Conditional Hardness for Satisfiable 3-CSPs

Ryan O'Donnell
Carnegie Mellon University
odonnell@cs.cmu.edu

Yi Wu
Carnegie Mellon University
yiwu@cs.cmu.edu

November 17, 2008

Abstract

In this paper we study a fundamental open problem in the area of probabilistic checkable proofs:

What is the smallest s such that $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,s}[O(\log n), 3]$?

In the language of hardness of approximation, this problem is equivalent to determining the smallest s such that getting an s-approximation for satisfiable 3-bit constraint satisfaction problems ("3-CSPs") is NP-hard.

The previous best upper bound and lower bound for s are $20/27 + \epsilon$ by Khot and Saket [19] and $5/8$ by Zwick [29]. In this paper we close the gap assuming Khot's d-to-1 Conjecture [15]. Formally, we prove that if Khot's d-to-1 Conjecture holds for any finite constant integer d, then $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,5/8+\epsilon}[O(\log n), 3]$ for any constant $\epsilon > 0$.

Our conditional result also solves Håstad's open question [12] on determining the inapproximability of *satisfiable* Max-`NTW` ("Not Two") instances and confirms Zwick's conjecture [29] that the 5/8-approximation algorithm for satisfiable 3-CSPs is optimal.

1 Introduction

1.1 The PCP Characterization of NP

The famous PCP (Probabilistic Checkable Proof) Theorem states that any language in NP has a proof system where the proofs can be probabilistically checked in a query-efficient way. The notation $\mathrm{PCP}_{c,s}(r(n), q(n))$ stands for the class of languages L verifiable by a proof system with the following parameters: for an input x of length n, the verifier uses $r(n)$ random bits and queries $q(n)$ bits in the proof to decide in polynomial time whether x is in L or not. The verifier has the following performance guarantees: i) if x is in L, there exists a proof that passes with probability c, and ii) if x is not in L, no proof passes with probability more than s. We call c the *completeness* and s the *soundness* of the verifier.

If the verifier decides which proof bits to query based only on x and the $r(n)$ random bits, the verifier is called *nonadpative*. On the other hand, if the verifier uses the results of previous queries to decide which proof bit to query next, the verifier is called *adaptive*. The notation aPCP and naPCP is used to distinguish languages verifiable by adaptive and nonadaptive verifiers. Adaptive verifiers can have better performance while nonadaptive verifiers have more natural implications for hardness of approximation for CSPs (see Theorem 1.3 for more discussion). We focus on nonadaptive proof systems in this paper.

Formally, the PCP Theorem [1, 2] states:

Theorem 1.1. $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,1/2}[O(\log n), O(1)]$.

In the PCP Theorem, the completeness c is 1; i.e., when the input x is in the language, there exists a proof that passes with probability 1. Such a verifier is said to have *perfect completeness*, which is a natural and desirable property of the proof system. As for the soundness, much effort is devoted to optimizing the tradeoff between $q(n)$ and s (as well as other parameters such as proof length, adaptivity, "free bit complexity", "alphabet size"...) [4, 11, 12, 27, 13, 19]. It is known that to achieve $c = 1$ and $s < 1$, the verifier must make at least 3 queries. This motivates the subject of study in this paper: optimizing the soundness s for 3-query nonadaptive PCP systems with perfect completeness. Formally, we examine the following question:

Question 1.2. *What is the smallest s such that* $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,s}[O(\log n), 3]$*?*

This problem was first studied by Bellare, Goldreich and Sudan [4] who achieved $s = 0.8999$. Håstad [12] further improved this result by achieving $s = 3/4 + \epsilon$ for every $\epsilon > 0$. Around the same time, Zwick [29] showed that $\mathrm{naPCP}_{1,5/8}[O(\log n), 3] \subseteq \mathrm{BPP}$ by giving a randomized polynomial-time 5/8-approximation algorithm for satisfiable 3-CSPs. This implies that unless $\mathrm{NP} \subseteq \mathrm{BPP}$, the best s for Question 1.2 must be bigger than 5/8. Zwick further conjectured that his algorithm is optimal:

Zwick's Conjecture: $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,5/8+\epsilon}[O(\log n), 3]$ *for all* $\epsilon > 0$.

See Section 1.2 for more discussion. No further progress was made for almost a decade, when Khot and Saket [19] showed that soundness $s = 20/27 + \epsilon \approx .741$ is achievable.

We note that certain relaxations of the problem are better understood. If we allow the verifier to be adaptive, Guruswami et al. [11] proved that $\mathrm{NP} \subseteq \mathrm{aPCP}_{1,1/2+\epsilon}[O(\log n), 3]$. If we allow an arbitrarily small loss of completeness for nonadaptive verifiers, Håstad [12] showed that $\mathrm{NP} \subseteq \mathrm{naPCP}_{1-\epsilon,1/2+\epsilon}[O(\log n), 3]$. By another result of Zwick [29], both of these results achieve optimal soundness assuming $\mathrm{NP} \not\subseteq \mathrm{BPP}$.

We think that Question 1.2 addresses an important missing part of our understanding of 3-query PCP systems. In addition, this question is equivalent to understanding the approximability of satisfiable 3-CSPs, as we now describe.

1.2 Max-kCSPs and Approximability

A k-bit Constraint Satisfaction Problem ("k-CSP") consists of a set of boolean variables, along with boolean constraints each of which involves at most k of these variables. Each boolean constraint in a k-CSP is some predicate of arity of at most k. Max-kCSP is the algorithmic problem of finding an assignment to the variables that maximizes the number of satisfied constraints. For a k-CSP instance, we use Opt to denote the maximum *fraction* of the constraints that can be satisfied. A k-CSP is called *satisfiable* if there exists an assignment that satisfies all the constraints; i.e., if Opt $= 1$. We can further specialize Max-kCSP by restricting the type of constraints to some predicate set Φ. For example, assuming the variables are called $x_1, \ldots, x_n$:

- Max-E3Lin: only the two predicates $x_i \oplus x_j \oplus x_k$, $\neg(x_i \oplus x_j \oplus x_k)$;
- Max-Cut: only the predicate $x_i \neq x_j$;
- Max-3Sat: only the $2 + 4 + 8$ predicates of the form ℓ_i, $\ell_i \vee \ell_j$, $\ell_i \vee \ell_j \vee \ell_k$, where ℓ_i denotes a "literal", either x_i or $\overline{x}_i$.

One additional, less familiar, example will be important for this paper:

- Max-NTW: only the 8 predicates of the form $\texttt{NTW}(\ell_i, \ell_j, \ell_k)$. Here NTW is the 3-ary predicate satisfied if and only the number of True inputs is zero, one, or three — i.e., "Not Two".

Algorithmically determining Opt for Max-kCSPs ($k \geq 2$) and for most "Max-Φ" problems is NP-hard [7, 6]. Much research therefore turns to the question of whether we can or cannot efficiently satisfy at least an $\alpha \cdot$ Opt fraction of the constraints. Most of the NP-hardness-of-approximation results are based on the following well-known connection between PCPs and hardness of approximation:

Theorem 1.3. *Let Φ be a set of predicates with arity no more than k. The following two statements are equivalent: i) It is* NP*-hard to distinguish whether a given Max-Φ instance has* Opt $\geq c$ *or has* Opt $\leq s$. *ii)* NP $\subseteq$ naPCP$_{c,s}(k, O(\log n))$, *where furthermore the verifier decides whether or not to accept based on applying a predicate from Φ to the proof bits it reads.*

Note that the *nonadaptiveness* is crucial in Theorem 1.3. If the verifier is adaptive in the above theorem, the equivalent hardness result would be for the more unnatural class of predicates Φ definable by *decision trees* of depth k.

As a direct application of the theorem, we have that Question 1.2 is equivalent to the following:

Question 1.4. *What is the smallest s such that it is* NP*-hard to distinguish whether a given Max-3CSP instance is satisfiable or has no assignment that satisfies more than an s fraction of constraints?*

We thus see why (unless NP $\subseteq$ BPP) Zwick's 5/8-approximation randomized algorithm for satisfiable Max-3CSP [29] mentioned earlier implies that the smallest s in Questions 1.2 and 1.4 must be bigger than 5/8. Further, Zwick's Conjecture is that $s = 5/8 + \epsilon$ is optimal for both Questions 1.2 and 1.4.

1.3 Optimal inapproximability, and Khot's Conjectures

For some important CSPs we have optimal (i.e., matching) approximation algorithms and NP-hardness-of-approximation results: Max-kLin(mod q) for $k \geq 3$ [12], Max-3Sat [12, 14, 30], and a few other Max-kCSP subproblems with $k \geq 3$ [12, 29, 28, 10]. All of the optimal hardness results are based on building a PCP system for a problem called Label-Cover (see Section 3 for a definition). For many other canonical problems such as Max-Cut and Max-2Sat, there is still a gap between the best known approximation algorithm and hardness result. To address this, Khot [15] proposed the *Unique Games Conjecture (UGC)* and *d-to-1 Conjectures* (see Section 3 for definitions). Assuming the UGC, we know optimal hardness-of-approximation results for several more problems, including Vertex-Cover [18], Max-Cut [16, 17, 23], and Max-2Sat [3]. A powerful recent result of Raghavendra [25] shows that for any Max-Φ CSP, the

optimal hardness factor — excluding the case when Opt = 1 — is equal to the integrality gap of a certain semidefinite program. Raghavendra's result uses the UGC.

Unfortunately, no hardness result based on the UGC can be applied to *satisfiable* Max-kCSPs and Max-Φ problems; i.e., problems with Opt = 1. The Unique Games Conjecture states that it is NP-hard to distinguish whether a "Unique Label-Cover" instance is *near satisfiable* or *far from satisfiable*; here "near satisfiable" cannot be replaced by "satisfiable", and this prevents us from getting any hardness result about satisfiable CSPs out of the UGC. We comment that the approximability of satisfiable Max-kCSP and Max-Φ can be very different from that of the near-satisfiable version. For example, satisfiable Max-3Lin instances can be solved exactly by a polynomial algorithm (Gaussian Elimination) whereas for near-satisfiable instances, i.e. Opt $= 1 - \epsilon$, it is NP-hard to do better than the trivial 2-approximation algorithm [12]. As another example, satisfiable Max-3CSPs instances have a 5/8-approximation algorithm [29] while near-satisfiable Max-3CSP instances are NP-hard to approximate beyond 1/2 [12].

To address the UGC's lack of perfect completeness, Khot additionally proposed the "d-to-1 Conjectures" [15]. The d-to-1 Conjecture states that it is NP-hard to distinguish whether a "d-to-1 Label-Cover instance" is satisfiable or far from satisfiable. The conjectures are parameterized by an integer constant $d \geq 2$. The bigger d is, the less restrictive are d-to-1 Label-Cover instances; hence for each d, the d-to-1 Conjecture implies the $(d+1)$-to-1 Conjecture. Prior to this work, the only application of the d-to-1 Conjectures was by Dinur et al. [8], who showed that the 2-to-1 Conjecture implies hardness of coloring 4-colorable graphs by $O(1)$ colors (and a few related results). In this paper we use a much weaker assumption: we only assume the d-to-1 Conjecture holds for *some* arbitrarily big (but constant) d.

1.4 Satisfiable Max-NTW

In Zwick's algorithm for satisfiable Max-3CSP, he observed that the bottleneck for improving the 5/8-approximation factor seemed to come from just one type of constraint: the NTW predicate described in Section 1.2. In the conclusion of Håstad's seminal paper on inapproximability [12] he posed only one concrete open question, a refinement of Zwick's Conjecture:

Question 1.5. *For each $\epsilon > 0$, given a satisfiable Max-NTW instance, is it NP-hard to find an assignment that satisfies more than an $5/8 + \epsilon$ fraction of the constraints?*

In other words, is *satisfiable* Max-NTW inapproximable beyond the the random assignment threshold of 5/8? (Note that Håstad proved this inapproximability for *near-satisfiable* Max-NTW instances in his paper.) A "yes" answer to this question is of course stronger than Zwick's Conjecture, since Max-NTW is a special Max-3CSP. As a result of Theorem 1.3, Question 1.5 is equivalent to deciding whether there is such a nonadaptive PCP system for an NP-complete language in which the verifier has perfect completeness, soundness $5/8+\epsilon$, and decides to accept or reject based on the NTW predicate (more precisely, the 8 NTW predicates gotten by allowing negated inputs). Constructing such a PCP system for d-to-1 Label-Cover is the main focus of the remaining paper.

2 Our Contribution and Methods

The main theorem in our paper is a "yes" answer to Håstad's open problem, Question 1.5, assuming that there exists a constant d for which Khot's d-to-1 Conjecture holds. Formally:

Theorem 2.1. *Suppose that Khot's d-to-1 Conjecture holds for some finite constant d. Then for any $\epsilon > 0$, there is a nonadaptive 3-query PCP system for NP that has perfect completeness and soundness $5/8 + \epsilon$; in addition, the verifier makes its decision based on an NTW predicate. Equivalently, given a satisfiable Max-NTW instance, it is NP-hard to satisfy more than a $5/8 + \epsilon$ fraction of the constraints.*

As discussed, this conclusion implies that the answer to Questions 1.2 and 1.4 is $s = 5/8 + \epsilon$, confirming Zwick's Conjecture:

Corollary 2.2. *Suppose that Khot's d-to-1 Conjecture holds for some finite constant d. Then* $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,5/8+\epsilon}[O(\log n), 3]$ *for every* $\epsilon > 0$.

2.1 Methods

Our proof is, in a way, similar to Håstad's inapproximability proof for Max-3Lin [12]. It uses the same overall framework: an "outer verifier" based on Label-Cover (in our case, d-to-1 Label-Cover) and an "inner verifier" based on a "Consistent-Dictators Test". There are two main challenges in the paper: i) *designing* an appropriate NTW-based Dictator Test, suitable for verifying an $O(1)$-to-1 Label-Cover instance; ii) analyzing the soundness of the proof system.

In [24], the authors proposed and analyzed a 3-query Dictator Test using the NTW predicate, with perfect completeness and soundness $5/8+\epsilon$. Unfortunately, it was a "single-function" test, generating queries from the space $\{-1,1\}^R \times \{-1,1\}^R \times \{-1,1\}^R$; as such, it was applicable only for use with Unique Label-Cover instances. But since the Unique Games Conjecture has imperfect completeness, the authors could not derive any new hardness-of-approximation result.

In this paper, we generalize the 3-query Dictator Test from [24]. Our new test, applicable for use with d-to-1 Label-Cover, generates queries according to a certain probability distribution $\mathcal{T}$ on the space $\{-1,1\}^R \times \{-1,1\}^{dR} \times \{-1,1\}^{dR}$. It is used to test that two functions $f : \{-1,1\}^R \to \{-1,1\}$ and $g : \{-1,1\}^{dR} \to \{-1,1\}$ are "consistent" Dictator functions.

In the resulting Fourier analysis of the PCP system, the main challenge is (as usual) to bound the expectation of a certain quadratic and cubic term. The analysis is more complicated compared with [24] and some very different techniques are used in this paper. We analyze the quadratic term using a novel argument about the positivity of certain linear operators.

As for the cubic term, we use Invariance Principle-style arguments as in [21, 20]. However, none of the results in [21, 20] can be applied as a black box in our proof. The reason is that our distribution $\mathcal{T}$ is *not* pairwise independent, and furthermore the correlation it has between $\{-1,1\}^R$ and $\{-1,1\}^{dR} \times \{-1,1\}^{dR}$ is actually 1. These facts introduce additional complications.

Our use of the Invariance Principle is notable in another way. Most other Invariance Principle proofs use it to pass from a given distribution to the Gaussian distribution. However, passing to the Gaussian distribution is not particularly useful for us. Instead we take full advantage of the fact that the Invariance Principle lets us pass to *any* distribution with the same pairwise correlations. Specifically, we find a different distribution on the *boolean cube* to work with, having the same (nonzero) pairwise correlations but without the difficult-to-analyze 3-wise correlations of the original distribution $\mathcal{T}$.

We hope the techniques in this paper will be useful in analyzing other satisfiable CSPs where the Unique Games Conjecture does not apply.

3 Khot's d-to-1 Conjectures, and PCP reductions

3.1 The d-to-1 Conjecture

To define d-to-1 Label-Cover and Khot's d-to-1 Conjecture, we first recall the basics of the Label-Cover problem.

Definition 3.1. *A Label-Cover instance $\mathcal{L}$ is defined by a tuple $(U, V, E, P, R_1, R_2, \Pi)$. Here U and V are the two vertex sets of a bipartite graph and E is the set of edges between U and V. P is an explicitly given probability distribution on E. R_1 and R_2 are integers with $1 \le R_1 \le R_2$. Π is a collection of "projections", one for each edge: $\Pi = \{\pi_e : [R_2] \to [R_1] \mid e \in E\}$. A labeling L is a mapping $L : U \to [R_1], V \to [R_2]$. We say that an edge $e = (u, v)$ is "satisfied" by labeling L if $\pi_e(L(v)) = L(u)$. We define:*

$$\mathrm{Opt}(\mathcal{L}) = \max_{\text{labelings } L} \quad \mathop{\mathbf{Pr}}_{e=(u,v)\sim P}[\pi_e(L(v)) = L(u)].$$

The fundamental inapproximability theorem of Raz [26] is the following:

Theorem 3.2. *For every constant $\eta > 0$ there is some constant $k(\eta) < \infty$ such that for Label-Cover instances $\mathcal{L}$ with $R_2 \geq k(\eta)$, it is NP-hard to distinguish the case* $\mathrm{Opt}(\mathcal{L}) = 1$ *from the case* $\mathrm{Opt}(\mathcal{L}) \leq \eta$.

We now define the d-to-1 property:

Definition 3.3. *A projection $\pi : [R_2] \to [R_1]$ is said to be "d-to-1" if for each element $i \in [R_1]$ we have $|\pi^{-1}(i)| \leq d$. We say the projection is "exactly d-to-1" if $R_2 = dR_1$, and $|\pi^{-1}(i)| = d$ for each i.*

We define (exact) d-to-1 Label-Cover to be the special case of Label-Cover in which each projection in Π is (exactly) d-to-1.

In fact, it is known that in Raz's Theorem one can take the Label-Cover instances to be exactly d-to-1; however, the d needs to be at least $\mathrm{poly}(1/\eta)$. Khot's d-to-1 Conjecture is that one can take d to be a constant *independent* of η. Formally, for each integer $d \geq 2$ we have:

Khot's d-to-1 Conjecture: *For every constant $\eta > 0$ there is some constant $k(\eta) < \infty$ such that for d-to-1 Label-Cover instances $\mathcal{L}$ with $R_2 \geq k(\eta)$, it is NP-hard to distinguish the case* $\mathrm{Opt}(\mathcal{L}) = 1$ *from the case* $\mathrm{Opt}(\mathcal{L}) \leq \eta$.

One could also make the "Exact d-to-1 Conjecture", which would be formally stronger than the d-to-1 Conjecture. Such a conjecture is easier to work with, and indeed the paper of Dinur et al. [8] on coloring hardness uses this conjecture instead. These conjectures have a downside, though, which is that it is not clear that the Exact d-to-1 Conjectures actually get weaker as d increases. By contrast, since a d-to-1 projection is by definition also $(d+1)$-to-1, we have that the d-to-1 Conjecture is stronger than the $(d+1)$-to-1 Conjecture for each d. Our results work with the original, weaker d-to-1 Conjecture, for any constant d. However, the difficulty added by working with d-to-1 rather than Exact d-to-1 is entirely notational and not conceptual; hence we strongly encourage the reader to imagine that $R_2 = dR_1$ and that all projections are exactly d-to-1 in the remainder of the work.

Finally, although we don't need it, we mention the Unique Games Conjecture for comparison purposes.

Khot's Unique Games Conjecture: *For every constant $\eta > 0$ there is some constant $k(\eta) < \infty$ such that for Exact 1-to-1 Label-Cover instances $\mathcal{L}$ with $R_1 = R_2 \geq k(\eta)$, it is NP-hard to distinguish the case* $\mathrm{Opt}(\mathcal{L}) \geq 1 - \eta$ *from the case* $\mathrm{Opt}(\mathcal{L}) \leq \eta$.

It is unknown whether the Unique Games Conjecture implies any of the d-to-1 Conjectures, or vice versa.

3.2 PCP System Framework

The high-level framework of our PCP system is similar to Håstad's for Max-3Lin [12]. Given is a d-to-1 Label-Cover instance $\mathcal{L} = (U, V, E, P, R_1, R_2, \Pi)$. A "proof" for $\mathcal{L}$ consists of a collection of truth tables of boolean functions, one for each vertex. More specifically, for each vertex $u \in U$, there is an associated boolean function $f_u : \{-1,1\}^{R_1} \to \{-1,1\}$ and for each vertex $v \in V$, there is an associated boolean function $g_v : \{-1,1\}^{R_2} \to \{-1,1\}$. (As is customary for Fourier analysis, from now on we represent "True" by -1 and "False" by 1.) The proof contains all the truth tables of these boolean functions and the length of it is $|U|2^{R_1} + |V|2^{R_2}$.

Our verifier checks the proof by following polynomial-time procedure:

- Pick an edge $e = (u, v)$ from distribution P.
- Generate a triple (x, y, z) from the distribution $\mathcal{T}_e$ on $\{-1,1\}^{R_1} \times \{-1,1\}^{R_2} \times \{-1,1\}^{R_2}$ (this distribution $\mathcal{T}_e$ will be specified later).
- Accept if $\mathtt{NTW}(f_u(x), g_v(y), g_v(z))$.

Folding. Actually, the above description is not completely accurate. Such a PCP will never work, since the "prover" can write the constantly 1 function for every f_u, g_v and such a proof always passes. To address this, our PCP uses the standard "folding trick" [4]. Note that this means our verifier actually uses all 8 possible NTW predicates, $\mathtt{NTW}(\pm a, \pm b, \pm c)$. The advantage of this trick is that we can assume all the functions h are *odd*, meaning that $h(-z) = -h(z)$ for all inputs z.

For the above PCP system (with $\mathcal{T}_e$ appropriately defined in terms of $\epsilon > 0$), we will show the following:

Completeness: If $\mathrm{Opt}(\mathcal{L}) = 1$, there is a proof which the verifier accepts with probability 1.

Soundness: If there exists a proof passing the verifier's test with probability exceeding $5/8 + \epsilon$, then $\mathrm{Opt}(\mathcal{L}) > \eta$, where $\eta > 0$ is a constant depending only on ϵ and d.

Together, this completeness and soundness gives our main result, Theorem 2.1.

4 The test distribution $\mathcal{T}_e$

Recall that the verifier first picks an edge $e = (u, v)$ in the d-to-1 Label-Cover instance. Then it generates $(x, y, z) \in \{-1,1\}^{R_1} \times \{-1,1\}^{R_2} \times \{-1,1\}^{R_2}$ according to a distribution $\mathcal{T}_e$, and accepts if $\mathtt{NTW}(f_u(x), g_v(y), g_v(z))$, where $f_u : \{-1,1\}^{R_1} \to \{-1,1\}$ and $g_v : \{-1,1\}^{R_2} \to \{-1,1\}$ are the odd functions whose truth tables the prover writes for vertices $u \in U$ and $v \in V$. In this section we will define the distribution $\mathcal{T}_e$.

For the picked edge e, we write $d_i = |\pi_e^{-1}(i)|$ for $i \in [R_1]$. By the d-to-1 projection property we know that $d_i \le d$ for each i. The verifier now views $f_u : \{-1,1\}^{R_1} \to \{-1,1\}$ as a function over an R_1-fold product set,

$$f_u : \mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^{R_1} \to \{-1,1\},$$

where each $\mathcal{X}^i = \{-1,1\}^{\{i\}}$ (a slightly complicated way to write $\{-1,1\}$). More importantly, the verifier also views $g_v : \{-1,1\}^{R_2} \to \{-1,1\}$ as a function over an R_1-fold product set,

$$g_v : \mathcal{Y}^1 \times \mathcal{Y}^2 \times \cdots \times \mathcal{Y}^{R_1} \to \{-1,1\},$$

where each $\mathcal{Y}^i = \{-1,1\}^{\pi_e^{-1}(i)}$. We will also write $\mathcal{Z}^i = \{-1,1\}^{\pi_e^{-1}(i)}$.

We construct $\mathcal{T}_e$ as a product probability distribution over the R_1-fold product set

$$\prod_{i=1}^{R_1} (\mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i) \cong \left(\prod_{i=1}^{R_1} \mathcal{X}^i\right) \times \left(\prod_{i=1}^{R_1} \mathcal{Y}^i\right) \times \left(\prod_{i=1}^{R_1} \mathcal{Z}^i\right), \tag{1}$$

thought of as the product of f_u's domain with two copies of g_v's domain. More specifically, for each i we have a distribution $\mathcal{T}_e^i$ on $\mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i$ and we think of this as a "correlated space" in the sense of Mossel [20], written $(\mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i; \mathcal{T}_e^i)$. (If we use the Exact d-to-1 Conjecture so that all d_i's are equal, the reader may note that all $\mathcal{T}_e^i$'s are the same.) We let $\mathcal{T}_e$ be the product distribution $\bigotimes_{i=1}^{R_1} \mathcal{T}_e^i$ over the domain (1), again thought of as a "correlated space" $(\prod_{i=1}^{R_1} \mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i; \mathcal{T}_e)$.

The distribution $\mathcal{T}_e^i$ on $\mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i$ will only depend on d_i, and further it will be symmetric under (simultaneous) permutations of the coordinates of $\mathcal{Y}^i$, $\mathcal{Z}^i$. We will think of it simply as a distribution on $\{-1,1\} \times \{-1,1\}^{d_i} \times \{-1,1\}^{d_i}$. Furthermore, the different distributions for various d_i will be very related. To define them, we will actually need to define several distributions on $\{-1,1\} \times \{-1,1\}^D \times \{-1,1\}^D$, which we will also write as $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. (Here $0 \le D \le d$.)

The first such distribution is the "Linearity Test" distribution on which Håstad's based his 3Lin Dictator Test:

Definition 4.1. *Define distribution $\mathcal{H}(D)$ generating $(\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_D, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_D) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ as follows: The bit $\boldsymbol{x}$ and the bits $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_D$ are independent and uniformly random; then for each $i \in [D]$, $\boldsymbol{z}_i$ is set to be $-\boldsymbol{x}\boldsymbol{y}_i$.*

(We use **boldface** to denote random variables.) Note that under $\mathcal{H}(D)$, the marginal distribution on $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_D)$ is also uniformly random.

As mentioned, Håstad's 3Lin verifier, which checks $\texttt{XOR}(f_u(\boldsymbol{x}), g_v(\boldsymbol{x}), g_v(\boldsymbol{z}))$, used a "tweaked" version of the distribution $\mathcal{H}(D)$. In standard completeness proofs we have that f_u and g_v are "matching" Dictator (or Long Code) functions, $f_u(x) = x_i$ and $g_v(y) = y_j$ with $\pi_e(j) = i$. In this case, a verifier has perfect completeness if the marginal distribution on each triple $(\boldsymbol{x}, \boldsymbol{y}_i, \boldsymbol{z}_i)$ is in the support of the verifier's predicate, in Håstad's case XOR. In order to prevent large parities from also passing the test with probability 1, Håstad added some noise to the distribution $\mathcal{H}(D)$: specifically, he rerandomized each coordinate $\boldsymbol{z}_i$ with some small probability δ. This meant that the marginal distribution on $(\boldsymbol{x}, \boldsymbol{y}_i, \boldsymbol{z}_i)$ was no longer completely supported on the domain of XOR, leading to a PCP with only near-perfect completeness, $1 - \delta/2$ (albeit with excellent soundness, close to 1/2).

We do not want to give up perfect completeness so we can't rerandomize the $\boldsymbol{z}_i$'s as Håstad does. We *do* have some slack that Håstad doesn't, though: the predicate NTW is also satisfied by the triple $(1, 1, 1)$, in addition to the four triples $(1, 1, -1)$, $(1, -1, 1)$, $(-1, 1, 1)$, $(-1, -1, -1)$ which satisfy XOR. We thus make the following tweak on $\mathcal{H}(D)$ by including $(1, 1, 1)$ as a possible value for $(\boldsymbol{x}, \boldsymbol{y}_i, \boldsymbol{z}_i)$.

Definition 4.2. *Define distribution $\mathcal{N}(D)$ generating $(\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_D, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_D) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ as follows: First, draw from $\mathcal{H}(D)$. Then pick a random integer $k \in [D]$ and set $\boldsymbol{y}_k = \boldsymbol{z}_k = \boldsymbol{x}$. Additionally, for $0 < \delta < 1$, define distribution $\mathcal{H}_\delta(D)$ to be the mixture distribution $\mathcal{H}_\delta(D) = (1 - \delta)\mathcal{H}(D) + \delta\mathcal{N}(D)$; i.e., one draws from $\mathcal{H}(D)$ with probability $1 - \delta$ and from $\mathcal{N}(D)$ with probability δ.*

The distribution $\mathcal{H}_\delta(D)$ is the key distribution for our verifier (note that $\mathcal{H}_\delta(1)$ is the distribution used by the authors in [24]). Among others, it has the following two essential properties: First, any triple $(\boldsymbol{x}, \boldsymbol{y}_i, \boldsymbol{z}_i)$ generated by $\mathcal{H}_\delta(D)$ is in the support of NTW. Second, under $\mathcal{H}_\delta(D)$ (and also under $\mathcal{N}(D)$), the marginal distribution on each of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ is uniform.

We are now ready to define the verifier's distribution $\mathcal{T}_e$:

Definition 4.3. *For each $i \in [R_1]$ we define $\mathcal{T}_e^i$ to be $\mathcal{H}_\delta(d_i)$, with*

$$\delta = (\epsilon/2)^2,$$

where the domain of $\mathcal{H}_\delta(d_i)$ is appropriately identified with the domain $\mathcal{X}^i \times \mathcal{Y}^i \times \mathcal{Z}^i$ of $\mathcal{T}_e^i$. As mentioned, $\mathcal{T}_e$ is the product of these distributions, $\mathcal{T}_e = \bigotimes_{i=1}^{R_1} \mathcal{T}_e^i$.

The expectation of functions under the key distribution $\mathcal{H}_\delta(D)$ is difficult to bound. The major reason for this is that for $D \geq 2$, there is *perfect correlation* between $\mathcal{X}$ and $\mathcal{Y} \times \mathcal{Z}$: given a draw $(\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_D, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_D)$ from $\mathcal{H}_\delta(D)$, the bit $\boldsymbol{x}$ is uniquely determined by $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_D, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_D$. (When $D \geq 3$ the bit $\boldsymbol{x}$ is the majority of the bits $-\boldsymbol{y}_i\boldsymbol{z}_i$; the reader can check that $\boldsymbol{x}$ is still determinable even when $D = 2$.) We mention that when $D = 1$ this correlation is *imperfect*, and this is what made the Invariance Principle-free analysis in [24] easier.

Our goal is to use Invariance Principle techniques to not only break this perfect correlation but drive it down to near 0. To do this we need to pass to a distribution $\mathcal{I}(D)$ with the same "1-wise" and "2-wise" correlations as $\mathcal{H}_\delta(D)$, but with almost no correlation between $\mathcal{X}$ and $\mathcal{Y} \times \mathcal{Z}$:

Definition 4.4. *Define distribution $\mathcal{I}(D)$ generating $(\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_D, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_D) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ as follows: First draw from $\mathcal{H}(D)$; then uniformly rerandomize the bit $\boldsymbol{x}$.*

By definition, $\mathcal{I}(D)$ and $\mathcal{H}(D)$ have the same marginal distribution on $\mathcal{Y} \times \mathcal{Z}$. Also, they have the same marginal distribution on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{X} \times \mathcal{Z}$; namely, uniform. In particular, $\mathcal{I}(D)$

has the uniform marginal distribution on $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. We now give the distribution $\mathcal{I}_\delta(D)$ which is "Invariant" with $\mathcal{H}_\delta(D)$, by adding the same tweak as before:

Definition 4.5. *Define distribution $\mathcal{I}_\delta(D)$ as the mixture distribution $\mathcal{I}_\delta(D) = (1-\delta)\mathcal{I}(D) + \delta\mathcal{N}(D)$.*

It is easy to check that $\mathcal{H}_\delta(D)$ and $\mathcal{I}_\delta(D)$ have the same "1-wise" and "2-wise" correlations; i.e., $\mathcal{H}_\delta(D)$ and $\mathcal{I}_\delta(D)$ have the same marginal distribution on $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $\mathcal{X}\times\mathcal{Y}$, $\mathcal{X}\times\mathcal{Z}$, and $\mathcal{Y}\times\mathcal{Z}$. In particular, their distributions on each of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ is uniform.

Crucially, though, the "3-wise correlations" of $\mathcal{H}_\delta(D)$ and $\mathcal{I}_\delta(D)$ are different; compare Lemma 5.4 to Equation (2) below.

5 Correlations and Influences

5.1 Correlations

We now recall the definition of *correlation* for correlated probability spaces, as introduced by Mossel [20].

Definition 5.1. *Let $(\Omega\times\Psi,\mu)$ be a (finite) correlated probability space, meaning that μ is a distribution on the finite product set $\Omega\times\Psi$ and that the marginals of μ on Ω and Ψ have full support. Define the "correlation" between Ω and Ψ to be*

$$\rho(\Omega,\Psi;\mu) = \max\Big\{ \operatorname*{\mathbf{Cov}}_{(\boldsymbol{\omega},\boldsymbol{\psi})\sim\mu}[f(\boldsymbol{\omega}),g(\boldsymbol{\psi})] \;\Big|\; f:\Omega\to\mathbb{R}, g:\Psi\to\mathbb{R}, \operatorname*{\mathbf{Var}}_{(\boldsymbol{\omega},\boldsymbol{\psi})\sim\mu}[f(\boldsymbol{\omega})] = \operatorname*{\mathbf{Var}}_{(\boldsymbol{\omega},\boldsymbol{\psi})\sim\mu}[g(\boldsymbol{\psi})] = 1\Big\}.$$

It is clear that in the definition of $\rho(\Omega,\Psi;\mu)$, we can equivalently maximize $|\mathbf{E}[fg]|$ over f restricted to have $\mathbf{E}[f]=0$, $\mathbf{E}[f^2]\le 1$ under μ's marginal on Ω; or, over similarly restricted g (or both).

For the distributions defined, we have the following correlation bounds (assuming $D\neq 0$) whose proofs are given in Section D.

Lemma 5.2. $\rho(\mathcal{X},\mathcal{Y};\mathcal{H}_\delta(D)) \le \delta$.

Lemma 5.3. $\rho(\mathcal{X}\times\mathcal{Y},\mathcal{Z};\mathcal{H}_\delta(D)) \le 1 - \frac{\delta^2}{D^2 2^{2D+1}}$.

Lemma 5.4. $\rho(\mathcal{X},\mathcal{Y}\times\mathcal{Z};\mathcal{I}_\delta(D)) \le \sqrt{\delta}$.

Some comments: If we did not tweak by $\mathcal{N}$ the distribution $\mathcal{H}$, we would have that $\rho(\mathcal{X}\times\mathcal{Y},\mathcal{Z};\mathcal{H}(D)) = 1$; this would completely prevent us from using Invariance Principle arguments. Even as it stands, for $\mathcal{H}_\delta$, we still have that

$$\rho(\mathcal{X},\mathcal{Y}\times\mathcal{Z};\mathcal{H}_\delta(D)) = 1 \tag{2}$$

and this causes some trickiness in our analysis. Also, the reader should notice that the correlation between $\mathcal{X}$ and $\mathcal{Y}\times\mathcal{Z}$ is small under $\mathcal{I}_\delta(D)$ as desired, which is not surprising since $\mathcal{I}(D)$ has independent marginals on $\mathcal{X}$ and $\mathcal{Y}\times\mathcal{Z}$.

In [20, Proposition 2.13], Mossel proved that the correlation of a product of correlated probability spaces is equal to the maximum correlation among the individual correlated spaces (excluding empty components). Hence from the above lemmas:

Lemma 5.5.

$$\rho\Big(\prod_{i=1}^{R_1}\mathcal{X}^i, \prod_{i=1}^{R_1}\mathcal{Y}^i; \mathcal{T}_e\Big) \le \delta \tag{3}$$

$$\rho\Big(\prod_{i=1}^{R_1}\mathcal{X}^i\times\mathcal{Y}^i, \prod_{i=1}^{R_1}\mathcal{Z}^i; \mathcal{T}_e\Big) \le 1 - \frac{\delta^2}{d^2 2^{2d+1}} \tag{4}$$

$$\rho\Big(\prod_{i=1}^{R_1}\mathcal{X}^i, \prod_{i=1}^{R_1}\mathcal{Y}^i\times\mathcal{Z}^i; \bigotimes_{i=1}^{R_1}\mathcal{I}_\delta(d_i)\Big) \le \sqrt{\delta} \tag{5}$$

Here we have used the fact that our verifier's overall distribution $\mathcal{T}_e$ is the product of the distributions $\mathcal{T}_e^i \cong \mathcal{H}_\delta(d_i)$, and that $d_i \leq d$ for each i.

5.2 Influences

In this section we recall basic notions from Fourier analysis, *influence* and the *Bonami-Beckner operator*; for more on Fourier analysis see, e.g., [22].

We first define a notion of the influence of a set of coordinates on a function f. Please note that the following definition is *not* standard (except in the case of singleton sets), but is useful for this paper:

Definition 5.6. *For a function $f : \{-1,1\}^n \to \mathbb{R}$ and set of coordinates $S \subseteq [n]$, we define the* influence of S on f *to be*

$$\mathrm{Inf}_S(f) = \sum_{U \supseteq S} \hat{f}(U)^2.$$

In the special case $S = \{i\}$ we write simply $\mathrm{Inf}_i(f)$, and this is the standard notion of the influence of a coordinate.

We next recall the Bonami-Beckner operator T_ρ acting on boolean functions:

Definition 5.7. *Let $0 \leq \rho \leq 1$. The* Bonami-Beckner *operator T_ρ is a linear operator mapping functions $g : \{-1,1\}^n \to \mathbb{R}$ into functions $T_\rho g : \{-1,1\}^n \to \mathbb{R}$ via*

$$(T_\rho g)(x) = \mathbf{E}[g(\boldsymbol{y})],$$

where in the expectation, $\boldsymbol{y}$ is formed from x by setting $\boldsymbol{y}_i = x_i$ with probability ρ and setting $\boldsymbol{y}_i$ to be a uniformly random bit with probability $1 - \rho$.

The operator T_ρ can alternately be defined by the following formula:

Proposition 5.8.

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S.$$

It is well known that for a "smoothed boolean function" (i.e., $T_{1-\gamma} f$, where $f : \{-1,1\}^n \to \{-1,1\}$), the sum of the influences of all coordinates is bounded. We will need a generalization of this, bounding the sum of influences of all constant-size sets:

Lemma 5.9. *For any function $f : \{-1,1\}^n \to \mathbb{R}$ with $\mathbf{E}[f^2] \leq 1$, and any parameters $0 < \gamma \leq 1/2$, $m \in \mathbb{N}$,*

$$\sum_{S \subseteq [n], |S| \leq m} \mathrm{Inf}_S(T_{1-\gamma} f) \leq (m/2\gamma)^m.$$

The proof appears in Section E.

6 Analysis of the verifier

In this section, we describe the completeness and soundness analysis for our verifier. For full details, see Section A.

Completeness Analysis: The completeness analysis is entirely standard; given a perfect labeling L for the d-to-1 Label-Cover instance we can take the "dictator proofs" $f_u = \chi_{L(u)}$, $g_v = \chi_{L(v)}$. This passes the verifier's test with probability 1, using the fact that all triples $(\boldsymbol{x}, \boldsymbol{y}_j, \boldsymbol{z}_j)$ generated by $\mathcal{H}_\delta(\delta)$ are in the support of `NTW`.

Soundness Analysis: This is the focus of the remainder of the paper. Our task is to show that for a given d-to-1 Label-Cover input $\mathcal{L}$, if $(f_u)_{u\in U}$ and $(g_v)_{v\in V}$ is any collection of odd functions causing the verifier to accept with probability more than $5/8+\epsilon$, then we must have $\mathrm{Opt}(\mathcal{L}) > \eta$, where $\eta > 0$ is a constant depending only on ϵ and d.

As usual, let us first arithmetize the probability a given proof passes. We have

$$\mathbf{Pr}[\text{verifier accepts}] = \mathop{\mathbf{Pr}}_{\substack{e=(u,v)\sim P\\ \mathcal{T}_e}} \mathbf{Pr}[\mathtt{NTW}(f_u(\boldsymbol{x}), g_v(\boldsymbol{y}), g_v(\boldsymbol{z}))] =$$

$$\mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P\\ \mathcal{T}_e}}[\tfrac{5}{8}+\tfrac{1}{8}(f_u(\boldsymbol{x})+g_v(\boldsymbol{y})+g_v(\boldsymbol{z}))+\tfrac{1}{8}(f_u(\boldsymbol{x})g_v(\boldsymbol{y})+f_u(\boldsymbol{x})g_v(\boldsymbol{z})+g_v(\boldsymbol{y})g_v(\boldsymbol{z}))-\tfrac{3}{8}f_u(\boldsymbol{x})g_v(\boldsymbol{y})g_v(\boldsymbol{z})].$$

Since f_u and g_v are odd for every (u,v), and since $\mathcal{T}_e$'s marginal distribution on each of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ is uniform, we conclude

$$\mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P\\ \mathcal{T}_e}}[\tfrac{1}{8}(f_u(\boldsymbol{x}) + g_v(\boldsymbol{y}) + g_v(\boldsymbol{z}))] = 0.$$

The next two terms are also straightforward to handle (recall that $\delta = (\epsilon/2)^2$):

Proposition 6.1. *For any $e = (u,v)$,*

$$\mathop{\mathbf{E}}_{\mathcal{T}_e}[\tfrac{1}{8}(f_u(\boldsymbol{x})g_v(\boldsymbol{y}) + f_u(\boldsymbol{x})g_v(\boldsymbol{z}))] \le \delta/4.$$

Proof. The joint distribution on $(\boldsymbol{x},\boldsymbol{y})$ and $(\boldsymbol{x},\boldsymbol{z})$ is the same, so it suffices to show

$$\mathop{\mathbf{E}}_{\mathcal{T}_e}[f_u(\boldsymbol{x})g_v(\boldsymbol{y})] \le \delta.$$

But this follows immediately from Lemma 5.5.(3) because f_u and g_v both have mean 0 and second moment 1, being odd boolean functions. □

So far we have shown

$$\mathbf{Pr}[\text{verifier acc.}] \le \tfrac{5}{8} + \tfrac{\delta}{4} + \tfrac{1}{8}\mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P\\ \mathcal{T}_e}}[g_v(\boldsymbol{y})g_v(\boldsymbol{z})] - \tfrac{3}{8}\mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P\\ \mathcal{T}_e}}[f_u(\boldsymbol{x})g_v(\boldsymbol{y})g_v(\boldsymbol{z})]. \tag{6}$$

The main effort goes into bounding the remaining two terms, especially the last one. We will prove the following theorems:

Theorem 6.2. *For any $e = (u,v)$, the fact that $g_v : \{-1,1\}^{R_2} \to \{-1,1\}$ is odd implies*

$$\mathop{\mathbf{E}}_{\mathcal{T}_e}[g_v(\boldsymbol{y})g_v(\boldsymbol{z})] \le \delta.$$

Theorem 6.3. *There exist constants $\gamma, \tau > 0$ depending only on d, δ such that the following holds. If for every $i \in [R_1]$ and every odd-cardinality set $S \subseteq \pi_e^{-1}(i)$ we have*

$$\min\{\mathrm{Inf}_i(T_{1-\gamma/2}f_u), \mathrm{Inf}_S(T_{1-\gamma/2}g_v)\} \le \tau, \tag{7}$$

then

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}_e}[f_u(\boldsymbol{x})g_v(\boldsymbol{y})g_v(\boldsymbol{z})]\right| \le 3\sqrt{\delta}.$$

Theorem 6.2 is proved in Section B using a novel argument about the positivity of certain linear operators. Theorem 6.3 is proved in Section C using an Invariance Principle-type proof.

With these theorems in hand we can conclude the proof of the verifier's soundness by familiar means. We give a randomized labeling for $\mathcal{L}$ based on "list-decoding" each vertex $w \in U \cup V$ to a short list of labels. We decode each $u \in U$ to the set of coordinates with influence at least τ on $T_{1-\gamma/2}f_u$. We decode each $v \in V$ to the union of all sets $S \subseteq [R_2]$ of odd cardinality at most d which have influence at least τ on $T_{1-\gamma/2}g_v$ (note that this decoding does not depend on the projections Π). By Lemma 5.9, these label lists are of bounded size. Then Theorem 6.3 ensures that if the last expression in (6) is nonnegligible, there must be a nonnegligible fraction of edges (u,v) for which (7) fails; by construction, the randomized labeling will satisfy a nonnegligible fraction of these edges in expectation. For full details, see Section A.

7 Discussion

Assuming Khot's d-to-1 Conjecture holds for some finite constant d, we have shown a tight PCP characterization of NP using 3 nonadaptive queries; more precisely, we have shown Zwick's Conjecture and settled Håstad's open problem about the inapproximability of satisfiable Max-NTW beyond the random assignment threshold. The methods in this paper illustrate how to analyze complicated d-to-1-based inner verifier distributions without pairwise independence. We hope these techniques will help in settling the approximability of other satisfiable CSPs, most notably satisfiable Max-NAE (Not-All-Equal).

An open technical question is whether the tradeoff we use between d and η in the d-to-1 Conjecture can be improved. Tracing through our proof reveals that we need $\eta = \exp(-2^{O(d^2)}/\epsilon^{O(d)})$ soundness for d-to-1 Label-Cover to achieve $5/8 + \epsilon$ hardness for Max-NTW. We did not put any effort into optimizing this dependence. It would be interesting to see if the doubly-exponential dependence of η on d could be improved, since Raz's Theorem is that the d-to-1 Conjecture is *true* if η needs only be $1/d^{\Omega(1)}$.

References

[1] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[2] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[3] P. Austrin. Balanced Max-2Sat might not be hardest. In *Proc. 39th ACM Symp. on the Theory of Computing*, pages 189–197, 2007.

[4] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and nonapproximability — towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[5] A. Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. *Ann. Inst. Fourier*, 18(2):193–204, 1968.

[6] N. Creignou. A dichotomy theorem for maximum generalized satisfiability problems. *Journal of Computing and Sys. Sci.*, 51(3):511–522, 1995.

[7] N. Creignou, S. Khanna, and M. Sudan. *Complexity classifications of boolean constraint satisfaction problems*. SIAM, Philadelphia, PA, 2001.

[8] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. In *Proc. 38th ACM Symp. on the Theory of Computing*, pages 344–353, 2006.

[9] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.

[10] V. Guruswami and S. Khot. Hardness of max 3SAT with no mixed clauses. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 154–162, 2005.

[11] V. Guruswami, D. Lewin, M. Sudan, and L. Trevisan. A tight characterization of NP with 3 query PCPs. *Electronic Colloq. on Comp. Complexity (ECCC)*, 5(34), 1998.

[12] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[13] J. Håstad and S. Khot. Query efficient PCPs with perfect completeness. *Theory of Computing*, 1(1):119–148, 2005.

[14] H. Karloff and U. Zwick. A 7/8-approximation algorithm for Max-3Sat? In *Proc. 38th IEEE Symp. on Foundations of Comp. Sci.*, pages 406–415, 1997.

[15] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on the Theory of Computing*, pages 767–775, 2002.

[16] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[17] S. Khot and R. O'Donnell. SDP gaps and UGC-hardness for MaxCutGain. In *Proc. 47th IEEE Symp. on Foundations of Comp. Sci.*, pages 217–226, 2006.

[18] S. Khot and O. Regev. Vertex Cover might be hard to approximate to within $2-\epsilon$. *Journal of Computing and Sys. Sci.*, 74(3):335–349, 2008.

[19] S. Khot and R. Saket. A 3-query non-adaptive PCP with perfect completeness. In *Proc. 21st IEEE Conference on Computational Complexity*, pages 159–169, 2006.

[20] E. Mossel. Gaussian bounds for noise correlation of functions. In *Proc. 46th IEEE Symp. on Foundations of Comp. Sci.*, 2008.

[21] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symp. on Foundations of Comp. Sci.*, pages 21–30, 2005. To appear, *Annals of Mathematics*.

[22] R. O'Donnell. Analysis of boolean functions lecture notes, 2007. http://www.cs.cmu.edu/ odonnell/boolean-analysis/.

[23] R. O'Donnell and Y. Wu. An optimal SDP algorithm for Max-Cut, and equally optimal Long Code tests. In *Proc. 40th ACM Symp. on the Theory of Computing*, pages 335–344, 2008.

[24] R. O'Donnell and Y. Wu. 3-bit dictator testing: 1 vs. 5/8, 2009. To appear in Proc. 20th ACM-SIAM Symp. on Discrete Algorithms.

[25] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc. 40th ACM Symp. on the Theory of Computing*, pages 245–254, 2008.

[26] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[27] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proc. 32nd ACM Symp. on the Theory of Computing*, pages 191–199, 2000.

[28] L. Trevisan, G. Sorkin, M. Sudan, and D. Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.

[29] U. Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proc. 9th ACM-SIAM Symp. on Discrete Algorithms*, pages 201–210, 1998.

[30] U. Zwick. Computer assisted proof of optimal approximability results. In *Proc. 13th ACM-SIAM Symp. on Discrete Algorithms*, pages 496–505, 2002.

Appendix

A Full completeness and soundness proofs

In this section we give the full details of the completeness and soundness of our verifier, modulo the analysis of the key quadratic and cubic terms.

We first amplify slightly on the completeness:

Completeness Analysis: Let $\mathcal{L}$ be a given d-to-1 Label-Cover input $\mathcal{L}$ with $\mathrm{val}(\mathcal{L}) = 1$; say $L : U \to R_1, V \to R_2$ is a perfect labeling. For each $u \in U$ the prover can take f_u to be the $L(u)$th dictator function, $\chi_{\{L(u)\}}$, and for each $v \in V$ can take g_v to be $L(v)$th dictator function $\chi_{\{L(v)\}}$. Note that these are odd functions and hence are unaffected by the folding. Now for any edge $e = (u, v)$ we have that $\pi_e(L(v)) = L(u)$; i.e., $L(v) \in \pi_e^{-1}(L(u))$. It follows from the definition of $\mathcal{T}_e$ that the relevant bits for f_u and g_v are generated solely from the distribution $\mathcal{T}_e^{L(u)}$, and each triple $(\boldsymbol{x}_{L(u)}, \boldsymbol{y}_{L(v)}, \boldsymbol{z}_{L(v)})$ it generates is in the support of `NTW`. Hence the proof passes with probability 1.

Next, we show soundness:

Soundness Analysis:

Proof. Supposing that the functions f_u and g_v cause the verifier to accept with probability exceeding $5/8 + \epsilon = 5/8 + 2\sqrt{\delta}$, we conclude from (6) and Theorem 6.2 that

$$\tfrac{5}{8} + 2\sqrt{\delta} < \tfrac{5}{8} + \tfrac{\delta}{4} + \tfrac{\delta}{8} - \tfrac{3}{8} \mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P \\ \mathcal{I}_e}} [f_u(\boldsymbol{x}) g_v(\boldsymbol{y}) g_v(\boldsymbol{z})],$$

which implies

$$\left| \mathop{\mathbf{E}}_{\substack{e=(u,v)\sim P \\ \mathcal{I}_e}} [f_u(\boldsymbol{x}) g_v(\boldsymbol{y}) g_v(\boldsymbol{z})] \right| > \tfrac{16}{3}\sqrt{\delta} - \delta > 4\sqrt{\delta}.$$

By an averaging argument, this implies that for at least a $\sqrt{\delta}$ fraction of the edges $e = (u, v)$ (under distribution P) we have

$$\left| \mathop{\mathbf{E}}_{\mathcal{I}_e} [f_u(\boldsymbol{x}) g_v(\boldsymbol{y}) g_v(\boldsymbol{z})] \right| > 3\sqrt{\delta}.$$

We call such edges "good". By Theorem 6.3, we know for every good edge there must exist some $i_e \in [R_1]$, and odd-cardinality set $S_e \subseteq \pi_e^{-1}(i_e)$ such that

$$\min\{\mathrm{Inf}_{i_e}(T_{1-\gamma/2} f_u), \mathrm{Inf}_{S_e}(T_{1-\gamma/2} g_v)\} \geq \tau. \tag{8}$$

As usual, we construct a randomized labeling strategy for $\mathcal{L}$. For each $u \in U$ we define

$$L_u = \{i \in [R_1] : \mathrm{Inf}_i(T_{1-\gamma/2} f_u) \geq \tau\}.$$

For each $v \in V$ we define

$$L_v = \{j \in [R_2] : j \in S, \mathrm{Inf}_S(T_{1-\gamma/2} g_v) \geq \tau, |S| \leq d, |S| \text{ is odd}\}.$$

Note that the definition of L_v does not depend at all on the projections $\pi_e \in \Pi$.

By Lemma 5.9 we know that for any of the ± 1-valued functions g_v we have $\sum_{|S| \leq d} \mathrm{Inf}_S(T_{1-\gamma/2} g_v) \leq (d/\gamma)^d$. Therefore at most $(d/\gamma)^d/\tau$ sets S can contribute in the definition of L_v and we conclude

$$|L_v| \leq d \cdot (d/\gamma)^d/\tau \qquad \text{for each } v \in V.$$

Similarly, we have

$$|L_u| \leq 1/(\gamma\tau) \qquad \text{for each } u \in U.$$

Next, by (8), whenever $e = (u, v)$ is a good edge we have that $i_e \in L_u$ and that S_e contributes to L_v. Since $|S_e|$ is odd, it is nonempty; hence we conclude there exists some $j_e \in L_v$ and $i_e \in L_u$ with $\pi_e(j_e) = i_e$.

Finally, we define the randomize labeling for $\mathcal{L}$ by choosing a random label from L_w for each $w \in U \cup V$ (if L_w is empty we can select an arbitrary label). We know that for every good edge $e = (u, v)$, the randomized labeling has at least a

$$\frac{1}{|L_u||L_v|} \geq \tau^2(\gamma/d)^{d+1}$$

chance of choosing i_e for u and j_e for v, thus satisfying e in $\mathcal{L}$. Since the good edges constitute more than a $\sqrt{\delta} = (\epsilon/2)$ fraction of edges under P, we conclude that the expected P-fraction of edge weight in $\mathcal{L}$ satisfied by the randomized labeling exceeds

$$\eta = (\epsilon/2)\tau^2(\gamma/d)^{d+1}.$$

Hence $\mathrm{Opt}(\mathcal{L}) > \eta$, a positive constant depending only on d and ϵ (since $\gamma, \tau > 0$ depend only on d and $\delta = (\epsilon/2)^2$), as desired. □

B Analyzing $\mathbf{E}[g_v(\boldsymbol{y})g_v(\boldsymbol{z})]$

This section is devoted to the proof of Theorem 6.2, which we repeat here for convenience:

Theorem 6.2 *For any $e = (u, v)$, the fact that $g_v : \{-1, 1\}^{R_2} \to \{-1, 1\}$ is odd implies*

$$\mathop{\mathbf{E}}_{\mathcal{T}_e}[g(\boldsymbol{y})g(\boldsymbol{z})] \leq \delta.$$

Recall that $\mathcal{T}_e = \bigotimes_{i=1}^{R_1} \mathcal{T}_e^i$, where $\mathcal{T}_e^i = \mathcal{H}_\delta(d_i)$ with $\mathcal{H}_\delta(d_i)$'s domain $\{-1, 1\} \times \{-1, 1\}^{d_i} \times \{-1, 1\}^{d_i}$ identified with $\mathcal{T}_e^i$'s domain $\mathcal{X}^i \times \mathcal{Y}^{\pi_e^{-1}(i)} \times \mathcal{Z}^{\pi_e^{-1}(i)}$, and $d_i = |\pi_e^{-1}(i)|$. Let us make notational simplifications. Clearly only the marginals on the $\mathcal{Y}$ and $\mathcal{Z}$ domains are relevant for the theorem; in this section we henceforth identity distributions with these marginals. Next, note that the statement of the theorem is invariant under permuting g's coordinates simultaneously with the distributions $\mathcal{N}_e^i$, using the permutation-invariance of the distributions $\mathcal{H}_\delta(d_i)$. Thus we can assume without loss of generality that $\pi_e^{-1}(1) = \{1, \ldots, d_1\}$, $\pi_e^{-1}(2) = \{d_1 + 1, \ldots, d_1 + d_2\}$, etc.; this lets us write, simply,

$$\mathcal{T}_e = \mathcal{H}_\delta(d_1) \otimes \mathcal{H}_\delta(d_2) \otimes \cdots \otimes \mathcal{H}_\delta(d_{R_1})$$

without worrying about coordinate positioning. Note also that coordinates i with $d_i = 0$ are irrelevant for the theorem, so we may assume without loss of generality that there are none. Finally, for the remainder of this section we consider $e = (u, v)$ to be fixed and abbreviate $\mathcal{T} = \mathcal{T}_e$, $\pi = \pi_e$, $g = g_v$.

The distribution $\mathcal{H}(D)$ on $\{-1, 1\}^D \times \{-1, 1\}^D$ simply generates $(\boldsymbol{y}, \boldsymbol{z})$ where $\boldsymbol{y}$ is uniformly random and $\boldsymbol{z} = \pm\boldsymbol{y}$ randomly. It is easy to see that the correlation $\rho(\{-1, 1\}^D \times \{-1, 1\}^D; \mathcal{H}(D)) = 1$, as $\mathbf{E}[\chi_S(\boldsymbol{y})\chi_S(\boldsymbol{z})] = 1$ for any set S with $|S|$ even. The reader can further check that the correlation $\rho(\{-1, 1\}^D, \{-1, 1\}^D; \mathcal{H}_\delta(D)) = 1 - \delta$; hence we cannot prove Theorem 6.2 in the same simple way we bounded the $\mathbf{E}[f_u g_v]$ terms in Proposition 6.1.

The key is to exploit the fact that the correlation between *odd*-cardinality characters under $\mathcal{H}(D)$ is 0, and is at most δ under $\mathcal{H}_\delta(D)$. Since our g is odd, each term in its Fourier expansion must have odd-size intersection with at least one block $\pi^{-1}(i)$; this is what ultimately lets us prove Theorem 6.2.

B.1 Matrix notation

For the proof, it will be necessary to change notation. Instead of looking at correlations of functions under a distribution, we will look at bilinear forms with matrix-vector notation. Let us define the matrix form of a distribution on $\{-1, 1\}^D \times \{-1, 1\}^D$ with respect to the Fourier basis.

Definition B.1. *Suppose $\mathcal{P}$ is a distribution on $\{-1, 1\}^D \times \{-1, 1\}^D$. The associated matrix form $M(\mathcal{P})$ is a $2^D \times 2^D$ matrix defined as follows: The rows and columns of $M(\mathcal{P})$ are indexed by subsets $S, T \subseteq [D]$. The (S, T) entry is defined by*

$$M(\mathcal{P})_{S,T} = \mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{P}}[\chi_S(\boldsymbol{y})\chi_T(\boldsymbol{z})].$$

As an example, the following is easy to check:

Proposition B.2. *For the distribution $\mathcal{H}(D)$ on $\{-1, 1\}^D \times \{-1, 1\}^D$, the matrix $M(\mathcal{H}(D))$ is a diagonal matrix with (S, S) entry equal to 0 if $|S|$ is odd and equal to 1 is $|S|$ is even.*

Henceforth in this section we will also identify functions $h : \{-1, 1\}^D \to \mathbb{R}$ with column vectors of length 2^D, with entries indexed by the subsets $S \subseteq [D]$ in the same order as in Definition B.1. The S-entry of vector h will be $\hat{h}(S)$. With this notation we have the fundamental relation

$$\mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{P}}[h_1(\boldsymbol{y})h_2(\boldsymbol{z})] = h_1^\top M(\mathcal{P})h_2.$$

The following lemma is easy to check.

Lemma B.3.

1. Suppose $\mathcal{P}_1$ and $\mathcal{P}_2$ are two distributions on $\{-1,1\}^D \times \{-1,1\}^D$, and $\mathcal{P} = c\mathcal{P}_1 + (1-c)\mathcal{P}_2$ is a mixture of the two distributions. Then $M(\mathcal{P}) = cM(\mathcal{P}_1) + (1-c)M(\mathcal{P}_2)$.

2. Suppose $\mathcal{P}_i$ is a distribution on $\{-1,1\}^{D_i} \times \{-1,1\}^{D_i}$, $i = 1,2$. Let $\mathcal{P}$ be the product distribution $\mathcal{P} = \mathcal{P}_1 \otimes \mathcal{P}_2$. Then $M(\mathcal{P})$ is the Kronecker product $M(\mathcal{P}) = M(\mathcal{P}_1) \otimes M(\mathcal{P}_2)$ (with the natural identification of indices $(S_1, S_2) \leftrightarrow S_1 \cup S_2$).

B.2 The proof of Theorem 6.2

Now we are ready to prove Theorem 6.2. Using our new notation, we have

$$\mathop{\mathbf{E}}_{\mathcal{T}}[g(\boldsymbol{y})g(\boldsymbol{z})] = g^\top M(\mathcal{T}_e)g = g^\top \left(\bigotimes_{i=1}^{R_1} M(\mathcal{H}_\delta(d_i)) \right) g.$$

Let us introduce the distribution $\mathcal{E}(D)$ on $\{-1,1\}^D \times \{-1,1\}^D$ which generates pairs $(\boldsymbol{y}, \boldsymbol{z})$ by choosing $\boldsymbol{y} \in \{-1,1\}^D$ uniformly at random and setting $\boldsymbol{z} = \boldsymbol{y}$. It is easy to check that $M(\mathcal{E}(D))$ is the *identity matrix*. Further introduce distribution

$$\mathcal{E}_\delta(D) = (1-\delta)\mathcal{H}(D) + \delta\mathcal{E}(D).$$

The proof of Theorem 6.2 is now an immediate consequence of the following two lemmas:

Lemma B.4.

$$g^\top \left(\bigotimes_{i=1}^{R_1} M(\mathcal{H}_\delta(d_i)) \right) g \le g^\top \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\delta(d_i)) \right) g.$$

Lemma B.5.

$$g^\top \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\delta(d_i)) \right) g \le \delta.$$

The first lemma is by linear algebra:

Proof. (Lemma B.4) It suffices to show that the matrix

$$\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\delta(d_i)) - \bigotimes_{i=1}^{R_1} M(\mathcal{H}_\delta(d_i)) \tag{9}$$

is positive semidefinite. We will show that for each $D \ge 1$ the matrices $M(\mathcal{E}_\delta(D)) - M(\mathcal{H}_\delta(D))$ and $M(\mathcal{E}_\delta(D)) + M(\mathcal{H}_\delta(D))$ are both positive semidefinite. This implies that the matrix in (9) is indeed positive semidefinite, by the basic matrix algebra result Lemma F.1 proved in Section F.

For notational simplicity, we henceforth omit showing the dependence on D. Since $M(\mathcal{E}_\delta) - M(\mathcal{H}_\delta) = \delta(M(\mathcal{E}) - M(\mathcal{N}))$, to show that $M(\mathcal{E}_\delta) - M(\mathcal{H}_\delta)$ is positive semidefinite we only need to show it for $M(\mathcal{E}) - M(\mathcal{N})$. For any $h : \{-1,1\}^D \to \mathbb{R}$ we have

$$h^\top M(\mathcal{N})h = \mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[h(\boldsymbol{y})h(\boldsymbol{z})] \le \sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[h(\boldsymbol{y})^2]} \sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[h(\boldsymbol{z})^2]}$$

by Cauchy-Schwarz. But note that the marginals of $\mathcal{N}$ on $\{-1,1\}^D \times \{-1,1\}^D$ are both uniform, and the same is true of $\mathcal{E}$. Hence

$$\sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[h(\boldsymbol{y})^2]} \sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[h(\boldsymbol{z})^2]} = \mathbf{E}[h^2] = \mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{E}}[h(\boldsymbol{y})h(\boldsymbol{z})] = h^\top M(\mathcal{E})h.$$

So we've established $h^\top M(\mathcal{N})h \le h^\top M(\mathcal{E})h$ for all h, and hence $M(\mathcal{E}) - M(\mathcal{N})$ is positive semidefinite as needed.

As for the matrix $M(\mathcal{E}_\delta)+M(\mathcal{H}_\delta)$, by definition it equals $2(1-\delta)M(\mathcal{H})+\delta(M(\mathcal{E})+M(\mathcal{N}))$. From Proposition B.2 we have that $M(\mathcal{H})$ is diagonal with only nonnegative numbers on the diagonal; hence it is positive semidefinite. Thus to show $M(\mathcal{E}_\delta)+M(\mathcal{H}_\delta)$ is positive semidefinite, it remains to show that $M(\mathcal{E})+M(\mathcal{N})$ is. But the proof for this is essentially identical to the above proof for $M(\mathcal{E})-M(\mathcal{N})$: we only need to start with

$$h^\top(-M(\mathcal{N}))h = \mathop{\mathbf{E}}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathcal{N}}[-h(\boldsymbol{y})h(\boldsymbol{z})]$$

and the minus sign disappears in the subsequent application of Cauchy-Schwarz. □

We prove Lemma B.5 via Fourier analysis:

Proof. (Lemma B.5) By Proposition B.2 and the fact that $M(\mathcal{E}(D))$ is the identity matrix, we have that $M(\mathcal{E}_\delta(D))$ is a diagonal matrix whose (S,S) entry is equal to 1 if $|S|$ is even and equal to δ if $|S|$ is odd. Unraveling definitions, it follows that

$$g^\top\left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\delta(d_i))\right)g = \sum_{S\subseteq[R_2]} \hat{g}(S)^2\cdot\delta^{\#\{i\in[R_1]:|S\cap\pi^{-1}(i)|\text{ is odd}\}}. \tag{10}$$

But g is an odd function, and therefore $\hat{g}(S)^2$ is nonzero only if $|S|$ is odd. But $|S|$ being odd implies that $|S\cap\pi^{-1}(i)|$ is odd for at least one i, and hence (10) is upper-bounded by

$$\sum_{S\subseteq[R_2]}\hat{g}(S)^2\cdot\delta = \mathbf{E}[g^2]\cdot\delta = \delta.$$

□

C Analyzing $\mathbf{E}[f_u(\boldsymbol{x})g_v(\boldsymbol{y})g_v(\boldsymbol{z})]$

This section is devoted to the proof of Theorem 6.3, which we repeat here for convenience.

Theorem 6.3 *There exist constants $\gamma,\tau>0$ depending only on d,δ such that the following holds. If for every $i\in[R_1]$ and every odd-cardinality set $S\subseteq\pi_e^{-1}(i)$ we have*

$$\min\{\mathrm{Inf}_i(T_{1-\gamma/2}f_u),\mathrm{Inf}_S(T_{1-\gamma/2}g_v)\}\le\tau, \tag{11}$$

then

$$\left|\mathop{\mathbf{E}}_{\mathcal{I}_e}[f_u(\boldsymbol{x})g_v(\boldsymbol{y})g_v(\boldsymbol{z})]\right|\le 3\sqrt{\delta}. \tag{12}$$

Outline of the proof: We make all of the same notational simplifications as at the beginning of Section B (except we don't drop the components with $d_i=0$); note that the notions of influence are also invariant under permutations of the coordinates. In particular we write $f=f_u$ and $g=g_v$ and retain that these are odd functions satisfying condition (11). Let us also write

$$\mathcal{H}_\delta := \mathcal{T} = \mathcal{H}_\delta(d_1)\otimes\mathcal{H}_\delta(d_2)\otimes\cdots\otimes\mathcal{H}_\delta(d_{R_1}),$$

and introduce the distribution

$$\mathcal{I}_\delta := \mathcal{I}_\delta(d_1)\otimes\mathcal{I}_\delta(d_2)\otimes\cdots\otimes\mathcal{I}_\delta(d_{R_1}).$$

The overall idea of the proof is to use Invariance Principle-type arguments to show that

$$\mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})]\approx\mathop{\mathbf{E}}_{\mathcal{I}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})].$$

We would then use the fact (Lemma 5.5.5)) that $\rho(\{-1,1\}^{R_1}, \{-1,1\}^{R_2} \times \{-1,1\}^{R_2}; \mathcal{I}_\delta) \leq \sqrt{\delta}$ to conclude that

$$\left| \mathop{\mathbf{E}}_{\mathcal{I}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})] \right| \leq \delta.$$

This conclusion is by definition of ρ, because f has mean zero and second-moment 1 (being an odd boolean functions) and because $g(\boldsymbol{y})g(\boldsymbol{z})$ has variance at most 1 (having range $\{-1,1\}$).

The above outline is a bit imprecise; at a more formal level, we need to break down the Invariance argument into two steps:

Theorem C.1. *There are positive constants $\gamma' \geq \gamma > 0$ (depending only on δ and d) such that*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \leq \sqrt{\delta}.$$

Theorem C.2. *Assuming $\tau > 0$ is small enough as a function of γ, δ, and d,*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \leq \sqrt{\delta}.$$

We also have

$$\left| \mathop{\mathbf{E}}_{\mathcal{I}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \leq \sqrt{\delta}. \tag{13}$$

This is again by $\rho(\{-1,1\}^{R_1}, \{-1,1\}^{R_2} \times \{-1,1\}^{R_2}; \mathcal{I}_\delta) \leq \sqrt{\delta}$, since $T_{1-\gamma}f$ has mean zero and second-moment at most 1, and because $T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})$ has variance at most 1, having range $[-1,1]$. Combining (13) with Theorems C.1 and C.2 yields the bound (12), completing the proof of the main Theorem 6.3.

In the remainder of this section we give the proofs of Theorems C.1 and C.2.

C.1 Proof of Theorem C.1

The proof of Theorem C.1 is quite similar to the proof of Mossel's Lemma 6.2 from [20]. However we cannot use that result directly for two reasons.

First, we do not have $\rho(\{-1,1\}^{R_1}, \{-1,1\}^{R_2}, \{-1,1\}^{R_2}; \mathcal{H}_\delta) < 1$ in the sense of Mossel; in fact, we have $\rho(\{-1,1\}^{R_1}, \{-1,1\}^{R_2} \times \{-1,1\}^{R_2}; \mathcal{H}_\delta) = 1$, as discussed in Section 4 directly after the definition of $\mathcal{T}_\epsilon$. It turns out we can evade this difficulty just by relying on the fact that $\rho(\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}, \times\{-1,1\}^{R_2}; \mathcal{H}_\delta) < 1$ (see Lemma 5.5.(4)).

The second reason we cannot use Mossel's Lemma 6.2 directly is that we need to keep a careful distinction between the usual boolean Bonami-Beckner operator T_ρ and Mossel's more general Bonami-Beckner operator.

We now proceed with the proof of Theorem C.1. We assume some familiarity with the Section 2 of Mossel's work [20], including the *Efron-Stein* decomposition of functions on product probability spaces.

Proof. As in Mossel's Lemma 6.2 we insert the Bonami-Beckner operators one-by-one. The proof is composed of the following three lemmas:

Lemma C.3. *By taking $\gamma' > 0$ small enough as a function of δ and d we ensure*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \leq \sqrt{\delta}/3. \tag{14}$$

Lemma C.4. *We also have*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \leq \sqrt{\delta}/3. \tag{15}$$

Lemma C.5. *By taking $\gamma > 0$ small enough as a function of γ', δ, and d we ensure*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \le \sqrt{\delta}/3. \tag{16}$$

We now prove the three lemmas, beginning with Lemma C.3.

Proof. (Lemma C.3) We have

$$\begin{aligned} \left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| &= \left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y}) \cdot (id - T_{1-\gamma'})g(\boldsymbol{z})] \right| \\ &= \left| \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{H}_\delta}[f(\boldsymbol{x})g(\boldsymbol{y}) \cdot U(id - T_{1-\gamma'})g(\boldsymbol{x},\boldsymbol{y})] \right|, \end{aligned} \tag{17}$$

where U denotes the *conditional expectation operator* (written "T" in [20]) for the correlated probability space $((\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}) \times (\{-1,1\}^{R_2}), \mathcal{H}_\delta)$, mapping functions on the latter space $\{-1,1\}^{R_2}$ into functions on the former space $\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}$.

We now consider the quantity inside the expectation in (17) to be a product of two functions on $\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}$, namely $F = fg$ and $G = U(id - T_{1-\gamma'})g$. We take the Efron-Stein decomposition of these two functions with respect to the (product) distribution $\mathcal{H}_\delta$ on $\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}$. Then by orthogonality of the Efron-Stein decomposition and Cauchy-Schwarz,

$$(17) = \left| \sum_{S\subseteq[R_1]} \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{H}_\delta}[F_S(\boldsymbol{x},\boldsymbol{y}) \cdot G_S(\boldsymbol{x},\boldsymbol{y})] \right| \le \sqrt{\sum_{S\subseteq[R_1]} \|F_S\|_2^2} \sqrt{\sum_{S\subseteq[R_1]} \|G_S\|_2^2}, \tag{18}$$

where the 2-norms $\|\cdot\|_2$ are with respect to $\mathcal{H}_\delta$'s marginal on $\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}$. By orthogonality again, the quantity $\sum_{S\subseteq[R_1]} \|F_S\|_2^2$ is just $\|F\|_2^2$, which is precisely 1 because F's range is $\{-1,1\}$. Hence we have

$$(18) \le \sqrt{\sum_{S\subseteq[R_1]} \|G_S\|_2^2}. \tag{19}$$

As Mossel shows, the conditional expectation operator U commutes with taking the Efron-Stein decomposition; i.e., $G_S = UG'_S$, where $G' = (id - T_{1-\gamma'})g$. Here the Efron-Stein decomposition is with respect to $\mathcal{H}_\delta$'s marginal distribution on the $\mathcal{Z}$-space $\{-1,1\}^{R_2}$, viz., the uniform distribution. It is also easy to check that this Efron-Stein decomposition of g has

$$g_S = \sum_{U\subseteq[R_2]:\pi(U)=S} \hat{g}(U)\chi_U.$$

It follows that applying the Bonami-Beckner operator $T_{1-\gamma'}$ to g also commutes with taking the Efron-Stein decomposition (technically, this uses $\gamma' < 1$). Hence we have $G_S = UG'_S = U(id - T_{1-\gamma'})g_S$. Substituting this into (19) yields

$$(19) = \sqrt{\sum_{S\subseteq[R_1]} \|U(id - T_{1-\gamma'})g_S\|_2^2}. \tag{20}$$

Recall from Lemma 5.5.(4) that the correlation

$$\rho((\{-1,1\}^{R_1} \times \{-1,1\}^{R_2}) \times (\{-1,1\}^{R_2}), \mathcal{H}_\delta) \le \rho_0 := 1 - \frac{\delta^2}{d^2 2^{2d+1}};$$

using this in Mossel's Proposition 2.12 we conclude that for each $S \subseteq [R_1]$,

$$\|U(id - T_{1-\gamma'})g_S\|_2 \le \rho_0^{|S|} \|(id - T_{1-\gamma'})g_S\|_2, \tag{21}$$

where the 2-norm on the right is with respect to the uniform distribution on $\{-1,1\}^{R_2}$.

Next, observe that

$$\begin{aligned}\|(id - T_{1-\gamma'})g_S\|_2^2 &= \sum_{U\subseteq[R_2]:\pi(U)=S} \left(1-(1-\gamma')^{2|U|}\right)\hat{g}(U)^2 \\ &\le \sum_{U\subseteq[R_2]:\pi(U)=S} \left(1-(1-\gamma')^{2d|S|}\right)\hat{g}(U)^2 = \left(1-(1-\gamma')^{2d|S|}\right)\|g_S\|_2^2,\end{aligned}$$

where we used that π is d-to-1. Substituting into (21) and then into (20), we determine

$$(20) \le \sqrt{\sum_{S\subseteq[R_1]} \rho_0^{2|S|}\left(1-(1-\gamma')^{2d|S|}\right)\|g_S\|_2^2}. \tag{22}$$

We now bound

$$\rho_0^{2|S|}\left(1-(1-\gamma')^{2d|S|}\right) \le \exp\left(-\frac{\delta^2}{d^2 2^{2d}}|S|\right)\cdot(2d|S|\gamma');$$

simple calculus shows that the maximum of this, over $|S|$, is at most

$$\frac{d^3 2^{2d+1}}{e\delta^2}\gamma'. \tag{23}$$

By choosing $\gamma' > 0$ small enough we can upper-bound (23) by $\delta/9$; specifically, we need to choose

$$\gamma' = (\delta/2^d)^{O(1)}.$$

Doing so, we get

$$(22) \le \sqrt{\sum_{S\subseteq[R_1]} (\delta/9)\|g_S\|_2^2} = (\sqrt{\delta}/3)\sqrt{\mathbf{E}[g^2]} = \sqrt{\delta}/3,$$

since g's range is $\{-1,1\}$. This completes the proof of Lemma C.3. □

The proof of Lemma C.4 is essentially identical. The two differences are: i) we interchange the roles of $\boldsymbol{y}$ and $\boldsymbol{z}$, which is okay because $\mathcal{H}_\delta$ is symmetric under this interchange; ii) we use $F = f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{z})$, which does not have $\mathbf{E}[F^2]$ precisely 1 but which still has $\mathbf{E}[F^2] \le 1$, since $T_{1-\gamma'}g$ is bounded in $[-1,1]$.

It remains to prove Lemma C.5. We cannot use the same method as in the previous lemmas, since the correlation $\rho(\{-1,1\}^{R_1}, \{-1,1\}^{R_2}\times\{-1,1\}^{R_2}; \mathcal{H}_\delta) = 1$. However the fact that we have already inserted $T_{1-\gamma'}$'s into the $\boldsymbol{y}$ and $\boldsymbol{z}$ spaces breaks the perfect correlation with $\boldsymbol{x}$, allowing us to effect the proof.

Proof. (Lemma C.5) By definition of the Bonami-Beckner operator we can write

$$\mathop{\mathbf{E}}_{\mathcal{H}_\delta}[f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] = \mathop{\mathbf{E}}_{\mathcal{H}_\delta^*}[f(\boldsymbol{x})g(\boldsymbol{y}^*)g(\boldsymbol{z}^*)],$$

where $\mathcal{H}_\delta^*$ is the distribution on $\{-1,1\}^{R_1}\times\{-1,1\}^{R_2}\times\{-1,1\}^{R_3}$ defined by first generating $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \sim \mathcal{H}_\delta$ and then forming $\boldsymbol{y}^*$, $\boldsymbol{z}^*$ by rerandomizing each bit in $\boldsymbol{y}$, $\boldsymbol{z}$ (respectively) independently with probability γ'. Note that $\mathcal{H}_\delta^*$ can also be written as a product distribution

$$\mathcal{H}_\delta^* = \mathcal{H}_\delta^*(d_1)\otimes\cdots\otimes\mathcal{H}_\delta^*(d_{R_1}),$$

where the distribution $\mathcal{H}_\delta^*(D)$ on $\{-1,1\}\times\{-1,1\}^D\times\{-1,1\}^D$ is defined in the analogous way.

The following lemma is proved in Section G:

Lemma C.6. *Assuming* $0 < D \le d$,

$$\rho(\{-1,1\},\{-1,1\}^D \times \{-1,1\}^D; \mathcal{H}^*_\delta(D)) \le 1 - {\gamma'}^{2d}/2.$$

As we've seen [20, Proposition 2.13], this lemma implies

$$\rho(\{-1,1\}^{R_1},\{-1,1\}^{R_2} \times \{-1,1\}^{R_2}; \mathcal{H}^*_\delta) \le \rho_0^* := 1 - {\gamma'}^{2d}/2. \tag{24}$$

It is now easy to complete the proof of the present lemma using the same proof technique used for Lemma C.3: One shows that

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}^*_\delta}[f(\boldsymbol{x})g(\boldsymbol{y}^*)g(\boldsymbol{z}^*)] - \mathop{\mathbf{E}}_{\mathcal{H}^*_\delta}[T_{1-\gamma}f(\boldsymbol{x})g(\boldsymbol{y}^*)g(\boldsymbol{z}^*)] \right| \le \sqrt{\delta}/3,$$

(which is equivalent to the bound (16)) by selecting γ small enough. Specifically, one needs

$$(\rho_0^*)^{2|S|}\left(1 - (1-\gamma)^{2d|S|}\right) \le \delta/9,$$

which can be achieved by taking

$$\gamma = (\delta/\gamma')^{O(d)} > 0.$$

□

Having proved Lemmas C.3–C.5, the proof of Theorem C.1 is complete. □

C.2 Proof of Theorem C.2

The technique for the proof is similar to the original Invariance Principle [21] and the "multidimensional" version of it due to Mossel [20]. Again, though, we cannot use the results in either of these papers directly. The reason is that one cannot execute a truncation argument passing from "smoothed functions" to "low-degree functions" in the trilinear setting; degree-truncated functions no longer need to have range $[-1,1]$, and this conflicts with the last step in our one-at-a-time replacement argument in Theorem C.1.

Instead we employ a novel but natural strategy: working directly with smoothed functions in the inductive proof of invariance. Our inductive proof of Theorem C.2 works as in [21, 20] by changing the distribution from $\mathcal{H}_\delta = \mathcal{H}_\delta(d_1) \otimes \cdots \otimes \mathcal{H}_\delta(d_{R_1})$ to $\mathcal{I}_\delta = \mathcal{I}_\delta(d_1) \otimes \cdots \otimes \mathcal{I}_\delta(d_{R_1})$ one component at a time. Specifically, we will show that for each $k \in [R_1]$,

$$\left| \mathop{\mathbf{E}}_{\bigotimes_{i=1}^{k-1}\mathcal{I}_\delta(d_i)\otimes\bigotimes_{i=k}^{R_1}\mathcal{H}_\delta(d_i)}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\bigotimes_{i=1}^{k}\mathcal{I}_\delta(d_i)\otimes\bigotimes_{i=k+1}^{R_1}\mathcal{H}_\delta(d_i)}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \le \Delta_k, \tag{25}$$

where

$$\Delta_k := \tau^{\gamma/6}\left(2^d \mathrm{Inf}_k(T_{1-\gamma/2}f) + \sum_{S\subseteq\pi^{-1}(k),|S|\text{ odd}} \mathrm{Inf}_S(T_{1-\gamma/2}g)\right).$$

If we sum this over all $k \in [R_1]$, the triangle inequality implies that

$$\begin{aligned}
&\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta}[T_{1-\gamma}f(\boldsymbol{x})T_{1-\gamma'}g(\boldsymbol{y})T_{1-\gamma'}g(\boldsymbol{z})] \right| \\
\le\ & \tau^{\gamma/6}\left(2^d \sum_{k=1}^{R_1} \mathrm{Inf}_k(T_{1-\gamma/2}f) + \sum_{S\subseteq[R_2],|S|\le d} \mathrm{Inf}_S(T_{1-\gamma/2}g)\right) \\
\le\ & \tau^{\gamma/6}(2^d(1/\gamma) + (d/\gamma)^d) \qquad \text{(by Lemma 5.9)} \\
\le\ & 2\tau^{\gamma/6}(d/\gamma)^d.
\end{aligned}$$

We now choose $\tau > 0$ small enough as a function of γ and d so that this bound is at most $\sqrt{\delta}$, completing the proof. This requires

$$\tau = \delta^{O(1/\gamma)} \cdot (\gamma/d)^{O(d/\gamma)}.$$

Thus it remains to show (25). Due to the rather severe notational complication, we will just show it for the case $k = 1$. The reader will see that the proof for $k = 2 \dots R_1$ is the same; the only fact used about the unchanged product distribution

$$\mathcal{I}_\delta(d_1) \otimes \cdots \otimes \mathcal{I}_\delta(d_{k-1}) \otimes \quad \cdot \quad \otimes \mathcal{H}_\delta(d_{k+1}) \otimes \cdots \otimes \mathcal{H}_\delta(d_{R_1})$$

is that its marginals on $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are uniform. What follows is the proof:

Lemma C.7. *Write $\mathcal{H}'_\delta$ for the distribution $\mathcal{H}_\delta(d_2) \otimes \cdots \otimes \mathcal{H}_\delta(d_{R_1})$. Then*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\delta(d_1) \otimes \mathcal{H}'_\delta} [T_{1-\gamma} f(\boldsymbol{x}) T_{1-\gamma'} g(\boldsymbol{y}) T_{1-\gamma'} g(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta(d_1) \otimes \mathcal{H}'_\delta} [T_{1-\gamma} f(\boldsymbol{x}) T_{1-\gamma'} g(\boldsymbol{y}) T_{1-\gamma'} g(\boldsymbol{z})] \right|$$
$$\leq \tau^{\gamma/6} \left(2^d \mathrm{Inf}_1(T_{1-\gamma/2} f) + \sum_{S \subseteq [d_1], |S| \ odd} \mathrm{Inf}_S(T_{1-\gamma/2} g) \right). \quad (26)$$

Proof. Let us write x' for strings $(x_2, \dots, x_{R_1})$, y' for strings $(y_{d_1+1}, \dots, y_{R_2})$, and strings z' similarly. We break up the Fourier expansion of f according to its dependence on x_1:

$$f(x) = F_\emptyset(x') + x_1 F_1(x').$$

Similarly, we break up the Fourier expansion of g according to its dependence on the bits $y_1, \dots, y_{d_1}$:

$$g = \sum_{S \subseteq [d_1]} \chi_S(y_1, \dots, y_{d_1}) G_S(y').$$

Here for any $S \subseteq [d_1]$ we have

$$G_S(y') = \sum_{Q \subseteq [R_2], Q \cap [d_1] = S} \hat{g}(Q) \chi_{Q \setminus S}(y'). \quad (27)$$

It is easy to check that we have

$$G_S(y') = \mathop{\mathbf{E}}_{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{d_1}} [g(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{d_1}, y') \chi_S(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{d_1})]$$

and therefore the function G_S is always bounded in $[-1, 1]$. Similarly $F_\emptyset$ and F_1 are bounded in $[-1, 1]$. We also observe that we have the Fourier expansions

$$T_{1-\gamma} f(x) = T_{1-\gamma} F_\emptyset(x') + (1-\gamma) x_1 T_{1-\gamma} F_1(x'), \quad (28)$$

$$T_{1-\gamma'} g(y) = \sum_{S \subseteq [d_1]} (1-\gamma')^{|S|} \chi_S(y_1, \dots, y_{d_1}) T_{1-\gamma'} G_S(y'). \quad (29)$$

We employ the following lemma, whose proof appears in Section H:

Lemma C.8. *For any functions $F : \mathcal{X} \to \mathbb{R}$, $G : \mathcal{Y} \to \mathbb{R}$, $H : \mathcal{Z} \to \mathbb{R}$,*

$$\mathop{\mathbf{E}}_{\mathcal{H}_\delta(D)} [F(\boldsymbol{x}) G(\boldsymbol{y}) H(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)} [F(\boldsymbol{x}) G(\boldsymbol{y}) H(\boldsymbol{z})] = \sum_{S \subseteq [D], |S| \ odd} (1-\delta) \hat{F}(\{1\}) \hat{G}(S) \hat{H}(S).$$

We now upper-bound the LHS of (26) by

$$\left|\mathop{\mathbf{E}}_{\mathcal{H}'_\delta}\left[\mathop{\mathbf{E}}_{\mathcal{H}_\delta(d_1)}[T_{1-\gamma}f(x_1,\boldsymbol{x}')T_{1-\gamma'}g(y_1,\dots,y_{d_1},\boldsymbol{y}')T_{1-\gamma'}g(z_1,\dots,z_{d_1},\boldsymbol{z}')]\right.\right.$$
$$\left.\left.-\mathop{\mathbf{E}}_{\mathcal{I}_\delta(d_1)}[T_{1-\gamma}f(x_1,\boldsymbol{x}')T_{1-\gamma'}g(y_1,\dots,y_{d_1},\boldsymbol{y}')T_{1-\gamma'}g(z_1,\dots,z_{d_1},\boldsymbol{z}')]\right]\right|.$$

Using Lemma C.8 and recalling (28) and (29), one can check that this is equal to

$$\begin{aligned}
&\left|\sum_{S\subseteq[d_1],|S|\text{ odd}}(1-\delta)(1-\gamma)(1-\gamma')^{2|S|}\mathop{\mathbf{E}}_{\mathcal{H}'_\delta}[T_{1-\gamma}F_1(\boldsymbol{x}')T_{1-\gamma'}G_S(\boldsymbol{y}')T_{1-\gamma'}G_S(\boldsymbol{z}')]\right| \\
\le &\sum_{S\subseteq[d_1],|S|\text{ odd}}(1-\gamma)(1-\gamma')^{2|S|}\cdot\mathop{\mathbf{E}}_{\mathcal{H}'_\delta}[|T_{1-\gamma}F_1(\boldsymbol{x}')T_{1-\gamma'}G_S(\boldsymbol{y}')T_{1-\gamma'}G_S(\boldsymbol{z}')|] \\
\le &\sum_{S\subseteq[d_1],|S|\text{ odd}}(1-\gamma)(1-\gamma')^{2|S|}\cdot\|T_{1-\gamma}F_1\|_3\|T_{1-\gamma'}G_S\|_3^2, \qquad (30)
\end{aligned}$$

where the last step uses Hölder's Inequality, and the norms $\|\cdot\|_3$ are with respect to the uniform distribution. This is indeed the only step using properties of the distribution $\mathcal{H}'_\delta$, the fact that its three marginals are uniform.

We now use the following lemma, whose proof in Section I is a straightforward application of the Hypercontractive Inequality of Bonami [5] and Gross [9]:

Lemma C.9. *For any function $f:\{-1,1\}^n\to[-1,1]$ and $0<\gamma<1$,*

$$\|T_{1-\gamma}f\|_3\le\|T_{1-\gamma/2}f\|_2^{(2+\gamma)/3}.$$

As F_1 and G_S are indeed bounded in $[-1,1]$, we can upper-bound

$$\|T_{1-\gamma}F_1\|_3\|T_{1-\gamma'}G_S\|_3^2\le\|T_{1-\gamma/2}F_1\|_2^{(2+\gamma)/3}\|T_{1-\gamma'/2}G_S\|_2^{(4+2\gamma')/3}. \qquad (31)$$

By expressing the Fourier coefficients of G_S using g's original Fourier coefficients (via (27)) we have

$$\begin{aligned}
\|T_{1-\gamma'/2}G_S\|_2^2 &= \sum_{Q\subseteq[R_2],Q\cap[d_1]=S}(1-\gamma'/2)^{2|Q|-2|S|}\hat{g}(Q)^2 \\
&\le \sum_{S\subseteq Q\subseteq[R_2]}(1-\gamma'/2)^{2|Q|-2|S|}\hat{g}(Q)^2=(1-\gamma'/2)^{-2|S|}\cdot\mathrm{Inf}_S(T_{1-\gamma/2}g),
\end{aligned}$$

where we also used $\gamma'\ge\gamma$ in the last step.

A similar calculation yields

$$\|T_{1-\gamma/2}F_1\|_2^2\le(1-\gamma/2)^2\cdot\mathrm{Inf}_1(T_{1-\gamma/2}f).$$

Plugging these last two bounds into (31) and then into (30), we upper-bound the LHS of (26) by

$$\sum_{S\subseteq[d_1],|S|\text{ odd}}\left(\mathrm{Inf}_1(T_{1-\gamma/2}f)\right)^{1/3+\gamma/6}\cdot\left(\mathrm{Inf}_S(T_{1-\gamma/2}g)\right)^{2/3+\gamma/3}.$$

By the hypothesis (11) of the overarching theorem we are proving, Theorem 6.3, either $(\mathrm{Inf}_1(T_{1-\gamma/2}f))^{\gamma/6}\le\tau^{\gamma/6}$ or $(\mathrm{Inf}_S(T_{1-\gamma/2}g))^{\gamma/6}\le\tau^{\gamma/3}$ for each S in the sum. In either case

we can bound the above by

$$\begin{aligned}
& \tau^{\gamma/6} \cdot \sum_{S\subseteq[d_1],|S| \text{ odd}} \left(\mathrm{Inf}_1(T_{1-\gamma/2}f)\right)^{1/3} \cdot \left(\mathrm{Inf}_S(T_{1-\gamma/2}g)\right)^{2/3} \\
\leq\ & \tau^{\gamma/6} \cdot \sum_{S\subseteq[d_1],|S| \text{ odd}} \left(\mathrm{Inf}_1(T_{1-\gamma/2}f) + \mathrm{Inf}_S(T_{1-\gamma/2}g)\right) \qquad (\text{using } a^{1/3}b^{2/3} \leq a+b \text{ for } a,b\geq 0) \\
\leq\ & \tau^{\gamma/6} \cdot \left(2^d \mathrm{Inf}_1(T_{1-\gamma/2}g) + \sum_{S\subseteq[d_1],|S| \text{ odd}} \mathrm{Inf}_S(T_{1-\gamma/2}g)\right).
\end{aligned}$$

This completes the proof of the inductive Lemma C.7 □

D Proofs of the correlation Lemmas 5.2–5.4

Proof. (Lemma 5.2) Recall that $\mathcal{H}_\delta(D) = (1-\delta)\mathcal{H}(D) + \delta\mathcal{N}(D)$. Note that the marginals of all distributions mentioned are uniform on $\mathcal{X}$ and on $\mathcal{Y}$. Now suppose that $f : \mathcal{X} \to \mathbb{R}$, $g : \mathcal{Y} \to \mathbb{R}$ are any functions with $\mathbf{E}[f] = \mathbf{E}[g] = 0$, $\mathbf{E}[f^2], \mathbf{E}[g^2] \leq 1$ (under the uniform distribution). Then

$$\mathop{\mathbf{E}}_{\mathcal{H}_\delta(D)}[f(\boldsymbol{x})g(\boldsymbol{y})] = (1-\delta)\mathop{\mathbf{E}}_{\mathcal{H}(D)}[f(\boldsymbol{x})g(\boldsymbol{y})] + \delta\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})g(\boldsymbol{y})] = \delta\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})g(\boldsymbol{y})],$$

because $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent under $\mathcal{H}(D)$, and

$$\delta\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})g(\boldsymbol{y})] \leq \delta\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})^2]}\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[g(\boldsymbol{y})^2]} \leq \delta$$

by Cauchy-Schwarz. This establishes $\rho(\mathcal{X},\mathcal{Y};\mathcal{H}_\delta(D)) \leq \delta$. □

Proof. (Lemma 5.3) For this we rely on Lemma 2.9 from Mossel's work [20] which implies that $\rho(\mathcal{X}\times\mathcal{Y},\mathcal{Z};\mathcal{H}_\delta(D)) \leq 1-\alpha^2/2$, where α is the least probability of an atom under $\mathcal{H}_\delta(D)$, so long as the distribution $\mathcal{H}_\delta(D)$ is "connected". It is easy to check that $\alpha = \frac{\delta}{d2^d}$ for $\mathcal{H}_\delta(D)$, so the proof is complete as long as we have connectedness.

"Connectedness" here refers to the undirected bipartite graph $G(\mathcal{X}\times\mathcal{Y},\mathcal{Z})$ which has an edge joining left-vertex $(x, y_1, \ldots, y_D)$ and right-vertex $(z_1, \ldots, z_D)$ if and only if

$$\mathop{\mathbf{Pr}}_{\mathcal{H}_\delta(D)}[(\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_D) = (x, y_1, \ldots, y_D), (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_D) = (z_1, \ldots, z_D).$$

We now show this graph is indeed connected. Recall that $\mathcal{H}_\delta(D) = (1-\delta)\mathcal{H}(D) + \delta\mathcal{N}(D)$. Let $(z_1, \ldots, z_D)$ be an arbitrary right-vertex. By definition of $\mathcal{N}(D)$, we know $(z_1, \ldots, z_d)$ is connected to $(x = z_1, y_1 = z_1, y_2 = -xz_2, \ldots, y_d = -xz_D)$. And by definition of $\mathcal{H}(D)$, the left-vertex $(x = z_1, y_1 = z_1, y_2 = -xz_2, \ldots, y_d = -xz_D)$ is connected to $(-1, z_2, \ldots, z_D)$. Hence $(\boldsymbol{z}_1, ..., \boldsymbol{z}_D)$ is connected to $(-1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_D)$. We can make the same argument for indices $2, \ldots, D$ and conclude that $(-1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_D)$ is connected to $(-1, -1, z_3, \ldots, \boldsymbol{z}_D)$, that this vertex is connected to $(-1, -1, -1, \boldsymbol{z}_4, \ldots, \boldsymbol{z}_D)$, etc., and hence that all right-vertices are connected to the right-vertex $(-1, \ldots, -1)$. Hence all right-vertices are connected. Therefore the whole graph, by the easy observation that no left-vertex is isolated. □

Proof. (Lemma 5.4) Recall that $\mathcal{I}_\delta(D) = (1-\delta)\mathcal{I}(D) + \delta\mathcal{N}(D)$. It may seem as though we should have $\rho(\mathcal{X},\mathcal{Y}\times\mathcal{Z};\mathcal{I}_\delta(D)) \leq \delta$ because the marginals of $\mathcal{I}(D)$ on $\mathcal{X}$ and $\mathcal{Y}\times\mathcal{Z}$ are independent. However this is incorrect because the marginals of $\mathcal{I}(D)$ and $\mathcal{N}(D)$ on $\mathcal{Y}\times\mathcal{Z}$ are not the same. Nevertheless we can still achieve an upper bound of $\sqrt{\delta}$, as follows.

Let $f : \mathcal{X} \to \mathbb{R}$ be any function with $\mathbf{E}[f] = 0$, $\mathbf{E}[f^2] \leq 1$ under the uniform distribution (which is the marginal of $\mathcal{I}(D)$, $\mathcal{N}(D)$, and hence $\mathcal{I}_\delta(D)$ on $\mathcal{X}$). Let also $G : \mathcal{Y}\times\mathcal{Z} \to \mathbb{R}$ be any function with $\mathbf{E}_{\mathcal{I}_\delta(D)}[G(\boldsymbol{y},\boldsymbol{z})] = 0$, $\mathbf{E}_{\mathcal{I}_\delta(D)}[G(\boldsymbol{y},\boldsymbol{z})^2] \leq 1$. This latter property yields

$$1 \geq \mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)}[G(\boldsymbol{y},\boldsymbol{z})^2] = (1-\delta)\mathop{\mathbf{E}}_{\mathcal{I}(D)}[G(\boldsymbol{y},\boldsymbol{z})^2] + \delta\mathop{\mathbf{E}}_{\mathcal{N}(D)}[G(\boldsymbol{y},\boldsymbol{z})^2] \geq \delta\mathop{\mathbf{E}}_{\mathcal{N}}[G(\boldsymbol{y},\boldsymbol{z})^2]. \qquad (32)$$

We now observe that

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y},\boldsymbol{z})] &= (1-\delta)\mathop{\mathbf{E}}_{\mathcal{I}(D)}[f(\boldsymbol{x})G(\boldsymbol{y},\boldsymbol{z})] + \delta\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})G(\boldsymbol{y},\boldsymbol{z})] \\ &\leq (1-\delta)\mathop{\mathbf{E}}_{\mathcal{I}(D)}[f(\boldsymbol{x})]\mathop{\mathbf{E}}_{\mathcal{I}(D)}[G(\boldsymbol{y},\boldsymbol{z})] + \delta\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})^2]}\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[G(\boldsymbol{y},\boldsymbol{z})^2]},\end{aligned}$$

where we have equality for the first summand because the marginals of $\mathcal{I}(D)$ on $\mathcal{X}$ and $\mathcal{Y}\times\mathcal{Z}$ are independent, and where we used Cauchy-Schwarz on the second summand. But we know that $\mathbf{E}_{\mathcal{I}(D)}[f(\boldsymbol{x})] = 0$ and that

$$\delta\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[f(\boldsymbol{x})^2]}\sqrt{\mathop{\mathbf{E}}_{\mathcal{N}(D)}[G(\boldsymbol{y},\boldsymbol{z})^2]} \leq \delta\cdot\sqrt{1}\cdot\sqrt{1/\delta},$$

using (32). Hence

$$\mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y},\boldsymbol{z})] \leq \sqrt{\delta}$$

and this completes the proof that $\rho(\mathcal{X},\mathcal{Y}\times\mathcal{Z};\mathcal{I}_\delta(D)) \leq \sqrt{\delta}$. □

E Proof of Lemma 5.9

Proof. We have

$$\sum_{S\subseteq[n],|S|\leq m}\mathrm{Inf}_S(T_{1-\gamma}f) = \sum_{|S|\leq m}\sum_{U\supseteq S}(1-\gamma)^{2|U|}\hat{f}(U)^2.$$

Each expression $(1-\gamma)^{2|U|}\hat{f}(U)^2$ is counted exactly $\sum_{i=0}^m\binom{|U|}{i}$ times, and this quantity is at most $(|U|+1)^m$ (since we can overcount the subsets of U of cardinality at most m by imagining picking m times from the set $U\cup\{\text{nothing}\}$). Hence the above is at most

$$\sum_{U\subseteq[n]}(|U|+1)^m(1-\gamma)^{2|U|}\hat{f}(U)^2.$$

Since $\sum_U\hat{f}(U)^2 = \mathbf{E}[f^2]\leq 1$ by hypothesis, we can complete the proof by showing that $(|U|+1)^m(1-\gamma)^{2|U|}\leq(m/2\gamma)^m$ always.

The result is clear for $m=0$ (assuming $0^0=1$) and is otherwise a simple exercise. We have

$$(u+1)^m(1-\gamma)^{2u}\leq(u+1)^m\exp(-2\gamma u) =: f(u),$$

and basic calculus implies that the maximum of $f(u)$ for $u\geq 0$ occurs when $u=(m/2\gamma)-1$ (which is nonnegative since $m\geq 1,\gamma\leq 1/2$). At this value of u we have $f(u) = (m/2\gamma)^m\exp(-(m-2\gamma))\leq(m/2\gamma)^m$ as required, where we again used $m-2\gamma\geq 0$. □

F Proof of the matrix algebra lemma

The following result of matrix algebra is likely known; we were unable to find a reference.

Lemma F.1. *Let A_i and B_i be $m_i\times m_i$ matrices, $i=1\ldots n$, and suppose that A_i-B_i and A_i+B_i are positive semidefinite. Then $\bigotimes_{i=1}^n A_i - \bigotimes_{i=1}^n B_i$ and $\bigotimes_{i=1}^n A_i + \bigotimes_{i=1}^n B_i$ are also positive semidefinite.*

Proof. We prove the claim by induction on n. The base case is immediate. For the inductive step, one begins by checking the identity

$$\bigotimes_{i=1}^{n+1}A_i - \bigotimes_{i=1}^{n+1}B_i = \tfrac{1}{2}(A_{n+1}+B_{n+1})\otimes\left(\bigotimes_{i=1}^n A_i - \bigotimes_{i=1}^n B_i\right) + \tfrac{1}{2}(A_{n+1}-B_{n+1})\otimes\left(\bigotimes_{i=1}^n A_i + \bigotimes_{i=1}^n B_i\right).$$

By induction, all four matrices appearing on the RHS above are positive semidefinite; hence the entire RHS is positive semidefinite. The same argument with the identity

$$\bigotimes_{i=1}^{n+1} A_i + \bigotimes_{i=1}^{n+1} B_i = \tfrac{1}{2}\left(A_{n+1}+B_{n+1}\right)\otimes\left(\bigotimes_{i=1}^{n} A_i + \bigotimes_{i=1}^{n} B_i\right) + \tfrac{1}{2}\left(A_{n+1}-B_{n+1}\right)\otimes\left(\bigotimes_{i=1}^{n} A_i - \bigotimes_{i=1}^{n} B_i\right).$$

completes the induction. □

G Proof of Lemma C.6

Proof. The proof is quite similar to the proof of Lemma 5.4 from Section D. Let $f : \mathcal{X} \to \mathbb{R}$ be any function with $\mathbf{E}[f] = 0$, $\mathbf{E}[f^2] \leq 1$ under the uniform distribution (which is the marginal of $\mathcal{H}^*_\delta(D)$ on $\mathcal{X}$). Let also $G : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$ be any function with $\mathbf{E}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)] = 0$, $\mathbf{E}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2] \leq 1$.

By the definition of $\mathcal{H}^*_\delta(D)$, there is a γ'^{2D} chance that all $2D$ bits in $\boldsymbol{y}$, $\boldsymbol{z}$ are rerandomized in forming $\boldsymbol{y}^*$, $\boldsymbol{z}^*$. Call this event V. Note that conditioned on V occurring, the bit $\boldsymbol{x}$ is independent of the strings $\boldsymbol{y}^*$, $\boldsymbol{z}^*$; also $\boldsymbol{x}$ is independent of V. Thus we have

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y}^*, \boldsymbol{z}^*)] && (33)\\
=\ &\left(1-\gamma'^{2D}\right)\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y}^*, \boldsymbol{z}^*) \mid \neg V] + \gamma'^{2D}\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y}^*, \boldsymbol{z}^*) \mid V]\\
=\ &\left(1-\gamma'^{2D}\right)\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y}^*, \boldsymbol{z}^*) \mid \neg V] + \gamma'^{2D}\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})]\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*) \mid V]\\
=\ &\left(1-\gamma'^{2D}\right)\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})G(\boldsymbol{y}^*, \boldsymbol{z}^*) \mid \neg V]\\
\leq\ &\left(1-\gamma'^{2D}\right)\sqrt{\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[f(\boldsymbol{x})^2]}\sqrt{\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2 \mid \neg V]}\\
\leq\ &\left(1-\gamma'^{2D}\right)\sqrt{\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2 \mid \neg V]}. && (34)
\end{aligned}$$

We also have

$$\begin{aligned}
1 \geq \mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2] &= \left(1-\gamma'^{2D}\right)\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2 \mid \neg V] + \gamma'^{2D}\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2 \mid V]\\
&\geq \left(1-\gamma'^{2D}\right)\mathop{\mathbf{E}}_{\mathcal{H}^*_\delta(D)}[G(\boldsymbol{y}^*, \boldsymbol{z}^*)^2 \mid \neg V].
\end{aligned}$$

Substituting this into (34) gives an upper bound of

$$\left(1-\gamma'^{2D}\right)/\sqrt{1-\gamma'^{2D}} = \sqrt{1-\gamma'^{2D}} \leq 1-\gamma'^{2D}/2 \leq 1-\gamma'^{2d}/2,$$

establishing that

$$\rho(\{-1,1\}, \{-1,1\}^D \times \{-1,1\}^D; \mathcal{H}^*_\delta(D)) \leq 1-\gamma'^{2d}/2$$

as claimed. □

H Proof of Lemma C.8

Proof. We use the Fourier expansion of each function in the LHS above, so F is expanded in terms of the characters χ_U, $U \subseteq [1]$, G in terms of characters $\chi_V \subseteq [D]$, and H in terms of

characters χ_W, $W \subseteq [D]$:

$$\text{LHS} = \sum_{U,V,W} \hat{F}(U)\hat{G}(V)\hat{H}(W) \left(\mathop{\mathbf{E}}_{\mathcal{H}_\delta(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] \right). \tag{35}$$

Both $\mathcal{H}_\delta(D)$ and $\mathcal{I}_\delta(D)$ are mixture distributions with δ weight on $\mathcal{N}(D)$: coupling the choice of mixing components and using linearity of expectation, we have

$$\begin{aligned} &\mathop{\mathbf{E}}_{\mathcal{H}_\delta(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}_\delta(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] \\ &\qquad = (1-\delta)\left(\mathop{\mathbf{E}}_{\mathcal{H}(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] - \mathop{\mathbf{E}}_{\mathcal{I}(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] \right). \end{aligned} \tag{36}$$

Notice that $\mathcal{H}(D)$ and $\mathcal{I}(D)$ have the same marginal distribution on $\mathcal{Y} \times \mathcal{Z}$. Hence for (36) to be nonzero, U must be nonempty — and therefore equal to $\{1\}$. When $U = \{1\}$ we have that

$$\mathop{\mathbf{E}}_{\mathcal{I}(D)}[\chi_U(\boldsymbol{x})\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] = \mathop{\mathbf{E}}_{\mathcal{I}(D)}[\boldsymbol{x}] \mathop{\mathbf{E}}_{\mathcal{I}(D)}[\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})] = 0.$$

As for the other term in (36), it is not hard to check that $\mathbf{E}_{\mathcal{H}(D)}[\boldsymbol{x}\chi_V(\boldsymbol{y})\chi_W(\boldsymbol{z})]$ is zero unless $V = W$ and $|V|$ is odd; in this case, the expectation is 1. Thus (35) is equal to the RHS, as claimed. □

I Proof of Lemma C.9

Proof. Even stronger we can prove that for $f' : \{-1,1\}^n \to [-1,1]$,

$$\|T_{1-\gamma'} f'\|_3 \le \|f'\|_2^{(2+2\gamma')/3}. \tag{37}$$

The result then follows from observing that

$$\|T_{1-\gamma} f\|_3 \le \|T_{1-\gamma/2} T_{1-\gamma/2} f\|_3 \le \|T_{1-\gamma/2} f\|_2^{(2+\gamma)/3},$$

where the first step uses $1-\gamma \le (1-\gamma/2)^2$ and the fact that $T_{1-\gamma/2}$ is a contraction in L^3, and the second step uses (37) (since $T_{1-\gamma/2} f$ is also bounded in $[-1,1]$).

To prove (37) we observe

$$\|T_{1-\gamma'} f'\|_3 = \mathbf{E}[|T_{1-\gamma'} f'|^3]^{1/3} \le \mathbf{E}[|T_{1-\gamma'} f'|^{2+2\gamma'}]^{1/3} = \|T_{1-\gamma'} f'\|_{2+2\gamma'}^{(2+2\gamma')/3},$$

using the fact that $|T_{1-\gamma'} f'| \le 1$. Now we use the Hypercontractive Inequality, observing that $1-\gamma' \le 1/\sqrt{1+2\gamma'}$, to conclude that the above is at most $\|f'\|_2^{(2+2\gamma')/3}$. □