On Approximability of Satisfiable k-CSPs: IV

Amey Bhangale
ameyrbh@gmail.com
Department of Computer Science and Engineering, University of California, Riverside.
Riverside, CA, USA

Subhash Khot*
ameyrbh@gmail.com
Department of Computer Science, Courant Institute of Mathematical Sciences, New York University.
New York City, NY, USA

Dor Minzer†
minzer.dor@gmail.com
Massachusetts Institute of Technology
Cambridge, MA, USA

ABSTRACT

We prove a stability result for general 3-wise correlations over distributions satisfying mild connectivity properties. More concretely, we show that if Σ, Γ and Φ are alphabets of constant size, and μ is a distribution over $\Sigma \times \Gamma \times \Phi$ satisfying: (1) the probability of each atom is at least $\Omega(1)$, (2) μ is pairwise connected, and (3) μ has no Abelian embeddings into $(\mathbb{Z}, +)$, then the following holds. Any triplets of 1-bounded functions $f : \Sigma^n \to \mathbb{C}$, $g : \Gamma^n \to \mathbb{C}$, $h : \Phi^n \to \mathbb{C}$ satisfying

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geqslant \varepsilon$$

must arise from an Abelian group associated with the distribution μ. More specifically, we show that there is an Abelian group $(H, +)$ of constant size such that for any such f, g and h, the function f (and similarly g and h) is correlated with a function of the form $\tilde{f}(x) = \chi(\sigma(x_1), \ldots, \sigma(x_n))L(x)$, where $\sigma : \Sigma \to H$ is some map, $\chi \in \hat{H}^{\otimes n}$ is a character, and $L : \Sigma^n \to \mathbb{C}$ is a low-degree function with bounded 2-norm.

En route we prove a few additional results that may be of independent interest, such as an improved direct product theorem, as well as a result we refer to as a "restriction inverse theorem" about the structure of functions that, under random restrictions, with noticeable probability have significant correlation with a product function.

In companion papers, we show applications of our results to the fields of Probabilistically Checkable Proofs, as well as various areas in discrete mathematics such as extremal combinatorics and additive combinatorics.

CCS CONCEPTS

• **Theory of computation → Problems, reductions and completeness**.

KEYWORDS

PCP, Analysis of Boolean Functions, Abelian Embeddings

ACM Reference Format:
Amey Bhangale, Subhash Khot, and Dor Minzer. 2024. On Approximability of Satisfiable k-CSPs: IV. 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/10.1145/3618260.3649610

1 INTRODUCTION

1.1 Studying 3-wise Correlations with Respect to a Distribution

Let Σ, Γ and Φ be alphabets of constant size, suppose μ is a distribution over $\Sigma \times \Gamma \times \Phi$, and let $f : \Sigma^n \to \mathbb{C}$, $g : \Gamma^n \to \mathbb{C}$, $h : \Phi^n \to \mathbb{C}$ be 1-bounded functions. What sort of triplets of functions f, g and h have a significant 3-wise correlation with respect to μ? In other words, what can be said about the functions f, g and h in the case that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geqslant \varepsilon, \quad (1)$$

where $\varepsilon > 0$ is thought of as a small constant? In [23], it is shown that if μ is connected, then this can only be the case if each one of f, g and h is correlated with a low-degree function. Here, we say that a distribution μ over $\Sigma_1 \times \Sigma_2 \times \Sigma_3$ is connected if for any partition of $\{1, 2, 3\}$ into two sets $I \cup J$, the bi-partite graph between $\text{supp}(\mu_I)$ and $\text{supp}(\mu_J)$ whose edges are all (a, b) if $(a, b) \in \text{supp}(\mu)$, is connected ($\mu_I$ is the marginal distribution of μ on the coordinates of I). In [2, 4], a strengthening of this result is proved, and it is shown that it suffices that the distribution μ does not admit any *Abelian embeddings*.

Definition 1.1. *An Abelian embedding of a distribution μ over $\Sigma \times \Gamma \times \Phi$ consists of an Abelian group $(H, +)$ and 3 maps $\sigma : \Sigma \to H$, $\gamma : \Gamma \to H$ and $\phi : \Phi \to H$ such that $\sigma(x) + \gamma(y) + \phi(z) = 0$ for all $(x, y, z) \in \text{supp}(\mu)$. We say that the embedding (σ, γ, ϕ) is non-trivial if at least one of the maps is not constant.*

Definition 1.2. *We say a distribution μ admits an Abelian embedding if it has a non-trivial Abelian embedding.*

In this language, the main result of [2, 4] asserts that if μ does not admit an Abelian embedding and the probability of each atom in μ is at least α thought of as a constant, then each one of f, g and h must be correlated with a low-degree function. As it can easily be seen, this result is strictly stronger than the corresponding result in [23] since any distribution μ which is connected does not admit an Abelian embedding. Moreover, as explained in [2, 4] this result is an if and only if, in the sense that in the presence of Abelian

*Supported by the NSF Award CCF-1422159, NSF CCF award 2130816, and the Simons Investigator Award.
†Supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.

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

ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649610

embedding one could design 1-bounded functions f, g and h for which (1) holds while at least one of the functions f, g and h only has $o(1)$-correlation with any low-degree function.

The main goal of this paper is to extend this understanding beyond the realm of distributions which do not have Abelian embeddings and prove structural results on functions f, g and h satisfying (1) in this more general setting. At a high level, one would like to say that such functions f, g and h could only arise as a result of using Abelian embeddings, using low-degree functions, or both. To prove such result however, we must focus our attention on distributions μ satisfying a very mild form of connectivity, which we refer to as pairwise connectedness.

Definition 1.3. *Let $\Sigma_1, \Sigma_2, \Sigma_3$ be finite alphabets, and let $P \subseteq \Sigma_1 \times \Sigma_2 \times \Sigma_3$. For a pair of distinct coordinates $i, j \in \{1, 2, 3\}$, we say P is $\{i, j\}$ connected if the bipartite graph $G = (\Sigma_i \cup \Sigma_j, E_{i,j})$, where $E_{i,j}$ is the set of label pairs that appear in some element of P, is connected.*

We say P is pairwise connected if it is pairwise connected for any two distinct $i, j \in \{1, 2, 3\}$.

We say a distribution μ is pairwise connected if $\mathrm{supp}(\mu)$ is pairwise connected.

At a high level, the notion of pairwise connectedness stems from the fact that if $\mathrm{supp}(\mu)$ is not pairwise connected, then there are examples of functions satisfying (1) without any useful structure for our purposes. Indeed, if $\mathrm{supp}(\mu)$ is not pairwise connected – without loss of generality it is not $\{1, 2\}$-connected, then we may find a non-trivial partition $\Sigma = \Sigma' \cup \Sigma''$ and $\Gamma = \Gamma' \cup \Gamma''$ so that in the support of μ there can only be pairs from $\Sigma' \times \Gamma'$ and $\Sigma'' \times \Gamma''$ on the first two coordinates. In this case, we may pick any pair of functions $s, s' : \{1, 2\}^n \to \mathbb{C}$ such that $s(a)s'(a) = 1$ for all $a \in \{1, 2\}^n$ (for example, one can take s whose absolute value is always 1, and s' to be its conjugate) and construct f, g, h as follows. For f, we set $f(x) = s(x')$ where for each i, $x'_i = 1$ if $x_i \in \Sigma'$ and otherwise $x'_i = 2$. For g, we similarly set $g(y) = s'(y')$ where for each i, $y'_i = 1$ if $y_i \in \Gamma'$ and otherwise $y'_i = 2$. For h, we take $h \equiv 1$. Thus, for any $(x, y, z) \in \mathrm{supp}(\mu)$ we have that

$$f(x)g(y)h(z) = s(x')s'(y') = 1,$$

as we have that $x' = y'$ by construction.

Henceforth, we will focus our attention on distributions μ which are pairwise connected. With this in mind, as explained earlier there are two ways of constructing functions f, g and h satisfying (1):

(1) If $\mathrm{supp}(\mu)$ admits a linear embedding, say for simplicity a cyclic group $(H, +) = (\mathbb{Z}_p, +)$ and maps $\sigma : \Sigma \to H, \gamma : \Gamma \to H$ and $\phi : \Phi \to H$ not all constant such that $\sigma(x) + \gamma(y) + \phi(z) = 0$, then one can take

$$f(x_1, \ldots, x_n) = e^{\frac{2\pi i}{|H|}(\sigma(x_1)+\ldots+\sigma(x_n))},$$
$$g(y_1, \ldots, y_n) = e^{\frac{2\pi i}{|H|}(\gamma(y_1)+\ldots+\gamma(y_n))},$$
$$h(z_1, \ldots, z_n) = e^{\frac{2\pi i}{|H|}(\phi(z_1)+\ldots+\phi(z_n))},$$

and note that $f(x)g(y)h(z) = 1$ pointwise, hence (1) holds. More generally, for a general Abelian group $(H, +)$ one can pick non-trivial characters $\chi_1, \ldots, \chi_n \in \hat{H}$, define

$$f(x_1, \ldots, x_n) = \prod_{j=1}^{n} \chi_j(\sigma(x_j)),$$
$$g(y_1, \ldots, y_n) = \prod_{j=1}^{n} \chi_j(\gamma(y_j)),$$
$$h(z_1, \ldots, z_n) = \prod_{j=1}^{n} \chi_j(\phi(z_j)),$$

and note again that $f(x)g(y)h(z) = 1$ pointwise hence (1) holds.

(2) In general, it may also be the case that for a distribution μ, low-degree functions also satisfy (1). Indeed, in that case one may try to find univariate 1-bounded functions $u : \Sigma \to \mathbb{C}$, $v : \Gamma \to \mathbb{C}$ and $w : \Phi \to \mathbb{C}$ for which

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} [u(x)v(y)w(z)] \right| \geqslant \Omega(1),$$

and then tensorize them to get

$$f(x_1, \ldots, x_n) = \prod_{j=1}^{d} u(x_j),\ g(y_1, \ldots, y_n) = \prod_{j=1}^{d} v(y_j),$$
$$h(z_1, \ldots, z_n) = \prod_{j=1}^{d} w(z_j),$$

which get value of $2^{-O(d)}$ in (1).

1.2 Main Results

With the above discussion in mind, one is tempted to conjecture that if μ is pairwise connected, then the only possible examples of triplets of functions f, g and h satisfying (1) must come from the above template.

1.2.1 The Stability Result. The main result of this paper is a stability result that formalizes this intuition, saying that under some mild assumptions on the distribution μ, if f, g and h are 1-bounded functions achieving significant 3-wise correlation as in (1), then f (and similarly g and h) must be correlated with a product of an embedding type function as in the first recipe, with a low-degree function as in the second recipe. The mild assumptions on μ correspond to it being pairwise connected (which is necessary, otherwise the statement is simply false), and for technical reasons we also need an additional assumption, namely that μ cannot be embedded in the Abelian group $(\mathbb{Z}, +)$. We remark though that this additional assumption is, as far as we know, not necessary, but removing it seems to require more ideas. With this in mind, a precise formulation of our main result is:

Theorem 1.4. *For all $m \in \mathbb{N}$, $\alpha > 0$ and $\varepsilon > 0$, there are $d \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that μ is a distribution over $\Sigma \times \Gamma \times \Phi$ such that*

(1) The probability of each atom in μ is at least α.
(2) The size of each one of Σ, Γ, Φ is at most m.
(3) The distribution μ is pairwise connected.
(4) μ does not admit an Abelian embedding into $(\mathbb{Z}, +)$.

Then, if $f: \Sigma^n \to \mathbb{C}$, $g: \Gamma^n \to \mathbb{C}$ and $h: \Phi^n \to \mathbb{C}$ are 1-bounded functions such that

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}[f(x)g(y)h(z)]\right| \geqslant \varepsilon,$$

then there are 1-bounded functions $u_1, \ldots, u_n: \Sigma \to \mathbb{C}$ and a function $L: \Sigma^n \to \mathbb{C}$ of degree at most d and 2-norm at most 1 such that

$$\left|\mathop{\mathbb{E}}_{x\sim\mu_x^{\otimes n}}\left[f(x)\cdot L(x)\prod_{i=1}^{n}u_i(x_i)\right]\right| \geqslant \delta.$$

Furthermore, there is $r \in \mathbb{N}$ depending only on m and an Abelian embedding (σ, γ, ϕ) of μ into an Abelian group $(H, +)$ of size at most r such that for all i, $u_i(x_i) = \chi_i(\sigma(x_i))$ where $\chi_i \in \widehat{H}$ is a character of H.

Quantitatively, we have that

$$d = \mathrm{poly}_{\alpha,m}(1/\varepsilon), \qquad \delta = 2^{-\mathrm{poly}_{\alpha,m}(1/\varepsilon)}.$$

The proof of Theorem 1.4 is quite long, and in Section 1.4 we give an overview of the steps we take in the proof. Some of the steps require ingredients that may be of independent interest, and which we explain next.

1.2.2 The Restriction Inverse Theorem. The proof of Theorem 1.4 uses a result which we refer to as the Restriction Inverse Theorem and present next.

Restrictions and Random Restrictions. Restrictions and random restrictions are vital to our argument to go through, and the Restriction Inverse Theorem can be thought of as a statement about them of independent interest. Given a function $f: (\Sigma^n, \mu^{\otimes n}) \to \mathbb{C}$, a set of coordinates $I \subseteq [n]$ and $\tilde{x} \in \Sigma^{\overline{I}}$, the restricted function $f_{\overline{I}\to\tilde{x}}$ is a function from Σ^I to $\mathbb{C}$ defined as

$$f_{\overline{I}\to\tilde{x}}(x') = f(x_I = x', x_{\overline{I}} = \tilde{x}),$$

where $(x_I = x', x_{\overline{I}} = \tilde{x})$ is the point in Σ^n whose I-coordinates are set according to x', and whose $\overline{I}$-coordinates are set according to $\tilde{x}$.

Random restrictions are restrictions in which either I, $\tilde{x}$ or both are chosen randomly. A typical setting we use is one where we have a parameter $\rho > 0$, and we pick $I \subseteq_\rho [n]$, by which we mean that we include each $i \in [n]$ in I with probability ρ; we then choose $\tilde{x} \sim \mu^{\overline{I}}$. For the purposes of this paper it is necessary to consider other (less standard) settings of random restrictions, but we will limit ourselves to this more typical setting for the purposes of this introduction.

Product functions. A function $f: \Sigma^n \to \mathbb{C}$ is called a product function if there are 1-bounded functions $f_1, \ldots, f_n: \Sigma \to \mathbb{C}$ such that

$$f(x_1 \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

It is clear that if f is a product function, then any restriction of it is still a product function. Thus, with probability 1, taking a random restriction of f yields a function which has perfect correlation with a product function. The Restriction Inverse Theorem is a statement about a deduction in the reverse direction: suppose f is a function that after random restriction it has a significant correlation with a product function. Is it necessarily the case that f itself is correlated with a product function?

As is usually the case with inverse-type questions, there are multiple regimes of parameters one may consider, and for us the most relevant regime is the so-called 1% regime. In this case, we have a parameter $\rho > 0$ (which is small but bounded away from 0) and a function $f: \Sigma^n \to \mathbb{C}$ such that

$$\Pr_{\substack{I\subseteq_\rho[n]\\ \tilde{x}\sim\mu^{\overline{I}}}}\left[\exists \text{ product function } p: \Sigma^I \to \mathbb{C}, \left|\langle f_{\overline{I}\to\tilde{x}}, p\rangle\right| \geqslant \varepsilon\right] \geqslant \varepsilon, \quad (2)$$

and we wish to deduce a structural result about f. As discussed, such situations may arise when f is a product function – or more generally when it is correlated with a product function. However, if f is a low-degree function (or even if it is just correlated with a low-degree function), a random restriction of f will be correlated with a constant function with noticeable probability, and hence with a product function. The Restriction Inverse Theorem essentially says that these are the only two ways that (2) can come about:

Theorem 1.5 (The Restriction Inverse Theorem, Informal). *For all $\varepsilon, \rho, \alpha > 0$ and $m \in \mathbb{N}$ there are $d \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose Σ is a finite alphabet of size at most m, μ is a distribution over Σ in which the probability of each atom is at least α, and $f: (\Sigma^n, \mu^{\otimes n}) \to \mathbb{C}$ is a 1-bounded function satisfying (2). Then there is a product function $p: \Sigma^n \to \mathbb{C}$ and a function $L: \Sigma^n \to \mathbb{C}$ of degree at most d and $\|L\|_2 \leqslant 1$ such that*

$$|\langle f, pL\rangle| \geqslant \delta.$$

We refer the reader to the full version of the paper for a more formal and general version of the Restriction Inverse Theorem. We remark that among other things, we also give explicit dependency of d and δ on ε and ρ. These quantitative aspects are important if one wishes to get decent quantitative bounds in Theorem 1.4, and we think they are also interesting in their own right.

1.2.3 The Direct Product Theorem. The proof of Theorem 1.5 (and thus, in turn, of Theorem 1.4) hinges on a direct product testing result, which may also be of independent interest. The problem of direct product testing has its roots in the field of probabilistic checkable proofs and in particular in hardness amplification. In this setting, one wishes to encode a function $f: [n] \to [R]$ (where n is thought of as very large) by local pieces that, on the one hand allows for local access to values of f. On the other hand, the encoding should be testable, in the sense that there is a test that only looks at a handful of locations of the encoding and determines whether it is an encoding of an actual function $f: [n] \to [R]$, or whether it is far from the encoding of any such function.

Our application calls for a particular direct product tester that has been extensively studied in the literature [5, 8, 9, 11, 13, 21, 22]. In this tester, the function f is encoded via its table of restrictions to sub-cubes of certain dimension. Namely, given a parameter $k \in \mathbb{N}$ (which for us will be equal to ρn, where ρ should be thought of as a very small constant), the direct product encoding of f is the mapping $F: \binom{[n]}{k} \to [R]^k$ defined by

$$F[A] = f|_A$$

for all $A \subseteq [n]$ of size k.

The test we associate with this encoding is determined by two parameters, $\alpha, \beta \in (0,1)$ that also should be thought of as small constants. Given a supposed table of restrictions $G\colon \binom{[n]}{k} \to [R]^k$, the test, which we call $\mathrm{DP}(\rho,\alpha,\beta)$, proceeds in the following way:

(1) Sample $C \subseteq [n]$ of size αk and sample $A, B \in \binom{[n]}{k}$ independently containing C.
(2) Sample $T \subseteq [n]$ of size βn.
(3) Query $G[A]$, $G[B]$ and check that $G[A]|_{C\cap T} = G[B]|_{C\cap T}$.

In other words, the tester selected two sets A, B that intersect on a sizable number of elements (at least αk), then a random subset of their shared elements and checks that the local assignments $G[A]$ and $G[B]$ agree on this random subset of shared elements.

Note that this test is complete, in the sense that if G is a legitimate direct product encoding, then it passes the test with probability 1. Thus, as is usually the case, the interesting aspect of this test is the soundness, which is equivalent to the following question. Suppose that the tester accepts a table $G\colon \binom{[n]}{k} \to [R]^k$ with probability at least s; is it necessarily the case that G is somewhat close to a legitimate direct product testing codewords?

In the so-called 99% regime, where the probability $s = 1-\varepsilon$ is thought of close to 1, this problem is completely understood, and in [8, 13] it is shown that in this case there is a function $f\colon [n] \to [R]$ such that for at least $1-O(\varepsilon)$ fraction of $A \in \binom{[n]}{k}$ it holds that $G[A] = f|_A$.

For us, the most so-called 1% regime is more relevant, wherein the probability $s = \varepsilon$ is thought of as close to 0. In this case, one can no longer expect a strong conclusion as in the 99% regime. Instead, naturally one would expect that in this case, there would have to be a function $f\colon [n] \to [R]$ such that for at least $\delta = \delta(\rho,\alpha,\beta,\varepsilon) > 0$ fraction of $A \in \binom{[n]}{k}$ it holds that $G[A] = f|_A$, but this is also too much to expect. Indeed, to see that take any $g\colon [n] \to [R]$, and for each A take $G[A]$ uniformly from $[R]^k$ with probability $1-\varepsilon$, and otherwise take it to be a string in $[R]^k$ of Hamming distance $r = \Theta(\log(1/\varepsilon))$ from $g|_A$. Using Chernoff's bound, one can prove that with high probability there is no function $f\colon [n] \to [R]$ satisfying the natural conclusion one expects, yet the tester passes with probability at least $\varepsilon^2(1-\beta)^{2r} = \mathrm{poly}(\varepsilon)$. The reason for that is that looking at two locations A, B queried by the tester, with probability ε^2 both of them get assigned strings close to $g|_A$ and $g|_B$ respectively, in which case with probability at least $(1-\beta)^{2r}$ the subset T excludes all coordinates on which either $G[A]$ and $g|_A$, or $G[B]$ and $g|_B$, disagree on.

Due to a rather versatile set of examples, results in the 1% regime are often more challenging to prove. Indeed, earlier results by [9, 13] managed to show that in this case there is a function $f\colon [n] \to [R]$ such that for at least $\delta = \delta(\rho,\alpha,\beta,\varepsilon,\eta) > 0$ fraction of $A \in \binom{[n]}{k}$ it holds that $\Delta(G[A], f|_A) \leqslant \eta k$. Here and throughout, $\Delta(x,y)$ represents the Hamming distance between strings x and y. The main drawback of this result is that the distance between $G[A]$ and $f|_A$ is linear in k, which is not good enough for our purposes. Indeed, for our application we need a result that gets a Hamming distance which is a constant $r = r(\rho,\alpha,\beta,\varepsilon) \in \mathbb{N}$ as opposed to a constant fraction.

In [5], such result was proved for a more specialized version of this test in the case of $\beta = 1$ and $R = 2$. Therein, both of the parameters α and ρ are thought of as constant, and it is proved that there are $r = r(\varepsilon,\alpha,\rho) \in \mathbb{N}$ and $\delta = \delta(\varepsilon,\alpha,\rho) > 0$ such that if G passes the test $\mathrm{DP}(\rho,\alpha,\beta=1)$ with probability at least ε, then there is a function $f\colon [n] \to \{0,1\}$ such that for at least δ fraction of $A \in \binom{[n]}{k}$ it holds that $\Delta(G[A], f|_A) \leqslant r$. Besides being a natural question of interest, the motivation of this result therein was to establish an earlier, less general version of the Restriction Inverse Theorem, Theorem 1.5 herein.

In this paper, we are once again in a situation that our proof of a restriction inverse theorem requires a direct product testing result, and the relevant test for us is the test $\mathrm{DP}(\rho,\alpha,\beta)$ above. Moreover, as herein we are concerned with getting good quantitative bounds, we no longer think of the parameters ρ,α,β as constants and thus try to get reasonable dependencies of r and δ on these parameters. For the purposes of this introductory section however, we do not mention these quantitative aspects and defer the interested reader to the full version of the paper. Thus, without a concern for these quantitative aspects our result reads:

Theorem 1.6 (The Direct Product Testing Theorem, Informal). *For all $\varepsilon,\rho,\alpha,\beta > 0$ there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds for all $R \in \mathbb{N}$. For $k = \rho n$, if $G\colon \binom{[n]}{k} \to [R]^k$ is a function that passes the test $\mathrm{DP}(\rho,\alpha,\beta)$ with probability at least ε, then there is a function $f\colon [n] \to [R]$ such that*

$$\Pr_{A \in \binom{[n]}{k}}\left[\Delta(f|_A, G[A]) \leqslant r\right] \geqslant \delta.$$

1.3 Applications and Motivations

In this section, we discuss some applications and motivating fields and type of problems Theorem 1.4 (and possible extensions of it) are likely to be related to. For some of them, we already have initial leads (and pursue them in subsequent papers as the current paper is already long enough as is), while for others the connection is more speculative.

1.3.1 Hardness of Approximation. Recall Mossel's result [23], asserting that in the case that μ is a connected distribution only the low-degree part of functions contributes to (1). For low-degree functions, one has the invariance principle of [24], and thus the combination of these two results can be seen as transforming expectations as in (1) to expectations over Gaussian space. This result has a few notable striking consequences in the field of hardness of approximation. Most notably, Raghavendra [26] uses precisely such ideas to show the relationship between dictatorship tests and Gaussian rounding scheme to semi-definite relaxations.

In this light, the result proved in this paper shows that only functions that are "characters times low-degree functions" can contribute to (1), and this suggests that an invariance principle that extends the invariance principle of [24] should exist. Indeed, in a future work [1] we are exploring this direction and will prove a more general such invariance principle, and discuss its relation to rounding schemes that combine semi-definite programming relaxations as well as linear programming relaxations. We believe such invariance principles will be crucial in the journey of understanding the approximability of satisfiable constraint satisfaction problems.

1.3.2 Higher Arity Predicates. The original motivation behind the question considered in this paper is the non-Abelian embedding hypothesis of [2], which is the following statement. Suppose $k \geqslant 3$ is an integer, $\Sigma_1, \dots, \Sigma_k$ finite alphabets and μ is a distribution over $\Sigma_1 \times \dots \times \Sigma_k$ in which the probability of each atom is at least $\alpha > 0$. We say μ admits an Abelian embedding if there is an Abelian group $(H, +)$ and maps $\sigma_i : \Sigma_i \to H$ for $i = 1, \dots, k$ such that $\sum_{i=1}^{k} \sigma_i(x_i) = 0$ for all $(x_1, \dots, x_k) \in \mathrm{supp}(\mu)$. We say μ admits a non-trivial Abelian embedding if at least one of the maps σ_i is non-constant.

Hypothesis 1.7. *In the above setting, if μ admits no non-trivial Abelian embeddings, then for all $\varepsilon > 0$ there is $\delta > 0$ such that if $f_i : \Sigma_i^n \to \mathbb{C}$ are 1-bounded functions with $\mathrm{Stab}_{1/2}(f_i; \mu_i^{\otimes n}) \leqslant \delta$ for at least one of the i's, then*

$$\left| \mathop{\mathbb{E}}_{(x_1,\dots,x_k)\sim\mu^{\otimes n}} \left[\prod_{i=1}^{k} f_i(x_i) \right] \right| \leqslant \delta.$$

In [2] a special case of this hypothesis is proved for a class of $k =$ 3-ary distributions, and in [4] this hypothesis is proved in general for all $k =$ 3-ary distributions. In these terms, the current paper does not signify any further progress towards establishing Hypothesis 1.7 beyond the case of 3-ary predicates, however we believe that the stability version proved herein will be crucial towards making further progress in this direction.

1.3.3 Gowers' Norms. Theorem 1.4 can be seen as an analog of the U_2-inverse theorem for Gowers uniformity norms [17] for general distributions. In the context of Gowers uniformity norms, the U_2-inverse theorem is a simple Fourier analytic computation only involving Fourier coefficients. Interestingly, at a point in our argument we too have to carry out such a computation (this is, however, a small part of the proof). It is tempting to speculate, and we have initial leads for this fact, that there should be higher order analogs of Gowers inverse theorems in the much more general setting of Theorem 1.4.

If true, such statements could be very useful to make progress on multiple problems in extremal combinatorics, and in particular in Szemerédi-type theorems [29]. This is so because it appears they are strong enough to facilitate density increment arguments. Indeed, as we explain next, in a companion paper we have used Theorem 1.4 to give effective bounds for the problem of finding restricted 3-arithmetic progressions in dense sets in $\mathbb{F}_p^n$, for a prime p.

1.3.4 Extremal Combinatorics. A set $A \subseteq \mathbb{F}_p^n$ is called somewhat restricted 3-AP free if it does not contain an arithmetic progression $x, x+a, x+2a$ where $x \in \mathbb{F}_p^n$ and $a \in \{0,1,2\}^n \setminus \{\vec{0}\}$. In a companion paper [3], we use Theorem 1.4 to give effective bounds on the density of restricted 3-AP sets:

Theorem 1.8. *There are absolute constants $C > 0$ and $1 \leqslant k \leqslant 10$ such that if $A \subseteq \mathbb{F}_p^n$ is a restricted 3-AP set, then*

$$\mu(A) = \frac{|A|}{p^n} \leqslant \frac{C}{\log^{(k)} n},$$

where $\log^{(k)} n$ is the k-fold iterated logarithm function.

Previously, the best known bound was $O(1/\log^* n)$, achieved by appealing to a quantitative version of the density Hales-Jewett theorem [25]. Theorem 1.8 makes progress on a question of Green [18] and on a question of Haszla, Holenstein and Mossel [23].

1.3.5 Multi-Player Parallel Repetition Theorems. Parallel repetition is a basic building block in the area of interactive protocols and in particular in applications in the field of hardness of approximation. In the setting of k-player games, we have a basic game Ψ involving a verifier and k players. The game consists of a set of questions X that are supposed to get labels from a finite alphabet Σ, a predicate $P : X^k \times \Sigma^k \to \{0,1\}$ that gives k-challenges and answers to them dictates whether these answers are deemed satisfactory or not, and a distribution μ over k-tuples of challenges. In the basic game Ψ, the verifier samples a challenge $(x_1, \dots, x_k) \sim \mu$, sends the question x_i to the ith player, receives an answer $a_i \in \Sigma$ from player i, and then accepts if and only if $P(x_1, \dots, x_k, a_1, \dots, a_k) = 1$. The value of the game, $\mathsf{val}(\Psi)$, is defined to be the maximum probability the verifier accepts under the best strategy for the players.

The t-fold repeated game, $\Psi^{\otimes t}$, is a game in which the verifier samples t sets of challenges, say $(x_{1,j}, \dots, x_{k,j}) \sim \mu$ for $j = 1, \dots, t$ independently, sends $(x_{i,1}, \dots, x_{i,t})$ to player i, receives from them answers $(a_{i,1}, \dots, a_{i,t})$ and accepts if and only if

$$P(x_{1,j}, \dots, x_{k,j}, a_{1,j}, \dots, a_{k,j}) = 1 \ \forall j = 1, \dots, t.$$

In words, the game is repeated for t-times, but in parallel, and the verifier confirms that each one of the executions of the basic game was accepting. It is clear that $\mathsf{val}(\Psi^{\otimes t}) \geqslant \mathsf{val}(\Psi)^t$, and the main question of interest in parallel repetition theorems is regarding the rate of decay of $\mathsf{val}(\Psi^{\otimes t})$ as a function of t; in particular is this decay exponential?

For 2-player games, i.e. for the case that $k = 2$, this problem is by now well understood, and it is known that the value of $\Psi^{\otimes t}$ is indeed exponentially decaying in t (however not in the most obvious manner); see [6, 12, 19, 27, 28]. The techniques that go into these proofs are either information theoretical, or analytical. In a sense, the analytical proofs are based on the well-known fact that the eigenvalues of a matrix tensorize when one tensorizes the matrix, as it turns out that, in a sense, the value of a game can be vaguely viewed as eigenvalues of a matrix.

For $k \geqslant 3$, the situation is much more complicated, and the only known bound for general games is due to Verbitsky [30] and gives rather weak bounds (as, once again, it relies on the density Hales-Jewett theorem).

Recently, the work of [10] identified a class of games referred to as "connected games" for which the information theoretic techniques from the setting of 2-player games still work, which sparked renewed interest in multi-player parallel repetition theorems. We remark that the notion of "connectedness" therein is very much similar to the notion of connectedness of distribution in our setting (which is much stronger than pairwise connectedness). This motivated a recent line of works [14–16, 20] that studied parallel repetition of 3-player games over binary questions. This line of work started with studying a game known as the GHZ game (which is well known in the physics literature and is a bottleneck to the techniques of [10]), proving polynomial decay for it, and using this as a stepping stone to prove polynomial decay parallel repetition theorems for more general classes of games.

We believe that the notion of Abelian embeddability should have a fundamental connection to the problem of parallel repetition in multiplayer games. In a sense, this question too is about "tensorization" of some value, but in this time one has to deal with k-dimensional tensors as opposed to matrices. Some evidence to that has been given in [7], wherein the authors give a very simple proof for the fact that the value of the GHZ game is exponentially vanishing with t (as opposed to just polynomial) which is inspired by Abelian embeddability. In a sense, the proof proceeds by identifying that the GHZ game actually entails within it a $(\mathbb{Z}_4,+)$-type additive structure. Then, using this fact along with powerful theorems from additive combinatorics, the authors give a structural result on the set of strategies for the players that perform well, which are then analyzed directly.

While being speculative, we believe that such connection should indeed exist, and in it the quantitative aspects of Theorem 1.4 should be highly relevant. At the current state, the quantitative bounds we get are not very good, but we believe that with more effort these could be improved to results that would be able to show $2^{-t^{\Omega(1)}}$ rate of decay in parallel repetition.

1.4 High Level Overview of the Proof of Theorem 1.4

In this section we give a high level overview of the proof of Theorem 1.4. As such, we often omit details, make simplifying assumptions and appeal to intuition in order to concentrate on the main ideas. The details appear in the full version of the paper.

At its core, our argument relies on the following intuition: if μ does not admit any Abelian embedding, then Theorem 1.4 is just equivalent to the main result of [2, 4]. Thus, one idea is to try to identify all Abelian embeddings of μ, define partial basis for $L_2(\Sigma^n;\mu_x^{\otimes n})$, $L_2(\Gamma^n;\mu_y^{\otimes n})$ and $L_2(\Phi^n;\mu_z^{\otimes n})$ based on these Abelian embeddings and then show that for f, g and h to satisfy (1), it must be the case that they correlated with a function from the span of this partial basis. The intuition is completing the partial bases into complete bases, once we "peel off" these embeddings based functions the rest of the functions in the bases are "oblivious" to the fact that μ admits Abelian embeddings. So, once we "peel off" these embedding functions, the situation is analogous to the case that μ does not have any Abelian embeddings, in which case the result of [2, 4] kick in.

Much of the effort in our proof goes into formalizing this rough idea, and once one is able to do that the rest of the proof is more streamline (but still requires a significant effort). Below, we give step by step description of the way we formalize this intuition.

1.4.1 Step 1: Master Embedding. The first issue is a distribution μ may have multiple linear embeddings, and they may interact in a non-trivial way. Indeed, given an Abelian embedding (σ,γ,ϕ) of μ into H, one can define a partial basis by composing characters of H with the embedding functions. But how do we know that different embeddings give us different basis elements? How do we combine these partial bases into something that makes sense and is convenient to work with?

Our first step is to identify that one may define a single embedding, which we refer to as the master embedding, that encapsulates within it all of the Abelian embeddings of μ. Indeed, we show that if μ does not have any $(\mathbb{Z},+)$ embedding, then there is a size M such that any Abelian embedding of μ "comes from" an Abelian embedding of μ into an Abelian group of size at most M. Hence, to include all Abelian embeddings it suffices to only look into embeddings of μ into Abelian groups of size at most M, and as there are only finitely many such embeddings we can just tensorize them. That is, letting $\sigma_i\colon \Sigma\to H_i$ be all possible σ's in linear embeddings of μ, where $(H_i,+)$ are Abelian groups, the master embedding of x is $\sigma_{\mathsf{master}}\colon \Sigma\to\prod_{i=1}^{R} H_i$ defined by

$$\sigma_{\mathsf{master}}(x) = (\sigma_1(x),\ldots,\sigma_R(x)),$$

and similarly one may define γ_{master} and ϕ_{master}. With the master embeddings in hand we now have a sensible way of defining a partial basis for functions in x, y and z by considering compositions of characters from H with the master embeddings.

At the present state, this partial basis is not very convenient. For example, it may well be the case that there are distinct $\chi,\chi'\in H$ such that $\chi\circ\sigma_{\mathsf{master}} = \chi'\circ\sigma_{\mathsf{master}}$. Indeed, this would be the case if the image of σ_{master} was a strict subgroup of H. More generally, linear dependencies within $\{\chi\circ\sigma_{\mathsf{master}}\}_{\chi\in\hat{H}}$ already start appearing as soon as the image of σ_{master} is not the entire group H, and this presents issues which we wish to avoid.

1.4.2 Step 2: Saturating the Master Embeddings. Our goal is therefore to arrange for the master embeddings σ_{master}, γ_{master} and ϕ_{master} to be *saturated*, meaning that the image of each one of them is the entire group H. To do so, we must change the distribution μ into a distribution μ' such that (a) on μ' the master embeddings are saturated, (b) there is a good enough relationship between 3-wise correlations over μ and 3-wise correlations over μ', and (c) we can deduce the conclusion of Theorem 1.4 on μ from the conclusion of Theorem 1.4 on μ'.

This transformation is achieved via the path trick, introduced in [2], which is ultimately just an application of the Cauchy-Schwarz inequality. The path trick is used in our arguments extensively, and often time the structure we need is quite subtle thereby requiring a very careful application of the path-trick. Nevertheless, below we explain at a high level the intuition behind the path trick and what it achieves.

Given a distribution μ, the path trick distribution (of length $2t+1$) with respect to x can be described as the following distribution μ':

1. Sample $(x_1,y_1,z_1)\sim\mu$.
2. Make a step from y: sample $(x_1',y_1',z_1')\sim\mu$ conditioned on $y_1'=y_1$.
3. Make a step from z: sample (x_2,y_2,z_2) conditioned on $z_2=z_1'$.
4. Repeat make a step from y/ make a step from z for t times.

Thus, the sequences $(y_1,y_1',y_2,y_2',\ldots,y_t,y_t',y_{t+1})$ of y's and the sequence $(z_1,z_1',z_2,z_2',\ldots,z_t,z_t',z_{t+1})$ of z's are generated (where $z_{i+1}=z_i'$ and $y_i'=y_i$), as well as $\vec{x}=(x_1,x_1',\ldots,x_t,x_t',x_{t+1})$ of x's. The output of the distribution ν is $(\vec{x},y_{t+1},z_1)$, and it is thought of as a 3-ary distribution over $\Sigma'\times\Gamma\times\Phi$ where $\Sigma'\subseteq\Sigma^{2t+1}$ is the set of feasible tuples of x in the procedure.

We refer to this procedure as the path trick since one may consider the bi-partite graph $G = (\Gamma \cup \Phi, E)$ whose edges are $(y, z) \in \Gamma \times \Phi$ that are in the support of $\mu|_{\Gamma\times\Phi}$. Thus, thinking of the x'es as labeling the edges of G, namely labeling an edge (y, z) by x if $(x, y, z) \in \mathsf{supp}(\mu)$, one gets that the above procedure generates a random path of length $2t + 1$ in the graph and record the labels of the edges that it traversed on.

Moving from the distribution μ to μ' has several benefits that have been used in our earlier papers:

(1) **Improving connectivity:** if μ is $\{2, 3\}$-connected, then for large enough t the support of μ' on the last two coordinates is full. Indeed, taking the random path view of the path trick, it is clear that as the graph G is connected, for sufficiently large t the same graph corresponding to μ' would be a complete bipartite clique.

(2) **Preserving properties of μ:** he distribution μ' preserves much of the properties of the distribution μ. In particular, if μ is pairwise connected then so is μ', and if μ does not admit any Abelian embeddings, then so does μ'.

(3) **The 3-wise correlations relations:** 3-wise correlations of functions over μ can be upper bounded by 3-wise correlations of functions related to the original functions over μ'. Indeed, assume for simplicity that the functions are real valued. If f, g and h achieve large 3-wise correlation in μ, then for $(x, y, z) \sim \mu^{\otimes n}$ one has that the values $h(z)$ and $f(x)g(y)$ are correlated, so looking at the above path we get that $h(z_{i+1}) \approx f(x_{i+1})g(y_{i+1})$ and $g(y_{i+1}) \approx f(x'_{i+1})h(z'_{i+1})$ and combining these we get that

$$\begin{aligned} h(z_{t+1}) &\approx f(x'_{t+1})h(z'_{t+1}) \\ &\approx f(x'_{t+1})f(x_t)g(y_t) \\ &\approx \ldots \\ &\approx f(x'_{t+1})f(x_t)\cdots f(x_2)f(x'_1)g(y_1). \end{aligned}$$

Thus, we expect g, h and $F(\vec{x}) = f(x'_{t+1})f(x_t)\cdots f(x_2)f(x'_1)$ to achieve a significant correlation in μ'. Indeed, this can be proved via an appropriate application of the Cauchy-Schwarz inequality.

For the purposes of this paper we need additional properties of the path trick transformations, which we explain next.

(1) **Abelian ebmeddings of μ lift to Abelian embeddings of μ':** not only does the path trick preserve lack of Abelian embeddings, but in fact if μ does admit Abelian embeddings, then μ' does not introduce any new ones. To be more precise, suppose that $\sigma\colon \Sigma \to H$, $\gamma\colon \Gamma \to H$ and $\phi\colon \Phi \to H$ are Abelian embeddings of μ. Then, these embeddings give rise to an Abelian embedding $\sigma_t\colon \Sigma' \to H$ with γ and ϕ of ν, as follows:

$$\sigma_t(\vec{x}) = \sum_{i=1}^{t} \sigma(x_i) - \sigma(x'_i) + \sigma(x_{t+1}). \tag{3}$$

With the notation above, we have that $\sigma(x_i)+\gamma(y_i)+\phi(z_i) = 0$, $\sigma(x'_i) + \gamma(y'_i) + \phi(z'_i) = 0$, and doing a proper addition/substraction one gets that

$$\sigma_t(\vec{x}) + \gamma(y_{t+1}) + \phi(z_1) = 0,$$

hence (σ_t, γ, ϕ) form an Abelian embedding of μ' into H.

(2) **The only Abelian embeddings of μ' are lifts of Abelian embeddings of μ:** all Abelian embeddings of μ' are precisely of this form. Namely, for any Abelian embedding (σ', γ, ϕ) of μ' into an Abelian group $(H, +)$ there is an Abelian embedding (σ, γ, ϕ) of μ into $(H, +)$ where σ satisfies a relation as in (3) where σ' plays the role of σ_t.This result has a few important consequences, and in particular it says that the path trick preserves master embeddings. Namely, if we start with a master embedding of μ, apply the path trick and the above transformation corresponding to it on the embeddings, then we will get the master embedding of μ'.

(3) **Saturating the embeddings:** it can be easily observed that if $(\sigma_{\mathsf{master}}, \gamma_{\mathsf{master}}, \phi_{\mathsf{master}})$ is a master embedding of μ (or for this purpose, any embedding of μ), then after the path trick we get the embedding $(\sigma_{\mathsf{master},t}, \gamma_{\mathsf{master}}, \phi_{\mathsf{master}})$ that clearly satisfies that $\mathsf{Image}(\sigma_{\mathsf{master},t}) \subseteq \mathsf{Image}(\sigma_{\mathsf{master}})$; this follows by looking at trivial paths that traverse the same edge back and fourth and use the same label of x all of the time. Moreover, it is clear that if $\mathsf{Image}(\sigma_{\mathsf{master}})$ was a subgroup of H then we would have that $\mathsf{Image}(\sigma_{\mathsf{master},t}) = \mathsf{Image}(\sigma_{\mathsf{master}})$. It is reasonable to expect that unless the set $\mathsf{Image}(\sigma_{\mathsf{master}})$ is indeed a subgroup, then for large enough t we would have that $\mathsf{Image}(\sigma_{\mathsf{master},t}) \subsetneq \mathsf{Image}(\sigma_{\mathsf{master}})$, in which case we enlarged the image of σ_{master} via the path trick.
Indeed, something along these lines is true. Namely, we show that by combination of path tricks along different directions (not only x) one can indeed always enlarge the image of an embedding so long as it is not a subgroup.[1]

In conclusion, using the path trick multiple times we can pass to a new distribution ν on which the embeddings are all saturated, 3-wise correlations over μ are upper bounded by 3-wise correlations over ν, and ν has improved connectivity – say that its support on the last two coordinates is full. It can be easily shown that in that case, the images of all of the components must be the same subgroup, and without loss of generality we assume it is the group H itself.

Note that in particular, the above properties mean that if

$$(\sigma_{\mathsf{master}}, \gamma_{\mathsf{master}}, \phi_{\mathsf{master}})$$

is a saturated master embedding of ν, then the distribution of $(\sigma_{\mathsf{master}}(x), \gamma_{\mathsf{master}}(y), \phi_{\mathsf{master}}(z))$ where $(x, y, z) \sim \nu$ has a full support on

$$\left\{ (a, b, c) \in H^3 \,\middle|\, a + b + c = 0 \right\},$$

which intuitively says that by moving from μ to ν we have "exposed" all of the Abelian structure in the distribution μ.

1.4.3 Step 3: Setting up a Basis Consisting of Embedding and Non-embedding Functions. Fix distributions μ over $\Sigma \times \Gamma \times \Phi$ and ν over $\Sigma' \times \Gamma' \times \Phi'$ as we have done so far, and suppose that (a power of) the 3-wise correlation of f, g and h over μ is upper bounded by the 3-wise correlation of F, G and H over ν.

Now that we have saturated the master embeddings in ν we can set up a partial for functions in $x \in \Sigma'$ as basis as before $\tilde{\chi}(x) = \chi(\sigma_{\mathsf{master}}(x))$ for all $\chi \in \hat{H}$ as before and get that now these

[1] In our formal proof this has to be done rather carefully as we wish to preserve the property that the alphabet of x is always a power of the original alphabet Σ.

functions are indeed linearly independent We can then complete it to a basis of $L_2(\Sigma'; \nu_x)$ by adding to it the functions $\psi_1, \dots, \psi_s$ that are orthogonal to all functions in $\mathsf{Span}(B_1)$, so that writing $B_1 = \left\{ \tilde{\chi} \mid \chi \in \hat{H} \right\}$ and $B_2 = \{\psi_1, \dots, \psi_s\}$ we have a basis $B_1 \cup B_2$ for $L_2(\Sigma'; \nu_x)$. Tensorizing, we get that $\left\{ v_{\vec{b}} \right\}_{\vec{b} \in (B_1 \cup B_2)^{\otimes n}}$ where $v_{\vec{b}} \colon \Sigma'^n \to \mathbb{C}$ is defined by $v_{\vec{b}}(x) = \prod_{i=1}^{n} v_{b_i}(x_i)$, is a basis for $L_2(\Sigma'^n; \nu_x^{\otimes n})$. Thus, we can write

$$F(x_1, \dots, x_n) = \sum_{\alpha \in (B_1 \cup B_2)^n} \widehat{F}(\alpha) v_\alpha(x).$$

We can define analogous bases for $L_2(\Gamma'; \nu_y)$ and $L_2(\Phi'; \nu_z)$. Now, each one of the functions F, G and H has an "embedding part", which is the parts of the monomials that use functions from B_1, and "non-embedding parts", which are monomials using functions from B_2. Intuitively, it should be the case that the more mass the functions have on the non-embedding parts, the smaller the 3-wise correlations would be; this is because that for uni-variate functions $u \colon \Sigma' \to \mathbb{C}$, $v \colon \Gamma' \to \mathbb{C}$ and $w \colon \Phi' \to \mathbb{C}$ of 2-norm 1, to achieve perfect 3-wise correlation it must be the case that $u(x) = v(y)w(z)$ in the support of ν, in which case u, v and w behave like an embedding function. We remark that there is a serious leap in this last step, which causes complications in later points of the argument. Later on, we refer to this gap as the Horn-SAT obstruction, and we will explain how it arises and how to overcome it later on.

In light of the above, it makes sense to define two notions of degrees for our partial basis. The first of which is the embedding degree of a monomial $v_{\vec{b}}$, $\mathsf{embeddeg}(v_{\vec{b}})$, which is the number of components v_{b_i} that come from the partial embedding basis B_1. The second of which is the non-embedding degree of a monomial $v_{\vec{b}}$, $\mathsf{nedeg}(v_{\vec{b}})$, which is the number of components of v_{b_i} that come from B_2.

1.4.4 *Step 4: Analyzing the Contribution of High Non-embedded Degree Components.*

The above discussion suggests that the parts of F, G and H of high non-embedding degree should contribute very little to their 3-wise correlation according to ν. Formally showing this, however, is quite tricky and this is where a considerable amount of effort in this paper is devoted to. Our argument here builds on an argument from [4].

To give some intuition for the argument we make several simplifying assumptions (some of which can be ensured, while other are not necessary). Assume that the marginal distribution of ν over y, z is uniform, and that the distribution of

$$(\sigma_{\mathsf{master}}(x), \gamma_{\mathsf{master}}(y), \phi_{\mathsf{master}}(z))$$

where $(x, y, z) \sim \nu$ is uniform over

$$\left\{ (a, b, c) \in H^3 \,\middle|\, a + b + c = 0 \right\}.$$

Further assume that the functions F, G and H are embedding homogenous and non-embedding homogenous functions, by which we mean that the embedding degree of each monomial of F is the same, and also the non-embedding degree of each monomial in F is the same; the same goes for G and H. Our argument here will be inductive on the number of coordinates n, and we show that the 3-wise correlation of functions F, G and H as above can be upper bounded by either the 3-wise correlation of $n-1$ variate functions of the same non-embedding degree, or by $(1 - \Omega(1))$ times the 3-wise correlation of $n - 1$ variate functions with non-embedding degree smaller by 1. Thus, iterating we would ultimately get a bound of $(1 - \Omega(1))^{\mathsf{nedeg}(F)}$ on the 3-wise correlations, which is small if the non-embedding degree of F is high.

In fact, we have two separate inductive arguments depending on if n is much larger than the non-embedding degree of F, or if it is comparable to it; we refer to this last case as the "near linear non-embedding degree case", and we now elaborate on each one of these cases.

The case that n is much larger than the non-embedding degree of F. In this case there is a variable, say the nth variable, such that in almost all of the mass of F lies on monomials in which the component of x_n is an embedding function. Using the homogeneity of F we can use find a decomposition of F as

$$\sum_{t \in T} \psi_t F_t(x_1, \dots, x_{n-1}) F'_t(x_n)$$

where each F'_t is either from B_1 or from B_2, and $\{F_t\}$, $\left\{F'_t\right\}$ form orthonormal sets and $\sum_{t \in T} |\psi_t|^2 = 1$. Similarly, we can find analogous decompositions form G and H as

$$\sum_{r \in R} \kappa_r G_r(y_1, \dots, y_{n-1}) G'_r(y_n), \qquad \sum_{s \in S} \rho_s H_s(z_1, \dots, z_{n-1}) H'_s(z_n).$$

Moreover, if F'_t is a function from B_1 then F_t has the same non-embedding degree as F, and if F'_t is from B_2 then F_t has one smaller non-embedding degree. The same goes for G and H, so to simplify presentation we consider the specialized case where

$$F(x) = \psi_1 F_1(x_1, \dots, x_{n-1}) F'_1(x_n) + \psi_2 F_2(x_1, \dots, x_{n-1}) F'_2(x_n),$$
$$G(y) = \kappa_1 G_1(y_1, \dots, y_{n-1}) G'_1(y_n) + \kappa_2 G_2(y_1, \dots, y_{n-1}) G'_2(y_n),$$
$$H(y) = \rho_1 H_1(z_1, \dots, z_{n-1}) G'_1(z_n) + \rho_2 H_2(z_1, \dots, z_{n-1}) H'_2(z_n),$$

where F'_1, G'_1 and H'_1 are embedding functions and F'_2, G'_2 and H'_2 are non-embedding functions. Thus, the coefficient ψ_2 is related to the mass x_n has on non-embedding functions and by choice is therefore small, and similarly we can expect it to be the case that κ_2 and ρ_2 are also small (which is true, but requires some preparatory work). Thus, the 3-wise correlation of F, G and H according to ν can be written as

$$\begin{aligned}
\mathop{\mathbb{E}}_{\nu^n}[FGH] = \;& \psi_1 \kappa_1 \rho_1 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1 G_1 H_1] \mathop{\mathbb{E}}_{\nu}\left[F'_1 G'_1 H'_1\right] \\
&+ \psi_1 \kappa_1 \rho_2 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1 G_1 H_2] \mathop{\mathbb{E}}_{\nu}\left[F'_1 G'_1 H'_2\right] \\
&+ \psi_1 \kappa_2 \rho_1 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1 G_2 H_1] \mathop{\mathbb{E}}_{\nu}\left[F'_1 G'_2 H'_1\right] \\
&+ \psi_1 \kappa_2 \rho_2 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1 G_2 H_2] \mathop{\mathbb{E}}_{\nu}\left[F'_1 G'_2 H'_2\right] \\
&+ \psi_2 \kappa_1 \rho_1 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_2 G_1 H_1] \mathop{\mathbb{E}}_{\nu}\left[F'_2 G'_1 H'_1\right] \\
&+ \psi_2 \kappa_1 \rho_2 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_2 G_1 H_2] \mathop{\mathbb{E}}_{\nu}\left[F'_2 G'_1 H'_2\right] \\
&+ \psi_2 \kappa_2 \rho_1 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_2 G_2 H_1] \mathop{\mathbb{E}}_{\nu}\left[F'_2 G'_2 H'_1\right] \\
&+ \psi_2 \kappa_2 \rho_2 \mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_2 G_2 H_2] \mathop{\mathbb{E}}_{\nu}\left[F'_2 G'_2 H'_2\right]. \qquad (4)
\end{aligned}$$

It turns out that terms the only term involving F_1' that does not vanish is $\mathbb{E}_\nu\left[F_1'G_1'H_1'\right]$. Indeed, as F_1' is a function from B_1 we may write it as a product of a function on y with a function on z, and thus expectations such as $\mathbb{E}_\nu\left[F_1'G_1'H_2'\right]$ can be written as expectation of product of a function of y and a function of z. Using independence, this product can be further written as the product of two expectations where at least one of these expectations is 0.

Thus, if the terms involving F_2 were not existent, then we would get the upper bound

$$\left|\mathop{\mathbb{E}}_{\nu^n}[FGH]\right| \leqslant |\psi_1\kappa_1\rho_1|\left|\mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1G_1H_1]\right|\left|\mathop{\mathbb{E}}_{\nu}\left[F_1'G_1'H_1'\right]\right| \leqslant \left|\mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_1G_1H_1]\right|,$$

and we have decreased the number of variables n by 1 (while keeping the non-embedding degree. In a sense, as ψ_2 is small this term indeed should constitute the majority of the contribution to $\mathbb{E}_{\nu^{\otimes n}}[FGH]$, but we cannot just ignore the contribution from the other terms.

A naive attempt at bounding the other term (and using the Cauchy-Scharz in a favorable way) can show that $|\mathbb{E}_{\nu^{\otimes n}}[FGH]|$ is at most the maximum of $\left|\mathbb{E}_{\nu^{\otimes n-1}}\left[F_iG_jH_k\right]\right|$ over i, j and k, however this bound is not good enough for us; indeed, if this maximum is achieved at anywhere other than $i = j = k = 1$ then the non-embedding degrees decrease, and in that case we must gain a factor of $(1 - \Omega(1))$ for our argument to go through.

The key to improve upon this naive attempt lies in what we refer to as the "additive base case". The additive base case is a statement about univariate functions that helps us to control the contribution form terms involving F_2' in a favorable way. Stated simply, the additive base case we use is the statement that if F' is a univariate non-embedding function, and G', H' are any univariate functions, then

$$\left|\mathop{\mathbb{E}}_{\nu}\left[F'(G'+H')\right]\right| \leqslant (1-\Omega(1))\|F'\|_2\|G'+H'\|.$$

The intuition for this inequality is that otherwise, the value of F' would be very close to the value of $\overline{G'}+\overline{H'}$, but this is only possible for embedding functions.

The point of the additive base case is that except for $\mathbb{E}_\nu\left[F_2'G_2'H_2'\right]$, the contribution of the terms not involving F_1' in (4) may be recasted as an expectation of the form dealt with in the additive base case. Indeed, if F_1', G_1' and H_1' were the simplest of embedding functions – namely constant functions – then this is rather clear, as these terms can be written as

$$\mathop{\mathbb{E}}_{\nu}\left[F_2'(aG_2+bH_2)\right]$$

for some coefficients a and b. In the case F_1', G_1' and H_1' are not the constant functions more effort is needed, and in particular one needs to guarantee that they "come" from the same character of H. Namely, that there is $\chi \in \hat{H}$ such that F_1', G_1' and H_1' are multiples of $\chi\circ\sigma_{\text{master}}$, $\chi\circ\gamma_{\text{master}}$ and $\chi\circ\phi_{\text{master}}$ respectively. As we are able to guarantee this fact, the contribution of these terms can still be associated with the additive base case, as essentially $G_1' = \overline{F_1'H_1'}$ and $H_1' = \overline{F_1'G_1'}$. Hence, that contribution can be re-written as

$$\mathop{\mathbb{E}}_{\nu}\left[F_1'F_2'(aG_2+bH_2)\right]$$

for some coefficients a and b, and this is still an expectation of the form handled by the additive base case.

Making an effective enough use of the additive base case, one can use 4 to either reduce n by 1 and keep the non-embedding degrees the same, or else reduce both n and the non-embedding degrees by 1 and then gain a factor of $1 - \Omega(1)$.

The near linear non-embedding degree case. We now consider the case that the non-embedding degree of F, G and H is comparable to n, the above argument no-longer works, and the reason is that the last term in (4) is no longer negligible, and at the same time we do not know how to give an effective upper bound on it using only the additive base case. Thus we must have a new base case that handles this last term, and intuitively one would expect the following base case to hold. Suppose that $\tilde{F}\colon\Sigma'\to\mathbb{C}$ is a function from B_2, and $\tilde{G}\colon\Gamma'\to\mathbb{C}$, $\tilde{H}\colon\Phi'\to\mathbb{C}$ are any functions of 2-norm 1. Then

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\nu}\left[\tilde{F}(x)\tilde{G}(y)\tilde{H}(z)\right]\right| \leqslant 1-\Omega(1).$$

The reason we expect this to be true is that otherwise, by compactness we would be able to find 3 such functions satisfying $\tilde{F}(x) = \tilde{G}(y)\tilde{H}(z)$ on the support of ν. Thus, the logs of these functions form an Abelian embedding so $\log(\tilde{F})$ must be an embedding function, in which case $\tilde{F}$ is also an embedding function in contradiction.

As stated, this argument is not quite correct, since it may be the case that the function F' gets the value 0 sometimes, in which case we cannot take logs and get away with it. This obstruction has already appeared in [4] wherein it was referred to as the "Horn-SAT obstruction", and here too we have to face it. In fact, as in our scenario we need to maintain many more properties of the distribution ν (compared to what was necessary in [4]), more care is needed to handle the Horn-SAT obstruction. Ultimately, the Horn-SAT obstruction is dealt with by stating a more complicated base case statement which we can guarantee to hold for the distribution ν while being useful enough to make our argument go through. For the simplicity of presentation however, we ignore this obstruction for now and explain the argument in the case the ideal base case holds.

Equipped with the ideal base case, we can give effective enough bounds on the last term in (4). In particular, if all of the contribution came from it, we would have been able to conclude that

$$\left|\mathop{\mathbb{E}}_{\nu^{\otimes n}}[FGH]\right| \leqslant (1-\Omega(1))\left|\mathop{\mathbb{E}}_{\nu^{\otimes n-1}}[F_2G_2H_2]\right|,$$

and iterating would finish the proof. One again however, there are other terms in (4) that need to be accounted for (the other terms involving F_2). To do so, ideally we would have liked (as in the additive base case) to re-arrange these terms so as to view their total contribution as an instantiation of the ideal base case, but this is not possible. Using a similar (but more complicated) argument we can still show that it is in fact the case that

$$\left|\mathop{\mathbb{E}}_{\nu^{\otimes n}}[FGH]\right| \leqslant (1-\Omega(1))\left|\mathop{\mathbb{E}}_{\nu^{\otimes n-1}}\left[\tilde{F}_2\tilde{G}_2\tilde{H}_2\right]\right|,$$

where $\tilde{F}_2, \tilde{G}_2$ and $\tilde{H}_2$ are functions of non-embedding degree at most 1 less of F_2, G_2 and H_2.

Overcoming the Horn-SAT Obstruction. A large part of our argument is devoted to gaining additional properties of ν, as well as other crucial reductions (for example, to allow us to assume homogeneity of F, G and H as above). One of the key properties achieved in this section is the so-called relaxed base case, which is a replacement for the ideal base case above that we are able to ensure.

A triplet of functions $u: \Sigma' \to \mathbb{C}$, $v: \Gamma' \to \mathbb{C}$ and $w: \Phi' \to \mathbb{C}$ is called a Horn-SAT embedding if $u(x) = v(y)w(z)$ in the support of ν. If u never vanishes, a Horn-SAT embedding can be transformed into an Abelian embedding, and thus (simply put) the Horn-SAT obstruction is really about the possible 0-patterns non-embedding functions may have. By careful manipulations (once again utilizing the path trick) we are able to find a set $\Sigma_{\text{modest}} \subseteq \Sigma'$ of size at least 2 such no Horn-SAT embedding can vanish on. Thus, we get that if u doesn't vanish on Σ_{modest} then it can never be a part of a Horn-SAT embedding. Therefore, it is natural to expect that if u has variance at least τ on Σ_{modest}, then

$$\left|\mathop{\mathbb{E}}_{\nu}[uvw]\right| \leqslant 1 - \theta(\tau)$$

where $\theta(\tau) > 0$ is some function of τ. This turns out to be true and useful, but there are many subtleties. For once, we need additional properties from Σ_{modest} to make this relaxed base case useful, and most important we need the symbols in Σ_{modest} to be mapped to the same group element in H by the master embedding. Secondly (and this has already appeared in [4]) we need a decent dependency between τ and $\theta(\tau)$.

1.4.5 Step 5: Reducing to Functions over H. We now return our functions F, G and H, armed with the knowledge that the contribution of high non-embedding degree parts if small. Thus, taking their parts of small non-embedding degree F', G' and H', we are able to conclude that $|\mathbb{E}_{\nu^{\otimes n}}[FGH] - \mathbb{E}_{\nu^{\otimes n}}[F'G'H']| \leqslant \frac{\varepsilon}{100}$, and so the 3-wise correlation of F', G' and H' according to ν is still significant.

We remark that as in our actual argument we will need the functions F', G' and H' to be bounded, so harsh truncations as we described do not fit the bill. Thus, we use a softer notion of truncations given by the *non-embedding noise operator.* For $\rho \in [0, 1]$ consider the Markov chain $\mathrm{T}_{\text{non-embed},\rho}$ on Σ' that on x, with probability ρ takes $x' = x$, and otherwise samples $x' \sim \nu|_x$ conditioned on $\sigma_{\text{master}}(x') = \sigma_{\text{master}}(x)$. When ρ is not specified, that is, when we write $\mathrm{T}_{\text{non-embed}}$, we mean that ρ is taken to be 0. Given such Markov chain one may consider the corresponding averaging operator on $L_2(\Sigma', \nu_x)$ given as

$$\mathrm{T}_{\text{non-embed},\rho} f(x) = \mathop{\mathbb{E}}_{x' \sim \mathrm{T}_{\text{non-embed},\rho} x}\left[f(x')\right],$$

and tensorize it to get an averaging operator

$$\mathrm{T}^{\otimes n}_{\text{non-embed},\rho}: L_2(\Sigma'^n, \nu_x^{\otimes n}) \to L_2(\Sigma'^n, \nu_x^{\otimes n}).$$

Simiarly, we can get averaging operators on the space$L_2(\Gamma'^n, \nu_y^{\otimes n})$ as well as on the space $L_2(\Phi'^n, \nu_z^{\otimes n})$, which by abuse of notation we also denote by $\mathrm{T}_{\text{non-embed},\rho}$. These averaging operators can be shown to essentially kill monomials of high non-embedding degree, hence serve as a replacement for harsh truncation arguments as above.

With these operators in hand, we can replace the harsh truncations above by $F' = \mathrm{T}^{\otimes n}_{\text{non-embed},\rho} F$, $G' = \mathrm{T}^{\otimes n}_{\text{non-embed},\rho} G$ and $H' = \mathrm{T}^{\otimes n}_{\text{non-embed},\rho} H$ (for suitably chosen ρ) and effectively be in the same situation as before, wherein we have functions F', G' and H' that have almost all of their mass on monomials with small non-embedding degrees, and also that

$$\left|\mathop{\mathbb{E}}_{\nu^{\otimes n}}[FGH] - \mathop{\mathbb{E}}_{\nu^{\otimes n}}\left[F'G'H'\right]\right| \leqslant \frac{\varepsilon}{100}.$$

We wish to transform the functions F', G' and H' into related bounded functions with non-embedding degree 0 for which the 3-wise correlation over ν is still significant. For that, we use a combination of random restrictions (so as the mass of F', G', and H' of small but not 0 non-embedding degree would almost all collapse to level 0), followed by averaging (to get rid of all monomials of positive non-embedding degree). Thus, we get functions $F''' = \mathrm{T}_{\text{non-embed}}(F'')$, $G''' = \mathrm{T}_{\text{non-embed}}(G'')$ and $H''' = \mathrm{T}_{\text{non-embed}}(H'')$ where F'', G'' and H'' are random restrictions of F', G' and H', so that with noticeable probability we have that

$$\mathop{\mathbb{E}}_{(x,y,z)\sim \nu^{\otimes n'''}}\left[F'''(x)G'''(y)H'''(z)\right] \geqslant \frac{\varepsilon}{2}$$

where n''' is the number of coordinates left alive after the random restriction. Now the functions F''', G''' and H''' can be viewed as functions defined over $H^{n'''}$, so the above expectation should be amendable to standard tools from discrete Fourier analysis.

1.4.6 Step 6: Applying the Linearity Testing Argument. Indeed, we re-cast the functions F''', G''' and H''' above as $F^\sharp: H^{n'''} \to \mathbb{C}$, $G^\sharp: H^{n'''} \to \mathbb{C}$ and $H^\sharp: H^{n'''} \to \mathbb{C}$ defined in the natural way (for example, $F^\sharp(a) = F(x)$ for x such that $\sigma_{\text{master}}(x_i) = a_i$ for each coordinate, where we know that the specific choice of x doesn't matter). Thus, from the distribution ν we get a corresponding distribution $\nu^\sharp$ over $\left\{(a,b,c) \in H^3 \,\middle|\, a+b+c=0\right\}$ whose support is full, and

$$\left|\mathop{\mathbb{E}}_{(a,b,c)\sim(\nu^\sharp)^{\otimes n'''}}\left[F^\sharp(a)G^\sharp(b)H^\sharp(c)\right]\right| \geqslant \frac{\varepsilon}{2}.$$

We now use random restrictions again, but for a different reason. Namely, we use random restrictions to shift from the distribution $\nu^\sharp$ to the uniform over $\left\{(a,b,c) \in H^3 \,\middle|\, a+b+c=0\right\}$, and get from $F^\sharp$, $G^\sharp$ and $H^\sharp$ restrictions $F^{\sharp'}$, $G^{\sharp'}$ and $H^{\sharp'}$ so that with noticeable probability

$$\left|\mathop{\mathbb{E}}_{a,b\in H^{n''''}}\left[F^{\sharp'}(a)G^{\sharp'}(b)H^{\sharp'}(-a-b)\right]\right| \geqslant \frac{\varepsilon}{4},$$

where n'''' is the number of coordinates left alive. In this case, a straightforward, classical Fourier analytic computation can be applied to relate the left hand side to the Fourier coefficients of $F^{\sharp'}$, $G^{\sharp'}$ and $H^{\sharp'}$ so that we get

$$\left|\sum_{\chi\in\hat{H}^{\otimes n''''}} \widehat{F^{\sharp'}}(\chi)\widehat{G^{\sharp'}}(\chi)\widehat{H^{\sharp'}}(\chi)\right| \geqslant \frac{\varepsilon}{4},$$

from which one can quickly conclude that there is χ such that $\left|\widehat{F^{\sharp\prime}}(\chi)\right| \geqslant \frac{\varepsilon}{4}$. In words, after a sequence of random restrictions, averaging and further random restriction, the function F is correlated with a function of the form $\chi \circ \sigma_{\text{master}}$. This is the type of result we are after, except that we wish to have such result for F and not for F after this sequence of operations.

1.4.7 Step 7: Going back to F via the Restriction Inverse Theorem. We have thus concluded that after random restriction, $F^{\sharp}$ is correlated with a function of the form $\chi \circ \sigma_{\text{master}}$ where $\chi \in \hat{H}^{n''''}$, and we wish to unravel the steps we took to get from F to $F^{\sharp}$ and conclude a structural result about F.

Noting that $\chi \circ \sigma_{\text{master}}$ is a product function, this is precisely a situation in which the restriction inverse theorem kicks in, and using a modified version of Theorem 1.5 we are able to conclude that $F^{\sharp}$ is correlated with a function of the form $L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}}$ where $\chi \in \hat{H}^{n'''}$ and L is a low-degree function of 2-norm at most 1. Thus, the same conclusion holds for F''' (as it is essentially the same function as $F^{\sharp}$.

Recalling that $F''' = \mathrm{T}_{\text{non-embed},0} F''$, we get that

$$\left|\langle \mathrm{T}_{\text{non-embed},0} F'', L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}} \rangle\right| \geqslant \frac{\varepsilon}{4},$$

but on the other hand we also have that

$$\begin{aligned}
&\left|\langle \mathrm{T}_{\text{non-embed},0} F'', L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}} \rangle\right| \\
&= \left|\langle F'', \mathrm{T}^{*}_{\text{non-embed},0}(L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}}) \rangle\right| \\
&= \left|\langle F'', \mathrm{T}_{\text{non-embed},0}(L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}}) \rangle\right| \\
&= \left|\langle F'', L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}} \rangle\right|,
\end{aligned}$$

where we used the fact that $\mathrm{T}_{\text{non-embed},0}$ is self-adjoint. Hence, we conclude that F'' is correlated with $L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}}$.

We now wish to unravel the last step of random restriction (that goes from F to F''), and for that we once again want to appeal to the restriction inverse theorem. However, the correlations we are talking about now are not quite about correlations with product functions. Amusingly, to circumvent this issue we apply more random restrictions. Intuitively, after a suitably chosen random restriction, the function L becomes close to constant, hence one expects the fact that F'' is correlated with $L \circ \sigma_{\text{master}} \cdot \chi \circ \sigma_{\text{master}}$ to convert to the fact that a random restriction of F'' is correlated with a restriction of $\chi \circ \sigma_{\text{master}}$ (which is a product function), and we show that this is indeed the case. Thus, we conclude that a random restriction of F'' is correlated with a function of the form $\chi \circ \sigma_{\text{master}}$. Noting that a random restriction of F'' is (overall) a random restriction of F (with different parameters), we are thus able to conclude from the restriction inverse theorem that F itself is correlated with a function of the form $L \cdot \chi \circ \sigma_{\text{master}}$.

1.4.8 Step 8: Going back to f via Properties of the Master Embedding. The last step in the proof of Theorem 1.4 is to use the structural result obtained for the function F to deduce a similar structural result for f. For that, we recall that (ignoring complex conjugates) the value of $F(x_1, \ldots, x_s)$ is $f(x_1) \cdots f(x_s)$, and, ignoring the low-degree part L for now, we know that F is correlated with $\chi \circ \sigma_{\text{master}}$ for some $\chi \in \hat{H}^{n}$. Recalling the relation 3, one quickly gets from it that

$$\chi \circ \sigma_{\text{master}}(x_1, \ldots, x_s) = \prod_{i=1}^{s} \chi \circ \sigma_{\text{master}}(x_i),$$

where, by abuse of notation, σ_{master} on the right hand side is the master embedding of μ (which is the original distribution, prior to any application of the path trick). Hence, the correlation between F and $\chi \circ \sigma_{\text{master}}$ translates to the fact that

$$\left| \mathop{\mathbb{E}}_{(x_1,\ldots,x_s)\sim \nu_x^{\otimes n}} \left[\prod_{i=1}^{s} f(x_i) \chi \circ \sigma_{\text{master}}(x_i) \right] \right| \geqslant \varepsilon' = \varepsilon'(\varepsilon) > 0$$

As discussed earlier, in [23] it is shown that if ν_x is a connected distribution, a correlation such as in the above can be noticeable only if $f \cdot \chi \circ \sigma_{\text{master}}$ is correlated with a low-degree function. Thus, the proof would be concluded if we are able to ensure connectivity of ν_x, which we are indeed able to. This requires some care in some of our earlier steps, and most notably in the way we apply the path trick. In fact, we are able to guarantee that the support of ν_x is full, that is, Σ^s.

Bringing the low-degree part L back, essentially the same argument works except that we need to apply a suitable random restriction beforehand to get rid of the low-degree part. Thus, the previous argument gives that with noticeable probability, a random restriction of $f \cdot \chi \circ \sigma_{\text{master}}$ is correlated with a low-degree function. Hence, after more random restrictions, we conclude that with noticeable probability a random restriction of $f \cdot \chi \circ \sigma_{\text{master}}$ is correlated with a constant function. Re-phrasing, this means that with noticeable probability a random restriction of f is correlated with a function of the form $\chi \circ \sigma_{\text{master}}$, and a final invocation of the restriction inverse theorem finishes the proof.

REFERENCES

[1] Amey Bhangale, Subhash Khot, and Dor Minzer. [n. d.]. A Mixed Invariance Principle and Applications to CSPs.

[2] 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

[3] Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. Effective Bounds for Restricted 3-Arithmetic Progressions in $\mathbb{F}_p^n$. *CoRR* abs/2308.06600 (2023). https://doi.org/10.48550/arXiv.2308.06600 arXiv:2308.06600

[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] 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.* ACM, 335–340. https://doi.org/10.1145/2746539.2746565

[7] Mark Braverman, Subhash Khot, and Dor Minzer. 2022. Parallel Repetition for the GHZ Game: Exponential Decay. *CoRR* abs/2211.13741 (2022). https://doi.org/10.48550/arXiv.2211.13741 arXiv:2211.13741

[8] Irit Dinur, Yuval Filmus, and Prahladh Harsha. 2019. Analyzing Boolean functions on the biased hypercube via higher-dimensional agreement tests. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms.* SIAM, 2124–2133.

[9] Irit Dinur and Elazar Goldenberg. 2008. Locally testing direct product in the low error range. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science.* IEEE, 613–622.

[10] Irit Dinur, Prahladh Harsha, Rakesh Venkat, and Henry Yuen. 2017. Multiplayer Parallel Repetition for Expanding Games. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA (LIPIcs, Vol. 67)*. 37:1–37:16. https://doi.org/10.4230/LIPIcs.ITCS.2017.37

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

[12] 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*. ACM, 624–633. https://doi.org/10.1145/2591796.2591884

[13] Irit Dinur and David Steurer. 2014. Direct product testing. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*. IEEE, 188–196.

[14] Uma Girish, Justin Holmgren, Kunal Mittal, Ran Raz, and Wei Zhan. 2021. Parallel Repetition for the GHZ Game: A Simpler Proof. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2021, August 16-18, 2021, University of Washington, Seattle, Washington, USA (Virtual Conference) (LIPIcs, Vol. 207)*. 62:1–62:19. https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2021.62

[15] Uma Girish, Justin Holmgren, Kunal Mittal, Ran Raz, and Wei Zhan. 2022. Parallel repetition for all 3-player games over binary alphabet. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*. ACM, 998–1009. https://doi.org/10.1145/3519935.3520071

[16] Uma Girish, Kunal Mittal, Ran Raz, and Wei Zhan. 2022. Polynomial Bounds on Parallel Repetition for All 3-Player Games with Binary Inputs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference) (LIPIcs, Vol. 245)*. 6:1–6:17. https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2022.6

[17] William T Gowers. 2001. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA* 11, 3 (2001), 465–588.

[18] Ben Green. [n. d.]. 100 open problems. *manuscript* ([n. d.]).

[19] 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

[20] Justin Holmgren and Ran Raz. 2020. A Parallel Repetition Theorem for the GHZ Game. *CoRR* abs/2008.05059 (2020). arXiv:2008.05059 https://arxiv.org/abs/2008.05059

[21] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. 2010. Uniform Direct Product Theorems: Simplified, Optimized, and Derandomized. *SIAM J. Comput.* 39, 4 (2010), 1637–1665. https://doi.org/10.1137/080734030

[22] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. 2012. New direct-product testers and 2-query PCPs. *SIAM J. Comput.* 41, 6 (2012), 1722–1768.

[23] Elchanan Mossel. 2010. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis* 19, 6 (2010), 1713–1756.

[24] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. 2010. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics* 171, 1 (2010), 295–341.

[25] DHJ Polymath. 2012. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics* (2012), 1283–1327.

[26] Prasad Raghavendra. 2008. Optimal algorithms and inapproximability results for every CSP?. In *Proceedings of the fortieth annual ACM symposium on Theory of computing (STOC)*. 245–254.

[27] 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

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

[29] Endre Szemerédi. 1975. On sets of integers containing no k elements in arithmetic progression. *Acta Arith* 27, 299-345 (1975), 21.

[30] Oleg Verbitsky. 1996. Towards the parallel repetition conjecture. *Theoretical Computer Science* 157, 2 (1996), 277–282.

Received 21-OCT-2023; accepted 2024-02-11