Breaking the ϵ-Soundness Bound of the Linearity Test over GF(2)*

Tali Kaufman† Simon Litsyn‡ Ning Xie§

Abstract

For Boolean functions that are ϵ-far from the set of linear functions, we study the lower bound on the rejection probability (denoted by $\text{REJ}(\epsilon)$) of the linearity test suggested by Blum, Luby and Rubinfeld. This problem is arguably the most fundamental and extensively studied problem in property testing of Boolean functions.

The previously best bounds for $\text{REJ}(\epsilon)$ were obtained by Bellare, Coppersmith, Håstad, Kiwi and Sudan. They used Fourier analysis to show that $\text{REJ}(\epsilon) \geq \epsilon$ for every $0 \leq \epsilon \leq 1/2$. They also conjectured that this bound might not be tight for ϵ's which are close to $1/2$. In this paper we show that this indeed is the case. Specifically, we improve the lower bound of $\text{REJ}(\epsilon) \geq \epsilon$ by an additive constant that depends only on ϵ: $\text{REJ}(\epsilon) \geq \epsilon + \min\{1376\epsilon^3(1-2\epsilon)^{12}, \frac{1}{4}\epsilon(1-2\epsilon)^4\}$, for every $0 \leq \epsilon \leq 1/2$. Our analysis is based on a relationship between $\text{REJ}(\epsilon)$ and the weight distribution of a coset code of the Hadamard code. We use both Fourier analysis and coding theory tools to estimate this weight distribution.

1 Introduction

Property testing [22, 12] studies the robust characterizations of various algebraic and combinatorial objects. It often leads to a new understanding of some well-studied problems and yields insight to other areas of computer science (see survey articles [11, 20, 21] for more on property testing). The first problem that was studied under the framework of property testing, as well as being one of the most extensively investigated property testing problems, is linearity testing. A Boolean function $f : \{0,1\}^m \to \{0,1\}$ is called *linear* if for all $x, y \in \{0,1\}^m$, $f(x) + f(y) = f(x+y)$, where addition is performed modulo 2. A function f is said to be ϵ-away from linear functions if one needs to change f's value on an ϵ-fraction of its domain to make f linear. Blum, Luby and Rubinfeld [9] considered the following randomized algorithm (henceforth referred to as the "BLR test") to test if a function is linear: Given a function $f : \{0,1\}^m \to \{0,1\}$, choose uniformly at random $x, y \in \{0,1\}^m$ and reject if $f(x) + f(y) \neq f(x+y)$. We call the probability of the test accepting linear functions the *completeness* of the test while the probability of rejecting non-linear functions *soundness*. Note that in general, among other things, soundness depends on the distance parameter ϵ.

In retrospect, it is quite surprising that the analysis of such a natural test turned out to be far from simple. Much effort has been devoted to understanding the rejection probability behavior of the BLR test [9, 3, 6, 4] due to its relation to the hardness of approximating some NP-hard

*A preliminary version of this work appeared in the Proceedings of RANDOM 2008.

†CSAIL, MIT, Cambridge, MA 02139. E-mail: kaufmant@mit.edu.

‡Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, ISRAEL. E-mail: litsyn@eng.tau.ac.il. Research supported by ISF Grant 1177/06

§CSAIL, MIT, Cambridge, MA 02139. E-mail: ningxie@csail.mit.edu. Research done while the author was at State Univ. of New York at Buffalo and visiting CSAIL, MIT. Research supported in part by NSF grant 0514771.

problems [10, 6, 7, 5]. Other sequence of works considered the optimal tradeoff between query complexity and soundness of some variants of the BLR test [28, 27, 24, 14, 25] and randomness needed for linearity tests over various groups [8, 26]. Many generalizations and extension of the BLR test were also studied; for example, testing linear consistency among multiple functions [2], testing polynomials of higher degree or polynomials over larger fields [22, 1, 17, 15, 23], and testing Long Codes [5, 13].

It is clear that the completeness of the BLR test is one, i.e., if f is linear, then the BLR test always accepts. The most important quantity for the BLR test (and for many other tests as well) is the soundness, since this parameter indicates how *robust* the test characterizes the objects being tested. The soundness analysis of the BLR test was found to be pretty involved. Indeed, various papers studied the following question: For every integer $m > 0$, real number $\epsilon \in [0, 1/2]$ and all Boolean functions $f : \{0,1\}^m \to \{0,1\}$ that are ϵ-away from linear functions, what is the minimum rejection probability of the BLR linearity test. We denote this lower bound by $\textsc{rej}(\epsilon)$. That is, if we let the probability that the BLR test rejects f by $\mathrm{Rej}(f)$ and denote the set of linear functions by LIN, then

$$\textsc{rej}(\epsilon) \stackrel{\text{def}}{=} \min_{\mathrm{dist}(f,\mathrm{LIN})=\epsilon} \mathrm{Rej}(f).$$

Understanding the behavior of $\textsc{rej}(\epsilon)$ as a function of ϵ is important not only because its relation to the hardness of approximating some NP-hard problems but also due to the fact that it is a natural and fundamental combinatorial problem. The hardest cases are those where $\frac{1}{4} \le \epsilon < \frac{1}{2}$.

In this paper, by combining Fourier analysis and coding theoretic tools, we improve the previously known best lower bound for $\textsc{rej}(\epsilon)$ by an additive term depending only on ϵ for all $\epsilon \in [1/4, 1/2)$. When combined with previously known bounds, our result shows that the celebrated Fourier analysis based soundness bound [4], $\textsc{rej}(\epsilon) \ge \epsilon$, is suboptimal by an additive term that depends only on ϵ for *all* $\epsilon \in (0, \frac{1}{2})$. In other words, we show that, for every constant $\epsilon \in [\frac{1}{4}, \frac{1}{2})$, there exists a constant $\delta(\epsilon) > 0$ that is independent of m such that $\textsc{rej}(\epsilon) \ge \epsilon + \delta$.

A key ingredient of our proof is viewing the Fourier coefficients in terms of the weight distribution of codewords and applying coding bounds to them. It is hoped that techniques developed in coding theory may find other places to improve results on Boolean functions obtained by Fourier analysis.

1.1 Related research

Blum, Luby and Rubinfeld [9] first suggested the BLR linearity test and showed, based on a self-correction approach, that $\textsc{rej}(\epsilon) \ge \frac{2}{9}\epsilon$ for every ϵ. Using a combinatorial argument, Bellare et al. [6] proved that $\textsc{rej}(\epsilon) \ge 3\epsilon - 6\epsilon^2$. This bound is optimal for small ϵ but is very weak for ϵ's that are close to $\frac{1}{2}$. Bellare and Sudan [7] further showed that $\textsc{rej}(\epsilon) \ge \frac{2}{3}\epsilon$ when $\epsilon \le \frac{1}{3}$ and $\textsc{rej}(\epsilon) \ge \frac{2}{9}$ when $\epsilon > \frac{1}{3}$. All these mentioned results hold over general fields. This series of works culminated in [4], where Fourier transform techniques found their first use in PCP-related analysis. The results obtained by [4] hold for the binary field and they are the following:

$$\textsc{rej}(\epsilon) \ge \begin{cases} 3\epsilon - 6\epsilon^2 & 0 \le \epsilon \le \frac{5}{16}; \\ \frac{45}{128} & \frac{5}{16} \le \epsilon \le \frac{45}{128}; \\ \epsilon & \frac{45}{128} \le \epsilon < \frac{1}{2}. \end{cases}$$

The results of [4] show that the bounds are tight for $\epsilon \le \frac{5}{16}$. Numerical simulation results of [4] suggested that the lower bound $\textsc{rej}(\epsilon) \ge \epsilon$ for $\epsilon > \frac{5}{16}$ may be improved, but not by too much. Kiwi [18] and Kaufman and Litsyn [16] gave alternative proofs for the fact that $\textsc{rej}(\epsilon) \ge \epsilon$

for every ϵ (up to an additive term of $O(\frac{1}{2^m})$). Their proofs are more coding theory oriented. Specifically, the proofs are based on studying the weight distribution of the Hadamard code and its ϵ-away coset as well as various properties of Krawtchouk polynomials.

1.2 The main result

In the following, we present our main result showing an improved bound for $\text{REJ}(\epsilon)$. Specifically, we prove

Theorem 1.1. *Let* $\Delta(\gamma) = \frac{5\gamma}{8} - \frac{\gamma^2}{32}$. *For all* ϵ, $1/4 \leq \epsilon < 1/2$ *and for all* γ, $0 < \gamma \leq 1$,

$$\text{REJ}(\epsilon) \geq \epsilon + \min\{4096\,(1-\Delta(\gamma))^3\,\epsilon^3(1-2\epsilon)^{12}, \frac{\gamma}{2}\epsilon(1-2\epsilon)^4\}.$$

As a simple corollary by plugging in $\gamma = 1/2$ and combining our new result with known bounds for $0 \leq \epsilon \leq \frac{1}{4}$ (i.e., $\text{REJ}(\epsilon) \geq 3\epsilon - 6\epsilon^2$), we get

Corollary 1.2. *For all* ϵ, $0 \leq \epsilon < 1/2$,

$$\text{REJ}(\epsilon) \geq \epsilon + \min\{1376\epsilon^3(1-2\epsilon)^{12}, \frac{1}{4}\epsilon(1-2\epsilon)^4\}.$$

Note that for every constant $\epsilon \in [\frac{1}{4}, \frac{1}{2})$, Theorem 1.1 improves upon $\text{REJ}(\epsilon) \geq \epsilon$ by an additive *constant*. Our result improves over all previously known bounds for every $\epsilon \in [\frac{45}{128}, \frac{1}{2})$, but only by a very small quantity. For example, for $\epsilon = 0.4$, our improvement of $\text{REJ}(\epsilon)$ is about 1.024×10^{-7}. We believe our bound can be further improved systematically (we remark that our current approach already gives bounds better than that stated in the Main Theorem for ϵ's such that $1/(1-2\epsilon)^2$ are far from integers). However, as the numerical results shown in [4], one can not expect to see too much improvement over $\text{REJ}(\epsilon) \geq \epsilon$. Our improvement over $\text{REJ}(\epsilon) \geq \epsilon$ vanishes at $\epsilon = \frac{1}{2}$. This is indeed as expected since we know that $\text{REJ}(\frac{1}{2}) = \frac{1}{2}$.[1]

1.3 Proof overview

The proof has three key ingredients. We use C to denote the Hadamard code of block length $n = 2^m$ whose codewords are exactly the set of all linear functions.

The coset code $C + f$. There are two equivalent ways of viewing the BLR test: one is to think of f as a Boolean function mapping $\{0,1\}^m$ to $\{0,1\}$ and the BLR test simply picks x and y uniformly at random and check if $f(x) + f(y) = f(x+y)$. This functional viewpoint leads naturally to the beautiful Fourier analysis approach of [4], which shows that $1 - 2\text{REJ}(\epsilon)$ can be exactly expressed as a cubic sum of Fourier coefficients of the function $(-1)^f$. Another way to study the BLR test, first suggested in [18] and followed by [16], is to treat f as a vector f (by abuse of notation) of length n with $n = 2^m$. Since the set of linear functions may be viewed as the set of codewords of the Hadamard code C, the BLR test can be viewed as picking a random weight-3 codeword from $C^\perp$ (which denotes the dual code of C) and checking if it is orthogonal to f. [2] We combine these two viewpoints together by reinterpreting the Fourier analytic result in

[1] Although there are functions that are at distance exactly $1/2$ from linear functions (e.g., the complements of all linear functions), the bound $\text{REJ}(\frac{1}{2}) = \frac{1}{2}$ is only known to be met by some functions asymptotically [4].

[2] Recall that the *scalar product* between two vectors $u, v \in \{0,1\}^m$ is $u \cdot v \stackrel{\text{def}}{=} \sum_{i=1}^m u_i v_i \pmod 2$. u and v are called *orthogonal* if $u \cdot v = 0$.

the coding theoretic setting. Our simple but important observation is that the Fourier coefficients of f are equivalent to the weights of the codewords in a *coset* of C. Therefore $1-2\text{REJ}(\epsilon)$ can be expressed as a simple function of the weight distribution of the code $C+f$. Specifically, $1-2\text{REJ}(\epsilon)$ can be written as a normalized sum of cubes $\sum_{c\in C} x_c^3$, each x_c is the weight of a codeword in $C+f$, where $C+f$ is an ϵ-away coset[3] of the Hadamard code C.

Maximization Problem. In order to obtain a lower bound on $\text{REJ}(\epsilon)$, we need to obtain an upper bound on a sum that involves the weight distribution of $C+f$. To this end, we reformulate our problem as a **Maximal Sum of Cubes Problem**, in which we look for an upper bound on the sum of cubes of a set of integers under certain constraints. The bound $\text{REJ}(\epsilon)=\epsilon$ obtained by [4] corresponds to the simple optimal configuration in which all the codewords of $C+f$ are of weight $\frac{1}{2}n$ except a constant number (i.e. $\frac{1}{(1-2\epsilon)^2}$) of them are of weight ϵn (one can use coding theory argument to show that there can't be more than $\frac{1}{(1-2\epsilon)^2}$ codewords of weight ϵn). Moreover, this is the unique configuration that meets the bound $\text{REJ}(\epsilon)=\epsilon$. Any deviation from the optimal configuration implies an improved lower bound on $\text{REJ}(\epsilon)$. Our strategy thus is to show that this optimal weight distribution is not achievable for $C+f$ due to some special properties of the code $C+f$. In particular, we will focus on the following two ways in which the optimal configuration may break down:

1. There exists a codeword of weight larger than $\frac{n}{2}$ in $C+f$.
2. The number of codewords in $C+f$ of weight at most $(\epsilon+\eta)n$ is less than $\frac{1}{(1-2\epsilon)^2}$, for some positive number η.

A natural tool to show that one of the above properties holds is the well-known Johnson Bound. Roughly speaking, the Johnson bound offers a bound on the maximum number of codewords of a specific weight in a code with some specific minimum distance. However, it turns out that the Johnson bound does *allow* the optimal configuration for the code $C+f$ (which yields $\text{REJ}(\epsilon)=\epsilon$ as discussed above), so we fail to get any improvement by applying the Johnson bound directly to $C+f$. The way we overcome this is by considering a new code $C|_{\mathcal{F}}$ of *shorter block length* and applying to it a slightly stronger variant of the commonly used Johnson bound (a variant which enables us to bound the number of codewords of *at least* (or *at most*) a specific weight). The possible switch from the code $C+f$ to the code $C|_{\mathcal{F}}$ turns out to be crucial in our analysis.

From the code $C+f$ to the code $C|_{\mathcal{F}}$. We consider the code $C|_{\mathcal{F}}$ of block length $n'=\epsilon n$, obtained from C by restricting it to the ϵn non-zero coordinates of f. This code is a linear code and in general has the same number of codewords as the original code $C+f$. Indeed we show that if it contains fewer codewords, then an improved lower bound on $\text{REJ}(\epsilon)$ is immediate. A nice property of this new code is that there is a one-to-one correspondence between the weight of a codeword in $C|_{\mathcal{F}}$ and the weight of the corresponding codeword in $C+f$. Since $C|_{\mathcal{F}}$ is a linear code, its minimum distance equals the minimum weight of its codewords. If this minimum weight is small, then by the one-to-one relation between the weights of $C+f$ and that of $C|_{\mathcal{F}}$, the heaviest codeword in $C+f$ will have a large weight, which yields an improved lower bound

[3]To make this clear, we remind the reader that the weight distribution of a code C is a set of integers that represent the numbers of codewords in C of different weights, where the weight of a codeword is the number of coordinates at which the codeword is non-zero. A vector f is ϵ-away from a code C if one needs to change an ϵ-fraction of f's coordinates to make it belong to C. An ϵ-away coset of C is obtained by adding a vector f to every codeword in C, where f is ϵ-away from C.

for $\text{REJ}(\epsilon)$ according to Condition 1 from above. However, if the maximum weight of $C + f$ is small, or equivalently, the minimum distance of $C|_{\mathcal{F}}$ is large, then by applying the Johnson bound to $C|_{\mathcal{F}}$, we get that the number of codewords lying between weight ϵn and $(\epsilon + \eta)n$ in $C + f$ is less than the optimal bound $\frac{1}{(1-2\epsilon)^2}$, which also yields an improved lower bound for $\text{REJ}(\epsilon)$ by Condition 2 mentioned before.

The intuitive reason that we benefit from applying the Johnson bound to $C|_{\mathcal{F}}$ rather than to $C + f$ is straightforward: The block length of $C|_{\mathcal{F}}$ is much smaller than the block length of $C + f$, but the number of codewords in $C|_{\mathcal{F}}$ is the same as $C + f$.[4]

The relations between the three codes in consideration, namely C, $C + f$, and $C|_{\mathcal{F}}$ (for a code C and a vector f that is ϵ-away from C), as well as the idea of looking at a restricted code of smaller block length in order to get better coding bounds, might have other applications.

1.4 Organization

Section 2 introduces necessary notation and definitions. In Section 3 we show that, for every f that is ϵ-away from linear, $\text{Rej}(f)$ can be expressed as a function of the weight distribution of a coset of the Hadamard code. Then we reformulate the problem of lower bounding $\text{REJ}(\epsilon)$ as a maximization problem in Section 4. In Section 5 we study the weight distribution of a restricted code of the coset code and then prove the Main Theorem in Section 6. Several technical claims appear in the Appendix.

2 Preliminaries

We write $[n]$ for the set $\{1, \ldots, n\}$, where n is a positive integer. Let v be a vector in $\{0,1\}^n$. We use $v(i)$ to denote the ith bit of v for every $1 \leq i \leq n$. The weight of v, denoted $\text{wt}(v)$ is the number of non-zero bits in v. A code C of block length n is a subset of $\{0,1\}^n$. C is called a *linear code* if C is a linear subspace. Let $u, v \in \{0,1\}^n$. The *distance* between u and v is defined to be the number of bits at which they disagree: $\text{dist}(u,v) = |\{i \in [n] \mid u(i) \neq v(i)\}| = \text{wt}(u - v)$. The minimum distance of a code C is $\min_{u,v \in C, u \neq v} \text{dist}(u,v)$. If C is a linear code, then the minimum distance of C equals the minimum weight of codewords in C. Let C be a code of block length n. The distance of $v \in \{0,1\}^n$ from code C is the minimum distance between v and codewords in C, i.e., $\text{dist}(v, C) \stackrel{\text{def}}{=} \min_{c \in C} \text{dist}(v, c)$. By abuse of notation, in the following, we use C to denote the Hadamard code and $C^{\perp}$ to denote its dual Hamming code.

Recall that a function $c : \{0,1\}^m \to \{0,1\}$ is *linear* if for all $x, y \in \{0,1\}^m$, $c(x) + c(y) = c(x + y)$. An equivalent characterization is: c is linear if and only if $c(x) = \alpha \cdot x = \sum_{i=1}^{m} \alpha_i x_i \pmod 2$ for some $\alpha \in \{0,1\}^m$, and we denote this linear function by c_α and denote the set of all such functions by LIN. Let $f, g : \{0,1\}^m \to \{0,1\}$. The (relative) distance between f and g is defined to be the fraction of points at which they disagree: $\text{dist}(f,g) \stackrel{\text{def}}{=} \Pr_{x \in \{0,1\}^m}[f(x) \neq g(x)]$. The distance between a function f and linear functions is the minimum distance between f and

[4] The reason we are able to improve the bound $\text{REJ}(\epsilon) \geq \epsilon$ by a constant is more subtle: For $\frac{1}{4} \leq \epsilon \leq \frac{1}{2}$, there is a "reciprocal" relationship between the *relative weights* of codeword in C and corresponding codeword in $C|_{\mathcal{F}}$; that is, the smaller the relative weight in C, the larger the relative weight in $C|_{\mathcal{F}}$, and vice versa. Note that the denominator of the expression in the Johnson bound is $\frac{d}{n} - 2\frac{w}{n}(1 - \frac{w}{n})$ after dividing by n^2. Therefore the Johnson bound will give better bounds when $\frac{w}{n}(1 - \frac{w}{n})$ gets smaller, or, when w/n is very close to either 0 or 1. By switching from C to $C|_{\mathcal{F}}$, $\frac{w}{n}$ is mapped to $\frac{w'}{n'}$. The advantage of changing to $C|_{\mathcal{F}}$ is that it makes the distance between $\frac{w'}{n'}$ and 1 smaller than the distance between $\frac{w}{n}$ and zero. This advantage disappears at $\epsilon = 1/2$, therefore we get no improvement at that point, as expected.

any linear function: $\mathrm{dist}(f, \mathrm{LIN}) \stackrel{\mathrm{def}}{=} \min_{g \in \mathrm{LIN}} \mathrm{dist}(f, g)$. A function f is said to be ϵ-*away* from linear functions if its distance from linear functions is ϵ, and is said to be ϵ-*far* from linear functions if the distance is at least ϵ.

Next we introduce some basic notions in Fourier analysis. We will focus on functions defined over the Boolean cube. Note that the set of functions $f : \{0,1\}^m \to \mathbb{R}$ forms a vector space of dimension 2^m. A convenient orthonormal basis for this vector space is the following collection of functions called characters: $\psi_\alpha(x) = (-1)^{\alpha \cdot x} = (-1)^{c_\alpha(x)}$, where $\alpha \in \{0,1\}^m$. Consequently, any $f(x) : \{0,1\}^m \to \mathbb{R}$ can be expanded as

$$f(x) = \sum_{\alpha \in \{0,1\}^m} \hat{f}_\alpha \psi_\alpha(x),$$

where $\hat{f}_\alpha = \langle f, \psi_\alpha \rangle \stackrel{\mathrm{def}}{=} \frac{1}{2^m} \sum_{x \in \{0,1\}^m} f(x)\psi_\alpha(x)$ is called the α-th Fourier coefficient of f. Define $h(x) = (-1)^{f(x)}$. Note that the range of $h(x)$ is $\{-1, 1\}$.

One can encode f as an $n = 2^m$ bit codeword in $\{0,1\}^n$ by enumerating all its values on the Boolean cube, and by abuse of notation we denote this codeword by f. The same encoding applied to the set of linear functions $\{c_\alpha\}_\alpha$ gives rise to the Hadamard code C, in which we (abusing notation again) denote the corresponding codewords by $\{c_\alpha\}_\alpha$.

For $0 \le \epsilon \le 1/2$, we let $\beta = 1 - 2\epsilon$.

We are going to use the following two elementary inequalities in our analysis. The proofs of these inequalities can be found in the Appendix.

Lemma 2.1. *For all real y with $0 \le y \le 1/2$,*

$$\frac{1}{1-y} - y \ge \frac{1}{\sqrt{1-2y^2}}.$$

Lemma 2.2. *Let γ be a constant with $0 \le \gamma \le 1$. Then for all real y with $0 \le y \le 1/2$,*

$$\frac{1}{(1-y)^2} - \frac{1}{1-2y^2} - \gamma \frac{y}{1-y} \ge (8 - 5\gamma)y^2.$$

3 The coset code $C + f$

Using Fourier analytic tools, Bellare et al. proved the following result in their seminal paper.

Lemma 3.1 ([4])**.** *Let $f : \{0,1\}^m \to \{0,1\}$ and $h(x) = (-1)^{f(x)}$. Then (recall that* $\mathrm{Rej}(f)$ *is the probability that BLR test rejects f)*

$$\mathrm{Rej}(f) = \frac{1}{2}\left(1 - \sum_{\alpha \in \{0,1\}^m} \hat{h}_\alpha^3\right).$$

Sometimes reformulating a Boolean function problem as a coding theoretic problem offers new perspectives. To this end, we need to introduce the standard notion of coset codes. Let D be a linear code of block length n and let $f \in \{0,1\}^n$ such that $f \notin D$, the f-*coset* of D is $D + f \stackrel{\mathrm{def}}{=} \{c + f \mid c \in D\}$. Note that $|D + f| = |D|$. The *weight distribution* or *spectrum* of a code D is $B^D = (B_0^D, B_1^D, \cdots, B_n^D)$, where $B_i^D = |\{c \in D \mid \mathrm{wt}(c) = i\}|$.

Now we switch the viewpoint from Boolean functions to vectors in the Boolean cube. That is, we transform Boolean function $f : \{0,1\}^m \to \{0,1\}$ into a vector $f \in \{0,1\}^n$ by evaluating

the Boolean function f on every point in the Boolean cube. Using the relation between linear functions and Hadamard code, we have the following coding theoretic formula for $\mathrm{Rej}(f)$:

Lemma 3.2. *Let $f \in \{0,1\}^n$, then*[5]

$$\mathrm{Rej}(f) = \frac{1}{2}\left(1 - \frac{1}{n^3}\sum_{i=0}^{n} B_i^{C+f}(n-2i)^3\right). \tag{1}$$

Proof. By the definition of Fourier coefficient,

$$\begin{aligned}\hat{h}_\alpha = \langle h, \psi_\alpha\rangle = \langle (-1)^f, (-1)^{c_\alpha}\rangle &= \frac{1}{2^m}\sum_{x\in\{0,1\}^m}(-1)^{f(x)+c_\alpha(x)} \\ &= \Pr_x[f(x) = c_\alpha(x)] - \Pr_x[f(x) \neq c_\alpha(x)] \\ &= 1 - \frac{2\mathrm{dist}(f, c_\alpha)}{n} = \frac{n - 2\mathrm{wt}(f + c_\alpha)}{n},\end{aligned} \tag{2}$$

where in the last step we use the fact that, for binary vectors u and v, $\mathrm{dist}(u,v) = \mathrm{wt}(u-v) = \mathrm{wt}(u+v)$. Lemma 3.1 now gives

$$\begin{aligned}\mathrm{Rej}(f) &= \frac{1}{2}\left(1 - \sum_{\alpha\in\{0,1\}^m}\hat{h}_\alpha^3\right) = \frac{1}{2}\left(1 - \sum_{\alpha\in\{0,1\}^m}\frac{(n-2\mathrm{wt}(f+c_\alpha))^3}{n^3}\right) \\ &= \frac{1}{2}\left(1 - \sum_{c\in C}\frac{(n-2\mathrm{wt}(f+c))^3}{n^3}\right) \\ &= \frac{1}{2}\left(1 - \frac{\sum_{i=0}^{n} B_i^{C+f}(n-2i)^3}{n^3}\right),\end{aligned}$$

where in the last step we change summation over codewords in C to summation over weights of the codewords in $C+f$. □

This relation between the Fourier coefficients of $(-1)^f$ and the weight distribution of the coset code $C+f$ as employed in (2) and (1) seems to be new and may find applications in other places.

Since $\mathrm{Rej}(f)$ is now expressed as a weight distribution of the coset code $C+f$, our next step is to study how the codewords in $C+f$ are distributed so that to make the rejection probability minimum.

4 Maximization problem

Note that we can rewrite Lemma 3.2 as

$$\mathrm{Rej}(f) = \frac{1}{2} - \frac{\sum_{c\in C}(n-2\mathrm{wt}(c+f))^3}{2n^3} = \frac{1}{2} - \frac{1}{2n^3}\sum_{c\in C} x_c^3,$$

[5] In fact we can state Lemma 3.2 in terms of x_c (see definition in next Section) directly and without invoking the notion of weight distribution of the coset code. However, we prefer to state it this way because studying the spectrum of the coset code was the starting point of this research and we hope this connection may find applications in other places.

where

$$x_c \stackrel{\text{def}}{=} n - 2\text{wt}(c+f),$$

for every $c \in C$ and there are n codewords in C. In order to prove an improved bound $\text{REJ}(\epsilon) \geq \epsilon + \delta$, all we need to show is, for *every* f that is ϵ-away from linear functions, $\text{Rej}(f) \geq \epsilon + \delta$. Hence our goal of getting a better *lower bound* than ϵ for $\text{REJ}(\epsilon)$ is equivalent to, for every vector f with $\text{dist}(f, C) = \epsilon n$, getting a better *upper bound* than $1 - 2\epsilon$ for $\frac{1}{n^3}\sum_{c\in C} x_c^3$. This observation motivates the following measure of improvement (gain) and reformulating the problem of lower bounding $\text{REJ}(\epsilon)$ as a **Maximal Sum of Cubes Problem**.

Definition 4.1. *Let $x_c = n - 2\text{wt}(c+f)$ for every $c \in C$. Define*

$$\text{GAIN}(f) = \frac{1}{n^3}\left((1-2\epsilon)n^3 - \sum_{c\in C} x_c^3\right).$$

Consequently, if $\text{dist}(f, C) = \epsilon n$, then $\text{Rej}(f) = \epsilon + \frac{1}{2}\text{GAIN}(f)$.

Since f is ϵ-away from C, it follows that $x_c \leq (1-2\epsilon)n$ for all $c \in C$. We further observe another constraint on the set of integers $\{x_c\}_{c\in C}$ is that their Euclidean norm is n^2.

Claim 4.2. *We have $\sum_{c\in C} x_c^2 = n^2$.*

This claim follows directly from Parseval's equality. An alternative proof, based on the norm-preserving property of the Hadamard matrix, was given in [16].

The following lemma shows, if these two constraints are the only constraints on $\{x_c\}_{c\in C}$, then the bound $\text{REJ}(\epsilon) \geq \epsilon$ is essentially optimal. However, as we will see in the next section, since $\{x_c\}_{c\in C}$ are related to the weight distribution of $C + f$, the properties of the code $C + f$ impose more constraints on $\{x_c\}_{c\in C}$, thus making this optimal bound unattainable.

Lemma 4.3. *Consider the following **Maximal Sum of Cubes Problem**: Let $0 < \beta \leq 1$ be a constant and n be a large enough integer. For a set of n integers $\{x_c\}_{c\in C}$, find the maximum of $\sum_{c\in C} x_c^3$ under the constraints:*

$$\sum_{c\in C} x_c^2 = n^2$$
$$\forall c \in C : x_c \leq \beta n.$$

The maximum is achieved at the following optimal configuration[6]*: $\frac{1}{\beta^2}$ of the x_c's are assigned the maximum value βn, and the rest are assigned the value zero. The maximum thus obtained is βn^3.*

Proof. Let $\{x_c\}_{c\in C}$ be a set of integers satisfying the constraints in the **Maximal Sum of Cubes Problem**. Then $x_c \leq \beta n$ for all $c \in C$. Since $x_c^2 \geq 0$, it follows that $x_c^3 \leq \beta n x_c^2$ for all $c \in C$ and furthermore, equality holds if and only if $x_c = 0$ or $x_c = \beta n$. Now summing over c in C implies that

$$\sum_{c\in C} x_c^3 \leq \beta n \sum_{c\in C} x_c^2 = \beta n^3. \tag{3}$$

Moreover, the equality in (3) is attained only if all of the values of x_c are either zero or βn. This is possible only if $\frac{1}{\beta^2}$ of the x_c's equal βn, and the rest equal zero. In that case $\sum_{c\in C} x_c^3 = \beta n^3$. □

[6] Another requirement necessary to attain the optimal bound is that $\frac{1}{\beta^2}$ is an integer. Therefore we already see some improvement upon $\text{REJ}(\epsilon) \geq \epsilon$ without any further calculation for all ϵ such that $\frac{1}{(1-2\epsilon)^2}$ is not an integer.

Note that in our setting $x_c = n - 2\mathrm{wt}(c+f)$, so $\beta = 1-2\epsilon$ and consequently $\sum_{c\in C} x_c^3 \le (1-2\epsilon)n^3$, where $0 \le \epsilon \le 1/2$. We will employ the following two lemmas on $\textsc{gain}(f)$ to obtain improvement upon the bound $\textsc{rej}(\epsilon) \ge \epsilon$.

Lemma 4.4. *Let $\{x_c\}_{c\in C}$ be a set of integers satisfying the constraints in the **Maximal Sum of Cubes Problem** stated in Lemma 4.3. If there exists an $x_{\tilde{c}}$ such that $x_{\tilde{c}} = -\delta n$ for some $\delta > 0$, then $\textsc{gain}(f) \ge \min\{2\delta^3, 2\beta^3\}$.*

Proof. We first consider the case that $\delta \le \beta$. Note that if we replace $x_{\tilde{c}}$ with $-x_{\tilde{c}}$ and keep other integers unchanged, then the new set of integers satisfy all the constraints in the **Maximal Sum of Cubes Problem**, so we have

$$\beta n^3 \ge \sum_{c\in C, c\neq\tilde{c}} x_c^3 + (-x_c)^3 = \sum_{c\in C} x_c^3 + 2|x_{\tilde{c}}|^3 = \sum_{c\in C} x_c^3 + 2(\delta n)^3.$$

It follows that $\textsc{gain}(f) \ge 2\delta^3$. Now consider the case that $\delta > \beta$. Note that $\sum_{c\neq\tilde{c}} x_c^2 = (1-\delta^2)n^2$ and $x_c \le \beta n$ for every $c \in C$, it follows immediately that $\sum_{c\neq\tilde{c}} x_c^3 \le \beta n \sum_{c\neq\tilde{c}} x_c^2 = \beta(1-\delta^2)n^3$. Therefore,

$$\textsc{gain}(f) = \frac{1}{n^3}(\beta n^3 - \sum_{c\in C} x_c^3) \ge \beta\delta^2 + \delta^3 \ge 2\beta^3. \qquad \square$$

Lemma 4.5. *Let $\eta > 0$ and $\{x_c\}_{c\in C}$ be a set of integers satisfying the constraints in the **Maximal Sum of Cubes Problem** stated in Lemma 4.3. If the number of $x_{\tilde{c}}$'s such that $x_{\tilde{c}} \ge (\beta-\eta)n$ is at most $\lfloor \frac{1}{\beta^2} \rfloor - 1$, then $\textsc{gain}(f) \ge \beta^2\eta$.*

Proof. Set $M = \lfloor \frac{1}{\beta^2} \rfloor$. Let $\{y_1, \ldots, y_n\}$ be a permutation of $\{x_c\}_{c\in C}$ such that $\beta n \ge y_1 \ge \cdots \ge y_n$. We have $y_1^2 + \cdots + y_n^2 = n^2$ and $y_M \le (\beta-\eta)n$. Define T to be: $T = y_1^2 + \cdots + y_{M-1}^2$. Then we have $T \le (M-1)(\beta n)^2 \le (\frac{1}{\beta^2} - 1)\beta^2 n^2$, and $y_M^2 + \cdots + y_n^2 = n^2 - T$. Therefore,

$$\begin{aligned}
\sum_{c\in C} x_c^3 = \sum_{i=1}^{n} y_i^3 &\le (\sum_{i=1}^{M-1} y_i^2)\beta n + (\sum_{i=M}^{n} y_i^2)(\beta-\eta)n \\
&= n^2(\beta-\eta)n + \eta n T \\
&\le n^2(\beta-\eta)n + \eta n(\frac{1}{\beta^2} - 1)\beta^2 n^2 \\
&= \beta n^3 - \beta^2\eta n^3.
\end{aligned}$$

$\square$

5 From the code $C+f$ to the code $C|_{\mathcal{F}}$

We denote by $\mathcal{F}$ the set of coordinates at which f is non-zero, i.e., $\mathcal{F} = \{i \in [n] \mid f(i) = 1\}$. Note that $|\mathcal{F}| = \mathrm{wt}(f)$. In the following we consider a code $C|_{\mathcal{F}}$ which will enable us to get some insight into the weight distribution of the code $C+f$.

First observe that, since we are only interested in the weight distribution of $C+f$, without loss of generality, we may assume that $\mathrm{wt}(f) = \epsilon n$. To see this, suppose that $c' \in C$ is the closest codeword to f (if there are more than one such codeword, then we may pick one arbitrarily). Since $\mathrm{dist}(f, C) = \epsilon n$, f can be written as $f = c' + c_{\epsilon n}$, with $\mathrm{wt}(c_{\epsilon n}) = \epsilon n$. Since C is a linear code, $C + f = \{c + f \mid c \in C\} = \{c + c' + c_{\epsilon n} \mid c \in C\} = \{\tilde{c} + c_{\epsilon n} \mid \tilde{c} \in C\} = C + c_{\epsilon n}$, where $\tilde{c} \stackrel{\text{def}}{=} c + c'$.

Definition 5.1. *Let C be a code of block length n and $f \in \{0,1\}^n$ be a vector of weight ϵn. We define the code $C|_{\mathcal{F}}$ of block length ϵn to be the code obtained by restricting code C to the non-zero coordinates of f. For convenience of notation, we will use $C' = C|_{\mathcal{F}}$ from now on and call it the "restricted code" of C.*

Recall that C is the Hadamard code of block length n. We denote the codeword $0^n \in C$ by c_0. For every codeword $c \in C$, we use c' to denote the corresponding codeword in the restricted code C'.

The following lemma shows a one-to-one correspondence between the weight of a codeword in $C + f$ and the weight of the corresponding codeword in C'.

Lemma 5.2. *Let c be a codeword in the Hadamard code C and $c' \in C'$ be the restriction of c to coordinates in $\mathcal{F}$. Let $x_c = n - 2\mathrm{wt}(c + f)$, then*

$$x_c = \begin{cases} (1-2\epsilon)n, & \text{if } c = c_0, \\ 4\mathrm{wt}(c') - 2\epsilon n, & \text{otherwise,} \end{cases}$$

where $0 \le \epsilon \le 1/2$.

Proof. For $c = c_0$, $\mathrm{wt}(c_0 + f) = \mathrm{wt}(f) = \epsilon n$, hence $x_{c_0} = (1-2\epsilon)n$. Next we consider the case that $c \neq c_0$. Since C is a Hadamard code, $\mathrm{wt}(c) = n/2$, i.e., there are $n/2$ ones and $n/2$ zeros in c. Because c has $\mathrm{wt}(c')$ ones in coordinates that are in $\mathcal{F}$, c has $n/2 - \mathrm{wt}(c')$ ones in coordinates that are in $[n] \setminus \mathcal{F}$. Note that f does not flip the bits in $[n] \setminus \mathcal{F}$, therefore $c + f$ also has $n/2 - \mathrm{wt}(c')$ ones in coordinates that are in $[n] \setminus \mathcal{F}$. Since $|f| = \epsilon n$, c has $\epsilon n - \mathrm{wt}(c')$ zeros in coordinates that are in $\mathcal{F}$, therefore $c + f$ has $\epsilon n - \mathrm{wt}(c')$ ones in coordinates that are in $\mathcal{F}$. It follows that $\mathrm{wt}(c + f) = n/2 - \mathrm{wt}(c') + \epsilon n - \mathrm{wt}(c') = (1/2 + \epsilon)n - 2\mathrm{wt}(c')$ and $x_c = 4\mathrm{wt}(c') - 2\epsilon n$. □

Lemma 5.3. *Either C' is a linear code or $\textsc{gain}(f) \ge 2(1-2\epsilon)^3$, for every $1/4 \le \epsilon \le 1/2$.*

Proof. Since C' is a restriction of linear code C, C' is a linear code if and only if all the codewords in C' are distinct. If C' is not a linear code, then there exist two distinct codewords c_1 and c_2 such that $c_1 = c_2$. This implies that there is a codeword c'_3 different from c'_0 such that $c'_3 = \vec{0}$. By Lemma 5.2, $x_{c'_3} = -2\epsilon n$. Since $2\epsilon \ge 1 - 2\epsilon$, by Lemma 4.4, $\textsc{gain}(f) \ge 2(1-2\epsilon)^3$. □

Since $2(1-2\epsilon)^3$ is always larger than the gain we are going to prove, from now on, we will focus on the case that C' is a linear code. Let $n' = \epsilon n$ be the block length of C', and d' be the minimum distance of C'. Note that C' contains n codewords. The following simple bound is useful.

Theorem 5.4 (Plotkin bound [19])**.** *Let D be a binary code of block length n and minimum distance d. If $d \ge n/2$, then D has at most $2n$ codewords.*

Applying this bound to the restricted code C' yields the following:

Claim 5.5. *For every $1/4 \le \epsilon < 1/2$, the minimum distance of C' satisfies $d' < n'/2$.*

Proof. Suppose $d' \ge n'/2$, then by the Plotkin bound stated in Theorem 5.4, C' has at most $2n' = 2\epsilon n < n$ codewords, a contradiction. □

6 Proof of the Main Theorem

In this section, we give a proof of the main theorem.

Theorem 1.1 (Main Theorem). *Let* $\Delta(\gamma) = \frac{5\gamma}{8} - \frac{\gamma^2}{32}$. *For all* ϵ, $1/4 \leq \epsilon < 1/2$ *and for all* γ, $0 < \gamma \leq 1$,

$$\textsc{rej}(\epsilon) \geq \epsilon + \min\{4096\,(1-\Delta(\gamma))^3\,\epsilon^3(1-2\epsilon)^{12}, \frac{\gamma}{2}\epsilon(1-2\epsilon)^4\}.$$

Our proof will rely on the following basic coding theorem which bounds the number of codewords of weight at least w. This is a slightly stronger variant of the well-known Johnson bound, for a proof see, e.g., the Appendix in [5].

Theorem 6.1 (Johnson bound). *Let* D *be a binary code of block length* n *and minimum distance* d. *Let* $B(n,d,w)$ *denote the maximum number of codewords in* D *of weight* at least w. *Suppose* $nd > 2w(n-w)$, *then*

$$B(n,d,w) \leq \frac{nd}{nd - 2w(n-w)}. \tag{4}$$

Remark on notation. *In the following, by abuse of notation, we write* $\textsc{gain}(\epsilon)$ *for the minimum of* $\textsc{gain}(f)$, *where* f *ranges over all functions that are* ϵ*-away from linear functions.*

The basic idea of the proof is the following. Since there is a one-to-one correspondence between the weight of the codewords in $C + f$ and that of C', we can safely work with the spectrum of C' for the purpose of proving lower bound on $\textsc{gain}(\epsilon)$. Because C' is a linear code, its minimum distance d is equal to the minimum weight of its codewords. If d is small (much smaller than $n'/2$), then there is a low weight codeword in C'. Consequently, there is an $x_c = -\delta n$ for some positive δ, which implies a large gain by Lemma 4.4. However, if d is large (very close to $n'/2$), then we can apply the Johnson bound in Theorem 6.1 to C' to show that the number of x_c such that $x_c \geq (1-2\epsilon-\eta)n$ is less than $\frac{1}{(1-2\epsilon)^2}$ for some positive η. This also implies a large gain by Lemma 4.5. Moreover, as shown below in Lemma 6.2, there is a trade-off relation between these two gains: If δ is small then η is large and vice versa. This trade-off enables us to prove that $\textsc{gain}(\epsilon) = \Omega(1)$ for every $1/4 \leq \epsilon < 1/2$.

Now we fill in the details of the proof. Let $1/4 \leq \epsilon < 1/2$ be a fixed constant, define a quadratic function

$$G(x) \stackrel{\text{def}}{=} (x - \frac{1}{4\epsilon})(x - \frac{1}{4\epsilon} + 1) + \frac{1 - 3(1-2\epsilon)^2}{4(1-2(1-2\epsilon)^2)}.$$

By Lemma 5.2, for all $c \neq c_0$ it holds that $4\mathrm{wt}(c') - 2\epsilon n = x_c \leq (1-2\epsilon)n$, or equivalently $\mathrm{wt}(c') \leq \frac{n}{4} = \frac{n'}{4\epsilon}$. Suppose the minimum distance of C' is $d = (1/2 - \delta')n'$. By Claim 5.5, δ' is positive.

Note that $x_{c_0} = (1-2\epsilon)n$ and for all $c \neq c_0$, $x_c \geq (1-2\epsilon-\eta)n$ iff $\mathrm{wt}(c') \geq (\frac{1}{4\epsilon} - \eta')n'$, where $\eta' = \frac{\eta}{4\epsilon}$. Therefore, in order to apply Lemma 4.5, it suffices to show that there are at most $\frac{1}{(1-2\epsilon)^2} - 2$ codewords in C' of weight at least $(\frac{1}{4\epsilon} - \eta')n'$ for some $\eta' > 0$. In the next lemma, we show a trade-off relation between δ' and η'. More specifically, there is a monotone decreasing function $F(\cdot)$ relates δ' and η'.

Lemma 6.2 (Trade-off Lemma). *For every* ϵ, $\frac{1}{4} \leq \epsilon < \frac{1}{2}$, *there exist two positive numbers* δ_0 *and* η_0 *which depend only on* ϵ, *and a function* F *which is parameterized only by* ϵ *and is monotone decreasing in* $[0, \eta_0]$, *such that the following holds: For all* δ' *with* $0 < \delta' < \delta_0$, *let* $\eta' = F(\delta')$, *and if the minimum distance of code* C' *is* $(\frac{1}{2} - \delta')n'$, *then* C' *has at most* $\frac{1}{(1-2\epsilon)^2} - 2$ *codewords of weight at least* $(\frac{1}{4\epsilon} - \eta')n'$.

Proof. Let $\eta' = F(\delta')$ where F will be defined later. We apply the Johnson bound stated in Theorem 6.1 to the restricted code C'. Plugging in minimum distance $d = (1/2 - \delta')n'$ and minimum weight $w' \stackrel{\text{def}}{=} (\frac{1}{4\epsilon} - \eta')n'$ into the right-hand side of (4), we impose that δ' and η' satisfy the following:

$$\frac{\frac{1}{2} - \delta'}{(\frac{1}{2} - \delta') - 2(\frac{1}{4\epsilon} - \eta')(1 - \frac{1}{4\epsilon} + \eta')} = \frac{1}{(1-2\epsilon)^2} - 2. \tag{5}$$

If we solve (5) to get $F(\delta') = \eta'$, then the statement in the lemma about η' is also true for all $\eta'' \leq \eta'$, provided η' is not too large[7]. By some elementary algebraic manipulations, we have

$$\begin{aligned}\delta' &= \frac{1}{2} - 2(\frac{1}{4\epsilon} - \eta')(1 - \frac{1}{4\epsilon} + \eta')\frac{1 - 2(1-2\epsilon)^2}{1 - 3(1-2\epsilon)^2} \\ &= \frac{2(1 - 2(1-2\epsilon)^2)}{1 - 3(1-2\epsilon)^2} G(\eta'). \end{aligned} \tag{6}$$

Note that since $1/4 \leq \epsilon < 1/2$, we have both $1 - 2(1-2\epsilon)^2$ and $1 - 3(1-2\epsilon)^2$ are positive. Therefore, whenever there are positive values η' to make $G(\eta')$ positive, the corresponding δ' will be positive as well.

Rewrite $G(\eta')$ as $G(\eta') = \eta'^2 - \bar{b}\eta' + \bar{c}$, where $\bar{b} = \frac{1}{2\epsilon} - 1 > 0$ and $\bar{c} = \frac{1-3(1-2\epsilon)^2}{4(1-2(1-2\epsilon)^2)} - \frac{1}{4\epsilon} + \frac{1}{16\epsilon^2}$. Since $\bar{b}^2 - 4\bar{c} = \frac{(1-2\epsilon)^2}{1-2(1-2\epsilon)^2} > 0$, there are two distinct real roots for $G(\eta') = 0$. Denote these two roots by η_1 and η_2 with $\eta_1 > \eta_2$. Then $G(\eta')$ assumes positive values for $\eta' > \eta_1$ and $\eta' < \eta_2$. Since $\eta_1 > \frac{1}{4\epsilon} - \frac{1}{2}$ but we are bounding the number of codewords of weight at least $w' = (\frac{1}{4\epsilon} - \eta')n' > \frac{1}{2}n'$, which requires $\eta' < \frac{1}{4\epsilon} - \frac{1}{2}$, so we only need to look at the region where $\eta' < \eta_2$. Therefore, we have:

$$\begin{aligned} &\text{There are positive } \eta' \text{ to make } G(\eta') \text{ positive} \\ \iff &\eta_2 > 0 \\ \iff &\bar{c} > 0 \\ \iff &\frac{1 - 3(1-2\epsilon)^2}{4(1-2(1-2\epsilon)^2)} - \frac{1}{4\epsilon} + \frac{1}{16\epsilon^2} > 0 \\ \iff &(\frac{1}{2\epsilon} - 1)^2 > \frac{(1-2\epsilon)^2}{1-2(1-2\epsilon)^2} \\ \iff &\epsilon > 1/6. \end{aligned}$$

That is, for all ϵ, $1/4 \leq \epsilon < 1/2$, $\eta_2 > 0$. Note that $G(\eta')$ is monotone decreasing in $[0, \eta_2]$, so the inverse of G exists, which we denote by G^{-1}. Finally, we set $F(\delta') = G^{-1}\left(\frac{1-3(1-2\epsilon)^2}{2(1-2(1-2\epsilon)^2)}\delta'\right)$, $\delta_0 = \frac{2(1-2(1-2\epsilon)^2)}{1-3(1-2\epsilon)^2}\bar{c} = \frac{2(1-2(1-2\epsilon)^2)}{1-3(1-2\epsilon)^2}\left(\frac{1-3(1-2\epsilon)^2}{4(1-2(1-2\epsilon)^2)} - \frac{1}{4\epsilon} + \frac{1}{16\epsilon^2}\right)$, and $\eta_0 = \eta_2$ to complete the proof. □

Combining this Trade-off Lemma with the two lemmas regarding $\textsc{gain}(\epsilon)$, Lemma 4.4 and Lemma 4.5, immediately implies a lower bound for $\textsc{gain}(\epsilon)$. However, such a lower bound is of the form some minimum over the interval $(0, \eta_0)$. Due to the monotonicity of the two gain functions we proved in Lemma 4.4 and Lemma 4.5, the next Lemma (Gain Lemma) shows that the lower bound is at least the minimum of the two gain functions at *any* η' in $(0, \eta_0)$, thus making

[7]That is, we require that $x \stackrel{\text{def}}{=} \frac{1}{4\epsilon} - \eta' > \frac{1}{2}$. Since the function $x(1-x)$ is monotone decreasing for $\frac{1}{2} < x < 1$, plugging some $\eta'' < \eta'$ into (5) will only make the LHS smaller thus changing the equality into an inequality.

it easier to obtain a closed form for $\textsc{gain}(\epsilon)$. Note that since F is monotone in $[0, \eta_0]$, the inverse of F exists in this interval and we denote it by F^{-1}.

Lemma 6.3 (Gain Lemma)**.** *Let* $1/4 \leq \epsilon < 1/2$. *For* all $\eta' \in (0, \eta_0)$, *let* $\delta' = F^{-1}(\eta')$, *then* $\textsc{gain}(\epsilon) \geq \min\{128(\epsilon\delta')^3, 4\epsilon(1-2\epsilon)^2\eta'\}$.

Proof. As before, we set $\delta = 4\epsilon\delta'$ and $\eta = 4\epsilon\eta'$. In the following, we consider ϵ to be any fixed value in $[\frac{1}{4}, \frac{1}{2})$. Suppose the minimum distance of C' is $(\frac{1}{2} - \delta')n'$. Then on one hand, there is an x_c, such that $x_c = -4\epsilon\delta' n = -\delta n$. On the other hand, by Lemma 6.2, there are at most $\frac{1}{(1-2\epsilon)^2} - 2$ codewords of weight at least $(\frac{1}{4\epsilon} - \eta')n'$ in C', which implies that there are at most $\frac{1}{(1-2\epsilon)^2} - 1$ x_c's such that $x_c \geq (1 - 2\epsilon - 4\epsilon\eta')n = (1 - 2\epsilon - \eta)n$ (recall that the codeword c_0 is of weight zero but satisfies $x_{c_0} = (1-2\epsilon)n$). Denote the gains as functions of η' given in Lemma 4.4 and Lemma 4.5 by $\textsc{gain}_\delta$ and $\textsc{gain}_\eta$, respectively. Then we have[8] $\textsc{gain}_\delta(\eta') = 2\delta^3 = 128\epsilon^3\delta'^3$ and $\textsc{gain}_\eta(\eta') = (1-2\epsilon)^2\eta = 4\epsilon(1-2\epsilon)^2\eta'$. Therefore $\textsc{gain}(\epsilon) \geq \min_{0<\eta'<\eta_0} \max\{\textsc{gain}_\delta(\eta'), \textsc{gain}_\eta(\eta')\}$. Because $G(\eta')$ is monotone decreasing in $[0, \eta_0]$, then δ' is monotone decreasing in $[0, \eta_0]$ by (6). Since $\textsc{gain}_\delta$ is monotone increasing in δ', it follows that $\textsc{gain}_\delta(\eta')$ is monotone decreasing in $[0, \eta_0]$. Also note that $\textsc{gain}_\eta(\eta')$ is monotone increasing in η'. Now at one end $\eta' = 0$, $\textsc{gain}_\delta(\eta') > 0$ and $\textsc{gain}_\eta(\eta') = 0$; at the other end $\eta' = \eta_0$, $\textsc{gain}_\delta(\eta') = 0$ and $\textsc{gain}_\eta(\eta') > 0$. Combining these facts we conclude that there exists an η'', $0 < \eta'' < \eta_0$, such that $\textsc{gain}_\delta(\eta'') = \textsc{gain}_\eta(\eta'') = \min_{0<\eta'<\eta_0} \max\{\textsc{gain}_\delta(\eta'), \textsc{gain}_\eta(\eta')\} \leq \textsc{gain}(\epsilon)$. By monotonicity of $\textsc{gain}_\delta(\eta')$ and $\textsc{gain}_\eta(\eta')$ again, $\textsc{gain}(\epsilon) \geq \textsc{gain}_\delta(\eta'') = \textsc{gain}_\eta(\eta'') \geq \min\{\textsc{gain}_\delta(\eta'), \textsc{gain}_\eta(\eta')\} = \min\{128(\epsilon\delta')^3, 4\epsilon(1-2\epsilon)^2\eta'\}$, for every $\eta' \in (0, \eta_0)$. □

In the following, our task is to derive an explicit bound for $\textsc{gain}(\epsilon)$. We begin with a simple lower bound for η_0.

Claim 6.4. *For all* $1/4 \leq \epsilon < 1/2$ *and let* η_0 *be as defined in the Trade-off Lemma (Lemma 6.2), then*

$$\eta_0 \geq \frac{(1-2\epsilon)^2}{2}.$$

Proof. By definition,

$$\eta_0 = \eta_2 = \frac{1}{2}\left(\frac{1}{2\epsilon} - 1 - \sqrt{\frac{(1-2\epsilon)^2}{1-2(1-2\epsilon)^2}}\right) = \frac{1-2\epsilon}{2}\left(\frac{1}{2\epsilon} - \frac{1}{\sqrt{1-2(1-2\epsilon)^2}}\right).$$

Now change the variable from ϵ to $y = 1-2\epsilon$ and apply Lemma 2.1, the desired bound follows. □

Set $\eta' = \gamma\frac{(1-2\epsilon)^2}{4}$, where $0 < \gamma \leq 1$ is a constant. Plugging η' into $G(\eta')$ and after some straightforward calculations, we get

$$G(\eta') = \frac{(1-2\epsilon)^2}{4}\left(\frac{\gamma^2}{4}(1-2\epsilon)^2 - \gamma\frac{1-2\epsilon}{2\epsilon} + \frac{1}{4\epsilon^2} - \frac{1}{1-2(1-2\epsilon)^2}\right).$$

[8] Here once again we focus on the worst case: If $\delta > \beta$, or equivalently, $\textsc{gain}_\delta(\eta') = 2\beta^3 = 2(1-2\epsilon)^3$, then the gain implied will be larger than that we are going to show.

By changing variable to $y = 1 - 2\epsilon$ and applying Lemma 2.2, we arrive at

$$\begin{aligned} G(\eta') &= \frac{y^2}{4}\left(\frac{\gamma^2}{4}y^2 - \gamma\frac{y}{1-y} + \frac{1}{(1-y)^2} - \frac{1}{1-2y^2}\right) \\ &\geq \frac{y^2}{4}\left(\frac{\gamma^2}{4}y^2 + (8-5\gamma)y^2\right) \\ &= 2(1-\Delta(\gamma))(1-2\epsilon)^4, \end{aligned}$$

where $\Delta(\gamma) \stackrel{\text{def}}{=} \frac{5\gamma}{8} - \frac{\gamma^2}{32}$. Therefore,

$$\delta' = \frac{2(1-2(1-2\epsilon)^2)}{1-3(1-2\epsilon)^2}G(\eta') \geq 2G(\eta') \geq 4(1-\Delta(\gamma))(1-2\epsilon)^4.$$

Plugging η' and δ' into Lemma 6.3, we get

$$\textsc{gain}(\epsilon) \geq \min\{8192(1-\Delta(\gamma))^3\epsilon^3(1-2\epsilon)^{12}, \gamma\epsilon(1-2\epsilon)^4\}.$$

This completes the proof of the Main Theorem.

Acknowledgment

N.X. is very grateful to Ronitt Rubinfeld for making his visit to MIT possible. We would like to thank Oded Goldreich, Ronitt Rubinfeld, Madhu Sudan and Luca Trevisan for encouragement, helpful discussions and valuable suggestions. We are also indebted to the anonymous referees (of both the preliminary conference version and the journal version of this work) for their extensive reports on errors and omissions, helpful comments which greatly improve the presentation of the paper, and suggestions that simplify several proof arguments.

References

[1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron. Testing low-degree polynomials over GF(2). In *Proceedings of RANDOM 2003*, pages 188–199, 2003.

[2] Y. Aumann, J. Håstad, M. Rabin, and M. Sudan. Linear-consistency testing. *Journal of Computer and System Sciences*, 62(4):589–607, 2001.

[3] L Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991. Earlier version in FOCS'90.

[4] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996. Earlier version in FOCS'95.

[5] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP and non-approximability - towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998. Earlier version in FOCS'95.

[6] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proc. 25th Annual ACM Symposium on the Theory of Computing*, pages 304–294, 1993.

[7] M. Bellare and M. Sudan. Improved non-approximability results. In *Proc. 26th Annual ACM Symposium on the Theory of Computing*, pages 184–193, 1994.

[8] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proc. 35th Annual ACM Symposium on the Theory of Computing*, pages 612–621, 2003.

[9] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993. Earlier version in STOC'90.

[10] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. *Journal of the ACM*, 43(2):268–292, 1996.

[11] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[12] O. Goldreich, S. Goldwaser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[13] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. Earlier version in STOC'97.

[14] J. Håstad and A. Wigderson. Simple analysis of graph tests for linearity and PCP. *Random Structures and Algorithms*, 22(2):139–160, 2003.

[15] C.S. Jutla, A.C. Patthak, A. Rudra, and D. Zuckerman. Testing low-degree polynomials over prime fields. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 423–432, 2004.

[16] T. Kaufman and S. Litsyn. Almost orthogonal linear codes are locally testable. In *Proc. 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 317–326, 2005.

[17] T. Kaufman and D. Ron. Testing polynomials over general fields. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 413–422, 2004.

[18] M. Kiwi. Algebraic testing and weight distributions of codes. *Theor. Comp. Sci.*, 299(1-3):81–106, 2003. Earlier version appeared as ECCC TR97-010, 1997.

[19] M. Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6:445–450, 1960.

[20] D. Ron. Property testing (a tutorial). In P.M. Pardalos, S. Rajasekaran, J. Reif, and J.D.P. Rolim, editors, *Handbook of Randomized Computing*, pages 597–649. Kluwer Academic Publishers, 2001.

[21] R. Rubinfeld. Sublinear time algorithms. In *Proceedings of the International Congress of Mathematicians (ICM)*, pages 1095–1110, 2006.

[22] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25:252–271, 1996.

[23] A. Samorodnitsky. Low-degree tests at large distances. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 506–515, 2007.

[24] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proc. 32nd Annual ACM Symposium on the Theory of Computing*, pages 191–199, 2000.

[25] A. Samorodnitsky and L. Trevisan. Gower uniformity, influence of variables and PCPs. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pages 11–20, 2006.

[26] A. Shpilka and A. Wigderson. Derandomizing homomorphism testing in general groups. *SIAM Journal on Computing*, 36(4):1215–1230, 2006. Earlier version in STOC'2004.

[27] M. Sudan and L. Trevisan. Probabilistically checkable proofs with low amortized query complexity. In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 18–27, 1998.

[28] L. Trevisan. Recycling queries in PCPs and linearity tests. In *Proc. 30th Annual ACM Symposium on the Theory of Computing*, pages 299–308, 1998.

A Proofs of Lemma 2.1 and Lemma 2.2

Lemma 2.1. *For all real y with $0 \leq y \leq 1/2$,*

$$\frac{1}{1-y} - y \geq \frac{1}{\sqrt{1-2y^2}}.$$

The following proof is simpler than our original one and was suggested by an anonymous referee.

Proof. Since $\frac{1}{1-y} - y = \frac{1-y+y^2}{1-y}$ it suffices to show that $\sqrt{1-2y^2} \geq \frac{1-y}{1-y+y^2}$, or equivalently that $(1-y+y^2)^2(1-2y^2) \geq (1-y)^2$. However,

$$\begin{aligned}
&(1-y+y^2)^2(1-2y^2) \\
&= (1-y+y^2)^2 - 2y^2(1-y+y^2)^2 \\
&= (1-y)^2 + 2y^2(1-y) + y^4 - 2y^2(1-y+y^2)^2.
\end{aligned}$$

To conclude, it suffices to show that

$$g(y) \stackrel{\text{def}}{=} 2(1-y) + y^2 - 2(1-y+y^2)^2 \geq 0.$$

Some simple calculations yield that $g(y) = y(2-5y+4y^2-2y^3) = y\left((2-y)(1-2y) + 2y^2(1-y)\right)$. The desired conclusion now follows since both $y(2-y)(1-2y)$ and $2y^3(1-y)$ are non-negative for all $0 \leq y \leq 1/2$. □

Lemma 2.2. *Let γ be a constant with $0 \leq \gamma \leq 1$. Then for all real y with $0 \leq y \leq 1/2$,*

$$\frac{1}{(1-y)^2} - \frac{1}{1-2y^2} - \gamma\frac{y}{1-y} \geq (8-5\gamma)y^2.$$

Proof. We break the proof into two parts: First we show that the inequality holds for $0 \leq y \leq 2/7$, then we prove it for the interval $2/7 \leq y \leq 1/2$.

Proposition A.1. *For all y and γ with $0 \le y \le \frac{2}{7}$ and $0 \le \gamma \le 1$,*

$$\frac{1}{(1-y)^2} - \frac{1}{1-2y^2} - \gamma\frac{y}{1-y} \ge (8-5\gamma)y^2.$$

Proof. By Taylor expansion,

$$\begin{aligned}
&\frac{1}{(1-y)^2} - \frac{1}{1-2y^2} - \gamma\frac{y}{1-y} - (8-5\gamma)y^2 \\
&= \sum_{k=0}^{\infty}(k+1)y^k - \sum_{k=0}^{\infty}(2y^2)^k - \gamma\sum_{k=1}^{\infty}y^k - (8-5\gamma)y^2 \\
&= (2-\gamma)y - (7-4\gamma)y^2 + (4-\gamma)y^3 + (1-\gamma)y^4 + \sum_{k=5}^{\infty}(k+1-\gamma)y^k - \sum_{k=3}^{\infty}(2y^2)^k \\
&\ge (2-\gamma)y - (7-4\gamma)y^2 + (4-\gamma)y^3 + (1-\gamma)y^4 - \frac{8y^6}{1-2y^2} \\
&\ge (2-\gamma)y - (7-4\gamma)y^2 + 3y^3 - \frac{8y^6}{1-2y^2}.
\end{aligned}$$

Since $0 \le y \le \frac{2}{7}$, $(2-\gamma)y \ge \frac{7}{2}(2-\gamma)y^2 = (7-\frac{7}{2}\gamma)y^2 \ge (7-4\gamma)y^2$, $\frac{8y^6}{1-2y^2} \le \frac{8y^6}{1-2(\frac{2}{7})^2} \le 10y^6$, and $3y^3 \ge 3(\frac{7}{2})^3y^6 \ge 10y^6$, this completes the proof of the Proposition. □

Proposition A.2. *For all y and γ with $\frac{2}{7} \le y \le \frac{1}{2}$ and $0 \le \gamma \le 1$,*

$$\frac{1}{(1-y)^2} - \frac{1}{1-2y^2} - \gamma\frac{y}{1-y} \ge (8-5\gamma)y^2.$$

Proof. Let $z = 1-2y$. After substituting z into the expression and some simplification, we see that proving the original inequality is equivalent to proving, for $0 \le z \le \frac{3}{7}$,

$$\frac{4}{(1+z)^2} - \frac{2}{2-(1-z)^2} - \gamma\frac{1-z}{1+z} \ge (2-\frac{5}{4}\gamma)(1-z)^2.$$

Or, after dividing $(1-z)^2$ on both sides,

$$\frac{4}{(1-z^2)^2} - \frac{2}{2(1-z)^2-(1-z)^4} - \gamma\frac{1}{1-z^2} \ge 2 - \frac{5}{4}\gamma.$$

Note that since $0 \le z \le \frac{3}{7}$, we have $\frac{\gamma}{1-z^2} \le \frac{\gamma}{1-(\frac{3}{7})^2} = \frac{49}{40}\gamma \le \frac{5}{4}\gamma$, so the only thing that remains to show is that $\frac{4}{(1-z^2)^2} - \frac{2}{2(1-z)^2-(1-z)^4} \ge 2$. Indeed,

$$\begin{aligned}
&\frac{4}{(1-z^2)^2} - \frac{2}{2(1-z)^2-(1-z)^4} \ge 2 \\
\iff &\frac{4}{(1-z^2)^2} - \frac{2}{1-(2z-z^2)^2} \ge 2 \\
\iff &\frac{2}{(1-z^2)^2} \ge \frac{2-z^2(2-z)^2}{1-z^2(2-z)^2} \\
\iff &2(1-z^2(2-z)^2) \ge (1-z^2)^2(2-z^2(2-z)^2) \\
\iff &2(1+2z-z^2) \ge 2(1+z)^2 - z^2(1+z)^2(2-z)^2 \\
\iff &(1+z)^2(2-z)^2 \ge 4 \\
\iff &z(1-z) \ge 0.
\end{aligned}$$

This finishes the proof of the Proposition. □

Now combining Proposition A.1 and Proposition A.2 together completes the proof of Lemma 2.2. □