \leq N$  such that  $|x_i - x_j| \leq d_0 - 1$ , erase the points  $x_i$  and  $x_j$  from the list  $x_1, \dots, x_N$ . Take  $C$  to be the remaining collection of points. The following two events hold with high probability:  $C \geq \Omega(N)$  and the minimal distance of  $C$  is in  $H^{-1}(1 - R) \cdot n \pm o(n)$ .

**2. Proof of Theorem 1.1**

We consider a code  $C$  of length  $n$  and distance  $d = \lfloor \delta n \rfloor$ . In the first part of the argument we find a nonnegative function  $\phi$  on  $\{0, 1\}^n$  with small support such that the adjacency matrix  $A$  of  $\{0, 1\}^n$  acts on  $\phi$  by multiplying it pointwise by the factor of at least  $n - 2d + 1$ . We chose this function to be the maximal eigenfunction of the Hamming ball of an appropriate radius around zero.

*Notation:* For  $0 \leq r \leq n$ , let  $B_r$  be the Hamming ball of radius  $r$  around 0 in  $\{0, 1\}^n$ . Let  $A_r$  be the adjacency matrix of the subgraph of  $\{0, 1\}^n$  induced by the vertices of  $B_r$  and let  $\lambda_r$  be the maximal eigenvalue of  $A_r$ .

Clearly,  $\lambda_r$  is an increasing function of  $r$ , with  $\lambda_0 = 0$  and  $\lambda_n = n$ . Let  $r_0$  be the smallest value of  $r$  for which the maximal eigenvalue of  $A_r$  is at least  $n - 2d + 1$ . It was shown in [9] (see also Lemma 3.3 in [23] for a direct argument) that  $\lambda_r \geq 2\sqrt{r(n-r)} - o(n)$ . Hence

$$r_0 \leq \left( \frac{1}{2} - \sqrt{\delta(1-\delta)} \right) \cdot n + o(n).$$

Abusing notation, we write  $r$  for  $r_0$  from now on.

*Definition 2.1:* Let  $\phi = \phi_r$  be the maximal eigenfunction of  $A_r$  with  $\|\phi_r\|_2 = 1$ .

(Recall that we consider normalized norms on  $\{0, 1\}^n$ , in particular  $\|\phi_r\|_2^2 = \langle \phi_r, \phi_r \rangle = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} \phi_r^2(x)$ .)

Since  $B_r$  is a connected graph which is invariant under permutations of the coordinates, the function  $\phi_r$  is uniquely defined. It is positive on  $B_r$  and symmetric ( $\phi_r(S) = \phi_r(|S|)$ , for  $0 \leq |S| \leq r$ ). We extend  $\phi_r$  to the whole space  $\{0, 1\}^n$  by setting  $\phi_r = 0$  outside  $B_r$  and, abusing notation, write  $\phi_r$  for this extension as well. We record the relevant properties of  $\phi_r$ :

- (1)  $\phi_r$  is supported on  $B_r$ .
- (2)  $\phi_r$  is nonnegative and symmetric.
- (3)  $A\phi_r \geq \lambda_r \cdot \phi_r \geq (n - 2d + 1) \cdot \phi_r$ , with all inequalities holding pointwise on  $\{0, 1\}^n$ .

We use these properties of  $\phi_r$  (from now on we write  $\phi$  for  $\phi_r$ ) to show that for the orthogonal projection  $\Lambda_r$  on the span of the Walsh–Fourier characters of weight at most  $r$  we have  $\dim(\langle \{\Lambda_{r(n)}\delta_x\}_{x \in C} \rangle) \geq \frac{1}{2d} \cdot |C|$ , proving the claim of the theorem.



Let the matrix  $M = M_r$  be defined as follows. The rows of  $M$  are indexed by the elements of  $C$  and the columns by the subsets of  $[n]$  of cardinalities  $0, \dots, r$ , arranged in increasing order of cardinalities (but otherwise arbitrarily). For  $y \in C$  and  $S \subseteq [n]$ , let  $M(y, S) = W_S(y) = (-1)^{\langle y, S \rangle}$  (viewing  $S$  as an element of  $\{0, 1\}^n$ ). Observe that the row of  $M$  indexed by  $x \in C$  contains the non-vanishing part of the Fourier expansion of  $2^n \cdot \Lambda_r(\delta_x)$ , and hence the rank of  $M$  equals  $\dim(\langle \{\Lambda_{r(n)}\delta_x\}_{x \in C} \rangle)$ . Let  $D$  be the  $|B_r| \times |B_r|$  diagonal matrix indexed by the subsets of  $[n]$  of cardinality at most  $r$ , with  $D(S, S) = \phi(S)$  for all  $|S| \leq r$ , and let  $\mathcal{M} = MDM^t$ . Then  $\mathcal{M}$  is a  $|C| \times |C|$  matrix whose rank is upper bounded by the rank of  $M$ . Hence it suffices to show that the rank of  $\mathcal{M}$  is at least  $\frac{1}{2d} \cdot |C|$ . For the remainder of the proof we write  $N$  for  $|C|$ , for typographic convenience.

Let  $\lambda_1, \dots, \lambda_N$  be the eigenvalues of  $\mathcal{M}$ . We will show that

$$\left( \sum_{i=1}^N \lambda_i \right)^2 \geq \frac{N}{2d} \cdot \sum_{i=1}^N \lambda_i^2.$$

This will imply that the number of non-zero eigenvalues is at least  $N/2d$ , proving the claim. In fact, let  $P$  be this number. Then, by the Cauchy–Schwarz inequality,  $(\sum_{i=1}^N \lambda_i)^2 = (\sum_{i=1}^P \lambda_i)^2 \leq P \cdot \sum_{i=1}^P \lambda_i^2 = P \cdot \sum_{i=1}^N \lambda_i^2$ . Writing  $\mathcal{M} = (m_{y,z})_{y,z \in C}$ , we can write this inequality in terms of the entries of  $\mathcal{M}$ :

$$\left( \sum_{y \in C} m_{y,y} \right)^2 \geq \frac{N}{2d} \cdot \sum_{y,z \in C} m_{y,z}^2.$$

Let  $f = 2^n \cdot \widehat{\phi}$ . By the definition of  $\mathcal{M}$ , for  $y, z \in C$  we have

$$m_{y,z} = \sum_{S:|S| \leq r} \phi(S)(-1)^{\langle y+z, S \rangle} = f(y+z).$$

So, we need to show that

$$(5) \quad (2d) \cdot N f^2(0) \geq \sum_{y,z \in C} f^2(y+z).$$

To do this we estimate  $\langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}}$  in two ways. On one hand,

$$\begin{aligned} \langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} &\geq (n - 2d + 1) \cdot \langle \phi * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} \\ &= (n - 2d + 1) \cdot \langle f^2, 1_C * 1_C \rangle \\ &= \frac{n - 2d + 1}{2^{2n}} \cdot \sum_{y,z \in C} f^2(y+z). \end{aligned}$$

We used the first and the second properties of  $\phi$  in the first step, and Parseval’s identity in the second step. Note that  $\phi = \widehat{f}$ . On the other hand,

$$\begin{aligned} \langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} &= \langle (n - 2|x|) \cdot f(x)^2, 1_C * 1_C \rangle \\ &= \frac{1}{2^{2n}} \sum_{y,z \in C} (n - 2|y + z|) f^2(y + z) \\ &\leq \frac{n}{2^{2n}} \cdot N f^2(0) + \frac{n - 2d}{2^{2n}} \cdot \sum_{y \neq z \in C} f^2(y + z) \\ &= \frac{2d}{2^{2n}} \cdot N f^2(0) + \frac{n - 2d}{2^{2n}} \cdot \sum_{y,z \in C} f^2(y + z). \end{aligned}$$

We used Parseval’s identity in the first step, and the fact that  $C$  has distance  $d$  in the third step. Combining the two estimates and simplifying gives (5). ■

### 3. Proof of Propositions 1.3 and 1.8

The proofs will follow the proof of Theorem 1.1. Somewhat imprecisely speaking, if either of Conjectures 1.2 or 1.7 holds, we would be able to replace the Hamming ball of radius  $r(n) = (\frac{1}{2} - \sqrt{\delta(1 - \delta)}) \cdot n + o(n)$  in the argument with a Hamming ball of radius  $r(n) - \Omega(n)$ . We use the same notation as in the proof of Theorem 1.1.

We start with a technical lemma, which provides a useful description of the symmetric function  $f_r = 2^n \widehat{\phi}_r$  (see Definition 2.1). We refer to Section 1.3.1 for the relevant notions in Fourier analysis on  $\{0, 1\}^n$ .

LEMMA 3.1: *Let  $0 \leq r < n$ . Then, for any  $x \in \{0, 1\}^n$  with  $|x| \neq \frac{n - \lambda_r}{2}$  we have*

$$f_r(x) = c \cdot \frac{K_{r+1}(x)}{n - \lambda_r - 2|x|},$$

where  $K_{r+1}$  is the appropriate Krawtchouk polynomial, and  $c$  is a positive constant.

*Proof.* We view  $\phi_r$  as a function on  $\{0, 1\}^n$ . Since it is symmetric and supported on  $B_r$ , we can write  $\phi_r = \sum_{i=0}^r a_i L_i$ , where  $L_i$  is the characteristic function of the Hamming sphere of radius  $i$  around zero, and the coefficients  $a_i$  are positive. Note that

$$A\phi_r = \lambda_r \cdot \phi_r + (r + 1)a_r L_{r+1}.$$

The extra summand on the right-hand side appears because  $\phi_r$ , viewed as a function on  $\{0, 1\}^n$ , vanishes on the Hamming sphere of radius  $r + 1$  around zero, while  $A\phi_r$  is positive on this set. Multiplying both sides of this equality by  $2^n$  and applying the Fourier transform we have, for  $x \in \{0, 1\}^n$ :

$$(n - 2|x|) \cdot f_r(x) = \lambda_r \cdot f_r(x) + (r + 1)a_r K_{r+1}(x),$$

which implies the claim of the lemma. ■

Lemma 3.1 implies the following facts about  $\lambda_r$ . These facts appear to be new (at least, we haven't been able to find them in the literature).

COROLLARY 3.2: *Let  $0 \leq r \leq n/2$ . Then*

- $\frac{n-\lambda_r}{2}$  is a root of  $K_{r+1}$ , with  $K_{r+1}$  viewed as the appropriate univariate real polynomial of degree  $r + 1$ .
- $\lambda_r \in 2\sqrt{r(n-r)} \pm o(n)$ .

*Proof.* In the notation of the proof of Lemma 3.1, we have  $f_r = \sum_{i=0}^r a_i K_i$ . Viewed as a univariate polynomial, this is a polynomial of degree  $r$ , and we have the identity  $(n - \lambda_r - 2k)f_r(k) = K_{r+1}(k)$ , for all integers  $k$  between 0 and  $n$ . This means that  $(n - \lambda_r - 2x) \cdot f_r(x) = K_{r+1}(x)$  for all real  $x$ , implying that  $\frac{n-\lambda_r}{2}$  is a root of  $K_{r+1}$ .

In particular,  $\lambda_r \leq n - 2x_{r+1}$ , where  $x_{r+1}$  is the minimal root of  $K_{r+1}$ . Using the known estimates on  $x_{r+1}$  (see, e.g., [18]) gives  $\lambda_r \leq 2\sqrt{r(n-r)} - o(n)$ . The second claim of the corollary follows from this and from the estimate

$$\lambda_r \geq 2\sqrt{r(n-r)} + o(n)$$

see [9]. ■

3.1. PROOF OF PROPOSITION 1.3. Fix  $0 < \delta < \frac{1}{2}$  and assume that Conjecture 1.2 holds for this value of  $\delta$ . For an integer  $n$ , let  $d = \lfloor \delta n \rfloor$ . Let  $C$  be a linear code of length  $n$  and distance  $d$ . Let  $r = r(n)$  be the minimal radius of a Hamming ball centered at 0 for which  $\lambda_r \geq n - 2(1 + \epsilon)d$ , where  $\epsilon \leq \epsilon_0$  and  $\epsilon_0 = \epsilon_0(\delta)$  is the constant specified by Conjecture 1.2. We proceed as in the proof of Theorem 1.1, using the same notation, but replacing the value  $r = r_0$  in that proof with the new value of  $r$  we have chosen. Computing  $\langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}}$

in two ways, we get, on the one hand,

$$\begin{aligned} \langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} &\geq \lambda_r \cdot \langle \phi * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} = \lambda_r \cdot \langle f^2, 1_C * 1_C \rangle \\ &= \frac{\lambda_r}{2^{2n}} \cdot \sum_{y,z \in C} f^2(y+z). \end{aligned}$$

On the other hand, assuming without loss of generality that  $(1 + \epsilon)d$  is an integer (which we can with a negligible loss), we have

$$\begin{aligned} \langle (A\phi) * \phi, \widehat{1_C}^2 \rangle_{\mathcal{F}} &= \langle (n - 2|x|) \cdot f(x)^2, 1_C * 1_C \rangle \\ &= \frac{1}{2^{2n}} \sum_{y,z \in C} (n - 2|y+z|) f^2(y+z) \\ &\leq \frac{n}{2^{2n}} \cdot N f^2(0) + \frac{n - 2d}{2^{2n}} \cdot \sum_{y,z \in C, y \neq z, |y-z| \leq (1+\epsilon)d} f^2(y+z) \\ &\quad + \frac{\lambda_r - 2}{2^{2n}} \cdot \sum_{y,z \in C, |y-z| > (1+\epsilon)d} f^2(y+z). \end{aligned}$$

Combining both estimates, rearranging, and writing for simplicity a larger expression than needed on the left-hand side of the following inequality, we get

$$n \cdot N f^2(0) + n \cdot \sum_{y \neq z \in C, |y-z| \leq (1+\epsilon)d} f^2(y+z) \geq \sum_{y,z \in C} f^2(y+z).$$

This means that one of the summands on the left-hand side is at least as large as  $\frac{1}{2} \cdot \sum_{y,z \in C} f^2(y+z)$ . We consider both of these possibilities.

(1)  $n \cdot N f^2(0) \geq \frac{1}{2} \cdot \sum_{y,z \in C} f^2(y+z)$ .

In this case, we can proceed as in the proof of Theorem 1.1 and deduce that the rank of the matrix  $\mathcal{M}$  is at least  $\frac{1}{2n} \cdot |C|$ . On the other hand, this rank is at most  $\sum_{i=0}^r |B_i| \leq 2^{H(\frac{r}{n}) \cdot n}$  (where we have used (4)). Hence

$$|C| \leq 2n \cdot 2^{H(\frac{r}{n}) \cdot n}.$$

Recalling that  $r = r(n)$  is the smallest integer with  $\lambda_r \geq n - 2(1 + \epsilon)d$ , and using the second claim of Corollary 3.2, we get that

$$r \leq \left(\frac{1}{2} - \sqrt{\delta(1 - \delta)}\right) \cdot n - an,$$

for some absolute constant  $a = a(\delta)$ . Hence we get an upper bound on  $|C|$  which is exponentially smaller than the first linear programming bound, completing the proof of the proposition in this case.

$$(2) \quad n \cdot \sum_{y \neq z \in C, |y-z| \leq (1+\epsilon)d} f^2(y+z) \geq \frac{1}{2} \cdot \sum_{y, z \in C} f^2(y+z).$$

This means that for some  $d \leq i \leq (1 + \epsilon)d$  we have

$$n^2 \cdot \sum_{y, z \in C, |y-z|=i} f^2(y+z) \geq \sum_{y, z \in C} f^2(y+z).$$

Let  $A$  be the  $|C| \times |C|$  matrix indexed by the elements of  $C$  with

$$A(y, z) = \begin{cases} f(y+z), & |y-z|=i, \\ 0, & \text{otherwise.} \end{cases}$$

Then the preceding inequality can be written as  $n^2 \cdot Tr(AM) \geq Tr(\mathcal{M}^2)$ .

Let  $F = \sum_{x \in |C|, |x|=i} f(x)W_x$ . Since  $C$  is a linear code, it is well-known (and easy to see) that the eigenvectors of  $A$  are the restrictions to  $C$  of the Walsh–Fourier characters  $\{W_u\}_{u \in \{0,1\}^n}$ , and the eigenvalue corresponding to  $W_u$  is  $F(u)$ . Since the restrictions of  $W_u$  and  $W_{u'}$  coincide iff  $u$  and  $u'$  are in the same coset of  $C^\perp$ , the function  $F$  is constant on the cosets of  $C^\perp$  in  $\{0, 1\}^n$ , and the distribution of  $F$  in  $\{0, 1\}^n$  is the same as the distribution of the eigenvalues of  $A$  in  $C$ , provided both  $\{0, 1\}^n$  and  $C$  are endowed with uniform probability measure. Let  $\alpha = (\alpha_1, \dots, \alpha_N)$  be the vector of eigenvalues of  $A$ , viewed as a function on  $C$ . By the preceding discussion, we have that  $\frac{\|\alpha\|_4}{\|\alpha\|_2} = \frac{\|F\|_4}{\|F\|_2}$ .

On the other hand,  $F$  is a symmetric function whose Fourier transform is supported on the set of vectors of weight  $i$  in  $C$ . The conditions of Conjecture 1.2 are satisfied, and since we have assumed the conjecture to hold we have

$$\frac{\|\alpha\|_4}{\|\alpha\|_2} = \frac{\|F\|_4}{\|F\|_2} \leq 2^{c\epsilon n + o(n)},$$

where  $c = c(\delta)$  is the constant specified in the conjecture.

Let  $\lambda = (\lambda_1, \dots, \lambda_N)$  be the vector of eigenvalues of  $\mathcal{M}$ . Assume that both  $\alpha$  and  $\lambda$  are arranged in decreasing order of values. Note that  $Tr(\mathcal{M}^2) = \|\lambda\|_2^2$ . We also have  $Tr(AM) \leq \langle \alpha, \lambda \rangle$  by [27].<sup>2</sup> Hence  $n^2 \cdot Tr(AM) \geq Tr(\mathcal{M}^2)$  implies  $n^2 \cdot \langle \alpha, \lambda \rangle \geq \|\lambda\|_2^2$ .

Taking everything into account, we have

$$\|\lambda\|_2^2 \leq n^2 \cdot \langle \alpha, \lambda \rangle \leq n^2 \cdot \|\alpha\|_4 \|\lambda\|_{\frac{4}{3}} \leq 2^{c\epsilon n + o(n)} \cdot \|\alpha\|_2 \|\lambda\|_{\frac{4}{3}} \leq 2^{c\epsilon n + o(n)} \cdot \|\lambda\|_2 \|\lambda\|_{\frac{4}{3}},$$

<sup>2</sup> Since  $C$  is a linear code, the matrices  $A$  and  $\mathcal{M}$  commute, so the result of [27] is not required. With that we state the argument in higher generality, to apply to general codes as well.

where in the second step we use Hölder’s inequality, and in the last step the elementary fact  $\|\alpha\|_2 \leq \|\lambda\|_2$ . So we get  $\|\lambda\|_2 \leq 2^{c\epsilon n + o(n)} \cdot \|\lambda\|_{\frac{4}{3}}$ . Let  $S \subseteq C$  be the support of  $\lambda$ . Note that  $|S| = \text{rank}(\mathcal{M})$ . Applying Hölder’s inequality once again (in the second step below), we have

$$\|\lambda\|_{\frac{4}{3}}^{\frac{4}{3}} = \langle \lambda^{\frac{4}{3}}, 1_S \rangle \leq \|\lambda^{\frac{4}{3}}\|_{\frac{3}{2}} \|1_S\|_3 = \|\lambda\|_2^{\frac{4}{3}} \left( \frac{|S|}{|C|} \right)^{\frac{1}{3}},$$

implying that  $\text{rank}(\mathcal{M}) = |S| \geq |C| \cdot \left( \frac{\|\lambda\|_{\frac{4}{3}}}{\|\lambda\|_2} \right)^4 \geq |C| \cdot 2^{-4c\epsilon n + o(n)}$ . On the other hand, we have  $\text{rank}(\mathcal{M}) = \text{rank}(M) \leq \sum_{i=0}^r \binom{n}{i} \leq 2^{H(\frac{r}{n}) \cdot n}$ . So, we get

$$\frac{1}{n} \log_2 |C| \leq H\left(\frac{r}{n}\right) + 4c\epsilon + o(1).$$

It remains to analyze the expression on the right-hand side of this inequality. Let  $\rho = \frac{r}{n}$ , let  $\rho_0 = \frac{1}{2} - \sqrt{\delta(1 - \delta)}$ , and let  $g(x) = 2\sqrt{x(1 - x)}$ . Note that  $g(\rho_0) = 1 - 2\delta$ . Ignoring negligible factors, which we do from now on in this calculation, we also have  $g(\rho) = 1 - 2\delta - 2\delta\epsilon$ . Assuming  $\epsilon$  is sufficiently small, and using first order approximations, we have  $\rho \approx \rho_0 - \frac{2\delta\epsilon}{g'(\rho_0)}$ , and hence  $H(\rho) \approx H(\rho_0) - \frac{H'(\rho_0)}{g'(\rho_0)} \cdot 2\delta\epsilon$ . It is easy to verify that  $c = c(\delta) = \frac{1}{8} \cdot \frac{H'(\rho_0)}{g'(\rho_0)} \cdot 2\delta$ , and hence

$$H\left(\frac{r}{n}\right) + 4c\epsilon \approx H(\rho_0) - 4c\epsilon = H\left(\frac{1}{2} - \sqrt{\delta(1 - \delta)}\right) - 4c\epsilon,$$

completing the proof of the proposition in this case. ■

3.2. PROOF OF PROPOSITION 1.8. Let  $0 < \delta < \frac{1}{2}$  and assume that Conjecture 1.7 holds for this value of  $\delta$ . We will assume that there is a sequence of codes  $C_n$  of length  $n$  and distance  $d = \lfloor \delta n \rfloor$  attaining the first linear programming bound and reach a contradiction. Assume then that  $n$  is large, and that  $C$  is a code of length  $n$ , distance  $d = \lfloor \delta n \rfloor$ , such that

$$\frac{1}{n} \log_2 |C| \geq H\left(\frac{1}{2} - \sqrt{\delta(1 - \delta)}\right) - o(1).$$

We follow the same argument as in the proof of Proposition 1.3. It is readily seen that everything works through if we show that if  $d \leq i \leq (1 + \epsilon)d$  is such that  $n^2 \cdot \sum_{y,z \in C, |y-z|=i} f^2(y + z) \geq \sum_{y,z \in C} f^2(y + z)$ , then the vector  $\alpha$  of eigenvalues of the distance- $i$  graph  $G(C, i)$  satisfies  $\frac{\|\alpha\|_4}{\|\alpha\|_2} \leq 2^{c\epsilon n + o(n)}$ . This will follow from Conjecture 1.7 if we show that  $G(C, i)$  is  $2^{o(n)}$ -balanced. This is what we proceed to show. We start with a technical lemma.

LEMMA 3.3: Let  $K_s$  be a Krawtchouk polynomial, for some  $1 \leq s \leq \frac{n}{2}$ . Let  $a$  be a root of  $K_s$ , and let  $g(x) = \frac{K_s(x)}{x-a}$ . Then:

- (1)  $\|g\|_2^2 = \frac{1}{2^n} \sum_{i=0}^n \binom{n}{i} g^2(i) \geq \frac{1}{n^2} \cdot \binom{n}{s}$ .
- (2) For all  $0 \leq k \leq n$  we have  $g^2(k) \leq O(n^3) \cdot \frac{2^n \binom{n}{s}}{\binom{n}{k}}$ .

*Proof.* Recall (see [17], Sections 1 and 5, for this and for additional properties of Krawtchouk polynomials used in this proof) that  $\|K_s\|_2^2 = \binom{n}{s}$ , and that all the roots of  $K_s$  lie in the interval  $(0, n)$ . Hence  $|k - a| < n$  for all  $0 \leq k \leq n$ , and the first claim of the lemma follows.

We pass to the second claim of the lemma. It is known that  $K_s$  has  $s$  simple roots. Let them be  $x_1 < x_2 < \dots < x_s$ , and let  $a = x_m$ , for some  $1 \leq m \leq s$ . There are two cases to consider. Either  $k$  lies inside the root region of  $K_s$ , that is  $x_1 \leq k \leq x_s$ , or not. We consider the first case. (The second case is similar and simpler.) Since  $\|K_s\|_2^2 = \binom{n}{s}$ , for all  $0 \leq i \leq n$  we have  $K_s^2(i) \leq \frac{2^n \binom{n}{s}}{\binom{n}{i}}$ . Hence if  $|k - a| > \frac{1}{2}$  the claim follows immediately. If  $|k - a| < \frac{1}{2}$  there are again two cases to consider,  $k > a$  and  $k < a$ . We consider the first case, the second is similar. Recall that the distance between any two consecutive roots of  $K_s$  is at least 2. Since  $a = x_m \leq k < x_m + \frac{1}{2}$ , this means that the point  $k + 1$  lies between  $x_m$  and  $x_{m+1}$  and it is at distance at least  $\frac{1}{2}$  from  $x_{m+1}$ . This implies that

$$\left| \frac{g(k)}{g(k+1)} \right| = \prod_{\ell \neq m} \frac{|k - x_\ell|}{|k+1 - x_\ell|} \leq \frac{x_s - x_m}{x_{m+1} - k - 1} \leq 2n.$$

It follows that

$$g^2(k) \leq 4n^2 \cdot g^2(k+1) \leq 4n^2 \cdot \frac{2^n \binom{n}{s}}{\binom{n}{k+1}} \leq 4n^3 \cdot \frac{2^n \binom{n}{s}}{\binom{n}{k}}. \quad \blacksquare$$

We proceed with the argument. Let  $(A_0, \dots, A_n)$  be the distance distribution vector of  $C$ , with  $A_k = \frac{1}{|C|} |\{(x, y) \in C \times C, |x - y| = k\}|$ . Note that

$$\sum_{y, z \in C, |y-z|=i} f^2(y+z) = |C| A_i f^2(i) \leq O(n^3) \cdot |C| A_i \cdot \frac{2^n \binom{n}{r+1}}{\binom{n}{i}},$$

where in the second step we have used Lemma 3.1, the first claim of Corollary 3.2, and the second claim of Lemma 3.3. On the other hand, we have

$$\begin{aligned} \sum_{y,z \in C} f^2(y+z) &= 2^n \cdot \langle 1_C * 1_C, f^2 \rangle = 2^{2n} \cdot \langle \widehat{1_C}^2, \widehat{f} * \widehat{f} \rangle_{\mathcal{F}} \\ &\geq (2^n \widehat{1_C}(0))^2 \cdot (f * f)(0) = |C|^2 \|f\|_2^2 \geq \frac{1}{n^2} |C|^2 \binom{n}{r+1}, \end{aligned}$$

where we have used Parseval’s identity in the second step, the fact that  $\widehat{f} = \phi$  is nonnegative in the third step, and the first claim of Lemma 3.3 in the fifth step. This means that  $n^2 \cdot \sum_{y,z \in C, |y-z|=i} f^2(y+z) \geq \sum_{y,z \in C} f^2(y+z)$  implies

$$A_i \geq \Omega\left(\frac{1}{n^5}\right) \cdot \frac{|C| \binom{n}{i}}{2^n} \geq 2^{-o(n)} \cdot 2^{H(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n} \cdot \frac{\binom{n}{i}}{2^n},$$

where we have used the assumption that  $C$  attains the first linear programming bound in the second step. Note that this means that the edge density of the graph  $G(C, i)$  is at least  $2^{-o(n)} \cdot 2^{H(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n} \cdot \frac{\binom{n}{i}}{2^n}$ .

(Observe that the preceding discussion provides an additional proof of the fact ([1]) that a linear code attaining the first linear programming bound must have (up to a negligible error) as many codewords of some weight close to its minimal distance as a random code of the same cardinality, see Section 1.1. It is in fact possible that the above inequality for  $A_i$  might have been derived directly from Corollary1 in [1], but we have not found a ready way to do so.)

Corollary1 in [1] also presents a complementary result: Let  $C'$  be a code with distance  $d = \lfloor \delta n \rfloor$ , and with distance distribution  $(A'_0, \dots, A'_n)$ . Then for any  $d \leq k \leq n/2$  we have

$$A'_k \leq 2^{o(n)} \cdot 2^{H(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n} \cdot \frac{\binom{n}{k}}{2^n}.$$

Let now  $C'$  be a subset of  $C$ , and consider the subgraph  $G(C', i)$  of  $G(C, i)$  induced by  $C'$ . The edge density of this subgraph is  $A'_i$ . Since any subset of a code with distance  $d$  is by itself a code with distance (at least)  $d$ , the edge density of  $G(C', i)$  is upper bounded by  $2^{o(n)} \cdot 2^{H(\frac{1}{2} - \sqrt{\delta(1-\delta)}) \cdot n} \cdot \frac{\binom{n}{i}}{2^n}$ . Combined with the above lower bound on the edge density of  $G(C, i)$ , this implies that  $G(C, i)$  is  $2^{o(n)}$ -balanced, and conditions of Conjecture 1.7 are satisfied, completing the proof of the proposition. ■



**4. Proof of Propositions 1.4 and 1.9**

We show that Conjectures 1.2 and 1.7 hold for random codes. (See Section 1.3.3 for the models of random codes we use.) In fact they hold in a strong sense, and we are allowed to replace the exponential expressions on the right-hand side of (2) and (3) by absolute constants. We do not attempt to compute the best possible values of these constants.

4.1. PROOF OF PROPOSITION 1.4. We start with a technical lemma.

LEMMA 4.1: *Let  $0 < R < 1$  be given. There exist positive constants  $\epsilon = \epsilon(R)$  and  $\alpha = \alpha(R)$  such that for any integer parameters  $d, i, t$  satisfying*

- $(1 - \epsilon)H^{-1}(1 - R) \cdot n \leq d \leq (1 + \epsilon)H^{-1}(1 - R) \cdot n,$
- $d \leq i \leq (1 + \epsilon)d,$
- $d \leq t \leq 2i,$

*the following holds: Let  $x \in \{0, 1\}^n$  be of weight  $t$ . Let  $D(x)$  be the set of all points  $y$  in  $\{0, 1\}^n$  for which  $|y| = |x - y| = i$ . Then*

$$|D(x)| \leq 2^{(1-R-\alpha)\cdot n}.$$

*Proof.* It is easy to see that  $|D(x)| = \binom{t}{\frac{t}{2}} \binom{n-t}{i-\frac{t}{2}}$ , with the understanding that the binomial coefficient  $\binom{b}{a}$  is 0 unless  $a, b$  are integers and  $0 \leq a \leq b$ .

Writing  $\delta = \frac{d}{n}$ ,  $\xi = \frac{i}{n}$ , and  $\tau = \frac{t}{n}$ , and using (4), it suffices to show that  $\tau + (1 - \tau)H\left(\frac{2\xi - \tau}{2 - 2\tau}\right) < 1 - R$  on the compact domain  $\Omega = \{(\delta, \xi, \tau) \subseteq \mathbb{R}^3\}$  given by

- $(1 - \epsilon)H^{-1}(1 - R) \leq \delta \leq (1 + \epsilon)H^{-1}(1 - R),$
- $\delta \leq \xi \leq (1 + \epsilon)\delta,$
- $\delta \leq \tau \leq 2\xi.$

It is easy to see that for any  $0 < R < 1$ , if  $\epsilon = \epsilon(R)$  is sufficiently small then all the partial derivatives of  $f(\xi, \tau) = \tau + (1 - \tau)H\left(\frac{2\xi - \tau}{2 - 2\tau}\right)$  are uniformly bounded from above on  $\Omega$ , and hence it suffices to prove that  $f(\delta, \tau) < 1 - R$  for  $\delta \leq \tau \leq 2\delta$ , where  $0 < R < 1$  and  $\delta = H^{-1}(1 - R)$ . Alternatively, it suffices to prove that the function  $g(\delta, \tau) = \tau + (1 - \tau)H\left(\frac{2\delta - \tau}{2 - 2\tau}\right) - H(\delta)$  is non-positive on  $\{(\delta, \tau) : 0 \leq \delta \leq \frac{1}{2}; \delta \leq \tau \leq 2\delta\}$ , and that  $g(\delta, \tau) = 0$  only if  $\delta = 0$  or  $\delta = \frac{1}{2}$ .

We will do this in two steps. First, we claim that  $g(\delta, \tau)$  does not increase in  $\tau$ , for any value of  $\delta$ . We have, after simplifying, that

$$\frac{\partial g}{\partial \tau} = 1 - H\left(\frac{2\delta - \tau}{2 - 2\tau}\right) - \frac{1 - 2\delta}{2 - 2\tau} \log_2 \left(\frac{2 - 2\delta - \tau}{2\delta - \tau}\right).$$

Let  $x = \frac{2\delta - \tau}{2 - 2\tau}$ . Then the above expression is  $1 - H(x) - (\frac{1}{2} - x) \log_2(\frac{1-x}{x})$ . It is easy to see that

$$H(x) + \left(\frac{1}{2} - x\right) \log_2\left(\frac{1-x}{x}\right) = \frac{1}{2} \log_2\left(\frac{1}{x(1-x)}\right) \geq 1,$$

and hence  $\frac{\partial g}{\partial \tau} \leq 0$ .

So it suffices to show that  $h(\delta) = g(\delta, \delta)$  is non-positive on  $0 \leq \delta \leq \frac{1}{2}$ , and that  $h(\delta) = 0$  only if  $\delta = 0$  or  $\delta = \frac{1}{2}$ . We have

$$\begin{aligned} h(\delta) &= \delta + (1 - \delta)H\left(\frac{\delta}{2 - 2\delta}\right) - H(\delta) \\ &= 1 + \frac{\delta}{2} \log_2(\delta) + (2 - 2\delta) \log_2(1 - \delta) - \frac{2 - 3\delta}{2} \log_2(2 - 3\delta). \end{aligned}$$

It is easy to see that  $h(0) = h(\frac{1}{2}) = 0$ . We will show that there exists  $0 < \delta_0 < \frac{1}{2}$  such that  $h' < 0$  for  $0 \leq \delta < \delta_0$  and  $h' > 0$  for  $\delta_0 < \delta \leq \frac{1}{2}$ . This will imply that  $h(\delta) < 0$  for any  $0 < \delta < \frac{1}{2}$ .

A simple calculation gives

$$h'(\delta) = \frac{1}{2} \log_2\left(\frac{\delta(2 - 3\delta)^3}{(1 - \delta)^4}\right).$$

Let  $P(\delta) = \delta(2 - 3\delta)^3 - (1 - \delta)^4$ . We need to show that there exists  $0 < \delta_0 < \frac{1}{2}$  such that  $P(\delta) < 0$  for  $0 \leq \delta < \delta_0$  and  $P(\delta) > 0$  for  $\delta_0 < \delta \leq \frac{1}{2}$ . It is easy to verify that  $P(\delta) = (1 - 2\delta)(14\delta^3 - 22\delta^2 + 10\delta - 1)$ . So it suffices to show this property for  $Q(\delta) = 14\delta^3 - 22\delta^2 + 10\delta - 1$ . The derivative  $Q'$  is a quadratic, and it is easy to see that it is strictly decreasing on  $[0, \frac{1}{2}]$ , and moreover that  $Q'(0) > 0$  and that  $Q'(\frac{1}{2}) < 0$ . This means that  $Q$  is unimodal—it increases up to some point in  $[0, \frac{1}{2}]$  and then decreases. Moreover,  $Q(0) = -1 < 0$  and  $Q(\frac{1}{2}) = \frac{1}{4} > 0$ . This verifies the required property for  $Q$ , and completes the proof of the lemma. ■

We proceed with the proof of the proposition. Let  $f$  be a function on  $\{0, 1\}^n$ , and let  $A$  be the Fourier support of  $f$ . By Proposition 1.1 in [15] we have

$$\left(\frac{\|f\|_4}{\|f\|_2}\right)^4 \leq \max_{x \in A+A} |\{(y, z) \in A \times A, y + z = x\}|.$$

So it suffices to show that there exists a positive constant  $K = K(R)$  such that the following holds with probability tending to 1 with  $n$  for a random linear code  $C$  of dimension  $\lfloor Rn \rfloor$ . Let  $d$  be the minimal distance of  $C$ , let  $d \leq i \leq (1 + \epsilon)d$ , where  $\epsilon = \epsilon(R)$  is the constant from Lemma 4.1, and let  $A$  be

the set of vectors of weight  $i$  in  $C$ . Assume  $A \neq \emptyset$ . Let  $f$  be a function whose Fourier transform is supported on  $A$ . Then

$$\max_{x \in A+A} |\{(y, z) \in A \times A, y + z = x\}| \leq K.$$

Recall that w.h.p. the minimal distance  $d$  of  $C$  satisfies

$$d_0 - o(n) \leq d \leq d_0 + o(n),$$

where  $d_0 = d_0(n) = H^{-1}(1 - R) \cdot n$ . Assume from now on that this is indeed the case. This means that the points  $x$  we need to consider are such that  $x = y + z$  for some points  $y, z \in \{0, 1\}^n$  with  $|y| = |z| = i$ , for some  $d_0 - o(n) \leq i \leq (1 + \epsilon)d_0$ . Using the union bound, it suffices to show for any suitable value of  $i$  and for any such point  $x$  the probability over  $C$  that

$$|\{(y, z) \in C \times C, |y| = |z| = i, y + z = x\}| > K,$$

for a sufficiently large constant  $K$  is  $o(\frac{1}{n2^n})$ .

Fix  $i$  and  $x$ . Let  $|x| = t$ . As in Lemma 4.1, let  $D(x)$  be the set of all points  $y$  in  $\{0, 1\}^n$  for which  $|y| = |x - y| = i$ . Note that

$$|\{(y, z) \in C \times C, |y| = |z| = i, y + z = x\}| \leq |D(x) \cap C|.$$

So it suffices to upper-bound  $|D(x) \cap C|$ . We proceed to do this. For a subset  $D \subseteq \{0, 1\}^n$  and for an integer parameter  $m$ , if  $|D \cap C| \geq 2^m$ , then  $C$  contains at least  $m$  linearly independent elements of  $D$ . It is well-known (and easy to see) that the probability of a random linear code  $C$  of a given cardinality to contain  $m$  given linearly independent vectors is at most  $(\frac{|C|}{2^n})^m$ , and hence, by the union bound,

$$\Pr_C\{|D \cap C| \geq 2^m\} \leq \binom{|D|}{m} \left(\frac{|C|}{2^n}\right)^m < \left(\frac{|C||D|}{2^n}\right)^m.$$

Let now  $D = D(x)$ . The parameters  $d, i, t$  satisfy the conditions of Lemma 3.3, and hence, by the lemma,  $\frac{|C||D|}{2^n} \leq 2^{-\alpha n}$ , for some positive constant  $\alpha = \alpha(R)$ . Hence, for  $m = \lceil 2/\alpha \rceil$ , we get

$$\Pr_C\{|D \cap C| \geq 2^m\} \leq 2^{-\alpha mn} \leq 2^{-2n} \leq o\left(\frac{1}{n2^n}\right),$$

as needed. To conclude, the proposition holds with  $K = 2^{\lceil 2/\alpha \rceil}$  and  $\epsilon_0 = \epsilon$ , where  $\alpha$  and  $\epsilon$  are given by Lemma 4.1. ■

4.2. PROOF OF PROPOSITION 1.9. Let  $0 < R < 1$ . In the following two lemmas we list some simple properties of random codes of prescribed cardinality  $N = \lfloor 2^{Rn} \rfloor$  chosen according to the model described in Section 1.3.3. (We believe all of these properties to be well-known, but we haven't been able to find a proper reference in the literature.)

Before stating the lemmas, let us make some remarks about the parameters  $\tau = \tau(R)$  and  $d_0 = d_0(R, n)$  in the model. The constant  $\tau$  is chosen to be sufficiently small so that all the constants appearing in the following statements which depend linearly on  $\tau$  will be smaller than  $\frac{1}{2}$  and so that  $\theta = \frac{\binom{n}{d_0} N}{2^n}$  is at most  $\frac{1}{2}$  as well. It is easy to see that it is possible to choose  $\tau$  appropriately. We also note that  $\theta$  is an absolute constant depending on  $R$  and that

$$H^{-1}(1 - R) \cdot n - o(n) \leq d_0 \leq H^{-1}(1 - R) \cdot n + o(n).$$

We will show that the following properties hold with probability tending to 1 with  $n$  for a random code  $C$  of length  $n$  and prescribed cardinality  $N = \lfloor 2^{Rn} \rfloor$ . We denote by  $d$  the minimal distance of  $C$ .

- LEMMA 4.2:     (1)  $|C| \geq (1 - O(\tau) - o(1)) \cdot N$ .  
 (2) For any  $d \leq k \leq \frac{n}{2}$  the number of pairs of points in  $C$  at distance  $k$  from each other lies between  $(1 - O(\tau) - o(1)) \cdot \frac{\binom{N}{2} \binom{n}{k}}{2^n}$  and  $(1 + o(1)) \cdot \frac{\binom{N}{2} \binom{n}{k}}{2^n}$ .  
 (3)  $d = d_0$ .

LEMMA 4.3: Let  $d \leq k \leq \frac{n}{2}$ . For  $x \in C$ , let

$$D_{x,k} := |\{y \in C, |x - y| = k\}|.$$

Then

$$\frac{\mathbb{E}_x D_{x,k}^2}{(\mathbb{E}_x D_{x,k})^2} \leq O\left(\frac{1}{\theta}\right),$$

where the expectations are taken with respect to the uniform probability distribution on  $C$ , and  $\theta = \frac{\binom{n}{d_0} N}{2^n}$ .

We will first prove the proposition assuming Lemmas 4.2 and 4.3 to hold, and then prove the lemmas. From now on let  $\epsilon = \epsilon(R)$  be the constant in Lemma 4.1.

We will need another technical lemma.

LEMMA 4.4: Let  $d_0 \leq i \leq (1 + \epsilon)d_0$ . For  $x, y \in \{0, 1\}^n$ , let

$$M_{x,y} = |\{z \in C, |x - z| = |y - z| = i\}|.$$

Then with probability at least  $1 - o(\frac{1}{n})$  we have

$$\max_{x,y \in C} M_{x,y} \leq K_1,$$

for some absolute constant  $K_1 = K_1(R)$ .

*Proof.* Let  $\tilde{C} = \{x_1, \dots, x_N\}$  be a list of  $N$  points chosen independently at random from  $\{0, 1\}^n$ . Clearly, it suffices to prove that there exists a constant  $K_1$  such that with high probability for any  $\{i, j\} \subseteq [N]$  we have  $|M_{x_i, x_j}| \leq K_1$ . By symmetry, and by the union bound, it suffices to show that

$$\Pr_{\tilde{C}}\{|M_{x_1, x_2}| > K_1\} < \frac{1}{nN^2},$$

where we may assume that  $|x_1 - x_2| \geq d_0$ . Fix  $x_1$  and  $x_2$ . Let  $|x_1 - x_2| = t$  for some  $d_0 \leq t \leq 2i$  and let  $D = D(x, y)$  be the set of all points in  $\{0, 1\}^n$  at distance  $i$  from both  $x$  and  $y$ . Then  $|D| = \binom{t}{\frac{t}{2}} \binom{n-t}{i-\frac{t}{2}}$ . Let  $\tilde{C}_1 = \{x_3, \dots, x_N\}$ . For any integer  $m \geq 1$  we have that

$$\Pr_{\tilde{C}}\{|M_{x,y}| \geq m\} = \Pr_{\tilde{C}_1}\{|\tilde{C}_1 \cap D| \geq m\} \leq \binom{|D|}{m} \cdot \left(\frac{|\tilde{C}_1|}{2^n}\right)^m < \left(\frac{ND}{2^n}\right)^m.$$

By Lemma 4.1,  $\frac{ND}{2^n} \leq 2^{-\alpha n}$ , for a positive absolute constant  $\alpha$ , and hence for  $K_1 = \frac{3}{\alpha}$ , we have  $\Pr\{|M_{x,y}| > K_1\} < 2^{-3n} < \frac{1}{nN^2}$ , and the claim of the lemma holds with this value of  $K_1$ . ■

We proceed to prove the proposition. Let  $d_0 \leq i \leq (1 + \epsilon)d_0$ , and let  $B$  be the adjacency matrix of the graph  $G(C, i)$ . Let  $\lambda$  be the vector of eigenvalues of  $B$ . Recall the notation  $D_{x,i} := |\{y \in C, |x - y| = i\}|$  and  $M_{x,y} = |\{z \in C, |x - z| = |y - z| = i\}|$ . We have

$$\begin{aligned} \|\lambda\|_2^2 &= \frac{1}{|C|} \sum_{i=1}^{|C|} \lambda_i^2 = \frac{1}{|C|} \sum_{x,y \in C} B^2(x, y) \\ &= \frac{1}{|C|} |\{(x, y) \in C \times C, |x - y| = i\}| \\ &= \mathbb{E}_{x \in C} D_{x,i}. \end{aligned}$$

For the following calculation, let  $g(z) = 1_{|z|=i}$ . Then for any  $x, y \in C$  we have  $B(x, y) = g(x + y)$ . Hence, using the Cauchy–Schwarz inequality in the fourth step below,

$$\begin{aligned} \|\lambda\|_4^4 &= \frac{1}{|C|} \sum_{i=1}^{|C|} \lambda_i^4 = \frac{1}{|C|} \sum_{x,y \in C} \left( \sum_{z \in C} B(x, z)B(z, y) \right)^2 \\ &= \frac{1}{|C|} \sum_{x,y \in C} \left( \sum_{z \in C} g(x + z)g(y + z) \right)^2 \\ &\leq \frac{1}{|C|} \sum_{x,y \in C} M_{x,y} \sum_{z \in C} g^2(x + z)g^2(y + z) \\ &\leq \frac{\max_{x,y \in C} M_{x,y}}{|C|} \cdot \sum_{x,y \in C} \sum_{z \in C} g^2(x + z)g^2(y + z) \\ &= \frac{\max_{x,y \in C} M_{x,y}}{|C|} \cdot \sum_{z \in C} \left( \sum_{x \in C} g^2(x + z) \right)^2 = \left( \max_{x,y \in C} M_{x,y} \right) \cdot \mathbb{E}_{z \in C} D_{z,i}^2. \end{aligned}$$

It follows that

$$\left( \frac{\|\lambda\|_4}{\|\lambda\|_2} \right)^4 \leq \left( \max_{x,y \in C} M_{x,y} \right) \cdot \frac{\mathbb{E}_z D_x^2}{(\mathbb{E}_z D_z)^2}.$$

Lemmas 4.3 and 4.4 imply that with probability tending to 1 with  $n$  the right-hand side of the last inequality is bounded from above by  $O(\frac{1}{\theta} \cdot K_1)$  for all  $d_0 \leq i \leq (1 + \epsilon)d_0$ , and the proposition holds with  $K = O(\frac{1}{\theta} \cdot K_1)$  and  $\epsilon_0 = \epsilon$ . ■

We proceed with the proofs of Lemmas 4.2 and 4.3.

4.2.1. *Proof of Lemmas 4.2 and 4.3.* We need the following technical claim.

LEMMA 4.5: Let  $\tilde{C} = \{x_1, \dots, x_N\}$  be a list of  $N$  points chosen independently at random from  $\{0, 1\}^n$ . Let  $M = \binom{N}{2}$ , and for  $0 \leq \ell \leq n$  let  $p_\ell = \frac{\binom{n}{\ell}}{2^n}$ .

- (1) For  $0 \leq \ell \leq n$  let  $X_\ell$  be the random variable counting the number of pairs of indices  $1 \leq i < j \leq N$  with  $|x_i - x_j| = \ell$ . Then  $\mathbb{E} X_\ell = Mp_\ell$ , and  $\sigma^2(X_\ell) = Mp_\ell(1 - p_\ell)$ .
- (2) For  $d_0 \leq \ell \leq \frac{n}{2}$ , let  $Y_\ell$  be the random variable counting the number of triples of distinct indices  $\{a, b, c\} \subseteq [N]$  such that one of the three pairwise distances among the points  $x_a, x_b, x_c$  is  $\ell$  and one is at most  $d_0 - 1$ . Then  $\mathbb{E} Y_\ell \leq O(\tau \cdot Mp_\ell)$ , and  $\sigma^2(Y_\ell) \leq O(\tau Mp_\ell + \tau M^{3/2} p_\ell^2)$ .

*Proof.* For  $\{i, j\} \subseteq N$  let  $Z_{\{i,j\}}$  be the indicator random variable which verifies whether  $|x_i - x_j| = \ell$ . Clearly

$$Pr_{\bar{C}}\{Z_{\{i,j\}} = 1\} = \frac{\binom{n}{\ell}}{2^n} = p\ell.$$

The first claim of the lemma follows from the fact that there are  $M$  such variables, and (as observed e.g., in [4]) they are pairwise independent.

We pass to the second claim of the lemma. For  $\{a, b, c\} \subseteq N$  let  $Z_{\{a,b,c\}}$  be the indicator random variable which verifies whether one of the three pairwise distances among the points  $x_a, x_b, x_c$  is  $\ell$  and one is at most  $d_0 - 1$ . We have, by the pairwise independence of the distances  $|x_a - x_b|$ ,  $|x_a - x_c|$ , and  $|x_b - x_c|$ , that

$$\begin{aligned} \mathbb{E}_{\bar{C}} Y_\ell &= \sum_{\{a,b,c\} \subseteq N} \mathbb{E}_{\bar{C}} Z_{\{a,b,c\}} \leq O\left(N^3 \cdot \frac{(\sum_{k=0}^{d_0-1} \binom{n}{k}) \cdot \binom{n}{\ell}}{2^{2n}}\right) \\ &\leq O\left(Mp\ell \cdot N \frac{\sum_{k=0}^{d_0-1} \binom{n}{k}}{2^n}\right) \leq O(\tau \cdot Mp\ell), \end{aligned}$$

where the last step follows from the choice of  $d_0$ .

Next, observe that for  $\{a, b, c\}, \{a', b', c'\} \subseteq N$  the random variables  $Z_{\{a,b,c\}}$  and  $Z_{\{a',b',c'\}}$  are independent unless  $|\{a, b, c\} \cap \{a', b', c'\}| \geq 2$ . Hence

$$\begin{aligned} \sigma^2(Y_\ell) &= \mathbb{E} Y_\ell^2 - (\mathbb{E} Y_\ell)^2 \\ &= \sum_{\{a,b,c\}, \{a',b',c'\}} (\mathbb{E} Z_{\{a,b,c\}} \cdot Z_{\{a',b',c'\}} - \mathbb{E} Z_{\{a,b,c\}} \cdot \mathbb{E} Z_{\{a',b',c'\}}) \\ &= \sum_{|\{a,b,c\} \cap \{a',b',c'\}| \geq 2} (\mathbb{E} Z_{\{a,b,c\}} \cdot Z_{\{a',b',c'\}} - \mathbb{E} Z_{\{a,b,c\}} \cdot \mathbb{E} Z_{\{a',b',c'\}}) \\ &\leq \sum_{|\{a,b,c\} \cap \{a',b',c'\}| \geq 2} \mathbb{E} Z_{\{a,b,c\}} \cdot Z_{\{a',b',c'\}} \\ &= \sum_{\{a,b,c\}} \mathbb{E}_{\bar{C}} Z_{\{a,b,c\}} + \sum_{|\{a,b,c\} \cap \{a',b',c'\}|=2} \mathbb{E} Z_{\{a,b,c\}} \cdot Z_{\{a',b',c'\}}. \end{aligned}$$

The first summand in the last expression is at most  $O(\tau \cdot Mp\ell)$ . We pass to the second summand. There are two possible cases we need to consider, depending on whether  $\sum_{k=0}^{d_0-1} \binom{n}{k}$  is smaller than  $\binom{n}{\ell}$ . Assume first that it is indeed smaller. For  $\{a, b, c, d\} \subseteq N$  let  $W_{\{a,b,c,d\}}$  be the indicator random variable which verifies whether there are two among the four points at distance at most  $d_0 - 1$ , and

each of the two remaining points is at distance  $\ell$  from one of the first two. It is easy to see that the second summand is upper bounded by

$$O\left(\sum_{\{a,b,c,d\} \subseteq N} \mathbb{E} W_{a,b,c,d}\right) \leq O\left(N^4 p_\ell^2 \frac{\sum_{k=0}^{d_0-1} \binom{n}{k}}{2^n}\right) \leq O(\tau N^3 p_\ell^2) = O(\tau M^{3/2} p_\ell^2).$$

If  $\sum_{k=0}^{d_0-1} \binom{n}{k}$  is larger than  $\binom{n}{\ell}$ , a similar computation shows that the second summand is upper bounded by  $O(\tau^2 \cdot M p_\ell)$ , completing the proof of the lemma. ■

We can now prove Lemma 4.2.

*Proof of Lemma 4.2.* The first claim of the lemma follows by the Chebyshev inequality from the first claim of Lemma 4.5 (and the definition of  $d_0$ ). The third claim of the lemma follows from its second claim. We pass to the second claim. Fix  $d \leq k \leq \frac{n}{2}$ . Observe first that by the first claim of Lemma 4.5 and the Chebyshev inequality, the number of pairs of points in  $\tilde{C}$  at distance  $k$  from each other lies between  $(1 - o(1)) \cdot M p_k$  and  $(1 + o(1)) \cdot M p_k$  with probability at least  $1 - o(\frac{1}{n})$ . Next, note that the number of pairs of points in  $\tilde{C}$  at distance  $k$  from each other removed in the erasure step is at most  $O(Y_k)$ . By the second claim of Lemma 4.5 and the Chebyshev inequality,  $Y_k \leq O(\tau M p_k)$  with probability at least  $1 - o(\frac{1}{n})$ , and the second claim of the lemma follows, by the union bound over all possible values of  $k$ . ■

We pass to the proof of Lemma 4.3.

*Proof of Lemma 4.3.* Fix  $d \leq k \leq \frac{n}{2}$ . We will show that

$$\Pr_C \left\{ \frac{\mathbb{E}_x D_{x,k}^2}{(\mathbb{E}_x D_{x,k})^2} \leq O\left(\frac{1}{\theta}\right) \right\} \geq 1 - o\left(\frac{1}{n}\right),$$

and the claim of the lemma will follow by the union bound over all possible values of  $k$ . For notational convenience we will write  $D_x$  for  $D_{x,k}$  in the remainder of the proof, and we will write ‘with high probability’ (w.h.p.) for probability at least  $1 - o(\frac{1}{n})$ .

First, we have that  $\sum_{x \in C} D_x = \sum_{x \in C} |\{y \in C, |x + y| = k\}|$  is the number of pairs of points in  $C$  at distance  $k$  from each other. As observed in the proof of the second claim of Lemma 4.2, this number is w.h.p. at least  $\Omega(M p_k)$ . This implies that  $\mathbb{E}_x D_x = \frac{1}{|C|} \sum_{x \in C} D_x \geq \Omega(N p_k)$ , and hence  $(\mathbb{E}_x D_x)^2 \geq \Omega(N^2 p_k^2)$ . We will show that w.h.p.  $\mathbb{E}_x D_x^2 \leq O(\frac{1}{\theta} \cdot N^2 p_k^2)$ , and this will imply the claim of the lemma.



Let  $\tilde{C} = \{x_1, \dots, x_N\}$  be a list of  $N$  points chosen independently at random from  $\{0, 1\}^n$ . For  $i \in [N]$ , let  $\tilde{D}_i$  be the number of indices  $j \in [N]$  so that  $|x_i - x_j| = k$ . It suffices to show that w.h.p.  $\sum_{i=1}^N \tilde{D}_i^2 \leq O(\frac{1}{\theta} \cdot N^3 p_k^2)$ . We proceed similarly to the proof of Lemma 4.5. Let  $S = \sum_{i=1}^N \tilde{D}_i^2$ . For  $\{i, j\} \subseteq [N]$  let  $Z_{\{i,j\}}$  be the indicator random variable which verifies whether  $|x_i - x_j| = k$ . We have that

$$\begin{aligned} \mathbb{E}_{\tilde{C}} S &= \mathbb{E}_{\tilde{C}} \sum_{i=1}^N \left( \sum_{j=1}^N Z_{i,j} \right)^2 = \sum_{i=1}^N \sum_{j_1, j_2=1}^N \mathbb{E}_{\tilde{C}} (Z_{i,j_1} \cdot Z_{i,j_2}) \\ &= \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_{\tilde{C}} Z_{i,j} + \sum_{i=1}^N \sum_{j_1 \neq j_2} \mathbb{E}_{\tilde{C}} Z_{i,j_1} \cdot E_{\tilde{C}} Z_{i,j_2} \leq O(N^2 p_k) + O(N^3 p_k^2). \end{aligned}$$

Note that since  $d_0 \leq k \leq \frac{n}{2}$ , we have that  $Np_k \geq Np_{d_0} \geq \theta$  and hence the bound above is at most  $O(\frac{1}{\theta} \cdot N^3 p_k^2)$ .

Next, we claim that  $\sigma^2(S) \leq O(N^4 p_k^3)$ . The argument for this estimate is very similar to that for the bound on the variance of  $Y_\ell$  in the proof of the second claim of Lemma 4.5 and we omit it. This bound on the variance implies, via Chebyshev’s inequality, that w.h.p.  $S \leq O(\frac{1}{\theta} \cdot N^3 p_k^2)$ , completing the proof of the lemma. ■

ACKNOWLEDGEMENT. We are grateful to Johan Håstad for his permission to publish his conjecture (see (1)) in our paper. We would also like to thank the anonymous referee for numerous and very useful comments.

### References

- [1] A. Ashikhmin, A. Barg and S. Litsyn, *Estimates of the distance distribution of codes and designs*, Institute of Electrical and Electronics Engineers. Transactions on Information Theory **47** (2001), 1050–1061.
- [2] A. Ashikhmin, A. Barg and S. G. Vladuts, *Linear codes with many light vectors*, Journal of Combinatorial Theory. Series A **96** (2001), 396–399.
- [3] C. Bachoc, D. C. Gijswijt, A. Schrijver and F. Vallentin, *Invariant semidefinite programs*, in *Handbook on Semidefinite, Conic and Polynomial Optimization*, International Series in Operations Research and Management Science, Vol. 166, Springer, New York, 2012, pp. 219–269.
- [4] A. Barg and G. D. Forney, *Random codes: Minimum distances and error exponents*, Institute of Electrical and Electronics Engineers. Transactions on Information Theory **48** (2002), 2568–2573.

- [5] A. Barg and D. B. Jaffe, *Numerical results on the asymptotic rate of binary codes*, in *Codes and Association Schemes*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 56, American Mathematical Society, Providence, RI, 2001, pp. 25–32.
- [6] N. Biggs, *Algebraic Graph Theory*, Cambridge Tracts in Mathematics, Vol. 67, Cambridge University Press, London, 1974.
- [7] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Research Reports. Supplements **10** (1973).
- [8] R. O’Donnel, *Analysis of Boolean Functions*, Cambridge University Press, New York, 2014.
- [9] J. Friedman and J-P. Tillich, *Generalized Alon–Boppana theorems and error-correcting codes*, *SIAM Journal on Discrete Mathematics* **19** (2005), 700–718.
- [10] D. C. Gijswijt, H. D. Mittelmann and A. Schrijver, *Semidefinite code bounds based on quadruple distances*, Institute of Electrical and Electronics Engineers. *Transactions on Information Theory* **58** (2012), 2697–2705.
- [11] J. Håstad, personal communication, 2016.
- [12] P. Ivanišvili and T. Tkocz, *Comparison of moments of Rademacher chaos*, *Arkiv för Matematik* **57** (2019), 121–128.
- [13] G. Kalai and N. Linial, *On the distance distribution of codes*, Institute of Electrical and Electronics Engineers. *Transactions on Information Theory* **41** (1995), 1467–1472.
- [14] A. Khintchine, *Über dyadische Brüche*, *Mathematische Zeitschrift* **18** (1923), 109–116.
- [15] N. Kirshner and A. Samorodnitsky, *On  $\ell_4 : \ell_2$  ratio of functions with restricted Fourier support*, *Journal of Combinatorial Theory. Series A* **172** (2020), 105–202.
- [16] N. Kirshner and A. Samorodnitsky, *A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres*, Institute of Electrical and Electronics Engineers. *Transactions on Information Theory* **67** (2021), 3509–3541.
- [17] I. Krasikov and S. Litsyn, *Survey of binary Krawchouk polynomials*, in *Codes and Association Schemes*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 56, American Mathematical Society, Providence, RI, 2001, pp. 199–212.
- [18] V. I. Levenshtein, *Universal bounds for codes and designs*, in *Handbook of Coding Theory*, North-Holland, Amsterdam, 1998, pp. 499–648.
- [19] N. Linial and A. Samorodnitsky, *Linear codes and sums of characters*, *Combinatorica* **22** (2002), 497–522.
- [20] J.H. van Lint, *Introduction to Coding Theory*, Graduate Texts in Mathematics, Vol. 86, Springer, Berlin, 1999.
- [21] J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, North-Holland Mathematical Library, Vol. 16, North-Holland, Amsterdam–New York–Oxford, 1977.
- [22] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr. and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte–MacWilliams inequalities*, Institute of Electrical and Electronics Engineers. *Transactions on Information Theory* **23** (1977), 157–166.
- [23] M. Navon and A. Samorodnitsky, *On Delsarte’s linear programming bounds for binary codes*, in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*, IEEE, Piscataway, NJ, 2005, pp. 327–336.

- [24] M. Navon and A. Samorodnitsky, *Linear programming bounds for codes via a covering argument*, *Discrete & Computational Geometry* **41** (2009), 199–207.
- [25] A. Samorodnitsky, *A modified logarithmic Sobolev inequality for the Hamming cube and some applications*, <https://arxiv.org/abs/0807.1679>.
- [26] A. Schrijver, *New code upper bounds from the Terwilliger algebra and semidefinite programming*, Institute of Electrical and Electronics Engineers. *Transactions on Information Theory* **51** (2005), 2859–2866.
- [27] C. M. Theobald, *An inequality for the trace of the product of two symmetric matrices*, *Mathematical Proceedings of the Cambridge Philosophical Society* **77** (1975), 265–267.