Approximation Resistance on Satisfiable Instances for Predicates with Few Accepting Inputs

Sangxia Huang

Abstract

For every integer $k \geq 3$, we prove that there is a predicate P on k Boolean variables with $2^{\tilde{O}(k^{1/3})}$ accepting assignments that is approximation resistant even on *satisfiable* instances. That is, given a *satisfiable* CSP instance with constraint P, we cannot achieve better approximation ratio than simply picking random assignments. This improves the best previously known result by Håstad and Khot where the predicate has $2^{O(k^{1/2})}$ accepting assignments.

Our construction is inspired by several recent developments. One is the idea of using direct sums to improve soundness of PCPs, developed by Chan [5]. We also use techniques from Wenner [32] to construct PCPs with perfect completeness without relying on the d-to-1 Conjecture.

1 Introduction

Consider a predicate $P : \{-1, 1\}^k \to \{0, 1\}$ on k Boolean variables (where for the input variables we use -1 for True and 1 for False), an instance of the Max-P problem consists of n Boolean variables, along with constraints of form $P(l_1, \cdots, l_k)$, where $l_1, \cdots, l_k$ are literals of k distinct variables, and the goal is to find a Boolean assignment to the variables that satisfies as many constraints as possible. A Max-P instance is called *satisfiable* if there exists an assignment that satisfies all the constraints simultaneously. Let $P^{-1}(1)$ be the set of accepting inputs of P.

One naive approximation algorithm for Max-P is to simply pick a random assignment. This gives an approximation ratio of $|P^{-1}(-1)|/2^k$. Somewhat surprisingly, it turns out that for some predicate P, the above naive algorithm gives the best possible performance assuming P $\neq$ NP. We call a predicate P *approximation resistant* if it is hard to achieve better approximation ratio than simply picking random assignments. In a celebrated result, Håstad [13] showed that Max-kLin, sets of linear equations in $\mathbb{Z}_2$ on $k \geq 3$ variables, is NP-hard to approximate better than $1/2 + \varepsilon$ for any $\varepsilon > 0$, while the random assignment algorithm achieves $1/2$. There has been much progress in understanding what kinds of predicates are approximation resistant, including characterization for predicates of small arity [13, 33, 11], as well as a handful of approximation resistant predicates of higher arities [13, 26, 11, 8].

The picture of approximation resistance becomes even clearer if we assume the Unique Games Conjecture (UGC) proposed by Khot [18], which states that it is NP-hard to distinguish whether certain Label Cover instance is *almost* satisfiable or far from satisfiable. Austrin and Mossel [4] proved that assuming the UGC, P is approximation resistant if the set of satisfying assignments $P^{-1}(1)$ contains the support of

a pairwise independent distribution. In [27, 28], Samorodnitsky and Trevisan showed approximation resistance for the following predicate assuming the UGC: the predicate is on $2^k - 1$ variables, denoted as $x^{(S)}$ for $\emptyset \neq S \subseteq \{1, \cdots, k\}$, and the predicate accepts if for all $S \subseteq [k]$, $|S| \geq 2$, we have $x^{(S)} = \prod_{i \in S} x^{(\{i\})}$. Let $K = 2^k - 1$. We denote the above predicate as HADAMARD_K. Note that HADAMARD_K has only $K+1$ accepting assignments over 2^K possible assignments, giving a density of $(K+1)/2^K$. Hast [12] proved that a predicate on $K \geq 3$ variables with at most $2\lfloor K/2 \rfloor + 1$ accepting inputs is not approximation resistant. Thus the result of Samorodnitsky and Trevisan is optimal in terms of the sparsity of the predicate.

In a recent breakthrough [5], Chan settled the NP-hardness of Max-HADAMARD_K (and, up to constant factor, Max k-CSP in general), bypassing the UGC. Chan introduced the idea of direct sums of probabilistically checkable proofs (PCPs) to improve soundness, which worked very well for predicates that are subgroups of a domain. In particular, the accepting assignments of the HADAMARD_K predicate is a subgroup under element-wise product, and Chan's result implies that it is approximation resistant assuming only P $\neq$ NP.

Let us now focus on the approximation of *satisfiable* instances. We call P approximation resistant on satisfiable instances if the best possible algorithm is still the random assignment algorithm even with the promise that there is an assignment that satisfies all constraints. In contrast to our understanding of approximation resistance as demonstrated above, approximation resistance on satisfiable instances is still largely a mystery. Most notably, if the constraints only involves linear equations, for instance $k\text{LIN}$ and HADAMARD_K, we can always find a satisfying assignment using Gaussian elimination if we are given satisfiable instances, whereas both of them are approximation resistant in general. Several other approximation algorithms for satisfiable instances were introduced [30], and in particular, it is known that predicates with fewer than $(k+1)$ satisfying assignments are never approximation resistant on satisfiable instances. Note that when k is even, the same statement holds even for non-satisfiable instances due to the aforementioned result by Hast [12].

On the hardness side, there have been only handful of results: Håstad [13] proved that k-SAT is approximation resistant for satisfiable instances. The sparsest such predicate known on k variables has $2^{O(k^{1/2})}$ accepting assignments, given by Håstad and Khot [15]. This situation is not particularly surprising, as there are quite a few differences between satisfiable instances and almost satisfiable instances. Some approximation resistant predicates, such as the $k\text{LIN}$ predicate discussed above, are not approximation resistant on satisfiable instances. In addition to this inherent structural difference, there are challenges in techniques as well. Many approximation resistance results are obtained via reduction from Unique Games. This immediately introduces problem with completeness because the Unique Games problem is not NP-hard for satisfiable instances.

To address this, Khot additionally proposed the "d-to-1 Conjectures" [18]. The conjecture states that it is NP-hard to distinguish whether a "d-to-1 Label Cover Instance" is satisfiable or far from satisfiable. O'Donnell and Wu proved a strong result in [24] that the Not-Two predicate (NTW) — predicate on three variables that accepts input whose number of -1's is not two — is approximation resistant on satisfiable instances assuming the d-to-1 conjecture for some d. Their approach was generalized by Tang [29] to Max-3CSP_q where q is a prime greater than 3, and Huang [16] to Boolean predicates of arity $k \geq 3$ that accepts a strict superset of inputs of odd parity.

Recently, Håstad [14] and Wenner [32] proved approximation resistance for the above predicates without assuming the d-to-1 conjecture. Their proofs are based on new analytic tools as well as Khot's Smooth-Label-Cover [17]. We note that several previous results that bypassed the UGC [17, 19, 9, 10] started from Smooth-Label-Cover, although it is not needed in Chan's recent result.

An interesting question is whether we could combine these recent developments to get approximation resistance result for Max-P on satisfiable instances for predicate P sparser than the one in Håstad and Khot [15]. From the PCP perspective, this requires PCPs that always accept correct proofs of correct statements. Not only is this a natural property to have, given the challenge of getting proofs with perfect completeness as discussed above, understanding approximability of k-CSP on satisfiable instances may also lead to new tools in both algorithm and hardness.

An immediate proposal to achieve tight lower-bound for Max k-CSP on satisfiable instances would be to construct predicates as in [16, 32], that is, adding a single additional accepting assignment to the $\textsc{Hadamard}_K$ predicate of arity $2^k - 1$. However, this simple approach does not work — the accepting inputs of $\textsc{Hadamard}_K$ forms a k-dimensional subspace, so if we add d new accepting inputs to it, we only need a $(k + d)$-dimensional subspace to include all those accepting inputs, and if we sample an assignment from the subspace induced by all constraints, the probability that we satisfy one clause is at least $1/2^{k+d}$ because 2^{k+d} is the maximum size of the subspace for each clause and there is at least one satisfiable assignment in it due to satisfiability of the whole instance. Thus whenever $d = o(2^k)$, the performance of the above sampling method beats simple random assignment, which only gives $(2^k + d)/2^{2^k}$.

The problem with adding accepting assignments to $\textsc{Hadamard}_K$ is that the resulting predicate does not have the group structure as in [5] any more. If we still take many rounds of direct sums as in [5], then to ensure perfect completeness, we need to accept many assignments that are products of the additional assignments we added and end up with a predicate that has more accepting assignments than we would want. On the other hand, as is demonstrated in [5], having more rounds of direct sum helps us to improve soundness dramatically and so if we are looking for sparse predicates that are approximation resistant, it would be natural to have more rounds of proofs in the direct sum. This paper is an attempt to strike a balance. We prove the following approximation resistance result.

Theorem 1.1. *There is a predicate of arity K with $2^{\tilde{O}(K^{1/3})}$ accepting assignments that is approximation resistance on satisfiable instances.*

This improves the best previous known result of $2^{O(K^{1/2})}$ of Håstad and Khot [15].

Our result is based on many ideas developed in a number of previous works, including [8, 32, 31, 5]. On the highest level, we use direct sum of several PCPs to get improved soundness result. However, as argued above, we also want to limit the number of PCPs involved. Therefore, we use long-code based PCP constructions that are already rather efficient, for example those used by Engebretsen and Holmerin [8]. In [31], Wenner showed how different types of noise operators behave similarly when the reduction is based on Smooth-Label-Covers. This is helpful when analyzing soundness of PCPs in that it allows us to move from correlated noise with perfect completeness to independent noise that are not perfect but easier to analyze. We also use a multivariate invariance theorem in [31], which extends methods of Mossel

et. al. [22, 21] to projection games. Similar techniques were developed also in other works such as [23] as well as in [5].

1.1 Organization of the Paper

The paper is organized as follows. Section 2 reviews some notation we use, including variants of Label-Cover, PCPs and Chan's new technique, and analysis of Boolean functions. We describe our PCP construction in Section 3. We analyze the soundness of our construction in Section 4.

2 Preliminaries

In this section, we introduce some notation. In Section 2.1, we discuss variants of Label-Cover problems, and in particular the Multi-Layer Smooth-Label-Cover that we use in the rest of the paper. We describe the general approach of proving approximation resistance via Label-Cover, as well as Chan's improvements in Section 2.2. In Section 2.3, we review basics of harmonic analysis of Boolean functions.

2.1 Variants of Label-Cover

We first recall the definition of the Label-Cover problem.

Definition 2.1. *A Label-Cover instance is defined by a tuple (U, V, E, L, R, Π). Here U and V are two sets of vertices of a bipartite multigraph, and E is the set of edges between them. L and R are label sets for vertices U and V, respectively. Π is a collection of projections, one for each edge e, $\pi_e : R \to L$. For a labeling $\sigma = (\sigma_U, \sigma_V)$ of the Label-Cover instance $\sigma_U : U \to L$, $\sigma_V : V \to R$, let its value be the fraction of edges $\{u, v\} \in E$ such that $\pi_{\{u,v\}}(\sigma(v)) = \sigma(u)$. The value of a Label-Cover instance is the maximum value of all possible assignments.*

The following theorem combines the celebrated PCP theorem [1, 2] with Raz's parallel repetition theorem [25] and shows hardness of Label Cover.

Theorem 2.2. *For every constant $\eta > 0$, there is some constant $C(\eta) < \infty$ such that for Label-Cover instances with $|R| \geq C(\eta)$, it is NP-hard to distinguish between the case where it has value 1 and where it has value no more than η.*

The Smooth-Label-Cover problem is a variant of Label-Cover first defined by Khot [17] for showing inapproximability of some coloring problems. We extend the definition of projection $\pi : R \to L$ to sets of labels $S \subseteq R$ by $\pi(S) := \{l \in L | \exists r \in S, \pi(r) = l\}$. We adopt the following definition of smoothness.

Definition 2.3 (Smoothness)**.** *A Label-Cover instance is (J, ξ)-smooth if for any set of labels $S \subset R$, $|S| \leq J$, we have*

$$\Pr_{e \sim E}[|\pi_e(S)| < |S|] \leq \xi. \tag{1}$$

Similar to Label-Cover, we have the following hardness result for Smooth-Label-Cover.

Theorem 2.4. *For every constant $\eta, J, \xi > 0$, there is some constant $D(\eta, J, \xi) < \infty$ such that for (J, ξ)-smooth Label-Cover instances with $|R| \geq D(\eta, J, \xi)$, it is NP-hard to distinguish between the case where it has value 1 and where it has value no more than η.*

Multi-layered Label-Cover was first devised in [6] to prove strong approximation hardness result for hypergraph vertex cover, and used in [8] for improving query efficiency and hardness of approximation result for Max CSP. Briefly speaking, a normal Label-Cover instance checks consistency of labeling between a pair of vertices, while in a k-layered Label-Cover instance, we consider tuples of $k-1$ independently sampled edges $(\{u_1, v_1\}, \{u_2, v_2\}, \cdots, \{u_{k-1}, v_{k-1}\})$, the k hybrid tuples of vertices $(u_1, \cdots, u_i, v_{i+1}, \cdots, v_{k-1})$ for $i = 0, \cdots, k-1$ and their corresponding labelings, and we require consistency between all pairs of tuples. Formally, given a Label-Cover instance as defined above, the constraint between pairs of labelings on tuples is defined as follows.

Definition 2.5. *Let $\vec{e} = (e_1, \ldots, e_{k-1}) \in E^{k-1}$ be a vector, and let $1 \le i < j \le k$. Define the mapping $\pi_{\vec{e}, j \to i} : L^{k-j} \times R^{j-1} \to L^{k-i} \times R^{i-1}$ as*

$$\begin{array}{rl} & (l_1, \ldots, l_{k-j}, r_{k-j+1}, \ldots, r_{k-1}) \\ \mapsto & (l_1, \ldots, l_{k-j}, \pi_{e_{k-j+1}}(r_{k-j+1}), \ldots, \\ & \pi_{e_{k-i}}(r_{k-i}), r_{k-i+1}, \ldots, r_{k-1}). \end{array}$$

It is not hard to see that the above definition preserves smoothness in the Layered Label-Cover instances.

Lemma 2.6. *For any (J, ξ)-smooth k-layered Label-Cover instance (U, V, E, L, R, Π), positive integers $1 < i \le k$, set of labelings $S \subseteq L^{k-i} \times R^{i-1}$ for vertex tuples with $|S| < J$, we have*

$$\Pr_{\vec{e} \sim E^{k-1}}[|\pi_{\vec{e}, i \to 1}(S)| < |S|] < \xi.$$

Proof. For any $\vec{e} \in E^{k-1}$ and $S = (S_1, \cdots, S_{k-1}) \subseteq L^{k-i} \times R^{i-1}$, observe that $|\pi_{\vec{e}, i \to 1}(S)| < |S|$ is equivalent to $\forall j \in \{k-i+1, \cdots, k-1\}, |\pi_{e_j}(S_j)| < |S_j|$. The result follows immediately. □

Combining all we have, we get the following hardness result for k-layered Smooth-Label-Cover problem.

Theorem 2.7. *For every constant $\eta, J, \xi, k > 0$, there is some constant $G(\eta, J, \xi, k) < \infty$ such that for (J, ξ)-smooth k-layered Label-Cover instances with $|R| \ge G(\eta, J, \xi, k)$, it is NP-hard to distinguish between the following two cases:*

YES: *There exist assignments $\sigma_m : U^{k-m} \times V^{m-1} \to L^{k-m} \times R^{m-1} (1 \le m \le k)$, such that for all $\vec{e} = (e_1, \ldots, e_{k-1}) \in E^{k-1}$ and all $i, j, 1 \le i < j \le k$, it holds that*

$$\begin{array}{rl} & \pi_{\vec{e}, j \to i}(\sigma_j(u_1, \ldots, u_{k-j}, v_{k-j+1}, \ldots, v_{k-1})) \\ = & \sigma_i(u_1, \ldots, u_{k-i}, v_{k-i+1}, \ldots, v_{k-1}). \end{array}$$

NO: *There are no two integers l and h ($1 \le l < h \le k$) such that there exist functions $P_h : U^{k-h} \times V^{h-1} \to L^{k-h} \times R^{h-1}$ and $P_l : U^{k-l} \times V^{l-1} \to L^{k-l} \times R^{l-1}$, such that for more than η fraction of $(e_1, \ldots, e_{k-1}) \in E^{k-1}$, we have*

$$\begin{array}{rl} & P_l(u_1, \ldots, u_{k-l}, v_{k-l+1}, \ldots, v_{k-1}) \\ = & \pi_{\vec{e}, h \to l}(P_h(u_1, \ldots, u_{k-h}, v_{k-h+1}, \ldots, v_{k-1})). \end{array}$$

Proof. The proof is similar to [8]. The completeness case is straightforward. For soundness, suppose there exists $1 \leq l < h \leq k$ and functions P_l, P_h, such that for more than η fraction of $(e_1, \ldots, e_{k-1}) \sim E^{k-1}$, we have

$$\begin{aligned} & P_l(u_1, \ldots, u_{k-l}, v_{k-l+1}, \ldots, v_{k-1}) \\ = \; & \pi_{\vec{e},h\to l}(P_h(u_1, \ldots, u_{k-h}, v_{k-h+1}, \ldots, v_{k-1})). \end{aligned}$$

Then there is a way to fix $e_1, \ldots, e_{h-1}, e_{h+1}, \ldots, e_k$ such that the probability that the above holds is no less than η. We conclude the proof by noting that the restriction of P_l and P_h on the h-th coordinate gives a labeling with value at least η for the original Label-Cover instance. □

2.2 PCP Reductions

In this section, we describe a reduction from Label-Cover which is now the standard technique in hardness of approximation. We also discuss direct sums of PCPs introduced by Chan [5].

Consider a Boolean predicate P of arity w. The reduction typically translates labelings for $u \in U$ and $v \in V$ to $2^{|L|}$ and $2^{|R|}$ Boolean variables, respectively. These variables are viewed as functions $f^u : \{-1,1\}^{|L|} \to \{-1,1\}$ and $g^v : \{-1,1\}^{|R|} \to \{-1,1\}$. We require that these functions are folded, that is, for any $x \in \{-1,1\}^{|L|}$, $y \in \{-1,1\}^{|R|}$, $f^u(-x) = -f^u(x)$ and $g^v(-y) = -g^v(y)$. For each pair of queries $(x, -x)$, we select one of them. If x is selected, then when $f(-x)$ is needed we return $-f(x)$ instead. Hence in the actual reduction we only use $2^{|L|-1}$ Boolean variables for each $u \in U$ and $2^{|R|-1}$ variables for each $v \in V$. This is also why we need to allow negated literals in the CSP instances. In a correct proof for a satisfiable Label-Cover instance, the functions are long codes for the corresponding labelings of u and v, that is, setting $f^u(x) = x_{\sigma_U(u)}$, and $g^v(y) = y_{\sigma_V(v)}$.

For an edge $\{u, v\}$ in the Label-Cover, we sample *queries*

$$(x^{(1)}, \cdots, x^{(m)}, y^{(m+1)}, \cdots, y^{(w)})$$

according to some carefully chosen *test distribution* $\mathcal{T}$. The distribution $\mathcal{T}$ has the property that for any $l \in L$ and $r \in R$ such that $\pi_{(u,v)}(r) = l$, the predicate P accepts $(x_l^{(1)}, \cdots, x_l^{(m)}, y_r^{(m+1)}, \cdots, y_r^{(w)})$ with probability 1 (or $1-\varepsilon$ for some small constant ε if we are considering non-perfect completeness).

Let the value of an edge be the following expectation

$$\mathop{\mathbf{E}}_{(x^{(1)},\cdots,x^{(m)},y^{(m+1)},\cdots,y^{(w)})\sim\mathcal{T}} \left[P(f^u(x^{(1)}), \cdots, f^u(x^{(m)}), g^v(y^{(m+1)}), \cdots, g^v(y^{(w)})\right]. \tag{2}$$

Observe that in the completeness case where the Label-Cover instance has an assignment that satisfies all the edges, then setting f^u and g^v to the long code of the labelings would give value 1 (or close to 1) for the above expectation.

In the soundness case, of course the f^u's and g^v's are not guaranteed to be long codes. Typically, when proving approximation resistance, we start the analysis by taking the Fourier expansion of predicate P in (2). The constant term in the expansion is exactly the density of P. We then show that if for some non-constant terms we have that $|\mathbf{E}[\prod f^u \prod g^v]| \geq \delta$ for some small constant $\delta > 0$, then we can find a *good* labeling for the Label-Cover instance we started with, allowing us to distinguish

between the YES case and the NO case. In some cases we can show that all possible non-constant terms — including those that do not appear in the expansion of P — are small, and this implies that predicate P is useless in the sense of [3], a stronger notion of inapproximability.

It is not hard to adapt the above reduction to k-Layered Label-Covers. Instead of encoding the labelings of single vertices as long codes, we encode labelings for the hybrid vertex tuples. The rest of the analysis is similar.

In [5], Chan introduced direct sum of PCPs to get improved hardness of approximation results and proved the first general criterion for approximation resistant predicate without assuming the UGC. The main result is the following.

Theorem 2.8 ([5]). *Let $k \geq 3$ be an integer, G a finite abelian group, and C a balanced pairwise independent subgroup of G^k. It is NP-hard to approximate Additive-CSP(C) better than $|C|/|G|^k + \varepsilon$ for any constant $\varepsilon > 0$.*

In the Boolean case, we have $G = \{-1, 1\}$ with product "$\cdot$" as the operation, C consists of the accepting assignments of some predicate P, and Additive-CSP(C) is exactly Max-P. Note that the balanced pairwise independence condition is similar to the condition of Austrin and Mossel [4].

The main idea in Chan's proof is that instead of sampling one edge, we sample c edges for some c to be determined. We also have c test distributions $\mathcal{T}_1, \ldots, \mathcal{T}_c$ corresponding to the edges we sampled, where each distribution $\mathcal{T}_i$ satisfies the same requirement as we had for $\mathcal{T}$ above. We sample queries $\{(x^{(1,i)}, \cdots, x^{(k,i)})\}_{i\in[c]}$. Note that for the same $m \in [k]$, whether the query $x^{(m,i)}$ is from $\{-1,1\}^{|L|}$ or $\{-1,1\}^{|R|}$ may vary among different $i \in [c]$ and is an important design choice. Depending on the edges we have sampled, we choose k functions to query as in the classical setting (some of those functions might be the same one). The functions now have larger domains, since the j-th function would take $\{x^{(j,i)}\}_{i\in[c]}$ as input. We also require that the functions are folded in each individual test — for any query $\{x^{(j,i)}\}_{i\in[c]}$ and $i \in [c]$ we have

$$f(x^{(j,1)}, \cdots, -x^{(j,i)}, \cdots, x^{(j,c)}) = \quad - f(x^{(j,1)}, \cdots, x^{(j,i)}, \cdots, x^{(j,c)}).$$

The intended solution now is that each function is the product of functions for the individual PCPs, and the functions in the individual PCPs are long codes of some legitimate labelings. Similar to the classical approach, we take the answers to the queries and accept if predicate P accepts.

It is not hard to see that for the completeness to hold, we would need that the set of satisfying assignments of P has some group structure — the entry-wise product of two satisfying assignments is still a satisfying assignment. Observe that the $\textsc{Hadamard}_K$ predicate satisfies this property. On the soundness side, we need to bound each term in the Fourier expansion of (2). The important observation (Lemma 5.3 in [5]) is that the absolute value of these terms are bounded by the absolute value of these terms in each individual PCP. Hence, all we need is to show that a term in (2) is small in at least one of the PCPs in the direct sum unless there is good labeling.

2.3 Efron-Stein Decomposition, Influence and Correlation

In this section, we recall basic notions from Fourier analysis, influence, the Bonami-Beckner operator, and correlation of correlated probability spaces.

Let (Ω, μ) be a finite probability space with $|\Omega| = q$. We assume that $\mu(x) > 0$ for every $x \in \Omega$. Let $\chi_0, \cdots, \chi_{q-1} : \Omega \to \mathbb{R}$ be an orthonormal basis for $L^2(\Omega, \mu)$ w.r.t. scalar product under μ. Let this basis be such that $\chi_0 = \mathbf{1}$ the identical one function. For $\sigma \in \mathbb{Z}_q^n$, define

$$\chi_\sigma(x_1, \cdots, x_n) = \prod_{i=1}^{n} \chi_{\sigma_i}(x_i).$$

Then $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$ forms an orthonormal basis for $L^2(\Omega^n, \mu^{\otimes n})$, and every function $f \in L^2(\Omega^n, \mu^{\otimes n})$ can be written as

$$f(x) = \sum_{\sigma \in \mathbb{Z}_q^n} \hat{f}(\sigma)\chi_\sigma(x).$$

We also make extensive use of the following Efron-Stein decomposition [7, 21]

Theorem 2.9. *Any function $f \in L^2(\Omega^n, \mu^{\otimes n})$ can be uniquely decomposed as*

$$f(x) = \sum_{S \subseteq [n]} f_S(x),$$

where

- *Function $f_S(x)$ depends only on $x_S = \{x_i | i \in S\}$.*
- *For every $S, T \subseteq [n]$, $S \setminus T \neq \emptyset$, $x' \in \Omega^n$, it holds that*

$$\mathbf{E}[f_S(x) | x_T = x'_T] = 0.$$

For $\sigma \in \mathbb{Z}_q^n$, let $Set(\sigma) = \{i | \sigma_i > 0\}$, and let $|\sigma| = |Set(\sigma)|$. It is easily verified that the Efron-Stein decomposition is related to the Fourier decomposition as follows

$$f_S(x) = \sum_{\substack{\sigma \in \mathbb{Z}_q^n \\ Set(\sigma) = S}} \hat{f}(\sigma)\chi_\sigma(x).$$

A useful notion of function is the influence of a coordinate.

Definition 2.10. *For $f \in L^2(\Omega^n, \mu^{\otimes n})$, $i \in [n]$, the influence of i on f is defined as*

$$\mathbf{Inf}_i(f) = \mathop{\mathbf{E}}_{x_{[n]\setminus i}} [\mathbf{Var}_{x_i}[f(x)]].$$

Note that when we refer to influence, it is always with respect to the underlying probability space $(\Omega^n, \mu^{\otimes n})$. We have the following characterization of influence in terms of Fourier decomposition and Efron-Stein decomposition.

Proposition 2.11. *For $f \in L^2(\Omega^n, \mu^{\otimes n})$ and $i \in [n]$,*

$$\mathbf{Inf}_i(f) = \sum_{\substack{\sigma \in \mathbb{Z}_q^n \\ i \in Set(\sigma)}} \hat{f}(\sigma)^2 = \sum_{S \ni i} \mathbf{E}[f_S^2].$$

Let the total influence $\mathbf{Inf}(f) = \sum_{i \in [n]} \mathbf{Inf}_i(f)$ be the sum of influences of all coordinates on f.

We next recall the Bonami-Beckner operator, or noise operator.

Definition 2.12. *Let $0 \leq \gamma \leq 1$. The Bonami-Beckner operator T_γ is a linear operator mapping $f \in L^2(\Omega^n, \mu^{\otimes n})$ to $T_\gamma f$ as follows*

$$(T_\gamma f)(x) = \mathbf{E}[f(y)],$$

where y is sampled by setting each bit independently to $y_i = x_i$ with probability $1-\gamma$, and otherwise sampled according to μ with probability γ.

Again we have the following Fourier/Efron-Stein characterization of T_γ.

Proposition 2.13. *For any $f \in L^2(\Omega^n, \mu^{\otimes n})$ and $0 \leq \gamma \leq 1$,*

$$T_\gamma f = \sum_{\sigma \in \mathbb{Z}_q^n} (1-\gamma)^{|\sigma|} \hat{f}(\sigma) \chi_\sigma.$$

We define noisy influence as $\mathbf{Inf}_i^{(\gamma)}(f) = \mathbf{Inf}_i(T_\gamma f)$, and similarly $\mathbf{Inf}^{(\gamma)}(f) = \sum \mathbf{Inf}_i^{(\gamma)}(f)$. We have the following bound for the total noisy influence of functions with range $[-1, 1]$.

Proposition 2.14. *For any $f : \Omega^n \to [-1, 1]$ and $0 < \gamma \leq 1$, we have*

$$\mathbf{Inf}^{(\gamma)}(f) = \sum_{i \in [n]} \mathbf{Inf}_i^{(\gamma)}(f) \leq \gamma^{-1}.$$

The following concept of *lifted* functions is useful in the context of projection games.

Definition 2.15. *Given function $f : \Omega^{nd} \to \mathbb{R}$ and d-to-1 mapping $\pi : R \to L$, define the lifted version of f as $\overline{f}^\pi : (\Omega^d)^n \to \mathbb{R}$ as naturally induced by π*

$$\overline{f}^\pi(\overline{x}) = f(x),$$

where $\overline{x}$ satisfies $\overline{x}_{r,t} = x_{(r,t)}$ for $r \in L, t \in [d]$.

In terms of influence, we have the following relation between f and $\overline{f}$.

Proposition 2.16. *For any r, we have*

$$\mathbf{Inf}_r(\overline{f}) \leq \sum_{r' : \pi(r') = r} \mathbf{Inf}_{r'}(f).$$

Proof. The claim follows by applying Proposition 2.11 and comparing the terms. □

The correlation for correlated probability spaces was introduced by Mossel [21]. Given a probability measure μ defined on $\Omega \times \Psi$, we say that Ω and Ψ are correlated spaces, and we use $(\Omega \times \Psi, \mu)$ to denote correlated spaces and the corresponding measure. We use the following definition of correlation.

Definition 2.17. *Let $(\Omega \times \Psi, \mu)$ be a correlated probability space, 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*

$$\rho(\Omega, \Psi; \mu) = \max_{\substack{f : \Omega \to \mathbb{R} \\ g : \Psi \to \mathbb{R}}} \left\{ |\mathbf{E}[fg]| \mid \mathbf{E}[f] = 0, \mathbf{E}[f^2] \leq 1, \mathbf{E}[g] = 0, \mathbf{E}[g^2] \leq 1 \right\},$$

where the expectation $\mathbf{E}[fg]$ is under μ, and $\mathbf{E}[f]$, $\mathbf{E}[f^2]$, $\mathbf{E}[g]$ and $\mathbf{E}[g^2]$ are under marginals of μ on corresponding spaces.

A useful fact for bounding correlation of probability spaces from [21] is that the correlation of a product of correlated probability space is equal to the maximum correlation among the individual correlated spaces (excluding empty components).

Lemma 2.18. *Let $\{(\Omega_i, \Psi, \mu_i)\}$ be a set of correlated probability spaces, then*

$$\rho\left(\prod_i \Omega_i, \prod_i \Psi_i; \prod_i \mu_i\right) \leq \max_i \rho(\Omega_i, \Psi_i; \mu_i).$$

We also need the following lemma when analyzing correlations. Intuitively, if we can decompose μ into a convex combination of two distributions and we can bound the correlation between Ω and Ψ in both of sub-distributions by some constant c, then barring special cases it seems reasonable that the correlation $\rho(\Omega, \Psi; \mu)$ should also be bounded by some function of c. More formally, we have the following lemma from [31].

Lemma 2.19. *Let $(\Omega, \Psi, \delta\nu + (1-\delta)\nu')$ be a correlated space such that the marginal distribution of at least one of Ω and Ψ is identical on both ν and ν'. Then*

$$\rho(\Omega, \Psi; \delta\nu + (1-\delta)\nu') \leq \sqrt{\delta\rho(\Omega, \Psi; \nu)^2 + (1-\delta)\rho(\Omega, \Psi; \nu')^2}.$$

Next we recall the definition of the conditional expectation operator.

Definition 2.20. *Let $(\Omega \times \Psi, \mu)$ be two correlated spaces. The conditional expectation operator $\mathcal{U}$ associated with (Ω, Ψ) is the operator mapping $f \in L^2(\Psi, \mu)$ to $Uf \in L^2(\Omega, \mu)$ by $(\mathcal{U}f)(x) = \mathbf{E}[f(Y)|X = x]$ for $x \in \Omega$ and $(X, Y) \in \Omega \times \Psi$ is distributed according to μ.*

An important property we need in the analysis is that the Efron-Stein decomposition commutes with the conditional expectation operator.

Proposition 2.21 ([20])**.** *Let $(\Omega \times \Psi, \mu) := (\prod \Omega_i \times \prod \Psi_i, \otimes \mu_i)$ be correlated space and let $\mathcal{U} := \otimes \mathcal{U}_i$ be the conditional expectation operator associated with Ω and Ψ. Suppose $f \in L^2(\Psi)$ has Efron-Stein decomposition $f(x) = \sum_{S \subseteq [n]} f_S(x_S)$. Then the Efron-Stein decomposition of $\mathcal{U}f$ satisfies $(\mathcal{U}f)_S = \mathcal{U}(f_S)$ for $S \subseteq [n]$.*

The following result shows that in the above setting, if the correlations between all Ω and Ψ are less than 1, then the L^2 norms of the high-degree terms of $\mathcal{U}f$ are small.

Proposition 2.22 ([20])**.** *Assume the setting of Proposition 2.21 and that for all i, we have $\rho(\Omega_i, \Psi_i; \mu_i) \leq \rho_i$. Then for all f, we have $\|\mathcal{U}(f_S)\|_2 \leq \left(\prod_{i \in S} \rho_i\right) \|f_S\|_2$.*

3 The Predicate, the PCP and Outline of Proof

Given the soundness parameter ε, the starting point of our reduction is a k-layer (J, ξ)-Smooth-Label-Cover, where J and ξ are constants solely dependent on ε that we will specify later.

The predicate. Fix some k, and let $[k] := \{1, 2, \ldots, k\}$. Let $\mathscr{S}_3 := \{S \subseteq [k] || S| = 3\}$, $\mathscr{S}_1 := \{S \subseteq [k] || S| = 1\}$. The predicate is on variables $\{x^{(S)}\}_{S \in \mathscr{S}_1 \cup \mathscr{S}_3}$ taking values from $\{-1, 1\}$. We call the variables $x^{(\{i\})}$ singleton variables and the remaining ones parity check variables. The predicate accepts if there exists $\vec{w} \in$

$\{-1,1\}^{\mathscr{S}_1 \cup \mathscr{S}_3}$ such that the number of -1's in $\vec{w}$ is no more than k, and $w_S x^{(S)} \cdot \prod_{i \in S} w_{\{i\}} x^{(\{i\})} = 1$ for all $S \in \mathscr{S}_3$.

We can view $\vec{w}$ as an error vector, and the predicate accepts inputs that are no more than Hamming distance k away from an assignment that satisfies all parity checks.

The predicate is on $k + \binom{k}{3}$ variables, and it has $O(2^k \cdot \binom{\binom{k}{3}+k}{k}) = 2^{O(k \log k)}$ accepting inputs, thus the density (assuming the predicate has arity K) is $2^{\tilde{O}(K^{1/3})}/2^K$, where the $\tilde{O}$ hides logarithmic factors.

Outline of proof. Before going into details about the construction of our PCPs, we first give an overview of our proof and explain the intuition behind the construction.

Our PCP design is based on Chan's idea of direct sums of PCPs [5] as described in Section 2.2. We prove that all non-constant terms in the Fourier expansion of (2) are small.

One crucial difference between Chan's proof and ours is that we require perfect completeness. This means that sometimes there would be perfect correlation between certain queries which makes it possible for provers to find good cheating strategies. In Chan's proof as well as in many related results where perfect completeness is not required, one can usually break this correlation by applying some independent noise to each query bit. However, in the case of perfect completeness, we cannot afford perturbing each bit independently, and thus we need to take extra care when designing test distributions. That is the main reason our predicate accepts inputs that *almost* satisfy all $\binom{k}{3}$ linear constraints. In some sense, these extra accepting inputs serve as noise that breaks up perfect correlations.

Another important property that Chan uses is the "group" structure of the predicate. This makes it relatively easy to take direct sums of a large number of PCPs, each handling a small number of non-constant terms from (2), without worrying too much about the completeness of the resulting PCP. Our predicate, however, does not satisfy this property due to the extra noise we added. It is certainly possible that if we take the sum of two assignments that are of distance k away from assignments that satisfies all linear equations, we end up with something that is distance $2k$ away from an assignment that satisfies all linear constraints, and that would break perfect completeness. To avoid this situation, we limit both the number of PCPs in the direct sum and in each PCP the distance from an assignment that satisfies all linear constraints. More specifically, in our construction the queries to each PCP are generated such that if the provers (of each individual PCP) answer according to some consistent long code, then the answers is at most distance 1 away from an assignment that satisfies all linear equations. When taking direct sum of the k PCPs, an answer that is the direct sum of k long codes would give us an answer that would be accepted by our predicate.

It remains to find a number of suitable PCPs. If we try to generalize previous approaches, for example those in [26, 8], to larger predicates such as $\textsc{Hadamard}_K$, one of the main adversarial strategies that we need to consider is that of inconsistent long codes. For example, consider a predicate P on variable $(x_1, \cdots, x_k)$ and a simple PCP reduction where we sample an edge $\{u, v\}$ and query functions f^u and g^v according to some test distribution $\mathcal{T}$ as described in Section 2.2. For simplicity, assume that the query to f^u corresponds to input variable x_1, and the remaining queries are on g^v. Suppose further that for a $\frac{1}{2} + \delta$ fraction of the accepting inputs of P, we have $x_2 x_3 x_4 = 1$ (both $\textsc{Hadamard}_K$ and the predicate we are studying in

this work have properties similar to this.) Let g^v be long code for some arbitrary label $r \in R$. Observe that the non-constant term $g^v(x_2)g^v(x_3)g^v(x_4)$ will always have expectation roughly order of δ simply due to the requirements on $\mathcal{T}$. In this case, we get a large non-constant term but it does not help us find a consistent labeling for Label-Cover. A similar argument can be made for Multi-layered Label-Cover. Chan's construction in [5] solves this problem by making sure that for each term, in at least one of the many PCPs in the direct sum the queries are on different functions. As discussed before, since we are aiming for fewer PCPs in the direct sum, it would be good if each PCP can carry out as many consistency checks as possible, and Multi-layered Label-Cover becomes a very natural choice. We also need to decide which query should be in which layer for each PCP so that we do not miss any sets of variables that has linear relations. This is mostly done in Section 4.1.

Now we describe the PCPs in more details.

The PCPs. Let $\mathcal{C} = \{\sigma_0, \ldots, \sigma_{k-1}\}$ be the set of cyclic permutations on $[k]$. The permutation σ_i maps i to k, $i+1$ to 1, and so on. We identify 0 with k, and thus σ_0 is the identity permutation. Each permutation corresponds to a PCP for a k-Layer Label-Cover instance, and the permutation decides which query should be in which layer in the Multi-Layered Label-Cover. As stated above, the final proof is the direct sum of these k PCPs.

We now describe the i-th PCP. It is based on a k-layered Label-Cover instance, and there are $k + \binom{k}{3}$ queries, one corresponding to an input variable. We denote the queries as $x^{(S)}$. For $S \in \mathscr{S}_1 \cup \mathscr{S}_3$, define $m_i(S) := \max \sigma_i(S)$ to be the maximum element of S under permutation σ_i. The query $x^{(S)}$ is in layer $m_i(S)$. Let $\mathcal{V}_i(S) := U^{k-m_i(S)} \times V^{m_i(S)-1}$ be the set of vertex tuples in layer $m_i(S)$. The proof has a function for each vertex tuple in $\mathcal{V}_i(S)$, and the input to the functions are $\{-1, 1\}$ strings indexed by the possible labelings $L^{k-m_i(S)} \times R^{m_i(S)-1}$ in layer $m_i(S)$. We denote the domain of the functions as $X_i^{(S)}$. In a correct proof of a correct labeling, the function would be a long code encoding a proper labeling for all vertices in the tuple. As described in Section 2.2, we require that all functions are folded.

The test distributions. We first define the test distributions for each individual PCP.

Fix $i \in [k]$ and consider the i-th PCP. For notational simplicity we omit i in the subscript for now. We first independently sample $k-1$ edges $\vec{e} = \{e_1, \ldots, e_{k-1}\}$. For $S \in \mathscr{S}_1$, sample $x^{(S)} \in X^{(S)}$ uniformly at random. For $S = \{s_1, s_2, s_3\} \in \mathscr{S}_3$, let $m = m(S)$ be the layer in which query $x^{(S)}$ is located, $m_j = m(s_j)$ for $j = 1, 2, 3$ be the layer query $x^{(\{s_j\})}$ is in, and set $x_r^{(S)} = \prod_{j=1}^{3} x^{(\{s_j\})}_{\pi_{\vec{e},m\to m_j}(r)}$ for all possible labelings $r \in L^{k-m} \times R^{m-1}$.

We then make use of the extra inputs allowed by the predicate to add some "noise" to the distributions. As discussed above, the resulting distribution must have the property that the output obtained by applying some consistent long codes is at most distance 1 away from an assignment that satisfies all $\binom{k}{3}$ equations. The idea is to perturb one of the $x^{(S)}$'s. For each $r \in L^{k-1}$, pick a uniformly random set $N_r \in \mathscr{S}_1 \cup \mathscr{S}_3$, and for each $t \in \pi^{-1}_{\vec{e},m(N_r)\to 1}(r)$, set $x_t^{(N_r)}$ to a uniform random bit independently with probability $1/2$.

We denote the test distribution by $\mathcal{T}$. For each $r \in L^{k-1}$, let $\mathcal{T}_r$ be the marginal distribution of the bits that map to r under $\pi_{\vec{e},l\to 1}$ for all $l \in [k]$. Observe that we have $\mathcal{T} = \otimes_{r \in L^{k-1}} \mathcal{T}_r$.

Let us start by analyzing the standard completeness case.

Lemma 3.1. *For any sampling of edges, let $f^{(S)}$ be the functions we are querying, and let $x^{(S)}$ be the corresponding queries. If the k-layered Label-Cover instance has a labeling that satisfies all the edges, then we can find $f^{(S)}$'s such that the answers $\{f^{(S)}(x^{(S)})\}_{S\in\mathscr{S}_1\cup\mathscr{S}_3}$ is at most Hamming distance 1 away from an assignment that satisfies all linear constraints on 3 singleton variables and 1 parity check variable.*

Proof. The argument is similar to a standard completeness argument.

Fix a labeling that satisfies all the edges. The proof in the PCP consists of long codes encoding the labeling of all hybrid vertex tuples.

Let $r \in L^{k-1}$ be the labeling for the vertex tuple in layer 1. The answers we get from the long codes is the same as returning one bit from each query generated according to $\mathcal{T}_r$. The claim follows by observing that for each tuple of bits produced as above, either it already satisfies all linear constraints, or it would satisfy all linear constraints after we flip the N_r-th bit. □

Denote the test distribution of the i-th PCP defined above as $\mathcal{T}_i$. The distribution of the final composed PCP is simply the product of the individual test distributions $\otimes_{i=1}^{k}\mathcal{T}_i$. The verifier samples the edges and the inputs to the functions, queries the functions (those that correspond to the chosen vertex tuples) and accepts if the answers returned by the functions are accepted by the predicate.

It is not hard to see from above discussions that the above PCP has perfect completeness.

Lemma 3.2. *If the k-layered Label-Cover instance has a labeling that satisfies all edges, then there exists a set of functions $\{f^{(S)}\}$ such that after querying $\{f^{(S)}\}$ the verifier accepts with probability 1.*

Proof. We let our final proof be the product of proofs of the k individual PCPs given by Lemma 3.1. Since the answer for each proof is at most distance 1 away from an assignment that satisfies all linear constraints, their product is at most distance k away, which is exactly what the verifier (and our predicate) accepts. □

4 Soundness

In this section, we analyze the soundness of our PCP. We set $\varepsilon_1 = \varepsilon/(7k^3+1)$, $\xi = \varepsilon_1^2$, $\rho_0 = 1 - 1/4\binom{k}{3}$, $J = 2\lceil \log_{\rho_0} \varepsilon_1 \rceil$, and γ such that $1-(1-\gamma)^{J/2} < \varepsilon_1$. Note that this gives $\rho_0^{J/2} \le \varepsilon_1$, and that all parameters depend only on k and ε. Also $\gamma < \varepsilon$.

As discussed in Section 3, we would like to prove that for all $\mathscr{S} \neq \emptyset$, the expectation

$$\mathbf{E}\left[\prod_{S\in\mathscr{S}} f^{(S)}(x^{(S)})\right], \tag{3}$$

is small unless there is good labeling.

Remark. The functions $f^{(S)}$ actually depend on the underlying edges we sampled. For notational convenience we suppress this dependency and save another layer of subscripts (of subscripts of subscripts).

As discussed in previous sections, we need to show that for each non-constant term, there is at least one PCP among those in the direct sum, such that if the expectation of the term under the PCP is large, we can find a good labeling for the underlying label cover instance by looking at the functions f restricted to that PCP.

Formally, we have the following lemma which is a reformulation of Lemma 5.3 in Chan [5].

Lemma 4.1. *Let $\mathcal{T} = \otimes_{i=1}^{k} \mathcal{T}_i$, where $\mathcal{T}_i$ is the test distribution for the i-th PCP. Suppose for some $\mathscr{S} \neq \emptyset$, we have*

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] \right| = \delta,$$

then for any $i \in [k]$, there exists functions $g^{(S)}$ whose inputs are query bits to the i-th PCP, such that

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}_i} \left[\prod_{S \in \mathscr{S}} g^{(S)}(x^{(S)}) \right] \right| \geq \delta.$$

Given $f^{(S)}$, we find $g^{(S)}$ by fixing query bits that are not in the i-th PCP in a way that does not lower the expectation.

Thus to bound each term, we need to carefully find an i, such that the test restricted to the i-th PCP has small expectation. We show how to choose such i in Section 4.1. We would be back to the traditional setting with Label-Covers and dictatorship testing from then on. In Section 4.2, we show that we can instead look at the distribution where each individual bit is further perturbed independently by some random noise. Then we show in Section 4.3 how to apply an invariance-type theorem from [32] in this new setting to get our soundness result.

4.1 Permutation Covering

Our k PCPs use cyclic permutations $C \in \mathcal{C}$ to decide the layer of each query and the inputs to the corresponding function. We first give a general definition of the crucial property we need from such sets of permutations.

Definition 4.2. *Let $\mathcal{P}$ be a set of permutations on $[k]$. We say that $\mathcal{P}$ covers $\mathscr{S}_1 \cup \mathscr{S}_3$ if for all $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, there exists a permutation $\sigma \in \mathcal{P}$, some $j, l_0 \in [k]$, such that*

$$\left| \{S \in \mathscr{S} | j \in S, \max \sigma(S) = l_0\} \right| \text{ is odd.}$$

We now reformulate the above definition and prove a necessary and sufficient condition for general sets of permutations $\mathcal{P}$ to cover $\mathscr{S}_1 \cup \mathscr{S}_3$.

For each set $S \in \mathscr{S}_1 \cup \mathscr{S}_3$, we construct a Boolean vector $v_S^{\mathcal{P}}$ as the following: the elements in the vector are indexed by a tuple $(i, l, j) \in [|\mathcal{P}|] \times [k] \times [k]$, and $v^{\mathcal{P}}_{S,(i,l,j)} = 1$ if $\max \sigma_i(S) = l$ and $j \in S$, and $v_{S,(i,l,j)} = 0$ otherwise.

Proposition 4.3. *The set of permutations $\mathcal{P}$ covers $\mathscr{S}_1 \cup \mathscr{S}_3$ iff the vectors $\{v_S^{\mathcal{P}}\}_{S \in \mathscr{S}_1 \cup \mathscr{S}_3}$ are linearly independent in $\mathbb{F}_2$.*

Proof. If the set $\mathcal{P}$ does not cover $\mathscr{S}_1 \cup \mathscr{S}_3$, then there exists a set $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, such that for any permutation $\sigma_i \in \mathcal{P}$ and $j, l_0 \in [k]$, we have that

$$\left| \{S \in \mathcal{P} | S \ni j, \max \sigma_i(S) = l_0\} \right| \text{ is even.}$$

Observe that for any $S \in \mathscr{S}_1 \cup \mathscr{S}_3$, the segment of $v_S^{\mathcal{P}}$ indexed by (i, l) for some fixed i and l is all zero if $\max \sigma_i(S) \neq l$, and otherwise it is exactly the character vector of

the set S. Therefore the above is equivalent to saying that for any $i \in [|\mathcal{P}|]$ and l_0, we have

$$\sum_{S \in \mathscr{S}} v^{\mathcal{P}}_{S,(i,l_0)} = 0,$$

where the summation is modulo 2. Since the above holds for all i and l_0, we have

$$\sum_{S \in \mathscr{S}} v^{\mathcal{P}}_{S} = 0,$$

or the vectors $\{v^{\mathcal{P}}_S\}_{S \in \mathscr{S}}$ are linearly dependent.

Note that all the above steps are equivalent statements. Thus the other direction also holds. □

As a side note, we can see from the above argument that it is necessary to have $\Omega(k)$ permutations in order to cover $\mathscr{S}_1 \cup \mathscr{S}_3$, because otherwise we would have $\Theta(k^3)$ vectors of dimension $o(k^3)$ and thus they could not be linearly independent.

We now prove that the set of all cyclic permutations $\mathcal{C} = \{\sigma_0, \ldots, \sigma_{k-1}\}$ covers $\mathscr{S}_1 \cup \mathscr{S}_3$.

Lemma 4.4. *The set of all cyclic permutations $\mathcal{C} = \{\sigma_0, \ldots, \sigma_{k-1}\}$ covers $\mathscr{S}_1 \cup \mathscr{S}_3$.*

Proof. For any given set $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, we show how to find the cyclic permutation σ and indices $j, l_0 \in [k]$ required in Definition 4.2.

For a set $S \in \mathscr{S}_1 \cup \mathscr{S}_3$, let

$$span(S) = \min_{\sigma_i \in \mathcal{C}} \max \sigma_i(S) - \min \sigma_i(S),$$

that is, the minimum distance between the largest and the smallest element under cyclic permutations. Note that for singleton sets $S \in \mathscr{S}_1$, we have $span(S) = 0$.

For a given set $\mathscr{S}$, let $S \in \mathscr{S}$ be a set with minimum span in $\mathscr{S}$ where we break ties arbitrarily. Pick i_0 such that $\sigma_{i_0}(S)$ contains 1 and $span(S) + 1$ as its minimum and maximum element. Let $\sigma := \sigma_{i_0}$ be the permutation we want, and let $l_0 = span(S) + 1$.

Now we select j. If $span(S) = 0$, then let $j = \sigma^{-1}(1)$ and we are done. This is because for any non-singleton set S', $\max \sigma(S) > 1$, and for any singleton set $S'' \neq S$, clearly $\sigma(S'') \neq \sigma(S)$. Thus S would be the only set containing j with $\max \sigma(S) = 1 = l_0$.

If $span(S) \neq 0$, then S has three elements, and there is no singleton set in $\mathscr{S}$. If there is any other non-singleton set $S'' \in \mathscr{S}$ with $\max \sigma(S'') = span(S) + 1$, then $\sigma(S'')$ and $\sigma(S)$ have the same maximum and minimum element, namely $span(S)+1$ and 1. That leaves us with the middle element. But since $S \neq S''$, the middle element must be different, so each of them appear only in one set, and setting j to the inverse of any of the middle elements under σ would work. Otherwise we take $j = \max S$. □

For $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, we consider the PCP corresponding to the cyclic permutation $\sigma_i \in \mathcal{C}$ covering $\mathscr{S}$ given by Lemma 4.4. We denote the PCP as PCP_i. As discussed before, we only need to show that if (3) is large even when restricted to PCP_i, we can find a good labeling for the Label-Cover instance we started with.

For notational simplicity, we only prove the case where $i = 0$, that is, for the identity permutation σ_0. Arguments for general cyclic permutations are entirely symmetric.

4.2 Introducing Independent Noise

In this section, we show that perturbing the queries does change the expectation of the terms by too much.

Formally, let $\mathcal{T}'_r$ be the distribution where we first sample according to $\mathcal{T}_r$, and then resample each bit independently with probability γ according to its marginal distribution in $\mathcal{T}_r$ — which in our case is uniform. Also define $\mathcal{T}' = \otimes_{r \in L^{k-1}} \mathcal{T}'_r$. We prove the following lemma which bounds the difference of expectation of (3) under $\mathcal{T}$ and $\mathcal{T}'$.

Lemma 4.5. *For any $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, we have*

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] - \mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] \right| < 7k^3 \varepsilon_1, \tag{4}$$

where $\varepsilon_1 = \varepsilon/(7k^3+1)$ is as defined at the beginning of Section 4.

Fix some $S_0 \in \mathscr{S}_1 \cup \mathscr{S}_3$. Let $\mathcal{T}^{(S_0)}$ be the distribution where under $\mathcal{T}$, we independently resample the bits in $x^{(S_0)}$ from the uniform distribution with probability γ. We first show in Lemma 4.6 below that the expectation under $\mathcal{T}$ is close to that under $\mathcal{T}^{(S_0)}$. Lemma 4.5 follows by applying similar arguments to all $x^{(S)}$'s one after another.

For $S \in \mathscr{S}_1 \cup \mathscr{S}_3$, let $m(S) = \max S$ be the maximum element in S. Recall that query $x^{(S)}$ is located in layer $m(S)$, and for $r \in L^{k-1}$, $\mathcal{T}_r$ is the distribution containing all bits in $\{x_t^{(S)} | S \in \mathscr{S}_1 \cup \mathscr{S}_3, \pi_{m(S)\to 1}(t) = r\}$, that is, the query bits that map to the same r. We use $\mathcal{T}_r^{(S_0)}$ to denote the marginal distribution of $\mathcal{T}^{(S_0)}$ on bits in $\{x_t^{(S_0)} | \pi_{m(S_0)\to 1}(t) = r\}$. Let $m = m(S_0)$.

Consider the difference of expectation between $\mathcal{T}$ and $\mathcal{T}^{(S_0)}$. If $f^{(S_0)}(x^{(S_0)})$ does not appear in the product, then there would be no difference. We now assume otherwise. The following lemma shows that introducing independent noise on one query does not change the expectation by too much.

Lemma 4.6. *For any $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, we have*

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] \right| < 7\varepsilon_1. \tag{5}$$

The proof follows the approach from [31], especially Lemma 3.15 through Lemma 3.17.

For notational simplicity let F' be the product of all terms but $f^{(S_0)}(x^{(S_0)})$ and we abbreviate $f^{(S_0)}$ as f. We use $\overline{X}^{(S_0)}$ to abbreviate $\prod_{S \in \mathscr{S}_1 \cup \mathscr{S}_3, S \neq S_0} X^{(S)}$. Similarly we define $\overline{X}_r^{(S_0)}$ for $r \in L^{k-1}$. The first step is to use Lemma 2.19 to bound the correlation $\rho(X^{(S_0)}, \overline{X}^{(S_0)}; \mathcal{T})$ and $\rho(X^{(S_0)}, \overline{X}^{(S_0)}; \mathcal{T}^{(S_0)})$. Since $\mathcal{T}$ is simply a product of $\mathcal{T}_r$ with different r's, by Lemma 2.18, we only need to bound $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mathcal{T}_r)$ and $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mathcal{T}_r^{(S_0)})$.

Claim 4.7. *For any $S_0 \in \mathscr{S}_3$, the correlation $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mathcal{T}_r)$ is upper-bounded by $1 - \frac{1}{4\binom{k}{3}} \triangleq \rho_0$. The same bound holds for $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mathcal{T}_r^{(S_0)})$.*

Proof. We divide $\mathcal{T}_r$ into two parts: (i) the set S_0 is chosen as N_r; or (ii) some set other than S_0 is chosen. It is not hard to verify that the marginal of $X_r^{(S_0)}$ after conditioning on either of them remains uniform and thus we can apply Lemma 2.19. Let μ be the conditional distribution assuming (i) happens, and ν be the one assuming (ii) happens. We have that $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \nu) = 1$. For the correlation of the other part, we have $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mu) = 1/2$, achieved by dictatorship functions. Therefore, the overall correlation is upper-bounded by $\sqrt{(1 - 1/\binom{k}{3}) + 1/\binom{k}{3} \cdot (1/2)^2} \leq \sqrt{1 - 1/2\binom{k}{3}} < 1 - 1/4\binom{k}{3}$.

Intuitively, the correlation under $\mathcal{T}_r^{(S_0)}$ could not exceed that under $\mathcal{T}_r$ since the noise we added are all independent. In particular, the part corresponding to $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \nu)$ becomes less than 1 due to lack of perfect correlation, and the part corresponding to $\rho(X_r^{(S_0)}, \overline{X}_r^{(S_0)}; \mu)$ remains the same. Thus the result follows by similar calculations as in $\mathcal{T}_r$. □

Take the Efron-Stein decomposition $f = \sum_{T \subseteq L^{k-1}} f_T$. More specifically, for $T \subseteq L^{k-1}$, we have that

$$f_T = \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ \pi_{m \to 1}(U) = T}} \hat{f}_U \chi_U.$$

Again for notational simplicity, we temporarily drop the subscript and write $\pi_{m \to 1}$ as π. We decompose the terms in the expectation in (5) as following

$$fF' = F' \sum_{T \subseteq L^{k-1}} f_T \tag{6}$$

$$= F' \sum_{\substack{T \subseteq L^{k-1} \\ |T| \leq J/2}} f_T + F' \sum_{\substack{T \subseteq L^{k-1} \\ |T| > J/2}} f_T. \tag{7}$$

We first bound the expectation of the high degree parts under both $\mathcal{T}$ and $\mathcal{T}^{(S_0)}$.

This is a standard correlation argument. We first consider the expectation under $\mathcal{T}$. Let $\mathcal{U}_{\mathcal{T}}$ be the conditional expectation operator mapping a function in $L_2(X^{(S_0)})$ to a function in $L_2(\overline{X}^{(S_0)})$ with respect to distribution $\mathcal{T}$. We have

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{T \subseteq L^{k-1} \\ |T| > J/2}} f_T \right] \right| = \left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{T \subseteq L^{k-1} \\ |T| > J/2}} \mathcal{U}_{\mathcal{T}} f_T \right] \right|. \tag{8}$$

Note that the expectation on the right hand side is in fact taken under the marginals

of $\mathcal{T}$ on $\overline{X}^{(S_0)}$. Applying Cauchy-Schwarz, we have

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[F'\sum_{\substack{T\subseteq L^{k-1}\\|T|>J/2}} f_T^{(S_0)}\right]\right| \le \sqrt{\mathop{\mathbf{E}}_{\mathcal{T}}\left[\sum_{\substack{T\subseteq L^{k-1}\\|T|>J/2}} \mathcal{U}_{\mathcal{T}}(f_T^{(S_0)})^2\right]}\sqrt{\mathop{\mathbf{E}}_{\mathcal{T}}[F'^2]} \tag{9}$$

$$\le \sqrt{\sum_{\substack{T\subseteq L^{k-1}\\|T|>J/2}} \left\|\mathcal{U}_{\mathcal{T}} f_T^{(S_0)}\right\|^2} \tag{10}$$

$$\le \sqrt{\sum_{\substack{T\subseteq L^{k-1}\\|T|>J/2}} \rho_0^{|T|}\left\|f_T^{(S_0)}\right\|^2} \le \rho_0^{J/2} \le \varepsilon_1, \tag{11}$$

where the inequality in (11) follows from Proposition 2.22 and that the norm in (11) is with respect to the marginal of $\mathcal{T}$ on $X^{(S_0)}$, which is uniform. The analysis for expectation under $\mathcal{T}^{(S_0)}$ is identical as it only involves correlation. Therefore

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}}[fF'] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}[fF']\right| \le \left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[F'\sum_{\substack{T\subseteq L^{k-1}\\|T|\le J/2}} f_T\right] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}\left[F'\sum_{\substack{T\subseteq L^{k-1}\\|T|\le J/2}} f_T\right]\right| + 2\varepsilon_1. \tag{12}$$

Now we turn to the low degree parts. Further unraveling the Efron-Stein decomposition, we have

$$F'\sum_{\substack{T\subseteq L^{k-1}\\|T|\le J/2}} f_T = F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U \tag{13}$$

$$= F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|=|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U + F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|>|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U. \tag{14}$$

Following terminology in [14, 32, 31], we refer to the first term as *shattered* term, and the second as non-shattered term. We would like to study these two terms separately. From (12), we have

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}}[fF'] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}[fF']\right| \tag{15}$$

$$\le 2\varepsilon_1 + \left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|=|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U\right] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}\left[F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|=|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U\right]\right|$$

$$+ \left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|>|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U\right] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}\left[F'\sum_{\substack{U\subseteq L^{k-m}\times R^{m-1}\\|U|>|\pi(U)|\\|\pi(U)|\le J/2}} \hat{f}_U\chi_U\right]\right|. \tag{16}$$

We first use smoothness to bound the non-shattered terms. The process is very similar to that in [31], and we get

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| > |\pi(U)| \\ |\pi(U)| \leq J/2}} \hat{f}_U \chi_U \right] \right| \leq 2\varepsilon_1. \tag{17}$$

The same argument holds under distribution $\mathcal{T}^{(S_0)}$. For the difference involving the shattered terms, we have

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| = |\pi(U)| \\ |\pi(U)| \leq J/2}} \hat{f}_U \chi_U \right] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| = |\pi(U)| \\ |\pi(U)| \leq J/2}} \hat{f}_U \chi_U \right] \right| \tag{18}$$

$$= \left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| = |\pi(U)| \\ |\pi(U)| \leq J/2}} \hat{f}_U \chi_U \right] - \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| = |\pi(U)| \\ |\pi(U)| \leq J/2}} (1-\gamma)^{|U|} \hat{f}_U \chi_U \right] \right| \tag{19}$$

$$= \left| \mathop{\mathbf{E}}_{\mathcal{T}} \left[F' \sum_{\substack{U \subseteq L^{k-m} \times R^{m-1} \\ |U| = |\pi(U)| \leq J/2}} (1-(1-\gamma)^{|U|}) \hat{f}_U \chi_U \right] \right| \tag{20}$$

$$\leq (1-(1-\gamma)^{J/2}) \leq \varepsilon_1. \tag{21}$$

The key step is (19) where we switch the distribution of the second term from $\mathcal{T}^{(S_0)}$ to $\mathcal{T}$. We rely crucially on the fact that $|U| = |\pi(U)|$. To see why this holds, denote the query to f as x (just for the current argument). Observe that x_t's are independent for $t \in U$ with different $\pi(t)$, so we first focus on the t's that map to the same $r \in U$. Looking at each $r \in \pi(U)$, $|\pi(U)| = |U|$ implies that there is a unique $t \in U$ such that $\pi(t) = r$, and thus perturbing x_t's where $\pi(t) = r$ with probability γ would give exactly a multiplicative factor of $(1-\gamma)$ to the expectation. Since each $r \in \pi(U)$ contributes a factor of $(1-\gamma)$, the final factor thus becomes $(1-\gamma)^{|\pi(U)|} = (1-\gamma)^{|U|}$.

Summing up the above, we have

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}}[fF'] - \mathop{\mathbf{E}}_{\mathcal{T}^{(S_0)}}[fF'] \right| \leq 7\varepsilon_1. \tag{22}$$

This completes the proof.

4.3 Influence Based Decoding

Suppose we have that for some $\mathscr{S} \subseteq \mathscr{S}_1 \cup \mathscr{S}_3$, the following term is large

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] \right| > \varepsilon_1, \tag{23}$$

then for at least an $\varepsilon_1/2$ fraction of all possible edge samplings, we have

$$\left| \mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{S \in \mathscr{S}} f^{(S)}(x^{(S)}) \right] \right| > \varepsilon_1/2. \tag{24}$$

In the rest of the proof, we focus on samplings of edges where (24) holds. We show how to extract a labeling for these edges.

Observe that after we fixed the edges, which function we query only depends on the layer of the query, so for the rest of this section, let f_l be the function on layer l. Also recall that $m(S) = \max S$ is the layer query $x^{(S)}$ is in, and thus in the PCP query $x^{(S)}$ goes to function $f_{m(S)}$. Let $l_m = \max_{S \in \mathscr{S}} m(S)$ be the maximum layer among queries that appears in $\mathscr{S}$.

For $l \in [k]$, denote the queries that appear on layer l as $\mathcal{L}_l := \{S \in \mathscr{S}_1 \cup \mathscr{S}_3 | \max S = l\}$, and let $\mathcal{L}_{\le l} := \cup_{l' \le l} \mathcal{L}_{l'}$, and similarly define $\mathcal{L}_{<l}$. We need the following observation on independence between queries.

Claim 4.8. *For any $l \in [k]$ and $S_0 \in \mathcal{L}_l$, $X^{(S_0)}$ and $\prod_{S \in \mathcal{L}_{<l}} X^{(S)}$ are independent under both $\mathcal{T}$ and $\mathcal{T}'$.*

Proof. We first consider $\mathcal{T}$. We can write $x^{(S_0)} = x_e \cdot x^{(\{l\})}$, where x_e depends on S_0, $\{x^{(S)}\}_{S \in \mathcal{L}_l}$ as well as choice of N_r's for $r \in L^{k-1}$, and the decision whether the bits in query $x^{(N_r)}$ are resampled, $x^{(\{l\})}$ is a uniform random string, and $\cdot$ denotes the element-wise product. Observe that $x^{(\{l\})}$ is independent of $\{x^{(S)}\}_{S \in \mathcal{L}_l}$, the N_r's and whether the bits are resampled, thus its marginal is still uniform no matter how we fix everything else, and so is the marginal of $x^{(S_0)}$. This implies that $X^{(S_0)}$ is independent of everything else and in particular $\prod_{S \in \mathcal{L}_{<l}} X^{(S)}$.

For $\mathcal{T}'$, note that the additional noise is applied independently to each bit, and we can use a similar argument as above to show that the marginal of $X^{(S_0)}$ is always uniform however we fix the other parameters. □

We rewrite the left hand side of (24) as

$$\mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{S \in \mathscr{S}} f_{m(S)}(x^{(S)}) \right] = \mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{l \in [k]} \prod_{S \in \mathcal{L}_l \cap \mathscr{S}} f_l(x^{(S)}) \right].$$

By our choice of permutation and Lemma 4.4, there exists l_0 and j_0 such that

$$|\{S \in \mathcal{L}_{l_0} \cap \mathscr{S} \mid S \ni j_0\}| \text{ is odd.}$$

Then flipping $x^{(\{j_0\})}$ while leaving all other $x^{(\{j'\})}$ unchanged changes the sign of the following

$$\prod_{S \in \mathcal{L}_{l_0} \cap \mathscr{S}} f_{l_0}(x^{(S)}),$$

and since the marginal of $x^{(\{j_0\})}$ is uniform, we have

$$\mathop{\mathbf{E}}_{\mathcal{T}'} \left[\prod_{S \in \mathcal{L}_{l_0} \cap \mathscr{S}} f_{l_0}(x^{(S)}) \right] = 0.$$

To complete the proof of soundness, we show that if

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{S\in\mathscr{S}} f_{m(S)}(x^{(S)})\right]\right| = \left|\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{l\in[k]}\prod_{S\in\mathcal{L}_l\cap\mathscr{S}} f_l(x^{(S)})\right]\right|$$

$$= \left|\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{l\in[k]}\prod_{S\in\mathcal{L}_l\cap\mathscr{S}} f_l(x^{(S)})\right] - \prod_{l\in[k]}\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{S\in\mathcal{L}_l\cap\mathscr{S}} f_l(x^{(S)})\right]\right| > \varepsilon_1/2, \qquad (25)$$

then there exists two layers $1 \le l < l_m \le k$ such that

$$\sum_{\substack{r_l\in L^{k-l}\times R^{l-1}\\ r_m\in L^{k-l_m}\times R^{l_m-1}\\ \pi_{l_m\to l}(r_m)=r_l}} \mathbf{Inf}^{(\gamma)}_{r_l}(f_l)\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m}) > \frac{\varepsilon_1^2}{4Z}, \qquad (26)$$

where $Z = Z(k,\gamma) := 2^{4k^3}k^6\gamma^{-2}$. This enables us to define a good labeling as the following: choose r_l with probability $\mathbf{Inf}^{(\gamma)}_{r_l}(f_l)/\mathbf{Inf}^{(\gamma)}(f_l)$, and similarly choose r_m with probability $\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m})/\mathbf{Inf}^{(\gamma)}(f_{l_m})$, then the probability that the labeling satisfies the edge is

$$\begin{aligned}
&\sum_{\substack{r_l\in L^{k-l}\times R^{l-1}\\ r_m\in L^{k-l_m}\times R^{l_m-1}\\ \pi_{l_m\to l}(r_m)=r_l}} \frac{\mathbf{Inf}^{(\gamma)}_{r_l}(f_l)}{\mathbf{Inf}^{(\gamma)}(f_l)}\frac{\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m})}{\mathbf{Inf}^{(\gamma)}(f_{l_m})}\\
>\ &\gamma^2 \sum_{\substack{r_l\in L^{k-l}\times R^{l-1}\\ r_m\in L^{k-l_m}\times R^{l_m-1}\\ \pi_{l_m\to l}(r_m)=r_l}} \mathbf{Inf}^{(\gamma)}_{r_l}(f_l)\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m})\\
\ge\ &\frac{\gamma^2\varepsilon_1^2}{4Z}.
\end{aligned}$$

This holds for at least $\varepsilon_1/2$ fraction of choices of edges, thus the expected value achieved by the above random labeling procedure is at least $\gamma^2\varepsilon_1^3/8Z$, a value depending only on k and ε.

The key step to proving (25) is to bound the following difference

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{S\in\mathscr{S}} f_{m(S)}(x^{(S)})\right] - \mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{l<l_m}\prod_{S\in\mathcal{L}_l\cap\mathscr{S}} f_l(x^{(S)})\right]\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{S\in\mathcal{L}_{l_m}\cap\mathscr{S}} f_{l_m}(x^{(S)})\right]\right|, \qquad (27)$$

where we recall that l_m is the highest layer of queries involved in $\mathscr{S}$. We can iteratively apply the bound on (27) to get (25). In order to establish (27), we use an invariance-type result from [31].

Theorem 4.9 ([31])**.** *Consider functions $\{f^{(t)} \in L^\infty(\Omega_t^n)\}_{t\in[m]}$ on a probability space $\mathcal{P} = (\prod_{t=1}^m \Omega_t, P)^{\otimes n}$ and a set $M \subsetneq [m]$. Furthermore, let $\mathcal{C}$ be the collection of minimal sets $C \subseteq [m]$, $C \not\subseteq M$, such that the spaces $\{\Omega_t\}_{t\in C}$ are dependent. Then*

$$\left|\mathbf{E}\left[\prod f^{(t)}\right] - \prod_{t\notin M}\mathbf{E}[f^{(t)}]\,\mathbf{E}\left[\prod_{t\in M} f^{(t)}\right]\right|$$

$$\le\ 2^{2m}\max_{C\in\mathcal{C}}\sqrt{\min_{r\in C}\mathbf{Inf}(f^{(r)})\sum_i\prod_{t\in C\setminus\{r\}}\mathbf{Inf}_i(f^{(t)})}\prod_{t\notin C}\left\|f^{(t)}\right\|_\infty.$$

To apply the above theorem, we first combine all functions that are not in the highest layer. Let $Q = \prod_{S\in\mathscr{S}\cap\mathcal{L}_{<l_m}} X^{(S)}$, and $q \in Q$ simply be concatenation of $\{x^{(S)}\}_{S\in\mathscr{S}\cap\mathcal{L}_{<l_m}}$. Define the combined function $F = \prod_{S\in\mathscr{S}\cap\mathcal{L}_{<l_m}} f_{m(S)}$. We still have by Claim 4.8 that Q and $X^{(S_0)}$ are independent for all $S_0 \in \mathcal{L}_{l_m}$. Then the first term in (27) becomes

$$\mathop{\mathbf{E}}_{\mathcal{T}'}\left[\prod_{S\in\mathscr{S}} f_{m(S)}(x^{(S)})\right] = \mathop{\mathbf{E}}_{\mathcal{T}}\left[T_\gamma F(q)\prod_{S\in\mathscr{S}\cap\mathcal{L}_{l_m}} T_\gamma f_{l_m}(x^{(S)})\right].$$

Let us set $M = \mathscr{S}\cap\mathcal{L}_{l_m}$. Consider the sets C in Theorem 4.9. Since Theorem 4.9 requires that $C \not\subseteq M$, we have that C must include variable q. Due to the independence in Claim 4.8, C must also include at least two variables from $\mathscr{S}\cap\mathcal{L}_{l_m}$. Applying Theorem 4.9, we have

$$\begin{aligned}&\left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[T_\gamma F(q)\prod_{S\in\mathscr{S}\cap\mathcal{L}_{l_m}} T_\gamma f_{l_m}(x^{(S)})\right] - \mathop{\mathbf{E}}_{\mathcal{T}}[T_\gamma F(q)]\mathop{\mathbf{E}}_{\mathcal{T}}\left[\prod_{S\in\mathscr{S}\cap\mathcal{L}_{l_m}} T_\gamma f_{l_m}(x^{(S)})\right]\right| \\ \le\ & 2^{2k^3}\sqrt{\mathbf{Inf}^{(\gamma)}(\overline{f}_{l_m})\sum_{r\in L^{k-1}}\mathbf{Inf}_r^{(\gamma)}(\overline{F})\mathbf{Inf}_r^{(\gamma)}(\overline{f}_{l_m})},\end{aligned}$$

where $\overline{F}$ and $\overline{f}_{l_m}$ are lifted versions of F and f_{l_m} as defined in Definition 2.15.

We have that $\mathbf{Inf}^{(\gamma)}(\overline{f}_{l_m}) \le \mathbf{Inf}^{(\gamma)}(f_{l_m}) \le \gamma^{-1}$. Now we need to relate $\mathbf{Inf}_r^{(\gamma)}(\overline{F})$ with $\mathbf{Inf}_r^{(\gamma)}(\overline{f}_{m(S)})$.

We use the following generalization of Lemma 6.5 from [21].

Lemma 4.10. *Let $(\prod_{i=1}^m \Omega_i^n, \mu)$ be correlated probability space, and $f_i : \Omega_i^n \to [-1,1]$ for $i = 1,\cdots,m$. Then for all r:*

$$\mathbf{Inf}_r(\prod_{i=1}^m f_i) \le m\sum_{i=1}^m \mathbf{Inf}_r(f_i).$$

The argument goes exactly the same so we omit the proof here.

Applying Lemma 4.10, we get $\mathbf{Inf}_r^{(\gamma)}(\overline{F}) \le k^3\sum_{S\in\mathscr{S}\cap\mathcal{L}_{<l_m}}\mathbf{Inf}_r^{(\gamma)}(\overline{f}_{m(S)})$.

Summing up and using Proposition 2.16, we get

$$\left|\mathop{\mathbf{E}}_{\mathcal{T}}\left[T_\gamma F(q)\prod_{S\in\mathscr{S}\cap\mathcal{L}_{l_m}} T_\gamma f_{l_m}(x^{(S)})\right] - \mathop{\mathbf{E}}_{\mathcal{T}}[T_\gamma F(q)]\mathop{\mathbf{E}}_{\mathcal{T}}\left[\prod_{S\in\mathscr{S}\cap\mathcal{L}_{l_m}} T_\gamma f_{l_m}(x^{(S)})\right]\right|$$

$$\le\ 2^{2k^3}\sqrt{k^3\gamma^{-1}\sum_{\substack{r\in L^{k-1}\\ S\in\mathscr{S}\cap\mathcal{L}_{<l_m}}}\mathbf{Inf}_r^{(\gamma)}(\overline{f}_{m(S)})\mathbf{Inf}_r^{(\gamma)}(\overline{f}_{l_m})} \tag{29}$$

$$\le\ 2^{2k^3}\sqrt{k^3\gamma^{-1}\sum_{\substack{1\le l<l_m\\ r_l\in L^{k-l}\times R^{l-1}\\ r_m\in L^{k-l_m}\times R^{l_m-1}\\ \pi_{l_m\to l}(r_m)=r_l}}\mathbf{Inf}_{r_l}^{(\gamma)}(f_l)\mathbf{Inf}_{r_m}^{(\gamma)}(f_{l_m})}. \tag{30}$$

Let $Z' = 2^{2k^3}\sqrt{k^3\gamma^{-1}}$, applying (30) to all layers, we get

$$\begin{aligned}&\left|\mathop{\mathbf{E}}_{T'}\left[\prod_{S\in\mathscr{S}} f^{(S)}(x^{(S)})\right]\right| \\ <\ & Z' \sum_{2\le l_m<k} \sqrt{\sum_{\substack{1\le l<l_m \\ r_l\in L^{k-l}\times R^{l-1} \\ r_m\in L^{k-l_m}\times R^{l_m-1} \\ \pi_{l_m\to l}(r_m)=r_l}} \mathbf{Inf}^{(\gamma)}_{r_l}(f_l)\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m})}.\end{aligned}$$

Thus if the left hand side of the above is larger than $\varepsilon_1/2$, then there exists $1 \le l < l_m \le k$ such that

$$\sum_{\substack{r_l\in L^{k-l}\times R^{l-1} \\ r_m\in L^{k-l_m}\times R^{l_m-1} \\ \pi_{l_m\to l}(r_m)=r_l}} \mathbf{Inf}^{(\gamma)}_{r_l}(f_l)\mathbf{Inf}^{(\gamma)}_{r_m}(f_{l_m}) > \left(\frac{\varepsilon_1/2}{kZ'}\right)^2 \cdot \frac{1}{k} = \frac{\varepsilon_1^2}{4Z}.$$

5 Acknowledgments

The author would like to thank Johan Håstad for inspiring discussions and invaluable advice; and Cenny Wenner for explaining many of his ideas, as well as numerous comments on improving this manuscript.

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

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

[3] Per Austrin and Johan Håstad. On the usefulness of predicates. In *IEEE Conference on Computational Complexity*, pages 53–63, 2012.

[4] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.

[5] Siu On Chan. Approximation resistance from pairwise independent subgroups. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:110, 2012.

[6] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM J. Comput.*, 34(5):1129–1146, 2005.

[7] B. Efron and C. Stein. The jackknife estimator of variance. *Annals of Statistics*, 9:586–596, 1981.

[8] Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. *Random Struct. Algorithms*, 33(4):497–514, 2008.

[9] Vitaly Feldman, Venkatesan Guruswami, Prasad Raghavendra, and Yi Wu. Agnostic learning of monomials by halfspaces is hard. In *FOCS*, pages 385–394, 2009.

[10] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing UGC from some optimal geometric inapproximability results. In *SODA*, pages 699–717, 2012.

[11] Gustav Hast. Beating a random assignment. In *APPROX-RANDOM*, pages 134–145, 2005.

[12] Gustav Hast. Beating a random assignment. *PhD Thesis*, 2005.

[13] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[14] Johan Håstad. On the NP-hardness of Max-Not-2. In *APPROX-RANDOM*, pages 170–181, 2012.

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

[16] Sangxia Huang. Approximation resistance on satisfiable instances for predicates strictly dominating parity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:40, 2012.

[17] Subhash Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *FOCS*, pages 23–32, 2002.

[18] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. ACM.

[19] Subhash Khot and Dana Moshkovitz. NP-hardness of approximately solving linear equations over reals. In *STOC*, pages 413–420, 2011.

[20] Elchanan Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *FOCS*, pages 156–165, 2008.

[21] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric And Functional Analysis*, 19:1713–1756, 2010. 10.1007/s00039-010-0047-x.

[22] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *FOCS*, pages 21–30, 2005.

[23] Ryan O'Donnell and John Wright. A new point of NP-hardness for unique games. In *STOC*, pages 289–306, 2012.

[24] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable 3-CSPs. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 493–502, New York, NY, USA, 2009. ACM.

[25] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[26] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 191–199, New York, NY, USA, 2000. ACM.

[27] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 11–20, New York, NY, USA, 2006. ACM.

[28] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. *SIAM J. Comput.*, 39(1):323–360, 2009.

[29] Linqing Tang. Conditional hardness of approximating satisfiable Max 3CSP$_q$. In *ISAAC*, pages 923–932, 2009.

[30] Luca Trevisan. Approximating satisfiable satisfiability problems. *Algorithmica*, 28(1):145–172, 2000.

[31] Cenny Wenner. Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width four. *Manuscript*, 2012.

[32] Cenny Wenner. Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width four - (extended abstract). In *APPROX-RANDOM*, pages 325–337, 2012.

[33] Uri Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *SODA*, pages 201–210, 1998.