Electronic Colloquium on Computational Complexity, Report No. 2 (2014)

Direct Sum Testing

Roee David* Irit Dinur* Elazar Goldenberg* Guy Kindler† Igor Shinkar*

January 6, 2014

Abstract

For a string $a \in \{0,1\}^n$ its *k-fold direct sum encoding* is a function f_a that takes as input sets $S \subseteq [n]$ of size k and outputs $f_a(S) = \sum_{i \in S} a_i$. In this paper we are interested in the Direct Sum Testing Problem, where we are given a function f, and our goal is to test whether f is close to a direct sum encoding, i.e., whether there exists some $a \in \{0,1\}^n$ such that $f(S) = \sum_{i \in S} a_i$ for most inputs S. By identifying the subsets of $[n]$ with vectors in $\{0,1\}^n$ in the natural way, this problem can be thought of as linearity testing of functions whose domain is restricted to the k'th layer of the hypercube.

We first consider the case $k = n/2$, and analyze for it a variant of the natural 3-query linearity test introduced by Blum, Luby, and Rubinfeld (STOC '90). Our analysis proceeds via a new proof for linearity testing on the hypercube, which extends also to our setting.

We then reduce the Direct Sum Testing Problem for general $k < n/2$ to the case $k = n/2$, and use a recent result on *Direct Product Testing* of Dinur and Steurer in order to analyze the test.

*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. Email: {roee.david,irit.dinur,elazar.goldenberg,igor.shinkar}@weizmann.ac.il.

†School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem, Israel. Email: gkindler@cs.huji.ac.il.

ISSN 1433-8092

Contents

1 Introduction

The *k-fold direct sum encoding* of a string $a \in \{0,1\}^n$ is the function $f : \binom{[n]}{k} \to \{0,1\}$ which takes as input subsets $S \subseteq [n]$ of size k, and whose output on such an S is $f(S) = \sum_{i \in S} a_i \pmod 2$. Direct sums were originally considered in theoretical computer science in the famous Yao XOR lemma [Yao82] for the purpose of hardness amplification in circuit complexity, and since then have been extensively studied. They can also potentially be used for gap amplification in PCP constructions, provided that we can devise and analyze local tests for them. We thus naturally arrive at the following problem, which is the focus of this paper.

Direct Sum Testing Problem: Efficiently test whether a given a boolean function $f : \binom{[n]}{k} \to \{0,1\}$ is (close to) a k-fold direct sum encoding.

This question is also very much related to that of testing whether a given function is close to a k-fold tensor power and we elaborate on this in Section 1.2.

If we represent subsets $S \in \binom{[n]}{k}$ by strings $x \in \{0,1\}^n$ of weight k, then the direct sum encoding f_a of a can be written as $f_a(x) = \sum_i x_i a_i \pmod 2$. In other words, f is the restriction of the linear function $x \mapsto \sum_{i \in [n]} a_i x_i \pmod 2$ to the k'th layer of the hypercube, which we denote by

$$L_k^n = \{x \in \{0,1\}^n : |x| = k\}.$$

Another definition for a function f to be linear on the hypercube is that it satisfies $f(x)+f(y) = f(x+y)$ for every pair of inputs x, y. This suggests the natural linearity-test considered in the paper of Blum, Luby, and Rubinfeld [BLR93]: Pick a pair of inputs x, y in the hypercube independently and uniformly at random, and check whether $f(x) + f(y) = f(x + y)$. A linear function clearly passes the test with probability 1, and it was shown in [BLR93] that a function that is far from all linear functions passes the test with probability bounded away from 1. This linearity test is quite fundamental, and appears in many different contexts in computational complexity including PCPs, locally testable codes, and hardness of approximation. The test is well studied, and has many known proofs and generalizations, including to the case where the domain is a group other than the hypercube. Most proofs, however, use the fact that the elements x and y are chosen by the test independently, and moreover, for all x, y their sum $x+y$ always belongs to the domain of the tested function (see Section 1.4 for more details).

In the direct sum testing problem setting we are interested in functions whose domain is L_k^n. Trying to apply the BLR-test on a function $f : L_k^n \to \{0,1\}$ we face an obstacle: if we pick $x, y \in L_k^n$ independently, often $x+y \notin L_k^n$, and so we cannot query f on that input. Even in the case $k = n/2$, where the expected weight of $x+y$ is also $n/2$, $x+y$ actually belongs to $L_{n/2}^n$ only with probability $O(1/\sqrt{n})$. To overcome this our test picks $x, y \in L_{n/2}^n$ randomly, *conditioned on* $x + y$ *being in* $L_{n/2}^n$. Since x and y are no longer independent and the domain does not have a group structure, the known linearity testing proofs do not seem to work for this setting.

The following is a formal definition of our direct sum test T_k^n for parameters n and k, where k is assumed to be even and bounded by $n/2$.

The Direct Sum Test - T_k^n
Given an oracle access to a function $f : L_k^n \to \{0,1\}$ do:

1. Pick $x, y \in L_k^n$ uniformly at random conditioned on $x + y \in L_k^n$.
2. Accept if and only if $f(x) + f(y) = f(x + y)$.

Step 1 of the test can be implemented by picking first $x \in L_k^n$ uniformly at random, then choosing $k/2$ coordinates inside x, and $k/2$ coordinates outside x uniformly at random, and

setting y to be 1 on these k coordinates and 0 elsewhere. Note that the test T_k^n only makes sense for even values of k, since otherwise it is impossible that x, y, and $x+y$ all belong to L_k^n.

Theorem 1.1 (Main Theorem)**.** *Let $n \in \mathbb{N}$, and let $k \leq n/2$ be an even integer. For any function $f : L_k^n \to \{0,1\}$ the following holds.*

1. *If f is linear, then T_k^n accepts with probability 1.*
2. *For all $\varepsilon > 0$ if $\Pr[T_k^n$ accepts $f] > 1 - \varepsilon$, then there exists a string $a \in \{0,1\}^n$ such that $\Pr_{x \in L_k^n}[f(x) = \sum_{i \in [n]} a_i x_i] > 1 - \delta$, where $\delta = \delta(\varepsilon)$ is such that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.*

1.1 Motivation - hardness amplification

The direct sum encoding was first considered in theoretical computer science in the context of hardness amplification for boolean circuits. Yao's XOR lemma [Yao82] (see also [GNW95]) shows that if a function is slightly hard to compute then its direct sum encoding is significantly harder to compute. In other words, the direct sum encoding amplifies the hardness of a function. Hardness amplification has been the subject of much research (see, e.g., [STV01, O'D02, HVV06, Tre03, IJKW10]). A closely related building block in the context of hardness amplification is the *direct product encoding*. The direct product encoding of a string $a \in \{0,1\}^n$ is a function $DP_a : \binom{[n]}{k} \to \{0,1\}^k$ that gets as input a k-element subset S and outputs $DP_a(S) := a|_S$.

The area of PCPs and hardness of approximation is another setting where hardness amplification is well studied. Here we deal with optimization problems, and the parameter that is amplified is the gap between the optimal value in the 'yes' and the 'no' cases. In these questions direct products play an important role. The celebrated parallel repetition theorem of Raz [Raz] shows that very strong amplification can be obtained by applying the direct product operation to games. Dinur's [Din07] gap amplification proof of the PCP theorem [AS98, ALM+98] proceeds by repeatedly performing a (derandomized, or punctured) direct product encoding.

In the aforementioned PCP constructions (as well as in other constructions that involve direct products [DR06, IKW09, DM11]) the analysis involves a so-called *direct product test*. Roughly speaking, a direct product test works by picking two intersecting sets, and checking that the function is consistent on their intersection. The analysis of such tests is far from trivial, and there has been a line of work investigating this [GS00, DR06, DG08, IKW09, DM11, DG10, DS13], especially in relation to PCP constructions.

Having reached a reasonable understanding of direct product tests it now seems possible to move to study direct sum tests, where a major motivation is that of alphabet reduction. Indeed, the large size of the alphabet in the direct product encoding makes it less useful in hardness amplification, and arguably the simplest way to reduce the alphabet size is XORing the entries of the output, resulting in the direct sum encoding. For example, in the gap amplification proof of the PCP theorem [Din07] each direct product step is followed by an alphabet reduction step that is rather complicated. If one were to replace the direct product by a direct sum, this step could potentially be avoided, thus leading to a significant simplification, as well as the potential of improving the parameters of PCP constructions. In order to obtain such constructions we need to devise an efficient test that checks whether a given function is (close to) a direct sum encoding.

Another important related question is that of understanding how the value of a multi-player game behaves under the direct sum operation. Here we imagine a twist on the parallel repetition of games setting, where the players are required to output the XOR of their answers (rather than their concatenation, as in the classical parallel repetition setup). This question is analogous to direct sum testing in a similar way that parallel repetition is analogous to direct product testing. The direct sum operation is particularly meaningful for XOR games, and is the core in a recent

breakthrough work of Chan [Cha13], who constructs a PCP with optimal amortized free bit complexity.

The PCP motivation also drives our quest to find a direct sum test that makes the absolute minimal number of queries, namely three. The fewer queries a test makes, the more useful it is for combination with other gadgets, leading to stronger inapproximability results.

1.2 Tensor Power Testing

The direct sum encoding is very much related to the tensor power operation, and our results imply a testing result for deciding whether a given function is a tensor power.

A function $f : [n]^k \to \{-1, 1\}$ is a *tensor power* if there is a function $b : [n] \to \{-1, 1\}$ such that $f = b^{\otimes k}$, i.e., $f(z) = \prod_{i=1}^{k} b(z_i)$ for all $z \in [n]^k$. It is a *tensor product* if there are k (possibly distinct) functions $b_1, \ldots, b_k : [n] \to \{-1, 1\}$ such that $f = b_1 \otimes \cdots \otimes b_k$.

One can see that by moving between $\{-1, 1\}$ notation and $\{0, 1\}$ notation the tensor power and the direct sum operations are very similar. Indeed, the only difference is that in the direct sum we consider k-element subsets $S \subset [n]$ whereas in the tensor product we consider k-tuples. When $k \ll \sqrt{n}$ this difference is negligible, which implies the testing results for tensor power. We suggest and analyze the following three query test.

The Tensor Power Test - TP_k^n
Given an oracle access to a function $f : [n]^k \to \{-1, 1\}$ do:

1. Pick $u, v, w \in [n]^{k/2}$ independently at random.
2. Pick three permutations $\pi_1, \pi_2, \pi_3 : [k] \to [k]$ independently at random.
3. Accept if and only if $f(uv \circ \pi_1) \cdot f(vw \circ \pi_2) = f(uw \circ \pi_3)$.

In the test above for a k-tuple $z \in [n]^k$ and a permutation $\pi : [k] \to [k]$ the notation $z \circ \pi$ denotes the k-tuple permuted by π, namely, $z \circ \pi = (z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(k)})$.

We prove the following theorem by reduction to Theorem 1.1.

Theorem 1.2. *Suppose $n, k \in \mathbb{N}$ and $\varepsilon > 0$ are such that $k^2/n = o(\varepsilon)$. Let $f : [n]^k \to \{-1, 1\}$ be a function that passes the test TP_k^n with probability at least $1 - \varepsilon$. Then there is some $b : [n] \to \{-1, 1\}$ such that*

$$\Pr_{z \in [n]^k}[f(z) = b^{\otimes k}(z) = \prod_{i=1}^{k} b(z_i)] \geq 1 - O(\varepsilon).$$

1.3 Technical contribution

Our proof of Theorem 1.1 first handles the case $k = n/2$, and then derives the result for $k < n/2$ via a reduction to $k = n/2$. For $k = n/2$ we have the following result.

Theorem 1.3 (Direct Sum Testing for $k = n/2$)**.** *Let $n \in \mathbb{N}$ be such that $n \equiv 0 \pmod 4$, and let $\varepsilon > 0$. For all functions $f : L_{n/2}^n \to \{0, 1\}$, if $\Pr[f(x) + f(y) = f(x + y)] > 1 - \varepsilon$ then there exists a string $a \in \{0, 1\}^n$ such that $\Pr_{x \in L_{n/2}^n}[f(x) = \sum_{i \in [n]} a_i x_i] > 1 - \delta$, where $\delta = \delta(\varepsilon) = \frac{\varepsilon}{3} \cdot (1 + o_n(1)) + O(\varepsilon^2)$.*[1]

In fact, we prove a 3-functions version of the above theorem: Given three functions $f_1, f_2, f_3 : L_{n/2}^n \to \{0, 1\}$ the test picks $x, y \in L_{n/2}^n$ with the same distribution as in $T_{n/2}^n$, and checks that $f_1(x) + f_2(y) = f_3(x + y)$.

[1] We denote by $o_n(1)$ a function that tends to zero as n tends to infinity.

Theorem 1.4. *Let $n \in \mathbb{N}$ be such that $n \equiv 0 \pmod 4$, and let $\varepsilon > 0$. Given three functions $f_1, f_2, f_3 : L^n_{n/2} \to \{0,1\}$ if $\Pr[f_1(x) + f_2(y) = f_3(x+y)] > 1 - \varepsilon$, then there exists some $i \in \{1,2,3\}$ and a string $a \in \{0,1\}^n$ such that $\Pr_{x \in L^n_{n/2}}[f_i(x) = \sum_{i \in [n]} a_i x_i] > 1 - O(\varepsilon)$.*

This theorem clearly implies Theorem 1.3 with weaker parameters (which we will fix later in the proof). We prove Theorem 1.4 by giving a new analysis for the BLR test on the entire hypercube and generalizing it for L^n_k with $k = n/2$. We give more details below.

Reducing Theorem 1.1 for $k < n/2$ to the case $k = n/2$: First, let us describe how the case $k < n/2$ in Theorem 1.1 is obtained by reduction to $k = n/2$. The key of the reduction is to notice that T^n_k actually performs T^{2k}_k on a random subset $u \subset [n]$ of size $2k$. That is, T^n_k is equivalent to a test that first chooses a random set u of $2k$ coordinates, sets x and y to be zero on coordinates outside of u, and on the u coordinates chooses them according to the distribution used by T^{2k}_k.

If a function f passes the test with probability close to 1, then for most choices of u the test passes with high probability when conditioned on the selection of u. By the $n/2$ case (namely Theorem 1.3) for each such u the restriction of f to inputs that are contained in u is close to some linear function $\phi^{(u)}$, i.e., there is a $2k$-string $\sigma^{(u)}$ such that $f(x) = \sum_{j \in u} \sigma^{(u)}_j x_j$ for most such x's. We then show that these "local" linear functions can be stitched together to a "global" linear function ϕ, by finding a global string $a \in \{0,1\}^n$ such that $\sigma^{(u)} = a|_u$. This is done by first showing that for most u, u' the strings $\sigma^{(u)}, \sigma^{(u')}$ are consistent on their common coordinates. Then, using a recent result by of Dinur and Steurer [DS13] on direct product testing, we conclude that these local consistencies between $\sigma^{(u)}$ and $\sigma^{(u')}$ imply the existence of such a global string. This implies existence of a "global" linear function $\phi : L^n_k \to \{0,1\}$ that is close to f.

A new analysis of linearity testing: Below we outline our analysis of linearity testing on the hypercube, and then explain its extension to the direct sum testing. Let $f : \{0,1\}^n \to \{0,1\}$ be a function such that $\Pr_{x,y}[f(x) + f(y) = f(x+y)] > 1 - \varepsilon$, and let δ be the relative distance of f from the nearest linear function. Our goal is to show that $\delta = O(\varepsilon)$.

The proof follows a two step approach. The first step shows a dichotomy: either δ is $O(\varepsilon)$ or it lies in $1/2 \pm O(\varepsilon)$. Indeed, if L is the closest linear function to f and $B_L = \{x \in \{0,1\}^n : f(x) \neq L(x)\}$ is the set of points in which $f(x) \neq L(x)$, the rejection probability of the test can be written as

$$\varepsilon = \Pr[f(x) + f(y) \neq f(x+y)] \geq \Pr[x, y, x+y \in B_L] + 3\Pr[x \in B_L, y, x+y \notin B_L]. \quad (1)$$

The first step follows easily from (1).

It is now left to rule out the possibility of f agreeing with every linear function on $1/2 \pm O(\varepsilon)$ fraction of the domain. This is done in the second step which is carried out by induction on the number of variables n, and works as follows. By averaging, there exist $x_n, y_n \in \{0,1\}$ such that when fixing the last coordinates of x and y to these values the test accepts f with high probability. Fixing the last bit to x_n naturally induces a function $f_1 : \{0,1\}^{n-1} \to \{0,1\}$ defined as $f(x) = f(x \circ x_n)$. Similarly, we define f_2, and f_3 by fixing the last bit to y_n and $x_n + y_n$ respectively. Hence, the functions $f_1, f_2, f_3 : \{0,1\}^{n-1} \to \{0,1\}$ pass the 3-function test with high probability. Our goal is now to prove that (i) If $\Pr[f_1(x) + f_2(y) = f_3(x+y)] > 1 - \varepsilon$, then one of the f_i's is close to a linear function; and (ii) If some f_i is close to a linear function, then f is close to a linear function.

In order to prove (i) we define a function $g : \{0,1\}^{n-1} \to \{0,1\}$ as $g = f_1 + f_2 + f_3$, and show that g passes the linearity test with high probability. By the induction hypothesis g must

be close to some linear function $L : \{0,1\}^{n-1} \to \{0,1\}$. We then show that each of the functions f_1, f_2, f_3 is either $O(\varepsilon)$-close to L or is $O(\varepsilon)$-close to $1 + L$. This implies that at least one of the three functions must be close to L, as otherwise the three functions would pass the test with very low probability

In order to prove (ii) let us suppose for concreteness that f_1 is $O(\varepsilon)$-close to a linear function L. Then f agrees with a linear function L on $(1-O(\varepsilon))$-fraction of the halfspace obtained by the fixing the last bit to x_n. By taking a random extension of L to the entire space $\{0,1\}^n$, we get that f agrees with some linear function in at least $3/4 - O(\varepsilon)$ fraction of the points. Therefore, the agreement of f with this linear function significantly deviates from $1/2$, and hence by the first step f must be $O(\varepsilon)$ close to a linear function.

Functions defined on $L^n_{n/2}$: In the above analysis looked at expressions of the form $\Pr[x \in B, y \in B]$ for some set B (see e.g. (1)). Since in the setting of $L^n_{n/2}$ the vertices x and y are not independent, estimating such a quantity is no longer a straightforward task. Similar estimations are also required in the induction step, although this was not mentioned explicitly in the sketch above.

We estimate this probability by considering the expansion properties of the underlying graph $J_n = (V_n, E_n)$, whose edges are (x, y) chosen by the test. Namely, the vertex set V_n of the graph is $L^n_{n/2}$ and there is an edge between x and y if and only if $x + y \in L^n_{n/2}$. We refer to J_n as the Johnson graph as it closely related to the Johnson scheme. We show that J_n satisfies the following expansion property.

Lemma 1.5. *Let $A \subseteq V_n$ be a subset of the vertices of J_n of size $|A| = \alpha|V|$. Pick an edge $(x, y) \in E_n$ of J_n uniformly at random. Then*

1. $\Pr[x \in A, y \notin A] = \alpha(1-\alpha) \pm \alpha(1-\alpha) \cdot \tilde{O}(n^{-1/4})$.
2. $\Pr[x, y \in A] = \alpha^2 \pm \alpha(1-\alpha) \cdot \tilde{O}(n^{-1/4})$.

This is proven by showing that J_n has very short mixing time. Specifically, we prove that if we start from an arbitrary vertex and make two random steps on the graph, then the distribution of the walk after two steps is $\tilde{O}(n^{-1/2})$-close to the uniform distribution. This is what we would expect to have from a random graph with such degree. We remark that although the spectrum of J_n is well known (see [GGL95]), the bounds on the expansion of J_n obtained from the spectral analysis are not sufficiently tight for our purpose, and we give a bespoke combinatorial analysis for this graph in order to obtain the result.

1.4 Related work

Since the original proof of [BLR93], the linearity test has received a lot of attention, and was extensively studied and generalized. Generalizations include testing linearity for groups other than the hypercube, testing low degree (rather than degree 1), and finding more randomness-efficient tests. See [BCLR08, BSVW03, BCH+96, SW04, AKK+05, KLX07, BKS+10, KS09].

Kopparty and Saraf [KS09] studied linearity testing on the hypercube for a large family of measures on distances between functions. That is, the distance between functions is defined as $\mathrm{dist}(f, g) = \Pr_{x \sim \mu}[f(x) \neq g(x)]$ for some predefined distribution μ, which is not necessarily the uniform measure. In particular, they show a linearity test that works for the distribution μ_p, where each bit of $x \in \{0,1\}^n$ is chosen to be 1 with probability p, and their proof also applies to the setting of L^n_k. A drawback of their test is that it makes $O(n/k)$ queries, and does not look like the natural "BLR-style" 3 query test.

Kaufman and Lubotzky [KL14] recently discovered an intriguing connection between high dimensional expanders and testing, a connection that served as a trigger for this work. They show that expansion of a k-dimensional simplicial complex $V \subseteq L_k^n$ on vertex set $[n]$ is equivalent to testing whether a function $f : V \to \{0,1\}$ is a linear extension of a function defined on L_{k-1}^n, i.e., whether there is some $g : L_{k-1}^n \to \{0,1\}$ such that $f(x) = \sum_{y \subseteq x, |y|=|x|-1} g(y)$ for all $x \in V$. The case of $k = 2$ coincides with our result (because a function $g : L_1^n \to \{0,1\}$ is just an n-bit string), and was analyzed by Linial and Meshulam in [LM06] in the language of simplicial complexes.

Motivated by constructions of short PCPs Ben-Sasson et al. [BSVW03] analyze a linearity test in which, just like in our result, the queries x and y are not independent. Their domain is the hypercube, and their goal was to minimize the number of random bits used by the test. The test works by choosing $x \in \{0,1\}^n$ uniformly at random, choosing $s \in S$ for some $S \subseteq \{0,1\}^n$ of size $n^{O(1)}$ uniformly at random, and setting $y = x + s$. The test accepts if and only if $f(x) + f(y) = f(x+y)$. They show that if S is a small biased set, then this indeed gives a good linearity test.

This idea was later generalized by Shpilka and Wigderson [SW04] to arbitrary groups Γ with generators S of size $|S| = O(1)$, where the Cayley graph $Cay(\Gamma, S)$ is an expander. They showed that the test described above performs nearly as well as the original BLR-test (depending on the expansion of the graph). The main difficulty in their work comes from the fact that x and y are not chosen independently, which is similar to our setting. They overcome this problem using the assumption that the underlying graph is an expander, which implies that if a function f is far from being linear, then the inconsistencies in the $f(x) + f(y) = f(x + y)$ test are "well spread", and hence it rejects such functions with non-negligible probability. Still, in their settings the domain of the function has a group structure, which seems to be crucial in their analysis.

Another natural generalization of linearity testing is checking whether a function is a low-degree polynomial. This was done by Alon et al. in [AKK+05], whose analysis was later improved by Bhattacharyya et al. [BKS+10]. The proof of Bhattacharyya et al. gives a new analysis of linearity test on the hypercube by induction. It seems to differ from our proof, and, in particular, we do not know if their proof can be extended to the setting of $L_{n/2}^n$.

1.5 Comparison with known proofs

In this section we explain why the BLR decoding-style analysis of linearity testing does not extend to our setting.

The combinatorial proofs of linearity testing, such the ones in [BLR93, BCLR08, SW04], take a function $f : \{0,1\}^n \to \{0,1\}$ that passes the linearity test with probability $1 - \varepsilon$, and define a *correction* function $g : \{0,1\}^n \to \{0,1\}$ by letting $g(x) = MAJ_{y \in \{0,1\}^n}\{f(y) + f(x+y)\}$. It is then claimed that for ε small enough the function g is linear. Then, using the fact that $\text{dist}(f, g) < O(\varepsilon)$ it follows that f is $O(\varepsilon)$-close to a linear function.

In our settings, even for the case $k = n/2$ this analysis does not apply, since if we take a function $f : L_{n/2}^n \to \{0,1\}$ that passes the $T_{n/2}^n$ test with high probability, and define the correction function $g : L_{n/2}^n \to \{0,1\}$ analogously, namely,

$$g(x) = MAJ\{f(y) + f(x+y) : y \in L_{n/2}^n \text{ such that } x + y \in L_{n/2}^n\},$$

then we can no longer assure that the function g is linear. To understand why the analysis above cannot work, note that in our setting the correction function g considers for every $x \in L_{n/2}^n$ the "local majority" vote over only a small fraction of the space, namely those vertices that intersect x on exactly $n/4$ of the coordinates. Therefore we cannot expect the global property of f to propagate after one step of majority voting. One could try to make more steps of corrections.

By the expansion properties of the underlying graph J_n this approach could potentially work, but it seems difficult to push through, and our proof takes a different approach.

The multiple-step correction approach looks similar to the work of Shpilka and Wigderson [SW04] discussed earlier in Section 1.4. In their setting every vertex x is tested only with a tiny fraction of the domain (induced by an underlying expander graph G). For every vertex x they considered the "local majority" of x. They use the expansion of G to show that if f passes the test with high probability, then iterating the "local majority" function would converge to the linear function closest to f. However, in order to prove convergence, they first defined a correction function using "global majority", and then prove the "local majority" converges to the same linear function.

1.6 Structure of the paper

We begin by presenting some notations in Section 2. In Section 3 we show a new analysis for the linearity test on the hypercube. In Section 4 we show that every function that passes the T_k^n test with probability 1, is in fact a linear function. In Section 5 we prove our main technical result, namely, Theorem 1.3. This is done by showing how to extend the proof for the hypercube to our setting. In Section 6 we show that the analysis of the T_k^n test for general case can be reduced to the case $k = n/2$, thus proving Theorem 1.1 for general $k < n/2$. In Section 8 we present the analysis of the vertex expansion of the Johnson graph.

We complement the discussion by presenting in Section 9 a different linearity test for functions $f : L_k^n \to \{0,1\}$ that works for all $k \leq n$, and makes $O(\max(\frac{n}{k}, \frac{n}{n-k}))$ queries.

2 Notations and Preliminaries

Notations: Let $n \in \mathbb{N}$. We denote by $L_k^n \subseteq \{0,1\}^n$ the set

$$L_k^n = \{x \in \{0,1\}^n : |x| = k\},$$

and by $L_{EVEN}^n \subseteq \{0,1\}^n$ the set

$$L_{EVEN}^n = \{x \in \{0,1\}^n : |x| \text{ is even}\}.$$

Note that L_{EVEN}^n is a subgroup of $\{0,1\}^n$.

Fact 2.1. *Let $n \in \mathbb{N}$, and let $k < n$. If k is even, then $\text{span}\langle L_k^n \rangle = L_{EVEN}^n$. If k is odd, then $\text{span}\langle L_k^n \rangle = \{0,1\}^n$.*

Definition 2.2. *Let $S \subseteq \{0,1\}^n$ be a subset (S is not necessarily a subgroup). A function $f : S \to \{0,1\}$ is said to be* linear *if there exists $a = (a_1, \ldots, a_n) \in \{0,1\}^n$ such that $f(x) = \sum_{i \in [n]} a_i x_i \pmod 2$ for all $x \in S$.*

In particular, as explained above, a function f is a direct sum if and only if it is a linear functions with domain L_k^n.

Note that if $\text{span}\langle S \rangle \neq \{0,1\}^n$, then the choice of $a \in \{0,1\}^n$ in Definition 2.2 may not be unique.

Fact 2.3. *Let $n \in \mathbb{N}$, and let $k < n$ be even. Suppose that $\phi : L_k^n \to \{0,1\}$ is a linear function. Then, there are precisely two strings $a, a' \in \{0,1\}^n$ such that $\phi(x) = \sum_{i \in [i]} a_i x_i \pmod 2$ and $\phi(x) = \sum_{i \in [n]} a'_i x_i \pmod 2$ for all $x \in L_k^n$. Specifically, the strings a and a' are complements of each other, i.e., $a_i = 1 - a'_i$ for all $i \in [n]$.*

We will also need the following claim on distances between distinct linear functions on L_k^n. Note that unlike in the hypercube settings this claim is not trivial in the L_k^n settings, and depends on k not being to close to 0 or n.

Proposition 2.4. *Let $p \in (0,1)$, let $n \in \mathbb{N}$ be an integer so that $pn \in \mathbb{N}$, and let $k = pn$. Then, for every pair of distinct linear functions $\phi_1 \neq \phi_2 : L_k^n \to \{0,1\}$ it holds that*

$$c \leq \Pr_{x \in L_k^n}[\phi_1(x) \neq \phi_2(x)] \leq 1 - c$$

for some constant $c = c(p) > 0$ that depends only on p.

In particular, for all $k \in (\varepsilon n, (1-\varepsilon)n)$ the distance between two distinct functions of L_k^n is bounded away above zero. We defer the proof to Appendix A.

We will also need the following definition of a δ-tester.

Definition 2.5. *Let $\mathbf{C}$ be a class of functions from a finite domain $\mathbf{D}$ to a finite range Σ. Let $\delta : (0,1] \to (0,1]$ be a function such that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. We say that T is a δ-test for the class $\mathbf{C}$ if*

1. *All functions in $\mathbf{C}$ are accepted by T with probability 1.*
2. *For every $\varepsilon > 0$, any function $f : \mathbf{D} \to \Sigma$ that passes the test T with probability $1 - \varepsilon$ is $\delta(\varepsilon)$-close to $\mathbf{C}$, i.e., there exists some $\phi \in \mathbf{C}$ such that $\Pr_{x \in \mathbf{D}}[f(x) \neq \phi(x)] < \delta(\varepsilon)$.*

Using this definition Theorem 1.1 considers the following class of functions.

Definition 2.6. *Define LIN_k^n to be the class of functions $f : L_k^n \to \{0,1\}$ for which there exists $a \in \{0,1\}^n$ such that for every $x \in L_k^n$ it holds that $f(x) = \sum_{i \in [n]} a_i x_i$.*

Similarly, by identifying k-subsets of $[n]$ with vectors in $\{0,1\}^n$ of weight k, the class of direct product functions can be written as follows.

Definition 2.7. *Define DP_k^n to be the class of functions $F : L_k^n \to \{0,1\}^k$ for which there exists $a \in \{0,1\}^n$ such that $F(x) = a_x$ for every $x \in L_k^n$, where a_x is the substring of a confined to the coordinates in which $x_i = 1$.*

Note that if T is a direct product tester in the sense of Definition 2.5 then for any function $F : L_k^n \to \{0,1\}^k$ that passes the test with high probability there exists a string $a \in \{0,1\}^n$ such that $F(x) \equiv a_x$ for most inputs $x \in L_k^n$. This is as opposed to a more relaxed definition of direct product tester (such as the ones in [DG08, IKW09]), where for most x's the Hamming distance between $F(x)$ and a_x is small.

3 Linearity Testing on the Hypercube

In this section we describe the idea behind the proof of Theorem 1.1 by discussing a simpler setting of the hypercube $\{0,1\}^n$. Our approach gives a new proof for linearity testing on $\{0,1\}^n$. In Section 5 we show how this proof can be modified in order to prove linearity testing in the setting of $L_{n/2}^n$.

Recall, the test gets as an oracle access a function $f : \{0,1\}^n \to \{0,1\}$ and tests whether f is close to a linear function. The test is defined as follows.

BLR Linearity Test on the Hypercube:
Given an oracle access to a function $f : \{0,1\}^n \to \{0,1\}$ do:

1. Select $x, y \in \{0,1\}^n$ independently.
2. Accept if and only if $f(x) + f(y) = f(x+y)$.

Theorem 3.1. *There exists some $\varepsilon_0 > 0$ small enough such that for every $\varepsilon \in (0, \varepsilon_0)$ and for all $n \in \mathbb{N}$ the following holds. Suppose that a function $f : \{0,1\}^n \to \{0,1\}$ passes the BLR test with probability $1 - \varepsilon$. Then f is $(\varepsilon/3 + 8\varepsilon^2/9)$-close to some linear function, namely, there exists a linear function $g : \{0,1\}^n \to \{0,1\}$ such that*

$$\Pr[f(x) \neq g(x)] < \varepsilon/3 + 8\varepsilon^2/9.$$

In fact, we prove a slightly stronger result. Suppose that we are given three functions $f_1, f_2, f_3 : \{0,1\}^n \to \{0,1\}$, and our goal is to check whether the functions are (close to) linear. Consider the following test. Given (an oracle access) to three functions $f_1, f_2, f_3 : \{0,1\}^n \to \{0,1\}$ the test works as follows.

Three Functions Testing Linearity on the Hypercube
Given an oracle access to three functions $f_1, f_2, f_3 : \{0,1\}^n \to \{0,1\}$ do:

1. Select $x, y \in \{0,1\}^n$ independently.
2. Accept if and only if $f_1(x) + f_2(y) = f_3(x+y)$.

Note that if we take a linear function f_1, and let $f_2 = f_3 = f_1 + 1$, then the above test accepts these functions. Therefore, if the test passes with high probability, it does not imply that all functions are close to linear. What we do prove is that at least one of the functions must be close to linear. Specifically, we prove the following theorem.

Theorem 3.2. *There exists some $\varepsilon_0 > 0$ small enough such that for every $\varepsilon \in (0, \varepsilon_0)$ and for all $n \in \mathbb{N}$ the following holds. Let $f_1, f_2, f_3 : \{0,1\}^n \to \{0,1\}$ be three functions. Suppose that they pass the three functions test with probability $1 - \varepsilon$. Then, there is $i \in \{1,2,3\}$ such that f_i is $2\varepsilon + O(\varepsilon^2)$-close to some linear function.*

3.1 First step towards the proof

Towards proving Theorem 3.1 we show first that if a function passes BLR test with probability $1-\varepsilon$, then for every linear function $L : \{0,1\}^n \to \{0,1\}$ it holds that either the distance between f and L is either close to 0 or close to $\frac{1}{2}$.

Lemma 3.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. For every linear function $L : \{0,1\}^n \to \{0,1\}$ let $\delta_L = \mathrm{dist}(L, f)$ be the distance of f from L. If $\Pr[f(x)+f(y) = f(x+y)] = 1 - \varepsilon$, then for every linear function L we have*

- *either $\delta_L \leq \varepsilon/3 + 8\varepsilon^2/9$*
- *or $\frac{1}{2} - (\varepsilon/3 + 8\varepsilon^2/9) \leq \delta_L \leq \frac{1}{2} + \varepsilon$.*

In particular, if $\phi : \{0,1\}^n \to \{0,1\}$ is a linear function such that $\mathrm{dist}(f, \phi) \notin [\frac{1}{2} - (\varepsilon/3 + 8\varepsilon^2/9), \frac{1}{2} + \varepsilon]$, then $\mathrm{dist}(f, \phi) \leq \varepsilon/3 + 8\varepsilon^2/9$.

Proof. Fix a boolean function $f : \{0,1\}^n \to \{0,1\}$, and suppose that it passes linearity test with probability $\Pr[f(x) + f(y) = f(x+y)] = 1 - \varepsilon$. For any linear function $L : \{0,1\}^n \to \{0,1\}$ define $B_L = \{x \in \{0,1\}^n : f(x) \neq L(x)\}$, and let $G_L = \{0,1\}^n \setminus B_L$. Using this notation the

distance between f and L is $\delta_L = \frac{|B_L|}{2^n}$, and the rejection probability of the test on f can be written as

$$\varepsilon = \Pr[\text{Test rejects } f] \geq \Pr[x, y, x+y \in B_L] + 3\Pr[x \in B_L, y, x+y \in G_L]. \tag{2}$$

Since $\Pr[\cdot] \geq 0$, it follow that each of the two terms is smaller than ε. The first term gives us the following bound.

$$\begin{aligned}
\Pr[x, y, x+y \in B_L] &\geq \Pr[x \in B_L] - \Pr[x \in B_L, y \notin B_L] - \Pr[x \in B_L, x+y \notin B_L] \\
&= \Pr[x \in B_L] - 2\Pr[x \in B_L, y \notin B_L] \\
&\geq \delta_L - 2\delta_L(1-\delta_L) \\
&\geq \delta_L(2\delta_L - 1).
\end{aligned}$$

Solving the inequality $\varepsilon \geq \Pr[x, y, x+y \in B_L] \geq \delta_L(2\delta_L - 1)$ for $\delta_L \geq 0$ we get

$$\delta_L \leq \frac{1}{2} + \varepsilon. \tag{3}$$

Similarly, the second term gives us

$$\begin{aligned}
\Pr[x \in B_L, y, x+y \in G_L] &= \Pr[x \in B_L] - \Pr[x \in B_L, y \in B_L] - \Pr[x \in B_L, x+y \in B_L] \\
&= \Pr[x \in B_L] - 2\Pr[x \in B_L, y \in B_L] \\
&\geq \delta_L - 2\delta_L^2 \\
&\geq \delta_L(1 - 2\delta_L).
\end{aligned}$$

Solving the quadratic inequality $\varepsilon \geq 3\Pr[x \in B_L, y, x+y \in G_L] \geq 3 \cdot \delta_L(1-2\delta_L)$ we get

$$\delta_L \leq \varepsilon/3 + 8\varepsilon^2/9 \qquad \text{or} \qquad \delta_L \geq \frac{1}{2} - (\varepsilon/3 + 8\varepsilon^2/9). \tag{4}$$

where we use the fact that $\sqrt{1-8\varepsilon/3} \geq 1 - 4\varepsilon/3 - 32\varepsilon^2/9$ holds for all $\varepsilon \in [0, 1/8]$. Lemma 3.3 follows by combining Equation (3) with Equation (4). □

3.2 Proof of Theorems 3.1 and 3.2

In this section we prove Theorems 3.1 and 3.2.

Proof. We prove both Theorems 3.1 and 3.2 by induction of n. For the base case for Theorem 3.1 note that if $n \leq \frac{\log(1/\varepsilon)}{2}$, then $\varepsilon < 2^{-2n}$, and hence for every $x, y \in \{0,1\}^n$ it holds that $f(x)+f(y) = f(x+y)$. This implies that f is a linear function. For the induction step we prove the following two lemmas.

Lemma 3.4. *Suppose that the statement of Theorem 3.1 holds for $n-1$. Then, the statement of Theorem 3.2 holds for $n-1$.*

Lemma 3.5. *Suppose that for some $n \in \mathbb{N}$ the statement of Theorem 3.2 holds for $n-1$. Then, the statement of Theorem 3.1 holds for n.*

Therefore, in order to prove Theorems 3.1 and 3.2 it is enough to prove the two foregoing lemmas. □

Proof of Lemma 3.4. We prove the lemma for the n-dimensional hypercube, and not for $n-1$ as stated in the lemma.

Let $f_1, f_2, f_3 : \{0,1\}^n \to \{0,1\}$ be three functions and suppose that $\Pr[f_1(x) + f_2(y) = f_3(x+y)] \geq 1-\varepsilon$. Define a function $g : \{0,1\}^n \to \{0,1\}$ as $g(x) = f_1(x) + f_2(x) + f_3(x)$. Note first that g is close to a linear function. Indeed,

$$\begin{aligned}
\Pr[g(x)+g(y) \neq g(x+y)] &= \Pr\left[\begin{array}{ccccc} & f_1(x)+f_1(y) & & & f_1(x+y) \\ + & f_2(x)+f_2(y) & \neq & + & f_2(x+y) \\ + & f_3(x)+f_3(y) & & + & f_3(x+y) \end{array}\right] \\
&= \Pr\left[\begin{array}{ccccc} & f_1(x)+f_2(y) & & & f_3(x+y) \\ + & f_2(x)+f_3(y) & \neq & + & f_1(x+y) \\ + & f_3(x)+f_1(y) & & + & f_2(x+y) \end{array}\right] \\
&\leq 3\Pr[f_1(x)+f_2(y) \neq f_3(x+y)] \\
&\leq 3\varepsilon,
\end{aligned}$$

and thus, by Theorem 3.1 for n the function g is $(\varepsilon + 8\varepsilon^2)$-close to some linear function.

Claim 3.6. *Let ϕ be a linear function such that $\mathrm{dist}(g,\phi) \leq \varepsilon + 8\varepsilon^2$. Then, the function f_1 is $(2\varepsilon + O(\varepsilon^2))$-close to either ϕ or to $\phi+1$, where $O()$ hides some absolute constant.*

Proof. We note first that $\Pr[f_1(x) + f_1(y) = g(x+y)] \geq 1 - 3\varepsilon$. Indeed

$$\begin{aligned}
\Pr[f_1(x)+f_1(y) \neq g(x+y)] &\leq \Pr[(f_2(y)+f_3(x+y)) + (f_3(x)+f_2(x+y)) = g(x+y)] + 2\varepsilon \\
&\leq \Pr[f_2(y)+f_3(x) = f_1(x+y)] + 2\varepsilon \\
&\leq 3\varepsilon.
\end{aligned}$$

Since $\phi : \{0,1\}^n \to \{0,1\}$ is a linear function such that $\mathrm{dist}(g,\phi) \leq \varepsilon + 8\varepsilon^2$, it follows that

$$\Pr_{x,y}[f_1(x) + \phi(x) = f_1(y) + \phi(y)] \geq 1 - 4\varepsilon - 8\varepsilon^2. \tag{5}$$

We claim that Equation (5) implies that f_1 is either close to ϕ or close to $\phi+1$. Indeed, let $f_1' : \{0,1\}^n \to \{0,1\}$ be defined as $f_1' = f_1 + \phi$. Then

$$\Pr[f_1'(x) = f_1'(y)] \geq 1 - 4\varepsilon - 8\varepsilon^2.$$

Therefore, by the "collision probability" argument f' is close to a constant function. Indeed, if we denote $p = \Pr[f_1'(x) = 1]$, then $p^2 + (1-p)^2 \geq 1 - 4\varepsilon - 8\varepsilon^2$, which implies that either $p \leq 2\varepsilon + O(\varepsilon^2)$ or $p \geq 1 - (2\varepsilon + O(\varepsilon^2))$. Therefore, f_1' is $(2\varepsilon + O(\varepsilon^2))$-close to a constant function, and hence f_1 is $(2\varepsilon + O(\varepsilon^2))$-close to either ϕ or to $\phi+1$. □

Similarly, the function f_2 and f_3 are also close to either ϕ or to $\phi+1$. It is left to prove that one of the f_i's must be linear. Indeed, if all f_i's were $2\varepsilon + C\varepsilon^2$ close to $\phi+1$, then

$$1 - \varepsilon \leq \Pr[f_1(x)+f_2(y) = f_3(x+y)] \leq \Pr[\phi(x)+\phi(y) = \phi(x+y)+1] + 3(2\varepsilon + C\varepsilon^2) = 3(2\varepsilon + C\varepsilon^2)$$

contradicting the assumption that ε is sufficiently small. Therefore, there must be some $i \in \{1,2,3\}$ such that f_i is $(2\varepsilon + O(\varepsilon^2))$-close to the linear function ϕ. This completes the proof of Lemma 3.4. □

We now turn to proof of Lemma 3.5.

Proof of Lemma 3.5. Let $f : \{0,1\}^n \to \{0,1\}$ be a linear function, and suppose that $\Pr[f(x) + f(y) = f(x+y)] \geq 1 - \varepsilon$. By averaging there are some bits $x_n, y_n \in \{0,1\}$ such that if we pick $x', y' \in \{0,1\}^{n-1}$ independently then

$$\Pr[f(x' \circ x_n) + f(y' \circ y_n) = f((x'+y') \circ (x_n + y_n))] \geq 1 - \varepsilon, \tag{6}$$

where $\circ$ denotes the concatenation of strings. Define three functions $f_1, f_2, f_3 : \{0,1\}^{n-1} \to \{0,1\}$ be letting

$$f_1(x') = f(x \circ x_n) \qquad f_2(x') = f(x' \circ y_n) \qquad f_3(x') = f(x' \circ (x_n + y_n))$$

Then, by Equation (6) we have

$$\Pr[f_1(x') + f_2(y') = f_3(x'+y')] \geq 1 - \varepsilon.$$

By the hypothesis of the lemma it follows that one of the functions f_i is $(2\varepsilon + O(\varepsilon^2))$-close to a linear function $\phi : \{0,1\}^{n-1} \to \{0,1\}$. Let us assume for concreteness that f_1 is this function. Our goal is to extend ϕ to an affine function ψ on $\{0,1\}^n$ so that $\Pr[f(x) = \psi(x)] \geq 3/4 - O(\varepsilon)$. Then, by Lemma 3.3 it will follow that ψ is linear, and f is close to ψ.

Let us assume first that $x_n = 0$. Define ψ randomly by choosing a random bit $b \in \{0,1\}$ and letting $\psi(x_1, \ldots, x_n) = \phi(x_1, \ldots, x_{n-1}) + b \cdot x_n$. Note that ψ agrees with ϕ on the subspace $\{x \in \{0,1\}^n : x_n = 0\}$, and for every $x \in \{0,1\}^n$ such that $x_n = 1$ it holds that $\Pr[\psi(x) = f(x)] = \frac{1}{2}$. Therefore, the expected agreement of f with ψ is

$$\mathbb{E}[\mathrm{agr}(f, \psi)] \geq \frac{1}{2}(\mathrm{agr}(f_1, \psi)) + \frac{1}{2} \cdot \frac{1}{2} \geq 3/4 - O(\varepsilon).$$

The case $x_n = 1$ is similar. Define ψ randomly by choosing a random bit $b \in \{0,1\}$ and letting $\psi(x_1, \ldots, x_n) = \phi(x_1, \ldots, x_{n-1}) + b \cdot (1 + x_n)$. Similarly, ψ agrees with ϕ on the affine subspace $\{x \in \{0,1\}^n : x_n = 1\}$, and for every $x \in \{0,1\}^n$ such that $x_n = 0$ it holds that $\Pr[\psi(x) = f(x)] = \frac{1}{2}$. Hence the expected agreement of f with ψ is

$$\mathbb{E}[\mathrm{agr}(f, \psi)] \geq \frac{1}{2}(\mathrm{agr}(f_1, \psi)) + \frac{1}{2} \cdot \frac{1}{2} \geq 3/4 - O(\varepsilon).$$

Therefore, if ε is sufficiently small, then by Lemma 3.3 the function ψ is linear, and $\mathrm{dist}(f, \psi) \leq \varepsilon/3 + 8\varepsilon^2/9$. The lemma follows. □

4 Proof of Theorem 1.1 for $\varepsilon = 0$

Before proving Theorem 1.1 we should first convince ourselves that Theorem 1.1 holds in even the simplest case $\varepsilon = 0$.

Proposition 4.1. *Let $n \in \mathbb{N}$ and let $k \leq n$ be even. Let $f : L_k^n \to \{0,1\}$ be a boolean function. Then f passes the T_k^n test with probability 1 if and only if f is a linear function.*

Proof. Clearly, every linear function passes the test with probability 1. For the other direction, let H be the set of all functions that pass the T_k^n test with probability 1. In order to prove the proposition we define an injective mapping $\psi : H \to \{0,1\}^{n-1}$. This implies that $|H| \leq 2^{n-1}$. By Fact 2.1 we have $\mathrm{span}\langle L_k^n \rangle = L_{EVEN}^n$, and hence there are 2^{n-1} distinct linear functions on L_k^n. This concludes the proof of Proposition 4.1.

In order to define a mapping ψ let us pick for each $i \in [n-1]$ an arbitrary vector $z^{(i)} \in L_k^n$ such that $z_i^{(i)} = 1$ and $z_{i+1}^{(i)} = 0$. Given a function $f \in H$, define $\psi(f) \in \{0,1\}^{n-1}$ by setting $(\psi(f))_i = f(z^{(i)}) + f(z^{(i)} + v_i)$, where $v_i = e_i + e_{i+1}$.[2] By the discussion above, it is enough to prove the following lemma.

[2] ψ is well defined since $z^{(i)} + v_i \in L_k^n$.

Lemma 4.2. *Let $\psi : H \to \{0,1\}^{n-1}$ be as above. Then, the mapping ψ is one-to-one.*

Proof of Lemma 4.2. It is easy to see that the mapping ψ is a homomorphism from H to $\{0,1\}^{n-1}$, that is, $\psi(f) + \psi(g) = \psi(f+g)$ for all $f, g \in H$. Thus, in order to prove that ψ is an injection, it is enough to show that the kernel of ψ is trivial.

Let $f \in H$ such that $\psi(f) = 0$. We claim that f is a constant function, and hence it must be the constant zero function (since the constant 1 function does not pass T_k^n with probability 1). A first step in this direction is the following claim which says that the definition of the mapping ψ is, in fact, independent of the choices of $z^{(i)}$.

Claim 4.3. *Let $i \in [n-1]$. For every $x, y \in L_k^n$ such that $x_i = y_i = 1$ and $x_{i+1} = y_{i+1} = 0$ it holds that $f(x) + f(x+v_i) = f(y) + f(y+v_i)$.*

Proof. For $i \in [n-1]$ define a graph $G_i = (V_i, E_i)$, where $V_i = \{x \in L_k^n : x_i = 1, x_{i+1} = 0\}$, and $(x,y) \in E_i$ if and only if $x + y \in L_k^n$. It is easy to check that the graph G_i is connected, and so it is enough to prove that $f(x) + f(x+v_i) = f(y) + f(y+v_i)$ for every two neighboring vertices $(x,y) \in E_i$. Indeed, since $x + y \in L_k^n$ we have

$$f(x) + f(x+v_i) = (f(x) + f(x+y)) + (f(x+y) + f(x+v_i)) = f(y) + f(y+v_i),$$

where the equalities $f(x) + f(x+y) = f(y)$ and $f(x+y) + f(x+v_i) = f(y+v_i)$ follows from the assumption that f passes the T_k^n test with probability 1. The claim follows. □

We also make the following observation.

Observation 4.4. *Let $x^{(0)} \in L_k^n$ be the vector such that $x_i^{(0)} = 1$ for $i \leq k$ and $x_i^{(0)} = 0$ for $i > k$. Then, for every $y \in L_k^n$ there is a sequence $x^{(1)}, \ldots, x^{(t)} \in \{0,1\}^n$ such that $x^{(j+1)} = x^{(j)} + v_{i_j}$ for some $i_j \in [n-1]$ and $x^{(t)} = y$.*

We are now ready to prove Lemma 4.2. By Claim 4.3 for every $i \in [n-1]$ and for every $x \in L_k^n$ such that $x_i = 1$ and $x_{i+1} = 0$ it holds that $f(x) + f(x+v_i) = f(z^{(i)}) + f(z^{(i)} + v_i) = 0$. Therefore, by the assumption that $\psi(f) = 0$ we get that $f(x) = f(x+v_i)$ for all $x \in L_k^n$ such that $x_i = 1$ and $x_{i+1} = 0$. By Observation 4.4 it follows that $f(x) = f(y)$ for every $x, y \in L_k^n$, i.e., f is a constant function, and so must be the constant 0 function. □

This completes the proof of Proposition 4.1. □

5 Proof of Theorem 1.3: Direct Sum Testing for $k = n/2$

In this section we prove Theorem 1.3. Recall that for n divisible by 4 the $T_{n/2}^n$ test is defined as follows.

$T_{n/2}^n$ - Direct Sum Test for $k = n/2$:
Given an oracle access to $f : L_{n/2}^n \to \{0,1\}$ do:

1. Pick $x, y \in L_{n/2}^n$ uniformly at random so that $x + y \in L_{n/2}^n$.
2. Accept if and only if $f(x) + f(y) = f(x+y)$.

Theorem 1.3 restated: *Let $n \in \mathbb{N}$ be such that $n \equiv 0 \pmod 4$, and let $\varepsilon > 0$. For all functions $f : L^n_{n/2} \to \{0,1\}$ if $\Pr[f(x)+f(y)=f(x+y)] > 1-\varepsilon$, then there exists a string $a \in \{0,1\}^n$ such that $\Pr_{x\in L^n_k}[f(x) = \sum_{i\in[n]} a_i x_i] > 1-\delta$, where $\delta = \delta(\varepsilon) = \frac{\varepsilon}{3}\cdot(1+O(\sqrt{\gamma_n}))+O(\varepsilon^2)$, and $\gamma_n = \tilde{O}(n^{-1/2})$ is the quantity from Lemma 1.5.*

As explained in Section 1.3 we prove a stronger three functions version of this theorem.

Direct Sum Test for three functions:
Given an oracle access to three functions $f_1, f_2, f_3 : L^n_{n/2} \to \{0,1\}$ do:

1. Select $x, y \in L^n_{n/2}$ such that $x + y \in L^n_{n/2}$.
2. Accept if and only if $f_1(x) + f_2(y) = f_3(x+y)$.

Theorem 1.4 restated: *Let $n \in \mathbb{N}$ be such that $n \equiv 0 \pmod 4$, and let $\varepsilon > 0$. For all functions $f_1, f_2, f_3 : L^n_{n/2} \to \{0,1\}$ if $\Pr[f_1(x)+f_2(y)=f_3(x+y)] > 1-\varepsilon$, then there exists $i \in \{1,2,3\}$ and a string $a \in \{0,1\}^n$ such that $\Pr_{x\in L^n_k}[f_i(x) = \sum_{i\in[n]} a_i x_i] > 1-\delta$, where $\delta = \delta(\varepsilon) = 4\varepsilon + O(\varepsilon^2)$.*

The proof is very similar to the proof of Theorems 3.1 and 3.2 from Section 3

Proof of Theorem 1.3. We prove Theorem 1.3 by induction of n, where in each step we increase n by 4, since the theorem assumes that $n \equiv 0 \pmod 4$. The base case $n \le \frac{\log(1/\varepsilon)}{2}$ holds by Proposition 4.1. For the induction step we prove the following two lemmas.

Lemma 5.1. *Suppose that the statement of Theorem 1.3 holds for some $n-4 \in \mathbb{N}$. Then, the statement of Theorem 1.4 holds for $n-4$.*

Lemma 5.2. *Suppose that for some $n \in \mathbb{N}$ the statement of Theorem 1.4 holds for $n-4$. Then, the statement of Theorem 1.3 holds for n.*

These two lemmas, clearly, prove Theorem 1.3. □

The rest of this section is devoted to proving Lemma 5.1 and Lemma 5.2. But first we prove an analogue of Lemma 3.3, saying that every function f that passes the test with probability close to 1 is either close to being linear or has distance close to 1/2 from every linear function.

Lemma 5.3. *Let $n \in \mathbb{N}$ be such that $n \equiv 0 \pmod 4$. Let $f : L^n_{n/2} \to \{0,1\}$ be a boolean function. For every linear function $L : L^n_{n/2} \to \{0,1\}$ let $\delta_L = \mathrm{dist}(L,f)$ be the distance of f from L. If f passes the $T^n_{n/2}$ test with probability $1-\varepsilon$, then for every linear function L we have*

- *either $\delta_L \le \frac{\varepsilon}{3}\cdot(1+O(\sqrt{\gamma_n}))+O(\varepsilon^2)$*
- *$\frac{1}{2} - \frac{\varepsilon}{3}\cdot(1+O(\sqrt{\gamma_n})) - O(\varepsilon^2) \le \delta_L \le \frac{1}{2}+\varepsilon+O(\sqrt{\gamma_n})$.*

In particular, if $\phi : L^n_{n/2} \to \{0,1\}$ is a linear function such that $\mathrm{dist}(f,\phi)$ significantly deviates from $\frac{1}{2}$, then f is $O(\varepsilon)$-close to ϕ.

Proof. The proof follows the lines of the proof of Lemma 3.3. The only difficulty comes from the dependence between the choices of x and y. We overcome the difficulty by using Lemma 1.5.

Similarly to the proof of Lemma 3.3 for every linear function $L : L^n_{n/2} \to \{0,1\}$ we define $B_L = \{x \in L^n_{n/2} : f(x) \ne L(x)\}$, and let $G_L = L^n_{n/2} \setminus B_L$. Using this notation the distance between f and L is $\delta_L = \frac{B_L}{\binom{n}{n/2}}$, and the rejection probability of the test on f is equal to

$$\varepsilon \ge \Pr[f(x)+f(y) \ne f(x+y)] \ge \Pr[x,y,x+y \in B_L] + 3\Pr[x \in B_L, y, x+y \in G_L]. \quad (7)$$

Since $\Pr[\cdot] \geq 0$, it follow that each of the two terms is smaller than ε. We use Lemma 1.5 to bound from below each term on the RHS separately. The first term is lower bounded as follows.

$$\begin{aligned}
\Pr[x, y, x+y \in B_L] &\geq \Pr[x \in B_L] - \Pr[x \in B_L, y \notin B_L] - \Pr[x \in B_L, x+y \notin B_L] \\
&= \Pr[x \in B_L] - 2\Pr[x \in B_L, y \notin B_L] \\
[\text{Using Lemma 1.5}] &\geq \delta_L - 2\delta_L(1-\delta_L) - 4\delta_L(1-\delta_L)\sqrt{\gamma_n}.
\end{aligned}$$

Solving the quadratic inequality $\varepsilon \geq \Pr[x, y, x+y \in B_L] \geq \delta_L(2\delta_L - 1) - 4\delta_L(1-\delta_L)\sqrt{\gamma_n}$ for $\delta_L \geq 0$ we get

$$\delta_L \leq \frac{1}{2} + \varepsilon + O(\sqrt{\gamma_n}). \tag{8}$$

Similarly, the second term gives us

$$\begin{aligned}
\varepsilon/3 \geq \Pr[x \in B_L, y, x+y \in G_L] &= \Pr[x \in B_L] - \Pr[x \in B_L, y \in B_L] - \Pr[x \in B_L, x+y \in B_L] \\
&= \Pr[x \in B_L] - 2\Pr[x \in B_L, y \in B_L] \\
[\text{Using Lemma 1.5}] &\geq \delta_L - 2\delta_L^2 - 4\delta_L(1-\delta_L)\sqrt{\gamma_n}.
\end{aligned}$$

Solving the quadratic inequality $\varepsilon/3 \geq (\delta_L - 2\delta_L^2 - 2\delta_L(1-\delta_L)\sqrt{\gamma_n})$ we get

$$\delta_L \leq \frac{\varepsilon}{3} \cdot (1 + O(\sqrt{\gamma_n})) + O(\varepsilon^2) \qquad \text{or} \qquad \delta_L \geq \frac{1}{2} - \frac{\varepsilon}{3} - O(\varepsilon^2) - O(\sqrt{\gamma_n}), \tag{9}$$

and the lemma follows. □

We now turn to the proof of Lemma 5.1

Proof of Lemma 5.1. The proof of Lemma 5.1 follows by the exactly the same arguments of the proof of Lemma 3.4.

The only difference is in the argument near the end of Claim 3.6, where we have a linear function ϕ that satisfies

$$\Pr_{x,y}[f_1(x) + \phi(x) = f_1(y) + \phi(y)] \geq 1 - 4\varepsilon - 8\varepsilon^2.$$

(see Equation (5)). In our settings the choices of x and y are not independent, and are chosen such that $|x \cap y| = n/4$. This corresponds to choosing a random edge (x, y) in the graph J_n described in Section 8. Replacing the "collision probability" argument with Corollary 8.4 we get that f_1 is $(4\varepsilon + O(\varepsilon^2))$-close either to ϕ or to $\phi + 1$.

The rest is exactly the same as in the proof of Lemma 3.4. □

Next we prove of Lemma 5.2.

Proof of Lemma 5.2. Let $f : L^n_{n/2} \to \{0, 1\}$ be a boolean function, and suppose that f passes the $T^n_{n/2}$ test with probability $1-\varepsilon$. We want to prove that f is close to some linear function. By Lemma 5.3 it is enough to prove that there is a linear function ϕ such that $\text{dist}(f, \phi)$ significantly deviates from 0.5.

Note that the distribution on the pairs $x, y \in L_{n/2}$ in the test can be equivalently described as follows.

1. Pick four distinct coordinates $I = (i_1, i_2, i_3, i_4) \in [n]^4$.
2. Pick $x', y' \in \{0, 1\}^{[n]\setminus I}$ of weight $(n-4)/2$ each such that $|x' \cap y'| = (n-4)/4$.

3. Extend x' to $x \in L_{n/2}$ by letting $x_I = (1,1,0,0)$.

4. Extend y' to $y \in L_{n/2}$ by letting $y_I = (1,0,1,0)$.

If f passes the test with probability $1-\varepsilon$, then, by averaging, there is some 4-tuple $I \in [n]^4$ such that conditioned on this choice of I the test passes with probability $1-\varepsilon$. Let us assume for simplicity that $I = (1,2,3,4)$. Then

$$\Pr_{x',y'}[f(1100 \circ x') + f(1010 \circ y') = f(0110 \circ (x'+y'))] > 1-\varepsilon, \tag{10}$$

where x', y' are chosen as in step 2, and $\circ$ denotes the concatenation of two strings.

Define three function $f_1, f_2, f_3 : L^{n-4}_{(n-4)/2} \to \{0,1\}$ by letting

$$f_1(x') = f(1100 \circ x') \qquad f_2(x') = f(1010 \circ x') \qquad f_3(x') = f(0110 \circ x').$$

Then, by Equation (10) it follows that the function f_1, f_2, f_3 pass the 3-function test with probability at least $1-\varepsilon$. By the hypothesis the statement of Theorem 1.4 holds for $n-4$, and thus one of the functions f_i is close to some linear function $\phi^{(I)} : L^{n-4}_{(n-4)/2} \to \{0,1\}$. That is, for a random $x' \in L^{n-4}_{(n/4)/2}$ we have

$$\Pr[f(1100 \circ x') = \phi^{(I)}(x')] > 1 - O(\varepsilon).$$

Next, we claim that $\phi^{(I)}$ can be extended to linear function that agrees with f on significantly more that 0.5 fraction of the points.

Claim 5.4. *The function $\phi^{(I)}$ can be extended to an affine function $\phi : L^n_{n/2} \to \{0,1\}$ such that $\Pr_{x \in L^n_{n/2}}[f(x) = \phi(x)] > 0.5 + (1 - O(\varepsilon))/32$.*

Therefore, since $\mathrm{dist}(f,\phi)$ significantly deviates from 0.5 by Lemma 5.3 the function ϕ is linear, and $\mathrm{dist}(f,\phi) \le \frac{\varepsilon}{3} \cdot (1 + O(\sqrt{\gamma_n})) + O(\varepsilon^2)$. This completes the proof of Lemma 5.2. □

We now return to the proof of Claim 5.4. The proof goes by choosing a random extension of $\phi^{(I)}$ similarly to the argument in the proof of Lemma 3.5.

Proof of Claim 5.4. Let $\{v_i = e_i + e_{i+1} : i = 1, \ldots, n-1\}$ be a basis of the subspace L^n_{EVEN}. By Fact 2.1 we have $L^n_{n/2} \subseteq L^n_{EVEN}$, and hence every $x \in L^n_{n/2}$ can be written as a linear combination of v_i's.

Consider the set $U = \{1100 \circ x' : x' \in L^{n-4}_{(n-4)/2}\} \subseteq L^n_{n/2}$. Note that every element of $x \in U$ can be written as $x = v_1 + \sum_{i=5}^{n-1} c_i v_i$ for some $c_i \in \{0,1\}$. Since $\phi^{(I)}$ is a linear function on U (or rather on $L^{n-4}_{(n-4)/2}$) there are some coefficients $(a_i \in \{0,1\} : i = 5, \cdots, n-1)$ such that $\phi^{(I)}(x) = \phi^{(I)}(v_1 + \sum_{i=5}^{n-1} c_i v_i) = \sum_{i=5}^{n-1} a_i \cdot c_i$. Let ϕ be a random linear extension of $\phi^{(I)}$ by choosing coefficients $a_1, \ldots, a_4$ uniformly at random, and letting $\phi(\sum_{i=1}^{n-1} c_i v_i) \stackrel{\text{def}}{=} a_1 + \sum_{i=1}^{n-1} a_i \cdot c_i$ (the free coefficient a_1 is also the multiplicand in the term $a_1 \cdot c_1$).[3] Also, for every $x \in L^n_{n/2} \setminus U$ it holds that $\Pr[\phi(x) = f(x)] = \frac{1}{2}$. Therefore, the expected agreement of ϕ with f is

$$\mathbb{E}[\mathrm{agr}(f,\phi)] \ge (1 - O(\varepsilon)) \cdot \frac{|U|}{L^n_{n/2}} + \frac{1}{2} \cdot (1 - \frac{|U|}{L^n_{n/2}}) \ge 0.5 + (1/2 - O(\varepsilon)) \frac{|U|}{L^n_{n/2}} \ge 0.5 + (1-\varepsilon)/32,$$

where the last inequality uses the fact that $\frac{|U|}{L^n_{n/2}} \ge 1/16$. Therefore, there is a choice of the random coefficients $a_1, \ldots, a_4 \in \{0,1\}$ such that $\Pr[f(x) = \phi(x)] \ge 0.5 + (1-\varepsilon)/32$, as required. □

[3] ϕ is indeed an extension of $\phi^{(I)}$ since every $x \in U$ is of the form $x = v_1 + \sum_{i=5}^{n-1} c_i v_i$, and thus $\phi(x) = \sum_{i=5}^{n-1} a_i \cdot c_i = \phi^{(I)}(x)$.

6 Direct Sum Testing for General $k < n/2$

In this section we prove Theorem 1.1 for all even $k < n/2$. The proof works by combining the direct sum tester for $k = n/2$ from Theorem 1.3 with a recent result of Dinur and Steurer [DS13] on direct product testing. They consider the following test for DP_k^n.

$DPT_{k,k'}^n$ - Direct Product Consistency Test:
Given an oracle access to a function $F : L_k^n \to \{0,1\}^k$ do:

1. Select $x \in L_k^n$ uniformly at random.
2. Select $y \in L_k^n$ such that $|x \cap y| = k'$.
3. Accept iff for every $i \in x \cap y$ it holds that $F(x)_i = F(y)_i$.

The following theorem asserts that $DPT_{k,k'}^n$ is indeed a test for the class DP_k^n.

Theorem 6.1 ([DS13]). *For all $p \in (0,1)$ and for all $n, k \in \mathbb{N}$ the test $DPT_{k,pk}^n$ is a $O(\varepsilon)$-test for the class DP_k^n, where the constant in the $O()$ notation depends only on p, but not on k or n.*

We use Theorem 6.1 together with Theorem 1.3 in order to prove Theorem 1.1 for all even $k < n/2$. Before reading the statement of the theorem it could be helpful to recall the definition of a δ-test (see Definition 2.5).

Theorem 6.2. *Let $n \in \mathbb{N}$, and let $k < n/2$ be even. Suppose that T_k^{2k} is a δ_1-test for the class LIN_k^{2k}, and suppose that $DPT_{2k-1,\frac{3k}{2}-1}^{n-1}$ is a δ_2-test for the class DP_{2k}^n, for some functions $\delta_1, \delta_2 : (0,1] \to (0,1]$ such that $\delta_1(\cdot)$ is a non-decreasing linear function. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ and for all $n, k \in \mathbb{N}$ the test T_k^n is a δ-test for the class LIN_k^n, where $\delta(\varepsilon) = 2(\delta_1(\varepsilon) + \delta_2(\varepsilon_2))$ and $\varepsilon_2 = O(\varepsilon)$.*

Plugging in $\delta_1(\varepsilon_1) = O(\varepsilon_1)$ from Theorem 1.3 and $\delta_2(\varepsilon_2) = O(\varepsilon_2)$ from Theorem 6.1 we get Theorem 1.1 with $\delta(\varepsilon) = O(\varepsilon)$ for all even values of $k \leq n/2$.

Proof. The distribution of the three queries made by the test T_k^n can be viewed as first selecting a random $u \in L_{2k}^n$, and then selecting $x, y \in L_k^u$ conditioned on $x + y \in L_k^u$, where L_k^u is the collection of all $x \in L_k^n$ such that $x \subseteq u$ (when we identify subsets of $[n]$ with their characteristic vectors). Therefore, if we condition on u, then the distribution on $x, y, x + y$ is identical to the distribution of T_k^{2k} which we analyzed in previous sections (where we ignore the zero padding of $x, y, x + y$ outside u).

Let f be a function that passes T_k^n with probability $1 - \varepsilon$. For each $u \in L_{2k}^n$ let

$$\varepsilon_u = \Pr[T_k^n \text{ rejects } u] = \Pr_{x,y,x+y \in L_k^u}[f(x) + f(y) \neq f(x+y)] .$$

We clearly have $\mathbb{E}_u[\varepsilon_u] = \varepsilon$. Since L_k^u is isomorphic to L_k^{2k} except that the strings in L_k^u have a padding of zeros outside u, we can invoke Theorem 1.3 to deduce the existence of a string $\sigma^{(u)} \in \{0,1\}^{2k}$ for which

$$\Pr_{x \in L_k^u}[f(x) = \sum_{i \in u} \sigma_i^{(u)} x_i \mod 2] \geq 1 - \delta_1(\varepsilon_u)$$

The next natural step would be to construct a direct product function that assigns each u with $\sigma^{(u)}$, and then to apply the direct product testing theorem. However, we first must resolve an ambiguity that stems from the fact that since k is even, both $\sigma^{(u)}$ and its complement give

rise to the same linear function on L_k^u, as explained in Fact 2.3. We resolve this ambiguity by considering only those u that contain a fixed coordinate $i_0 \in [n]$, and by letting the string $\sigma^{(u)}$ be the one that assigns 0 to the coordinate i_0. The coordinate $i_0 \in [n]$ is chosen so that

$$\Pr[T_k^n \text{ accepts } f | i_0 \in u] \geq 1 - \varepsilon,$$

where such a coordinate is guaranteed to exist by averaging.

Let $L_{2k,0}^n$ denote the set of elements in L_{2k}^n that contain i_0. Now, we define a new function $F : L_{2k,0}^n \to \{0,1\}^{2k}$ by letting

$$F(u) = \sigma^{(u)} \qquad \forall u \in L_{2k,0}^n.$$

We claim that F passes the following test DPT_0 with high probability.

1. Select $u \in L_{2k,0}^n$ uniformly at random.
2. Select $v \in L_{2k,0}^n$ uniformly at random conditioned on $|u \cap v| = \frac{3k}{2}$.
3. Accept if and only if $F(u)_j = F(v)_j$ for every $j \in u \cap v$.

The following two claims complete the proof of Theorem 6.2.

Claim 6.3. *The function F passes the test DPT_0 with probability $1 - \varepsilon_2$ for $\varepsilon_2 = O(\varepsilon)$. This implies that there exists a string $a \in \{0,1\}^n$ such that*

$$\Pr_{u \in L_{2k,0}^n}[F(u) = a_u] > 1 - \delta_2(\varepsilon_2),$$

where a_u denotes the restriction of a to the coordinates $j \in [n]$ such that $u_j = 1$.

The following claim asserts that f has a large agreement with the global linear function represented by the string $a \in \{0,1\}^n$ from Claim 6.3.

Claim 6.4. *Let $a \in \{0,1\}^n$ be the string from Claim 6.3. Then*

$$\Pr_{x \in L_k^n}[f(x) = \sum_{j \in x} a_j] > 1 - 2(\delta_1(\varepsilon) + \delta_2(\varepsilon)).$$

This concludes the proof of Theorem 6.2. □

Proof of Claim 6.3. Let $u, v \in L_{2k}^n$ be the choices of the test DPT_0 such that $|u \cap v| = \frac{3k}{2}$, and suppose that the DPT_0 test rejects on this pair, i.e., $F(u)_{u \cap v} \neq F(v)_{u \cap v}$. Hence, by Fact 2.3, since $i_0 \in u \cap v$ and $F(u)_{i_0} = F(v)_{i_0}$, the linear functions $\phi^{(u)}, \phi^{(v)}$ defined by $\phi^{(u)}(x) = \sum_i \sigma_i^{(u)} x_i$ and $\phi^{(v)}(x) = \sum_i \sigma_i^{(u')} x_i$ are not identical on $L_k^{u \cap v}$. Therefore, by Proposition 2.4, it follows that for a random $x \in L_k^{u \cap v}$ we have $\Pr_{x \in L_k^{u \cap v}}[\phi^{(u)}(x) \neq \phi^{(v)}(x)] \geq c$ for some absolute constant $c > 0$.

Let ε_2 denote the probability that the test DPT_0 rejects. Then

$$\begin{aligned}
\Pr_{\substack{u,v \sim DPT_0 \\ x \in L_k^{u \cap v}}}[\phi^{(u)}(x) = f(x) = \phi^{(v)}(x)] &\leq \Pr[F(u)|_{u \cap v} = F(v)|_{u \cap v}] \\
&\quad + \Pr[\sum_{j \in x} F(u)_j = \sum_{j \in x} F(v)_j \text{ and } F(u)|_{u \cap v} \neq F(v)|_{u \cap v}] \\
&\leq (1 - \varepsilon_2) + \varepsilon_2 \cdot (1 - c) = 1 - c\varepsilon_2. \qquad (11)
\end{aligned}$$

On the other hand, we claim that

$$\mathbb{E}_{u,v\sim DPT_0}\Pr[\phi^{(u)}(x)=f(x)=\phi^{(v)}(x)]\geq \mathbb{E}_{u,v}[1-\delta_1(\varepsilon_u)-\delta_1(\varepsilon_v)]=1-2\mathbb{E}_u[\delta_1(\varepsilon_u)]. \quad (12)$$

Indeed, for any $u\in L^n_{2k}$ we have $\Pr[f(x)=\phi^{(u)}(x)]\geq 1-\delta_1(\varepsilon_u)$, and similarly for v. Equation (12) follows by union bound on u and v. Taking expectation over all choices u,v of the test DPT_0, we get the inequality.

It remains to recall that $\delta_1(\cdot)$ is a non-decreasing linear function, and so $\mathbb{E}_u[\delta_1(\varepsilon_u)]\leq \delta_1(\mathbb{E}_u[\varepsilon_u])\leq\delta_1(\varepsilon)$, where the last inequality is because the choice of i_0 to ensures that $\mathbb{E}_{u\in L^n_{2k,0}}[\varepsilon_u]\leq \mathbb{E}_{u\in L^n_{2k}}[\varepsilon_u]=\varepsilon$. Combining (11) and (12) we get $\varepsilon_2\leq 2\delta_1(\varepsilon)/c=O(\varepsilon)$, as required.

We have shown that the test DPT_0 accepts on F with probability $1-O(\varepsilon)$. We can now ignore the i_0'th coordinate in all $u\in L^n_{2k,0}$ and think of F as a function $F:L^{n-1}_{2k-1}\to\{0,1\}^{2k-1}$, and so, the test DPT_0 above corresponds to the direct product consistency test with parameters $(n-1,2k-1,\frac{3k}{2}-1)$. Since F passes the test with probability $1-\varepsilon_2$, by the assumption of the theorem there exists a string $a\in\{0,1\}^n$ such that

$$\Pr_{u\in L^n_{2k,0}}[F(u)=a_u]>1-\delta_2(\varepsilon_2),$$

as required. □

Proof of Claim 6.4. Let $a\in\{0,1\}^n$ be the string from Claim 6.3 satisfying $\Pr_{u\in L^n_{2k,0}}[F(u)=a_u]>1-\delta_2(\varepsilon_2)$. We first show that if we first pick $u\in L^n_{2k,0}$ and then pick $x\in L^u_k$ at random, then with high probability we have $f(x)=\sum_{j\in x}a_j$. More precisely, we show that

$$\Pr_{\substack{u\in L^n_{2k,0}\\ x\in L^u_k}}[f(x)=\sum_{j\in x}a_j]>1-(\delta_1(\varepsilon)+\delta_2(\varepsilon_2)). \quad (13)$$

In order to prove (13) recall that $\Pr_{x\in L^u_k}[f(x)=\phi^{(u)}(x)=\sum_{j\in x}F(u)_j]\geq 1-\delta_1(\varepsilon_u)$ for each $u\in L^n_{2k,0}$. So in expectation over $u\in L^n_{2k,0}$ we have

$$\Pr_{\substack{u\in L^n_{2k,0}\\ x\in L^u_k}}[f(x)=\phi^{(u)}(x)=\sum_{j\in x}F(u)_j]\geq 1-\mathbb{E}_{u\in L^n_{2k,0}}[\delta_1(\varepsilon_u)]\geq 1-\delta_1(\mathbb{E}_u[\varepsilon_u])\geq 1-\delta_1(\varepsilon).$$

By the assumption on a for random $u\in L^n_{2k,0}$ we have $F(u)=a_u$ with probability at least $1-\delta_2(\varepsilon_2)$. Therefore, by union bound we get that

$$\Pr_{\substack{u\in L^n_{2k,0}\\ x\in L^u_k}}[f(x)=\sum_{j\in x}F(u)_j=\sum_{j\in x}a_j]\geq 1-(\delta_1(\varepsilon)+\delta_2(\varepsilon_2)),$$

which implies (13).

To obtain the statement of Claim 6.4 the distribution of x needs to be uniform in L^n_k. Nonetheless, the distribution x in (13) can be written as a convex combination of the uniform distribution on L^n_k with probability 0.5, and another distribution that puts more weight on x's that contain i_0. Thus, Equation (13) implies that if we pick $x\in L^n_k$ uniformly at random, then

$$\Pr[f(x)=\sum_{j\in x}a_j]>1-2(\delta_1(\varepsilon)+\delta_2(\varepsilon_2)),$$

and the claim follows. □

7 Testing Tensor Powers

In this section we show that for $k \ll n$, Theorem 1.1 also implies a tester for tensors. A function $f : [n]^k \to \{-1, 1\}$ is a tensor power if there is some $b : [n] \to \{-1, 1\}$ such that $f = b^{\otimes k}$, i.e.

$$f(z_1, \ldots, i_k) = b(z_1) \cdot b(z_2) \cdots b(z_k).$$

We will use the following notation. For two tuples $u, v \in [n]^{k/2}$ we denote their concatenation by $uv \in [n]^k$. For a tuple $z \in [n]^k$ and a permutation $\pi : [k] \to [k]$ we let $z \circ \pi \in [n]^k$ be the permuted k-tuple $(z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(k)}) \in [n]^k$. For convenience, we repeat the tensor power test described in the introduction.

The Tensor Power Test - TP_k^n
Given an oracle access to a function $f : [n]^k \to \{-1, 1\}$ do:

1. Pick $u, v, w \in [n]^{k/2}$ independently uniformly at random.
2. Pick three permutations $\pi_1, \pi_2, \pi_3 : [k] \to [k]$ independently at random.
3. Accept if and only if $f(uv \circ \pi_1) \cdot f(vw \circ \pi_2) = f(uw \circ \pi_3)$.

We prove the following theorem.

Theorem 1.2 restated: *Suppose $n, k \in \mathbb{N}$ and $\varepsilon > 0$ are such that $k^2/n = o(\varepsilon)$. Let $f : [n]^k \to \{-1, 1\}$ be a function that passes the test TP_k^n with probability at least $1 - \varepsilon$. Then there is some $b : [n] \to \{-1, 1\}$ such that*

$$\Pr_{z \in [n]^k}[f(z) = b^{\otimes k}(z) = b(z_1) \cdot b(z_2) \cdots b(z_k)] \geq 1 - O(\varepsilon).$$

Proof. The proof is by reduction to Theorem 1.1. Given a function $f : [n]^k \to \{-1, 1\}$, we define $g : \binom{[n]}{k} \to \{0, 1\}$. The idea is to define $g(S)$ so that $(-1)^{g(S)} = f(a)$, where a is some ordering of the subset S. However, since f might give different values to different orderings, this is not well defined, and instead we consider all possible orderings of S. We write $S = \{a_1 < a_2 < \cdots < a_k\}$ and let $(S \circ \pi) := (a_{\pi(1)}, \ldots, a_{\pi(k)})$ for every permutation $\pi : [k] \to [k]$. We define $g(S)$ to equal 1 if $\Pr_\pi[f(S \circ \pi) = -1] > 1/2$, and let $g(S) = 0$ otherwise. In other words

$$(-1)^{g(S)} = \text{majority}_\pi f(S \circ \pi),$$

where majority stands for the most common value. We claim that the function g passes the direct sum test with probability $1 - O(\varepsilon)$. Indeed, the queries $x, y, x + y$ chosen by the direct sum test distribution T_k^n, are equivalently described as choosing three disjoint $k/2$-element subsets $U, V, W \subseteq [n]$ and setting $x = U \cup V$, $y = V \cup W$ and then $x + y = (U \cup V)\Delta(V \cup W) = U \cup W$ (where we identify subsets on $[n]$ with their indicator vectors). We couple this choice in the direct sum test with a choice of $uv = (UV \circ \pi_1), vw = (VW \circ \pi_2), uw = (UW \circ \pi_3)$ in f, where π_1, π_2, π_3 are random permutations.

Since TP_k^n randomizes the order, whenever $g(x)+g(y) \neq g(x+y)$ there is constant probability that TP_k^n fails by making random choices that coincide with the majority (this occurs for all three queries with probability at least $1/8$). Therefore, the probability of rejection of TP_k^n on f is at least

$$\Pr[TP_k^n \text{ rejects } f] \geq \Omega(1) \cdot \Pr[T_k^n \text{ rejects } g] - \Pr[\text{COLLISION}],$$

where COLLISION denotes the even that at least one of the $u, v, w \in [n]^k$ has the same value on different coordinates. By the assumption that $k^2/n = o(\varepsilon)$ it follows that $\Pr[\text{COLLISION}] = o(\varepsilon)$, and hence the direct sum test rejects on g with probability at most $O(\varepsilon)$.

We deduce from Theorem 1.1 that g is $1 - O(\varepsilon)$ close to a direct sum, i.e., that there is some $a \in \{0,1\}^n$ such that $\Pr_S[g(S) = \sum_{i \in S} a_i \mod 2] \geq 1 - O(\varepsilon)$.

Note that the probability over $S \in \binom{[n]}{k}$ and $\pi : [k] \to [k]$ that $f(S \circ \pi) \neq (-1)^{g(S)}$ is at most $O(\varepsilon)$, as otherwise this would cause TP_k^n to reject with too high a probability.

By the assumption that $k^2/n = o(\varepsilon)$ for a random $z \in [n]^k$ no COLLISION happens with probability $1 - o(\varepsilon)$, and hence such z can be chosen by picking a random $S \in \binom{[n]}{k}$ and $\pi : [k] \to [k]$, and setting $z = (S \circ \pi)$. Therefore,

$$\Pr_{z \in [n]^k}[f(z) = (-1)^{\sum_{i \in [k]} a_{z_i}}] \geq (1 - o(\varepsilon)) \Pr_{z \leftarrow S, \pi}[f(z) = (-1)^{\sum_{i \in S} a_i}] - o(\varepsilon) = 1 - O(\varepsilon),$$

and so, for $b \in \{-1,1\}^n$ defined as $b_i = (-1)^{a_i}$ we have $f(z) = b(z_1) \cdot b(z_2) \cdots b(z_k)$ with probability $1 - O(\varepsilon)$. This completes the proof of the theorem. □

8 Vertex Expansion of the Johnson Graph

In order to analyze the direct sum test $T_{n/2}^n$ it will be convenient to consider the following graph, which is a natural interpretation of the Johnson scheme.

Definition 8.1. *For $n \in \mathbb{N}$ such that $n \equiv 0 \pmod 4$ define a graph $J_n = (V_n, E_n)$ by letting $V = L_{n/2}^n$, and setting an edge between two vertices x and y if and only if $|x \cap y| = n/4$.*

Remark The degree of the graph is almost linear. Indeed, if we denote by $N = \binom{n}{n/2} = \Theta(\frac{2^n}{\sqrt{n}})$ the number of vertices in J_n, then every vertex J_n has degree $\binom{n/2}{n/4}^2 = \Theta(\frac{2^n}{n}) = \Theta(\frac{N}{\sqrt{\log(N)}})$.

Note that we can describe linearity test in terms of the graph J_n as follows.

$T_{n/2}^n$ test - restated in terms of J_n
Given an oracle access to a function $f : V_n \to \{0,1\}$ do:

1. Select a random triangle $\{x, y, x+y\} \subseteq V_n$ in J_n.
2. Accept if and only if $f(x) + f(y) = f(x+y)$.

Below we claim that the mixing time of J_n is 2. More precisely, if we start from an arbitrary vertex $x \in V_n$ and perform a random walk x, y, z of length 2, then the distribution of z is $\tilde{O}(1/\sqrt{n})$ close to the uniform distribution on V_n, i.e., it is almost independent of the starting vertex x.

Lemma 8.2. *For $n \in \mathbb{N}$ such that $n \equiv 0 \pmod 4$ consider the graph $J_n = (V_n, E_n)$. Fix a vertex $x \in V_n$. Let $y \in N(x)$ be a random neighbor of x, and let $z \in N(y)$ be a random neighbor of y. Then, the distribution of z is $\tilde{O}(1/\sqrt{n})$-close in total variation distance to the uniform distribution on V_n.*

Proof. Since the graph J_n is vertex transitive, we may assume that $x = \underbrace{11\ldots1}_{n/2}\underbrace{00\ldots0}_{n/2}$. Then the distance of the distribution of z from the uniform distribution is

$$\text{dist}_{TV}(z, U) = \frac{1}{2} \sum_{w \in L_{n/2}^n} \left| \Pr[z = w] - \frac{1}{\binom{n}{n/2}} \right|$$

We partition the sum according to the weight of w on the coordinates $\{1,\ldots,n/2\}$. Then

$$\mathrm{dist}_{TV}(z,U) \quad = \quad \frac{1}{2}\sum_{i=0}^{n/2}\left|\Pr[|z_{[1\ldots n/2]}|=i]-\frac{\binom{n/2}{i}\binom{n/2}{n/2-i}}{\binom{n}{n/2}}\right|$$

Let $y\in N(x)$ be a neighbor of x. Then, y has $n/4$ many 1's in x, and $n/4$ many 1's outside x. Assume for concreteness that $y=\underbrace{11\ldots1}_{n/4}\underbrace{00\ldots0}_{n/4}\underbrace{11\ldots1}_{n/4}\underbrace{00\ldots0}_{n/4}$. Then, in order to choose the 1's of $z\in N(y)$ we do the following.

1. Pick j coordinates from $\{1,\ldots,n/4\}$.
2. Pick $i-j$ coordinates from $\{n/4+1,\ldots,n/2\}$.
3. Pick $n/4-j$ coordinates from $\{n/2+1,\ldots,3n/4\}$.
4. Pick $n/4-(i-j)$ coordinates from $\{3n/4+1,\ldots,n\}$.

Therefore $\Pr[|z_{[1\ldots n/2]}|=i]=\sum_{j=0}^{i}\frac{\binom{n/4}{j}\binom{n/4}{i-j}\binom{n/4}{n/4-j}\binom{n/4}{n/4-(i-j)}}{\binom{n/2}{n/4}^2}$, and so

$$\mathrm{dist}_{TV}(z,U)=\frac{1}{2}\sum_{i=0}^{n/2}\left|\sum_{j=0}^{i}\frac{\binom{n/4}{j}^2\binom{n/4}{i-j}^2}{\binom{n/2}{n/4}^2}-\frac{\binom{n/2}{i}^2}{\binom{n}{n/2}}\right|.$$

The following proposition completes the proof of Lemma 8.2.

Proposition 8.3. *For all $n\in\mathbb{N}$ we have*

$$\sum_{i=0}^{n/2}\left|\sum_{j=0}^{i}\frac{\binom{n/4}{j}^2\binom{n/4}{i-j}^2}{\binom{n/2}{n/4}^2}-\frac{\binom{n/2}{i}^2}{\binom{n}{n/2}}\right|=\tilde{O}(1/\sqrt{n}).$$

We defer the proof of the proposition to Appendix B. □

As a corollary, we conclude that J_n is a vertex-expander in the following strong sense.

Lemma 1.5 restated. *For $n\in\mathbb{N}$ such that $n\equiv 0 \pmod 4$ let $\gamma_n=\tilde{O}(n^{-1/2})$ be the distance of z from the uniform distribution in Lemma 8.2. Let $A\subseteq V_n$ be a subset of the vertices of J_n of size $|A|=\alpha|V|$. Pick an edge $(x,y)\in E_n$ of J_n uniformly at random. Then*

1. $\Pr[x\in A, y\notin A]=\alpha(1-\alpha)\pm 2\alpha(1-\alpha)\cdot\sqrt{\gamma_n}$.
2. $\Pr[x,y\in A]=\alpha^2\pm 2\alpha(1-\alpha)\cdot\sqrt{\gamma_n}$.

That is, the vertex expansion of J_n is very close to that of a random graph/complete graph.

Proof. Denote by $\tau=\frac{|E[A,\overline{A}]|}{|E_n|}$ the fraction of edges in the cut $(A,\overline{A})$. Note that τ is equal to

$$\tau=2\alpha\cdot\mathbb{E}_{x\in A}\left[\frac{\deg_{\overline{A}}(x)}{\deg(x)}\right], \tag{14}$$

where $deg_{\overline{A}}(x)=|N(x)\cap\overline{A}|$ denotes the number of neighbors of x in $\overline{A}$.

For the first item, let us pick a random vertex $x_0 \in V_n$, and perform a random walk x_0, x_1, x_2 of length 2. Let $E_{A\overline{A}} \in \{0,1,2\}$ be a random variable counting the number of times that the walk crosses the cut $(A, \overline{A})$. Then

$$\mathbb{E}[E_{A\overline{A}}] = 2\tau.$$

On the other hand

$$\begin{aligned}\mathbb{E}[E_{A\overline{A}}] &= \Pr[x_0 \in A, x_2 \in \overline{A}] + \Pr[x_0 \in \overline{A}, x_2 \in A] \\ &\quad +2\Pr[x_0, x_2 \in A, x_1 \in \overline{A}] + 2\Pr[x_0, x_2 \in \overline{A}, x_1 \in A].\end{aligned} \tag{15}$$

Using Lemma 8.2 we have

$$\Pr[x_0 \in A, x_2 \in \overline{A}] = \Pr[x_0 \in A] \cdot \Pr[x_2 \in \overline{A} | x_0 \in A] \geq \alpha \cdot (1 - \alpha - \gamma_n).$$

By reversibility, we also have $\Pr[x_0 \in \overline{A}, x_2 \in A] = \Pr[x_0 \in A, x_2 \in \overline{A}]$.

Next we claim that $\Pr[x_0, x_2 \in \overline{A}, x_1 \in A] \geq \frac{\tau^2}{4\alpha}$. Indeed, in order to estimate $\Pr[x_0, x_2 \in \overline{A}, x_1 \in A]$ we may pick a vertex $x_1 \in V_n$ uniformly at random, and then pick two neighbors $x_0, x_2 \in N(x_1)$ independently. Then

$$\begin{aligned}\Pr[x_0, x_2 \in \overline{A}, x_1 \in A] &= \Pr[x_1 \in A] \Pr[x_0, x_2 \in \overline{A} | x_1 \in A] \\ &= \alpha \cdot \mathbb{E}_{x \in A}\left[\left(\frac{\deg_{\overline{A}}(x)}{\deg_V(x)}\right)^2\right] \\ \text{[Cauchy-Schwartz inequality]} \quad &\geq \alpha \cdot \mathbb{E}_{x \in A}\left[\frac{\deg_{\overline{A}}(x)}{\deg_V(x)}\right]^2 \\ \text{[using Equation (14)]} \quad &= \frac{\tau^2}{4\alpha}.\end{aligned}$$

Analogously, we have $\Pr[x_0, x_2 \in A, x_1 \in \overline{A}] \geq \frac{\tau^2}{4(1-\alpha)}$. Therefore, using Equation (15) we get

$$2\tau \geq 2\alpha(1 - \alpha - \gamma_n) + \tau^2\left(\frac{1}{2\alpha} + \frac{1}{2(1-\alpha)}\right).$$

Solving the quadratic inequality (in τ) gives us

$$(\tau - 2\alpha(1-\alpha))^2 \leq 4\alpha^2(1-\alpha) \cdot \gamma_n.$$

By symmetry, we may change the roles of A and $\overline{A}$ to get

$$(\tau - 2\alpha(1-\alpha))^2 \leq 4\alpha(1-\alpha)^2 \cdot \gamma_n.$$

The statement of the first item follows from the fact that $\Pr[x \in A, y \notin A] = \tau/2$.

For the second item of the lemma we have

$$\Pr[x, y \in A] = \Pr[x \in A] - \Pr[x \in A, y \notin A].$$

Plugging the estimation from the first item, the lemma follows. □

Corollary 8.4. *Let $A \subseteq V_n$ be a set of vertices of density $\alpha = \frac{|A|}{|V|}$. Suppose that if we pick a random edge $(x, y) \in E$ then the probability that one endpoint is in A and the other is not in A is at most ε for some $\varepsilon > 0$ sufficiently small. Then either $\alpha < \varepsilon + O(\varepsilon^2)$ or $\alpha > 1 - (\varepsilon + O(\varepsilon^2))$.*

Proof. Suppose that $n \in \mathbb{N}$ be large enough such that $\gamma_n < 1/16$. (Otherwise the statement is trivial for sufficiently large multiplicative constant in the $O(\varepsilon)$ term.) By Lemma 1.5 we have

$$1 - \varepsilon < \Pr[x, y \in A] + \Pr[x, y \in \overline{A}] \leq \alpha^2 + (1-\alpha)^2 \pm 4\alpha(1-\alpha) \cdot \sqrt{\gamma_n}.$$

This is equivalent to $\alpha(1-\alpha) < \frac{\varepsilon}{2 - 4\sqrt{\gamma_n}} < \varepsilon$. Solving the quadratic inequality in α gives the desired result. □

8.1 Testing Tensor Products

It is natural to extend the above result to a tester for tensor products, i.e. for a tester deciding if a given function is equal to $b_1 \otimes b_2 \otimes \cdots \otimes b_k$. For this, we propose the following test

The Tensor product Test - $TProd_k^n$
Given an oracle access to a function $f : [n]^k \to \{-1, 1\}$ do:

1. Pick $u_1, u_2 \in [n]^k$ independently at random.
2. Pick $v_1, v_2 \in [n]^k$ independently at random by setting
3. Pick three permutations $\pi_1, \pi_2, \pi_3 : [k] \to [k]$ independently at random.
4. Accept if and only if $f(uv \circ \pi_1) \cdot f(vw \circ \pi_2) = f(uw \circ \pi_3)$.

9 A Different Direct Sum Tester

For odd values of k, as well as for the case $k > \frac{2}{3}n$ one can see that no 3-query test can succeed. In this section we present a different linearity test for functions $f : L_k^n \to \{0, 1\}$ that works for all $k \leq n$, and makes $O(\max(\frac{n}{k}, \frac{n}{n-k}))$ queries. Therefore, the test makes constant number of queries if $\frac{k}{n}$ is bounded away from 0 and 1. The test has a very simple analysis and does not rely on direct product test from Theorem 1.1. A similar result is already known from the work of Kopparty and Saraf [KS09].

As mentioned in the introduction, Kopparty and Saraf [KS09] extend the BLR linearity testing result to a large family of distributions on the hypercube, which they call "uniformly-correlatable distributions". That is, for a distribution μ on $\{0, 1\}^n$ the goal is to test whether a function $f : \{0, 1\}^n \to \{0, 1\}$ is close to some linear function, and the distance between functions is defined as $\text{dist}(f, g) = \Pr_{x \sim \mu}[f(x) \neq g(x)]$. We may give the following interpretation to their test. Given a function $f : \{0, 1\}^n \to \{0, 1\}$ the test constructs a probabilistic oracle $g : \{0, 1\}^n \to \{0, 1\}$ such that if we want to query g in a uniformly random point x, it is enough to make a small number of queries to f on inputs $r_1, \ldots, r_t$ such that the marginal distribution of each r_i is equal to μ. Then, they simulate a linearity test on the (randomized) function g with respect to the uniform distribution, and if the test passes with high probability, they deduce that g is close to a linear function, and also f is close to a linear function.

Our tester is similar to the tester of Kopparty and Saraf, but the analysis we provide seems to differ from theirs. In particular, our proof also constructs a probabilistic oracle as above, but does not rely on the fact that the uniform distribution on L_k^n is "uniformly-correlatable", and only uses the fact that every element in $\text{span}\langle L_k^n \rangle$ is a sum of $O(\max(\frac{n}{k}, \frac{n}{n-k}))$ vectors of L_k^n.

Theorem 9.1. *Let $p \in (0, 1)$, let $n \in \mathbb{N}$ be sufficiently large such that $pn \in \mathbb{N}$, and let $k = pn$. There exists a test T_p such that given an oracle access to a function $f : L_k^n \to \{0, 1\}$ makes $\max(O(\frac{1}{p}), O(\frac{1}{1-p}))$ queries to f, and satisfies the following conditions.*

1. *If f is linear, then T_p accepts with probability 1.*
2. *For all $\varepsilon > 0$, if $\Pr[T_p \text{ accepts } f] > 1 - \varepsilon$, then there exists a string $a \in \{0, 1\}^n$ such that $\Pr_{x \in L_k^n}[f(x) = \sum_{i \in [n]} a_i x_i] > 1 - \delta$, where $\delta = \delta(\varepsilon)$ is such that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.*

Proof. Let us assume for simplicity that $p = k/n \leq 0.5$. (The case $p > 0.5$ is handled analogously.) Then, our goal is to design a linearity test for L_k^n that makes $O(n/k)$ queries. Recall (Fact 2.1) that if k is even, then $\text{span}\langle L_k^n \rangle = L_{EVEN}^n$, and otherwise $\text{span}\langle L_k^n \rangle = \{0, 1\}^n$. Furthermore, note that for every $x \in \text{span}\langle L_k^n \rangle$ there are $t = \lceil 1/p \rceil$ vectors $x_1, \ldots, x_t \in L_k^n$ such

that $\sum_i x_i = x$. This simple observation allows us to define the following test T_p that makes $3t = 3\lceil 1/p \rceil$ queries.

T_p - Direct Sum Test for L_k^n:
Given an oracle access to a function $f : L_k^n \to \{0,1\}$ do:

Set $t = \lceil 1/p \rceil$. With probability $1/2$ apply the subroutine T^1, and with probability $1/2$ apply the subroutine T^2, where T^1 and T^2 are defined as follows.

T^1:

1. Pick $x, y \in \text{span}\langle L_k^n \rangle$ independently uniformly at random. Let $z = x + y$.
2. Pick $x_1, \ldots, x_t \in L_k^n$ such that $\sum_i x_i = x$.
3. Pick $y_1, \ldots, y_t \in L_k^n$ such that $\sum_i y_i = y$.
4. Pick $z_1, \ldots, z_t \in L_k^n$ such that $\sum_i z_i = z$.
5. Accept if and only if $\sum_i f(x_i) + \sum_i f(y_i) + \sum_i f(z_i) = 0$.

T^2:

1. Pick $x \in L_k^n$, and pick $y \in \text{span}\langle L_k^n \rangle$ independently uniformly at random. Let $z = x + y$.
2. Pick $y_1, \ldots, y_t \in L_k^n$ such that $\sum_i y_i = y$.
3. Pick $z_1, \ldots, z_t \in L_k^n$ such that $\sum_i z_i = z$.
4. Accept if and only if $f(x) + \sum_i f(y_i) + \sum_i f(z_i) = 0$.

We interpret our test as follows. Given a function $f : L_k^n \to \{0,1\}$ the test constructs a probabilistic oracle to a function $G : \text{span}\langle L_k^n \rangle \to \{0,1\}$ such that whenever we want to query G in a uniformly random point x, we query f on a constant locations $x_1, \ldots, x_t$ that sum up to x, and set $G(x) = \sum_i f(x_i)$.

Using this notation the test T^1 can be interpreted as testing whether $G(x)+G(y) = G(x+y)$ for uniformly random chosen $x, y \in \text{span}\langle L_k^n \rangle$. Similarly, the test T^2 can be interpreted as testing whether $f(x) + G(y) = G(x + y)$ for uniformly random chosen $x \in L_k^n$ and $y \in \text{span}\langle L_k^n \rangle$.

Next, we define a function $g : \text{span}\langle L_k^n \rangle \to \{0,1\}$ to be the "rounding" of G. That is, for all $x \in \text{span}\langle L_k^n \rangle$ we set $g(x) = 1$ if $\Pr[G(x) = 1] > 0.5$, and $g(x) = 0$. Note that the domain of g has a subgroup structure. We claim below that g passes the linearity test with high probability. Using [BLR93] this will imply that g is close to some linear function.

Claim 9.2. *Pick $x, y \in \text{span}\langle L_k^n \rangle$ independently uniformly at random. Then*

$$\Pr_{x,y}[g(x) + g(y) = g(x + y)] > 1 - 4\varepsilon.$$

Therefore, there exists a linear function $\phi : \text{span}\langle L_k^n \rangle \to \{0,1\}$, such that $dist(g, \phi) = O(\varepsilon)$.

Proof. For every $x \in \text{span}\langle L_k^n \rangle$ we measure how decisive G is on x by letting $\varepsilon_x = \Pr[G(x) \neq g(x)]$. We first show that for uniformly random $x \in \text{span}\langle L_k^n \rangle$ we have $\mathbb{E}_x[\varepsilon_x] < \varepsilon$.

Indeed, let $x \in \text{span}\langle L_k^n \rangle$. For any fixed values of $y \in \text{span}\langle L_k^n \rangle$, $b_2 = G(y)$, and $b_3 = G(x+y)$ we have $\Pr_G[G(x) = b_2 + b_3] \leq 1 - \varepsilon_x$ (the randomness is only over the choice of $G(x)$). Therefore by taking the expectation we have

$$1 - \varepsilon \leq \Pr[G(x) + G(y) = G(x + y)] = \mathbb{E}_x \left[\Pr_{G,y}[G(x) = G(y) + G(x + y)] \right] \leq \mathbb{E}_x[1 - \varepsilon_x],$$

This implies that $\mathbb{E}_w[\varepsilon_w] < \varepsilon$.

Since each of the query of T^1 is uniformly distributed in $\mathrm{span}\langle L_k^n \rangle$, it follows that with probability at least $1 - 3\varepsilon$, we have $G(x) = g(x)$, $G(y) = g(y)$, and $G(x+y) = g(x+y)$. This completes the first part of the claim.

Since the domain of g is a subgroup, it follows from [BLR93] that there exists a linear function $\phi : \mathrm{span}\langle L_k^n \rangle \to \{0,1\}$, such that $dist(g, \phi) = O(\varepsilon)$. The claim follows. □

Next, we use the assumption that f passes the test T^2 with high probability in order to prove that f is $O(\varepsilon)$-close to the restriction of ϕ to L_k^n.

Claim 9.3. *Let ϕ be the linear function from Claim 9.2. Then f is $O(\varepsilon)$-close to the restriction of ϕ to L_k^n.*

Proof. Consider the oracle G as defined above, and note that the test T^2 checks that $f(x) + G(y) = G(y+x)$ for $x \in L_k^n$‘ and $y \in \mathrm{span}\langle L_k^n \rangle$ chosen uniformly at random.

Note that by Claim 9.2 for a uniformly random $y \in \mathrm{span}\langle L_k^n \rangle$ we have $\Pr[G(y) = g(y) = \phi(y)] > 1 - O(\varepsilon)$, and similarly $\Pr[G(x+y) = g(x+y) = \phi(x+y)] > 1 - O(\varepsilon)$.

Since by the assumption we have $\Pr[f(x) = G(y) + G(x+y)] > 1 - \varepsilon$, it follows that a random $x \in L_k^n$ with probability $1 - O(\varepsilon)$ we have $f(x) = \phi(y) + \phi(x+y)$. By linearity of ϕ this implies that

$$\Pr_x[f(x) = \phi(x)] > 1 - O(\varepsilon),$$

as required. □

This completes the proof of Theorem 9.4. □

Note that the proof of Theorem 9.4 is more general, and generalizes to all subsets $V \subseteq \{0,1\}^n$ such that each element of $\mathrm{span}\langle V \rangle$ can be obtained by summing a small number of elements from V.

Theorem 9.4. *Let $n \in \mathbb{N}$, and let $V \subseteq \{0,1\}^n$. Suppose that for every $x \in \mathrm{span}\langle V \rangle$ there are t elements $x^{(1)}, \ldots, x^{(t)} \in V$ such that $\sum_{i=1}^t x^{(i)} = x$. Then, there exists a test T_V such that given an oracle access to a function $f : V \to \{0,1\}$ makes $3t$ queries to f, and satisfies the following conditions.*

1. *If f is linear on V, then T_V accepts with probability 1.*
2. *For all $\varepsilon > 0$, if $\Pr[T_V$ accepts $f] > 1 - \varepsilon$, then there exists a string $a \in \{0,1\}^n$ such that $\Pr_{x \in V}[f(x) = \sum_{i \in [n]} a_i x_i] > 1 - \delta$, where $\delta = \delta(\varepsilon)$ is such that $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$.*

10 Open Problems and Future Work

Extending the proof to low degree polynomials. A natural generalization of linearity testing is whether a given function is close to a low degree polynomial. Alon et al. [AKK+05], and later Bhattacharyya et al. [BKS+10], studied this question for functions defined on the hypercube. Do these results extend to functions defined only on L_k^n? That is, given a function $f : L_k^n \to \{0,1\}$, test whether there is a low degree polynomial $p : \{0,1\}^n \to \{0,1\}$ such that $f(x) = p(x)$ for most points $x \in L_k^n$.

Reducing the Direct Product Testing Problem to Direct Sum Testing. We showed a direct product tester implies the existence of a direct sum tester. Is there a reduction going in the other direction? That is, given the fact we have a direct sum tester does it imply the existence of a direct product tester.

The Tensor Product Testing Interpretation. In Theorem 1.2 we proved tensor power testing for $k = o(\sqrt{n})$. It would be interesting to extended to the theorem for all values of k. Another interesting question in this direction is to test whether a given function $f : [n]^k \to \{-1, 1\}$ is close to a tensor product of k possibly distinct functions $b^{(1)} \otimes \cdots \otimes b^{(k)}$.

Low acceptance probability regime. Linearity testing was analyzed in the low acceptance regime (see [BCH+96, KLX07]), where the goal is to understand functions that pass the test with probability which is slightly larger compared to the probability that a random function passes it. It is shown in [BCH+96] that if f defined on the hypercube passes the BLR test with probability slightly above 0.5, then it has a non trivial correlation with some linear function. Understanding the analogous question in the L_k^n setting seems very interesting. Interestingly, for small values of k, probability 0.5 is not the correct threshold to look at. See Appendix C for details.

Derandomized Direct Sum Testing As explained in the introduction, one of the motivations for the direct sum testing question came from the potential of constructing new PCPs that rely on direct sums as opposed to direct products which incur an increase in alphabet size. Since the direct sum encoding of an n-bit string has length n^k, it is not very efficient, and a derandomized version of the direct sum test is well called for. Similar results for direct products are already known [Din07, IJKW10, DM11] and can provide a good starting point. This question is closely related to the recent work of Kaufman and Lubotzky [KL14], who discovered an interesting connection between the high dimensional expanders and property testing.

Acknowledgement

We are thankful to Oded Goldreich for many helpful discussions, and for encouraging us to include Section 9 in the paper.

References

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron, *Testing reed-muller codes*, IEEE Transactions on Information Theory **51** (2005), no. 11, 4032–4039.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy, *Proof verification and the hardness of approximation problems*, J. ACM **45** (1998), no. 3, 501–555.

[AS98] Sanjeev Arora and Shmuel Safra, *Probabilistic checking of proofs: A new characterization of NP*, J. ACM **45** (1998), no. 1, 70–122.

[BCH+96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan, *Linearity testing in characteristic two*, IEEE Transactions on Information Theory **42** (1996), no. 6, 1781–1795.

[BCLR08] Michael Ben-Or, Don Coppersmith, Mike Luby, and Ronitt Rubinfeld, *Non-abelian homomorphism testing, and distributions close to their self-convolutions*, Random Struct. Algorithms **32** (2008), no. 1, 49–70.

[BKS+10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman, *Optimal testing of reed-muller codes*, Proceedings of The 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, 2010.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld, *Self-testing/correcting with applications to numerical problems*, J. Comput. Syst. Sci. **47** (1993), no. 3, 549–595.

[BSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson, *Randomness-efficient low degree tests and short pcps via epsilon-biased sets*, Proceedings of the 35th annual ACM Symposium on Theory of Computing, 2003, pp. 612–621.

[Cha13] Siu On Chan, *Approximation resistance from pairwise independent subgroups*, Proceedings of the 45th annual ACM Symposium on Theory of Computing, 2013, pp. 447–456.

[DG08] Irit Dinur and Elazar Goldenberg, *Locally testing direct product in the low error range*, Proceedings of the 49Th IEEE Symposium On Foundations Of Computer Science, 2008, pp. 613–622.

[DG10] ———, *The structure of winning strategies in parallel repetition games*, Proceedings of the 14th RANDOM, 2010, pp. 518–530.

[Din07] Irit Dinur, *The PCP theorem by gap amplification*, J. ACM **54** (2007), no. 3.

[DM11] Irit Dinur and Or Meir, *Derandomized parallel repetition via structured pcps*, Computational Complexity **20** (2011), no. 2, 207–327.

[DR06] Irit Dinur and Omer Reingold, *Assignment testers: Towards a combinatorial proof of the PCP theorem*, SIAM J. Comput. **36** (2006), no. 4, 975–1024.

[DS13] Irit Dinur and David Steurer, *Direct product testing*, ECCC TR13-179 (2013).

[GGL95] R. L. Graham, M. Grötschel, and L. Lovász (eds.), *Handbook of combinatorics (vol. 1)*, MIT Press, Cambridge, MA, USA, 1995.

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson, *On Yao's XOR lemma*, http://eccc.hpi-web.de/report/1995/050/.

[GS00] Oded Goldreich and Shmuel Safra, *A combinatorial consistency lemma with application to proving the pcp theorem*, SIAM J. Comput. **29** (2000), no. 4, 1132–1154.

[Hoe63] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), no. 301, 13–30.

[HVV06] Alexander Healy, Salil P. Vadhan, and Emanuele Viola, *Using nondeterminism to amplify hardness*, SIAM Journal on Computing **34** (2006), no. 4, 903–931.

[IJKW10] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson, *Uniform direct product theorems: Simplified, optimized, and derandomized*, SIAM J. Comput. **39** (2010), no. 4, 1637–1665.

[IKW09] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson, *New direct-product testers and 2-query PCPs*, Proceedings of the 41st annual ACM symposium on Theory of computing (New York, NY, USA), ACM, 2009, pp. 131–140.

[KL14] Tali Kaufman and Alexander Luobtzky, *High dimensional expanders and property testing*, Proceedings of the 5th ITCS, 2014.

[KLX07] Tali Kaufman, Simon Litsyn, and Ning Xie, *Breaking the epsilon-soundness bound of the linearity test over GF(2)*, SIAM Journal on Computing **39** (2007), no. 5, 1988–2003.

[KS09] Swastik Kopparty and Shubhangi Saraf, *Tolerant linearity testing and locally testable codes*, Proceedings of the 12th RANDOM, 2009, pp. 601–614.

[LM06] Nathan Linial and Roy Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica **26** (2006), no. 4, 475–487.

[O'D02] Ryan O'Donnell, *Hardness amplification within np*, Proceedings of the 34th annual ACM symposium on Theory of computing (New York, NY, USA), STOC '02, ACM, 2002, pp. 751–760.

[Odl95] Andrew Odlyzko, *Asymptotic enumeration method*, Handbook of Combinatorics (1995), 1063–1229.

[Raz] Ran Raz, *A parallel repetition theorem*, SIAM Journal on Computing **27**, no. 3, 763–803.

[STV01] Madhu Sudan, Luca Trevisan, and Salil P. Vadhan, *Pseudorandom generators without the xor lemma*, J. Comput. Syst. Sci. **62** (2001), no. 2, 236–266.

[SW04] Amir Shpilka and Avi Wigderson, *Derandomizing homomorphism testing in general groups*, Proceedings of the 36th annual ACM Symposium on Theory of Computing, ACM, 2004, pp. 427–435.

[Tre03] Luca Trevisan, *List-decoding using the xor lemma*, Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (Washington, DC, USA), FOCS '03, IEEE Computer Society, 2003, pp. 126–.

[Yao82] Andrew C. Yao, *Theory and application of trapdoor functions*, Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (Washington, DC, USA), SFCS '82, IEEE Computer Society, 1982, pp. 80–91.

A Proof of Proposition 2.4

Proof. It is clearly enough to prove that for any non-constant linear function $\phi : L_k^n \to \{0,1\}$ it holds that $c \leq \Pr_{x \in L_k^n}[\phi(x) = 1] \leq 1 - c$.

Note first that trivially $\Pr[\phi(x) = 1] \in (c_1, 1 - c_1)$ where $c_1 = \frac{1}{\binom{n}{pn}} > \exp(-\Omega(n))$. We prove below that $\Pr[\phi(x) = 1] \in (c_2, 1 - c_2)$ for $c_2 \geq \min\{\frac{p}{4(1-p)}, \frac{1-p}{4p}\} - \exp(-\Omega(n))$ where the constant in the $\Omega()$ notation depends only on p. Taking $c = \max\{c_1, c_2\}$ clearly implies the statement of the proposition.

Let $s \in \{0,1\}^n$ be a string that represents the function ϕ, i.e., $\phi(x) = \sum_{i \in [n]} s_i x_i \pmod 2$ for all $x \in L_k^n$, and let $m = |s|$. Note that we may assume without loss of generality that $1 \leq m \leq n/2$.[4] Observe that the probability over $x \in L_k^n$ that $\phi(x) = 1$ equals to the probability

[4] Otherwise, if $m > n/2$, then we may consider the function $\bar{\phi}$ defined by $\bar{s} = (1 - s_1, \ldots, 1 - s_n) \in \{0,1\}^n$, and then $\bar{\phi}(x) = \phi(x) + k \pmod 2$ for all $x \in L_k^n$, and so we have either $\Pr[\bar{\phi}(x) = 1] = \Pr[\phi(x) = 1]$ or $\Pr[\bar{\phi}(x) = 1] = 1 - \Pr[\phi(x) = 1]$.

that for a fixed string $y \in L_k^n$ and a random string $\sigma \in L_m^n$ it holds that $\sum_{i\in[n]} \sigma_i y_i = 1 \pmod 2$. Therefore, it is enough to prove that for a random $\sigma \in L_m^n$ it holds that

$$\Pr\left[\sum_{i\in[n]} \sigma_i y_i = 1 \pmod 2\right] \in (c, 1-c).$$

Let us first choose $m-1$ coordinates of σ at random, and denote them by $\sigma' \in L_{m-1}^n$.

Claim A.1. *With probability at least* $1-\exp(-\Omega(n))$ *we have* $|\{i \in [n] \setminus \sigma' : i \in y\}| \geq pn/4$, *and* $|\{i \in [n] \setminus \sigma' : i \notin y\}| \geq pn/4 \geq (1-p)n/4$.

Proof. If $m < \min\{pn/2, (1-p)n/2\}$, then by the assumption that $|y| = pn$ the claim holds trivially. Otherwise, by standard concentration bounds (see, e.g., [Hoe63]) with probability $1-\exp(-\Omega(n))$ we have $|\sigma' \cap y| \leq 1.5pm < 3pn/4$ and $|\sigma' \cap ([n] \setminus y)| \leq 1.5(1-p)m < 3(1-p)n/4$ which clearly implies the claim. □

Hence, if we let $\ell_1 = |\{i \in [n] \setminus \sigma' : i \in y\}|$ and $\ell_2 = |\{i \in [n] \setminus \sigma' : i \notin y\}|$, then with high probability we have $\ell_1 \in (pn/4, pn)$ and $\ell_2 \in ((1-p)n/4, (1-p)n))$. Therefore, the probability that the last bit of σ is chosen to be in y is between $\frac{p}{4(1-p)}$ and $\frac{1-p}{4p}$. This completes the proof of the proposition. □

B Proof of Proposition 8.3

Proposition 8.3 restated: *For all* $n \in \mathbb{N}$ *we have*

$$\sum_{i=0}^{2n} \left| \frac{\binom{2n}{i}^2}{\binom{4n}{2n}} - \sum_{j=0}^{i} \frac{\binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2} \right| = \tilde{O}(1/\sqrt{n}).$$

Proof. By symmetry between i and $2n-i$ it is enough to bound the sum running from 0 to n. Note also that we may consider only $i \geq n - \sqrt{n}\log(n)$. Indeed, we have

$$\sum_{j=0}^{i} \frac{\binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2} \leq \left(\sum_{j=0}^{i} \frac{\binom{n}{j}\binom{n}{i-j}}{\binom{2n}{n}} \right)^2 = \left(\frac{\binom{2n}{i}}{\binom{2n}{n}} \right)^2,$$

and hence, the contribution of $i \leq n - \sqrt{n}\log(n)$ is upper bounded by

$$\sum_{i=0}^{n-\sqrt{n}\log(n)} \frac{\binom{2n}{i}^2}{\binom{4n}{2n}} + \frac{\binom{2n}{i}^2}{\binom{2n}{n}^2} = 2^{-\Omega(\log^2(n))} \ll \tilde{O}(1/\sqrt{n}).$$

Similarly, the contribution of the terms where $j \notin [n/2 - \sqrt{n}\log(n), n/2 + \sqrt{n}\log(n)]$ is $2^{-\Omega(\log^2(n))} \ll \tilde{O}(1/\sqrt{n})$. Therefore, it is enough to bound the sum

$$\sum_{i=n-\sqrt{n}\log(n)}^{n} \left| \frac{\binom{2n}{i}^2}{\binom{4n}{2n}} - \sum_{j=n/2-\sqrt{n}\log(n)}^{n/2+\sqrt{n}\log(n)} \frac{\binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2} \right| \tag{16}$$

In order to do it we use the following two estimations of binomial coefficients, which follow from Stirling's formula (see, e.g., [Odl95]). For all n the binomial coefficient $\binom{2n}{n}$ can be estimated as

$$\binom{n}{\lfloor n/2 \rfloor} = \frac{2^n}{\sqrt{\pi n/2}} \cdot \left(1 \pm O(\frac{1}{n})\right). \tag{17}$$

and for all k, n such that $|k - n/2| < 2\sqrt{n}\log(n)$ the binomial coefficient $\binom{n}{k}$ can be estimated as

$$\binom{n}{k} = \binom{n}{\lfloor n/2 \rfloor} \cdot \exp\left(-\frac{(n/2-k)^2}{n/2}\right) \cdot \left(1 \pm O(\frac{\log(n)}{\sqrt{n}})\right), \tag{18}$$

where in both estimates the $O()$ notation hides some absolute constant independent of k and n. This immediately gives us an estimate on the term $\frac{\binom{2n}{i}^2}{\binom{4n}{2n}}$

Claim B.1. *For all n sufficiently large and for $i \in [n - \sqrt{n}\log(n), n]$ we have*

$$\frac{\binom{2n}{i}^2}{\binom{4n}{2n}} = \sqrt{\frac{2}{\pi n}} \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot \left(1 \pm \tilde{O}(\frac{1}{\sqrt{n}})\right).$$

Proof.

$$\begin{aligned}
\frac{\binom{2n}{i}^2}{\binom{4n}{2n}} &= \frac{\binom{2n}{n}^2 \cdot \exp\left(-\frac{2(n-i)^2}{n}\right)}{\binom{4n}{2n}} \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})) \\
&= \frac{2^{4n}/(\pi n)}{2^{4n}/\sqrt{2\pi n}} \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})) \\
&= \sqrt{\frac{2}{\pi n}} \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})).
\end{aligned}$$

□

Now we move to the sum $\frac{\sum_j \binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2}$.

Claim B.2. *For all n sufficiently large, for $i \in [n - \sqrt{n}\log(n), n]$ and for $j \in [n/2 - \sqrt{n}\log(n), n/2 + \sqrt{n}\log(n)]$ we have*

$$\frac{\sum_j \binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2} = \sqrt{\frac{2}{\pi n}} \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot \left(1 \pm \tilde{O}(\frac{1}{\sqrt{n}})\right).$$

Proof. By the assumption that $i \in [n - \sqrt{n}\log(n), n]$ and $j \in [n/2 - \sqrt{n}\log(n), n/2 + \sqrt{n}\log(n)]$ we have $(i - j) \in [n/2 - 2\sqrt{n}\log(n), n/2 + \sqrt{n}\log(n)]$, and so using Equation (18) we get

$$\begin{aligned}
\binom{n}{j}\binom{n}{i-j} &= \binom{n}{n/2}^2 \cdot \exp\left(-\frac{(n/2-j)^2 + (n/2-(i-j))^2}{n/2}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})) \\
&= \binom{n}{n/2}^2 \cdot \exp\left(-\frac{4(j-i/2)^2}{n}\right) \cdot \exp\left(-\frac{(n-i)^2}{n}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_j \frac{\binom{n}{j}^2 \binom{n}{i-j}^2}{\binom{2n}{n}^2} &= \sum_j \frac{\binom{n}{n/2}^4}{\binom{2n}{n}^2} \cdot \exp\left(-\frac{8(j-i/2)^2}{n}\right) \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})) \\
&= \frac{4}{\pi n} \cdot \exp\left(-\frac{2(n-i)^2}{n}\right) \cdot \sum_j \exp\left(-\frac{8(j-i/2)^2}{n}\right) \cdot (1 \pm \tilde{O}(\frac{1}{\sqrt{n}})).
\end{aligned}$$

Note that although the sum $\sum_j \exp\left(-\frac{8(j-i/2)^2}{n}\right)$ runs over j from $n/2-\sqrt{n}\log(n)$ to $n/2+\sqrt{n}\log(n)$, we may extend it the sum from $-\infty$ to $+\infty$ by losing $(1+O(2^{-\log^2(n)}))$ multiplicative factor. We estimate the sum from $-\infty$ to $+\infty$ by the moving to the corresponding integral.

$$\sum_{j=-\infty}^{\infty} \exp\left(-\frac{8(j-i/2)^2}{n}\right) = \sum_{j=-\infty}^{\infty} \exp\left(-\frac{8j^2}{n}\right) = \int_{-\infty}^{\infty} \exp\left(-\frac{8x^2}{n}\right) dx + O(1) = \sqrt{\frac{\pi n}{8}} + O(1).$$

Finally, we get

$$\begin{aligned}\sum_j \frac{\binom{n}{j}^2\binom{n}{i-j}^2}{\binom{2n}{n}^2} &= \frac{4}{\pi n}\cdot\exp\left(-\frac{2(n-i)^2}{n}\right)\cdot\left(\sqrt{\frac{\pi n}{8}}+O(1)\right)\cdot(1\pm\tilde{O}(\frac{1}{\sqrt{n}})) \\ &= \sqrt{\frac{2}{\pi n}}\cdot\exp\left(-\frac{2(n-i)^2}{n}\right)\cdot(1\pm\tilde{O}(\frac{1}{\sqrt{n}})),\end{aligned}$$

as required. □

Combining Claims B.1 and B.2 the sum in Equation (16) can be bounded by

$$\begin{aligned}\sum_{i=n-\sqrt{n}\log(n)}^{n}\left|\frac{\binom{2n}{i}^2}{\binom{4n}{2n}}-\sum_j\frac{\binom{n}{j}^2\binom{n}{i-j}^2}{\binom{2n}{n}^2}\right| &\le \sqrt{\frac{2}{\pi n}}\cdot\sum_i\exp\left(-\frac{2(n-i)^2}{n}\right)\cdot\left|(1+\tilde{O}(\frac{1}{\sqrt{n}}))-(1-\tilde{O}(\frac{1}{\sqrt{n}}))\right| \\ &\le \sqrt{\frac{2}{\pi n}}\cdot\tilde{O}(\frac{1}{\sqrt{n}})\cdot\sum_{i=n-\sqrt{n}\log(n)}^{n}1 \\ &= \tilde{O}(\frac{1}{\sqrt{n}}).\end{aligned}$$

This completes the proof of Proposition 8.3. □

C Low Error Acceptance Probability Regime

In this section we give a randomized construction of a function $f: L_4^n \to \{0,1\}$ that passes the T_4^n test with probability larger than 0.51, but has negligible correlation with all linear functions. The function is defined as follows.

Let $g: L_2^n \to \{0,1\}$ be a random function. For every $x \in L_4^n$ pick $x_1, x_2 \in L_2^n$ at random such that $x = x_1 + x_2$ and let $f(x) = g(x_1) + g(x_2)$.

Claim C.1. $\Pr[T_4^n \text{ accepts } f] > 0.51$.

Proof. Pick $x, y \in L_4^n$ at random according to the distribution of T_4^n, and let $z = x + y$. If the partitions $x = x_1 + x_2$ and $y = y_1 + y_2$ and $z = z_1 + z_2$ are consistent, i.e., if $x_1 = y_1$, $x_2 = z_1$ and $y_2 = z_2$, then by definition of f it satisfies $f(x) + f(y) = f(x + y)$, and so the test clearly accepts. This happens with some constant probability bounded away from zero (taking $1/3^3$ is enough). □

On the other hand, f has only negligible correlation with any linear function. This can be shown by proving that for each linear function the correlation is small with high probability, and then use union bound.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de