Linearity Testing in Characteristic Two[*]

M. BELLARE [†] D. COPPERSMITH [‡] J. HÅSTAD [§] M. KIWI [¶] M. SUDAN [||]

Abstract

Let $\mathsf{Dist}(f,g) = \Pr_u [\, f(u) \neq g(u) \,]$ denote the relative distance between functions f, g mapping from a group G to a group H, and let $\mathsf{Dist}(f)$ denote the minimum, over all linear functions (homomorphisms) g, of $\mathsf{Dist}(f,g)$. Given a function $f\colon G \to H$ we let $\mathsf{Err}(f) = \Pr_{u,v} [\, f(u)+f(v) \neq f(u+v) \,]$ denote the rejection probability of the BLR (Blum-Luby-Rubinfeld) linearity test. *Linearity testing* is the study of the relationship between $\mathsf{Err}(f)$ and $\mathsf{Dist}(f)$, and in particular the study of lower bounds on $\mathsf{Err}(f)$ in terms of $\mathsf{Dist}(f)$.

The case we are interested in is when the underlying groups are $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. In this case the collection of linear functions describe a Hadamard code of block length 2^n and for an arbitrary function f mapping $\mathrm{GF}(2)^n$ to $\mathrm{GF}(2)$ the distance $\mathsf{Dist}(f)$ measures its distance to a Hadamard code (normalized so as to be a real number between 0 and 1). The quantity $\mathsf{Err}(f)$ is a parameter that is "easy to measure" and linearity testing studies the relationship of this parameter to the distance of f.

The code and corresponding test are used in the construction of efficient probabilistically checkable proofs and thence in the derivation of hardness of approximation results. In this context, improved analyses translate into better non-approximability results. However, while several analyses of the relation of $\mathsf{Err}(f)$ to $\mathsf{Dist}(f)$ are known, none is tight.

We present a description of the relationship between $\mathsf{Err}(f)$ and $\mathsf{Dist}(f)$ which is nearly complete in all its aspects, and entirely complete (i.e. tight) in some. In particular we present functions $L, U\colon [0,1] \to [0,1]$ such that for all $x \in [0,1]$ we have $L(x) \leq \mathsf{Err}(f) \leq U(x)$ whenever $\mathsf{Dist}(f) = x$, with the upper bound being tight on the whole range, and the lower bound tight on a large part of the range and close on the rest.

Part of our strengthening is obtained by showing a new connection between the linearity testing problem and Fourier analysis, a connection which may be of independent interest. Our results are used by Bellare, Goldreich and Sudan to present the best known hardness results for Max-3SAT and other MaxSNP problems [7].

Index Terms — Probabilistically checkable proofs, approximation, program testing, Hadamard codes, error detection, linearity testing, MaxSNP.

---

[*] A preliminary version of this paper appeared in Proceedings *of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995.

[†] Department of Computer Science & Engineering, Mail Code 0114, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093. `mihir@cs.ucsd.edu`.

[‡] Research Division, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA. `copper@watson.ibm.com`.

[§] Department of Computer Science, Royal Institute of Technology, 10044 Stockholm, Sweden. `johanh@nada.kth.se`. Part of this work was done while the author was visiting MIT.

[¶] Depto. de Ingeniería Matemática, Fac. de Cs. Físicas y Matemáticas, U. de Chile. `mkiwi@dim.uchile.cl`. This work was done while the author was at the Dept. of Applied Mathematics, Massachusetts Institute of Technology. Supported by AT&T Bell Laboratories PhD Scholarship, NSF Grant CCR–9503322, FONDECYT grant No. 1960849, and Fundación Andes.

[||] Research Division, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA. `madhu@watson.ibm.com`.

1 Introduction

One of the contributions of computational complexity theory has been to re-examine the classical notion of what constitutes a proof of a mathematical statement. The complexity class NP introduced the notion of an *efficiently verifiable* proof. It asks that the proof, which is a sequence of written symbols, not only be verifiable, but be verifiable quickly, namely in polynomial time. Over the last decade or so, researchers have furthered this avenue in many ways. One of the many notions that has been developed, and explored, allows the verifier (of the claimed proof) to be probabilistic in its actions. The new verifier is also allowed to err in its judgment, as long as it doesn't do so too often— proofs of false statements can be accepted with small probability. (This probability is measured over coin tosses made by the verifier and not over any distribution over theorems/proofs.) As a tradeoff, the notion restricts the access of the verifier into the proof, allowing a verifier to only *query* or probe the proof in a small number of bits and studies the behavior of the number of bits that are needed to be probed in any proof system as a function of the error probability. Such a proof system, i.e., the verifier and its associated format for valid proofs, is referred to as a probabilistically checkable proof system — PCP, for short. Along with the development of this notion, the research has also yielded a series of technical developments, which have constructed PCP verifiers which examine only a constant number of bits, C, of a purported proof and reject proofs of incorrect statements with probability $\frac{1}{2}$. This constant is a universal constant, and independent of the length of the theorem or the proof. The new proof systems, do require valid proofs to be longer than traditional (deterministic) proof systems would allow for. However the size of the new proofs are only polynomially larger than the size of the traditional proofs.

Apart from the inherent interest in the construction and performance of PCP systems, a major motivating factor for the study of PCP systems is their use in the derivation of non-approximability results for combinatorial optimization problems. The theory of NP-completeness has been employed as an important tool in the analysis of the complexity of finding *optimal* solutions to discrete (or combinatorial) optimization problems. For many optimization problems, the NP-complete or NP-hard ones, this theory can be used to show that no polynomial time solution solves this problem *exactly*, unless NP = P. However the possibility that solutions to these problems which approximate the optimum to within a relative error of ϵ for every $\epsilon > 0$ may be found in polynomial time, remained open. A new connection uses the PCP constructions mentioned above to show that for many interesting problems, even such approximate solutions can not be found in polynomial time unless NP = P. This connection further serves to motivate the study of PCP systems and in particular, their efficiency (for instance, the parameter C above) since improved efficiency translates into stronger non-approximability results.

The prime motivation for the problem to be studied in this paper is these PCP constructions and the ensuing hardness of approximation results. However, a full explanation of the details of these results is beyond the scope of this paper — in fact, we will not even attempt to formalize the definitions above. The interested reader is directed towards any of a number of surveys which have appeared on this topic.[1] Fortunately, the problem to be studied in this paper can be formulated cleanly without reference to the above mentioned results and furthermore have an interesting implication in a coding theoretic setting. We first describe this setting and then procedd to formally define the problem of interest.

Central to many of the construction of efficient PCPs has been the construction and analysis

---

[1] It isn't possible to provide an exhaustive list of the dozen or so surveys available but if you are on the web check out `http://www-cse.ucsd.edu/users/mihir/pcp.html`.

of error-correcting codes and probabilistic "error-detection" algorithms for these error-correcting codes. These algorithms function as follows: Given a word w which is supposed to be a codeword of some error-correcting code, the algorithm probabilistically chooses a small (sometimes constant) number of bits of the word w to examine, computes a (simple) boolean function of these bits and outputs a verdict ACCEPT/REJECT. The guarantee obtainable from such algorithms is weaker than the guarantee expected ¿from classical error-detection algorithms. In particular, the guarantees behave as follows: Given a valid codeword, the algorithm must output ACCEPT with probability 1. On the other hand if the input is far from any valid codeword (i.e., the distance is more than some specified constant fraction of the minimum distance of the code), then the algorithm must output REJECT with some positive probability, bounded away from 0. Most of the codes used in these constructions are well-known ones, with Hadamard Codes and variants of Reed-Solomon being the most commonly used ones. Much of the technical development in this area is directed towards the construction and analysis of the probabilistic error-correcting algorithms. This area of study, collectively referred to as *testing* in the PCP literature is the origin of the problem considered in this paper.

It is a feature of the area that while tests are easy to specify, they are notoriously hard to analyze, especially to analyze well. Yet, good analyses are, for several reasons, worth striving for. There is, first, the inherent mathematical interest of getting the best possible analysis and understanding of a well-defined combinatorial problem. But, there is a more pragmatic reason: better analyses typically translate into improved (increased) factors shown non-approximable in hardness of approximation results.

The specific problem considered here is called the *linearity testing* problem. We wish to look at a particular test, called the BLR test, with was the first ever proposed. Our focus is the case of most importance in applications, when the underlying function maps between groups of characteristic two. Several analyses have appeared, yet none is tight. Each improved analysis implies improved factors shown non-approximable in hardness of approximation results.

Let us begin by describing the linearity testing problem and past work more precisely.

1.1 The Problem

The linearity testing problem is a problem related to homomorphisms between groups. Let G, H be finite groups. A function $g\colon G \to H$ is said to be *linear* if $g(u) + g(v) = g(u{+}v)$ for all $u, v \in G$. (That is, g is a group homomorphism.) We will use the notation $u \overset{R}{\leftarrow} G$ to represent a random variable u chosen uniformly at random from the (finite) group G. Here are some basic definitions:

- $\textsc{Lin}(G, H)$ — Set of all linear functions of G to H
- $\mathsf{Dist}(f, g) \overset{\text{def}}{=} \Pr_{u \overset{R}{\leftarrow} G}[\, f(u) \neq g(u)\,]$ — (relative) distance between $f, g\colon G \to H$
- $\mathsf{Dist}(f) \overset{\text{def}}{=} \min\{\, \mathsf{Dist}(f, g) : g \in \textsc{Lin}(G, H)\,\}$ — Distance of f to its closest linear function.

The BLR Test. Blum, Luby and Rubinfeld [9] suggest a probabilistic method to "test" if a function f is really a *linear* function. This test, henceforth referred to as the BLR test, is the following [9]— Given a function $f\colon G \to H$, pick $u, v \in G$ at random and reject if $f(u) + f(v) \neq f(u{+}v)$. Let

$$\mathsf{Err}(f) \overset{\text{def}}{=} \Pr_{u,v \overset{R}{\leftarrow} G}[\, f(u) + f(v) \neq f(u{+}v)\,]$$

denote the probability that the BLR test rejects f. The issue in linearity testing is to study how $\mathsf{Err}(f)$ behaves as a function of $x = \mathsf{Dist}(f)$. In particular, one would like to derive good lower bounds on $\mathsf{Err}(f)$ as a function of x.

$\textsc{Rej}(\cdot)$. A convenient way to capture the above issues is via the *rejection probability function* $\textsc{Rej}_{G,H}\colon [0,1] \to [0,1]$ of the test. It associates to any number x the minimum value of $\mathsf{Err}(f)$, taken over all functions f of distance x from the space of linear functions. Thus,

$$\textsc{Rej}_{G,H}(x) \quad \stackrel{\text{def}}{=} \quad \min\{\, \mathsf{Err}(f) \,:\, f\colon G \to H \text{ s.t. } \mathsf{Dist}(f) = x \,\} \, .$$

The graph of $\textsc{Rej}_{G,H}$ —namely $\textsc{Rej}_{G,H}(x)$ plotted as a function of x— is called the *linearity testing curve.*[2] This curve depends only on the groups G, H.

By definition it follows that $\textsc{Rej}_{G,H}(x) > 0$ if $x > 0$. However it is not easy to see if any other quantitative statements can be made about $\textsc{Rej}_{G,H}(x) > 0$ for larger values of x. The most general problem in linearity testing is to determine the function $\textsc{Rej}_{G,H}(\cdot)$ for given G, H. Much of the work that has been done provides information about various aspects of this function.

THE KNEE OF THE CURVE. At first glance, it may be tempting to believe that $\textsc{Rej}_{G,H}(\cdot)$ will be a monotone non-decreasing function. One of the most surprising features of $\textsc{Rej}_{G,H}$ is that this is not necessarily true. It turns out (and we will see such an example presently) that there exist groups G, H such that $\textsc{Rej}_{G,H}(\frac{1}{4}) \geq \frac{3}{8}$, but $\textsc{Rej}_{G,H}(\frac{2}{3}) = \frac{2}{9}$. The threshold of $x = \frac{1}{4}$ turns out to be significant in this example and an important parameter that emerges in the study of linearity testing is how low $\textsc{Rej}_{G,H}(x)$ can be for $x \geq \frac{1}{4}$. In this paper we call this parameter, identified in [2, 6, 7, 8], the *knee* of the curve. Formally:

$$\textsc{Knee}_{G,H} \quad \stackrel{\text{def}}{=} \quad \min\{\, \textsc{Rej}(x) \,:\, x \geq \tfrac{1}{4} \,\} \, .$$

1.2 Error detection in Hadamard codes

In this paper we look at the performance of the BLR test when the underlying groups are $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$ for some positive integer n. For notational simplicity we now drop the groups G, H from the subscripts, writing $\textsc{Rej}(x)$ and $\textsc{Knee}$— it is to be understood that we mean $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$.

This special case is of interest because of the following reason: In this case the family of functions $\textsc{Lin}(\mathrm{GF}(2)^n, \mathrm{GF}(2))$ actually defines a Hadamard code of block length 2^n. Notice that every linear function l is specified by a vector α from $\mathrm{GF}(2)^n$ such that $l(x) = \langle \alpha, x \rangle$ (where $\langle \alpha, x \rangle = \sum_{i=1}^n \alpha_i x_i$ denotes the inner product of vectors α, x). Thus we can associate with each of the 2^n linear functions l, a codeword which is the 2^n bit sequence $(l(x) \,:\, x \in \mathrm{GF}(2)^n)$. Any two distinct codewords differ in exactly 2^{n-1} positions, making this a $(2^n, 2^n, 2^{n-1})$-code. For further details see MacWilliams and Sloane [18, pages 48–49].

For an arbitrary function f, the parameter $\mathsf{Dist}(f)$ simply measures its distance to the above mentioned Hadamard code, normalized by 2^n. Estimating $\mathsf{Dist}(f)$ is thus related to the classical task of error-detection. The parameter $\mathsf{Err}(f)$ on the other hand simply defines a quantity that can be estimated to fairly good accuracy by a probabilistic algorithm, which probes f in a few places (or reads a few bits of the purported codeword). The algorithm repeats the following step several times: It picks random $x, y \in \mathrm{GF}(2)^n$ and tests to see if $f(x) + f(y) = f(x+y)$. At the end it reports the average number of times this test fails. It can be verified easily that this provides an

---

[2] Actually the function $\textsc{Rej}_{G,H}(x)$ is only defined for finitely many values, namely the integral multiples of $\frac{1}{|G|}$, and undefined for infinitely many values. Thus the linearity testing curve is not really a curve in the real plane, but simply describes a function of finitely many points.

estimate on $\mathsf{Err}(f)$, and the accuracy of this estimate improves with the number of iterations. The advantage of this algorithm is that it probes f in very few places in order to compute its output (in particular the number of probes can be independent of n). The aim of Linearity Testing is to turn this estimate on $\mathsf{Err}(f)$ into an estimate on $\mathsf{Dist}(f)$. This would thus yield an algorithm which probes f in few places and yet yields some reasonable estimates on $\mathsf{Dist}(f)$, and in particular solves the earlier mentioned probabilistic error-detection task. This is the ingredient which makes this test useful in the applications to PCPs and motivates our study.

1.3 Previous work

The first investigation of the shape of the linearity testing curve, by Blum, Luby and Rubinfeld [9], was in the general context where G, H are arbitrary finite groups. Their analysis showed that $\mathrm{REJ}_{G,H}(x) \geq \frac{2}{9}\,x$ [9]. (They indicate that this is an improvement of their original analysis obtained jointly with Coppersmith.) Interest in the tightness of the analysis begins with Bellare, Goldwasser, Lund and Russell [6] in the context of improving the performance of PCP systems. They showed that $\mathrm{REJ}_{G,H}(x) \geq 3\,x - 6\,x^2$. It turns out that, with very little effort, the result of [9] can be used to show that $\mathrm{REJ}_{G,H}(x) \geq \frac{2}{9}$ for $x \geq \frac{1}{3}$. This claim appears in Bellare and Sudan [8], without proof. A proof is included in the appendix of this paper, for the sake of completeness. Of the three bounds above, the last two bounds supercede the first, so that the following theorem captures the state of knowledge.

Theorem 1.1 [6, 9, 10] Let G, H be arbitrary finite groups. Then:
(1) $\mathrm{REJ}_{G,H}(x) \geq 3x - 6x^2$.
(2) $\mathrm{KNEE}_{G,H} \geq \frac{2}{9}$.

As indicated above, an improved lower bound for the knee would lead to better PCP systems. But in this general setting, we can do no better. The following example of Coppersmith [10] shows that the above value is in fact tight in the case of general groups. Let m be divisible by three. Let f be a function from $\mathcal{Z}_m^n$ to $\mathcal{Z}_m$ such that $f(u) = 3k$, if $u_1 \in \{3k-1, 3k, 3k+1\}$. Then, $\mathsf{Dist}(f) = \frac{2}{3}$. Furthermore, $f(u) + f(v) \neq f(u{+}v)$ only if $u_1 = v_1 = 1 \pmod 3$, or $u_1 = v_1 = -1 \pmod 3$, i.e. $\mathsf{Err}(f) = \frac{2}{9}$.

This leads into our research. We note that the problem to which linearity testing is applied in the proof system constructions of [2, 6, 7, 8] is that of testing Hadamard codes (in the first three works) and the long code (in the last work). But this corresponds to the above problem in the special case where $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. (G is regarded as an additive group in the obvious way. Namely, the elements are viewed as n-bit strings or vectors over $\mathrm{GF}(2)$, and operations are component-wise over $\mathrm{GF}(2)$.) For this case, the example of Coppersmith does *not* apply, and we can hope for better results.

1.4 New results and techniques

As pointed out earlier we focus on the case where the domain and range are of characteristic two and in particular $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. We provide two new analyses of $\mathrm{REJ}(x)$ in this case.

FOURIER ANALYSIS. We establish a new connection between linearity testing and Fourier analysis. We provide an interpretation of $\mathsf{Dist}(f)$ and $\mathsf{Err}(f)$ in terms of the Fourier coefficients of an appropriate transformation of f. We use this to cast the linearity testing problem in the language of Fourier series. This enables us to use Fourier analysis to study the BLR test. The outcome is the following:

Theorem 1.2 For every real number $x \leq \frac{1}{2}$, $\textsc{Rej}(x) \geq x$.

Apart from lending a new perspective to the linearity testing problem, the result exhibits a feature which distinguishes it from all previous results. Namely, it shows that $\textsc{Rej}(x) \to \frac{1}{2}$ as $x \to \frac{1}{2}$.[3] (According to the previous analysis, namely Theorem 1.1, $\textsc{Rej}(x)$ may have been bounded above by $\frac{2}{9}$ for all $x \geq \alpha$, where α is the larger root of the equation $3z - 6z^2 = \frac{2}{9}$.) Furthermore we can show that the analysis is tight (to within $o(1)$ factors) at $x = \frac{1}{2} - o(1)$.

This result can also be combined with Part (1) of Theorem 1.1 to show that $\textsc{Knee} \geq \frac{1}{3}$. However this is not tight. So we focus next on finding the right value of the knee.

Combinatorial analysis. The analysis to find the knee is based on combinatorial techniques. It leads us to an isoperimetric problem about a 3-regular hypergraph on the vertices of the n-dimensional hypercube. We state and prove a Summation Lemma which provides a tight isoperimetric inequality for this problem. We then use it to provide the following exact value of the knee of $\textsc{Rej}(x)$.

Theorem 1.3 $\textsc{Knee} = \frac{45}{128}$.

Tightness of the analysis. We provide examples to indicate that, besides the knee value, the lower bounds on $\textsc{Rej}(x)$ as indicated by our and previous results are tight for a number of points. In particular, the curve is tight for $x \leq \frac{5}{16}$, and the bound at $x = \frac{1}{2} - o(1)$ is matched up to within $o(1)$ factors (i.e., there exist functions $f_n : \mathrm{GF}(2)^n \to \mathrm{GF}(2)$ such that as n goes to ∞, $\mathsf{Err}(f_n)$ and $\mathsf{Dist}(f_n)$ go to $\frac{1}{2}$).

Other results. The isoperimetric inequality underlying Theorem 1.3 turns out to reveal other facts about $\textsc{Rej}(x)$ as well. In particular it helps establish a tight *upper bound* on $\mathsf{Err}(f)$ as a function of $\mathsf{Dist}(f)$. This result is presented in Section 3.

Also, while the main focus of this paper has been the BLR test, we also present in Section 5 a more general result about testing for total degree one in characteristic two. The purpose is to further illustrate the strength and elegance of the Fourier analysis technique, as well as its more general applicability to the problem of analyzing program testers.

Graph. Figure 1 summarizes the results of this work. The points $\{ (\mathsf{Dist}(f), \mathsf{Err}(f)) : f \}$ lie in the white region of the first graph. The dark shaded region represents the forbidden area before our work, and the light shaded region represents what we add to the forbidden area. Note we both extend the lower bound and provide upper bounds. The dots are actual computer constructed examples; they indicate that perhaps the lower bound may be improved, but not by much.[4] In particular, the knee value is tight. Furthermore the upper bound is tight.

The second graph indicates lower bounds on $\textsc{Rej}(x)$. The line $\frac{2}{9}x$ represents the result of [9]. The parabola is the curve $3x - 6x^2$ representing the result of [6]. The curve $\frac{2}{3}x$ when $x \leq \frac{1}{3}$ and $\frac{2}{9}$ when $x > \frac{1}{3}$ represents the result of [8]. Our additions are the 45 degree line of x and the horizontal line at $\frac{45}{128}$ for the new knee value.

---

[3] Note that $\mathsf{Dist}(f) \leq \frac{1}{2}$ for all $f\colon G \to H$ because we are working over GF(2), so only the portion $x \in [0, \frac{1}{2}]$ of the curve is interesting.

[4] More precisely, we have a randomized procedure that with high probability can construct, for each plotted point, a function f such that $(\mathsf{Dist}(f), \mathsf{Err}(f))$ is arbitrarily close to the point in question.

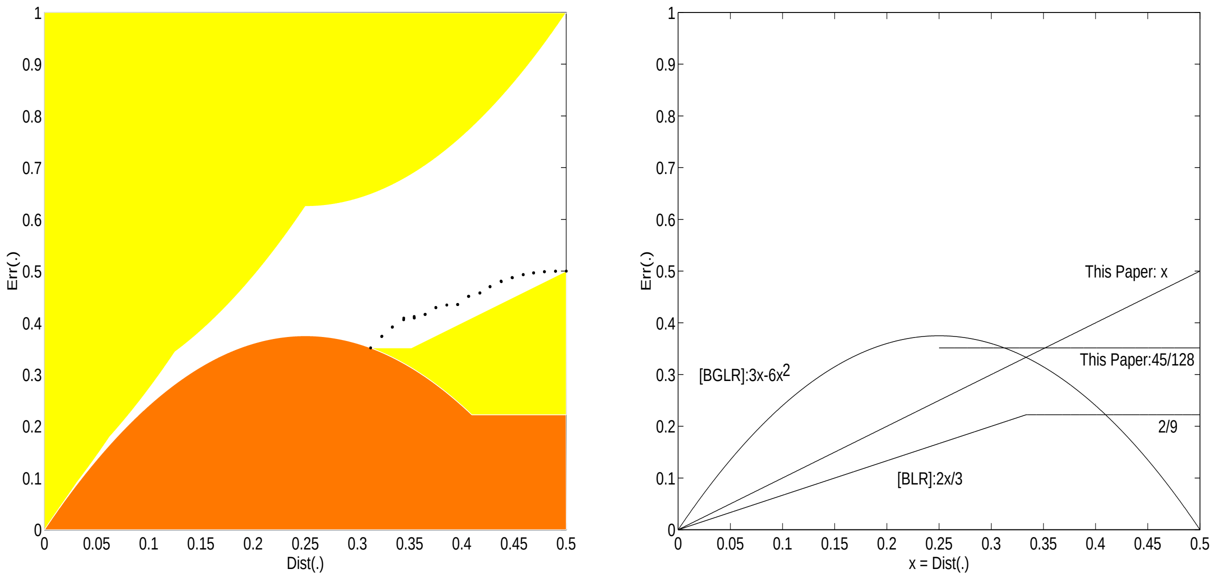

Figure 1: The points $(\mathsf{Dist}(f), \mathsf{Err}(f))$ in the plane, and the successive lower bounds. See text for discussion.

1.5 Application to MaxSNP hardness

As mentioned earlier, the construction of PCP systems have led to new results showing the non-approximability of many combinatorial optimization problems. This surprising connection, initiated by Feige, Goldwasser, Lovász, Safra and Szegedy [11], showed how to turn the results on constructions of efficient PCP systems into results which showed that for certain combinatorial optimization problems finding an ϵ approximate solution is also an NP-hard task. A subsequent result, due to Arora, Lund, Motwani, Sudan and Szegedy [2] managed to use a similar idea to show that an analogous result holds for a large collection of problems called MaxSNP hard problems. The result says that for every MaxSNP hard problem, there is a constant $\epsilon > 0$, such that the task of finding solutions which approximate the optimum to within a relative error of ϵ for this problem, is also NP-hard. Subsequently, initiated by the work of Bellare, Goldwasser, Lund and Russell [6], a series of works have improved the above results by constructing more efficient PCP systems and thereby showing stronger hardness of approximation results for MaxSNP hard problems.

Usage of the linearity test in the construction of efficient PCPs, and thence in the derivation of hardness of approximability results for MaxSNP problems, begins in [2] and continues in [6, 8, 7]. In the first three cases, it is used to test the Hadamard code; in the last case, to test a different code called the long code. In all cases the underlying problem is the one we have considered above, namely linearity testing with $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$.

The MaxSNP hardness result of [6] used only two things: The lower bound $\textsc{Rej}(x) \geq 3x - 6x^2$ of Theorem 1.1, and the best available lower bound k on the knee. They were able to express the non-approximability factor for Max-3SAT as an increasing function $g_1(k)$ depending solely on k. The lower bound on the knee that they used was $\textsc{Knee} \geq \frac{1}{6}$ derived from Part (1) of Theorem 1.1 and [9]. Their final result was that approximating Max-3SAT within $\frac{113}{112} \approx 1.009$ is NP-hard.

Improved proof systems were built by [8]. Again, their non-approximability factor had the form

$g_2(k)$ for some function g_2 depending only on the best available lower bound k on the knee. They used KNEE $\geq \frac{2}{9}$ to show that approximating Max-3SAT within $\frac{74}{73} \approx 1.014$ is NP-hard. Theorem 1.3 would yield direct improvements to the results of [6, 8] with no change in the underlying proof systems or construction. However, better proof systems are now known, namely the long code based ones of [7]. The analysis in the latter uses *both* our results (namely Theorem 1.3 and Theorem 1.2). They show that approximating Max-3SAT within 1.038 is NP-hard. They also exploit our analyses to derive strong non-approximability results for other MaxSNP problems (like Max-2SAT and Max-Cut) and for Vertex Cover.

Thus, the applications of [6, 8] motivated our consideration of the linearity testing problem. In the process we proved more than these works needed. But, interestingly, later [7] found our results useful in the same context.

1.6 Relationship to other work

As mentioned earlier, there are a variety of problems which are studied under the label of *testing*. In particular, a variety of tasks address the issue of testing variants of Reed-Solomon codes. These tests, referred to in the literature as *low-degree tests* are used in a variety of ways in proof systems. We briefly explain, first, what are the other problems and results in low degree testing and why they differ from ours; second how the usage of these in proof systems is different from the usage of linearity tests.

LOW DEGREE TESTING. We are given a function $f\colon F^n \to F$, where F is a field, and we are given a positive integer d. In the low *individual degree* testing problem we are asked to determine whether f is close to some polynomial p of degree d in each of its n variables. When specialized to the case of $d = 1$, this task is referred to as *multi-linearity testing*. In the low *total degree* testing problem we are asked to determine whether f is close to some polynomial p of total degree d in its n variables.[5] Multi-linearity tests were studied by [4, 11]. Low individual degree tests were studied by [3, 5, 12, 19]. Total degree tests were studied by [2, 13, 14, 20].

What we are looking at, namely linearity testing over GF(2), is a variant of the total degree testing problem in which the degree is $d = 1$, F is set to GF(2), and the constant term of the polynomial p is forced to 0. Even though a significant amount of work has been put into the analysis of the low degree tests by the above mentioned works, the analysis does not appear to be tight for any case. In particular one cannot use those results to derive the results we obtain here. In fact the tightness of the result obtained here raises the hope that similar techniques can be used to improve the analysis in the above testers.

THE ROLE OF TESTING IN PCP SYSTEMS. An important tool in the construction of proof systems is a tool referred to as *recursion* [3]. Roughly, the tool provides an analog of the process of construction of *concatenated* error-correcting codes, to the realm of PCPs. A PCP proof system constructed by recursion consists of several levels of different atomic PCPs. The PCP at each level of recursion typically uses some form of low-degree testing, the kind differing from level to level.

The use of multi-linearity testing was initiated by Babai, Fortnow and Lund [4]. For efficiency reasons, researchers beginning with Babai, Fortnow, Levin and Szegedy [5] then turned to low individual degree testing. This testing is used in the "higher" levels of the recursion. Linearity testing showed up for the first time in the lowest level of the recursion, in the checking of the

---

[5] To illustrate the difference between individual and total degree, note that $f(x_1, \ldots, x_n) = x_1 x_2$ is multi-linear but not linear.

Hadamard code in [2]. The proof systems of [7] use all these different testers, but, as we explained, the final non-approximability factors obtained can be expressed only in terms of the shape of the linearity testing curve.

RECENT WORK. Kiwi [16] provides improved analysis for the linearity testing problem over all finite fields. He obtains this result by providing another new interpretation of the linearity testing problem, this time by relating it to a weight enumeration problem of a linear code studied as a function of the minimum distance of its dual code.

Håstad [15] has shown a tester for a different code, namely the "long code" of [7], and an analysis for the test is again based on a Fourier Transform based approach. The analysis once again provides significant improvements to non-approximability results for the clique problem.

1.7 Discussion

The main argument behind the analysis of the BLR test given in [9] is the following: given f taking values from one finite group G into another finite group, start by defining a function g_f whose value at u is $\textsc{Plurality}\{\, f(u+v) - f(v) \,:\, v \in G \,\}$.[6] Then, show that if $\mathsf{Err}(f)$ is sufficiently small, three things happen. First, an overwhelming majority of the values $\{\, f(u+v) - f(v) \,:\, v \in G \,\}$ agree with $g_f(u)$, second, g_f is linear, and last, g_f is close to f. This argument is called the plurality argument. The assumption that the rejection probability of the test is small seems to be an essential component of this argument.

The arguments used in most of the previous works on low-degree testing are based on the plurality argument. So far, these type of arguments have been unable to show a non-trivial relation between the probability that a given function fails a test, and its distance from a family of low-degree polynomials, when the probability that the test fails is high (i.e., larger than $\frac{1}{2}$). Our discrete Fourier analysis approach does not exhibit the properties discussed above, and this may be one of the reasons for its success.

Our approach was somewhat inspired by the coding theoretic statement of the linearity testing problem; however the final analysis does not bring this out clearly. Kiwi's [16] approach brings the connection out much more explicitly and suggests that further exploration of the relationship to coding theory may prove fruitful.

2 Fourier Analysis of the Linearity Test

In this section we prove Theorem 1.2 and discuss how tight it is.

CONVENTIONS. In the rest of this work, unless explicitly said otherwise, F denotes GF(2). Furthermore, whenever we write LIN it is to be understood that we are referring to $\textsc{Lin}(F^n, F)$. Throughout this section, if an element b of F appears as an exponent, e.g. $(-1)^b$, it is to be understood as a real number. Thus $(-1)^b$ takes the value 1 or -1 depending on whether b is 0 or 1 respectively.

The main result of this section is based on an application of discrete Fourier analysis techniques to the study of the BLR test. More precisely, we view a function $f\colon F^n \to F$ as a real valued function, and define a function h which is a simple transformation of f. We prove that if the distance from f to its nearest linear function is large, then the Fourier coefficients of h cannot be

---

[6] The *plurality* of a multiset is the most commonly occurring element in the multiset (ties are broken arbitrarily).

very large. Furthermore, we show that the smaller the Fourier coefficients of h are, the higher the probability that f will fail the BLR test.

In the rest of this section, we first review the basic tools of discrete Fourier analysis that we use, and then give a precise formulation of the argument discussed above.

DISCRETE FOURIER TRANSFORM. We consider the family of all real-valued functions on F^n as a 2^n-dimensional real vector space. For functions ϕ, θ: $F^n \to R$, let $\langle \phi, \theta \rangle = \frac{1}{|F|^n} \cdot \sum_{u \in F^n} \phi(u)\theta(u)$ denote the inner product of the functions ϕ and θ. The family of functions $\{ \psi_\alpha : \alpha \in F^n \}$, where $\psi_\alpha(u) = (-1)^{\alpha \cdot u}$, $\alpha \cdot u = \sum_{i=1}^n \alpha_i u_i$, form an orthonormal basis for this linear space (i.e., $\langle \psi_\alpha, \psi_\alpha \rangle = 1$ and $\langle \psi_\alpha, \psi_\beta \rangle = 0$ if $\alpha \neq \beta$). Thus every function ϕ can be uniquely expressed as linear combination of the ψ_α's, namely, $\phi = \sum_{\alpha \in F^n} \widehat{\phi}_\alpha \psi_\alpha$. The coefficient $\widehat{\phi}_\alpha$ is referred to as the α-th Fourier coefficient of ϕ. By the orthonormality of the basis $\{ \psi_\alpha : \alpha \in F^n \}$ it follows that:

$$\widehat{\phi}_\alpha = \langle \phi, \psi_\alpha \rangle. \tag{1}$$

Also the orthonormality of the basis yields the following identity known as Parseval's equality:

$$\langle \phi, \phi \rangle = \sum_{\alpha \in F^n} (\widehat{\phi}_\alpha)^2. \tag{2}$$

The *convolution* of two functions ϕ and θ, denoted $\phi * \theta$, is a function mapping F^n to the reals and defined as follows: $(\phi * \theta)(x) = \frac{1}{|F|^n} \cdot \sum_{u+v=x} \phi(u)\theta(v)$. Note that the convolution operator is associative. Lastly we need the following identity, called the *convolution identity*, which shows the relationship between the Fourier coefficients of two functions ϕ and θ and the Fourier coefficients of their convolution:

$$\forall \alpha \in F^n, \quad \left(\widehat{\phi * \theta}\right)_\alpha = \widehat{\phi}_\alpha \widehat{\theta}_\alpha. \tag{3}$$

LOWER BOUND. To lower bound $\mathsf{Err}(f)$ we use discrete Fourier analysis techniques. We start by establishing a relation between the Fourier coefficients of a transformation of the function f, and $\mathsf{Dist}(f)$, i.e., the distance from f to the linear function closest to f. The transformation is given by the function h: $F^n \to R$, defined as $h(u) = 1$ if $f(u) = 0$ and $h(u) = -1$ otherwise. Over GF(2), h can be expressed as $h(\cdot) = (-1)^{f(\cdot)}$ and this is a crucial element of the following two lemmas. The first lemma shows that if $\mathsf{Dist}(f)$ is large, the Fourier coefficients of h are small.

Lemma 2.1 Suppose f: $F^n \to F$ and $\alpha \in F^n$. Let $h(u) = 1$ if $f(u) = 0$ and -1 otherwise. Then $\widehat{h}_\alpha \leq 1 - 2\,\mathsf{Dist}(f)$.

Proof: Let $l_\alpha(u) = \alpha \cdot u = \sum_{i=1}^n \alpha_i u_i$. Clearly, $l_\alpha \in$ LIN and $\psi_\alpha = (-1)^{l_\alpha}$.

$$\begin{aligned}
\widehat{h}_\alpha &= \langle (-1)^f, \psi_\alpha \rangle && \text{(Using (1))} \\
&= \langle (-1)^f, (-1)^{l_\alpha} \rangle \\
&= \tfrac{1}{|F|^n} \cdot \textstyle\sum_{u \in F^n} (-1)^{f(u)+l_\alpha(u)} \\
&= \Pr_u [\, f(u){=}l_\alpha(u) \,] - \Pr_u [\, f(u){\neq}l_\alpha(u) \,] \\
&= 1 - 2\,\mathsf{Dist}(f, l_\alpha) \\
&\leq 1 - 2\,\mathsf{Dist}(f) \; . \quad \blacksquare
\end{aligned}$$

∎

Our next lemma connects the other parameter, $\mathsf{Err}(f)$, to the value of a convolution of h. This lemma uses the identity $h(\cdot) = (-1)^{f(\cdot)}$ and hence the fact that we are working over GF(2). (In what follows, we use a bold-faced 0, to denote the vector of all 0's to enable distinguishing it from the scalar 0.)

Lemma 2.2 Suppose $f\colon F^n \to F$ and $\alpha \in F^n$. Let $h(u) = 1$ if $f(u) = 0$ and -1 otherwise. Then

$$\mathsf{Err}(f) = \tfrac{1}{2}\left(1 - (h * h * h)(\mathbf{0})\right).$$

Proof: Notice that over GF(2), $f(u) + f(v) + f(u+v)$ is always 0 or 1. Furthermore, the BLR test accepts on random choice u, v if $f(u) + f(v) + f(u+v) = 0$. Alternatively, we can consider the expression $h(u)h(v)h(u+v) = (-1)^{f(u)+f(v)+f(u+v)}$ and observe that the test accepts if this expression is 1 and rejects if this expression is -1. Thus the expression $\frac{1}{2}(1 - h(u)h(v)h(u+v))$ is an indicator for the rejection event in the BLR test, i.e., $\frac{1}{2}(1 - h(u)h(v)h(u+v))$ is 1 if the BLR test rejects and 0 otherwise. Thus we have

$$\mathsf{Err}(f) = \frac{1}{|F|^{2n}} \sum_{u,v \in F^n} \tfrac{1}{2}\left(1 - h(u)h(v)h(u+v)\right) = \tfrac{1}{2}\left(1 - \frac{1}{|F|^{2n}} \sum_{u,v \in F^n} h(u)h(v)h(u+v)\right).$$

From the definition of convolution it follows that $(h * h * h)(\mathbf{0}) = \frac{1}{|F|^{2n}} \sum_{u,v \in F^n} h(u)h(v)h(u+v)$. Thus we derive

$$\mathsf{Err}(f) = \tfrac{1}{2}\left(1 - (h * h * h)(\mathbf{0})\right).$$

∎

The proof of Theorem 1.2 now follows easily using Properties (1), (2), and (3).

Proof of Theorem 1.2: From Lemma 2.2 it suffices to analyze $(h * h * h)(\mathbf{0})$.

$$\begin{aligned}
(h * h * h)(\mathbf{0}) &= \textstyle\sum_{\alpha \in F^n} \left(\widehat{h * h * h}\right)_\alpha \psi_\alpha(\mathbf{0}) && \text{(Using } \psi_\alpha\text{'s as a basis)} \\
&= \textstyle\sum_{\alpha \in F^n} \left(\widehat{h * h * h}\right)_\alpha && \text{(Since } \psi_\alpha(\mathbf{0}) = 1, \text{ for every } \alpha.) \\
&= \textstyle\sum_{\alpha \in F^n} (\widehat{h}_\alpha)^3 && \text{(Using (3))} \\
&\le \left(\max_{\alpha \in F^n} \widehat{h}_\alpha\right) \left(\textstyle\sum_{\alpha \in F^n} (\widehat{h}_\alpha)^2\right) && \\
&= \left(\max_{\alpha \in F^n} \widehat{h}_\alpha\right) && \text{(Using (2) and } \langle h, h \rangle = 1.) \\
&\le 1 - 2\,\mathsf{Dist}(f) && \text{(Using Lemma 2.1).}
\end{aligned}$$

Now using Lemma 2.2, we have

$$\mathsf{Err}(f) = \tfrac{1}{2}\left(1 - (h * h * h)(\mathbf{0})\right) \ge \tfrac{1}{2}\left(1 - (1 - 2\,\mathsf{Dist}(f))\right) = \mathsf{Dist}(f).$$

∎

The next lemma complements Theorem 1.2. This lemma is a slightly more refined version of the bound $\textsc{Rej}(x) \ge 3x - 6x^2$ derived in [6]. To state it we first define the *slack* between functions f and l by

$$\mathsf{sl}(f, l) \overset{\text{def}}{=} \Pr_{u,v \overset{R}{\leftarrow} F^n}\left[f(u) \ne l(u), f(v) \ne l(v), f(u+v) \ne l(u+v) \right].$$

Lemma 2.3 For all $f\colon F^n \to F$ and all $l \in \textsc{Lin}$,

$$\mathsf{Err}(f) \;=\; 3\,\mathsf{Dist}(f,l) - 6\,\mathsf{Dist}(f,l)^2 + 4\,\mathsf{sl}(f,l)\ .$$

Proof: Since f takes values in $F = \mathrm{GF}(2)$, $f(u) + f(v) \neq f(u{+}v)$ if and only if f differs from l in exactly one of the points $\{u, v, u{+}v\}$ or in all of the points $\{u, v, u{+}v\}$. Thus $\mathsf{Err}(f) =$

$$3\,\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v), f(u{+}v)=l(u{+}v)\,\right] + \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)\neq l(v), f(u{+}v)\neq l(u{+}v)\,\right]\ .$$

Furthermore, observe that

$$\begin{aligned} &\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v), f(u{+}v)=l(u{+}v)\,\right] \\ &\quad= \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v)\,\right] - \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v), f(u{+}v)\neq l(u{+}v)\,\right] \\ &\quad= \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v)\,\right] - \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(u{+}v)\neq l(u{+}v)\,\right] \\ &\qquad+ \mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)\neq l(v), f(u{+}v)\neq l(u{+}v)\,\right]\ . \end{aligned}$$

Hence,

$$\begin{aligned} \mathsf{Err}(f) \;=\; & 3\,\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v)\,\right] - 3\,\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(u{+}v)\neq l(u{+}v)\,\right] \\ & + 4\,\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)\neq l(v), f(u{+}v)\neq l(u{+}v)\,\right]\ . \end{aligned}$$

By definition, the last term on the RHS above is $4\,\mathsf{sl}(f,l)$. Moreover, the events $\{\ (u,v) : f(u)=l(u)\ \}$, $\{\ (u,v) : f(v)=l(v)\ \}$, $\{\ (u,v) : f(u{+}v)=l(u{+}v)\ \}$ are pairwise independent. Hence, $\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(u{+}v)\neq l(u{+}v)\,\right] = (1 - \mathsf{Dist}(f,l))^2$ and $\mathrm{Pr}_{u,v}\left[\, f(u)\neq l(u), f(v)=l(v)\,\right] = \mathsf{Dist}(f,l)\,(1 - \mathsf{Dist}(f,l))$. Performing a simple algebraic manipulation, suffices to conclude the proof of the lemma. ∎

TIGHTNESS DISCUSSION. We now discuss how tight the results of this section are. Throughout the rest of this discussion let $x \in [0,1]$ be such that $x\,|F|^n$ is an integer.

Case 1: $x > \frac{1}{2}$.
Then there is no function $f\colon F^n \to F$ such that $\mathsf{Dist}(f) = x$ (since the expected distance from a randomly chosen linear function to f is at most $\frac{1}{2}\,(1 + \frac{1}{|F|^n})$).

Case 2: $x = \frac{1}{2}$.
Randomly choose f so $f(u) = X_u$, where X_u is a random variable distributed according to a Bernoulli distribution with parameter $p \in [\frac{1}{2}, 1]$.[7] A Chernoff bound (see [1, Appendix A]) shows that with overwhelming probability $0 \leq x - \mathsf{Dist}(f) = o(1)$. Moreover, Chebyschev's inequality (see [1, Ch. 4]) implies that with high probability $|\mathsf{Err}(f) - (3\,p(1-p)^2 + p^3)\,| = o(1)$. Thus, if $p = \frac{1}{2}$, Theorem 1.2 is almost tight in the sense that $\textsc{Rej}(x)$ is almost x.

Case 3: $x \leq \frac{5}{16}$.
We will show that in this case the bound $\textsc{Rej}(x) \geq 3x - 6x^2$ is tight. Indeed, for u in F^n let $\lfloor u \rfloor_k \stackrel{\text{def}}{=} u_1 \cdots u_k$. If $S = \{\, u \in F^n : \lfloor u \rfloor_4 \in \{1000, 0100, 0010, 0001, 1111\}\,\}$, then for any function f which equals 1 in $x\,|F|^n$ elements of S, and 0 otherwise, it holds that $\mathsf{Dist}(f) = \mathsf{Dist}(f,0) = x$ and $\mathsf{sl}(f,0) = 0$. Hence, Lemma 2.3 implies that $\mathsf{Err}(f) = 3\,x - 6\,x^2$.

Figure 1, gives evidence showing that Theorem 1.2 is close to being optimal for x in the interval $[\frac{5}{16}, \frac{1}{2}]$. But, as the next two sections show, there is room for improvements.

---

[7] A Bernoulli distribution with parameter p corresponds to the distribution of a $\{0,1\}$-random variable with expectation p.

3 The Summation Lemma

This section is devoted to proving a combinatorial result of independent interest, but necessary in the tighter analysis of the linearity test that we give in Section 4. We also apply this result to obtain a tight upper bound on the probability that the BLR test fails.

First, recall that the lexicographic order in F^n is the total order relation $\leq$ such that $u \leq v$ if and only if $\sum_i u_i 2^{-i} \leq \sum_i v_i 2^{-i}$ (arithmetic over the reals).

Loosely stated, we show that given three subsets A, B, C of F^n, the number of triplets (u, v, w) in $A \times B \times C$ such that $u+v+w = 0$, is maximized when A, B, C are the lexicographically smallest $|A|, |B|, |C|$ elements of F^n respectively.

The following lemma, independently proved by D. J. Kleitman [17], gives a precise statement of the above discussed fact.

For convenience we introduce the following notation: for every nonnegative integer n and $A, B, C \subseteq F^n$ let

$$\Phi_n(A,B,C) \quad = \quad \{\, (u,v,w) \in A \times B \times C \,:\, u+v+w = 0 \,\} \,,$$

and let

$$\varphi_n(A,B,C) \quad = \quad \tfrac{1}{|F|^{2n}} \left|\{\, (u,v,w) \in A \times B \times C \,:\, u+v+w = 0 \,\}\right| \,.$$

Also, for $S \subseteq F^n$ we let S^* denote the collection of the lexicographically smallest $|S|$ elements of F^n.

Lemma 3.1 (Summation Lemma) For every $A, B, C \subseteq F^n$,

$$\varphi_n(A,B,C) \quad \leq \quad \varphi_n(A^*,B^*,C^*) \,.$$

Proof: We proceed by induction. The case $n = 1$ can be easily verified. For the inductive step, we first define, for every $i \in \{1, \ldots, n\}$, a transformation that sends $S \subseteq F^n$ to $S^{(i)} \subseteq F^n$. This transformation consists in lexicographically ordering the elements of S whose i-th component is 0 and 1 respectively. The transformation does not change the number of elements of S with i-th component 0 and 1 respectively. Consider $i \in \{1, \ldots, n\}$ and $b \in F$. Let $f_{i,b}$ be the function that embeds F^{n-1} onto $\{u \in F^n : u_i = b\}$ in the natural way, i.e. for $u = (u_j)_{j \neq i} \in F^{n-1}$, $(f_{i,b}(u))_j = u_j$ if $j \neq i$, and b otherwise. For $S \subseteq F^n$, let $S_b^{(i)}$ be the natural projection into F^{n-1} of the elements of S whose i-th coordinate is b, i.e. $S_b^{(i)} = \{\, (u_j)_{j \neq i} \in F^{n-1} \,:\, f_{i,b}(u) \in S \,\}$. Furthermore, let

$$S^{(i)} \quad = \quad f_{i,0}\left((S_0^{(i)})^*\right) \bigcup f_{i,1}\left((S_1^{(i)})^*\right) \,.$$

Observe that $|S| = |S_0^{(i)}| + |S_1^{(i)}|$. Moreover, lexicographically ordering a set does not change its cardinality, thus $|(S_0^{(i)})^*| = |S_0^{(i)}|$ and $|(S_1^{(i)})^*| = |S_1^{(i)}|$. Since $f_{i,0}$ and $f_{i,1}$ are injective and their ranges are disjoint it follows that $|S^{(i)}| = |S|$.[8]

---

[8] The following example might help in clarifying the notation so far introduced: if $n = 3$ and $S = \{010, 011, 100, 101, 111\}$, then $S_0^{(2)} = \{10, 11\}$, $S_1^{(2)} = \{00, 01, 11\}$, $(S_0^{(2)})^* = \{00, 01\}$, $(S_1^{(2)})^* = \{00, 01, 10\}$, and $S^{(2)} = \{000, 001, 010, 011, 110\}$.

Note that addition (in F^n) of two lexicographically small elements of F^n yields a lexicographically small element of F^n. Thus, it is reasonable to expect that for every $A, B, C \subseteq F^n$ and $i \in \{1, \ldots, n\}$, $\varphi_n(A, B, C) \leq \varphi_n(A^{(i)}, B^{(i)}, C^{(i)})$. We will now prove this latter inequality. Indeed, note that

$$\begin{aligned} \varphi_n(A, B, C) &= \varphi_{n-1}(A_0^{(i)}, B_0^{(i)}, C_0^{(i)}) + \varphi_{n-1}(A_1^{(i)}, B_1^{(i)}, C_0^{(i)}) \\ &\quad + \varphi_{n-1}(A_1^{(i)}, B_0^{(i)}, C_1^{(i)}) + \varphi_{n-1}(A_0^{(i)}, B_1^{(i)}, C_1^{(i)}) . \end{aligned}$$

Applying the inductive hypothesis to each term on the RHS above shows that

$$\begin{aligned} \varphi_n(A, B, C) &\leq \varphi_{n-1}((A_0^{(i)})^*, (B_0^{(i)})^*, (C_0^{(i)})^*) + \varphi_{n-1}((A_1^{(i)})^*, (B_1^{(i)})^*, (C_0^{(i)})^*) \\ &\quad + \varphi_{n-1}((A_1^{(i)})^*, (B_0^{(i)})^*, (C_1^{(i)})^*) + \varphi_{n-1}((A_0^{(i)})^*, (B_1^{(i)})^*, (C_1^{(i)})^*) . \end{aligned}$$

In the previous inequality, the RHS is $\varphi_n(A^{(i)}, B^{(i)}, C^{(i)})$. Hence, $\varphi_n(A, B, C) \leq \varphi_n(A^{(i)}, B^{(i)}, C^{(i)})$ as claimed. We will now show that we can assume that for all $i \in \{1, \ldots, n\}$, $A^{(i)} = A$, $B^{(i)} = B$, and $C^{(i)} = C$. Indeed, if this was not the case, we can repeat the above argument by considering $A^{(i)}$, $B^{(i)}$, $C^{(i)}$ instead of A, B, C. To prove that this iterative process is guaranteed to eventually stop let $u \in F^n$ also represent the integer with binary expansion u. Note that if $A^{(i)} \neq A$, or $B^{(i)} \neq B$, or $C^{(i)} \neq C$, then $\sum_{u \in S} u > \sum_{u \in S^{(i)}} u$ for some $S \in \{A, B, C\}$. Hence the aforementioned iterative process stops in at most $\sum_{S \in \{A,B,C\}} \sum_{u \in S} u$ steps.

One would like to conclude the proof of the lemma by claiming that, if for all i, $A^{(i)} = A$, $B^{(i)} = B$, and $C^{(i)} = C$, then A, B, C are equal to A^*, B^*, C^* respectively. We will show that the latter claim is 'almost' true, in the sense that if e denotes $(1, 0, \ldots, 0) \in F^n$, e' denotes $(0, 1, \ldots, 1) \in F^n$, and $V = \{ (u_1, \ldots, u_n) \in F^n : u_1 = 0 \}$ then the following holds:

$$\text{If for every } i \in \{1, \ldots, n\},\ S = S^{(i)},\ \text{then } S = S^* \text{ or } S = (V \setminus \{e\}) \cup \{e'\} .$$

We prove the above fact by contradiction. Assume that $S \neq S^*$ and $S \neq (V \setminus \{e\}) \cup \{e'\}$. Since $S = S^{(1)}$, then either $(1, 0, \ldots, 0, 1) \in F^n$ is in S or $(0, 1, \ldots, 1, 0) \in F^n$ is not in S. Suppose that $(1, 0, \ldots, 0, 1) \in F^n$ is in S. Since $S = S^{(1)}$ and $S \neq S^*$, we know that $e \notin S$. Thus, $(1, 0, \ldots, 0) \in F^{n-1}$ is in $S_1^{(n)}$ and $(0, 1, \ldots, 1) \in F^{n-1}$ is not in $S_1^{(n)}$. Hence, $(S_1^{(n)})^* \neq S_1^{(n)}$. It follows that $S \neq S^{(n)}$, a contradiction. Suppose now that $(0, 1, \ldots, 1, 0) \in F^n$ is not in S. Since $S = S^{(1)}$ and $S \neq S^*$, we know that $e' \in S$. Thus, $(1, 0, \ldots, 0) \in F^{n-1}$ is in $S_0^{(n)}$ and $(0, 1, \ldots, 1) \in F^{n-1}$ is not in $S_0^{(n)}$. Hence, $(S_0^{(n)})^* \neq S_0^{(n)}$. It follows that $S \neq S^{(n)}$, again a contradiction.[9]

Thus far we have shown that in order to upper bound $\varphi_n(A, B, C)$ we can restrict our attention to the sets A, B, C that are either in lexicographically smallest order or take the form $(V \setminus \{e\}) \cup \{e'\}$. To conclude the lemma we need to consider three cases. These cases depend on how many of the sets A, B, C are in lexicographically smallest order.

Case 1: exactly two of the sets A, B, C are in lexicographically smallest order.
Without loss of generality assume $A = A^*$, $B = B^*$, and $C = (V \setminus \{e\}) \cup \{e'\}$. Then

$$\varphi_n(A, B, C) = \varphi_n(A, B, V) + \varphi_n(A, B, \{e'\}) - \varphi_n(A, B, \{e\}) .$$

Note that $\varphi_n(A, B, \{e\}) = \max\{0, |A \cap V| + |B \cap V| - |V|\} + \max\{0, |A \setminus V| + |B \setminus V| - |V|\}$ and $\varphi_n(A, B, \{e'\}) = \min\{|A \setminus V|, |B \cap V|\} + \min\{|A \cap V|, |B \setminus V|\}$. Hence, $\varphi_n(A, B, \{e\}) \geq \varphi_n(A, B, \{e'\})$.

---

[9] Observe that we only required that $S^{(1)} = S^{(n)} = S$.

Thus, $\varphi_n(A,B,C) \leq \varphi_n(A,B,V)$. To conclude, observe that $C^* = V$ and recall that $A = A^*$ and $B = B^*$.

Case 2*:* exactly one of the sets A, B, C is in lexicographically smallest order.

Without loss of generality, we assume that $A = A^*$ and $B = C = (V \setminus \{e\}) \cup \{e'\}$. If $A = F^n$ or $A = \emptyset$, then it is obvious that $\varphi_n(A,B,C) = \varphi_n(A^*,B^*,C^*)$ and we are done. Thus, we also assume that $A \neq F^n$ and $A \neq \emptyset$. Then,

$$\begin{aligned}\varphi_n(A,B,C) &= \varphi_n(A,V,V) - \varphi_n(A,V,\{e\}) - \varphi_n(A,\{e\},V) \\ &\quad + \varphi_n(A,V,\{e'\}) + \varphi_n(A,\{e'\},V) + \varphi_n(A,\{e\},\{e\}) \\ &\quad - \varphi_n(A,\{e\},\{e'\}) - \varphi_n(A,\{e'\},\{e\}) + \varphi_n(A,\{e'\},\{e'\}) \;.\end{aligned}$$

Note that $\varphi_n(A,V,\{e\}) = \varphi_n(A,\{e\},V) = |A \cap V|$, $\varphi_n(A,V,\{e'\}) = \varphi_n(A,\{e'\},V) = |A \setminus V|$. Since $A = A^*$ and $A \neq F^n$, then $\varphi_n(A,\{e'\},\{e\}) = \varphi_n(A,\{e'\},\{e\}) = 0$. Since $A = A^*$ and $A \neq \emptyset$, then $\varphi_n(A,\{e'\},\{e'\}) = \varphi_n(A,\{e\},\{e\}) = 1$. Thus, $\varphi_n(A,B,C) = \varphi_n(A,V,V) - 2\,|A \cap V| + 2\,|A \setminus V| + 2$. Since $A = A^*$ and $A \neq F^n$, then $|A \setminus V| < |V|$, and if $|A \setminus V| \neq 0$, then $|A \cap V| = |V|$. Since $A = A^*$ and $A \neq \emptyset$, then $|A \cap V| > 0$, and if $|A \setminus V| = 0$, then $|A \cap V| = |A|$. Hence, $\varphi_n(A,B,C) \leq \varphi_n(A,V,V)$. To conclude, observe that $B^* = C^* = V$ and recall that $A = A^*$.

Case 3*:* none of the sets A, B, C is in lexicographically smallest order.

In this case $A = B = C = (V \setminus \{e\}) \cup \{e'\}$. Thus,

$$\begin{aligned}\varphi_n(A,B,C) &= \varphi_n(V,V,V) - \max\{0, |A \cap V| + |B \cap V| - |V|\} \\ &\quad - \max\{0, |A \cap V| + |C \cap V| - |V|\} - \max\{0, |B \cap V| + |C \cap V| - |V|\} \;.\end{aligned}$$

Hence, $\varphi_n(A,B,C) \leq \varphi_n(V,V,V)$. To conclude, observe that $A^* = B^* = C^* = V$. ∎

By definition, a subspace V of F^n is such that if $u, v \in V$, then $u+v \in V$. This motivates using

$$\tfrac{1}{|S|^2}|\Phi_n(S,S,S)| \;,$$

as a measure of how *close* the set $S \subseteq F^n$ is to being a subspace. The larger this quantity is, the closer the set S is off being a subspace. From this point of view, the Summation Lemma implies that the collection of the lexicographically smallest m elements of F^n is the subset of F^n (of cardinality m) that more closely resembles a subspace.

Lemma 3.2 Suppose $f\colon F^n \to F$. Let $x = \mathsf{Dist}(f)$. Let k be the unique integer such that $2^{-k} \leq x < 2^{-k+1}$, and let $\delta = 2^{-k}$. Then

$$\mathsf{Err}(f) \;\leq\; 3\,x - 6\,x^2 + 4\,\delta^2 + 12\,(x-\delta)^2 \;.$$

Proof: Let l be the closest linear function to f, and let $S = \{\, u : f(u) \neq l(u) \,\}$. Note that $\mathsf{sl}(f,l) = \varphi_n(S,S,S)$, thus by Lemma 2.3 we have that

$$\mathsf{Err}(f) \;=\; 3\,\delta - 6\,\delta^2 + 4\,\varphi_n(S,S,S) \;.$$

By the Summation Lemma, $\varphi_n(S,S,S) \leq \varphi_n(S^*,S^*,S^*)$. The lemma will follow once we show that $\varphi_n(S^*,S^*,S^*) = \delta^2 + 3\,(x-\delta)^2$. Indeed, let V be the lexicographically smallest $\delta\,|F|^n$ elements of F^n. Note that V is a subspace, $V \subset S^*$, and $|S^*| = |S| = x\,|F|^n$. Since $\varphi_n(S^* \setminus V, V, V)$, $\varphi_n(V, S^* \setminus V, V)$, $\varphi_n(V, V, S^* \setminus V)$, and $\varphi_n(S^* \setminus V, S^* \setminus V, S^* \setminus V)$ are all equal to 0 we get that

$$\begin{aligned}\varphi_n(S^*, S^*, S^*) &= \varphi_n(S^* \setminus V, S^* \setminus V, V) + \varphi_n(S^* \setminus V, V, S^* \setminus V)\\ &\quad + \varphi_n(V, S^* \setminus V, S^* \setminus V) + \varphi_n(V, V, V)\,.\end{aligned}$$

Note that $\varphi_n(V, V, V) = \delta^2$. Moreover, $\varphi_n(S^* \setminus V, S^* \setminus V, V)$, $\varphi_n(S^* \setminus V, V, S^* \setminus V)$, and $\varphi_n(V, S^* \setminus V, S^* \setminus V)$ are all equal to $(x-\delta)^2$. Thus, $\varphi_n(S^*, S^*, S^*) = \delta^2 + 3\,(x-\delta)^2$ as we claimed.

We will now prove that the bound of Lemma 3.2 cannot be improved. Indeed, let $x \in [0, \frac{1}{2}]$ be such that $x\,|F|^n$ is an integer. Let S be the lexicographically smallest $x\,|F|^n$ elements of F^n. Consider the function $f\colon F^n \to F$ which evaluates to 1 on every element of S and to 0 otherwise, i.e. f is the characteristic function of S. We will prove that the closest linear function to f is the zero function, hence $\mathsf{Dist}(f) = x$. But, first note that since $S = S^*$, then $\varphi_n(S, S, S) = \varphi_n(S^*, S^*, S^*)$. Hence, from the proof of Lemma 3.2, it follows that $\mathsf{Err}(f)$ meets the upper bound of the statement of Lemma 3.2. To prove that the closest linear function to f is the zero function we argue by contradiction. We consider the following two cases:

Case 1*:* $x \in [0, \frac{1}{4}]$.
Here, the zero function is at distance x from f. If some other linear function was at distance less than x from f, then such linear function would be at distance less than $2x \le \frac{1}{2}$ from the zero function. A contradiction, since two distinct linear functions are at distance $\frac{1}{2}$.

Case 2*:* $x \in (\frac{1}{4}, \frac{1}{2}]$.
Let V be the largest subspace of F^n contained in S, and let V' be the smallest subspace of F^n that contains S. Recall that the cardinality of a subspace of F^n is a power of two. Thus, since S is the set of the lexicographically smallest $x|F|^n$ elements of F^n, then $|V| = \frac{1}{4}|F|^n$ and $|V'| = \frac{1}{2}|F|^n$. For the sake of contradiction, assume $l\colon F^n \to F$ is a nonzero linear function whose distance to f is less than x. Note that a linear function which is nonzero over a subspace of F^n must evaluate to 1 in exactly half the elements of that subspace. In particular, l evaluates to 1 on half the elements of F^n.

Case 2.1*:* l evaluates to 0 over V.
Recall that f evaluates to 0 outside of S and to 1 over S. Moreover, l evaluates to 1 in exactly half the elements of F^n. Thus, l disagrees with f in every element of V and in at least $\frac{1}{2}|F|^n - |S \setminus V|$ of the elements not in V. Hence, the distance between f and l is at least $\frac{1}{4} + (\frac{1}{2} - (x - \frac{1}{4})) \ge x$, a contradiction.

Case 2.2*:* l does not evaluate to 0 over V.
Then, l evaluates to 1 in exactly half the elements of V and half the elements of V'. Thus, l disagrees with f in half the elements of V and in at least $|S \setminus V| - \frac{1}{2}(|V'| - |V|)$ of the elements of $S \setminus V$. Moreover, l evaluates to 1 on half the elements of F^n and on half the elements of V'. Hence, since f evaluates to 0 on the elements of $F^n \setminus V'$, it follows that l disagrees with f in $\frac{1}{2}|F^n| - \frac{1}{2}|V'|$ of the elements of $F^n \setminus V'$. Thus, the distance between f and l is at least $\frac{1}{2}|V| + (|S \setminus V| - \frac{1}{2}(|V'| - |V|)) + (\frac{1}{2}|F^n| - \frac{1}{2}|V'|) = \frac{1}{8} + ((x - \frac{1}{4}) - \frac{1}{8}) + (\frac{1}{2} - \frac{1}{4}) = x$, again a contradiction.

4 Combinatorial analysis of the linearity test

We now prove Theorem 1.3, i.e. that $\textsc{Knee} = \frac{45}{128}$. To prove that $\textsc{Knee} \geq \frac{45}{128}$ we associate to a function $f\colon F^n \to F$ a function $g_f\colon F^n \to F$, whose value at u is $\textsc{Plurality}\{\, f(u{+}v) - f(v) : v \in F^n \,\}$. Then, if $\mathsf{Err}(f)$ is sufficiently small three things occur: (i) An overwhelming majority of the values $\{\, f(u+v) - f(v) : v \in F^n \,\}$ agree with $g_f(u)$, (ii) g_f is linear, (iii) g_f is close to f. This argument was first used in [9] while studying linearity testing over finite groups. We will show how this argument can be tightened in the case of linearity testing over GF(2).

More precisely, the proof of Theorem 1.3 is a consequence of the following three lemmas:

Lemma 4.1 For all $f\colon F^n \to F$, then $\mathsf{Err}(f) \geq \frac{1}{2}\,\mathsf{Dist}(f, g_f)$.

Lemma 4.2 For all $f\colon F^n \to F$, if g_f is linear, then $\mathsf{Err}(f) \geq 2\,\mathsf{Dist}(f, g_f) \cdot [1 - \mathsf{Dist}(f, g_f)]$.

Lemma 4.3 For all $f\colon F^n \to F$, if $\mathsf{Err}(f) < \frac{45}{128}$, then g_f is linear.

We first show that Theorem 1.3 follows from the above stated results. Assume $\textsc{Knee} < \frac{45}{128}$, then, there is a function $f\colon F^n \to F$, such that $\mathsf{Err}(f) < \frac{45}{128}$ and $x = \mathsf{Dist}(f) \geq \frac{1}{4}$. By Lemma 2.3, $\mathsf{Err}(f) \geq 3x - 6x^2$, thence we need only consider the case in which x is at least $\frac{5}{16}$. Moreover, by Lemma 4.3, g_f is a linear function. Thus, $\mathsf{Dist}(f, g_f) \geq x \geq \frac{5}{16}$, which together with Lemmas 4.1 and 4.2 imply that $\mathsf{Err}(f) \geq \min_{x\in[5/16,1]} \max\left\{\frac{1}{2}x, 2(1-x)x\right\} = \frac{3}{8}$, a contradiction. Hence, $\textsc{Knee} \geq \frac{45}{128}$. In our tightness discussion part of Section 2 we showed that there exists a function $f\colon F^n \to F$ such that $\mathsf{Dist}(f) = \frac{5}{16}$ and $\mathsf{Err}(f) = \frac{45}{128}$. Hence, $\textsc{Knee} = \frac{45}{128}$ as we wanted to prove.

The rest of this section is dedicated to proving Lemmas 4.1 through 4.3.

The proofs of Lemmas 4.1 and 4.2 are based on an observation which is implicit in [14]. This observation crucially depends on the fact that f takes values over $F = \mathrm{GF}(2)$. It says that for every $u \in F^n$,

$$\mathrm{Pr}_v\left[\, f(u{+}v){-}f(v){=}g_f(u) \,\right] \;\;\geq\;\; \tfrac{1}{2} \;.$$

Hence, if $f(u){\neq}g_f(u)$, then $f(u) \neq f(u{+}v) - f(v)$ at least half of the time, which implies

$$\mathrm{Pr}_{u,v}\left[\, f(u){+}f(v){\neq}f(u{+}v) \mid f(u){\neq}g_f(u) \,\right] \;\;\geq\;\; \tfrac{1}{2} \;. \qquad (4)$$

Proof of Lemma 4.1: Simple conditioning says that $\mathsf{Err}(f)$ is at least

$$\mathrm{Pr}_{u,v}\left[\, f(u){+}f(v){\neq}f(u{+}v) \mid f(u){\neq}g_f(u) \,\right] \cdot \mathsf{Dist}(f, g_f) \;.$$

But by (4) we know this is at least $\frac{1}{2}\,\mathsf{Dist}(f, g_f)$. ■

Proof of Lemma 4.2: Assume g_f is linear. As observed in the proof of Lemma 2.3

$$\begin{aligned}\mathsf{Err}(f) \;&=\; 3\,\mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(v){=}g_f(v), f(u{+}v){=}g_f(u{+}v) \,\right] \\ &\qquad + \mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(v){\neq}g_f(v), f(u{+}v){\neq}g_f(u{+}v) \,\right] \;.\end{aligned}$$

Since g_f is linear, $\mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(v){=}g_f(v), f(u{+}v){=}g_f(u{+}v) \,\right] =$

$$\mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(u){+}f(v){\neq}f(u{+}v) \,\right] - \mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(v){\neq}g_f(v), f(u{+}v){\neq}g_f(u{+}v) \,\right] \;.$$

Hence,

$$\begin{aligned}\mathsf{Err}(f) &= 3\,\mathrm{Pr}_{u,v}\left[\, f(u){+}f(v){\neq}f(u{+}v) \mid f(u){\neq}g_f(u) \,\right] \cdot \mathsf{Dist}(f,g_f) \\ &\quad - 2\,\mathrm{Pr}_{u,v}\left[\, f(u){\neq}g_f(u), f(v){\neq}g_f(v), f(u{+}v){\neq}g_f(u{+}v) \,\right] \ .\end{aligned}$$

In this last expression, the first term can be lower bounded, as in the proof of Lemma 4.1, by $\frac{3}{2}\,\mathsf{Dist}(f,g_f)$. The second term is $2\,\mathsf{sl}(f,g_f)$. Thus, we have $\mathsf{Err}(f) \geq \frac{3}{2}\,\mathsf{Dist}(f,g_f) - 2\,\mathsf{sl}(f,g_f)$. Finally, applying Lemma 2.3, we get that $\mathsf{Err}(f) \geq 3\,\mathsf{Dist}(f,g_f) - 3\,\mathsf{Dist}(f,g_f)^2 - \frac{1}{2}\,\mathsf{Err}(f)$. The lemma follows. ■

Proof of Lemma 4.3: By contradiction. Assume g_f is not linear. Then there are x, y such that $g_f(x) + g_f(y) \neq g_f(x{+}y)$. Note that by construction $g_f(0) = 0$, thus x and y are distinct and nonzero. Hence, $x, y, x{+}y$ are distinct. Since $g_f(x) + g_f(y) \neq g_f(x{+}y)$ it cannot be that $g_f(x), g_f(y), g_f(x{+}y)$ are all zero. Without loss of generality, we assume that $g_f(x{+}y) = 1$. We now show that we can also assume that $g_f(x) = g_f(y) = 1$. Indeed, if f satisfies the latter assumption we are done. Otherwise, since $g_f(x) + g_f(y) \neq g_f(x{+}y) = 1$, we have that $g_f(x) = g_f(y) = 0$. Let $l\colon F^n \to F$ be a linear function such that $l(x) = l(y) = 1$ (such function exists since x, y are distinct and nonzero). Set $f' = f + l$ and observe that $\mathsf{Err}(f') = \mathsf{Err}(f)$ and $g_{f'} = g_f + l$. Hence, $\mathsf{Err}(f') < \frac{45}{128}$, $g_{f'}(x) + g_{f'}(y) \neq g_{f'}(x{+}y)$, and $g_{f'}(x) = g_{f'}(y) = g_{f'}(x{+}y) = 1$. So, we can continue arguing about f' instead of f.

Set $S = \{0, x, y, x{+}y\}$. We will begin by investigating nonlinearity on cosets of S. For every $s \in F^n$, define f_s to be the function from S to F, such that $f_s(u) = f(s{+}u)$. For every $s, t \in F^n$, let

$$p_{s,t} = \mathrm{Pr}_{u,v \xleftarrow{R} S}\left[\, f_s(u){+}f_t(v){\neq}f_{s+t}(u{+}v) \,\right].$$

By interchanging the orders of expectations we see that

$$\mathsf{Err}(f) = E_{s,t \xleftarrow{R} F^n}\left[\, p_{s,t} \,\right] \ . \tag{5}$$

Now $p_{s,t}$ depends only on the values of f on the cosets $s + S$, $t + S$, and $s + t + S$. We classify these cosets according to the pattern of values of f on the coset. Define the trace of f at w as

$$tr_f(w) \ = \ [f(w), f(w{+}x), f(w{+}y), f(w{+}x{+}y)] \ .$$

We partition the elements w of F^n according to the values that the trace of f at w takes,

$$\begin{aligned} H_0 &= \{\, w : tr_f(w) \text{ equals } [0,0,0,0] \text{ or } [1,1,1,1] \,\} \\ H_x &= \{\, w : tr_f(w) \text{ equals } [0,0,1,1] \text{ or } [1,1,0,0] \,\} \\ H_y &= \{\, w : tr_f(w) \text{ equals } [0,1,0,1] \text{ or } [1,0,1,0] \,\} \\ H_{x+y} &= \{\, w : tr_f(w) \text{ equals } [0,1,1,0] \text{ or } [1,0,0,1] \,\} \\ H_{odd} &= \{\, w : tr_f(w) \text{ has an odd number of 1's} \,\} \ , \end{aligned}$$

and define their relative measures $h_0 = |H_0|/|F|^n$, $h_x = |H_x|/|F|^n$, $h_y = |H_y|/|F|^n$, $h_{x+y} = |H_{x+y}|/|F|^n$, and $h_{odd} = |H_{odd}|/|F|^n$. Notice that if $s \in H_z$ then the whole coset $s + S$ is in H_z, for any of the five sets H_z.

By symmetry we may assume that $h_x \leq h_y \leq h_{x+y}$.

The condition $g_f(x+y) = 1$ implies

$$\Pr_{u \xleftarrow{R} F^n} [\, f(u+x+y) = f(u) \,] \le \tfrac{1}{2},$$

whence

$$h_0 + h_{x+y} + \tfrac{1}{2} h_{odd} \;\; \le \;\; \tfrac{1}{2} \, , \tag{6}$$

since for each coset $w+S$ in H_{odd}, half the elements $w+u$ satisfy $f(w+u) = f(w+u+x+y)$, while all elements w of H_0 and H_{x+y} satisfy $f(w) = f(w+x+y)$.

So no single set among the four H_0, H_x, H_y, or H_{x+y} is too large; each of h_0, h_x, h_y, h_{x+y} is bounded by $\frac{1}{2}$. If f were strictly linear, one of these four sets would cover all of F^n. As it is, the interaction of several substantial sets among H_0, H_x, H_y, H_{x+y}, or the presence of a large H_{odd}, will force a large nonlinearity on f, and will give the desired lower bound on $\mathsf{Err}(f)$.

To quantify this interaction between sets, we partition $F^n \times F^n$ into six sets as follows:

$\mathcal{A}$ = Set of all (s,t) such that $\{s, t, s+t\}$ are all in the same set, either H_0 or H_x or H_y or H_{x+y}

$\mathcal{B}$ = Set of all (s,t) such that two of $\{s, t, s+t\}$ are in the same set H_0 or H_x or H_y or H_{x+y}, and the other one is in H_{odd}

$\mathcal{C}$ = Set of all (s,t) such that at least two of $\{s, t, s+t\}$ are in H_{odd}

$\mathcal{D}$ = Set of all (s,t) such that $\{s, t, s+t\} \subset H_0 \cup H_x \cup H_y \cup H_{x+y}$ with exactly two elements from the same set H_0, H_x, H_y or H_{x+y}

$\mathcal{E}$ = Set of all (s,t) such that one of $\{s, t, s+t\}$ is in H_{odd}, the other two are from different sets in H_0, H_x, H_y and H_{x+y}

$\mathcal{F}$ = Set of all (s,t) such that $\{s, t, s+t\}$ are from different sets H_0, H_x, H_y, H_{x+y}

The following tables illustrate the above defined partition.

(s,t)	H_0	H_x	H_y	H_{x+y}	H_{odd}
H_0	$\mathcal{A}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$
H_x	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{E}$
H_y	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{E}$
H_{x+y}	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{E}$
H_{odd}	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{C}$

$s + t \in H_0$

(s,t)	H_0	H_x	H_y	H_{x+y}	H_{odd}
H_0	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{E}$
H_x	$\mathcal{D}$	$\mathcal{A}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$
H_y	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{E}$
H_{x+y}	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{E}$
H_{odd}	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{C}$

$s + t \in H_x$

(s,t)	H_0	H_x	H_y	H_{x+y}	H_{odd}
H_0	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{E}$
H_x	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{E}$
H_y	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{A}$	$\mathcal{D}$	$\mathcal{B}$
H_{x+y}	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{E}$
H_{odd}	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{C}$

$s + t \in H_y$

(s,t)	H_0	H_x	H_y	H_{x+y}	H_{odd}
H_0	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{E}$
H_x	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{E}$
H_y	$\mathcal{F}$	$\mathcal{F}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{E}$
H_{x+y}	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{A}$	$\mathcal{B}$
H_{odd}	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{C}$

$s + t \in H_{x+y}$

(s,t)	H_0	H_x	H_y	H_{x+y}	H_{odd}
H_0	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{C}$
H_x	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{C}$
H_y	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{E}$	$\mathcal{C}$
H_{x+y}	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{B}$	$\mathcal{C}$
H_{odd}	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$

$s + t \in H_{odd}$

We now proceed to show a lower bound for $\mathsf{Err}(f)$ which depends on the relative size of the sets $\mathcal{A}$,$\mathcal{B}$,$\mathcal{C}$,$\mathcal{D}$,$\mathcal{E}$, and $\mathcal{F}$. Indeed, observe that if (s,t) is in $\mathcal{B}$, then $p_{s,t}$ is at least $\frac{1}{4}$. (We calculate an example: suppose s and $s+t$ are both in H_x, with $tr_f(s) = [0,0,1,1]$ and $tr_f(s+t) = [1,1,0,0]$, while t is in H_{odd}, with $tr_f(t) = [1,1,0,1]$. If f were linear on the cosets $s+S, t+S, s+t+S$, and $tr_f(s)$, $tr_f(s+t)$ were as given, then $tr_f(t)$ would necessarily be $[1,1,0,0]$, and t would be in H_x. The value $tr_f(t)$ differs from $[1,1,0,0]$ in the last position, corresponding to $x+y$. Thus whenever $v = x+y$ we will have $f(s{+}u) + f(t{+}v) \neq f(s{+}t{+}u{+}v)$. This happens for $\frac{1}{4}$ of the random choices of (u,v).) With similar arguments one can show that if (s,t) is in $\mathcal{C}$, then $p_{s,t}$ is at least $\frac{3}{8}$. And, if (s,t) is in $\mathcal{D}$, $\mathcal{E}$, or $\mathcal{F}$, then $p_{s,t}$ is $\frac{1}{2}$. Hence, if for a set $T \subseteq F^n \times F^n$ we let $\mu(T) = |T|/|F|^{2n}$, then (5) yields

$$\mathsf{Err}(f) \quad \geq \quad \tfrac{1}{4}\,\mu(\mathcal{B}) + \tfrac{3}{8}\,\mu(\mathcal{C}) + \tfrac{1}{2}\left[\mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F})\right] .$$

Recalling that $\mu(\mathcal{C}) = 1 - (\mu(\mathcal{A}) + \mu(\mathcal{B}) + \mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F}))$, allows us to conclude that

$$\mathsf{Err}(f) \quad \geq \quad \tfrac{3}{8} - \tfrac{1}{8}\left(3\,\mu(\mathcal{A}) + \mu(\mathcal{B})\right) + \tfrac{1}{8}\left[\mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F})\right] . \tag{7}$$

We now derive from (7) another lower bound for $\mathsf{Err}(f)$ which will depend solely on $h_0, h_x, h_y, h_{x+y}, h_{odd}$, and $\mu(\mathcal{F})$.

We first need the following identities relating the measure of the sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, and $\mathcal{F}$, to h_0, h_x, h_y, h_{x+y}, and h_{odd}. Consider the probability that randomly chosen s and t are in the same set H_0, H_x, H_y, or H_{x+y}, plus the corresponding probabilities for $(s, s+t)$ and $(t, s+t)$; expressing this sum of probabilities in two ways yields

$$3\mu(\mathcal{A}) + \mu(\mathcal{B}) + \mu(\mathcal{D}) \quad = \quad 3\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) . \tag{8}$$

Consider the probability that s and t are in two different sets H_0, H_x, H_y, or H_{x+y}, plus the corresponding probabilities for $(s, s+t)$ and $(t, s+t)$; expressing this sum of probabilities in two ways yields:

$$2\mu(\mathcal{D}) + \mu(\mathcal{E}) + 3\mu(\mathcal{F}) \quad = \quad 3\left((1-h_{odd})^2 - \left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right)\right) . \tag{9}$$

Adding $-\frac{1}{8}$ of (8) and $\frac{1}{8}$ of (9) to (7), gives

$$\mathsf{Err}(f) \quad \geq \quad \tfrac{3}{8} + \tfrac{3}{8}(1-h_{odd})^2 - \tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) - \tfrac{1}{4}\mu(\mathcal{F}) . \tag{10}$$

We now proceed to upper bound $\mu(\mathcal{F})$. We divide the analysis into two cases.

Case 1: $h_x + h_y - h_0 - h_{x+y} > \frac{1}{4}$.

By case assumption and since $h_{x+y} \geq h_y$ we have that $h_x \geq h_x + h_y - h_0 - h_{x+y} > \frac{1}{4}$. So, $h_x, h_y, h_{x+y} \in (\frac{1}{4}, \frac{1}{2}]$. As in Section 3, for $A, B, C \subseteq F^n$ we let

$$\varphi_n(A,B,C) \;=\; \tfrac{1}{|F|^{2n}} \left|\{\, (u,v,w) \in A\times B\times C \,:\, u{+}v{+}w = 0 \,\}\right| .$$

Observe now, that for each element (u,v) of $\mathcal{F}$, $\{u, v, u{+}v\}$ either contains an element from H_0 or contains one element from each of the sets H_x, H_y, and H_{x+y}.

The contribution to $\mathcal{F}$ of the elements (u,v), where $\{u, v, u{+}v\}$ contain elements from each of the sets H_x, H_y, and H_{x+y}, is upper bounded by $6\,\varphi_n(H_x, H_y, H_{x+y})$. The Summation Lemma implies that $\varphi_n(H_x, H_y, H_{x+y}) \leq \varphi_n(H_x^*, H_y^*, H_{x+y}^*)$. Note that h_x, h_y, h_{x+y} completely characterize H_x^*, H_y^*, H_{x+y}^*. Thus since $h_x, h_y, h_{x+y} \in (\frac{1}{4}, \frac{1}{2}]$ we have that

$$\begin{aligned}
\varphi_n(H_x^*, H_y^*, H_{x+y}^*) &= \tfrac{1}{4} - \tfrac{1}{2}(h_x + h_y + h_{x+y}) + h_x h_y + h_x h_{x+y} + h_y h_{x+y} \\
&= \tfrac{1}{4} - \tfrac{1}{2}[(h_0 + h_{odd}) + (h_x + h_y + h_{x+y})](h_x + h_y + h_{x+y}) \\
&\quad + h_x h_y + h_x h_{x+y} + h_y h_{x+y} \\
&= \tfrac{1}{4} - \tfrac{1}{2}\,(h_0 + h_{odd})(h_x + h_y + h_{x+y}) - \tfrac{1}{2}\,(h_x^2 + h_y^2 + h_{x+y}^2) \; .
\end{aligned}$$

Hence, $6\,\varphi_n(H_x, H_y, H_{x+y}) \leq \frac{3}{2} - 3\,(h_0 + h_{odd})(h_x + h_y + h_{x+y}) - 3\,(h_x^2 + h_y^2 + h_{x+y}^2)$.

Furthermore, the contribution to $\mathcal{F}$ of the elements (u,v), where $\{u, v, u{+}v\}$ contains an element of H_0 is upper bounded by

$$3\,\varphi_n(H_0, H_x, H_y \cup H_{x+y}) + 3\,\varphi_n(H_0, H_y, H_x \cup H_{x+y}) + 3\,\varphi_n(H_0, H_{x+y}, H_x \cup H_y) \; ,$$

which is at most $3h_0\,(h_x{+}h_y{+}h_{x+y})$. Putting it all together, we have

$$\mu(\mathcal{F}) \;\leq\; \tfrac{3}{2} - 3h_{odd}(h_x + h_y + h_{x+y}) - 3(h_x^2 + h_y^2 + h_{x+y}^2) \; ,$$

which jointly with (10) implies that

$$\begin{aligned}
\mathsf{Err}(f) &\geq \tfrac{3}{8} + \tfrac{3}{8}(1 - h_{odd})^2 - \tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) \\
&\quad -\tfrac{3}{8} + \tfrac{3}{4}h_{odd}(h_x + h_y + h_{x+y}) + \tfrac{3}{4}(h_x^2 + h_y^2 + h_{x+y}^2) \\
&= \tfrac{3}{8} - \tfrac{3}{8}h_{odd}^2 - \tfrac{3}{4}h_0 h_{odd} - \tfrac{3}{4}h_0^2 \\
&\geq \tfrac{3}{8} - \tfrac{3}{8}(h_{odd} + 4h_0)^2 \; .
\end{aligned}$$

We conclude the analysis of this case by noting that

$$\begin{aligned}
\tfrac{1}{4} &\geq 1 - 3(h_x + h_y - h_0 - h_{x+y}) \\
&\geq 1 - h_x - h_y - h_{x+y} + 3h_0 \\
&= h_{odd} + 4h_0 \; ,
\end{aligned}$$

where the first inequality follows by case assumption, and the second one because $h_x \leq h_y \leq h_{x+y}$, so that

$$\mathsf{Err}(f) \geq \tfrac{3}{8} - \tfrac{3}{8}\left(\tfrac{1}{4}\right)^2 = \tfrac{45}{128} \; .$$

Case 2: $h_x + h_y - h_0 - h_{x+y} \leq \frac{1}{4}$.

To each element (u,v) in $\mathcal{F}$, associate the unique tuple $(u',v') \in \{u,v,u{+}v\} \times \{u,v,u{+}v\}$, such that $(u',v') \in H_0 \times H_{x+y} \cup H_x \times H_y$. This scheme associates to each element of $H_0 \times H_{x+y} \cup H_x \times H_y$ at most 6 elements of $\mathcal{F}$. Thus, $\mu(\mathcal{F}) \leq 6\,(h_0 h_{x+y} + h_x h_y)$. Which jointly with (10) implies

$$\begin{aligned}
\mathsf{Err}(f) &\geq \tfrac{3}{8} + \tfrac{3}{8}(1-h_{odd})^2 - \tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) - \tfrac{3}{2}\left(h_0 h_{x+y} + h_x h_y\right) \\
&= \tfrac{3}{8} + \tfrac{3}{8}[(h_0 + h_{x+y}) + (h_x + h_y)]^2 - \tfrac{3}{4}[(h_0 + h_{x+y})^2 + (h_x + h_y)^2] \\
&= \tfrac{3}{8} - \tfrac{3}{8}\left(h_x + h_y - h_0 - h_{x+y}\right)^2 \ .
\end{aligned}$$

The analysis of this case concludes by observing that

$$\begin{aligned}
\tfrac{1}{4} &\geq h_x + h_y - h_0 - h_{x+y} \\
&= 1 - h_{odd} - 2(h_0 + h_{x+y}) \\
&\geq 0\,,
\end{aligned}$$

where the first inequality is by case assumption, and the latter one follows from (6), so that again

$$\mathsf{Err}(f) \geq \tfrac{3}{8} - \tfrac{3}{8}\left(\tfrac{1}{4}\right)^2 = \tfrac{45}{128}\ .$$

■

5 Total degree one testing in characteristic two

Although the main purpose of our work is to give a near optimal analysis of the BLR test, we now describe and analyze a way of testing for total degree one over GF(2). Our purpose is to further illustrate the strength and elegance of the Fourier analysis technique, as well as its more general applicability to the problem of analyzing program testers.

As usual, let $F = \mathrm{GF}(2)$. Note that a total degree one polynomial p is either a linear function or a linear function plus a constant. Thus, since F is of characteristic two, $p(u){+}p(v){+}p(w) = p(u{+}v{+}w)$ for all $u,v,w \in F^n$. The latter is satisfied only if p is of total degree one. In analogy to the case of linearity testing, define

- $\textsc{Deg}_1$ — Set of all polynomials of total degree one from F^n to F
- $\mathsf{Dist}_1(f) \overset{\text{def}}{=} \min\{\, \mathsf{Dist}(f,p) : p \in \textsc{Deg}_1 \,\}$ — Distance of f to its closest polynomial of total degree one.

Again, assume we are given oracle access to a function f mapping F^n to F. We want to test that f is close to a polynomial of total degree 1 from F^n to F, and make as few oracle queries as possible.

THE TOTAL DEGREE 1 TEST. The test is the following — Pick $u,v,w \in F^n$ at random, query the oracle to obtain $f(u), f(v), f(w), f(u{+}v{+}w)$, and reject if $f(u) + f(v) + f(w) \neq f(u{+}v{+}w)$. Let

$$\mathsf{Err}_1(f) \ \overset{\text{def}}{=}\ \Pr_{u,v,w \overset{R}{\leftarrow} F^n}\left[\, f(u) + f(v) + f(w) \neq f(u{+}v{+}w) \,\right]\ ,$$

be the probability that the test rejects f. Also let

$$\textsc{Rej}_1(x) \ \overset{\text{def}}{=}\ \min\{\, \mathsf{Err}_1(f) : f\colon F^n \to F \text{ s.t. } \mathsf{Dist}_1(f) = x \,\}\ .$$

In order to understand how good this test is we need to lower bound $\mathsf{Err}_1(f)$ in terms of $x = \mathsf{Dist}_1(f)$. The techniques discussed in this work gives us tools for achieving this goal. In fact, applying these techniques we will show that if $h(\cdot) = (-1)^{f(\cdot)}$ (f viewed as a real valued function), then $|h_\alpha| \le 1 - 2x$, for all α in F^n. Indeed, note that all functions in $\textsc{Deg}_1$ are of the form $l_\alpha(\cdot) + \beta$, where β is in F and l_α denotes the function that sends u to $\langle \alpha, u\rangle = \sum_{i=1}^{n} \alpha_i u_i$ (arithmetic over F). Then, as in Lemma 2.1, we have that $\widehat{h}_\alpha = 1 - 2\,\mathsf{Dist}(f, l_\alpha) \le 1 - 2\,x$. Moreover, since $\mathsf{Dist}(f, l_\alpha) + \mathsf{Dist}(f, l_\alpha + 1) = 1$, we also have that $\widehat{h}_\alpha = 2\,\mathsf{Dist}(f, l_\alpha + 1) - 1 \ge 2\,x - 1$, which proves the claim.

Arguing as in the proofs of Lemma 2.2 and Theorem 1.2 yields

$$\mathsf{Err}_1(f) = \tfrac{1}{2}\left(1 - (h * h * h * h)(0)\right) = \tfrac{1}{2}\left(1 - \textstyle\sum_{\alpha\in F^n} (\widehat{h}_\alpha)^4\right) .$$

Hence, the previously derived bound on the absolute value of the Fourier coefficients of h and Parseval's equality imply that

$$\mathsf{Err}_1(f) \ge \tfrac{1}{2}\left(1 - (1-2\,x)^2 \sum_{\alpha\in F^n} (\widehat{h}_\alpha)^2\right) = 2\,x\,(1-x) .$$

Finally, note that since f takes values over GF(2), then $f(u){+}f(v){+}f(w){\ne}f(u{+}v{+}w)$ if and only if f differs from every $p \in \textsc{Deg}_1$ in exactly one of the points $\{u, v, w, u{+}v{+}w\}$, or in exactly three of the points $\{u, v, w, u{+}v{+}w\}$. This observation leads to a generalization of Lemma 2.3 that allows to show that $\mathsf{Err}_1(f) \ge 8\,x\,(1-x)\,(\frac{1}{2} - x)$.

We have shown the following:

Lemma 5.1 $\textsc{Rej}_1(x) \ge \max\left\{\ 8\,x\,(1-x)\,(\tfrac{1}{2}-x)\ ,\ 2\,x\,(1-x)\ \right\}$.

Acknowledgments

J. H. thanks Mike Sipser for making his visit to MIT possible. M. K. thanks Dan Kleitman, Carsten Lund, Mike Sipser, and Dan Spielman for several interesting and helpful discussions. We thank Sanjeev Arora and Ronitt Rubinfeld for comments on an earlier draft. Part of this work was done while M.B. was at the IBM T. J. Watson Research Center.

References

[1] N. Alon and J. H. Spencer. The probabilistic method. John Wiley & Sons, Inc., 1992.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy. Proof verification and intractability of approximation problems. Proceedings *of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[3] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. Proceedings *of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[4] L. Babai, L. Fortnow and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, Vol. 1, 3–40, 1991.

[5] L. Babai, L. Fortnow, L. Levin and M. Szegedy. Checking computations in polylogarithmic time. Proceedings *of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

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

[7] M. Bellare, O. Goldreich and M. Sudan. Free bits and non-approximability. Proceedings *of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995.

[8] M. Bellare and M. Sudan. Improved non-approximability results. Proceedings *of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

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

[10] D. Coppersmith. Notes, summer 1990.

[11] U. Feige, S. Goldwasser, L. Lovász, S. Safra and M. Szegedy. Approximating clique is almost NP-complete. Proceedings *of the* 32nd *Symposium on Foundations of Computer Science*, IEEE, 1991.

[12] K. Friedl, Zs. Hátsági and A. Shen. Low-degree testing. Proceedings *of the* 5th *Annual Symposium on Discrete Algorithms*, ACM-SIAM, 1994.

[13] K. Friedl and M. Sudan. Some Improvements to Total Degree Tests. *Proceedings of the Third Israel Symposium on Theory and Computing Systems*, IEEE, 1995.

[14] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. Proceedings *of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

[15] J. Håstad. Testing of the long code and hardness for clique. To appear in Proceedings *of the* 28th *Annual Symposium on Theory of Computing*, ACM, 1996.

[16] M. Kiwi. Probabilistically checkable proofs and the testing of Hadamard-like codes. PhD thesis, Massachusetts Institute of Technology, Cambridge, January 1996.

[17] D. J. Kleitman. Private communication, October 1995.

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

[19] A. Polishchuk and D. Spielman. Nearly Linear Size Holographic Proofs. Proceedings *of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

[20] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials and their applications to program testing. IBM Technical Report RC 19156, 1993. To appear in *SIAM Journal on Computing*.

A BLR test analysis implied by previous work

Consider the function f that takes values from the finite group G into another finite group H. As suggested by [9] we define the function g_f that at $u \in G$ equals the most commonly occurring value in the multiset $\{\, f(u{+}v) - f(v) : v \in G \,\}$ (ties broken arbitrarily). In [9] it is shown that if $\mathsf{Err}(f) < \frac{2}{9}$, then g_f is linear, and for all $v \in G$, $\Pr_{u \overset{R}{\leftarrow} G}[\, g_f(v){=}f(u{+}v) - f(u) \,] > \frac{2}{3}$. Thus,

$$\mathsf{Err}(f) \;\geq\; \mathsf{Dist}(f, g_f) \cdot \Pr_{u,v \overset{R}{\leftarrow} G}[\, f(v){\neq}f(u{+}v) - f(u) \mid g_f(v){\neq}f(v) \,] \;\geq\; \tfrac{2}{3}\,\mathsf{Dist}(f, g_f) \;.$$

In other words, as observed in [8], if $\mathsf{Err}(f) < \frac{2}{9}$, then $\mathsf{Dist}(f) \leq \frac{3}{2}\,\mathsf{Err}(f)$.