Characterizing Direct Product Testing via Coboundary Expansion*

Mitali Bafna
mitalib@mit.edu
Massachusetts Institute of Technology
Cambridge, USA

Dor Minzer
dminzer@mit.edu
Massachusetts Institute of Technology
Cambridge, USA

ABSTRACT

A d-dimensional simplicial complex X is said to support a direct product tester if any locally consistent function defined on its k-faces (where $k \ll d$) necessarily come from a function over its vertices. More precisely, a direct product tester has a distribution μ over pairs of k-faces (A, A'), and given query access to $F\colon X(k) \to \{0,1\}^k$ it samples $(A, A') \sim \mu$ and checks that $F[A]|_{A\cap A'} = F[A']|_{A\cap A'}$. The tester should have (1) the "completeness property", meaning that any assignment F which is a direct product assignment passes the test with probability 1, and (2) the "soundness property", meaning that if F passes the test with probability s, then F must be correlated with a direct product function.

Dinur and Kaufman showed that a sufficiently good spectral expanding complex X admits a direct product tester in the "high soundness" regime where s is close to 1. They asked whether there are high dimensional expanders that support direct product tests in the "low soundness", when s is close to 0.

We give a characterization of high-dimensional expanders that support a direct product tester in the low soundness regime. We show that spectral expansion is insufficient, and the complex must additionally satisfy a variant of coboundary expansion, which we refer to as *Unique-Games coboundary expanders*. Conversely, we show that this property is also sufficient to get direct product testers. This property can be seen as a high-dimensional generalization of the standard notion of coboundary expansion over non-Abelian groups for 2-dimensional complexes. It asserts that any locally consistent Unique-Games instance obtained using the low-level faces of the complex, must admit a good global solution.

CCS CONCEPTS

• **Theory of computation** → **Complexity theory and logic**.

KEYWORDS

High dimensional expansion, Agreement testing, Probabilistically checkable proofs

*The full version of the paper is given in [2].

STOC '24, June 24–28, 2024, Vancouver, BC, Canada

ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649714

ACM Reference Format:
Mitali Bafna and Dor Minzer. 2024. Characterizing Direct Product Testing via Coboundary Expansion. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3618260.3649714

1 INTRODUCTION

The problem of testing direct product functions lies at the intersection of many areas within theoretical computer science, such as error correcting codes, probabilistically checkable proofs (PCPs), hardness amplification and property testing. In its purest form, one wishes to encode a function $f\colon [n] \to \{0,1\}$ using local views in a way that admits local testability/local correction. More precisely, given a parameter $1 \leqslant k < n$, the encoding of f using subsets of size k can be viewed as $F\colon \binom{[n]}{k} \to \{0,1\}^k$ that to each subset $A \subseteq [n]$ of size k assigns a vector of length k describing the restriction of f to A.[1] We refer to this encoding as the direct product encoding of f according to the Johnson graph (for reasons that will become apparent shortly). The obvious downside of this encoding scheme is, of course, that its length is much larger than the description of f (roughly n^k vs $\Theta(n)$). However, as this encoding contains many redundancies, one hopes that it more robustly stores the information in the function f, thereby being more resilient against corruptions.

1.1 Direct Product Testing with 2 Queries

Indeed, one of the primary benefits of the above direct product encoding is that it admits local testers using a few queries. These testing algorithms also go by the name "agreements testers" or "direct product testers", and are often very natural to design. A direct product tester for the above encoding, which we parameterize by a natural number $1 \leqslant t \leqslant k$ and denote by $\mathcal{T}_t$, proceeds as follows:

(1) Choose two subsets $A, A' \subseteq [n]$ uniformly at random conditioned on $|A \cap A'| = t$.
(2) Query $F[A], F[A']$ and check that $F[A]$ and $F[A']$ agree on $A \cap A'$.

These type of testers have been first considered and used by Goldreich and Safra [23] in the context of the PCP theorem. They later have been identified by Dinur and Reingold [18] as a central component in gap amplification. To get some intuition to this test, note that a direct product function clearly passes the test with probability 1. Thus, we say that the tester has perfect completeness. The soundness of the test – namely the probability that a table F which is far from a direct product encoding passes the test – is more difficult to analyze. Intuitively, querying F at a single location gives the

[1]To be more specific, one identifies $A = \{a_1, \ldots, a_k\}$ with the ordered tuple $(a_1, \ldots, a_k)$ where $a_1 < a_2 < \ldots < a_k$ and defines $F[A] = (f(a_1), \ldots, f(a_k))$.

value of a (supposed) f on k inputs. Thus, if F is far from any direct product function, the chance this will be detected should grow with k. Formalizing this intuition is more challenging however, and works in the literature are divided into two regimes: the so-called 99% regime, and the 1% regime. To be more precise, suppose the table F passes the direct product tester $\mathcal{T}_t$ with probability at least $s > 0$; what can be said about its structure?

In the 99% regime, namely the case where $s = 1 - \varepsilon$ is thought of as close to 1, results in the literature [18, 20] show that F has to be close to a direct product function. More specifically, for $t = \Theta(k)$ the result of [20] asserts that there exists $f\colon [n] \to \{0, 1\}$ such that $F[A] = f|_A$ for $1 - O(\varepsilon)$ fraction of the k-sets A. A structural result of this form is a useful building block in several applications. It can be used to construct constant query PCPs with constant soundness; it also serves as a building block in other results within complexity theory; see for instance [8, 13].

The 1% regime, namely the case where $s = \delta$ is thought of as a small constant, is more challenging. In this case, the works [14, 26] show that F has to be correlated with a direct product function. More specifically, these works show that for (say) $t = \sqrt{k}$ if $\delta \geqslant 1/k^{\Omega(1)}$, then there exists $f\colon [n] \to \{0, 1\}$ such that for at least $\delta^{O(1)}$ fraction of the k-sets A, we have that

$$\Delta(F[A], f|_A) \leqslant k^{-\Omega(1)},$$

where for two strings $x, y \in \{0, 1\}^k$, $\Delta(x, y) = \frac{\#\{i \in [k] \mid x_i \neq y_i\}}{k}$ denotes the fractional Hamming distance between them.[2] The motivation for studying this more challenging regime of parameters stems mainly from the perspective of hardness amplification (where one wishes to show that if a given task is somewhat hard, then repeating this task k-times in parallel gets exponentially harder) as well as from the study of PCPs with small soundness. Indeed, in [26] the authors show that direct product testers similar to the above facilitate soundness amplification schemes for PCPs with similar performance to parallel repetition theorems [7, 19, 25, 35, 36]. Direct product testers in the low soundness regime have additional applications in property testing, as well as in the study of the complexity of satisfiable constraint satisfaction problems [3–6].

1.2 Size Efficient 2-Query Direct Product Testing

In the context of PCPs and hardness amplification, one typically thinks of the parameter n as very large, and k as a large constant number. With this in mind, representing an assignment $f\colon [n] \to \{0, 1\}$ using its direct product encoding incurs a polynomial blow-up in size. Indeed, this type of step is often the only step in the PCP reduction that introduces a polynomial (as opposed to just linear) blow-up in the instance size. In this light, a natural question is whether it is possible to perform hardness amplification with a significantly smaller blow up in the encoding/instance size. Efficient schemes of this type are often referred to as "derandomized direct product tests", "derandomized hardness amplification" or "derandomized parallel repetition theorems".

In [26] a more efficient hardness amplification procedure is proposed. Therein, instead of considering all k-sets inside $[n]$, the domain $[n]$ is thought of as a vector space $\mathbb{F}_q^d$ and one considers all subspaces of dimension $\log_q(k)$. It is easy to see that the encoding size then becomes $n^{\Theta(\log k)}$, making it more efficient. The paper [26] shows that direct product testers analogous to the tester above work in this setting as well; they essentially match all of the results achieved by the Johnson scheme. Building upon [26], Dinur and Meir [16] show how to establish parallel repetition theorems using the more efficient direct product encoding via subspaces. This parallel repetition theorem works for structured instances, which the authors show to still capture the entire class NP.

High dimensional expanders (HDX), which have recently surged in popularity, can be seen as sparse models of the Johnson graph. This leads us to the main problem considered in this paper, due to Dinur and Kaufman [15]:

> Do high dimensional expanders facilitate direct product testers in the low soundness regime?

The main goal of this paper is to investigate the type of expansion properties that are necessary and sufficient for direct product testing with low soundness. It is known that there are HDXs of size $O_k(n)$ and $O_k(1)$ degree, and if any of these objects facilitates a direct product tester with small soundness, they would essentially be the ultimate form of derandomized direct product testers.[3] To state our results, we first define the usual notion of spectral high dimensional expansion, followed by our variant of the well-known notion of coboundary expansion.

1.2.1 High Dimensional Local Spectral Expanders. A d-dimensional complex is composed of $X(0) = \{\emptyset\}$, a set of vertices $X(1)$, which is often identified with $[n]$ and a set of i-uniform hyperedges, $X(i) \subseteq \binom{X(1)}{i}$, for each $i = 2, \ldots, d$. A d-dimensional complex $X = (X(0), X(1), \ldots, X(d))$ is called simplicial if it is downwards closed. Namely, if for every $1 \leqslant i \leqslant j \leqslant d$, and every $J \in X(j)$, if $I \subseteq J$ has size i, then $I \in X(i)$. The size of a complex is the total number of hyperedges in X. The degree of a vertex $v \in X(1)$ is the number of faces in $X(d)$ containing it, and the degree of a complex X is the maximum of the degree over all the vertices in $X(1)$.

We need a few basic notions regarding simplicial complex, and we start by presenting the notion of links and spectral expansion.

Definition 1.1. *For a d-dimensional simplicial complex $X = (X(0), X(1), \ldots, X(d))$, $0 \leqslant i \leqslant d - 2$ and $I \in X(i)$, the link of I is the $(d - i)$-dimensional complex X_I whose faces are given as*

$$X_I(j - i) = \{J \setminus I \mid J \in X(j), J \supseteq I\}.$$

For a d-dimensional complex $X = (X(0), X(1), \ldots, X(d))$ and $I \in X$ of size at most $d - 2$, the graph underlying the link of I is the graph whose vertices are $X_I(1)$ and whose vertices are $X_I(2)$.

[2] In contrast to the 99% regime, in this case one has to settle with agreement with f only on a small portion of the k-sets, and furthermore this agreement is not perfect; it is on $(1 - o(1))$ fraction of the elements in the k-sets. As discussed in [14, 26], qualitatively speaking (namely, up to the precise parameters) this is the best type of results possible.

[3] We remark that to be useful, it seems that a derandomized direct product tester would need to roughly have equal degrees. This is because in applications, each one of the values of the encoded function $f\colon [n] \to \{0, 1\}$ is "equally important". In that case, the requirement of being a $O(n)$ sized direct product tester is equivalent to having $O(1)$ degree.

Distributions over the complex. It is convenient to equip a complex X with a measure μ_k for each one of its levels $X(k)$. For $k = d$ we consider the measure μ_d which is uniform over $X(d)$; for each $k < d$, the measure μ_k is the push down measure of μ_d: to generate a sample according to μ_k, sample $D \sim \mu_d$ and then $K \subseteq D$ of size k uniformly. Abusing notation, we will refer to all of the measures μ_k simply as μ, as the cardinality of the sets in discussion will always be clear from context. The set of measures in the link of I is the natural set of measures we get by conditioning μ on containing I.

Equipped with measures over complexes, we may now define the notion of spectral HDX.

Definition 1.2. *A d-dimensional simplicial complex X is called a γ one-sided (two-sided) local spectral expander if for every $I \in X$ of size at most $d-2$, the second eigenvalue (singular value) of the normalized adjacency matrix of the graph underlying the link of I is at most γ.*

In this work, we will only be concerned with simplicial complexes that are very strong spectral expanders. With this regard, following the works of [21, 32, 33] one can show that for every $\gamma > 0$ and every $d \in \mathbb{N}$ there exists an infinite family of d-dimensional complexes of linear size that are γ one-sided or two-sided local expanders (see [15, Lemma 1.5]).

1.2.2 Results in the High Soundness Regime. Dinur and Kaufman [15] were the first to consider the question of direct product testing over HDX. They showed that a sufficiently good high dimensional spectral expander admits a direct product tester in the high soundness regime. The tester they consider is essentially the same as the tester in the Johnson scheme; one thinks of k which is much larger than 1 but much smaller than the dimension of the complex d. The tester has parameters $1 \leqslant s \leqslant k/2$ and is given oracle access to a table $F\colon X(k) \to \{0,1\}^k$, and proceeds as follows:

Agreement-Test 1 (F, k, s)**.**

(1) Sample $D \sim \mu_d$.
(2) Sample $I \subseteq D$ of size s uniformly.
(3) Sample $I \subseteq A, A' \subseteq D$ of size k uniformly.
(4) Accept if $F[A]|_I = F[A']|_I$.

Henceforth, we refer to this test as the (k,s) direct product tester over X. Dinur and Kaufman consider the case where $s = k/2$, and proved that for every $\varepsilon > 0$, provided that γ is sufficiently small, if $F\colon X(k) \to \{0,1\}^k$ passes the above test with probability at least $1-\varepsilon$, then there exists $f\colon X(1) \to \{0,1\}$ such

$$\Pr_{A\sim\mu_k}[F[A] \equiv f|_A] \geqslant 1 - O(\varepsilon).$$

A follow-up work by Dikstein and Dinur [9] further refined this result, and investigated more general structures that support direct product testing in the high soundness regime.

A problem related to direct product testing, called the list agreement testing problem, was considered in the high soundness regime by Gotlib and Kaufman [24]. In the list agreement testing problem, each face is assigned a list of $m = O(1)$ functions, and one performs a local test on these lists to check that they are consistent. With this in mind, the result of Gotlib and Kaufman [24] asserts that under certain structural assumptions on the lists, if the underlying complex has sufficiently good coboundary expansion, then one can design a 3-query list agreement tester that is sound. The list agreement problem will play an important role in the current work, and while we do not know how to use the result of Gotlib and Kaufman for our purposes, their work inspired us to look at connections between agreement testing and notions of coboundary expansion.

1.3 Main Results

Despite considerable interest, no positive nor negative results are known regarding the question of whether HDX support direct product testers in the low-soundness regime. In fact, the majority of applications of HDX are in the high soundness regime, with the first construction of c^3-locally testable codes [12] and quantum LDPC codes [22, 30, 34]. At a first glance, this seems surprising: very good expander graphs give rise to objects in the low-soundness regime, and high dimensional expanders are essentially their higher order analogs.

The main contribution of this work is an explanation to this phenomenon. We show that, to facilitate direct product testers in the low-soundness regime, a high dimensional spectral expander must posses a property that may be seen as a generalization of *coboundary expansion* [31]. On the other hand, we also show that coboundary expansion is sufficient to get direct product testers. Thus, to construct constants degree, sparse complexes facilitating direct product testing, one should first come up with local spectral expanders that are also coboundary expanders.

Below, we state our main results regarding the soundness of the test, which give analysis of the (k,s) tester defined above assuming expansion properties of the complex X. In a concurrent and independent work, Dikstein and Dinur [10] established related results.

1.3.1 Coboundary Expansion. For convenience, we follow the presentation of coboundary expansion from [11]. Suppose we have a function $f\colon X(2) \to \mathbb{F}_2$. The function f is said to be consistent on the triangle $\{u,v,w\} \in X(3)$ if it holds that $f(\{u,v\}) + f(\{v,w\}) + f(\{u,w\}) = 0$. What can we say about the structure of functions f which are consistent with respect to $1-\xi$ measure of the triangles? Clearly, if f is a function of the form $f(\{u,v\}) = g(u) + g(v)$ for some $g\colon X(1) \to \mathbb{F}_2$, then it is consistent with respect to all triangles. In the case that X is a coboundary expander, the converse is also true: any f which is $(1-\xi)$ triangle consistent is $O(\xi)$-close to a function of this form.

More broadly, the notion of coboundary expansion often refers to a property of higher dimensional faces, and to more general groups beyond $\mathbb{F}_2$. We refrain from defining these notions precisely and instead turn to our variant of coboundary expansion, which we show governs the soundness of direct product testing.

1.3.2 Unique-Games Coboundary Expansion. Our notion of coboundary expansion replaces the group $\mathbb{F}_2$ with non-Abelian groups, more specifically with the permutation groups S_m; we also need to consider higher dimensional faces. Some definitions in this spirit have been made, for example in [17, 24], and our notion is inspired by theirs.

Given a d-dimensional complex X and an integer $t \leqslant d/3$, we consider the graph $G_t[X] = (X(t), E_t(X))$ whose vertices are the t-faces of X, namely $X(t)$, and (u,v) is an edge if $u \cup v \in X(2t)$. We say $T = (u, v, w)$ is a triangle in $G_t[X]$ if each of $u, v, w \in X(t)$ and $u \cup v \cup w \in X(3t)$.

Definition 1.3. *Let X be a d-dimensional complex and let t be an integer such that $t \leqslant d/3$. Let $\pi\colon E_t(X) \to S_m$ be a function that satisfies $\pi(u,v) = \pi(v,u)^{-1}$ for all $(u,v) \in E_t[X]$. We say that π is consistent on the triangle (u,v,w) in $G_t[X]$ if $\pi(u,v)\pi(v,w) = \pi(u,w)$.*

We say that π is $(1-\xi)$-consistent on triangles if sampling $T \sim \mu_{3t}$ and then splitting T as a triangle $u \cup v \cup w$ uniformly where $|u| = |v| = |w| = t$,

$$\Pr_{\substack{T\sim\mu_{3t}\\ T=u\cup v\cup w}}\left[\pi(u,v)\pi(v,w) = \pi(u,w)\right] \geqslant 1-\xi.$$

One way to think of this definition is as a locally consistent instance of Unique-Games. Indeed, a π as above specifies a Unique-Games (UG) instance on the graph $G_t[X]$ whose constraints are locally consistent on triangles. The goal in this UG instance may be thought of assigning elements from $[m]$ to the vertices of $G_t[X]$, namely finding an assignment $A\colon X(t) \to [m]$, so as to maximize the fraction of edges (u,v) for which $A(u) = \pi(u,v)A(v)$.

With this definition in mind, we can now present a simplified version of our notion of coboundary expansion. One way to arrive at a locally consistent UG instance as in Definition 1.3 is to first pick some function $g\colon X(t) \to S_m$ and then define $\pi(u,v) = g(u)g(v)^{-1}$. Thus, a natural question is whether there are other ways to construct locally consistent UG instances on $G_t[X]$. In simple terms, our simplified notion of UG coboundary expansion asserts that this is essentially the only way to arrive at instances of this form. More precisely:

Definition 1.4. *We say that a d-dimensional simplicial complex X is an (m,r,ξ,c) UG coboundary expander if for all $t \leqslant r$ and for all functions $f\colon E_t[X] \to S_m$ that are $(1-\xi)$-consistent on triangles, there is $g\colon X(t) \to S_m$ such that*

$$\Pr_{u\cup v\sim\mu_{2t}}\left[\pi(u,v) = g(u)g(v)^{-1}\right] \geqslant 1-c.$$

We remark that if a complex X is an (m,r,ξ,c) UG coboundary expander, then given a $(1-\xi)$-locally consistent instance of Unique-Games on $G_t[X]$ for some $t \leqslant r$, one may find an assignment satisfying at least $1-c$ fraction of the constraints. Indeed, by definition, given the constraint map π we may find $g\colon X(t) \to S_m$ such that $\pi(u,v) = g(u)g(v)^{-1}$ with probability at least $1-c$ over the choice of $u \cup v \sim \mu_{2t}$. Thus, taking the labeling $A(v) = g(v)(1)$, we see that A satisfies all edges on which $\pi(u,v) = g(u)g(v)^{-1}$.

The first result of this paper asserts that a spectral HDX which is a UG coboundary expander admits a direct product tester in the low soundness regime.

Theorem 1.5. *Suppose that a simplicial complex X is a sufficiently good spectral and UG coboundary expander. If $F\colon X(k) \to \{0,1\}^k$ passes the $(k,\sqrt{k})$ direct product test on X with probability δ, then there is $f\colon X(1) \to \{0,1\}$ such that*

$$\Pr_{A\sim\mu_k}\left[\Delta(f|_A, F[A]) = o(1)\right] \geqslant \Omega_\delta(1).$$

In words, being a UG coboundary expander is a sufficient condition for a spectral expander to support a low soundness direct product tester. As far as we know, however, this condition may not be necessary; below, we present a condition which is both necessary and sufficient. Nevertheless, we chose to present its simpler to state version, Definition 1.4, as we find it more appealing, intuitive and resembling non-Abelian variants of the usual notion of coboundary expansion.

Remark 1.6. *The usual definition of coboundary expansion in the literature refers to Abelian groups such as $\mathbb{F}_2$, see for example [11, 24, 27–29]. In the $\mathbb{F}_2$ setting, coboundary expansion for the base graph can be seen as a UG instance over $\mathbb{F}_2$, but it is often phrased in topological notions using the boundary and coboundary maps; these definitions extend well to higher dimensional faces. Coboundary expansion has also been defined for non-Abelian groups [17, 24, 29], however, as far as we know, these definitions coincide with ours only for the case that $t = 1$ in Definition 1.4.*

1.3.3 A Necessary and Sufficient Condition for Low Soundness Direct Product Testing.

We now move on to stating a more complex version of Definition 1.4 which is both necessary and sufficient for low soundness direct product testing. Let us again consider the graph $G_t[X]$ and a $(1-\xi)$ triangle consistent assignment of permutations on the edges $\pi\colon E_t[X] \to S_m$. However, unlike before, these permutations are guaranteed to satisfy an additional premise. Precisely, suppose that each face $u \in X(t)$ is assigned a list of m elements from $\{0,1\}^t$, say $L(u) = (L_1(u), \ldots, L_m(u))$, and each face $T \in X(3t)$ is also assigned a list $L'(T) = (L'_1(T), \ldots, L'_m(T))$. In words, we would like the permutations π to be consistent with the lists with respect to concatenations. Towards this end, we introduce a convenient but informal notation to compare strings. Given $u, v \in X(t)$ that are disjoint and strings $L_i(u), L_i(v) \in \{0,1\}^t$, we shall think of $L_i(u)$ as an assignment to the vertices in u and of $L_i(v)$ as an assignment to the vertices in v. Thus, the notation $L_i(u) \circ L_i(v)$ will be a string in $\{0,1\}^{2t}$ which encodes the assignment to $u \cup v$ provided by the concatenation of the two assignments. More generally, given u, v disjoint and list assignments $L(u), L(v)$ we define

$$L(u) \circ L(v) = (L_1(u) \circ L_1(v), \ldots, L_m(u) \circ L_m(v)).$$

Lastly, given a list $L(u)$ as above and $\pi \in S_m$, we define $\pi L(u) = (L_{\pi(1)}(u), \ldots, L_{\pi(m)}(u))$.

Definition 1.7. *Let $L\colon X(t) \to (\{0,1\}^t)^m$, $L'\colon X(3t) \to (\{0,1\}^{3t})^m$, and $\xi > 0$. We say π is $(1-\xi)$-consistent with the lists L and L' if choosing $T \sim \mu_{3t}$ and a splitting $T = u \cup v \cup w$ into a triangle, we have that*

$$\Pr_{\substack{T\sim\mu_{3t}\\ T=u\cup v\cup w}}\left[L'(T) = L(u) \circ \pi(u,v)L(v) \circ \pi(u,w)L(w)\right] \geqslant 1-\xi.$$

We say that π is $(1-\xi)$-strongly triangle consistent if there are lists L and L' such that π is $(1-\xi)$-consistent with respect to the lists L and L'.

It is easy to see that if π is $(1-\xi)$-strongly triangle consistent, then π is $(1-O(\xi))$-triangle consistent. Thus, the class of triangle consistent functions π is more restrictive. With the notion of strong triangle consistency we are now ready to state a weaker variant of Definition 1.4; the only difference between the two definitions is that in the definition below, we only require that any strongly triangle consistent assignment admits a global structure. More precisely:

Definition 1.8. *We say that a d-dimensional simplicial complex X is a weak (m, r, ξ, c) UG coboundary expander if the following condition is satisfied for all $t \leqslant r$. Suppose $\pi\colon E_t[X] \to S_m$ is a $(1-\xi)$-strongly triangle consistent function. Then there exists $g\colon X(t) \to S_m$ such that*

$$\Pr_{u \cup v \sim \mu_{2t}}[\pi(u,v) = g(u)g(v)^{-1}] \geqslant 1 - c.$$

The parameter r in Definition 1.8 is often referred to as the level at which UG coboundary expansion holds. With the notion of weak UG coboundary expansion, we can now state a stronger version of Theorem 1.5. Roughly speaking, the following two results asserts that for a sufficiently good spectral simplicial complex X, the direct product tester over X works in the low soundness regime if and only if X is a weak UG coboundary expander with sufficiently good parameters.

Theorem 1.9. *The following results hold for any simplicial complex X.*

(1) ***Weak UG-coboundary is Necessary:*** *If a simplicial complex X is a sufficiently good spectral expander which is not a UG coboundary expander, then there is $\delta > 0$ such that for sufficiently large k, there is $F\colon X(k) \to \{0,1\}^k$ that passes the $(k, \sqrt{k})$ direct product tester with probability δ and yet for all $f : X(1) \to \{0,1\}$ we have that*

$$\Pr_{A \sim \mu_k}[\Delta(F[A], f|_A) = o(1)] = o(1).$$

(2) ***Weak UG-coboundary is Sufficient:*** *For all $\varepsilon, \delta > 0$, if a simplicial complex X is a sufficiently good spectral expander and a weak UG coboundary expander on level $O(1)$, then the direct product test over X with respect to sufficiently large k has soundness δ. Namely, if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$ direct product tester with respect to X with probability at least δ, then there is $f : X(1) \to \{0,1\}$ such that*

$$\Pr_{A \sim \mu_k}[\Delta(F[A], f|_A) \leqslant \varepsilon] \geqslant \Omega(1).$$

We refer the reader to the full version for more formal statements and their proofs. We use our necessary result above to conclude that some of the best known sparse spectral expanders – namely some LSV complexes – do not support direct product testers in the low soundness regime precisely because they fail to satisfy coboundary expansion. As the result of Dinur and Kaufman [15] asserts that LSV complexes admit direct product testers in the high soundness regime, we conclude that the low soundness regime is qualitatively different.

In the above theorem, the structure for F we get is relatively weak though, and only asserts that with significant probability over the choice of $A \sim \mu_k$, we have that $F[A]_i = f(i)$ for $(1-\varepsilon)$ fraction of $i \in A$. In the next theorem, we show that if the level r on which coboundary expansion holds is linear in k, then the conclusion of Theorem 1.9 can be strengthened to say that with significant probability over $A \sim \mu_k$, it holds that $F[A]_i = f(i)$ for all but constantly many of $i \in A$.[4]

[4]For that, we need to consider a direct product tester with intersection parameter s, which is significantly smaller than k but is linear in it. Indeed, it is easy to see that the conclusion of Theorem 1.10 would fail if either $s \leqslant k^{0.99}$ or $s \geqslant k/100$.

Theorem 1.10. *If a simplicial complex X is a sufficiently good spectral expander, and for $k \in \mathbb{N}$ it holds that X is a sufficiently good weak UG coboundary expander on level $\Omega(k)$, then the direct product test over X with respect to k has soundness δ. Namely, for all $\delta > 0$ there is $\eta > 0$ such that if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \eta k)$ direct product tester with respect to X with probability at least δ, then there is $f : X(1) \to \{0,1\}$ such that*

$$\Pr_{A \sim \mu_k}\left[\Delta(F[A], f|_A) \leqslant O(1/k)\right] \geqslant \Omega(1).$$

In the full version, we examine several well known complexes. We show that dense complexes such as the complete and the Grassmann complex are UG coboundary expanders. On the flip side we use well-known theorems that some LSV complexes are not coboundary expanders, to show that they fail to support direct product testers.

2 PROOF OF THEOREM 1.9: UG COBOUNDARY IS SUFFICIENT

In this section, we prove the "sufficient" part of Theorem 1.9, formally stated below.

Theorem 2.1. *There is $c > 0$ such that for all $\varepsilon, \delta > 0$ there is $\xi, \eta > 0$ and $m, r \in \mathbb{N}$ such that for sufficiently large k, sufficiently large d and γ small enough function of d, the following holds. If a d-dimensional simplicial complex X is a γ-spectral expander and (m, r, ξ, c) weak UG coboundary expander, then the direct product test over X with respect to sufficiently large k has soundness δ. Namely, if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$ direct product tester with respect to X with probability at least δ, then there is $f : X(1) \to \{0,1\}$ such that*

$$\Pr_{A \sim \mu_k}[\Delta(F[A], f|_A) \leqslant \varepsilon] \geqslant \eta.$$

Remark 2.2. *We remark here that in the above theorem, we require $\delta \geqslant 1/\log k$, $d \geqslant \mathrm{poly}(k)\exp(1/\delta)$, $r = \exp(\mathrm{poly}(1/\delta))$ and $\xi = \mathrm{poly}(\delta)$, which is equal to $1/(\log r)^c$ for some $c \in (0,1)$. In in the full version we improve the latter dependence to show that UG coboundary expansion of $(m, r, \exp(-o(r)), c)$ is sufficient.*

We begin by setting up some notations that will be helpful throughout the proof. Given a global function $f : [d] \to \{0,1\}$ and a set $B \subseteq [d]$ we let $f(B)$ denote the assignment to B using f. For a function $f : [d] \to \{0,1\}$ and an assignment $F : X(k) \to \{0,1\}^k$ we let $\mathrm{Agr}(f, F)$ denote the subset of $X(k)$ where $f(s) = F(s)$ and $\mathrm{agr}(f, F)$ denote the probability of this event under the measure μ_k. Furthermore for $\nu \in (0,1)$ let $\mathrm{Agr}_\nu(f, F)$ denote the subset of $X(k)$ where $f(B)$ and $F(B)$ agree on $(1-\nu)$-fraction of the elements in B and $\mathrm{agr}_\nu(f, F)$ denotes the probability of this event under μ_k.

2.1 High Level Structure of the Proof

The proof of Theorem 2.1 follows the outline given in the introduction. For convenience we break it into two parts, encapsulated in the following two lemmas. In the first lemma we implement the first four steps in the plan and reduce the problem of direct product testing to the problem of "list agreement" testing. In this problem, for each d-face D in a complex X we have a list $L[D]$ of $O(1)$ functions, and we test whether these lists are in 1-to-1 correspondence according to the up-down-up walk on the complex. More precisely, the problem is defined as follows:

List-Agreement-Test 1.
Input: a list $L(D)$ for each $D \in X(d)$ and a parameter $\eta \in (0,1)$.

(1) Choose random $B \sim X(d/2)$.
(2) Choose independently random $A, A' \supseteq_d B$ from $X(d)$.
(3) Accept iff both lists are non-empty and $L[A]|_B \neq_{<\eta} L[A']|_B$.

With the list agreement problem formally defined, we can now state the lemma encapsulating the first few steps in the argument, saying that an assignment that passes the direct product test with probability bounded away from 1 implies a natural list assignment passing the list agreement test with probability close to 1.

Lemma 2.3. *For all $\delta > 0$, for sufficiently large $k \in \mathbb{N}$, $d \geqslant \mathrm{poly}(k)2^{\mathrm{poly}(1/\delta)}$, sufficiently small γ compared to d and $\tau = O(\delta^{68})$, the following holds. Suppose that X is a d-dimensional simplicial complex which is a γ-spectral expander, and $F : X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$-agreement-test 1 with probability δ. Then, there exists $2^{-1/\delta^{1200}} \leqslant \eta' \leqslant \delta^{101}$ and lists $(L[D])_{D \in X(d)}$ satisfying:*

(1) ***Short, non-empty lists:*** *With probability $1 - O(\tau)$ over the choice of $D \sim X(d)$, the list $L[D]$ is non-empty and has size at most $O(1/\delta^{12})$.*
(2) ***Good agreement:*** *For all $D \in X(d)$ and every $f \in L[D]$, we have that $agr_\nu(f, F|_D) \geqslant \Omega(\delta^{12})$ for $\nu = 1/k^{\Omega(1)}$.*
(3) ***Distance in the lists:*** *With probability at least $1 - O(\tau)$ over the choice of $D \sim X(d)$, the list $L[D]$ has distance at least $\delta^{-100}\eta'$.*

Furthermore the lists above pass the List-Agreement-Test 1 with parameter η', with probability $1 - \tau$.

Armed with the conversion of our assignment F to lists that pass the list agreement test with probability close to 1, we implement the next three steps in the introduction. Namely, we show that if X is a sufficiently good UG coboundary expander, then we can use the lists above to define a locally consistent instance of Unique-Games on low levels of the complex and apply UG coboundary expansion to deduce the existence of a global solution.

Lemma 2.4. *Assume there exists a collection of lists $\{L[D]\}_{D \in X(d)}$ that satisfy the premise of Lemma 2.3, and assume that X is a γ-spectral expander for $\gamma < 1/\mathrm{poly}(d)$ and a weak $(O(1/\delta^{12}), t, O(\sqrt{\tau}), c)$ UG coboundary expander for $t = \Theta\left(\frac{\tau\delta^{12}}{\eta'}\right)$. Then there exists $G : X(1) \to \{0,1\}$ such that*

$$\Pr_{D \sim X(d)} [\Delta(G(D), L[D]) \leqslant \delta] \geqslant 1 - O(c^{1/2} + \tau^{1/4} + \gamma).$$

Here, the distance between a function $G(D)$ and a list of functions $L[D]$ is the minimal distance between $G(D)$ and any function in the list.

The proof of Theorem 2.1 now readily follows from the above two lemmas.

Proof of Theorem 2.1. In the setting of Theorem 2.1, first assume that $\varepsilon = \delta$ (otherwise we lower both of them to be the minimum of ε and δ)). Apply Lemma 2.3 and then Lemma 2.4 to conclude that there is a function $G\colon X(1) \to \{0,1\}$ such that

$$\Pr_{D \sim \mu_d} [\Delta(G(D), L[D]) \leqslant \varepsilon] \geqslant \frac{1}{2}.$$

Fix $D \in X(d)$ such that $\Delta(G(D), L[D]) \leqslant \varepsilon$, and let $f \in L[D]$ be such that $\Delta(G(D), f) \leqslant \varepsilon$. Sampling $A \subseteq_k D$, we have by the "good agreement" property of the list that $F[A] \neq_{<\nu} f|_A$ with probability at least $\Omega(\delta^{12})$. By Chernoff's bound we have that $G(D)|_A \neq_{<2\varepsilon} f|_A$ with probability $1 - o(1)$. It follows that with probability at least $\Omega(\delta^{12})$ over $A \subset_k D$, $F[A]$ and $G(A)$ differ on at most $2\varepsilon + \nu \leqslant 3\varepsilon$ fraction of the coordinates of A. Since the fraction of good Ds is $\geqslant 1/2$, $\Delta(F[A], G[A]) \leqslant 3\varepsilon$ on at least $\Omega(\delta^{12})$ fraction of $X(k)$ as required. □

2.2 Auxiliary Claims

Our proof requires a few basic auxiliary probabilistic claims, which we record here. The first claim asserts that if the distance between two functions $f, g\colon [d] \to \{0,1\}$, then choosing a random subset $A \subseteq_k [d]$, we have that the distance between $f|_A$ is also very close to R. More precisely:

Claim 2.5. *Suppose $R \in (0,1)$, and let $f, g\colon [d] \to \{0,1\}$ be functions such that $\Delta(f,g) = R$. Then, for $\frac{1}{R^2} \leqslant k \leqslant d$ we have that:*

(1) $\Pr_{A \subseteq_k [d]} \left[\Delta_A(f,g) > 2R\right] \leqslant 2^{-\Omega(Rk)}$.
(2) $\Pr_{A \subseteq_k [d]} \left[\Delta_A(f,g) < R/2\right] \leqslant 2^{-\Omega(Rk)}$.

Proof. Both of the items are immediate consequences of Chernoff's inequality. The arguments are essentially identical, and we give a proof of the first item only.

To see this, sample $A \subseteq [d]$ by including each element in A with probability k/d. Let $I \subseteq [d]$ be the set of $i \in [d]$ such that $f(i) \neq g(i)$, and for each $i \in I$ define the random variable Z_i to be the indicator of $i \in A$. Define $Z = \sum_{i \in I} Z_i$, and note that $\Delta_A(f,g) = \frac{1}{|A|}Z$. Noting that $\mathbb{E}[Z] = \frac{k|I|}{d} = Rk$, by the Chernoff bound we get that $\Pr\left[Z \geqslant 1.1Rk\right] \leqslant 2^{-\Omega(Rk)}$; also, by another application of the Chernoff bound we get that $|A| \geqslant 0.9k$ with probability $1 - 2^{-\Omega(k)}$. It follows that except with probability $2^{-\Omega(Rk)}$ we have that $\Delta_A(f,g) \leqslant \frac{1.1Rk}{0.9k} \leqslant 2R$. The probability that $|A| = k$ is $\Omega(1/\sqrt{k})$, and conditioned on that A is distributed as $A \subseteq_k [d]$, hence we get that the probability in the first item is at most $O\left(\sqrt{k}2^{-\Omega(Rk)}\right) = 2^{-\Omega(Rk)}$. □

The second claim asserts that if two functions f and g are relatively far, then there are not many k-sets A on which they roughly agree. More precisely:

Claim 2.6. *Suppose that $F\colon X(k) \to \{0,1\}^k$ is an assignment, that $D \in X(d)$ is a face and that $f, g\colon D \to \{0,1\}$ are functions such that $\Delta(f,g) > C\nu$, where $\nu \in (0,1), C \geqslant 6$. Then*

$$\Pr_{A \subseteq_k D}\left[A \in Agr_\nu(f,F) \cap Agr_\nu(g,F)\right] \leqslant 2^{-\Omega(Ck\nu)}.$$

Proof. By Claim 2.5, sampling $A \subseteq_k D$ we get that $\Delta_A(f,g) \geqslant \frac{C\nu}{2}$ with probability $1 - 2^{-\Omega(C\nu k)}$; we claim that such A cannot both be in $\mathrm{Agr}_\nu(f,F)$ and in $\mathrm{Agr}_\nu(g,F)$. Indeed, otherwise we would get that

$$\frac{C\nu}{2} \leqslant \Delta(f|_A, g|_A) \leqslant \Delta(F[A], f|_A) + \Delta(F[A], g|_A) \leqslant 2\nu,$$

and contradiction since $C \geqslant 6$. □

2.3 Proof of Lemma 2.3: Reduction from Agreement to List Agreement Testing

2.3.1 Localizing to a Johnson. The first step of the proof is to localize to a random d-face $D \sim \mu_d$, and show that with probability close to 1, the assignment F passes the direct product test inside T with noticeable probability. More precisely:

Lemma 2.7. *If (k,s)-Agreement-Test 1 on F passes with probability δ, then*

$$\Pr_{D \sim X(d)}\Big[(k,s) - \textit{direct product test passes with probability } \geqslant \delta^2/16 \textit{ inside } D\Big] \geqslant 1 - o(1).$$

Proof. Let $\mathcal{D}_1$ be the distribution on (A, A', I) induced by Agreement Test 1, and consider the following distribution $\mathcal{D}_2$ over (A, A', I):

(1) Sample $B \sim \mu_{\sqrt{d}}$.
(2) Sample $I \subseteq D$ of size s uniformly.
(3) Sample $I \subseteq A, A' \subseteq B$ of size k uniformly.

Note that conditioned on $|A \cap A'| = s$, the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are identical. Thus, as the probability of this event is $1 - O(k^2/\sqrt{d}) = 1 - o(1)$ in both distributions, it follows that the statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is $o(1)$. Therefore,

$$\Pr_{(A,A',I)\sim D_2}\left[F[A]|_I = F[A']|_I\right] \geqslant \delta - o(1).$$

Denote by $\mathcal{D}_2(B)$ the distribution on (A, A', I) conditioned on sampling B, and by p_B the probability that $F[A]|_I = F[A']|_I$ if B was chosen. By an averaging argument, with probability at least $\frac{\delta}{4}$ over the choice $B \sim \mu_{\sqrt{d}}$ we have that $p_B \geqslant \frac{\delta}{2}$; we call such B good, and denote the set of good B's by $\mathcal{B}$.

We get that

$$\Pr_{D\sim\mu_d}\left[\Pr_{B \subseteq_{\sqrt{d}} D}[B \in \mathcal{B}] \geqslant \frac{\delta}{8}\right] \geqslant 1 - O\left(\frac{1}{\sqrt{d}} + \gamma\right) = 1 - o(1).$$

Fix a d-face D satisfying the above event. Thus, picking $B \subset_{\sqrt{d}} D$ and $(A, A', I) \sim \mathcal{D}_2(B)$ passes the direct product test with probability at least $\frac{\delta^2}{8}$. Let this distribution be $\mathcal{D}_2(D)$. As before, letting the distribution $\mathcal{D}_1(A)$ be the distribution over $(A, A', I) \sim D_1$ conditioned on sampling D, the statistical distance between $\mathcal{D}_1(D)$ and $\mathcal{D}_2(D)$ is $o(1)$. Therefore we get that,

$$\Pr_{D\sim\mu_d}\Big[(k,s) - \text{direct product test passes w.p.} \geqslant \delta^2/8 - o(1) \text{ inside } D\Big] \geqslant 1 - o(1)$$

which completes the proof. □

We refer to a d-face $D \in X(d)$ for which the event in Lemma 2.7 holds as good, and thus conclude that $1 - o(1)$ fraction of the d-faces are good. Note that the above argument would also work for $d/2$-faces, and thus we similarly define the notion of good $d/2$-faces.

2.3.2 Getting a List on Each Good Johnson and Generating a Gap. Fix a good d-face D, and consider the assignment F when restricted to k-sets inside k. For notational convenience, we denote this restricted assignment by F_D. Thus, the event in Lemma 2.7 translates to saying that the direct product tester over the Johnson scheme passes inside D with noticeable probability. Thus, using direct product testing results over the Johnson scheme, we may "explain" this consistency via correlations of F_D with true direct product functions. Towards this end, we use a result due to [14] (see also [26], who state a version that is more convenient for our purposes).

Theorem 2.8. *Suppose that F_D passes the $(k, \sqrt{k})$ direct product test in D with probability ε. Then there is a function $g : [d] \to \{0,1\}$ such that*

$$\Pr_{A\subseteq_k D}\left[\Delta(g|_A, F_D[A]) \leqslant 1/k^{\Omega(1)}\right] \geqslant \Omega(\varepsilon^6).$$

Theorem 2.8 by itself is not enough for us, and we need an idea that is often useful in conjunction with such results: list decoding. We wish to consider all direct product functions that are correlated with F_D and have these as the lists. Alas, there is a technical issue: the number of direct product functions that are correlated with F_D need not be bounded in terms of ε, the probability that the test passes. To remedy this issue we require the notion of η-covers, defined below.

Definition 2.9. *Let $\mathcal{F} \subseteq \mathcal{G}$ be two families of functions from $[d]$ to $\{0,1\}$. We say that $\mathcal{F}$ is an η-cover for $\mathcal{G}$ if for any $g \in \mathcal{G}$ there exists $f \in \mathcal{F}$ such that $\Delta(f,g) \leqslant \eta$.*

We are now ready to present a procedure that, given a good d-face D, generates a short list of functions that "explain" most of the probability that F_D passes the direct product test inside D, and which is also short. The procedure takes as input a restriction of the assignment F to a face D, which below we denote by G, and finds one by one direct product functions that are correlated with G, following by randomizing G at appropriate places.

Algorithm 1. The short list algorithm.

Input: $G : \binom{[d]}{k} \to \{0,1\}^k$, $\delta > 0$, $r \in \mathbb{N}$, $\eta \in (0,1)$.
Output: List of functions $\{f_1, \ldots, f_m\}$ from $[d] \to \{0,1\}$.
Operation:

(1) Set $t = k^{-c}$ for $0 < c < 1$, $\delta_0 = \Theta(\delta^6)$, $\widetilde{G}_0 = G$, and initialize $L_1, I_1 = \emptyset$.
(2) For $i \in \{0, \ldots, \lfloor 1/\delta^{80} \rfloor\}$:
   - If there exists f with $\mathrm{agr}_t(f, \widetilde{G}_i) > \delta_i$, add i to I_1 and f_i to L_1.
   - Obtain $\widetilde{G}_{i+1}$ by randomizing $\widetilde{G}_i$ on k-sets $A \in \mathrm{Agr}_t(f, \widetilde{F}_i)$.
   - $\delta_{i+1} = \delta_i - \delta^{100}$.
(3) Create lists I_2, L_2 as follows: for all $i \in I_1$, add i to I_2 and f_i to L_2 iff $i \geqslant r$.
(4) Construct a graph G whose vertices are L_2, and $f, g \in L_2$ are adjacent if $\Delta(f,g) < \eta$. Take a maximal independent set in G and add the corresponding functions to L_3.
(5) Output L_3.

The following lemma summarizes some of the basic properties of the short list algorithm. We will use the parameters and notation specified in the algorithm throughout this section.

LEMMA 2.10. *When ran on $G = F_D$ for a good d-face D with parameter $\Theta(\delta^2)$ in place of δ, setting $\delta' = \Theta(\delta^{12})$, with probability $1 - o(1)$ Algorithm 1 outputs a list $L = \{(i, f_i)\}_{i \in I}$ with $I \subset \{0, \dots, 1/\delta'^{80}\}$ such that,*

(1) $0 \neq |I_1| \leqslant \frac{2}{\delta'}$.
(2) *For all $i \in I_1$, $agr_t(f_i, G) > \delta' - i\delta'^{100} - o(1)$.*
(3) *If $i \notin I_1$ then for all g, $agr_t(g, \widetilde{G}_i) < \delta_i$.*
(4) *For all $i \in \lfloor 1/\delta'^{80} \rfloor$ and $B \subseteq_{d/2} A$, if $g\colon B \to \{0,1\}$ is a function such that $\min_{j \in I_1, j \geqslant i} \Delta(g, f_j|_B) > \Omega(\log(1/\delta')t)$ and $agr_t(g, \widetilde{G}_{i+1}|_B) < \theta$, then $agr_t(g, G|_B) < \theta + \exp(-\Omega(tk \log(1/\delta')))$.*

PROOF. First note that by Theorem 2.8 we get that there is at least one function with $\operatorname{agr}_t(f) \geqslant \delta'$, therefore the list is non-empty. Let us start by proving the upper bound on the size.

Proof of (1): At the i^{th} iteration we add a function to the list only if $\operatorname{agr}_t(f_i, \tilde{G}_i) > \delta_i$ which is always at least $\delta' - \delta'^{20}$. Let $\mathcal{R} \subseteq \binom{D}{k}$ be the k-sets that have been randomized in the algorithm so far, so $|\mathcal{R}| \geqslant (\delta' - \delta'^{20})\binom{d}{k}$. Using the Chernoff bound we get that every function $g\colon D \to \{0,1\}$ satisfies:

$$\Pr\left[\frac{|\operatorname{Agr}_t(g) \cap \mathcal{R}|}{|\mathcal{R}|} > \frac{2\binom{k}{tk}}{2^k}\right] \leqslant \exp\left(-\frac{\binom{k}{tk}}{2^k}\delta\binom{d}{k}\right) \leqslant \exp(-(d/4)^k).$$

Therefore by a union bound we get that with probability $1 - o(1)$, for all functions on D the above holds, and we condition on this event. Hence, the contribution of $\mathcal{R}$ to the agreement of function found in later steps in the procedure is always at most $o(1)$. Thus, each newly found function in the process increases the measure of $\mathcal{R}$ by at least $\delta' - \delta'^{20} - o(1) \geqslant \delta'/2$. Therefore, with probability $1 - o(1)$ the process terminates after at most $2/\delta'$ steps, which is thus also an upper bound on the list size I_1.

Proof of (2): If we inserted f into the list at step i, then $\operatorname{agr}_t(f, \tilde{G}_i) \geqslant \delta' - i\delta'^{100}$. As we have already argued, with probability $1 - o(1)$ at most $o(1)$ of this agreement comes from k-sets in which $\tilde{G}_i$ was randomized, and it follows that $\operatorname{agr}_t(f, G) \geqslant \delta' - i\delta'^{100} - o(1)$.

Proof of (3): If $i \notin I_1$ then the process terminated before step i, meaning that the assignment at that time no longer was δ_i-correlated with any direct product function.

Proof of (4): Denote by $\mathcal{R}_i$ the collection of all k-sets in which the assignment has been randomized in steps prior to the $i+1$th iteration, and consider $\widetilde{G}_{i+1}$. By Claim 2.6 for all $j \geqslant i, j \in I_1$ we get,

$$\Pr_{A \subseteq_k B}\left[A \in \operatorname{Agr}_t(g, G|_B) \cap \operatorname{Agr}_t(f_j|_B, G|_B)\right] \leqslant \exp(-\Omega(tk \log(1/\delta'))), \tag{1}$$

and so

$$\Pr_{A \subseteq_k B}\left[A \in \operatorname{Agr}_t(g, G|_B) \cap \mathcal{R}_i\right] \leqslant 1/\delta' \cdot \exp(-\Omega(tk \log(1/\delta'))) + o(1) \tag{2}$$

$$\leqslant \exp(-\Omega(tk \log(1/\delta'))) \tag{3}$$

It follows from the above that

$$\begin{aligned}\Pr_{A \subseteq_k B}\left[A \in \operatorname{Agr}_t(g, G|_B)\right] &= \Pr_{A \subseteq_k D}\left[A \in \operatorname{Agr}_t(g, G|_B) \cap \mathcal{R}_i\right] \\ &+ \Pr_{A \subseteq_k D}\left[A \in \operatorname{Agr}_t(g, G|_B) \cap \overline{\mathcal{R}_i}\right] \\ &\leqslant \exp(-\Omega(tk \log(1/\delta'))) \\ &+ \Pr_{A \subseteq_k D}\left[A \in \operatorname{Agr}_t(g, \tilde{G}_{i+1}|_B) \cap \overline{\mathcal{R}_i}\right] \\ &\leqslant \exp(-\Omega(tk \log(1/\delta'))) \\ &+ \Pr_{A \subseteq_k D}\left[A \in \operatorname{Agr}_t(g, \tilde{G}_{i+1}|_B)\right],\end{aligned}$$

which is at most $\theta + \exp(-\Omega(tk \log(1/\delta')))$. □

We will now consider the run of the short list algorithm on a d-face with various options for parameters, and its relationship with direct product functions on $d/2$ sub-faces. We will especially care about the relationship between the functions in the list of the d-face $D \in X(d)$, and direct product functions on its $d/2$-faces that have large correlation with the assignment F. In a sense, we will want to show that these are "the same functions"; ultimately, this is where the local consistency of the lists comes from.

Towards this end, we will run the algorithm above for D faces, and denote the outputted list by $L[D]$, For $d/2$ sub-faces of D, we will let $L[B]$ be an η-cover for functions that have sufficient agreement with $F|_B$. The following lemma summarizes the properties of such runs of the short list algorithm:

LEMMA 2.11. *Let $\varepsilon, \delta > 0$, $\eta = 2^{-1/\delta^{1200}}$, let k be sufficiently large and let $d \geqslant \operatorname{poly}(k)\exp(\operatorname{poly}(1/\delta))$. Suppose that F_D passes the $(k, \sqrt{k})$ direct product tester inside D with probability at least δ. Then choosing $r, i \sim \lfloor 1/\delta^{80} \rfloor$ uniformly and running Algorithm 1 with parameters r and $\eta' = \delta^{-100i}\eta$ on D and on all $d/2$ sub-faces, with probability $1 - O(\delta^{68})$ the algorithm outputs a list $L[D]$ such that:*

(1) ***Non-empty, short list:*** *$0 \neq |L[D]| \leqslant 1/\delta'$, where $\delta' = \Theta(\delta^6)$.*
(2) ***Significant correlation:*** *For all $f \in L$, $agr_t(f, F_D) \geqslant \delta_r := \delta' - r\delta'^{100}$, where $t = k^{-\Omega(1)}$.*
(3) ***Large distance in the list:*** *$\Delta(L[D]) > \delta^{-100}\eta'$.*
(4) ***Downwards consistent:***

$$\Pr_{B \subseteq_{d/2} D}[\forall f \in L[D], \exists g \in L[B] \textit{ with } \Delta(f|_B, g) \leqslant \eta'] \geqslant 1 - o(1)$$

In words, for each function in the list of D, projecting it onto a random $B \subseteq_{d/2} D$ yields a function which is very close to a function in the list of B.

(5) ***Upwards consistent:***

$$\Pr_{B \subseteq_{d/2} D}[\forall g \in L[B], \exists f \in L[D] \textit{ with } \Delta(g, f|_B) \leqslant 2\eta'] \geqslant 1 - o(1)$$

In words, choosing a random $B \subseteq D$, every function in the list $L[B]$ is close to a projection of some function from the list $L[D]$.

For each $B \subseteq_{d/2} D$, $L[B]$ is an η'-cover for functions on B with $agr_t(g, F|_B) > \delta_r - \delta^{200}$.

The first four items in Lemma 2.11 are not too hard to establish; the fifth item however requires more care, and this is where we are going to utilize results from random sub-instances of max-k-CSPs. In particular, we require the following theorem from [1]:

Theorem 2.12. *For all $\gamma, \tau \in (0,1)$, $k \in \mathbb{N}$ and $d \geqslant \mathrm{poly}(k/\tau)\exp(1/\gamma^2)$,*[5] *consider a k-CSP with $\binom{n}{k}$ constraints that each depend on a unique k-set of variables. If $q \geqslant \mathrm{poly}(k/\tau\gamma)$ then:*

$$\Pr_{Q \subset_q [d]}\left[|\mathrm{val}(\Psi|_Q) - \mathrm{val}(\Psi)| \leqslant \gamma\right] \geqslant 1 - \tau.$$

Using the above, we get that,

Lemma 2.13. *For all $\zeta \in (0,1)$, $d \geqslant \mathrm{poly}(k)\exp(1/\zeta^2)$, and all functions $G : \binom{[d]}{k} \to \{0,1\}^k$ that satisfy $agr_t(g, G) \leqslant \alpha$ for all $g\colon [d] \to \{0,1\}$, the following holds:*

$$\Pr_{B \subseteq_{d/2}[d]}[\max_g agr_t(g|_B, G|_B) < \alpha + \zeta] \geqslant 1 - \mathrm{poly}(1/d).$$

Proof. Consider the following Max-k-CSP $\Psi = ([d], \mathcal{F})$. The constraints in $\mathcal{F}$ are as follows: for every k-subset I we have the constraint $f_I : \{0,1\}^k \to \{0,1\}$ defined as,

$$f_I(x) = \begin{cases} 1, & \text{if } \Delta(x, G[A]) \leqslant t, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the value of Ψ is $\mathrm{val}(\Psi) = \max_g \mathrm{agr}_t(g, G)$. Applying Theorem 2.12 with $\tau = 1/d^c$ for small enough $c > 0$, we get that with probability $1-\tau$ over the choice of $B \subseteq_{d/2} [d]$, $\mathrm{val}(\Psi|_B) \leqslant \mathrm{val}(\Psi) + \zeta$, which is at most $\alpha+\zeta$. Noting that $\mathrm{val}(\Psi|_B) = \max_g \mathrm{agr}_t(g|_B, G|_B)$ finishes the proof. □

We are now ready to prove Lemma 2.11.

Proof of Lemma 2.11. The proofs of (1) and (2) are immediate from point (1) and (2) of Lemma 2.10.

Proof of (3): Consider the lists produced by the algorithm and consider the pairwise distances $\Delta(f_i, f_j)$ for $f_i, f_j \in L[D]$. Since $|L_2| \leqslant 1/\delta'$ there are at most $1/\delta'^2$ different pairwise distances, therefore with probability $1 - O(\delta^{68})$ over $i \in \{0, \dots, 1/\delta^{80}\}$ we have that for all $i \neq j$ either $\Delta(f_i, f_j) < \eta'$ or $> \delta^{-100}\eta'$. In that case, a maximal independent set L_3 obtained in the foruth step of the short list algorithm satisfies that for all $f_i, f_j \in L_3$, $\Delta(f_i, f_j) > \delta^{-100}\eta'$.

Proof of (4): We get that for each $f \in L[D]$, with probability $1 - o(1)$ over the choice of $B \subseteq_{d/2} D$ we have that $\mathrm{agr}_t(f|_B, F|_B) \geqslant \delta_r - o(1)$. Thus, by the upper bound on the size of $L[D]$ and the union bound we get

$$\Pr_{B \subseteq_{d/2} D}\left[\forall f \in L[D], \mathrm{agr}_t(f|_a) \geqslant \delta_r - o(1)\right] \geqslant 1 - o(1).$$

By the property of η'-covers we conclude that

$$\Pr_{B \subseteq_{d/2} D}[\forall f \in L[D], \exists g \in L[B] \text{ with } \Delta(f|_B, g) \leqslant \eta'] \geqslant 1 - o(1).$$

[5]For general dense k-CSPs they incur a $\exp(2^{2^k})$ dependence in d, which comes from the fact that there can be 2^{2^k} constraints in Ψ that can be satisfied by setting a particular set of variables $I \subset_k [d]$ to a fixed assignment $z \in \{0,1\}^k$. In our setting, there could only be one constraint that gets satisfied by such fixing, and therefore we do not incur this triple-exponential dependence on k (though this wouldn't matter for us in any case).

Proof of (5): Note that the list L_2 has size at most $1/\delta'$, hence with probability at least $1 - O(\delta^{74})$ over the choice of r, we get that $r+1 \notin I_1$. This means that we have a gap: $\forall h, \mathrm{agr}_t(h, \widetilde{G}_{r+1}) < \delta_{r+1}$. Condition on r being chosen so that this holds; by Lemma 2.13 we get that

$$\Pr_{B \subseteq_{d/2} D}\left[\max_h \mathrm{agr}_t(h|_B, \widetilde{G}_{r+1}|_B) < \delta_{r+1} + \delta^{200}\right] \geqslant 1 - o(1).$$

Fix a B where the above holds, let $L[B]$ be an η'-cover as in the statement of the lemma, and take $g \in L[B]$. Assume for contradiction that $\Delta(f|_B, g) > \eta' + \Omega(\log(1/\delta')t)$ for all $f \in L_3$. By the maximality of the independent set L_3, we get that for all $f \in L_2 \setminus L_3$, there exists $f' \in L_3$ such that $\Delta(f, f') < \eta'$. Therefore if g is $\Omega(t\log(1/\delta')) + \eta$-far from all $f \in L_3$, then it is $\Omega(\log(1/\delta')t)$-far from all $f' \in L_2$ and in particular from all $f_j \in L_1$ for $j \geqslant r$, $j \in I_1$. Since $\mathrm{agr}_t(g, \widetilde{G}_{r+1}|_B) < \delta_{r+1} + \delta^{200}$, we may apply the fourth item in Lemma 2.10 to get that $\mathrm{agr}_t(g, G|_B) < \delta_{r+1} + \delta^{200} + \exp(-\Omega(tk\log(1/\delta'))) < \delta_r - \delta^{200}$, for $G = F_D$, which is a contradiction to g being in $L[B]$. □

2.3.3 Consistency of the Local Lists. In this section, we finish the proof of Lemma 2.3. Fix parameters as therein, let $\mathcal{D}$ be the set of good faces (namely, faces in which the $(k, \sqrt{k})$ agreement test passes with probability at least $\delta' = \delta^2/16$), and recall that by Lemma 2.7 we have that $\mu_d(\mathcal{D}) \geqslant 1 - o(1)$.

Let $\eta = 2^{-1/\delta^{1200}}$. We sample r and i integers between 1 and $\lceil 1/\delta^{80} \rceil$ uniformly, set $\eta' = \delta^{-100i}\eta$ and run the short list algorithm on each $D \in \mathcal{D}$ with the parameters r and i. For each D, with probability $1 - O(\delta^{68})$ (over the choice of r, i) we get a list $L[D]$ as in Lemma 2.11. It follows by linearity of expectation and an averaging argument that we may choose r and i such that we get lists $L[D]$ for at least $1 - O(\delta^{68})$ of $D \in \mathcal{D}$ such that $L[D]$ satisfies the conditions of Lemma 2.11, and we fix such r and i henceforth. Below, we refer to a good D that additionally has a list $L[D]$ satisfying the conditions of Lemma 2.11 as very good, and we note that the probability that D is very good is at least $1 - O(\delta^{68}) - o(1) = 1 - O(\delta^{68})$. For each $B \in X(d/2)$, we fix $L[B]$ to be an η' cover of the collection of functions $g\colon B \to \{0,1\}$ such that $\mathrm{agr}_t(g, F|_B) \geqslant \delta_r = \delta' - r\delta'^{100}$.

The first three items in the statement of Lemma 2.3 clearly hold by Lemma 2.11, and in the rest of the argument we argue that the list agreement test passes. Towards this end, consider a generation of queries for the list agreement test. Namely, sample $B \sim \mu_{d/2}$ and independently sample $D, D' \supset_d B$. We say a triple (D, B, D') is good if:

(1) The d-faces D and D' are very good.
(2) It holds that $\Delta(L[D]|_B), \Delta(L[D']|_B) > \frac{1}{2}\delta^{-100}\eta'$.
(3) For all $f \in L[D]$, there exists $g \in L[B]$ with $\Delta(f|_B, g) < \eta'$, and for all $g \in L[B]$ there exists $f \in L[D]$ with $\Delta(g, f|_B) < 2\eta'$. The same holds when D is replaced by D'.

Note that since marginally, each one of D and D' is distributed according to μ_d, we get that the first item holds with probability $1 - O(\delta^{68})$. Note that the marginal distribution of (B, D) is the same as sampling $D \sim \mu_d$, and then $B \subseteq_{d/2} D$. Thus, if the first item holds, then $\Delta(L[D]) \geqslant \delta^{-100}\eta'$, hence by Claim 2.5 we get that the second item holds with probability $1 - o(1)$. Lastly, if the first item holds, then by Lemma 2.11 we get that the third item holds with probability $1-o(1)$. Overall by the union bound, we get that all of the events above holds together with probability at least $1-O(\delta^{68})$.

To finish the proof, we argue that if (D, B, D') is good, then the list agreement test passes on it. For that, we show that for each $f \in L[D]$ there exists a unique $f' \in L[D']$ s.t. $\Delta_B(f, f') \leqslant 3\eta'$ and vice versa. We show the argument only in one of the directions, and the other direction is identical. Take $f \in L[D]$ and consider $f|_B$; by the η'-cover property we can find a $g \in L[B]$ with $\Delta(f|_B, g) \leqslant \eta'$. By the third property above, for g we may find $f' \in L[D']$ with $\Delta(g, f'|_B) \leqslant 2\eta'$, so by the triangle inequality $\Delta_B(f, f') \leqslant 3\eta$. Next, we show the uniqueness of f'. For any $f'' \in L[D'] \setminus \{f'\}$, by the second property above $\Delta(f''|_B, f'|_B) \geqslant \frac{1}{2}\delta^{-100}\eta'$, so

$$\Delta(f''|_B, f|_B) \geqslant \Delta(f''|_B, f'|_B) - \Delta(f'_B, f|_B) \geqslant \frac{1}{2}\delta^{-100}\eta' - 3\eta' \geqslant 100\eta'.$$

2.4 List Agreement Testing Using UG Coboundary Expansion: Proof of Lemma 2.4

The goal of this section is to prove Lemma 2.4. Throughout this section, we fix lists $\{L[D]\}_{D \in X(d)}$ satisfying the premise of Lemma 2.3. We refer to a d-face $D \in X(d)$ for which the properties in Lemma 2.3 are satisfied as good, and note that the measure of the set of good d-faces under μ_d is at least $1 - O(\tau)$. Our first goal is to define a locally consistent instance of Unique-Games on which we can apply coboundary expansion. At the moment though we have assignments only to the d-faces, and our UG coboundary expansion only holds for much lower levels. Thus, we will first show how to project our list assignments to lower levels.

2.4.1 Global Consistency of the List Sizes. We begin with establishing several basic claims that will be useful in the projection process. Their proofs can be found in the full version of the paper. The following claim asserts that almost all of the lists $L[D]$ have the same size. More precisely,

Claim 2.14. *There exists $\ell \leqslant \mathrm{poly}(1/\delta)$ such that* $\Pr_{D\sim\mu_d}\left[|L(D)| \neq \ell\right] \leqslant 10\tau$.

We pick ℓ to be the list size parameter from Claim 2.14. In the next claim we prove that the fact that the list $L[D]$ typically has a large distance implies that its projection onto a sub-face has the same size.

Claim 2.15. *For $t \geqslant 102\frac{\log(\ell/\tau)}{\delta^{-100}\eta'}$ it holds that* $\Pr_{\substack{D\sim\mu_d\\ B\subset_t D}}\left[|L(D)|_B| \neq \ell\right] \leqslant O(\tau)$.

2.4.2 Majority decoding. Next, we show that for t that is not too large, for a typical t-face B, almost all of the d-faces D have the same projection of $L[D]$ onto B. More precisely:

Claim 2.16. *For $t \leqslant \frac{\tau}{\eta'}$, with probability at least $1 - O(\sqrt{\tau})$ over the choice of $B \sim \mu_t$ it holds that*

$$\Pr_{D, D' \supseteq_d B}[L[D]|_B = L[D']|_B] \geqslant 1 - O(\sqrt{\tau}).$$

With Claim 2.16 in hand, one may naturally project the lists that we have on d faces to t-faces in a way that "preserves their essence". More precisely, take a parameter t in the range

$$102\frac{\delta^{100}\log(\ell/\tau)}{\eta'} \leqslant t \leqslant \frac{\tau}{\eta'}. \tag{4}$$

For each $B \in X(t)$ define a list for B using weighted majority

$$L[B] := \mathrm{Maj}_{D\supset_d B}\left(L[D]|_B\right),$$

where the weight of D is $\Pr_{D'\supseteq_d B}\left[D' = D\right]$

Claim 2.17. *For t in the range as in (4), we have that:*

1. $\Pr_{B\sim\mu_t}\left[|L[B]| = \ell\right] \geqslant 1 - O(\sqrt{\tau})$.
2. *Choosing $B \sim \mu_t$, with probability at least $1 - O(\sqrt{\tau})$ it holds that* $\Pr_{D\supseteq_d B}\left[L[D]|_B = L[B]\right] \geqslant 1 - O(\sqrt{\tau})$.

2.4.3 Designing the Unique Games Instance and Proving Triangle Consistency. Fix a t as in (4). Our next goal is to define a Unique-Games instance on the weighted graph G whose vertices are $X(t)$ and whose edge correspond to $2t$-faces: the edges are (u, v) where $u \cup v \in X(2t)$, and its weight is proportional to $\mu_{2t}(u \cup v)$. We remark that strictly speaking, we only define a partial Unique-Games instance on the subset of t-faces B where $|L[B]| = \ell$. By Claim 2.17 these t-faces constitute almost all of $X(t)$, and we encourage the reader to ignore this point.[6]

List ordering, permutations and concatenation. Towards this end, we fix an ordering for each one of the lists constructed thus far (both for d-faces as well as for t-faces). Thus, we will think of the list of $B \in X(t)$ as $L[B] = (L_1[B], \ldots, L_\ell[B])$. For a permutation $\pi \in S_\ell$, we define $\pi(L[B]) = (L_{\pi(1)}[B], \ldots, L_{\pi(\ell)}[B])$. For $u, v \in X(t)$ such that $u \cup v \in X(2t)$ and $\pi \in S_\ell$, we denote

$$L[u] \circ \pi(L[v]) = \left(L_1[u] \circ L_{\pi(1)}[v], \ldots, L_\ell[u] \circ L_{\pi(\ell)}[v]\right),$$

and think of it as a list of assignments to $u \cup v$.

Defining the constraints of the Unique Games instance. Consider the set of $2t$-faces $W \in X(2t)$, and note that one has the analog of Claim 2.17 for these as well, and thus we fix lists $L[W]$ satisfying Claim 2.17 for $2t$-faces as well. Let $\mathcal{W} \subseteq X(2t)$ be the collection of all $2t$-faces for the items in Claim 2.17 hold.

We now define a Unique-Games instance Ψ over G by describing the constraints on the graph G. For each edge (u, v), we put a constraint as follows. If $u \cup v \notin \mathcal{W}$, we put an arbitrary constraint. Else, we put a constraint between u and v if $L[u \cup v]|_u = L[u]$ and $L[u \cup v]|_v = L[v]$. Note that in that case, there is a natural 1-to-1 correspondence between $L[u]$, $L[u \cup v]$ and $L[v]$, and we fix it to be the constraint between $L[u]$ and $L[v]$. Stated otherwise, the constraint on (u, v) is the unique permutation $\pi = \pi(u, v) \in S_\ell$ such that $L[u \cup v] = L[u] \circ \pi(L[v])$ (when both sides are thought of as assignments to $u \cup v$). We think of edges as being directed, and note that then $\pi(u, v) = \pi(v, u)^{-1}$.

The following claim asserts that Ψ is $(1 - O(\sqrt{\tau}))$ strongly triangle consistent.

Claim 2.18. $\Pr_{\substack{Z\sim\mu_{3t}\\ Z=u\cup v\cup w}}$ [*(u, v, w) is strongly consistent in* Ψ] $\geqslant 1 - O(\sqrt{\tau})$.

Proof. We use Claim 2.17 for $3t$-faces, and denote the set of $3t$-faces for which the items there hold by $\mathcal{Z}$. Thus, $\mu_{3t}(\mathcal{Z}) \geqslant 1 - O(\sqrt{\tau})$. Note that sampling $D \sim \mu_d$, then $Z \subseteq_{3t} D$ and then writing $Z = u \cup v \cup w$, with probability $1 - O(\sqrt{\tau})$ we have that there is a 1-to-1 correspondence between the list of each one of u, v, w,

[6] Alternatively, one may think of picking an arbitrary list of size ℓ for every $B \in X(t)$ where $|L[B]| \neq \ell$.

the lists of $u \cup v, v \cup w, u \cup w$, the list of Z and the list of D. We get a 1-to-1 correspondence between the list of u and the list of $v \cup w$, and we denote it by $\pi(u, v \cup w)$, and all of these correspondences are consistent. In particular, we get that $\pi(u,w) = \pi(u,v) \circ \pi(v,w)$ (as both can be thought of as re-alignments of the list of w to concatenate with the list of u so that they agree with $L[u \cup w]$), and hence $\pi(w,u)\pi(u,v)\pi(v,w) = \mathrm{id}$. This proves triangle consistency, and strong triangle consistency readily follows. □

2.4.4 Applying UG Coboundary Expansion. By Claim 2.18 we get that Ψ is $(1 - O(\sqrt{\tau}))$ strongly triangle consistent, and applying the Unique-Games coboundary expansion we get that there is $g\colon X(t) \to S_m$ such that

$$\Pr_{u \cup v \in X(2t)} \left[\pi(u,v) = g(u)g(v)^{-1}\right] \geqslant 1 - c. \tag{5}$$

We now pick an element from the list of each u. More precisely, define $h\colon X(t) \to [\ell]$ defined by $h(v) = g(v)(1)$. Note that if (u,v) is an edge such that the event in (5) holds, then

$$\pi(u,v)(h(v)) = \pi(u,v)g(v)(1) = g(u)(1) = h(u).$$

In other words, for each vertex u we picked an assignment from the list of u in a locally consistent way. We may thus define the assignment $R(u) = L[u]_{h(u)}$; our goal is to show that there is a global function on $X(1)$ that agrees with many of these selections. Towards this end, we first show that R passes the standard direct product test with probability close to 1.

Lemma 2.19. *We have that*

$$\Pr_{\substack{D \sim \mu_d \\ Q \subseteq_{t/2} D \\ Q \subseteq B, B' \subseteq_t D}} \left[R(B)|_Q = R(B')_Q\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2}).$$

Proof. Sample $Z \sim \mu_{3t}$, and write $Z = u \cup v \cup w$, $Z = u \cup v' \cup w'$ independently. Note that by the strong triangle consistency, get that with probability at least $1 - O(\sqrt{\tau} + c)$ we have that

$$\begin{aligned} L[Z] &= L[u] \circ \pi(u,v)L[v] \circ \pi(u,w)L[w] \\ &= L[u] \circ \pi(u,v')L[v'] \circ \pi(u,w')L[w'] \end{aligned}$$

and the edges $(u,v), (u,w), (u,v'), (u,w')$ are satisfied. In that case, we conclude that $R(u) \circ R(v) \circ R(w)$ and $R(u) \circ R(v') \circ R(w')$ correspond to the same function in the list of Z, and so we get that

$$\Pr_{\substack{Z \sim \mu_{3t} \\ Z = u \cup v \cup w = u \cup v' \cup w'}} \left[R(u) \circ R(v) \circ R(w) = R(u) \circ R(v') \circ R(w')\right] \geqslant 1 - O(\sqrt{\tau} + c). \tag{6}$$

For $Z \in X(3t)$, we associate splittings as $Z = u \cup v \cup w$ points in the multi-slice

$$\binom{[3t]}{t,t,t} = \left\{x \in \{0,1,2\}^{3t} \mid \forall j \in \{0,1,2\}, \left|\{i \mid x_i = j\}\right| = t\right\}$$

by identifying u with the set of coordinates equal to 0, v with the set of coordinates equal to 1 and w with the set of coordinates equal to 2. We define $\tilde{R}_Z(x) = R(u) \circ R(v) \circ R(w)$. For each $j \in \{0,1,2\}$, consider the Markov chain T_j on $\binom{[3t]}{t,t,t}$ that from x moves to y where the set of coordinates that are 0 are kept, and the rest are randomized. Then (6) implies that

$$\Pr_{\substack{Z \sim \mu_{3t} \\ x \in \binom{[3t]}{t,t,t}, y \sim \mathrm{T}_0 x}} \left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1 - O(\sqrt{\tau} + c)$$

and analogously we have the same statement for T_1 and T_2, hence by the union bound

$$\Pr_{\substack{Z \sim \mu_{3t} \\ x \in \binom{[3t]}{t,t,t}, y \sim \mathrm{T}_2\mathrm{T}_1\mathrm{T}_0 x}} \left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1 - O(\sqrt{\tau} + c).$$

Therefore, for at least $1 - O(\tau^{1/4} + \sqrt{c})$ of Z, we have that

$$\Pr_{x \in \binom{[3t]}{t,t,t}, y \sim \mathrm{T}_2\mathrm{T}_1\mathrm{T}_0 x} \left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1 - O(\tau^{1/4} + \sqrt{c}),$$

and we call such Z decisive. Fix a decisive Z; the Markov chain $\mathrm{T}_2\mathrm{T}_1\mathrm{T}_0$ has second eigenvalue at most $1 - \Omega(1)$, and thus from the above it follows that

$$\Pr_{x, y \in \binom{[3t]}{t,t,t}} \left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1 - O(\tau^{1/4} + \sqrt{c}),$$

and we define $R(Z)$ to be the most popular value of $\tilde{R}_Z(x)$. Concluding, for decisive Z we get

$$\Pr_{Z = u \cup v \cup w} \left[R(z) = R(u) \circ R(v) \circ R(w)\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2}).$$

Fix a decisive Z, and consider the following direct product tester over Z: choose $Q \subseteq_{t/2} Z$, and then $Q \subseteq B, B' \subseteq_t Z$ such that $B \cap B' = Q$. With probability at least $1 - O(\tau^{1/4} + c^{1/2})$ we get that $R(B)_Q = R(Z)|_Q = R(B')|_Q$. Noting that sampling $Z \sim \mu_{3t}$ and then generating Q, B, B' yields a distribution of (B, Q, B') that is $O(t^2/d) = o(1)$ close to the distribution of Q, B, B' in the direct product tester in the lemma, so the conclusion follows. □

2.4.5 Concluding the Global Structure. With Lemma 2.19 in hand, we apply the direct-product testing theorem from [15] (for the 99% regime) to get that there exists a global function $G : X(1) \to \{0,1\}$ such that

$$\Pr_{B \sim \mu_t} \left[G|_B = R(B)\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma).$$

In the next lemma we analyze the agreement of G with our lists $L[D]$ for $D \in X(d)$, thereby completing the proof of Lemma 2.4.

Claim 2.20. $\Pr_{D \sim \mu_d} \left[\Delta(G|_D, L[D]) \leqslant \frac{100 \log(2\ell)}{t}\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma)$

Proof. Sample $D \sim \mu_d$ and then $B \subseteq_t D$. Then $L[B] = L[D]|_B$ with probability at least $1 - O(\tau)$, and $G|_B \in L[B]$ with probability $1 - O(\tau^{1/4} + c^{1/2} + \gamma)$, hence

$$\Pr_{B \subseteq_t D \sim \mu_d} \left[\Delta(G|_B, L[D]|_B) = 0\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma).$$

We get that with probability at least $1 - O(\tau^{1/4} + c^{1/2} + \gamma)$ over the choice of D, it holds that $\Delta(G|_B, L[D]|_B) = 0$ with probability at least $1/2$ over the choice of B, and we argue that event in question holds for each such D. To see that, first note that fixing $f : D \to$

$\{0,1\}$ such that $\Delta(G|_D, f) \geqslant 100\log(\ell)/t$, it holds that $G|_B = f|_B$ with probability at most

$$\left(1 - \frac{100\log(2\ell)}{t}\right)^t \leqslant (2\ell)^{-100}.$$

Thus, if $\Delta(G|_D, L[D]) \geqslant 100\log(\ell)/t$, then by the union bound $\Delta(G|_B, L[D]|_B) = 0$ with probability at most $(2\ell)^{-99} < 1/2$. □

ACKNOWLEDGEMENTS

We thank Mehtaab Sawhney for suggesting the use of the weak regularity lemma and Nikhil Vyas for bringing [1] to our attention, which led to a considerable simplification on our earlier attempts at proving Lemma 2.13. We thank Tali Kaufman for helpful comments on an earlier version of this paper. We also thank the Simons institute for hosting us during the "Analysis and TCS: New Frontiers" program, where part of the work was carried out.

Mitali Bafna was supported in part by the Computer Science Department, CMU and a gift from CYLAB, CMU. Dor Minzer was supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.

REFERENCES

[1] Noga Alon, Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. 2002. Random sampling and approximation of MAX-CSP problems. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, John H. Reif (Ed.). ACM, 232–239. https://doi.org/10.1145/509907.509945

[2] Mitali Bafna and Dor Minzer. 2023. Characterizing Direct Product Testing via Coboundary Expansion. *CoRR* abs/2308.09668 (2023). https://doi.org/10.48550/ARXIV.2308.09668 arXiv:2308.09668

[3] Amey Bhangale, Subhash Khot, and Dor Minzer. 2022. On approximability of satisfiable k-CSPs: I. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*. 976–988. https://doi.org/10.1145/3519935.3520028

[4] Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. On Approximability of Satisfiable k-CSPs: II. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*. 632–642. https://doi.org/10.1145/3564246.3585120

[5] Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. On Approximability of Satisfiable k-CSPs: III. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*. 643–655. https://doi.org/10.1145/3564246.3585121

[6] Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. On Approximability of Satisfiable k-CSPs: IV. *CoRR* abs/2307.16248 (2023). https://doi.org/10.48550/arXiv.2307.16248 arXiv:2307.16248

[7] Mark Braverman and Ankit Garg. 2015. Small Value Parallel Repetition for General Games. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*. 335–340. https://doi.org/10.1145/2746539.2746565

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

[9] Yotam Dikstein and Irit Dinur. 2019. Agreement Testing Theorems on Layered Set Systems. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*. 1495–1524. https://doi.org/10.1109/FOCS.2019.00088

[10] Yotam Dikstein and Irit Dinur. 2023. Agreement theorems for high dimensional expanders in the small soundness regime: the role of covers. *Electron. Colloquium Comput. Complex.* TR23-119 (2023).

[11] Yotam Dikstein and Irit Dinur. 2023. Coboundary and cosystolic expansion without dependence on dimension or degree. arXiv:2304.01608 [math.CO]

[12] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. 2022. Locally testable codes with constant rate, distance, and locality. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, Stefano Leonardi and Anupam Gupta (Eds.). ACM, 357–374.

[13] Irit Dinur, Yuval Filmus, and Prahladh Harsha. 2019. Analyzing Boolean functions on the biased hypercube via higher-dimensional agreement tests: [Extended abstract]. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*. 2124–2133. https://doi.org/10.1137/1.9781611975482.128

[14] Irit Dinur and Elazar Goldenberg. 2008. 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*. IEEE Computer Society, 613–622. https://doi.org/10.1109/FOCS.2008.26

[15] Irit Dinur and Tali Kaufman. 2017. 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*. IEEE Computer Society, 974–985.

[16] Irit Dinur and Or Meir. 2011. Derandomized Parallel Repetition via Structured PCPs. *Comput. Complex.* 20, 2 (2011), 207–327. https://doi.org/10.1007/s00037-011-0013-5

[17] Irit Dinur and Roy Meshulam. 2019. Near Coverings and Cosystolic Expansion - an example of topological property testing. *Electron. Colloquium Comput. Complex.* TR19-126 (2019). ECCC:TR19-126 https://eccc.weizmann.ac.il/report/2019/126

[18] Irit Dinur and Omer Reingold. 2006. Assignment testers: Towards a combinatorial proof of the PCP theorem. *SIAM J. Comput.* 36, 4 (2006), 975–1024.

[19] Irit Dinur and David Steurer. 2014. Analytical approach to parallel repetition. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*. 624–633. https://doi.org/10.1145/2591796.2591884

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

[21] Shai Evra and Tali Kaufman. 2016. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 36–48.

[22] Shai Evra, Tali Kaufman, and Gilles Zémor. 2020. Decodable quantum LDPC codes beyond the square root distance barrier using high dimensional expanders. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, Sandy Irani (Ed.). IEEE, 218–227.

[23] Oded Goldreich and Shmuel Safra. 2000. A combinatorial consistency lemma with application to proving the PCP theorem. *SIAM J. Comput.* 29, 4 (2000), 1132–1154.

[24] Roy Gotlib and Tali Kaufman. 2023. List Agreement Expansion from Coboundary Expansion. In *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA (LIPIcs, Vol. 251)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 61:1–61:23.

[25] Thomas Holenstein. 2009. Parallel Repetition: Simplification and the No-Signaling Case. *Theory Comput.* 5, 1 (2009), 141–172. https://doi.org/10.4086/toc.2009.v005a008

[26] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. 2009. New direct-product testers and 2-query PCPs. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, Michael Mitzenmacher (Ed.). ACM, 131–140.

[27] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. 2014. Ramanujan Complexes and Bounded Degree Topological Expanders. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*. 484–493. https://doi.org/10.1109/FOCS.2014.58

[28] Tali Kaufman and David Mass. 2022. Double Balanced Sets in High Dimensional Expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference)*. 3:1–3:17. https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2022.3

[29] Tali Kaufman and David Mass. 2022. Unique-Neighbor-Like Expansion and Group-Independent Cosystolic Expansion. *CoRR* abs/2211.09482 (2022). https://doi.org/10.48550/arXiv.2211.09482 arXiv:2211.09482

[30] Anthony Leverrier and Gilles Zémor. 2022. Quantum Tanner codes. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*. IEEE, 872–883.

[31] Nathan Linial and Roy Meshulam. 2006. Homological connectivity of random 2-complexes. *Combinatorica* 26, 4 (2006), 475–487.

[32] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. 2005. Explicit constructions of Ramanujan complexes of type $\tilde{A}_d$. *Eur. J. Comb.* 26, 6 (2005), 965–993.

[33] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. 2005. Ramanujan complexes of type $\tilde{A}_d$. *Israel journal of Mathematics* 149 (2005), 267–299.

[34] Pavel Panteleev and Gleb Kalachev. 2022. Asymptotically good Quantum and locally testable classical LDPC codes. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, Stefano Leonardi and Anupam Gupta (Eds.). ACM, 375–388.

[35] Anup Rao. 2011. Parallel Repetition in Projection Games and a Concentration Bound. *SIAM J. Comput.* 40, 6 (2011), 1871–1891. https://doi.org/10.1137/080734042

[36] Ran Raz. 1998. A Parallel Repetition Theorem. *SIAM J. Comput.* 27, 3 (1998), 763–803. https://doi.org/10.1137/S0097539795280895