Algorithms and Certificates for Boolean CSP Refutation: Smoothed Is No Harder Than Random

Venkatesan Guruswami
venkatg@berkeley.edu
UC Berkeley
USA

Pravesh K. Kothari
praveshk@cs.cmu.edu
Carnegie Mellon University
USA

Peter Manohar
pmanohar@cs.cmu.edu
Carnegie Mellon University
USA

ABSTRACT

We present an algorithm for strongly refuting *smoothed* instances of all Boolean CSPs. The smoothed model is a hybrid between worst and average-case input models, where the input is an arbitrary instance of the CSP with only the negation patterns of the literals re-randomized with some small probability. For an n-variable smoothed instance of a k-arity CSP, our algorithm runs in $n^{O(\ell)}$ time, and succeeds with high probability in bounding the optimum fraction of satisfiable constraints away from 1, provided that the number of constraints is at least $\tilde{O}(n)(\frac{n}{\ell})^{\frac{k}{2}-1}$. This matches, up to polylogarithmic factors in n, the trade-off between running time and the number of constraints of the state-of-the-art algorithms for refuting *fully random* instances of CSPs.

We also make a surprising connection between the analysis of our refutation algorithm in the significantly "randomness starved" setting of semi-random k-XOR and the existence of even covers in *worst-case* hypergraphs. We use this connection to positively resolve Feige's 2008 conjecture – an extremal combinatorics conjecture on the existence of even covers in sufficiently dense hypergraphs that generalizes the well-known Moore bound for the girth of graphs. As a corollary, we show that polynomial-size refutation witnesses exist for arbitrary smoothed CSP instances with number of constraints a polynomial factor below the "spectral threshold" of $n^{k/2}$, extending the celebrated result for random 3-SAT of Feige, Kim and Ofek.

CCS CONCEPTS

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

KEYWORDS

CSP refutation, Smoothed CSPs, Even covers

ACM Reference Format:
Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. 2022. Algorithms and Certificates for Boolean CSP Refutation: Smoothed Is No Harder Than Random. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC '22), June 20–24, 2022, Rome, Italy*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3519935.3519955

STOC '22, June 20–24, 2022, Rome, Italy

ACM ISBN 978-1-4503-9264-8/22/06.
https://doi.org/10.1145/3519935.3519955

1 INTRODUCTION

Worst-case complexity theory paints a grim picture for solving Constraint Satisfaction Problems (CSPs). For a large class [10, 23] of Max CSPs with k-ary Boolean predicates (k-CSPs), the Exponential Time Hypothesis (ETH) [17] implies that for sparse instances, i.e., with $m = O(n)$ constraints in n variables, there is no sub-exponential time approximation algorithm that beats simply returning a random assignment. While fully-dense instances (i.e., $m \geqslant O(n^k)$) admit [7] a polynomial time approximation scheme (PTAS), ETH implies that lowering m to just $\sim n^{k-1}$ makes the problem APX-hard [16] even for sub-exponential time algorithms. In fact, for instances with $m \leqslant O(n^{k-1})$, we suspect that even efficiently verifiable *certificates* of non-vacuous upper bounds on the value, i.e., max fraction of constraints satisfiable, do not exist.

The study of *random* CSPs, on the other hand, offers a stark contrast. Max k-CSPs with any strictly super-linear number of, say, $m \geqslant n^{1.1}$ randomly generated constraints[1] admit [3, 9, 25] sub-exponential time *tight refutation*[2] algorithms. These are based on *spectral methods* that exploit problem structure in non-trivial ways. Further, when $m \sim \tilde{O}(n^{k/2}) \ll n^{k-1}$, such algorithms in fact yield a PTAS for certifying the value of the input instance correctly. In fact, a considerably more fine-grained, predicate-specific and likely sharp picture [8, 20] of the trade-off between running time and number of constraints has emerged in the last decade. Adding to this rich theory is the fascinating work of [14] that shows that random CSPs admit polynomial-time verifiable certificates of non-trivial upper bounds on the value even when $m \sim n^{k/2-\delta_k}$ – i.e., when number of constraints are polynomially smaller than the threshold for efficient refutation.

How does the complexity landscape of CSPs – for both algorithms and certificates – interpolate between these two extremes? Is the worst-case understanding too pessimistic? Is the average-case understanding too idealistic? And are the sophisticated algorithmic tools and the structural properties that govern their success for random CSPs relevant to more general instances?

Refutation algorithms in the smoothed model. To formally study these questions, in 2007, Feige [12] introduced a natural "hybrid" model in between worst-case and random instances (in the spirit of the pioneering work of Spielman and Teng [28]). In this *smoothed model*, an instance is generated by starting from an arbitrary (i.e., worst-case) instance, and then negating each literal in each clause independently with some small, constant probability. In contrast to random CSPs where the clause structure

[1] i.e., uniformly random and independently chosen variables and "literal patterns" in each constraint.
[2] Such algorithms correctly *certify* an upper bound on the value within an arbitrarily small additive ε w.h.p.

(i.e., k-tuples describing the constraints) and the literal patterns (i.e., which variables are negated in a constraint) are chosen uniformly at random and independently, the clause structure in smoothed CSPs is completely arbitrary (i.e., worst-case) and only a small constant fraction of the literal patterns are random. In [12], Feige combined semidefinite programming with a new combinatorial certificate based on a natural notion of cycles in hypergraphs, and proved that polynomial algorithms succeed in weakly refuting (i.e., certifying a $1 - o_n(1)$ upper bound on value, Definition 1.2) *smoothed* 3-SAT formulas with $m \geqslant \tilde{O}(n^{1.5})$ constraints.

Feige's techniques, however, appear fundamentally limited to weak refutation and specialized to 3-CSPs. As a result, there is no known strong refutation algorithm (i.e., certifying a $1 - \Omega(1)$ upper bound on value) for smoothed instances of 3-SAT and no known (even weak) refutation algorithm for smoothed instances of any nontrivial 4-CSP.

In this work, we develop new techniques that yield strong refutation algorithms for all Boolean k-CSPs with (a possibly sharp) trade-off between running time and number of constraints matching that of random k-CSPs [25], up to polylogarithmic factors. In particular, our results show that the algorithmic task of strong refutation in the significantly "randomness starved" setting of smoothed instances is no harder than in a fully random instance.

Refutation witnesses below spectral threshold: Feige's conjecture. The work [14] (and extensions [30]), prove that there are efficiently verifiable witnesses of unsatisfiability for *fully random* k-CSPs with $n^{\frac{k}{2}-\delta_k}$ constraints for some constant $\delta_k > 0$; when $k = 3$, this threshold is $n^{1.4}$. These witnesses are based on certain natural analogs of cycles in hypergraphs called *even covers*. In an effort to understand if such witnesses exist in more general instances, Feige [13] conjectured a trade-off between number of constraints and size of a smallest even cover. This conjecture formally generalizes the Moore bound [5] on girth of graphs to hypergraphs.

In this work, we prove Feige's conjecture by a new *spectral double counting* argument that relates sub-exponential time smoothed refutation algorithms and the existence of even covers in hypergraphs. As a consequence, we derive that there are efficiently verifiable witnesses of unsatisfiability for smoothed instances of all k-CSPs with $m \sim n^{k/2-\delta_k}$ constraints, for some constant δ_k, which is polynomially smaller than the threshold at which efficient refutation algorithms exist even for random k-CSPs.

Summary. Taken together, our main results can be interpreted as suggesting that the worst-case picture of complexity of CSPs arises entirely because of *islands of pathology*: most instances "around" the worst-case hard ones are in fact essentially as easy as random, for both refutation algorithms as well as existence of refutation witnesses. Further, in a precise sense, the difficulty of worst-case instances can be attributed to the worst-case literal patterns, rather than the clause structure.

Our contribution is shown visually in Figure 1. Figure 1 plots the time vs. # constraints trade-off for refuting random and smoothed 3-SAT instances (along with the analogous trade-off for approximation schemes for worst case instances). Our contribution is the smoothed case (blue line), which shows that smoothed 3-SAT instances can be refuted with the same trade-off as random ones

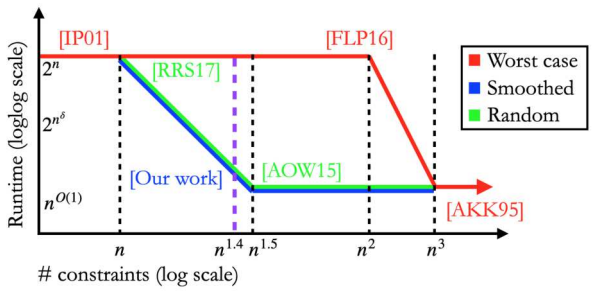

Figure 1: Time vs. # constraints trade-off for refuting random and smoothed 3-SAT instances, and for approximation schemes for worst-case instances. The smoothed case is our contribution. We also prove that refutation witnesses exist for smoothed instances at the purple line, i.e., $n^{1.4}$ constraints.

(green line). We also show that there exist efficiently verifiable refutation witnesses for smoothed instances at $n^{1.4}$ constraints (purple line), matching the result for random instances due to [14].

Our results. We now discuss our results on algorithms and certificates, as well as the interconnected techniques and insights that go into them. Let us recall the standard notation to talk about CSPs.

Definition 1.1 (k-ary Boolean CSPs, random, semirandom, and smoothed instances)**.** A CSP instance ϕ on n variables with a k-ary predicate $P : \{\pm 1\}^k \to \{0,1\}$ is a set of m constraints on n variables of the form $P(\xi(C)_1 x_{C_1}, \xi(C)_2 x_{C_2}, \ldots, \xi(C)_k x_{C_k}) = 1$. Here, $C = (C_1, C_2, \ldots, C_k)$ ranges over a collection $\mathcal{H}$ of *scopes* (a.k.a. clause structure) of k-tuples of n variables such that $C_i \neq C_j$ for any i, j and $\xi : \mathcal{H} \to \{\pm 1\}^k$ are "literal negation patterns" one for each C in $\mathcal{H}$. The *value* of ϕ, $\mathrm{val}(\phi)$, is the maximum fraction of constraints satisfied by any assignment to the n variables.

In a *random* (sometimes, *fully random* in order to disambiguate from related models) instance, $\mathcal{H}$ is a collection of m uniformly random and independently chosen k-tuples and the $\xi(C)$'s are chosen uniformly at random and independently from $\{\pm 1\}^k$ for each C.

In a *semirandom* instance, $\mathcal{H}$ is arbitrary (i.e., worst-case) and $\xi(C) \in \{\pm 1\}^k$ are uniformly at random and independent for each C.

In a *smoothed* instance, $\mathcal{H}$ is arbitrary (i.e., worst-case) and $\xi(C) \in \{\pm 1\}^k$ are obtained by starting with arbitrary (i.e., worst-case) $\xi'(C) \in \{\pm 1\}^k$ for each C and then for each C, i, setting $\xi(C)_i = \xi'(C)_i$ with probability 0.99 and $\xi(C)_i = -\xi'(C)_i$ with probability 0.01, independently.

We note that the semirandom model is more general than the random model, and the smoothed model is more general than the semirandom model.

Definition 1.2 (Weak, Strong and Tight refutation algorithms)**.** A refutation algorithm takes as input a CSP instance ϕ and outputs a value $\text{alg-val}(\phi) \in [0,1]$ with $\text{alg-val}(\phi) \geqslant \mathrm{val}(\phi)$ for all ϕ. For a distribution $\mathcal{D}$ over ϕ, we say that the refutation algorithm *weakly refutes* instances drawn from $\mathcal{D}$ if with high probability over $\phi \sim \mathcal{D}$, $\text{alg-val}(\phi) < 1$. We also define *strong refutation* ($\text{alg-val}(\phi) <$

$1-\delta$ for some absolute constant $\delta > 0$) and *ε-tight refutation* (alg-val$(\phi) <$ val$(\phi) + \varepsilon$ for arbitrarily small ε) analogously.

Algorithms for smoothed refutation. Our first main result gives a (possibly sharp) trade-off between running time and number of constraints for strongly refuting *smoothed* CSP instances.

Theorem 1 (Smoothed refutation, informal)**.** *For every $\ell = \ell(n)$, there is a $n^{O(\ell)}$-time strong refutation algorithm for smoothed CSPs with $m \geqslant m_0 = \tilde{O}(n) \cdot \left(\frac{n}{\ell}\right)^{\left(\frac{\ell}{2}-1\right)}$ constraints. That is, for any CSP instance ϕ with $m \geqslant m_0$ constraints, with probability 0.99 over the smoothing ϕ_s of ϕ, the algorithm outputs* alg-val$(\phi_s) \leqslant 1-\delta$ *for some absolute constant $\delta > 0$.*

Here, $t = t(P) \leqslant k$ is the "degree of uniformity" of P – the smallest integer $t \leqslant k$ such that there is no t-wise uniform distribution on $\{\pm 1\}^k$ supported entirely on the satisfying assignments $P^{-1}(1) \subseteq \{\pm 1\}^k$.

In order to understand the trade-off described by the theorem, let us apply it to two examples.

Example 1.3. For k-SAT, P is the Boolean OR function. We thus have $t(P) = k$, as the uniform distribution on odd-parity strings is supported on $P^{-1}(1)$ and is $(k-1)$-wise uniform. Our result gives a polynomial time algorithm to strongly refute smoothed instances of k-SAT whenever the number of constraints $m \geqslant \tilde{O}(n^{\frac{k}{2}})$. More generally, for any $\delta > 0$, in time $2^{O(n^\delta)}$ the algorithm strong refutes smoothed instances with $\geqslant \tilde{O}(n^{(1-\delta)\frac{k}{2}+\delta})$ constraints.

As a second example, consider the "Hadamard predicate" P on $k = 2^{2^q-1}$ bits where $P(x) = 1$ if and only if x is a codeword of the truncated Hadamard code. Hadamard CSPs naturally appear in the design of query efficient PCPs. Here, $t(P) = 3 \ll k$, so our theorem gives a polynomial-time algorithm to strongly refute smoothed instances of the Hadamard CSP with at least $\tilde{O}(n^{1.5})$ constraints, and a 2^{n^δ}-time algorithm for instances with at least $\tilde{O}(n^{1.5-\delta/2})$ constraints $\forall \delta \in (0, 1]$.

Comparison with prior results. Theorem 1 can be directly compared to works on refuting random, semirandom and smoothed (in the order of increasing generality) CSPs.

Building on [3, 9], Raghavendra, Rao and Schramm [25] proved the same trade-off (up to a polylog(n) factor in m) between running time and number of constraints required as in Theorem 1 for the significantly simpler special case of *fully random* CSPs – when the clause structure and the literal patterns are chosen uniformly at random from the respective domains. Our result shows that the same trade-off holds for *smoothed* instances – i.e., with worst-case clause structure and small random perturbations of worst-case literal patterns. All known efficient refutation algorithms, including ours and that of [25], can in hindsight be interpreted as an analysis of the canonical sum-of-squares (SoS) relaxation for the max k-CSP problem. For random CSPs (and thus also for the more general smoothed instances we study) the trade-off we obtain is known to be essentially tight [8, 20] for such "SoS-encapsulated" algorithms: this fact is often taken as evidence of sharpness of this trade-off.

Much less is known about refuting CSPs in the more general *semirandom* and *smoothed* models. Feige [12] gave a *weak* refutation algorithm for refuting smoothed and semirandom instances of 3-SAT. His techniques apply to all 3-CSPs but do not seem to extend to either strong refutation or 4-CSPs. More recently, in a direct precursor to this work, Abascal, Guruswami and Kothari [1] gave a polynomial time algorithm for refuting *semirandom* instances of all CSPs – thus obtaining one of the extreme points (corresponding to $\ell = O(1)$) in the trade-off in Theorem 1 above. Theorem 1 relies on a key idea from their work (row bucketing) along with several new ideas discussed below.

Algorithms for refuting *semirandom* k-XOR. Our main technical result is an algorithm for *tight* refutation of *semirandom* instances of k-XOR. Theorem 1 then follows by a simple blackbox reduction that relies on a dual polynomial introduced in [3]. For the special case of k-XOR, a semirandom instance ψ is completely described by an arbitrary k-uniform instance hypergraph $\mathcal{H}$ and a collection of "right-hand sides" $b_C \in \{\pm 1\}$, one for each $C \in \mathcal{H}$. One can associate to ψ a homogeneous degree k polynomial $\psi(x)$ on the hypercube $\{\pm 1\}^n$ that computes the "advantage over 1/2" of an assignment x; that is, the value of the associated instance is $\frac{1}{2} + \max_{x \in \{\pm 1\}^n} \psi(x)$. Tight refutation corresponds to certifying that $\psi(x) \leqslant \varepsilon$ for arbitrary $\varepsilon > 0$.

$$\psi(x) = \frac{1}{m} \sum_{C \in \mathcal{H}} b_C \prod_{i \in C} x_i \,.$$

Theorem 1.4 (Strongly refuting semirandom k-XOR, informal)**.** *For every $k \in \mathbb{N}$ and $\ell = \ell(n)$ and every $\varepsilon > 0$, there is a $n^{O(\ell)}$ time ε-tight refutation algorithm for homogeneous degree k polynomials that succeeds with probability at least 0.99 over the draw of the coefficients i.i.d. uniform on $\{-1, 1\}$, whenever the associated hypergraph $\mathcal{H}$ has $m \geqslant n\left(\frac{n}{\ell}\right)^{\frac{k}{2}-1} \cdot \mathrm{poly}(\frac{\log n}{\varepsilon})$ hyperedges.*

In particular, for every $\delta > 0$, we obtain a $2^{O(n^\delta)}$-time ε-tight refutation algorithm for semirandom k-XOR instances with $m \gg \tilde{O}(n) \cdot n^{(1-\delta)\left(\frac{k}{2}-1\right)} \mathrm{poly}(\frac{1}{\varepsilon})$-constraints.

Prior works and brief comparison of techniques. The trade-off above (up to polylog(n) factors in m) matches the one obtained for refuting fully random k-XOR [25]. Our techniques, however, necessarily need to be significantly different, as the analysis in [25] (and related works it built on [3, 9, 11]) crucially rely on the randomness of the hypergraph $\mathcal{H}$. In particular, the refutation in [25] uses the spectral norm of a certain "symmetric tensor power" of the canonical matrix obtained from the instance. They analyze this matrix using a technical tour-de-force argument using the trace moment method.[3] A couple of follow-up works have attempted to simplify the analyses in [25]. Wein, Alaoui and Moore [29] succeeded in giving a simpler proof (introducing the *Kikuchi matrix*, a variant of which is central to this work) for the case of random k-XOR for *even* k, and they also suggest that a natural generalization of their Kikuchi matrix for random odd k will work (their suggestion does not pan out, as we prove in Appendix A). In a recent work, Ahn [2] simplified some aspects of the analysis of the "symmetric tensor power" matrix in the analysis of [25]. To summarize, the tools in prior works on random CSPs for analyzing the spectra of relevant correlated random matrices seem to use the randomness of the hypergraph both heavily and in a rather opaque manner.

[3] Just the technical argument in [25] runs over 20 pages!

For the more general setting of semirandom k-XOR refutation, the best known result [1] obtained an extreme point in the trade-off (i.e., the case of $\ell = O(1)$). That work analyzes the $\infty \to 1$-norm of the canonical matrix associated with the CSP instance. In this special case when $\ell = O(1)$, it turns out that handling 3-XOR instances allows deriving all larger k as a corollary. For the case of 3-XOR, their analysis relies on a new *row bucketing* step according to the *butterfly degree* of a pair of vertices (a new notion that they define), along with a certain pseudo-random vs structure decomposition for arbitrary 3-uniform hypergraphs associated with the 3-XOR instance.

To prove Theorem 1.4, we build on [1] and introduce a few new tools. For even k, the Kikuchi matrix of [29] analyzed using the row bucketing idea (with an appropriate generalization of the butterfly degree) of [1] yields a correct trade-off (see Sections 2.1 and 2.2). The case of odd k turns out to be significantly more challenging (as has always been the case in CSP refutation) and needs new ideas. We introduce a variant of the Kikuchi matrix for this purpose. Unlike the case of even k (and the algorithm in [1]), the spectral norm of this matrix is provably too large to yield a refutation – even for *random* instances. Indeed, this is why the strategy suggested by [29] does not pan out, as we show in Appendix A. Instead, we use the spectral norm of a matrix obtained by pruning away appropriately chosen rows. We then show that the number of pruned rows is not too large, and so does not contribute too much to the $\infty \to 1$-norm of the full matrix.

The row pruning step motivates a definition of *regularity*, a collection of natural pseudorandom properties that relate to *well-spreadness* in the intersection structure of the hyperedges in the instance hypergraph.[4] We then show that the hyperedges in every k-uniform hypergraph can be decomposed, via a *regularity decomposition* lemma, into k'-uniform hypergraphs for $k' \leqslant k$, along with some "error" hyperedges, such that (i) each of the k'-uniform hypergraphs satisfies regularity, and (ii) refuting all of these k'-XOR instances provides a refutation for the original instance. We explain our row pruning and the regularity decomposition steps in more detail in Section 2.

Short refutations below spectral threshold: proving Feige's conjecture. In a one-of-a-kind result, Feige, Kim and Ofek [14] (henceforth, FKO) proved that with high probability over the draw of a fully random 3-SAT instance ψ, there is a polynomial size *witness* that weakly refutes ψ if ψ has $m \sim \tilde{O}(n^{1.4})$ constraints. Formally, there is a polynomial time *non-deterministic* refutation algorithm that succeeds in finding a refutation with high probability over the drawn of a fully random 3-SAT instance with $m \sim \tilde{O}(n^{1.4})$ constraints. On the other hand, all known polynomial time *deterministic* refutation algorithms require the input random instance to have $\Omega(n^{1.5})$ constraints – this bound is often called the *spectral threshold*. The fastest known refutation algorithm [25] for instances with $\sim n^{1.4}$ constraints runs in time $2^{n^{0.2}}$, matching the SoS lower bound [20]. Thus, intriguingly, the FKO result shows the existence of polynomial time verifiable refutation witnesses (i.e., certificates of an upper bound of $1 - o_n(1)$ on the value) at a constraint density at which there are no known $2^{n^{o(1)}}$-time refutation algorithms. Does such a "gap" between thresholds for existence vs efficient computability of refutation witnesses persist for semirandom and smoothed instances, i.e., instances with *worst-case* constraint hypergraphs?

In 2008, Feige [13] made an elegant conjecture on the existence of even covers in sufficiently dense hypergraphs. This conjecture can be interpreted as generalizing to hypergraphs the classical Moore bound on the girth of graphs with a given number of edges. If true, Feige's conjecture implies that the FKO result holds for all semirandom and smoothed CSP instances – in particular, the FKO result does not rely on the properties of the underlying hypergraph at all. Let us explain this conjecture below.

Definition 1.5 (Even Cover and Girth)**.** For a k-uniform hypergraph $\mathcal{H}$ on $[n]$, an *even cover* of length t is a collection of t distinct hyperedges $C_1, C_2, \ldots, C_t$ in $\mathcal{H}$ such that every vertex in $[n]$ appears in an even number of C_i's. The *girth* of $\mathcal{H}$ is the length of the smallest even cover in $\mathcal{H}$.

Conjecture 1.6 (Feige's conjecture, Conjecture 1.2 in [13])**.** *Every k-uniform hypergraph $\mathcal{H}$ on $[n]$ with $m \geqslant m_0 = O(n)\left(\frac{n}{\ell}\right)^{\frac{k}{2}-1}$ hyperedges has an even cover of length $O(\ell \log n)$.*

A brief history of the conjecture. For $k = 2$, an even cover is a 2-regular subgraph (and thus a union of cycles) in a graph and thus, the conjecture above reduces to the question of determining the maximum girth (the length of the smallest cycle) in a graph with n vertices and $nd/2$ edges for parameter d. The best known bound is due to Alon, Hoory and Linial [5] who proved that for every graph on n vertices with $nd/2$ edges for $d > 2$, there is a cycle of length at most $c \log_{d-1} n$ for $c \leqslant 2$. The best known lower bound on the girth is $c \log_{(d-1)} n$ for $c \geqslant 4/3$ by Margulis [22] and Lubotzky, Philips and Sarnak [21] via explicit constructions of Ramanujan graphs. Obtaining a tight bound on c has been an outstanding open problem for the last 3 decades.

Much less is known for hypergraphs. When k even and $\ell = O(1)$, Naor and Verstraete [24] proved the conjecture. They were motivated by a natural coding theory interpretation: viewing each hyperedge as describing the non-zero coefficients of linear equations over $\mathbb{F}_2$, an even cover is a *sparse linear dependency* and thus, the conjecture gives the rate-distance trade-off for linear codes with column-sparse parity check matrices. In the more challenging case when k is odd, the bounds for $\ell = O(1)$ case in [24] were improved to essentially optimal ones in [13]. For $\ell \gg 1$, the best previous bound for 3-uniform hypergraphs is due to a simple argument of Alon and Feige [4] (Lemma 3.3), who proved that every 3-uniform hypergraph with $\tilde{O}(n^2/\ell)$ hyperedges has an even cover of size ℓ (this is off by $\sim \sqrt{n}$ factor in m). For 3-uniform hypergraphs with $m \gg n^{1.5+\varepsilon}$ (and the case when $m \gg n^{k/2}$ in general), [18] proved that there are even covers of size $O(1/\varepsilon)$. Finally, Feige and Wagner [15] proved some variants ("generalized girth problems") in order to build tools to approach this conjecture.

To summarize, prior to this work, the conjecture was known to be true only for $\ell = O(1)$. For larger ℓ, the only approach was the combinatorial strategy introduced in [15]. In this work, we prove Feige's conjecture (up to poly $\log n$ slack in m) via a new *spectral double counting argument*.

[4] This is closely related to the notion of spread encountered in recent work on the sunflower conjecture [6, 26].

Theorem 2 (Feige's conjecture is true, informal). *For every $k \in \mathbb{N}$ and $\ell = \ell(n)$, every k-uniform hypergraph $\mathcal{H}$ with $m \geqslant m_0 = \tilde{O}(n) \cdot (\frac{n}{\ell})^{\frac{k}{2}-1}$ hyperedges has an even cover of size $O(\ell \log n)$.*

Our spectral double counting argument[5] is heavily derived from our analysis for smoothed refutation using our Kikuchi matrices; indeed, our proof of Theorem 2 mirrors our steps in the analysis of our refutation algorithm. In fact, in a precise sense (as we explain in Section 2.3), our approach gives a tight connection between even covers in hypergraphs and simple cycles (and in turn, the spectral norm of the corresponding adjacency matrix) in the "Kikuchi graph" built from the hypergraph.

Combining with our smoothed refutation algorithms (Theorem 1) we immediately obtain a generalization of the FKO result that yields a polynomial time non-deterministic refutation algorithm for smoothed instances of all k-ary CSPs with number of constraints m polynomially below the spectral threshold of $n^{k/2}$.

Theorem 3 (Non-deterministic refutation, informal). *There is a* non-deterministic *polynomial time algorithm that weakly refutes smoothed instances of any k-CSP with $m \geqslant m_0 = \tilde{O}(n^{\frac{k}{2}-\frac{k-2}{2(k+8)}})$-constraints. For the special case of $k = 3$, $m_0 = \tilde{O}(n^{1.4})$.*

2 OVERVIEW OF OUR TECHNIQUES

In this section, we illustrate our key ideas by giving essentially complete proofs of some special cases of our main results along with expository comments.

This overview is structured as follows: we will first give an essentially complete proof for refuting *semirandom* instances of *even-arity* k-XOR. As has been the trend in all the refutation results, the even-arity case happens to be significantly simpler but allows us to showcase two key ideas:

(1) The power of the Kikuchi matrix. In fact, this work can be thought of as a paean to the beautiful structure and the applications of the Kikuchi matrix and its variant that we introduce for odd-arity k-XOR. Combined with the *row bucketing* idea from [1], we can easily resolve the case of even arity k-XOR. The Kikuchi matrix was introduced by [29] to give a simpler proof of the result of [25] for refuting *fully random* instances of *even-arity* k-XOR. They left open the question of finding an analogous proof for the odd-arity case (again, for fully random CSPs) and even suggested an approach. Their approach, however, does not pan out, as we prove in Appendix A. Our Kikuchi matrix for the odd-arity case along with our analysis technique (that does not directly work with spectral norms) allows us to prove sharp trade-offs for refuting random CSPs and with additional ideas, make them work even for the significantly randomness starved semirandom and smoothed settings.

(2) The connection between refutations obtained via an appropriate norm of the Kikuchi matrix in the randomness-starved semirandom setting and the existence of *even covers* in *worst-case* hypergraphs. In this overview, we will use this connection to give a single page proof of Feige's conjecture for k-hypergraphs for k *even*. We note that this gives an interesting instance of the phenomenon where the analysis of an algorithm in a reduced-randomness setting can be used to infer a purely combinatorial property of worst-case structures.

We will then discuss our ideas for odd-arity case at a high-level by focusing on 3-XOR. As is usual in CSP refutation, even for the special case of *fully random* instances, refuting odd-arity XOR is significantly more challenging [3, 9, 11]. We introduce several new ideas to tackle the semirandom (and thus also the smoothed) case: (1) a new, suitable variant of the Kikuchi matrix, (2) the idea of *row pruning* combined with *row bucketing*, and (3) a new *regularity decomposition* for arbitrary hypergraphs.

Our proof of Feige's conjecture for odd-k-uniform hypergraphs is conceptually similar to the even case – in that it mimics the refutation argument closely – but needs all the new machinery for refutation introduced above for handling semirandom odd-arity k-XOR and must use the trace moment method (instead of the matrix Bernstein) in the step that upper bounds the spectral norm of appropriate sequence of matrices produced in our analysis. The combinatorial argument required in analyzing the trace method turns out to be somewhat more intricate in the odd arity case. We will not discuss it in this overview.

Our reduction from smoothed CSP refutation to semirandom CSP refutation is short and elementary. We will not discuss this argument in this overview.

2.1 Random 4-XOR Via the Kikuchi Matrix of [29]

Let's start by defining the Kikuchi matrix and showing how it gives a simple refutation algorithm with the optimal trade-off for random instances of even-arity k-XOR. We will focus on $k = 4$ here.

Definition 2.1 (Kikuchi Matrix). Let $N = \binom{n}{\ell}$. For a 4-XOR instance described by $\mathcal{H}$ and b_C's for $C \in \mathcal{H}$, let $A_C \in \mathbb{R}^{N \times N}$ be the matrix indexed by all possible subsets of $[n]$ of size exactly ℓ. The entry of A_C at any (S, T) where $S, T \in \binom{[n]}{\ell}$ is defined by:

$$A_C(S,T) = \begin{cases} b_C & \text{if } S \oplus T = C \\ 0 & \text{otherwise} \end{cases}$$

Here, $S \oplus T$ is the symmetric difference of the sets S, T. The level ℓ Kikuchi matrix of the instance is then simply $A = \sum_{C \in \mathcal{H}} A_C$.

Quadratic forms of the Kikuchi matrix. The quadratic forms of this matrix are closely related to the polynomial $\phi(x)$ associated with the input 4-XOR instance: namely, $\phi(x) := \frac{1}{m} \sum_{C \in \mathcal{H}} b_C \prod_{i \in C} x_i$. Notice that the non-zero entries of the matrix A correspond to pairs of sets (S, T) such that the symmetric difference of S, T is one of the clauses in the input 4-XOR instance. Observe that if $S \oplus T = C$, then $|S \cap C| = 2$, $|T \cap C| = 2$, and $|S \cap T| = \ell - 2$. In particular, each b_C appears in $\binom{4}{2} \cdot \binom{n-4}{\ell-2}$ different entries of A. Now, let $x^{\odot \ell}$ be the $\binom{n}{\ell}$-dimensional vector of degree ℓ monomials in x. That is, the entries of $x^{\odot \ell}$ are indexed by subsets of size ℓ of $[n]$ and the S-th entry of $x^{\odot \ell}$ is given by $\prod_{i \in S} x_i$. Then,

[5] Subsequent to our posting of this paper, Tim Hsieh and Sidhanth Mohanty were able to use our spectral double counting technique with the non-backtracking walk matrix of a graph to recover the sharpest known result (matching [5]) for the Moore bound for irregular graphs. We believe a similar approach might also help achieve sharper results for size of smallest even covers in hypergraphs.

we must have:

$$\binom{4}{2}\cdot\binom{n-4}{\ell-2}\phi(x)=\frac{1}{m}\left(x^{\odot\ell}\right)^{\top}Ax^{\odot\ell} \tag{2.1}$$

This immediately provides a certificate of upper bound on the value of the input instance as it must hold that

$$\max_{x\in\{-1,1\}^n}\phi(x)\leqslant\frac{1}{6m}\cdot\binom{n-4}{\ell-2}^{-1}\binom{n}{\ell}\|A\|_2\leqslant O\Big(\frac{n^2}{m\ell^2}\Big)\cdot\|A\|_2\,, \tag{2.2}$$

where $\|A\|_2$ is the spectral norm of the matrix A. If we can show that $\|A\|_2\leqslant\tilde{O}(\ell)$ w.h.p. over the draw of the hypergraph $\mathcal{H}$ and the b_C's, then, whenever $m\gg\tilde{O}(n)\cdot\frac{n}{\ell}$, the spectral norm of A provides a certificate that $\phi(x)\leqslant 0.01$ for every $x\in\{\pm1\}^n$.

It is in the ease of establishing such an upper bound on the spectral norm that the choice of Kikuchi matrix really shines! Observe that A_C's are a sequence of *independent, random* matrices and thus, one can try to apply off-the-shelf matrix concentration inequalities to bound the spectral norm of A. Instead of using the matrix Chernoff inequality as in [29], we will use the matrix Bernstein inequality below as it turns out to generalize better.

Fact 2.2 (Matrix Bernstein Inequality)**.** *Let $M_1,M_2,\ldots,$ be independent random $N\times N$ matrices with mean 0 such that $\|M_i\|_2\leqslant R$ almost surely. Let $\sigma^2=\max\{\left\|\mathbb{E}[\sum_i M_iM_i^\top]\right\|_2,\left\|\mathbb{E}[\sum_i M_i^\top M_i]\right\|_2\}$ be the* variance *term. Then, with probability at least $1-1/n^{100}$,*

$$\left\|\sum_i M_i\right\|_2\leqslant O(R\log N+\sigma\sqrt{\log N})\,.$$

Spectral norm of the Kikuchi matrix. Let's analyze $\|A\|_2$ using this inequality. First, observe that any row of A_C has at most 1 non-zero entry of magnitude 1. Since the spectral norm of a matrix is upper bounded by the maximum ℓ_1 norm of any of its rows, this immediately yields that $\|A_C\|_2\leqslant 1$. Let's now compute the "variance" term. Here's the key observation about the Kikuchi matrix that makes this analysis so simple: the matrix A_C^2 is *diagonal* for every C. To see this, observe that the entry at any (S,T) of this matrix is given by by $\sum_U\mathbb{E}\,A_C(S,U)A_C(U,T)$. A term in the summation is non-zero only if $S\oplus U=U\oplus T=C$ which can happen if and only if $T=S$.

Let's now compute the diagonals of $\mathbb{E}\sum_C A_C^2$. Notice that $A_C^2(S,S)$ equals either 1 or 0 for every C. Thus, $\sum_C A_C^2(S,S)=\deg(S)$ where

$$\deg(S):=|\{C\mid|S\cap C|=2\}|\,,$$

and so the variance term σ^2 is $\max_S\deg(S)$.

How large can this be? Since each constraint contributes $\binom{4}{2}\cdot\binom{n-4}{\ell-2}$ non-zero entries to A, $\sum_{S\in\binom{n}{\ell}}\deg(S)=\binom{4}{2}\cdot\binom{n-4}{\ell-2}m$. Thus, on an average $\deg(S)$ is $\approx m\ell^2/n^2$. When $m\sim n^2/\ell$, this is $\sim\ell$.

When $\mathcal{H}$ is a *random hypergraph* with $\sim n^2/\ell$ hyperedges, we expect $\deg(S)$ to not deviate too much from its expectation. In fact, using the Chernoff bound yields $\deg(S)\leqslant O(\ell\log n)$ for all S whp. Since $N=\binom{n}{\ell}$, this yields that $\|A\|_2\leqslant O(\log N)+O(\sqrt{\ell\log n\cdot\log N})=\tilde{O}(\ell)$ on as desired.

2.2 Semirandom Instances of 4-XOR Via Row Bucketing from [1]

Let us now conduct a post-mortem of the above proof to see where we used the randomness of the hypergraph $\mathcal{H}$. Even after fixing $\mathcal{H}$, the A_C's are independent random matrices, with all the randomness coming from the b_C's. Thus, we can still apply the matrix Bernstein inequality. The only point in the proof where we used the randomness of the hypergraph $\mathcal{H}$ was to establish that $\deg(S)=O(\ell\log n)$ for every S. So, our proof immediately extends to semirandom instances where the instance hypergraph $\mathcal{H}$ is such that $\deg(S)=O(\ell\log n)$ for every S.

This bound is delicate: when $\deg(S)=\Omega(\ell^2)$, we obtain no non-trivial refutation guarantee and even $\deg(S)\sim\ell^{1.1}$ results in a suboptimal trade-off. On the other hand, in arbitrary $\mathcal{H}$, $\deg(S)$ can be as large as m (but no larger). Further, this is a "real" issue (and not an artefact of the use of Matrix Bernstein inequality): when $\deg(S)$ is large, so is the spectral norm of A.

Key observation: only sparse vectors cause large quadratic forms. Our way forward builds on that of [1] who recently gave a polynomial time algorithm for (strongly) refuting semirandom instances of k-XOR with $\geqslant\tilde{O}(n^{k/2})$ constraints. The key observation is when $\deg(S)$ is large, the spectral norm of A is high but intuitively, the "offending" large quadratic forms are induced only by "sparse" vectors, i.e., vectors where the ℓ_2 norm is contributed by a small fraction of the coordinates. On the other hand, we only care about upper bounding quadratic forms of A on vectors where all coordinates are ±1 and are thus are maximally "non-sparse" or "flat".

Row bucketing. We can formalize this observation via *row bucketing*. Let $d_0\sim m\cdot\ell^2/n^2$ be the average value of $\deg(S)$. Let's partition the row indices in $\binom{n}{\ell}$ into multiplicatively close buckets $\mathcal{F}_0,\mathcal{F}_1,\cdots,\mathcal{F}_t$ so that for each $i\geqslant1$,

$$\mathcal{F}_i=\left\{S\mid 2^{i-1}d_0<\deg(S)\leqslant 2^id_0\right\}\,.$$

and $\mathcal{F}_0=\{S\mid\deg(S)\leqslant d_0\}$. Then, since $\deg(S)\leqslant m$ and $d_0\geqslant1$ (as $m\sim n^2/\ell$), we can take $t\leqslant\log_2 m$. Further, by Markov's inequality, $|\mathcal{F}_i|\leqslant 2^{-i}\binom{n}{\ell}=2^{-i}N$. For each $i,j\leqslant t$, let $A_{i,j}$ be the matrix obtained by zeroing out all rows not in $\mathcal{F}_i$ and all columns not in $\mathcal{F}_j$ from the Kikuchi matrix A. Then, $A=\sum_{i,j\leqslant t}A_{i,j}$.

The key observation is the following: while $A_{i,j}$ has non-zero rows and columns where $\deg(S)$ is larger by a 2^i (2^j, respectively) factor than the average, we are compensated for this by a reduction in the number of non-zero rows and columns.

Let $y\in\mathbb{R}^N$ be any vector with entries in $\{\pm1\}^N$, and let $y_{\mathcal{F}_i}$ be the vector obtained by zeroing out all coordinates of y that are not indexed by elements of $\mathcal{F}_i$. Then, we must have:

$$\max_{y\in\{\pm1\}^N}y^\top A_{i,j}y=\max_{y\in\{\pm1\}^N}(y_{\mathcal{F}_i})^\top A_{i,j}(y_{\mathcal{F}_j})\leqslant\sqrt{|\mathcal{F}_i||\mathcal{F}_j|}\cdot\left\|A_{i,j}\right\|_2\,. \tag{2.3}$$

We apply the Matrix Bernstein inequality in a similar manner to the previous analysis. The "variance" term grows by a factor of $\max\{2^i,2^j\}$ over the bound obtained for the random case. As a result, the spectral norm of $A_{i,j}$ is higher by a factor of $\max\{2^{i/2},2^{j/2}\}$. On the other hand, the effective ℓ_2 norm of the vector drops by at $2^{-(i+j)/2}$. The trade-off "breaks in our favor" and the dominating

term in the bound is $A_{0,0}$ – the spectral norm of which is at most of the same order as that of the A in the case of the previous random 4-XOR analysis! We thus obtain that $\max_{y\in\{\pm1\}^N} y^T Ay$ is $\tilde{O}(\frac{n^2}{m\ell^2}\cdot\ell)$, and so we certify that $\phi(x)\leqslant 0.01$ for every $x\in\{\pm1\}^n$.

2.3 Proving Feige's Conjecture for 4-Uniform Hypergraphs

We now discuss how the analyses of the Kikuchi matrix from the previous section relates to Feige's conjecture on even covers in 4-uniform (and in general, any even-uniform) hypergraphs. A priori, such a connection may appear rather surprising that the analysis of a super-polynomial size matrix introduced for refuting k-XOR can shed light on a purely combinatorial combinatorial fact. But we will soon see that this is yet another instance of the Kikuchi matrix doing its magic.

Recall that Feige's conjecture suggests a trade-off between the number of hyperedges and an appropriate notion of *girth* (i.e., length of the smallest cycle, or *even cover*) in hypergraphs that generalizes the classical Moore bound [5], which asserts that every graph on n vertices with $nd/2$ edges has a cycle of length $\leqslant 2\log_{d-1}(n)$. To explain our *spectral double counting* argument to prove this conjecture, it is helpful to first use it to prove a (significantly weaker) version of the Moore bound and then generalize to hypergraphs H via the "Kikuchi graph" derived from H.

Proposition 2.3 (Weak Moore bound in irregular graphs)**.** *Every graph G on n vertices and $nd/2$ edges for $d\geqslant O(\log_2^3(n))$ has a cycle of length $\leqslant\lceil 2\log_2 n\rceil$.*

Our *spectral double counting* argument counts the number of edges of G in two different ways: let A be the 0-1 adjacency matrix of G. Then, the quadratic form $\mathbf{1}^\top A\mathbf{1} = nd$. We will show that if G does not have a cycle of size $\leqslant 2\lceil\log_2 n\rceil$, then, all ±1-coordinate quadratic forms of A are at most $n\cdot\tilde{O}(\sqrt{d})$. Together, these two bounds yields the desired contradiction.

Claim 2.4 (Trace Method in the absence of even covers). Let A be the 0-1 adjacency matrix of a graph G on n vertices with $nd/2$ edges with no cycle of length $\leqslant 2r$ for $r=\lceil\log_2 n\rceil$. Then, for every $y\in\{\pm1\}^n$,

$$y^\top Ay\leqslant n\sqrt{d}\cdot O(\log_2^{1.5}(n))\,.$$

Notice that this claim immediately yields a contradiction if $nd > n\sqrt{d}\cdot O(\log_2^{1.5}(n))$, which holds if $d\geqslant O(\log_2^3 n)$, thus proving Proposition 2.3. Let's now see how to prove this claim.

Proof. The average degree of vertices in G is d. Let $\mathcal{F}_i=\{v\mid 2^i d\leqslant\deg(v)\leqslant 2^{i+1}d\}$ for each $1\leqslant i\leqslant\log_2 n$. Let $A_{i,j}$ be obtained by zeroing out all rows not in $\mathcal{F}_i$ and all columns not in $\mathcal{F}_j$ from A. Then, $A=\sum_{i,j}A_{i,j}$.

By a similar observation as in the previous subsection, we have:

$$y^\top Ay\leqslant\sum_{i,j}\sqrt{|\mathcal{F}_i||\mathcal{F}_j|}\left\|A_{i,j}\right\|_2\,. \tag{2.4}$$

Let's now bound $\left\|A_{i,j}\right\|_2$. The idea is to use the trace moment method on the matrix $A_{i,j}$: for every r, $\mathrm{tr}((A_{i,j}A_{i,j}^\top)^r)\geqslant\left\|A_{i,j}\right\|_2^{2r}$. This method is typically employed in analyzing the spectral norm of *random* matrices. But notice that $A_{i,j}$ is a *fixed* matrix – nothing random in it. Nevertheless, our key observation is if G has no cycle of length $\leqslant 2r$, then one can derive the same *exact upper bound* on $\mathrm{tr}(A_{i,j}^{2r})$ *as if it was a random "signing"* of the adjacency matrix of G.

We have:

$$\begin{aligned}&\mathrm{tr}((A_{i,j}A_{i,j}^\top)^r)\\&=\sum_{v_1,v_2,\ldots,v_{2r}}A_{i,j}(v_1,v_2)A_{i,j}(v_3,v_2)\cdots A_{i,j}(v_{2r-1},v_{2r})A_{i,j}(v_1,v_{2r})\,.\end{aligned}$$

The term corresponding to $(v_1,v_2,\ldots,v_{2r})$ contributes a non-zero value (of at most 1) to the right hand side above if and only if the sequence $\{v_i,v_{i+1}\}$ is an edge, say e_i in G for each $i\leqslant 2r$. Consider now the multiset of edges $E'=\{e_1,e_2,\ldots,e_r\}$. Since these are edges on a walk, viewing the e_i's as subsets of $[n]$ of size exactly 2, we must have that $\oplus_{i=1}^{2r}e_i=0$. Let's now prune E' by removing any e_i,e_j that are equal. We must be able to remove all edges in this procedure, as otherwise we are left with a 2-regular induced subgraph inside G, and so G must have a cycle of length $\leqslant 2r$. Thus, each edge of G occurs an even number of times in the multiset E'.

Let's now use this observation to count the number of returning walks beginning with a fixed vertex v_1. For each edge, we "match" its first occurrence along the walk with the last occurrence. There are $\frac{2r!}{r!2^r}$ different ways to select this matching. Given a matching, there are at most r distinct choices of edges to be made. We make these choices inductively along the path from v_1 to v_{2r}. At each step we can make a new choice (i.e., we are not traversing an edge that is already matched to a previously chosen edge) given our previous choices, there are at most $\Delta=\max\{2^i,2^j\}d$ choices for the edge. Summing up over all choices for v_1, we obtain that the number of non-zero contributing $2r$ length walks is at most $n\cdot\Delta^r 2^r r!$. Thus,

$$\begin{aligned}\left\|A_{i,j}\right\|_2&\leqslant\max\{2^{i/2},2^{j/2}\}\cdot n^{1/2r}d^{1/2}2^{1/2}\sqrt{r}\\&\leqslant 2d^{1/2}\max\{2^{i/2},2^{j/2}\}\sqrt{2\log_2 n}\,,\end{aligned}$$

for $r=2\lceil\log_2 n\rceil$ and large enough n.

Plugging back in (2.4) yields that

$$y^\top Ay\leqslant 2\sum_{i\leqslant j}2^{-(i+j)/2}n2^{j/2}\cdot\sqrt{2d\log_2 n}\leqslant nd^{1/2}O(\log_2^{1.5}n)\,.\quad\square$$

Let's summarize the idea of the proof: analyzing the quadratic forms on the hypercube of adjacency matrix with row bucketing yields a (significantly weaker but still non-trivial) bound on the girth of a graph with a given number of edges. This argument can possibly be sharpened (to only an absolute constant factor loss) by switching to the non-backtracking walk matrix of G (instead of the adjacency matrix) and dropping the row bucketing step. The above loose argument, however, generalizes to hypergraphs as we show below.

Lemma 2.5 (Feige's Conjecture for 4-Uniform Hypergraphs)**.** *Every 4-uniform hypergraph $\mathcal{H}$ on $[n]$ with $m\geqslant O(\frac{n^2}{\ell}\log_2^3 n)$ hyperedges has an even cover of length $O(\ell\log_2 n)$.*

For every $C\in\mathcal{H}$, let $b_C=1$ and consider the Kikuchi matrix A of the 4-XOR instance specified by $\mathcal{H}$ and b_C's. Equivalently, A is simply the adjacency matrix of the "Kikuchi graph" on vertex set $\binom{[n]}{\ell}$ where edges correspond to pairs (S,T) such that $S\oplus T=C$ for some $C\in\mathcal{H}$. The idea is to repeat the argument for the adjacency

matrix above but this time on the Kikuchi graph. The "win" in this scheme is a reduction of the problem on hypergraphs to a related problem on the associated Kikuchi graph that is significantly easier to reason about.

As in the previous section, each $C \in \mathcal{H}$ corresponds to $\binom{4}{2} \cdot \binom{n-4}{\ell-2}$ different non-zero entries in A and in particular, we have for $x = 1^n$,

$$(x^{\odot \ell})^\top A x^{\odot \ell} = 6\binom{n-4}{\ell-2}|\mathcal{H}|\,.$$

Our proof exactly mirrors the proof of the above weak Moore bound for graphs. We will show that if $\mathcal{H}$ has no even cover of length $2r$ for $r = 0.5\log_2 N$, then, $y^\top A y \leqslant \binom{n}{\ell}\tilde{O}(\ell)$ for any $y \in \{-1,1\}^N$.

Let $\deg(S) = |\{C \mid |S \cap C| = 2\}|$. Write $A = \sum_{i,j} A_{i,j}$ where $A_{i,j}$ has all rows not in $\mathcal{F}_i = \{S \mid 2^{i-1}d_0 < \deg(S) \leqslant 2^i d_0\}$ ($\mathcal{F}_0 = \{S \mid \deg(S) \leqslant d_0\}$) and all columns not in $\mathcal{F}_j$ zeroed out, where $d_0 \sim m\ell^2/n^2$. Note that $\deg(S) \leqslant m$ so the number of buckets is at most $\lceil \log_2 m \rceil$. We can now argue:

$$(y^\top A y) \leqslant \sum_{i,j} \|A_{i,j}\|_2 \cdot \sqrt{|\mathcal{F}_i||\mathcal{F}_j|}\,.$$

In the previous section, when b_C's were independent, random bits, we used the matrix Bernstein inequality to bound $\|A_{i,j}\|_2$. Here, b_C's are fixed (and equal to 1) so, of course, that strategy cannot work. Instead, our proof uses the trace moment method as in the proof of the weak Moore bound.

Proposition 2.6. *Suppose $\mathcal{H}$ has no even cover of length $2r$ for $r \leqslant \log_2 N$. Then, $\|A_{i,j}\|_2 \leqslant O(\ell \log_2 n)$.*

Proof of Proposition. As before, we use $\|A_{i,j}\|_2^{2r} \leqslant \mathrm{tr}((A_{i,j}A_{i,j}^\top)^r))$ for any $r \in \mathbb{N}$. We then have:

$$\begin{aligned}&\mathrm{tr}((A_{i,j}A_{i,j}^\top)^r))\\ &= \sum_{S_1,S_2,\ldots,S_{2r}} A_{i,j}(S_1,S_2)\cdot A_{i,j}(S_3,S_2)\cdots A_{i,j}(S_{2r-1},S_{2r})A_{i,j}(S_{2r+1},S_{2r})\,,\end{aligned}$$

where we adopt the convention that $S_{2r+1} = S_1$. Let us now analyze the right hand side of this equality. Each term in the RHS corresponds to a $2r$-tuple $(S_1, S_2, \ldots, S_{2r})$ of sets from $\binom{[n]}{\ell}$ can contributes either 0 or 1.

If a term corresponding to $(S_1, S_2, \ldots, S_{2r})$ contributes a +1, then, for each $i \leqslant 2r$, there must be a $C_i \in \mathcal{H}$ such that $S_i \oplus S_{i+1} = C_i$. Thus, each non-zero term is in bijection with $(S_1, C_1, C_2, \ldots, C_{2r})$. On the other hand, we must have that $\emptyset = \oplus_{i=1}^{2r} S_i \oplus S_{i+1} = \oplus_{i=1}^{2r} C_i$, as each S_i appears twice in $\oplus_{i=1}^{2r} S_i \oplus S_{i+1}$, and thus the total symmetric difference is $\emptyset$. Hence, a non-zero term $(S_1, C_1, C_2, \ldots, C_{2r})$ must satisfy $\oplus_{i=1}^{2r} C_i = \emptyset$.

Let us analyze such a $2r$-tuple of hyperedges. By removing equal pairs repeatedly as in the previous proof, we can conclude that since $\mathcal{H}$ has no even cover of length $\leqslant 2r$, each hyperedge in $\mathcal{H}$ occurs an even number of times in the (multi)set $\{C_1, C_2, \ldots, C_{2r}\}$.

We now count the number of $(S_1, C_1, \ldots, C_{2r})$ such that each C_i occurs an even number of times. Since C_i's occur in pairs, we can match the first occurrence of the hyperedge in the ordered set $(C_1, C_2, \ldots, C_{2r})$ to the last. There are $\leqslant 2^r r!$ different ways of selecting this matching. Given S_1 and the matching, there are at most r unique C_i's to choose. When making a choice of C_i (say), S_i is already determined by the previous choices. Thus, we have at most $\deg(S_i) \leqslant \Delta \leqslant \max\{2^i, 2^j\}d_0$ unique choices for the hyperedge C. In total, there are $\leqslant N \cdot 2^r r! \Delta^r$ non-zero terms, and so

$$\begin{aligned}\|A_{i,j}\|_2 &\leqslant N^{1/2r} 2^{1/2}\sqrt{r}\max\{2^{i/2}, 2^{j/2}\}\sqrt{d_0}\\ &\leqslant \max\{2^{i/2}, 2^{j/2}\}2\sqrt{\log_2 N}\sqrt{d_0}\,,\end{aligned}$$

for $r = 0.5\log_2 N$ and large enough n. The remaining calculation now mimics the one for Proposition 2.3 (recalling that $d_0 \sim m\ell^2/n^2$), and finishes the proof of Lemma 2.5 □

2.4 Refuting Semirandom 3-XOR Via Row Pruning

The case of odd arity XOR refutation is lot more challenging. Even in the well-studied special case of random CSP refutation and the special case of $\ell = O(1)$ (i.e., polynomial time refutation), the case of odd arity CSPs turns out to be significantly more challenging than the even case. So let us start by focusing on the case of random 3-XOR first.

As in the case of 4-XOR, we would like to begin by finding a simpler argument (compared to [25]) for the special case of *random* 3-XOR using some appropriate variant of the Kikuchi matrix. In fact, [29] attempted this by introducing a variant of the Kikuchi matrix, and suggested an explicit approach (see Section F.1 of [29]) to prove that the spectral norm of that matrix yields a refutation, but this does not work (see Appendix A). Indeed, we do not know of any reasonable variant of the Kikuchi matrix whose spectral norm yields a refutation for even *fully random* 3-XOR instances with the expected trade-off.

Instead, we will introduce a variant of the Kikuchi matrix and use it to give a refutation algorithm for *random* 3-XOR instances by relying not on the spectral norm (which is too large) but, instead, the spectral norm of a "pruned" version of the matrix. We will then discuss the remaining key ideas of *regularity decomposition* combined with row bucketing to refute semirandom odd-arity XOR.

Bipartite 3-XOR. The Kikuchi matrix we introduce relates directly to a polynomial obtained by applying the standard "Cauchy-Schwarz trick" to the input polynomial. Consider the polynomial $\psi(x) = \frac{1}{m}\sum_{C\in\mathcal{H}} b_C \prod_{i\in C} x_i$ associated with a 3-XOR instance described by a 3-uniform hypergraph $\mathcal{H}$ with m hyperedges and "right-hand sides" b_C's. For each $C \in \mathcal{H}$, let $C_{\min}$ be the minimum indexed element in C (using the natural ordering on $[n]$). Then,

$$\max_{x\in\{\pm1\}^n} \psi(x) \leqslant \max_{x,y\in\{\pm1\}^n} \frac{1}{m}\sum_{C\in\mathcal{H}} b_C y_{C_{\min}} x_{C\setminus C_{\min}}\,,$$

where each y_u is formally a new variable, but we think of y_u as equal to x_u. Let us reformulate this expression a bit: let $\mathcal{H}_u = \{C \mid C' = (C,u) \in \mathcal{H}, C'_{\min} = u\}$. Then,

$$\max_{x\in\{\pm1\}^n} \psi(x) \leqslant \max_{x,y\in\{\pm1\}^n} \frac{1}{m}\sum_{u\in[n]} y_u \sum_{C\in\mathcal{H}_u} b_{u,C} x_C\,.$$

One can think of the RHS as the polynomial associated with a *bipartite* instance of the 3-XOR problem on $2n$ variables, since every constraint uses one y variable and two x variables. Our refutation algorithm works for such bipartite instances more generally.

For such a bipartite instance, using the Cauchy-Schwarz inequality, we can derive:

$$\max_{x,y\in\{\pm1\}^n}\left(\frac{1}{m}\sum_{u\in[n]}y_u\sum_{C\in\mathcal{H}_u}b_{u,C}x_C\right)^2 \leqslant \left(\frac{n}{m^2}\sum_u\sum_{C,C'\in\mathcal{H}_u}b_{u,C}b_{u,C'}x_Cx_{C'}\right) \leqslant \frac{nm}{m^2}+\frac{n}{m^2}\left(\sum_u\sum_{C\neq C'\in\mathcal{H}_u}b_{u,C}b_{u,C'}x_Cx_{C'}\right) := \frac{n}{m}+f(x) \quad (2.5)$$

The first term on the RHS is $\leqslant \varepsilon^2/2$ if $m \geqslant 2n/\varepsilon^2$. The second term produces a $\leqslant$ 4-XOR instance.

We thus end up with a 4-XOR instance – an even arity instance – albeit with significantly less randomness than required in the argument from previous section. So, we need some different tools to refute such instances. The first of this is the following variant of the Kikuchi matrix that is designed specifically for "playing well" with the symmetries produced by the squaring step above.

Our Kikuchi matrix. Our Kikuchi matrix is indexed by subsets of size ℓ on a universe of size $2n$ – corresponding to two labeled copies of each of the original n x variables. For each $C \in \mathcal{H}$, let $C^{(1)}$ be the subset of $[n]\times[2]$ where every variable is labeled with "1", and similarly for $C^{(2)}$. This trick is done to ensure that the clauses $x_{C^{(1)}}x_{C'^{(2)}}$ form a 4-XOR instance, as now $C^{(1)}$ and $C'^{(2)}$ *by definition* cannot intersect.

For even k, the "independent" pieces in the Kikuchi matrix were the matrices A_C, one for each $C \in \mathcal{H}$. For odd k, the independence pieces will be A_u – one for each y_u because of the loss of independence due to the Cauchy-Schwarz step above.

Definition 2.7 (Kikuchi Matrix, 3-XOR)**.** Let $N = \binom{[2n]}{\ell}$. For each $u \in [n]$, let $A_u \in \mathbb{R}^{N\times N}$ be defined as follows: for each $S,T \subseteq [n]\times[2]$ of size ℓ, we will set $A_u(S,T)$ to be non-zero if there are $C,C' \in \mathcal{H}_u$ such that $S \oplus T = C^{(1)} \oplus C'^{(2)}$ and $1 = |S\cap C^{(1)}| = |S\cap C'^{(2)}| = |T\cap C^{(1)}| = |T\cap C'^{(2)}|$. That is, $A_u(S,T)$ is non-zero if each of S,T contain one variable each from $C^{(1)}$ and $C'^{(2)}$ each. In that case, we will set $A_u(S,T) = b_{u,C}\cdot b_{u,C'}$. Finally, set $A = \sum_u A_u$.

Equivalently, $A_u(S,T)$ is non-zero if there are $C,C' \in \mathcal{H}_u$ such that if the 1-labeled (respectively, 2-labeled) elements in S,T have symmetric difference C (C', respectively). This construction is important for the success of our row pruning step (which we will soon discuss) and at the same time ensures that every pair (C,C') of constraints in $\mathcal{H}_u$ contributes an equal number of non-zero entries in the Kikuchi matrix A. We note that if we do not introduce the 2 copies of each variable, the number of times a pair (C,C') appears in the matrix would depend on $|C\cap C'|$.

The quadratic forms of A relate to the value of the underlying 4-XOR instance: for $D = 4\binom{2n-4}{\ell-2}$,

$$\mathrm{val}(\phi)^2 \leqslant \varepsilon^2/4 + \mathrm{val}(f) \leqslant \varepsilon^2/4 + \frac{n}{m^2D}\left(\max_{z\in\{\pm1\}^N} z^\top Az\right).$$

Bounding $z^\top Az$. In the even arity case, we were able to obtain a refutation at this point by simply using the spectral norm of A to bound the right hand side above. However, this turns out to provably fail here. To see why, let us define the relevant notion of degree – the count of the number of non-zero entries in each row of A_u:

$$\deg(S) = |\{C,C' \in \mathcal{H}_u \mid |S\cap C^{(1)}| = |S\cap C'^{(2)}| = 1\}|$$

If we were to apply the matrix Bernstein inequality, the "almost sure" upper bound on A_u for all u is at least as large as $\sim \max_S \sqrt{\deg(S)}$ and it's not too hard to show that there are S for which this bound is at least ℓ. As a result, the best possible spectral norm upper bound that we can hope to obtain on A is $\Omega(\ell\log_2 N) = \tilde{\Omega}(\ell^2)$ – a bound that gives us no non-trivial refutation algorithm.

Row pruning. The key observation that "rescues" this bad bound is that $\deg(S)$ cannot be large for too many rows. To see why, consider the random variable that selects a uniformly random $S \in \binom{[2n]}{\ell}$ and outputs $\deg(S)$. This can be well approximated (for our purposes) by random set where every element is included independently with probability $\sim \ell/2n$. The expectation of $\deg(S)$ on this distribution is $O(1)$. By relying on the fact that $|C\cap C'| = \emptyset$ in $\mathcal{H}_u$ for almost all pairs with high probability, $\mathbf{Var}[\deg(S)] = O(1)$. A Chernoff bound yields that the fraction of S for which $|\{C \in \mathcal{H}_u \mid |S\cap C| > O(\log n)\}|$ is inverse polynomially small in n. A union bound on all u then shows the fraction of rows that are "bad" for any u is at most an inverse polynomial.

It turns out we can ignore such "bad" rows with impunity. This is because we are interested in certifying upper bounds on quadratic forms of A over "flat" vectors again and we can argue that removing "bad" rows cannot appreciably affect them. For the "residual matrix", we can now apply the matrix Bernstein inequality and finish off the proof! The execution here requires *row bucketing* with respect to a combinatorial parameter called the butterfly degree (generalizing a similar notion in [1]) that controls the variance term in the analysis.

Extending to semirandom instances. Looking back, the previous analysis uses that the graphs $\mathcal{H}_u$'s obtained from the random 3-uniform hypergraph $\mathcal{H}$ satisfy a "spread" condition: there are few to none distinct pairs $C,C' \in \mathcal{H}_u$ such that $C\cap C' \neq \emptyset$. This notion of *regularity* is the precise pseudo-random property of $\mathcal{H}$ that is enough for our argument (i.e. the row pruning step) above to go through.

For the case of 3-XOR, such a regularity property is relatively easy to ensure by a certain ad hoc argument: if too many pairs $C,C' \in \mathcal{H}_u$ happen to share a variable, then, "resolving" them yields a system of 2-XOR constraints. Refutation in the special case of 2-XOR is easy using the Grothendieck inequality; this has been observed in several works, including [1, 12]. Indeed, this was roughly the strategy employed in the recent work [1] for the case of $\ell = O(1)$ for semirandom k-XOR. In fact, in the $\ell = O(1)$ regime, it turns out that one can reduce k-XOR for all k to the case of 3-XOR and get the right trade-off; thus, such a decomposition for 3-XOR is enough for the argument of [1] to go through for all k.

2.5 Handling k-XOR for $k > 3$: Hypergraph Regularity

When $\ell \gg O(1)$, the case of higher arity k does not reduce to $k = 3$. Once again, working through the case of random k-XOR

inspires our more general argument. We work with a generalization of the Kikuchi matrix introduced in the previous section for the case of $k = 3$. When analyzing the row pruning step, we need to rely on certain tail inequalities for low-degree polynomials that depends on the "spread" of the hypergraph defined by the indices of the non-zero coefficients in the polynomial. We use the result of Schudy and Sviridenko [27] that builds on an influential line of work on concentration inequalities for polynomials with combinatorial structure in the monomials begun by [19]. Our application of this inequality is rather delicate and as a result, we need a significantly stricter notion of *regularity* – we call this (ε, ℓ)-regularity – for our row pruning argument to go through.

Hypergraph regularity decomposition. Roughly speaking the notion of (ε, ℓ)-regularity (indexed by the parameter ℓ and an accuracy bound ε) we need demands that each for each subset $Q \subseteq [n]$, the number of hyperedges $C \in \mathcal{H}_u$ such that $Q \subseteq C$ is bounded above by an appropriate function of m, n and ℓ. Random hypergraphs $\mathcal{H}$ satisfy such a regularity property naturally.

In order to handle arbitrary hypergraphs, we introduce a new *regularity decomposition* for hypergraphs. Our regularity decomposition is based on a certain *bipartite contraction* operation that takes a bipartite hyperedge $(u, C) \in \mathcal{H}$ and a subset $Q \subseteq C$ and replaces it with $((u, Q), C \setminus Q)$. This operation should be thought of as "merging" all the elements in Q and u into a new single element (u, Q) and obtaining a smaller arity hyperedge in a variable extended space.

We give a greedy (and efficient) algorithm that starts from a k-uniform hypergraph and repeatedly applies bipartite contraction operations to obtain a sequence of k'-uniform hypergraphs for $k' \leqslant k$ along with some "error" hyperedges, with the property that each of the k'-uniform hypergraphs produced are (ε, ℓ)-regular. Each of the k'-uniform hypergraphs produced is naturally associated with a k'-XOR instance related to the input k-XOR instance. We show that refuting each of these output instances yields a refutation for the original k-XOR instance.

Cauchy-Schwarz even in the even-arity setting. Unlike in the case of 3-XOR where the resulting bipartite 3-XOR instance had an equal number of y and x variables above, the bipartite k'-XOR instances produced via our regularity decomposition are *lopsided* – the number of y variables can be polynomially larger in n than the number n of the x variables. A naive bound on the number of constraints required to refute such instances is too large to yield the required trade-off, even in the case for even k.

Instead (and in contrast to all previous works on CSP refutation), we show that an appropriate application of the "Cauchy-Schwarz" trick above to even-arity k-XOR instances allows us to "kill" the y_u's appearing in the polynomial, leaving us with only a polynomial in the x_i's. This is a rather different usage of the technique – in prior works (and as in the case of 3-XOR highlighted above), it was instead used to build the right "square" matrices for obtaining spectral refutations of the associated CSP instances when k is odd.

ACKNOWLEDGMENTS

Venkatesan Guruswami is supported in part by NSF grants CCF-1814603 and CCF-1908125, and a Simons Investigator Award. Pravesh K. Kothari is supported by NSF CAREER Award #2047933 and an award from the Google Research Scholar program. Peter Manohar is supported in part by an ARCS Scholarship, NSF Graduate Research Fellowship (under Grant No. DGE1745016) and NSF CCF-1814603. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

A ANALYZING THE [29] APPROACH FOR RANDOM 3-XOR

In this section, we will prove the approach suggested by [29] (in their Appendix F.1, F.2) for strongly refuting random k-XOR with k odd does not yield the right trade-off for m as a function of n, ℓ. Our proof reduces to showing that a certain matrix defined in [29] does not have small spectral norm. For simplicity, we present the argument for $k = 3$.

First, we give a brief overview of their approach. Let ψ be a random 3-XOR instance in n variables and m clauses, with hypergraph $\mathcal{H}$ and coefficients $\{b_C\}_{C \in \mathcal{H}}$. We will assume that each pair $C_1 \neq C_2 \in \mathcal{H}$ has $|C_1 \cap C_2| \leqslant 1$, as this holds with high probability provided that $m \ll n^2$ (and recall that we are working in the regime of $m \sim n^{1.5}$ or smaller, as for $m \gg n^{1.5}$ there is a polynomial-time refutation [1]).

The construction of [29] is as follows. First, partition the hyperedges $\mathcal{H}$ arbitrarily into $\mathcal{H}_1, \ldots, \mathcal{H}_n$, such that if $C \in \mathcal{H}_u$ then $u \in C$. From now on, we shall think of $\mathcal{H}$ as $\cup_{u=1}^n \mathcal{H}_u$. We note that our lower bound will hold regardless of the choice of the partition here.

Next, let ψ be the polynomial $\psi(x) := \frac{1}{m} \sum_{C \in \mathcal{H}} b_C x_C$, where $x_C := \prod_{i \in C} x_i$. Applying the Cauchy-Schwarz inequality, we have that

$$\psi(x)^2 \leqslant \frac{1}{m} \sum_{u=1}^{n} x_u^2 + \frac{n}{m^2} \sum_{u=1}^{n} \sum_{C \neq C' \in \mathcal{H}_u} b_C b_{C'} x_{C \setminus \{u\}} x_{C' \setminus \{u\}}$$
$$= \frac{n}{m} + f(x),$$

where $f(x) := \frac{n}{m^2} \sum_{u=1}^{n} \sum_{C \neq C' \in \mathcal{H}_u} b_C b_{C'} x_C x_{C'}$.

We now recall the following definition from [29].

Definition A.1. Let $\ell \in \mathbb{N}$, and let $\mathcal{H} = \cup_{u=1}^n \mathcal{H}_u$ be a 3-uniform hypergraph. For $\vec{S}, \vec{T} \in [n]^\ell$ and $C_1 = \{u, v_1, w_1\}, C_2 = \{u, v_2, w_2\} \in \mathcal{H}_u$ with $\{v_1, w_1\} \cap \{v_2, w_2\} = \emptyset$, we write $\vec{S} \overset{C_1, C_2}{\leftrightarrow} \vec{T}$ if there exist $i \neq j \in [\ell]$ such that (1) $\vec{S}_t = \vec{T}_t$ for all $t \neq i, j$, and (2) $\{\vec{S}_i, \vec{S}_j\}$ contains exactly one element from each of $\{v_1, w_1\}$ and $\{v_2, w_2\}$, and $\{\vec{T}_i, \vec{T}_j\}$ contains the other two remaining elements. We note that if $\vec{S} \overset{C_1, C_2}{\leftrightarrow} \vec{T}$ for some C_1, C_2, then we cannot have $\vec{S} \overset{C_1', C_2'}{\leftrightarrow} \vec{T}$ for any other pair C_1', C_2'.

Let $A_u \in \mathbb{R}^{n^\ell \times n^\ell}$ be the matrix where $A_u(\vec{S}, \vec{T}) = b_{C_1} b_{C_2}$ if $\vec{S} \overset{C_1, C_2}{\leftrightarrow} \vec{T}$ for some $C_1 \neq C_2 \in \mathcal{H}_u$, and 0 otherwise, and let $A := \sum_{u=1}^n A_u$.

It is simple to observe that $\max_{x \in \{\pm 1\}^n} f(x) \leqslant \frac{n}{m^2} \cdot O(\frac{n^2}{\ell^2}) \|A\|_2$, as $\frac{m^2}{n} f(x) = \frac{1}{4\binom{\ell}{2}(n-4)^{\ell-2}} (x^{\otimes \ell})^\top A x^{\otimes \ell}$ for all $x \in \{\pm 1\}^n$ because each pair $C_1 \neq C_2 \in \mathcal{H}_u$ "appears" exactly $4\binom{\ell}{2}(n-4)^{\ell-2}$ times in

the matrix A. Thus, in order to get the correct $m = n^{1.5}/\sqrt{\ell}$ trade-off, we need to show that $\|A\|_2 \leqslant O(\ell)$, with high probability over $\mathcal{H}$ and the b_C's.

We prove that $\|A\|_2$ is in fact *large* with high probability, and so the above approach of [29] fails. Formally, we prove that with high probability, the matrix A has a spectral norm $\Omega(\min(\ell^2, \frac{m^2}{n^2}))$, which has the following implications. If the minimum is $\frac{m^2}{n^2}$, then the upper bound certified on f is $\Omega(n/\ell^2)$, and thus the upper bound certified on ψ is $\Omega(\sqrt{n}/\ell)$. This is not useful, as it is greater than 1 when $\ell \ll n$. If the minimum is ℓ^2, then we certify a good upper bound on f (and therefore also ψ) only if $m \geqslant n^{1.5}$, which is higher than the desired threshold of $n^{1.5}/\sqrt{\ell}$.

Proposition A.2. *Let ψ be a random 3-XOR instance with n variables and m constraints, with constraint hypergraph $\mathcal{H} = \cup_{u=1}^{n} \mathcal{H}_u$ and coefficients $\{b_C\}_{C\in\mathcal{H}}$. Suppose that $2n \leqslant m \ll n^2$. Let $\ell \leqslant n$. Then, $\|A\|_2 \geqslant \binom{\ell'}{2}$, where $\ell' := \min(\lceil \frac{m}{2n} \rceil, \ell)$.*

We note that the Proposition A.2 holds regardless of the choice of the partitioning of $\mathcal{H}$ into the $\mathcal{H}_u$'s, and also for any choice of the b_C's (and so, in particular, for random b_C's). We also note that Proposition A.2 holds for arbitrary $\mathcal{H}$, provided that $|C_1 \cap C_2| \leqslant 1$ for all $C_1 \neq C_2 \in \mathcal{H}$; this holds with high probability for a random $\mathcal{H}$, provided that $m \ll n^2$.

Proof. With high probability over $\mathcal{H}$, we may assume that $|C_1 \cap C_2| \leqslant 1$ for all $C_1 \neq C_2 \in \mathcal{H}$ for all $C_1 \neq C_2 \in \mathcal{H}$. We proceed, assuming that this holds.

As $m \geqslant 2n$, there must exist some variable $u \in [n]$ that appears in at least $\frac{m}{n}$ constraints. Hence, there must exist at least $\lceil \frac{m}{2n} \rceil$ constraints that include u and all have the same sign $b \in \{\pm 1\}$.

Let $\ell' := \min(\lceil \frac{m}{2n} \rceil, \ell)$. By the above, we have ℓ' constraints $\{C_i\}_{i\in[\ell']} = \{\{u, v_i, w_i\}\}_{i\in[\ell']}$ such that $b_{C_i} = b$ for all i. Furthermore, by assumption on $\mathcal{H}$, we have $|C_i \cap C_j| \leqslant 1$ for all $i \neq j \in [\ell']$. As $u \in C_i \cap C_j$, it thus follows that $\{v_i, w_i\} \cap \{v_j, w_j\} = \emptyset$. Let $z \in [n]$ be arbitrary. Let $\mathcal{R}$ denote the set of tuples $(r_1, \ldots, r_{\ell'}, z, \ldots, z) \in [n]^\ell$ such that $r_i \in \{v_i, w_i\}$ for all $i \in [\ell']$. We note that the element z merely pads each tuple in $\mathcal{R}$ to have length exactly ℓ when $\ell' < \ell$.

Let M be the submatrix of A indexed by the tuples in $\mathcal{R}$. Note that M is a $2^{\ell'} \times 2^{\ell'}$ matrix, as $|R| = 2^{\ell'}$. Let $\vec{S} = (r_1, \ldots, r_{\ell'}, z, \ldots, z)$ be a row in M. We will show that each row of M has exactly $\binom{\ell'}{2}$ nonzero entries, each of which is 1.

First, let us consider the contribution to M from A_u. Fix a row $\vec{S} \in \mathcal{R}$. For each pair of indices $i \neq j \in [\ell']$, we can replace the i-th and j-th elements of $\vec{S}$ with the elements of $\{v_i, w_i\}$ and $\{v_j, w_j\}$ not used in $\vec{S}$, and this will yield some $\vec{T} \in \mathcal{R}$ with $\vec{S} \overset{\{u,v_i,w_i\},\{u,v_j,w_j\}}{\leftrightarrow} \vec{T}$. Hence, $A_u(\vec{S}, \vec{T}) = b^2 = 1$. Any other $\vec{T} \in \mathcal{R}$ will differ from $\vec{S}$ by at least 2 elements, and thus we must have $A_u(\vec{S}, \vec{T}) = 0$ for such $\vec{T}$.

Next, let us consider the contribution to M from $A_{u'}$ for $u' \neq u$. Fix a row $\vec{S} \in \mathcal{R}$. It suffices to only consider $\vec{T}$ obtained by swapping the i-th and j-th entries of $\vec{S}$, for some $i \neq j \in [\ell']$, as above. If $A_{u'}(\vec{S}, \vec{T})$ is nonzero, then we must have $\vec{S} \overset{\{u',v_i,w_i\},\{u',v_j,w_j\}}{\leftrightarrow} \vec{T}$, and thus that $\{u', v_i, w_i\}, \{u', v_j, w_j\} \in \mathcal{H}_{u'}$. However, this implies that $|\{u, v_i, w_i\}, \{u', v_i, w_i\}| = 2 > 1$, which contradicts our assumption on $\mathcal{H}$.

We have thus shown that the matrix M is $2^{\ell'} \times 2^{\ell'}$, with each row having exactly $\binom{\ell'}{2}$ nonzero entries, all of which are 1. It thus follows that $\|A\|_2 \geqslant \|M\|_2 \geqslant (1^{2^{\ell'}})^\top M 1^{2^{\ell'}}/2^{\ell'} = \binom{\ell'}{2}$, which finishes the proof. □

REFERENCES

[1] Jackson Abascal, Venkatesan Guruswami, and Pravesh K. Kothari. 2021. Strongly refuting all semi-random Boolean CSPs. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*. SIAM, 454–472.

[2] Kwangjun Ahn. 2020. A Simpler Strong Refutation of Random k-XOR. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference (LIPIcs, Vol. 176)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2:1–2:15.

[3] Sarah R. Allen, Ryan O'Donnell, and David Witmer. 2015. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*. 689–708.

[4] Noga Alon and Uriel Feige. 2009. On the power of two, three and four probes. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA, 346–354.

[5] Noga Alon, Shlomo Hoory, and Nathan Linial. 2002. The Moore bound for irregular graphs. *Graphs Combin.* 18, 1 (2002), 53–57.

[6] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. 2020. Improved bounds for the sunflower lemma. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*. ACM, 624–630.

[7] Sanjeev Arora, David R. Karger, and Marek Karpinski. 1995. Polynomial time approximation schemes for dense instances of *NP*-hard problems. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing, 29 May-1 June 1995, Las Vegas, Nevada, USA*. ACM, 284–293.

[8] Boaz Barak, Siu On Chan, and Pravesh Kothari. 2015. Sum of Squares Lower Bounds from Pairwise Independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*. 97–106.

[9] Boaz Barak and Ankur Moitra. 2016. Noisy Tensor Completion via the Sum-of-Squares Hierarchy. In *Proceedings of the 29th Annual Conference on Learning Theory*. 417–445.

[10] Siu On Chan. 2013. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*. 447–456.

[11] Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. 2004. Strong Refutation Heuristics for Random k-SAT. In *Approximation, Randomization, and Combinatorial Optimization, Algorithms and Techniques (Lecture Notes in Computer Science, Vol. 3122)*. Springer, 310–321.

[12] Uriel Feige. 2007. Refuting smoothed 3CNF formulas. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*. 407–417.

[13] Uriel Feige. 2008. Small linear dependencies for binary vectors of low weight. In *Building bridges*. Bolyai Soc. Math. Stud., Vol. 19. Springer, Berlin, 283–307. https://doi.org/10.1007/978-3-540-85221-6_9

[14] Uriel Feige, Jeong Han Kim, and Eran Ofek. 2006. Witnesses for non-satisfiability of dense random 3CNF formulas. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*. 497–508.

[15] Uriel Feige and Tal Wagner. 2016. Generalized Girth Problems in Graphs and Hypergraphs.

[16] Dimitris Fotakis, Michael Lampis, and Vangelis Th. Paschos. 2016. Sub-exponential Approximation Schemes for CSPs: From Dense to Almost Sparse. In *33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016, February 17-20, 2016, Orléans, France (LIPIcs, Vol. 47)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 37:1–37:14.

[17] Russell Impagliazzo and Ramamohan Paturi. 2001. On the Complexity of k-SAT. *J. Comput. Syst. Sci.* 62, 2 (2001), 367–375.

[18] Domingos Dellamonica Jr., Penny E. Haxell, Tomasz Luczak, Dhruv Mubayi, Brendan Nagle, Yury Person, Vojtech Rödl, Mathias Schacht, and Jacques Verstraëte. 2012. On Even-Degree Subgraphs of Linear Hypergraphs. *Comb. Probab. Comput.* 21, 1-2 (2012), 113–127. https://doi.org/10.1017/S0963548311000575

[19] Jeong Han Kim and Van H Vu. 2000. Concentration of multivariate polynomials and its applications. *Combinatorica* 20, 3 (2000), 417–434.

[20] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. 2017. Sum of squares lower bounds for refuting any CSP. In *STOC*. ACM, 132–145.

[21] A. Lubotzky, R. Phillips, and P. Sarnak. 1988. Ramanujan graphs. *Combinatorica* 8, 3 (1988), 261–277. https://doi.org/10.1007/BF02126799

[22] G. A. Margulis. 1988. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii* 24, 1 (1988), 51–60.

[23] Dana Moshkovitz and Ran Raz. 2010. Two-query PCP with subconstant error. *J. ACM* 57, 5 (2010), Art. 29, 29. https://doi.org/10.1145/1754399.1754402

[24] Assaf Naor and Jacques Verstraëte. 2008. Parity check matrices and product representations of squares. *Combinatorica* 28, 2 (2008), 163–185. https://doi.org/10.1007/s00493-008-2195-2

[25] Prasad Raghavendra, Satish Rao, and Tselil Schramm. 2017. Strongly refuting random CSPs below the spectral threshold. In *STOC*. ACM, 121–131.

[26] Anup Rao. 2019. Coding for Sunflowers. *CoRR* abs/1909.04774 (2019). arXiv:1909.04774 http://arxiv.org/abs/1909.04774

[27] Warren Schudy and Maxim Sviridenko. 2012. Concentration and Moment Inequalities for Polynomials of Independent Random Variables. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms* (Kyoto, Japan) *(SODA '12)*. Society for Industrial and Applied Mathematics, USA, 437–446.

[28] Daniel A. Spielman and Shang-Hua Teng. 2003. Smoothed analysis: motivation and discrete models. In *Algorithms and data structures*. Lecture Notes in Comput. Sci., Vol. 2748. Springer, Berlin, 256–270. https://doi.org/10.1007/978-3-540-45078-8_23

[29] Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. 2019. The Kikuchi Hierarchy and Tensor PCA. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*. IEEE Computer Society, 1446–1468.

[30] David Witmer. 2017. *Refutation of random constraint satisfaction problems using the sum of squares proof system.* Ph.D. Dissertation. Carnegie Mellon University.