Towards a General Direct Product Testing Theorem

Elazar Goldenberg
The Academic College of Tel Aviv-Yaffo, Israel
elazargo@mta.ac.il

Karthik C. S.[1]
Weizmann Institute of Science, Israel
karthik.srikanta@weizmann.ac.il

Abstract

The Direct Product encoding of a string $a \in \{0,1\}^n$ on an underlying domain $V \subseteq \binom{[n]}{k}$, is a function $\mathsf{DP}_V(a)$ which gets as input a set $S \in V$ and outputs a restricted to S. In the Direct Product Testing Problem, we are given a function $F : V \to \{0,1\}^k$, and our goal is to test whether F is close to a direct product encoding, i.e., whether there exists some $a \in \{0,1\}^n$ such that on most sets S, we have $F(S) = \mathsf{DP}_V(a)(S)$. A natural test is as follows: select a pair $(S, S') \in V$ according to some underlying distribution over $V \times V$, query F on this pair, and check for consistency on their intersection. Note that the above distribution may be viewed as a weighted graph over the vertex set V and is referred to as a test graph.

The testability of direct products was studied over various domains and test graphs: Dinur and Steurer (CCC '14) analyzed it when V equals the k-th slice of the Boolean hypercube and the test graph is a member of the Johnson graph family. Dinur and Kaufman (FOCS '17) analyzed it for the case where V is the set of faces of a Ramanujan complex, where in this case $V = O_k(n)$. In this paper, we study the testability of direct products in a general setting, addressing the question: what properties of the domain and the test graph allow one to prove a direct product testing theorem?

Towards this goal we introduce the notion of coordinate expansion of a test graph. Roughly speaking a test graph is a coordinate expander if it has global and local expansion, and has certain nice intersection properties on sampling. We show that whenever the test graph has coordinate expansion then it admits a direct product testing theorem. Additionally, for every k and n we provide a direct product domain $V \subseteq \binom{n}{k}$ of size n, called the Sliding Window domain for which we prove direct product testability.

2012 ACM Subject Classification Theory of computation → Probabilistic computation

Keywords and phrases Property Testing, Direct Product, PCP, Johnson graph, Ramanujan Complex, Derandomization

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2018.11

Acknowledgements We are truly grateful to Irit Dinur for her constant support throughout this project and for her many illuminating and helpful discussions and comments. We also thank the anonymous reviewers for their detailed and useful feedback.

[1] This work was supported by Irit Dinur's ERC-CoG grant 772839.

38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2018).
Editors: Sumit Ganguly and Paritosh Pandya; Article No. 11; pp. 11:1–11:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The direct product encoding of a function is a way to aggregate multiple values of the input function using a single query. Justifying the vague intuition that it is much harder to compute multiple values of a function rather then a single value of it, the direct product encoding has been successfully used in several contexts of hardness amplification. The hardness can either measure the fraction of inputs on which every reasonable-time algorithm fails to compute the input function, or the fraction of unsatisfied assignments of a given CNF-formula or the communication complexity of the function.

In most of the PCP constructions an assignment to the given input is broken into many tiny pieces. Each small piece is encoded individually and then one should be able to test whether these tiny pieces could be stitched together into a global assignment. This testability task is referred to as an agreement test, and instantiations of it include low degree tests such as the plane vs. plane [12], the line vs. line test [1] and the cube vs. cube test [2], and the direct product test used in [8].

More concretely, we associate the direct product encoding of strings of size n, with some underlying domain[2] V which is a collection of subsets of $[n]$ of cardinality k. Given a string $a \in \{0,1\}^n$ its direct product encoding on the domain V, denoted by $\mathsf{DP}_V(a)$, is defined as follows: For every set $S \in V$ we define $\mathsf{DP}_V(a)(S) = a|_S$ (where $a|_S$ is the restriction of a to the coordinates in S). In this paper we study the testability of this encoding, namely: Given $F : V \to \{0,1\}^k$ we want to decide whether F agrees with some $\mathsf{DP}_V(a)$ on most sets S while querying F only on a few locations, specifically two. In other words, we focus on two-query tests in the paper where we pick a pair of subsets (both in the domain) according to some fixed distribution and then check if the two subsets agree on their intersection. We say that a domain V admits a direct product testing theorem if there exists a two-query test $\mathcal{T}$ satisfying the following: For every $\varepsilon \geq 0$ and $F : V \to \{0,1\}^k$ if $\mathcal{T}$ accepts F with probability $1-\varepsilon$, then we have $F(S) = \mathsf{DP}_V(a)(S)$ for some $a \in \{0,1\}^n$ on $1 - O(\varepsilon)$-fraction of the sets S in V, where the constant behind the O notation is independent of $|V|$ and k.

This question was studied under various domains. Dinur and Steurer [9] analyzed a two-query test under the domain $V = \binom{[n]}{k}$. Recently, Dinur and Kaufman [6] studied this question in a much shrunken domain, which is obtained by considering the set of the faces of a high dimensional expander. However, both of these proofs are tailored to the structure of their own domain and cannot be (trivially) generalized to other domains. It is natural to ask whether a more generalized argument can be applied covering both of these domains, and on which domains it may be applied. The main question we are investigating is as follows:

Which domains admit a two-query direct product testing theorem?

Let us elaborate more about the previous proofs. The proofs given by [9] and [6] first analyze the testability in the high error regime, i.e. when the acceptance probability is slightly bounded away from 0. They show that any function that passes the test with non-negligible probability ε must agree with some legal codeword $\mathsf{DP}_V(a)$ on $\Omega(\varepsilon)$ fraction of sets. Then they analyze the test in the low error regime, i.e. when the acceptance probability of the test is close to 1. Finally they stitch local tiny agreements into a single codeword and show that the agreement is almost everywhere.

[2] For the ease of presentation, we only consider domains which are a subset of $\binom{[n]}{k}$ in this section. However, in the rest of the paper we consider V which is a collection of subsets of $[n]$, and all our results are proved for this more general case.

We would like to establish a direct product testing theorem using a more straightforward approach: we decode a string from the input function F using the majority operator and then show that if the test passes with high probability then F is close to the direct product encoding of the decoded string. More precisely, given the input function F, we define a string $a \in \{0,1\}^n$ as follows: for every coordinate $i \in [n]$ we set a_i to be the majority value of $F(S)_i$, where the majority is taken over the sets that contain i. Next we show that if F passes the test with probability $1-\varepsilon$ then F must be $1-O(\varepsilon)$-close to $\mathsf{DP}_V(a)$. We remark that Dinur and Reingold [8] indeed followed this proof strategy, however, their proof admits only a relaxed notion of closeness between the input function and the direct product encoding of the decoded string (namely, that on most sets S, $F(S)$ and $\mathsf{DP}_V(a)(S)$ agree *only* on most of the coordinates in S).

Observe that any two-query test on a domain V gives rise to a weighted graph whose vertex set is V and the weight we assign for each pair (S,S') is the probability of this pair being picked by the test[3]. We refer to this graph as the test graph. We say that a test graph yields a tester for the domain V, if for every $\varepsilon \geq 0$ and every function $F : V \to \{0,1\}^k$ the following holds: if the test accepts F with probability $1-\varepsilon$, then F must be $1-O(\varepsilon)$-close to some $\mathsf{DP}_V(a)$. Here the test corresponds to picking an edge (S,S') at random (according to the distribution of weights on the edges) and accepting if and only if $F(S)|_{S\cap S'} = F(S')|_{S\cap S'}$.

Another proof insight that we desire is the explicit use of the properties of the underlying test graph. For example, one property that the test graph must satisfy to be a tester is that for most edges (S,S') the intersection between S and S' is linear in k. Assume not, then we consider the following construction of F: We start from $F = \mathsf{DP}_V(a)$ for some $a \in \{0,1\}^n$ and then for each $S \in V$ we reset the value of $F(S)_i$ for some random $i \in S$. Then for most sets (S,S') with small intersection the test accepts but F is far from any direct product codeword. Another property that the test graph must have is some notion of expansion. Summing up, our more refined question is as follows:

> *What properties of the test graph yields a tester for its underlying domain?*

1.1 Our Results

Our conceptual contributions in this paper are two-fold. First, we introduce a notion called *coordinate expansion* which captures the properties of direct product testable domains. Second, we introduce the sliding window domain which is of size exactly *equal* to the universe and is direct product testable. Our main technical contribution is showing that domains having coordinate expansion with certain parameters admit a direct product theorem.

1.1.1 A General Direct Product Theorem

We introduce below the notion of coordinate expansion. Informally, a coordinate expander has both global and local expansion properties, and has good intersection properties.

▶ **Definition 1** ((λ,ρ)-Coordinate Expander)**.** Let $G = (V,E)$ be a test graph, where $V \subseteq \binom{[n]}{k}$. For $i \in [n]$ let $V_i = \{S \in V | i \in S\}$ and G_i be the subgraph of G induced by the vertices in V_i. The graph G is called (λ,ρ)-coordinate expander if:

[3] In this paper we analyze test graphs which are undirected.

1. $\lambda(G) < \lambda$ (where $\lambda(G) = \max\{|\lambda_2(A_G)|, |\lambda_{|V|}(A_G)|\}$ and A_G is the normalized adjacency matrix of G).
2. For every $i \in [n]$ we have that $\lambda(G_i) < \lambda$ and for each $S \in V_i$ the probability that a uniformly random neighbor S' of S is in V_i is at least ρ.
3. For every subset S and $T \subseteq S$, satisfying $|T| \geq 2/\rho$, the probability that for a uniformly random neighbor S' of S we have $|S' \cap T| \leq \rho|T|/2$ is upper bounded bounded by λ .

Notice that condition 1 implies that the test graph must be a good expander (in the traditional sense). Moreover, condition 2, implies that on certain local subsets (i.e., subsets containing a common coordinate) of vertices, the induced subgraph must be expanding as well. Finally, condition 3 implies that the neighbors of every subset S samples well every subset T of S.

Observe that condition 2 is necessary for the test graph to be a direct product tester. To see this, consider a test graph that does not satisfy this property, namely, there exists a coordinate $i \in [n]$ for which: there exits a set $B_i \subset V_i$ such that $\Pr_{S' \in V_i}[S' \notin B_i | S \in B_i] = o(1)$. Then, we show that the test graph does not yield a tester. Indeed, consider the following construction of F: we first choose $F = \mathsf{DP}_V(a)$ for some $a \in \{0,1\}^n$. Then for every $S \in B_i$ we change the value of $F(S)_i$ to $1 - a_i$. Clearly, the distance of F from a direct product encoding equals $\delta := |B_i| / |V|$. However, the rejection probability equals:

$$2 \cdot \Pr_{S' \sim S}[S \in B_i \text{ and } S' \in V_i \backslash B_i] \leq 2 \cdot \Pr[S \in B_i] \cdot \Pr[S' \in V_i] \cdot \Pr_{S' \in V_i}[S' \notin B_i | S \in B_i] = o(1) \cdot \delta.$$

Then, we show our main technical result that coordinate expansion implies direct product testing (for a certain range of parameters).

► **Theorem 2.** *Let $\rho \geq 1/2$ and $\lambda \leq 1/33$. Let $G = (V, E)$ be a test graph, $V \subseteq \binom{[n]}{k}$, let $\varepsilon \geq 0$, and $F : V \to \{0,1\}^k$. Let G be a (λ, ρ)-coordinate expander. If F passes the test implied by the test graph G with probability $1 - \varepsilon$ then F is $1 - O(\varepsilon)$-close to $\mathsf{DP}_V(a)$ for some $a \in \{0,1\}^n$.*

The overview of the above proof is given in Section 1.2. Also, as an application of the above theorem, we show[4] a direct product theorem for the test graph isomorphic to the Johnson graph $J(n, k)$ when k is close to $n/2$, where $J(n, k)$ is a graph whose vertex set is the set of all subsets of $[n]$ of cardinality k, and two subsets have an edge if their intersection is equal to $k/2$. This should be compared to [9], where they show the direct product for the Johnson graph for all the layers up to $n/2$ (i.e., for all $J(n, k)$ where $k \leq n/2$).

The main open problem stemming from our work is to improve the parameters in Theorem 2. In particular, does the following hold?

► **Open Problem 3.** *Does $(1/2, 1/2)$-coordinate expansion imply a direct product theorem?*

A positive resolution of the above open question would imply direct product testability on the test graph isomorphic to the Johnson graph for every layer of the Boolean hypercube (completely recovering the results in [9]). It even implies a direct product testability on a new domain: where the subsets are stemming from d-dimensional subspaces of $\mathbb{F}_2^m$ and two subsets are connected by an edge if they intersect on a $(d-1)$-dimensional subspace (this is referred to as the Grassmann graph). Finally, we would like to recall that Theorem 2

[4] The claim as written here is slightly inaccurate. Please refer to Appendix B for a precise statement.

states that $(1/33, 1/2)$-coordinate expansion implies a direct product theorem, i.e., in order to positively resolve Open Problem 3, we might need to improve the analysis in the proof of Theorem 2 to accommodate test graphs with weaker expansion properties.

In fact, if we can resolve Open Problem 3 in a slightly stronger way i.e., if we show that for some small enough constant $\gamma > 0$, we have $(1/2 + \gamma, 1/2)$-coordinate expansion implies a direct product theorem then we recover the testability result of [6] on Ramanujan complexes. Summarizing, we view the study of coordinate expansion as providing a unified framework to prove direct product theorems. Also, it might be useful in the future to establish direct product testability for new domains (in a black-box manner).

1.1.2 Sliding Window Domain

In this subsubsection, we define a new direct product testable domain which we call the sliding window domain, and also discuss about the necessary and sufficient structure that a domain (and test graph) should have, in order to admit direct product testing.

For every n, k, the sliding window domain $\mathcal{A} \subseteq \binom{[n]}{k}$ is the collection of all contiguous k-sized subsets (windows) of $[n]$, i.e., $\mathcal{A} = \{\{i, \ldots, i + k - 1\} \mid i \in [n]\}$, where the addition is done modulo n. Two vertices (i.e., subsets in $\mathcal{A}$) have an edge in the test graph, if their intersection is non-empty. Notice that $|\mathcal{A}| = n$ and yet we show that it admits a direct product theorem (see Theorem 9 for a simple proof).

Let us put the above result in context with the recent breakthrough of Dinur and Kaufman [6]. In [6], the authors obtain a direct product testable domain (subset of $\binom{[n]}{k}$) of size $O(2^{k^2} n)$. The domain arises from the highly non-trivial object called Ramanujan complex. Such a domain is studied because apart from admitting a direct product theorem over a domain of size linear in the universe (i.e., n), it also has other desirable properties such as *distance amplification* which are needed for applications in gap and hardness amplification. Thus, our direct product testing result (Theorem 9) provides a conceptual clarification that if one is only interested in direct product testing as a property testing question, then there is a very simple domain of size n, namely the sliding window domain, which is testable.

Roughly speaking, a domain (subset of $\binom{[n]}{k}$) has distance amplification if for every two strings of relative distance δ, the relative distance between their direct product encoding is $\Omega(k\delta)$. This seems to be a crucial property for PCP applications of direct product testing. Thus, the construction of the sliding window domain provides a conceptual clarification as to why we need high dimensional expanders: we can obtain direct product testing from simple constructions like the sliding window domain and we can obtain distance amplification from known constructions of vertex expanders (see Appendix D); but to obtain both simultaneously, [6] needed high dimensional expanders. We leave it as an open question whether there exists a simple construction admitting both direct product testability and distance amplification.

▶ **Open Problem 4.** *Is there a (relatively) simple domain of linear size in the universe (i.e., n) for which we have both direct product testing and distance amplification?*

Lack of Global Expansion. We would like to now briefly discuss about the minimal structure of the domain (and the test graph) sufficient to prove a direct product theorem. This is highlighted by the sliding window domain, an in particular by the proof of its testability (Lemma 10 to be precise). Notice that $G_{\mathcal{A}}$ has very bad edge-expansion/vertex-expansion but is a very good local expander, i.e., the induced subgraph containing any particular coordinate has good expansion (in fact is a clique). Lemma 10 guarantees that in such situations[5] the domain admits direct product testing if for every vertex in the test graph, and every element in that vertex, the probability of retaining that element when moving to a uniformly random neighbor is bounded from below by a positive constant. The probability of retaining a coordinate when moving to a random neighbor is $1/2$ in $\mathcal{A}$, and thus $\mathcal{A}$ admits a direct product theorem. Therefore, $\mathcal{A}$ demonstrates that direct product testing does not require the test graph to be an expander (like the Johnson/Ramanujan graph) but only needs to have certain local expansion properties. Finally, recall that we had earlier argued that local expansion is necessary (to justify the need for condition 2 in Definition 1) for direct product testing.

Finally, it seems that conditions 1 and 3 in coordinate expansion are not (necessarily) needed for direct product testing, but are merely artifacts of our proof (Theorem 2). However, these conditions might imply distance amplification[6] and are typically guaranteed in structured domains of interest (namely, Johnson, Grassmannian, and Ramanujan).

1.2 Technical Contribution: Proof Overview of Theorem 2

For the sake of convenience, through out this subsection, we fix $V = \binom{[n]}{k}$ and the test would pick pairs (S, S') that intersect on $k/2$ elements and checks for agreement. As suggested above there is a natural way to decode any function $F : V \to \{0,1\}^k$ using the majority operator: define a string $a \in \{0,1\}^n$ by setting a_i to be the majority value of $F(S)_i$ for all $S \ni i$. We define $B = \{S | F(S) \neq \mathsf{DP}_V(a)(S)\}$, i.e., B is the subset of the domain that disagrees with the direct product encoding of the decoded string. Also for $S \in B$ we call $i \in S$ conflicting if $F(S)_i \neq a_i$. Our goal is to show that the test rejects with probability $\Omega(|B| \,/\, |V|)$ as $|B|/|V|$ is the relative distance between F and $\mathsf{DP}_V(a)$.

Indeed fix $S \in B$, then it must contain at least one conflicting coordinate, say i. Observe that with probability $1/2$ we also have that $i \in S'$. Now if S' were a random set containing i, then since at least half of the elements that contain i agree with the majority value, the test rejects with probability $1/2$. And the overall rejection probability of the test would be at least $\frac{|B|}{4|V|}$ and we are done.

However, S' is not a random set that contains i, it intersects with S on further $k/2 - 1$ coordinates. Therefore, it may well be that among the neighbors of S that contain i we do not see the majority value so often. A natural way to overcome this is by aggregating all Ss' that contain i and disagree with the majority value on i. We could try to show that if we start from some member of this set then with constant probability we reach S' that contains i and resides outside of this aggregated set (by using the local expansion property). But this leads into another problem: using this argument sets S that contain many conflicting coordinates are counted many times, whereas sets that contain few conflicting coordinates are counted much less.

Our analysis proceeds by studying the variance of the number of conflicting coordinates in the following manner. We first sort the set B based on the number of their conflicting coordinates. Let B_L (resp. B_H) be the first (resp. last) third of the elements in B according the sorting. We first show that if the number of conflicting coordinates of each member in B_L is much smaller than it is in B_H, then the test rejects with probability $\Omega(\frac{|B|}{|V|})$. To show this, we prove that whenever the test picks $S \in B_H$ then with constant probability S' is in

[5] Lemma 10 can be generalized to accommodate test graphs which are locally subgraphs that strongly satisfy the expander mixing lemma.

[6] This would be an interesting question to resolve in either direction.

$B_L \cup \{V \setminus B\}$ (by using the global expansion property). Moreover, there is a large subset Γ of conflicting coordinates in S which are also in S' (follows from condition 3 in Definition 1). However, S' has few conflicting coordinates in total (by our choice of S'), and thus, there must be a coordinate in Γ that agrees with the majority value on S' but disagrees on it on S and hence the test rejects the edge (S, S').

On the other hand, if the number of conflicting coordinates does not vary a lot among these sets, then we analyze the test by selecting (at random) a single conflicting coordinate in S and focusing on the rejection probability based only on the value of the selected coordinate.

1.3 Related Work

The question of testing the direct product was studied extensively when the underlying domain $V = \binom{[n]}{k}$ [10, 8, 5, 9, 11]. In this setting, Goldreich and Safra [10] proposed a constant query test. Dinur and Reingold [8] suggested the two-query test mentioned above and analyzed it in the high acceptance regime but with a relaxed distance measure.

The state of the art in this context is the result of Dinur and Steurer[7] [9] dealing with the domain $V = \binom{[n]}{k}$ where k varies between 2 and $n/2$. They analyze the aforementioned two-query test with $k/2$-intersection size. They analyze it in the high acceptance regime and show that $\binom{[n]}{k}$ indeed admits a direct product testing theorem. The proof is quite involved and in particular analyzes first the low acceptance regime. Recently, in a breakthrough paper, Dinur and Kaufman [6] analyzed the two-query test when the underlying domain is obtained from the set of faces of a Ramanujan complex. Their approach crucially relies on the result of [9].

We remark that the direct product testability question was further analyzed in the low acceptance regime under the domain $\binom{[n]}{k}$, see [5, 11, 7] and also under the domain where the universe is $\mathbb{F}_2^m$, and the domain is the set of all subspaces of $\mathbb{F}_2^m$ [11].

1.4 Organization of the Paper

Section 2 lists the notations and technical tools that we use in the paper. In Section 3 we formalize the notion of direct products and their testing. In Section 4 we prove our main technical result, namely, that whenever the underlying test graph is a (λ, ρ)-coordinate expander it admits a direct product testing theorem. Finally, in Section 5 we introduce the sliding window domain for which we show a direct product theorem.

2 Preliminaries

In this section, we list the notations and technical tools used in this paper.

Notations. We use the following notations throughout the paper. We denote the set $\{1, \ldots, n\}$ by $[n]$. For any $n, k \in \mathbb{N}$, with $k \leq n$, we denote by $\binom{[n]}{k}$, all subsets of $[n]$ of cardinality k. For any set S, we denote by $\mathcal{P}(S)$ the power set of S, i.e., the set of all subsets of S. For any graph $G(V, E)$ and any two subsets $S, T \subseteq V$, we denote by $E(S, T)$ the set of all edges between S and T. For any $x, y \in \{0,1\}^n$, we denote by $\Delta(x, y)$ the relative Hamming distance between x and y given by the fraction of coordinates in which x and y differ.

[7] The result in [9] is stated in the language of tuples, i.e., the domain is a subset of $[n]^k$, but their result also holds when the domain is a collection of k-sized subsets of $[n]$. See [4] for more details.

Johnson Graph Family. For every $n, k, t \in \mathbb{N}$ such that $t \leq k \leq n$, $J(n,k,t)$ is a graph which is a member of the Johnson graph family, whose vertex set is $\binom{[n]}{k}$, and whose edge set is $\{(S,S') \mid S, S' \in \binom{[n]}{k}, |S \cap S'| = t\}$.

Expander Mixing Lemma. The following is a standard claim concerning the expansion of two sets in expander graphs. For completeness we include a proof in Section A:

▶ **Claim 5.** *Let $G = (V,E)$ be a d-regular graph and A be its adjacency matrix. Let λ be its second largest eigenvalue in absolute value. Let $S, T \subseteq V$ satisfying: $|S| \leq |V|/2$ then:*

$$\Pr_{(u,v)}[v \in T | u \in S] \leq \frac{|T|}{|V|} + \frac{\lambda}{d}\sqrt{\frac{|T|}{|S|}},$$

where the probability is given by first picking u uniformly at random from S, and then picking v according to A. Furthermore, let μ be a distribution on S satisfying that for every two elements $b, b' \in S$: $\mu(b) \leq c\mu(b')$, then:

$$\Pr_{(u,v)}[v \in T \mid u \sim \mu] \leq \frac{|T|}{|V|} + \frac{\lambda}{d} \cdot \sqrt{\frac{c\,|T|}{|S|}}$$

3 Direct Product Testing: The Setting

In this section, we formalize the notion of direct products and their testing. Specifically, we formalize the notion of direct product testing through test graphs, which is slightly non-standard but it helps in introducing the notion of coordinate expansion in a later section succinctly.

For every subset S of $[n]$, let $\mathcal{F}_S$ be the class of all functions whose domain is S and range is $\{0,1\}$. Let $V \subseteq \mathcal{P}([n])$ be the domain of the direct product. Let $\mathcal{F}_V$ be the class of all functions whose domain is V and maps every subset S in V to a function in $\mathcal{F}_S$. The direct product encoding is a function $\mathsf{DP}_V : \{0,1\}^n \to \mathcal{F}_V$ defined as follows: for every string $a \in \{0,1\}^n$, and every subset $S \in V$, let $\mathsf{DP}_V(a)_S$ be defined as the projection function which maps S to a_S, the string a restricted to only the coordinates in S.

▶ **Definition 6.** For two functions $F, G \in \mathcal{F}_V$ we define their relative distance as:

$$\Delta(F,G) = \frac{|\{S \in V | F(S) \neq G(S)\}|}{|V|}.$$

For a function F and a set of functions $\tilde{G}$ we define the distance between F and $\tilde{G}$ as the minimal distance between F and some function $G \in \tilde{G}$. If $\Delta(F, \tilde{G}) \leq \delta$, we say that F is $1-\delta$-close to $\tilde{G}$, otherwise, it is δ-far from $\tilde{G}$.

For every function $F \in \mathcal{F}_V$, we define $\mathsf{dec}(F)$ as follows: Given F construct $a^F \in \{0,1\}^n$ in the following way,

$$a_i^F = \operatorname*{\mathsf{maj}}_{\substack{S \in V \\ S \ni i}}(F(S)_i).$$

Then, we define $\mathsf{dec}(F) := \mathsf{DP}_V(a^F)$.

Let G_V be a graph whose vertex set is V. Then we interpret G_V as a test graph on functions defined on $\mathcal{F}_V$ in the following sense:

Test $\mathcal{T}(G_V)$:
Input: A function $F \in \mathcal{F}_V$.
Procedure: Pick an edge (S, S') in G_V uniformly at random.
Output: Accept if and only if $F(S)|_{S\cap S'} = F(S')|_{S\cap S'}$.

It is important to note that we allow self loops and multiple edges between a pair of vertices. Also, we can generalize the above direct product testing setting to the case when V is a multiset of $\mathcal{P}([n])$, and the results in this paper still hold. However, we choose not to handle this more general setting for the sake of clarity of presentation. The above remark also applies to the case of studying test graphs which are not regular in degree, that are not considered in this paper. Finally, throughout the paper, we drop the subscript V in G_V, if V is clear from the context.

4 Direct Product Testing: Coordinate Expansion

In this section we prove our main technical result, namely, that whenever the underlying test graph is a (λ, ρ)-Coordinate Expander (defined next) it admits a direct product testing theorem.

▶ **Definition 7** ((λ, ρ)-Coordinate Expander)**.** Let $n \in \mathbb{N}$ and let $G = (V, E)$ be a test graph, where $V \subseteq \mathcal{P}([n])$. For $i \in [n]$ let $V_i = \{S \in V | i \in S\}$ and G_i be the subgraph of G induced by the vertices in V_i. Let $\lambda(G) = \max\{|\lambda_2(A_G)|, |\lambda_{|V|}(A_G)|\}$, where A_G is the normalized adjacency matrix of G. The graph G is called (λ, ρ)-coordinate expander if:

1. $\lambda(G) < \lambda$ and for every $i \in [n]$ we have $\lambda(G_i) < \lambda$.
2. For every $i \in [n]$ and for each $S \in V_i$ we have $\Pr_{S'\sim S}[S' \in V_i] \geq \rho$.
3. For every subset S and $T \subseteq S$, satisfying $|T| \geq 2/\rho$, we have $\Pr_{S'\sim S}[|S' \cap T| \leq \rho|T|/2] \leq \lambda$.

Informally, a domain is a coordinate expander if the test graph is an expander and every induced subgraph of the test graph containing a fixed coordinate is also an expander[8], and it has good correlation/intersection properties – i.e., for any subset S and coordinate $i \in S$, an uniformly random neighbor of S contains i with constant probability (say $\rho > 0$), and for every S in the domain, and any subset T of S, the number of elements of T that we see in a random neighbor of S is close to the expected number, which is $\rho \cdot |T|$. Below, we see that coordinate expansion of the test graph implies a direct product theorem for the underlying domain.

▶ **Theorem 8.** *Let $n \in \mathbb{N}$, and let $\rho \geq 1/2$ and $\lambda \leq 1/33$ be some constants. Let $G = (V, E)$ be a graph, $V \subseteq \mathcal{P}([n])$, let $\varepsilon \geq 0$, and $F \in \mathcal{F}^V$. Let G be a (λ, ρ)-coordinate expander. If F passes $\mathcal{T}(G)$ with probability $1 - \varepsilon$ then F is $1 - O(\varepsilon)$-close to $\mathsf{dec}(F)$.*

Proof. Let $F^* := \mathsf{dec}(F) = \mathsf{DP}_V(a^F)$. We define $B, C \subseteq V$ as follows:

$$B = \{S \mid F(S) \neq F^*(S)\} \text{ and } C = V \setminus B.$$

Let $\beta = |B| / |V|$. Given a subset $S \in V$ we say that a coordinate i is conflicting if the value of $F(S)$ at i does not equal a^F_i. For a set S denote by $B(S)$ the set of conflicting coordinates in S. We show that $\mathcal{T}(G)$ rejects with probability at least $\Omega(\beta)$.

[8] Actually, the property of an expander that we need is that for any two sets of vertices S, T in the graph, the number of edges between S and T is roughly equal to $\alpha|S||T|$, where α is the density of the edge set of the graph.

Let us sort in ascending order the elements of B based on the number of coordinates on which they disagree with F^*. For a parameter $0 \leq p \leq 1$ we define the set $B_{\geq p}$ as the set of last $(1-p)\,|B|$ elements of B (and similarly the set $B_{\leq p}$ is the set of the first $p\,|B|$ elements of B). We denote by m_p the number of conflicting coordinates of the $p\,|B|$-th element of B.

Let $c = 3/40$. We consider two cases based on $m_c, m_{1/2}$ and m_{1-c}.

Case 1: $m_{1-c} > \frac{2}{\rho} m_{1/2}$ or $m_{1/2} > \frac{2}{\rho} m_c$

For both the possibilities we have similar arguments, which is why they are clubbed under one case, but will be handled separately for ease of presentation.

Case 1A: $m_{1-c} > \frac{2}{\rho} m_{1/2}$

The probability that an uniformly random $S \in V$ is in $B_{\geq 1-c}$ is $c\beta$. Now by Claim 5, we get that

$$\Pr[S' \in B_{>1/2} | S \in B_{\geq 1-c}] < \beta/2 + \lambda\sqrt{\frac{1}{2c}},$$

so with probability at least $1 - \beta/2 - \lambda\sqrt{\frac{1}{2c}}$ if $S \in B_{\geq 1-c}$ then $S' \in B_{\leq 1/2} \cup C$.

Now, by the third property of (λ, ρ)-coordinate expander, the probability that $|S' \cap B(S)| \leq \frac{\rho}{2}|B(S)|$ is at most λ. Notice that the probability that $|S' \cap B(S)| \leq m_{1/2}$ is at least the probability that $|S' \cap B(S)| \leq \frac{\rho}{2}|B(S)|$ (because $m_{1/2} < \frac{\rho}{2} m_{1-c} \leq \frac{\rho}{2}|B(S)|$). Hence we have that the probability that $|S' \cap B(S)| \leq m_{1/2}$ is at most λ.

Overall, using union bound, conditioned on $S \in B_{\geq 1-c}$, the probability that $S' \in B_{\leq 1/2} \cup C$ and $|S' \cap B(S)| > m_{1/2}$ is at least $1 - \beta/2 - \lambda\sqrt{\frac{1}{2c}} - \lambda$. But in such a case since $S' \in B_{\leq 1/2} \cup C$ we get $|B(S')| \leq m_{1/2}$, so there exists at least one coordinate $i \in S \cap S'$ on which $F(S')_i = a_i^F$ but $F(S)_i \neq a_i^F$, so the test rejects. In total $\mathcal{T}$ rejects with probability at least $c\beta\left(1 - \beta/2 - \lambda\sqrt{\frac{1}{2c}} - \lambda\right) \geq c\beta\left(1/2 - \lambda\left(\sqrt{\frac{1}{2c}} + 1\right)\right)$ (where we used a trivial bound that $\beta \leq 1$). Notice that $1/2 - \lambda\left(\sqrt{\frac{1}{2c}} + 1\right) > 0$ holds for $c = 3/40$ whenever $\lambda \leq 0.13$.

Case 1B: $m_{1/2} \geq \frac{2}{\rho} m_c$

In this case we would like to mimic the proof strategy of the previous case. That is we would like to show that with non-zero constant probability a random neighbor in $B_{\geq 1/2}$ is in $B_{\leq c} \cup C$. By an application of Claim 5, we get:

$$\Pr[S' \in B_{>c} | S \in B_{\geq 1/2}] < (1-c)\beta + \lambda\sqrt{2-2c},$$

so with probability at least $1 - (1-c)\beta - \lambda\sqrt{2-2c}$ if $S \in B_{\geq 1/2}$ then $S' \in B_{\leq c} \cup C$.

Now, by the third property of (λ, ρ)-coordinate expander, the $\Pr[|S' \cap B(S)| \leq \frac{\rho}{2}|B(S)|]$ is at most λ. Notice that $m_c \leq \frac{\rho}{2} m_{1/2} \leq \frac{\rho}{2} \cdot |B(S)|$ and thus $\Pr[|S' \cap B(S)| \leq \frac{\rho}{2}|B(S)|] \geq \Pr[|S' \cap B(S)| \leq m_c]$. Therefore we have $\Pr[|S' \cap B(S)| \leq m_c] \leq \lambda$.

Overall, using union bound, conditioned on $S \in B_{\geq 1/2}$, the probability that $S' \in B_{\leq c} \cup C$ and $|S' \cap B(S)| > m_c$ is at least $1 - (1-c)\beta - \lambda\sqrt{2-2c} - \lambda$. But in such a case since $S' \in B_{\leq c} \cup C$ we get $|B(S')| \leq m_c$, so there exists at least one coordinate $i \in S \cap S'$ on which $F(S')_i = a_i^F$ but $F(S)_i \neq a_i^F$, so the test rejects. In total $\mathcal{T}$ rejects with probability at least $\frac{\beta}{2}\left(1 - (1-c)\beta - \lambda\sqrt{2-2c} - \lambda\right) \geq \frac{\beta}{2}\left(c - \lambda\left(\sqrt{2-2c} + 1\right)\right)$ (where we used a trivial bound that $\beta \leq 1$). Notice that $\left(c - \lambda\left(\sqrt{2-2c} + 1\right)\right) > 0$ holds for $c = 3/40$ whenever $\lambda \leq 0.03177$.

Case 2: $m_{1-c} \leq \frac{4}{\rho^2} m_c$

Define $B_{(c,1-c)}$ as the set $B \setminus (B_{\leq c} \cup B_{\geq 1-c})$. Observe that in $B_{(c,1-c)}$ the number of conflicting coordinates is between m_c and $4m_c/\rho^2$. Now we would like to consider a different test $\mathcal{T}'(G)$ that selects S, S' according to G. If $S \notin B_{(c,1-c)}$ then $\mathcal{T}'$ accepts. Otherwise, it picks uniformly at random $i_0 \in B(S)$ and checks for consistency *only* on i_0, namely: It rejects iff $i_0 \in S'$ and $F(S)_{i_0} \neq F(S')_{i_0}$. Clearly the rejection probability of $\mathcal{T}'(G)$ is at most the rejection probability of $\mathcal{T}(G)$. We conclude the proof by showing that $\mathcal{T}'(G)$ rejects F with probability $\Omega(\beta)$.

With probability $(1-2c)\beta$ the test $\mathcal{T}'$ selects $S \in B_{(c,1-c)}$ and we would like to analyze the rejection probability conditioned on that. For this sake we bound the probability of the following events:

- E_1 is the event where $S' \in B_{\leq c} \cup B_{\geq 1-c}$.
- E_2 is the event where $i_0 \in S'$ and $S' \notin \tilde{B}_{i_0}$ where $\tilde{B}_i = \{S \in B_{(c,1-c)} | F(S)_i \neq a_i^F\}$.

If the event E_2 occurs but E_1 does not, then it must be the case that $F(S')_{i_0} = a_{i_0}^F$. Hence $\mathcal{T}'$ rejects. As a consequence $\Pr[\mathcal{T}' \text{ rejects}] \geq (1-2c)\beta(\Pr[E_2|S \in B_{(c,1-c)}] - Pr[E_1|S \in B_{(c,1-c)}])$. Thus it suffices to show that $(\Pr[E_2|S \in B_{(c,1-c)}] - Pr[E_1|S \in B_{(c,1-c)}])$ is a positive constant bounded away from 0.

To bound the probability for the event E_1 we use Claim 5: The probability of E_1 conditioned on $S \in B_{(c,1-c)}$ is at most $2c\beta + \lambda\sqrt{\frac{2c}{1-2c}}$.

Since the graph G is a (λ, ρ)-coordinate expander then for each $i \in S$, we have that $\Pr[i \in S'] \geq \rho$, in particular this is true for i_0, hence: $\Pr[i_0 \in S'] \geq \rho$.

Now we divide the event E_2 into disjoint events depending on the value of i_0 and bound the rejection probability of $\mathcal{T}'$ conditioned on specific value of i_0. Fix $i \in [n]$ and assume that $\mathcal{T}'$ selects $S, S' \in V_i$ and sets $i_0 = i$ (so $S \in \tilde{B}_i$). We denote by β_i the fraction $\frac{|\tilde{B}_i|}{|V_i|}$. Observe that $\beta_i \leq 1/2$, since otherwise the majority value would become the value of $F(S)_i$, but we have $S \in \tilde{B}_i$.

Note, that under the assumption that $\mathcal{T}'$ selects $i_0 = i$ and $S \in \tilde{B}_i$, sets S with few conflicting coordinates are more likely to be chosen than those who have many of them. However, since by our assumption the number of conflicting coordinates is between m^* and $\frac{4}{\rho^2} m^*$, then sets with m^* conflicting coordinates are only $4/\rho^2$-times more probable than those having $\frac{4}{\rho^2} m^*$-conflicting coordinates. Denote by μ the distribution of picking $S \in \tilde{B}_i$ assuming that $\mathcal{T}'$ selects $i_0 = i$. By an application of Claim 5 we get:

$$\Pr_{S \sim \mu, S'}[S' \in \tilde{B}_i] \leq \frac{|\tilde{B}_i|}{|V_i|} + \lambda\sqrt{\frac{4}{\rho^2}} \leq \frac{1}{2} + 2\lambda/\rho$$

So we get that,

$$\Pr[E_2|S \in B_{(c,1-c)}] = \left(1 - \Pr_{S \sim \mu, S'}[S' \in \tilde{B}_i \mid i_0 \in S']\right) \cdot \Pr[i_0 \in S'] \geq \frac{\rho}{2} - 2\lambda.$$

Summing up, we get:

$$\begin{aligned}
\Pr[\mathcal{T}' \text{ rejects}|S \in B_{(c,1-c)}] &\geq \Pr[E_2|S \in B_{(c,1-c)}] - \Pr[E_1|S \in B_{(c,1-c)}] \\
&\geq \frac{\rho}{2} - 2\lambda - \left(2c + \lambda\sqrt{\frac{2c}{1-2c}}\right) \\
&\geq \frac{1}{4} - 2c - \lambda\left(2 + \sqrt{\frac{2c}{1-2c}}\right),
\end{aligned}$$

a constant bounded away from 0 for $c = 3/40$ whenever $\lambda < 0.04$. ◀

In Appendix B, we consider the test graph $J(n, k, k/2)$ and show a direct product theorem when k is close to $n/2$.

5 Sliding Window Domain

In this section, we introduce the sliding window domain for which we show a direct product theorem.

Construction. Let $k, n \in \mathbb{N}$ such that $k \leq n$. Let $\mathcal{A}$ be a collection of n subsets of $[n]$ of Hamming weight k.

$$\mathcal{A} = \{\{i, \ldots, i + k - 1\} \mid i \in [n]\},$$

where the addition is done[9] modulo n.

Testability. The domain of our direct product test is $\mathcal{A}$. The corresponding test is as follows:

Test $\mathcal{T}$:
Input: A function $F : \mathcal{A} \to \{0, 1\}^k$.
Procedure: Pick uniformly at random $S \in \mathcal{A}$. Then pick uniformly at random $S' \in \mathcal{A}$ such that $S \cap S' \neq \emptyset$.
Output: Accept if and only if $F(S)|_{S \cap S'} = F(S')|_{S \cap S'}$.

The test graph $G_{\mathcal{A}}$ of the above is given by the vertex set $\mathcal{A}$ and the edge set $\{(S, S') \mid S \cap S' \neq \emptyset\}$. The correctness of the above test is shown below. We would like to emphasize that $|\mathcal{A}| = n$ and yet admits a direct product theorem.

▶ **Theorem 9.** *Let $\varepsilon \geq 0$ and $F \in \mathcal{F}_{\mathcal{A}}$. If F passes $\mathcal{T}(G_{\mathcal{A}})$ with probability $1 - \varepsilon$ then F is $(1 - 4\varepsilon)$-close to* $\mathsf{dec}(F)$.

Proof. We will in fact prove a more general direct product testing result.

▶ **Lemma 10.** *Let $n \in \mathbb{N}$ and $G = (V, E)$ be a d-regular graph where $V \subseteq \mathcal{P}([n])$, let $\varepsilon \geq 0$, and $F \in \mathcal{F}_V$. For every $i \in [n]$, let the induced subgraph of V_i in G be a clique (with self loops). Additionally, let $c > 0$ be a constant such that for every $S \in V$ and every $i \in S$, the probability that a uniformly random neighbor S' of S in G contains i is at least c. If F passes $\mathcal{T}(G)$ with probability $1 - \varepsilon$ then F is $(1 - \frac{2\varepsilon}{c})$-close to* $\mathsf{dec}(F)$.

Now we show that the above lemma gives the proof of the theorem. Let $\mathcal{A}_i = \{S \in \mathcal{A} \mid i \in S\}$. Note that for every $i \in [n]$, the induced subgraph of $\mathcal{A}_i$ in G is a clique (with self loops) because any two subsets in $\mathcal{A}_i$ have i in their intersection and thus have non-empty intersection. Also for every $S \in \mathcal{A}$ and every $i \in S$, the probability that a uniformly random neighbor S' of S in G contains i is at least $1/2$. Thus, from Lemma 10 the theorem follows. ◀

We complete the proof of the above theorem by showing Lemma 10 below.

[9] Strictly speaking, the addition is done modulo n and then the resulting number is incremented by one.

Proof of Lemma 10. Let $F^* := \mathsf{dec}(F) = \mathsf{DP}_V(a^F)$. Let $B \subseteq V$ be defined as follows:

$$B = \{S \mid F(S) \neq F^*(S)\}.$$

Let $C_i \subseteq V_i$ be defined as follows:

$$C_i = \{S \in V_i \mid F(S)_i = a_i^F\}.$$

By definition of a_i^F, it is clear that $|C_i| \geq |V_i|/2$.

Since F passes $\mathcal{T}(G)$ with probability $1-\varepsilon$ this implies that the number of edges that fail $\mathcal{T}(G)$ is at most $\varepsilon \cdot \frac{|V|d}{2}$.

Fix $S \in B$. Fix $i \in [n]$ (arbitrarily) such that $F(S)_i \neq F^*(S)_i$. Now observe that whenever $S' \in C_i$, the test $T(G)$ rejects the edge (S,S') in G because $F(S)_i \neq a_i^F = F(S')_i$. This implies that there are at least $|C_i| \geq |V_i|/2 \geq {}^{cd}/_2$ many edges incident on S that fail the test $T(G)$. Therefore, there are in total at least $|B| \cdot {}^{cd}/_4$ edges that fail the test. Recall that the total number of rejected edges is at most $\varepsilon \cdot \frac{|V|d}{2}$. Thus we have that $|B|/|V| \leq \frac{2\varepsilon}{c}$. The proof is concluded by noting that the distance between F and F^* is exactly $|B|/|V|$. ◀

Note that Lemma 10 holds even when the induced subgraph of V_i in G is a clique without self loops. In Appendix C, we provide a couple of direct product theorems on domains that are known in literature as an immediate consequence of this lemma.

Lack of Global Expansion. Notice that $G_{\mathcal{A}}$ has very bad edge-expansion/vertex-expansion but is a very good local expander, i.e., the induced subgraph containing any particular coordinate has good expansion (in fact is a clique). Lemma 10 guarantees that and thus $\mathcal{A}$ admits a direct product theorem. Therefore, $\mathcal{A}$ demonstrates that direct product testing does not require the test graph to be an expander (like the Johnson/Ramanujan graph) but only to have certain local expansion properties.

Sub-linear Size Domains. We remark here that we could consider subsets $\tilde{\mathcal{A}}$ of $\mathcal{A}$ of size smaller than n which *still* admit a direct product theorem. For example consider $\tilde{\mathcal{A}}$ as follows:

$$\tilde{\mathcal{A}} = \{\{{}^{ik}/_2, \ldots, {}^{ik}/_2 + k - 1\} \mid i \in [2n/k]\},$$

and the test graph $G_{\tilde{\mathcal{A}}}$ is given by the vertex set $\tilde{\mathcal{A}}$ and the edge set $\{(S,S') \mid S \cap S' \neq \emptyset\}$. It is easy to see that $\tilde{\mathcal{A}}$ admits a direct product theorem by applying Lemma 10. Again, we emphasize that $|\tilde{\mathcal{A}}| = 2n/k$ and yet admits a direct product theorem.

Comparison with Dinur and Kaufman. One might wonder that if direct product testing results can be established on linear sized direct product domains using simple constructions such as the sliding window domain then, why did [6] work so hard and use extremely heavy objects such as high dimensional expanders to obtain linear sized direct product domains. This is because for applications to gap and hardness amplification, it is desirable that a direct product domain also has distance amplification (defined below) and high dimensional expanders have distance amplification whereas the sliding window domain does not.

▶ **Definition 11** (Distance Amplification, [6]). A direct product domain $V \subseteq \binom{[n]}{k}$ is said to have distance amplification if for every $x, y \in \{0,1\}^n$ such that $\delta := \Delta(x,y) < {}^1/_k$, we have that $\Delta(\mathsf{DP}_V(x), \mathsf{DP}_V(y)) = \Omega(k\delta)$.

Thus, the construction of the sliding window domain provides a conceptual clarification as to why we need high dimensional expanders: we can obtain direct product testing from simple constructions like the sliding window domain and we can obtain distance amplification from known constructions of vertex expanders (see Appendix D); but to obtain both simultaneously, [6] needed high dimensional expanders.

References

1 Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. In *IN 29TH STOC*, pages 485–495, 1997.

2 Amey Bhangale, Irit Dinur, and Inbal Livni Navon. Cube vs. Cube Low Degree Test. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pages 40:1–40:31, 2017. doi:10.4230/LIPIcs.ITCS.2017.40.

3 Andries E. Brouwer, Sebastian M. Cioabă, Ferdinand Ihringer, and Matt McGinnis. The smallest eigenvalues of Hamming graphs, Johnson graphs and other distance-regular graphs with classical parameters. *Journal of Combinatorial Theory, Series B*, 2018. doi:10.1016/j.jctb.2018.04.005.

4 Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct Sum Testing. *SIAM J. Comput.*, 46(4):1336–1369, 2017. doi:10.1137/16M1061655.

5 Irit Dinur and Elazar Goldenberg. Locally Testing Direct Product in the Low Error Range. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 613–622, 2008. doi:10.1109/FOCS.2008.26.

6 Irit Dinur and Tali Kaufman. High Dimensional Expanders Imply Agreement Expanders. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 974–985, 2017. doi:10.1109/FOCS.2017.94.

7 Irit Dinur and Inbal Livni Navon. Exponentially Small Soundness for the Direct Product Z-Test. In *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, pages 29:1–29:50, 2017. doi:10.4230/LIPIcs.CCC.2017.29.

8 Irit Dinur and Omer Reingold. Assignment Testers: Towards a Combinatorial Proof of the PCP Theorem. *SIAM J. Comput.*, 36(4):975–1024, December 2006. doi:10.1137/S0097539705446962.

9 Irit Dinur and David Steurer. Direct Product Testing. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 188–196, 2014. doi:10.1109/CCC.2014.27.

10 Oded Goldreich and Shmuel Safra. A Combinatorial Consistency Lemma with Application to Proving the PCP Theorem. *SIAM J. Comput.*, 29(4):1132–1154, 2000. doi:10.1137/S0097539797315744.

11 Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New Direct-Product Testers and 2-Query PCPs. *SIAM J. Comput.*, 41(6):1722–1768, 2012. doi:10.1137/09077299X.

12 Ran Raz and Shmuel Safra. A Sub-constant Error-probability Low-degree Test, and a Sub-constant Error-probability PCP Characterization of NP. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 475–484, New York, NY, USA, 1997. ACM. doi:10.1145/258533.258641.

A Missing Proofs

Proof of Claim 5. We prove only the furthermore part. The first part follows by plugging $c = 1$. For a set $S \subseteq V$ we denote by $\mathbf{1}_S$ the characteristic vector of S. Let $p_\mu \in \mathbb{R}^n$ be the distribution vector that describes μ. First observe that:

$$\Pr_{(u,v)}\left[v \in T \mid u \sim \mu\right] = \frac{1}{d} \cdot (p_\mu)^t \cdot A \cdot \mathbf{1}_T,$$

where the probability is taken over u that is drawn according to μ and v is a uniformly random neighbor of u. Note that $\mathbf{1}_V$ is an eigenvector of A corresponding to the largest eigenvalue (in absolute value) of d. We decompose the vectors: $p_\mu, \mathbf{1}_T$ as follows:

$$p_\mu = \frac{1}{|V|}\mathbf{1}_V + \vec{p} \text{ and } \mathbf{1}_T = \gamma \mathbf{1}_V + \vec{t}.$$

Note that $\gamma = \frac{|T|}{|V|}$ and $\vec{p}, \vec{t}$ are both orthogonal to $\mathbf{1}_V$ and let $\beta = \frac{|S|}{|V|}$. In these notations:

$$\begin{aligned}
\frac{1}{d} \cdot (p_\mu)^t \cdot A \cdot \mathbf{1}_T &= \frac{1}{d}\left(\frac{1}{|V|}\mathbf{1}_V + \vec{p}\right)^t \cdot A \cdot (\gamma \mathbf{1}_V + \vec{t}) \\
&= \gamma + \langle \vec{p}A, \vec{t}\rangle \\
&\leq \gamma + \frac{\lambda}{d}\|\vec{p}\| \cdot \left\|\vec{t}\right\|,
\end{aligned}$$

where in the last step we used the Cauchy-Schwarz inequality and the fact that $\|\vec{p}A\| \leq \lambda \|\vec{p}\|$. Now since the value of each coordinate of p_μ is upper bounded by $\frac{c}{|S|}$ we get: $\|\vec{p}\|^2 = \|p_\mu\|^2 - \frac{1}{|V|^2}\|\mathbf{1}_V\|^2 \leq \frac{c}{|S|} - \frac{1}{|V|}$, and $\left\|\vec{t}\right\|^2 = (\gamma(1-\gamma))\,|V|$. So:

$$\begin{aligned}
\Pr_{(u,v)}\left[v \in T \mid u \sim \mu\right] &= \frac{1}{d} \cdot (p_\mu)^t \cdot A \cdot \mathbf{1}_T \\
&\leq \gamma + \frac{\lambda}{d} \cdot \sqrt{\left(\frac{c}{|S|} - \frac{1}{|V|}\right)\gamma(1-\gamma)\,|V|} \\
&\leq \gamma + \frac{\lambda}{d} \cdot \sqrt{\frac{c\gamma\,|V|}{|S|}} \\
&= \frac{|T|}{|V|} + \frac{\lambda}{d} \cdot \sqrt{\frac{c|T|}{|S|}}
\end{aligned}$$

◀

B Application of Theorem 8: $\Omega(n)$-slice of the Hypercube

In this section, we consider the test graph $J(n, k, k(0.5+\varepsilon))$, where ε is some small constant. The domain of the direct product encoding is $\binom{[n]}{k}$. The pair (S, S') is connected by an edge if and only if: $|S \cap S'| = k/2(0.5+\varepsilon)$. We show that:

▶ **Claim 12.** *Let $\varepsilon = 1/64$. Let $n \in \mathbb{N}$, let $1/2 - \varepsilon \leq c \leq 1/2$ be a constant and let $k = c \cdot n$, then the graph $J(n, k, k \cdot (1/2+\varepsilon))$ is $(1/33, 1/2)$-coordinate expander for large enough n.*

Proof.

1. The proof of the second largest eigenvalue in absolute value was recently confirmed in [3] and we use it below. Note that $\lambda_0 = \binom{k}{k/2+\varepsilon k}\binom{n-k}{k/2-\varepsilon k}$, which is the degree of the graph. Theorem 3.10 in [3] states that λ_1 below is the second largest eigenvalue in absolute value when

$$\lambda_1 = -\binom{1}{0}\binom{k}{k/2+\varepsilon k}\binom{n-k-1}{k/2+\varepsilon k-1} + \binom{1}{1}\binom{k-1}{k/2+\varepsilon k}\binom{n-k}{k/2+\varepsilon k} \leq \lambda_0/33$$

2. Fix $i \in [n]$. Then the graph G_i is isomorphic to $J(n-1, k-1, k/2-1+\varepsilon k)$. Therefore by the first item $\lambda(G_i) < \lambda$. Clearly, for every value of $i \in [n]$ and for each $S \in V_i$ the probability that $i \in S'$ equals $1/2 + \varepsilon$.
3. We verify here the proof for $t > 16$. The case when $t = 4, 8, 12, 16$ can be routinely calculated and verified.

$$\begin{aligned}
\Pr[|T \cap S'| \le t/4] &\le \frac{t}{4} \cdot \frac{\binom{t}{t/4} \cdot \binom{k-t}{k/2-t/4}}{\binom{k}{k/2}} \\
&\le \frac{t}{4} \cdot \frac{\binom{t}{t/4} \cdot \binom{k-t}{k/2-t/2}}{\binom{k}{k/2}} \\
&\le \frac{t}{4} \cdot (1.01)2^{H(1/4)t}(1.01)\frac{2^{k-t}}{\sqrt{k-t}} \cdot \frac{2\sqrt{k/2}}{2^k} \\
&\le \frac{t}{2} \cdot (1.02)\sqrt{\frac{k}{k-t}} \cdot 2^{-.43t} \\
&< 1/33
\end{aligned}$$

Where in third line we used Stirling's approximation that for all $n \ge 16$ to derive: $\frac{2^n}{2\sqrt{n/2}} \le \binom{n}{n/2} \le (1.01)\frac{2^n}{\sqrt{\pi n/2}}$ and $\binom{n}{\epsilon n} \le (1.01)2^{H(\epsilon)n}$. ◀

As a corollary we get that we test the direct product encoding when the domain V equals $\binom{[n]}{k}$ for values of k which are close to $n/2$. Recall that [9] established this result for all $k \le n/2$.

C Simple Applications of Lemma 10

In this subsection, we consider two direct product domains, namely $\binom{[n]}{n/2}$ and $\binom{[n]}{2}$ and prove a direct product theorem for these domains when the test graph is a clique and a member of Johnson graph family respectively.

C.1 $n/2$ slice of the Hamming cube

A natural two-query test on the $n/2$ slice of the Hamming cube is as follows:

Test $\mathcal{T}$:
Input: A function $F : \binom{[n]}{n/2} \to \{0,1\}^{n/2}$.
Procedure: Pick uniformly and independently at random $S, S' \in \binom{[n]}{n/2}$.
Output: Accept if and only if $F(S)|_{S \cap S'} = F(S')|_{S \cap S'}$.

We now interpret the above test in the language established in Section 3. In the above test, the domain V of the direct product is $\binom{[n]}{n/2}$ and the test graph G is a clique with self loops. Therefore, for every $i \in [n]$, the induced subgraph of V_i in G is a clique (with self loops). And, for every $S \in V$ and every $i \in S$, the probability that a uniformly random neighbor S' of S in G contains i is $1/2$. Thus, from Lemma 10 we have that for any $F \in \mathcal{F}_V$, if F passes $\mathcal{T}(G)$ with probability $1-\varepsilon$ then F is $(1-4\varepsilon)$-close to $\mathsf{dec}(F)$.

C.2 $J(n, 2, 1)$ of the Johnson Graph Family

For the domain $\binom{[n]}{2}$, we note that if we pick two elements from $\binom{[n]}{2}$ uniformly and independently at random then they have empty intersection with probability almost 1. Therefore, the same test as for the $n/2$ slice of the Hamming cube does not work here. Nonetheless, there is still a natural two-query test for the domain $\binom{[n]}{2}$ described as follows:

Test $\mathcal{T}$:
Input: A function $F : \binom{[n]}{2} \to \{0,1\}^2$.
Procedure: Pick uniformly at random $S \in \binom{[n]}{2}$. Then pick uniformly at random $S' \in \binom{[n]}{2}$ such that $|S \cap S'| = 1$.
Output: Accept if and only if $F(S)|_{S\cap S'} = F(S')|_{S\cap S'}$.

We now interpret the above test in the language established in Section 3. In the above test, the domain V of the direct product is $\binom{[n]}{2}$ and the test graph G is $J(n,2,1)$. Note that for every $i \in [n]$, the induced subgraph of V_i in G is a clique (without self loops) because any two distinct subsets in V_i have i in their intersection and thus have intersection size equal to 1. Also for every $S \in V$ and every $i \in S$, the probability that a uniformly random neighbor S' of S in G contains i is $1/2$. Thus, from Lemma 10 we have that for any $F \in \mathcal{F}_V$, if F passes $\mathcal{T}(G)$ with probability $1-\varepsilon$ then F is $(1-4\varepsilon)$-close to $\mathsf{dec}(F)$.

D Linear Sized Domains having Distance Amplification

In this section, we show how to construct a collection of sets which have distance amplification. To do so we rely on the existence of vertex expanders.

▶ **Definition 13** (Vertex Expansion). Let $G(V,E)$ be a d-regular graph. For every subset $S \subseteq V$ let $\partial(S) = \{u \in V \setminus S \mid \exists v \in S \text{ such that } (u,v) \in E\}$. The vertex isoperimetric constant $h(G)$ is defined as follows:

$$h(G) = \min_{0 \le |S| \le |V|/d} \frac{|\partial(S)|}{|S| \cdot d}.$$

We say that G is a vertex expander if $h(G)$ is a constant bounded away from 0.

▶ **Theorem 14** (Folklore). *For all $d > 2$, a random d-regular graph is a vertex expander with high probability.*

Given a d-regular graph $G(V,E)$ (where $n := |V|$) which is a vertex expander with vertex isoperimetric constant $\gamma > 0$, we show how to construct $\mathcal{A}_G \subseteq \binom{[n]}{d}$ of cardinality n such that $\mathcal{A}_G$ has distance amplification. We identify the vertices in V with $[n]$ and construct $\mathcal{A}_G$ as follows:

$$\mathcal{A}_G = \{\partial(\{v\}) \mid v \in V\}.$$

▶ **Claim 15.** *$\mathcal{A}_G$ has distance amplification.*

Proof. Fix distinct $x, y \in \{0,1\}^n$. Let $\delta := \Delta(x,y) \le 1/d$. Let $R \subseteq [n]$ be the set of coordinates on which x and y differ. Clearly, $|R| \le n/d$. The number of subsets in $\mathcal{A}_G$ that contain an element in R is at least $\gamma d|R|$. Therefore we have $\Delta(\mathsf{DP}_{\mathcal{A}_G}(x), \mathsf{DP}_{\mathcal{A}_G}(y)) \ge \gamma\delta d$. ◀