Goldreich's PRG: Evidence for near-optimal polynomial stretch

Ryan O'Donnell*
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA
odonnell@cs.cmu.edu

David Witmer†
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA
dwitmer@cs.cmu.edu

***Abstract*—Furthering the study of cryptography in NC^0, we give new evidence for the security of Goldreich's candidate pseudorandom generator with near-optimal, polynomial stretch. Our evidence consists both of security against subexponential-time $\mathbb{F}_2$-linear attacks as well as subexponential-time attacks using SDP hierarchies such as Sherali–Adams$^+$ and Lasserre/Parrilo. More specifically, instantiating Goldreich's generator with the predicate $P(x_1, \dots, x_5) = x_1 + x_2 + x_3 + x_4 x_5 \pmod 2$ we get a candidate 5-local PRG which stretches n bits to $n^{1.499}$ bits and which is secure against both $\mathbb{F}_2$-linear attacks and attacks based on the Lasserre/Parrilo SDP hierarchy. Previous works with such small locality only gave stretch $n^{1.249}$ and were only shown to be secure against $\mathbb{F}_2$-linear attacks. Our result is essentially optimal, as known SDP/spectral techniques show the generator would not be secure if used with stretch $\Theta(n^{3/2} \log n)$.**

More generally, when (a slight variant of) Goldreich's generator is used with a local predicate $P(x)$ which is $(t-1)$-wise independent, we show that one can allow stretch $n^{t/2-\epsilon}$ for any $\epsilon > 0$ while resisting subexponential-time attacks based on the Sherali–Adams$^+$ SDP hierarchy. Again, this is amount of stretch is (potentially) optimal due to known SDP/spectral methods which succeed at stretch $\Theta(n^{t/2} \log n)$. Finally, for a large family of predicates we also extend this result to security against the much stronger Lasserre/Parrilo SDP hierarchy.

I. Introduction

A major goal of cryptography is the construction of very efficient, secure cryptographic primitives; e.g., one-way functions (OWFs) or pseudorandom generators (PRGs). One interpretation of "very efficient" — suggested as early as the mid-'80s [1] — is "highly parallelizable" or "in NC^1". An even more ambitious goal, suggested in works by Goldreich [2] and by Cryan and Miltersen [3] from the early 2000's, is that of *cryptography in* NC^0. By this is meant the possibility of, say, PRGs $f : \{0,1\}^n \to \{0,1\}^m$ $(m > n)$ with *constant locality*. We say that f is *k-local* if each output bit $f(x)_j$ depends on at most k input bits x_i.

*Supported by NSF grants CCF-0747250 and CCF-1116594 and a grant from the MSR–CMU Center for Computational Thinking.

†Supported by NSF grants CCF-0747250 and CCF-1116594, a grant from the MSR–CMU Center for Computational Thinking, and an NSF Graduate Research Fellowship.

A celebrated work of Applebaum, Ishai, and Kushilevitz [4] showed that under standard cryptographic assumptions (e.g., hardness of factoring or lattice problems) there are secure PRGs $f : \{0,1\}^n \to \{0,1\}^m$ computable in NC^0 with locality as small as 4. Unfortunately, the PRGs they construct have only sublinear stretch; i.e., $m \le n + o(n)$. This deficiency is inherent in the [4] methodology, and in fact it's known [5] that a 4-local PRG can achieve stretch at best $O(n)$. On the other hand, it would be quite desirable to have a cryptographically secure PRG with constant locality and *polynomial* stretch; i.e., $m = n^{1+\epsilon}$ for a positive constant ϵ. An example application would be secure two-party computation with only constant overhead [6].

Goldreich's generator.: The main candidate for such a constant-locality, polynomial-stretch PRG was proposed by Goldreich [2]; see also [7]. Goldreich's suggestion was the following (for more precise details, see Section II): To construct a potential k-local OWF/PRG mapping n bits to $m \ge n$ bits, first fix a Boolean predicate $P : \{0,1\}^k \to \{0,1\}$ and also fix once and for all a list $S^1, \dots, S^m \in [n]^k$ of randomly chosen k-tuples. (Alternatively, it may be enough that the associated n-vertex, m-edge, k-uniform hypergraph be a sufficiently good expander.) Then on input $x \in \{0,1\}^n$, the jth output bit of the generator is defined to be P applied to the S^j bits of x. It's also fruitful to think of the output of this generator as a random "planted instance" of the k-CSP (constraint satisfaction problem) with constraint predicate P. There is one twist to this CSP viewpoint, though: rather than a traditional CSP instance specifying m satisfied constraints such as

$$P(x_1, x_9, x_2, x_{11}, x_{15}) = 1$$
$$P(x_7, x_3, x_n, x_{12}, x_2) = 1$$
$$P(x_{45}, x_5, x_{n-2}, x_8, x_1) = 1$$
$$\dots$$

the output of Goldreich's generator should be viewed as a list of m constraints together with a 0/1 "right-hand side" specifying whether or not the constraint is

satisfied; e.g.,

$$P(x_1, x_9, x_2, x_{11}, x_{15}) = 0$$
$$P(x_7, x_3, x_n, x_{12}, x_2) = 1$$
$$P(x_{45}, x_5, x_{n-2}, x_8, x_1) = 0$$
$$\dots$$

(Alternatively, this can be viewed as a CSP with both P and $\neg P$ constraints.) Roughly speaking, the generator is a OWF if these random planted CSP instances are hard to solve, and it's a PRG if these random planted CSP instances are hard to distinguish from completely random instances (i.e., where the right-hand sides are uniformly random).

Naturally, the security of Goldreich's candidate PRG depends on the predicate P as well as the stretch m. A number of *negative* results are known; for example, if P is an $\mathbb{F}_2$-linear function (i.e., an XOR predicate) then we don't even get a OWF for any m, since one can efficiently invert a system of $\mathbb{F}_2$-linear equations. Further negative results are reviewed in Section II-I, but the most important one to mention is that if P fails to be "t-wise independent" — equivalently, if P has a nonzero Fourier coefficient of degree at most t — then Goldreich's PRG is not secure when $m = \widetilde{\Theta}(n^{t/2})$.

These negative results imply that if we want a k-local PRG with superlinear stretch we'll need a non-linear predicate P of arity $k \geq 5$ which is at least 3-wise independent. There is essentially only one such predicate with $k = 5$, which we call "TSA" (standing for "Tri-Sum-And", the name given to the predicate in the inapproximability work [8]):

$$\text{TSA}(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2 + x_3 + x_4 x_5 \pmod 2.$$

Positive evidence.: It is plausible that Goldreich's generator, instantiated with TSA, is a 5-local PRG with stretch $m = O(n^{1.499})$. In this paper we present multiple forms of evidence supporting this possibility. Before stating our results, we briefly review some previous work supporting the security of Goldreich's generator. In 2003, Mossel, Shpilka, and Trevisan [5] showed that a variant of Goldreich's generator (using a not-completely-random 5-uniform hypergraph), when instantiated with the TSA predicate and $m = O(n^{1.249})$, strongly fools all $\mathbb{F}_2$-linear tests. More precisely, it's an "ϵ-biased generator" with $\epsilon = 2^{-n^{\Omega(1)}}$. This gives some evidence of security, with a stretch which is polynomial but short of the potential $O(n^{1.499})$. Recent work of Applebaum, Bogdanov, and Rosen [9] extended this result to hold for all so-called "non-degenerate" predicates P, even for purely random k-uniform hypergraphs.

The property of ϵ-biasedness is only one necessary condition for PRGs. In the context of "CSP-like" PRGs it's natural to investigate attacks involving traditional algorithmic methods for CSPs. Cook, Etesami, Miller, and Trevisan [10] (building on [11]) showed that a family of "myopic backtracking" (DPLL-like) algorithms requires exponential time to invert the TSA-based generator when $m = n$. It should be mentioned that these sorts of algorithms also fail when the predicate P is purely $\mathbb{F}_2$-linear, even though Goldreich's generator is easy to break in this case.

Finally, we mention that Applebaum has recently shown [12] using standard cryptographic techniques that if Goldreich's generator with TSA is indeed a PRG (or even a OWF) with stretch even $n^{1.01}$, then for every $b, c > 1$ there exist PRGs with stretch n^c, distinguishing probability $1/n^b$, and locality $b^{O(\log c)}$. Unfortunately, this locality is rather large: in practice, the $O(\cdot)$ hides an infeasibly large constant, and in theory, getting negligible distinguishing probability requires super-constant locality.

A. Our results

In this work we give evidence showing that Goldreich's generator, when instantiated with a $(t-1)$-wise independent predicate P, may be secure with polynomial stretch almost as large as the known barrier: $m = O(n^{t/2-\delta})$ for any $\delta > 0$. In this section we will give informal statements of our results; more precise statements will appear in the sections that follow.

Let's begin with the particular case of $P = \text{TSA}$. Improving on the results in [5], [9], we show:

Theorem I.1. *Goldreich's generator with $P = \text{TSA}$ and $m = O(n^{1.499})$ is a 5-local ϵ-biased generator for $\epsilon = 2^{-n^{\Omega(1)}}$.*

The amount of stretch in this theorem is essentially optimal, as it is known (see Theorem II.11) that Goldreich's generator cannot be cryptographically secure with a 5-local predicate for $m = \widetilde{\Theta}(n^{3/2})$, due to the existence of an attack based on SDP/spectral methods.

As mentioned, being ϵ-biased (i.e., secure against $\mathbb{F}_2$-linear combinations of output bits) is only one very particular requirement for a cryptographic PRG. Especially for Goldreich's "CSP-like" construction, more strong evidence for the PRG's security would come from the failure of "traditional algorithmic tools for CSPs". We propose *semidefinite programming (SDP) hierarchies* as a natural and powerful class of algorithmic attacks to rule out. Briefly, we will be considering the "basic" SDP hierarchy known as Sherali–Adams$^+$ (SA$^+$) [13], [14], [15], as well as the extremely powerful Lasserre/Parrilo/Sum-of-Squares SDP hierarchy [16], [17]. (For more details, see Section II.) Both of these hierarchies are parameterized by a "rounds/degree" parameter $r \in \mathbb{N}$; as r increases we

get stronger and stronger SDPs but the running time increases as $n^{O(r)}$.

Considering SDP-based attacks on Goldreich's PRG is very natural, due to their strength in solving CSPs. For example, the SA$^+$ hierarchy is known to encapsulate many "local" CSP algorithms such as the "k-consistency" algorithm. In particular, constantly many rounds of SA$^+$ are known to decide satisfiability of any "bounded width" CSP [18] (even "robustly" [19]). They are also known to decide satisfiability of *any* CSP instance whose primal instance has constant treewidth [20]. Raghavendra's deep theory of CSPs [21] also shows that SA$^+$ gives essentially the optimal approximation algorithm for all CSPs assuming the Unique-Games Conjecture. The Lasserre/Parrilo hierarchy is known to be even more powerful (see, e.g., [22]), with constantly many rounds sufficing to well-solve all known instances of the Unique-Games problem itself. Particularly relevant to the cryptographic considerations in this paper is the following fact: the attack showing that Goldreich's PRG is not secure for $m = \widetilde{\Theta}(n^{t/2})$ if P is not t-wise independent relies on an SDP/spectral algorithm (implementable with SA$^+$). The only deficiency of SDP hierarchies in the context of CSPs seems to be that they are fooled by purely linear predicates; i.e., they cannot simulate Gaussian elimination.

In light of the power of SDP algorithms in the context of CSPs, the following result of ours may be considered good evidence in favor of the security of Goldreich's PRG:

Theorem I.2. *For $t \geq 3$, fix a $(t-1)$-independent predicate P and any $\delta > 0$. Suppose we instantiate Goldreich's PRG with $m = O(n^{t/2-\delta})$ and with certain "bad tuples" removed. (With high probability there are only $o(m)$ such bad tuples, and they may be initially recognized and removed in polynomial time, once and for all.) Then the PRG is perfectly secure against the attack based on computing the SA$^+$ relaxation value, even for $n^{\Omega(\delta)}$ rounds.*

Again, we remark that the stretch $m = O(n^{t/2-\delta})$ is essentially optimal. Our analysis for this theorem follows work of [23] fairly closely.

Finally, we can significantly strengthen Theorem I.2 for TSA and for a large family of TSA-like predicates. For these predicates we can get near-optimal stretch with perfect security against the much stronger Lasserre/Parrilo SDP hierarchy, simply by using a k-*partite* random hypergraph structure. The following theorem was jointly observed by the authors together with Boaz Barak, Siu On Chan, and Li-Yang Tan:

Theorem I.3. *If Goldreich's PRG is instantiated with P of the the form $P(x) = x_1 + \cdots + x_t + Q(x_{t+1}, \ldots, x_k)$ (mod 2), $m = O(n^{t/2-\delta})$, and a random n-vertex m-edge, k-*partite *hypergraph, then it is perfectly secure against attacks based on computing the Lasserre/Parrilo relaxation value, even for $n^{\Omega(\delta)}$ rounds.*

We remark that it is a seemingly very difficult question to obtain this result, even for $P = \text{TSA}$, when the random k-partite hypergraph is replaced simply with a random k-uniform hypergraph. On the other hand, for practical cryptographic purposes there seems to be no reason *not* to use a k-partite hypergraph structure for the PRG; in particular, our Theorem I.1 continues to hold in this setting (and in fact is slightly easier to prove).

B. Organization

In Section II, we give some definitions and background. We prove Theorem I.1 in Section III, Theorem I.2 in Section IV, and Theorem I.3 in Section V.

II. Definitions and preliminaries

A. Distributions on graphs and hypergraphs

We will need the following two natural distributions on graphs:

1) Let $G(n,m)$ be the uniform distribution on multigraphs with n vertices and m edges. We will also consider the set of edges to be ordered.
2) Let $B(n,m)$ be the uniform distribution on bipartite multigraphs with $2n$ vertices such that the vertices are partitioned into two sets of n vertices and every edge contains exactly one vertex from each set. Again, the set of edges is ordered.

We will call these the uniform and bipartite models, respectively. Also, we will use $G(n,m)$ to refer to the set of all graphs with n vertices and m edges and likewise call the set of all bipartite graphs with $2n$ vertices and m edges $B(n,m)$.

We will also consider the analogous distributions on hypergraphs:

1) Define $H(n,m,k)$ to be the uniform distribution over k-uniform hypergraphs with n vertices and m possibly duplicated hyperedges. We will considered both the individual hyperedges and the set of hyperedges to be ordered.
2) For $r \leq k$, define $H^{(r)}(n,m,k)$ to be the uniform distribution over r-partite k-uniform hypergraphs with rn vertices and m hyperedges, i.e., n-vertex, m-hyperedge, k-uniform hypergraphs with a partition of the vertices into r sets of size n such that each hyperedge contains at least one vertex from each set. Both the individual hyperedges and the set of hyperedges are ordered.

We will call these the uniform and r-partite models, respectively. Also, we will use $H(n,m,k)$ to refer to the set of all k-uniform hypergraphs with n vertices

and m edges and likewise call the set of all r-partite k-uniform hypergraphs with rn vertices and m edges $H^{(r)}(n, m, k)$.

Finally, given a hypergraph H, let V_H be the vertex set of H and E_H be the set of hyperedges of H.

B. Goldreich's generator

Goldreich [2] suggested constructing a PRGs/OWF $f : \{0,1\}^n \to \{0,1\}^m$ in the following manner given a predicate $P : \{0,1\}^k \to \{0,1\}$:

1) Draw H from $H(n, m, k)$. We will consider H to be chosen in advance and fixed. Let S^j be the tuple corresponding to the jth hyperedge of H.
2) Set the jth output bit of f, $f(x)_j$, to be $P(x_{S_1^j}, x_{S_2^j}, \ldots, x_{S_k^j})$.

We will denote the instance of Goldreich's generator associated with hypergraph H and predicate P as $f_{H,P}$.

Alternatively, we can also construct $f : \{0,1\}^{nk} \to \{0,1\}^m$ by drawing H from $H^{(k)}(n, m, k)$ in Step 1. We will consider both the uniform and k-partite models. In addition, we will consider allowing H to be semirandom, in the sense that we will allow alterations to be made to H in polynomial time.

C. Properties of predicates

Two properties of a predicate, independence and algebraic degree, affect the security of its corresponding PRG.

Definition II.1. A predicate $P : \{0,1\}^k \to \{0,1\}$ is $(t-1)$-wise independent if for all $|S| \subseteq [k]$ such that $|S| \leq t-1$, $\hat{P}(S) = 0$, where $\hat{P}(S) = \mathbf{E}_{x\sim\{0,1\}^k}[P(x)(-1)^{\sum_{i\in S} x_i}]$ is the Fourier coefficient of P on S.

Recall that any $P : \{0,1\}^k \to \{0,1\}$ can be expressed as a unique multilinear polynomial over $\mathbb{F}_2^k$.

Definition II.2. The algebraic degree of a predicate $P : \{0,1\}^k \to \{0,1\}$ is the degree of its representation as a polynomial over $\mathbb{F}_2^k$.

D. The TSA predicate and generalizations

We define the 5-ary predicate $\mathrm{TSA}(x_1, x_2, x_3, x_4, x_5)$ as follows:

$$\mathrm{TSA}(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2 + x_3 + x_4 \cdot x_5 \mod 2.$$

Observe that the TSA predicate is 2-wise independent and has $\mathbb{F}_2$ degree 2.

We will show that Goldreich generators based on TSA are secure against linear tests with stretch up to $\Theta(n^{3/2-\delta})$. In addition, we will consider the following generalization of TSA:

$$\mathrm{XORAND}_{t,u}(x) = x_1 + \cdots + x_t + x_{t+1} \cdots x_{t+u}.$$

This predicate is $(t-1)$-wise independent and has $\mathbb{F}_2$ degree u.

E. Security of PRGs

In general, we call a PRG secure if no algorithm can clearly distinguish its output from a uniform random string:

Definition II.3. A pseudorandom generator $f : \{0,1\}^n \to \{0,1\}^m$ is ϵ-secure if for any efficient algorithm $\mathcal{A}$, the distinguishing advantage

$$\left| \Pr_{x\sim\{0,1\}^n}[\mathcal{A}(f(x)) = 1] - \Pr_{x\sim\{0,1\}^m}[\mathcal{A}(x) = 1] \right| \leq \epsilon.$$

$m - n$ is called the stretch of f.

Showing that a function satisfies this condition for every algorithm is hard, so we restrict our attention to studying the security of Goldreich's PRG against two particular classes of algorithms: $\mathbb{F}_2$-linear tests and SDPs.

F. $\mathbb{F}_2$-Linear Tests

$\mathbb{F}_2$-linear tests have been widely studied as attacks on PRGs [3], [5], [9]. In this case, $\mathcal{A}$ is of the form $L(f) = \sum_{i\in S} f_i$ for some $S \subseteq [m]$, where the sum is taken mod 2. Following [9], define $\mathrm{bias}(f, L)$ as

$$\left| \Pr_{x\sim\{0,1\}^n}[L(f(x)) = 1] - \Pr_{x\sim\{0,1\}^m}[L(x) = 1] \right|$$

and

$$\mathrm{bias}(f) = \max_L \{\mathrm{bias}(f, L)\}$$

where L is any linear test. We will omit the f when the function we are considering is clear from the context. Define $\mathrm{size}(L) = |S|$.

G. Distinguishers for Goldreich's function based on SDPs

As described in the introduction, we can think of a function $f_{H,P}(x)$ as having a corresponding CSP

$$\begin{aligned} P(x_{S_1^1}, \ldots, x_{S_k^1}) &= f_{H,P}(x)_1 \\ P(x_{S_1^2}, \ldots, x_{S_k^2}) &= f_{H,P}(x)_2 \\ &\vdots \\ P(x_{S_1^m}, \ldots, x_{S_k^m}) &= f_{H,P}(x)_m. \end{aligned}$$

One particularly powerful class of algorithms for solving CSPs is semidefinite programs (SDPs). SDPs can also be used as distinguishers for $f_{H,P}$: Given $y \in \{0,1\}^m$, construct the CSP

$$\begin{aligned} P(x_{S_1^1}, \ldots, x_{S_k^1}) &= y_1 \\ P(x_{S_1^2}, \ldots, x_{S_k^2}) &= y_2 \\ &\vdots \\ P(x_{S_1^m}, \ldots, x_{S_k^m}) &= y_m. \end{aligned}$$

If $y = f_{H,P}(x)$ for some x, the CSP has a solution. However, the CSP is highly unsatisfiable if y is chosen uniformly at random. Indeed, it is well-known (e.g., [23]) that with high probability over the choice of H only $\frac{|P^{-1}(1)|}{2^k}m$ constraints can be satisfied. If for y chosen uniformly at random, an SDP returns value m, i.e., believes that all constraints can be satisfied, then this $f_{H,P}$ is perfectly secure against attacks based on computing the value of this SDP. Many known attacks against Goldreich's generator use SDPs (see Section II-I).

1) The Sherali–Adams$^+$ hierarchy: In the Sherali–Adams$^+$ hierarchy, denoted SA$^+$, gives probability distributions on assignments to small sets of variables that are consistent on their intersections. Formally, let μ_S be a distribution over $\{0,1\}^S$ and let $\{\mu_S\}$ be a family of such distributions. For $T \subseteq S$ and an assignment $\alpha \in \{0,1\}^S$, we denote by $\alpha|_T$ the assignment induced by α on T.

Definition II.4. A family of distributions $\{\mu_S\}$ is t-locally consistent if for all $T \subseteq S \subseteq [n]$ such that $|S| \leq t$ and for all $\alpha \in \{0,1\}^T$, the marginal of μ_S on T is identical to μ_T, i.e.,

$$\sum_{\substack{\beta \in \{0,1\}^S \\ \beta|_T = \alpha}} \mu_S(\beta) = \mu_T(\alpha).$$

The Sherali-Adams LP hierarchy requires solutions to form a family of locally consistent distributions $\{\mu_S\}$. For t rounds of the hierarchy, any solution must be a t-locally consistent distribution. In the SA$^+$ hierarchy, we additionally require that there exist vectors $\{v_{i,b}\}_{i\in[n],b\in\{0,1\}}$ such that $\mathbf{Pr}_{\mu_{ij}}[i = b \wedge j = c] = \langle v_{i,b}, v_{j,c}\rangle$. To show security against SA$^+$, we need to give consistent local distributions and accompanying vectors that are supported on satisfying assignments to the corresponding CSP's constraints. Such a solution will have objective function value equal to m even though only about $\frac{|P^{-1}(1)|}{2^k}m$ constraints are satisfiable.

2) The Lasserre hierarchy: The Lasserre hierarchy is a powerful class of SDPs for solving polynomial optimization problems. We define it as in [22], [24].

Consider a polynomial optimization problem of the form

$$\inf\{p(x) \mid x \in \mathbb{R}^n, q_1(x) \geq 0, \ldots, q_k(x) \geq 0\}, \quad (1)$$

where p and the q_i's are polynomials. Let $\mathbb{R}[X]_d$ be the set of polynomials in X over $\mathbb{R}$ of degree at most d. We call a polynomial s a sum of squares (SOS) if it can be expressed as the sum of squares of some polynomials. The degree-d Lasserre relaxation for (1) is then

$$\inf\{\widetilde{\mathbf{E}}(p) \mid \widetilde{\mathbf{E}}(1) = 1,\ \widetilde{\mathbf{E}}(sq_i) \geq 0,\ \widetilde{\mathbf{E}}(s) \geq 0\ \forall \text{ SOS } s\}$$

where $\widetilde{\mathbf{E}} : \mathbb{R}[X]_d \to \mathbb{R}$ is a linear map. This relaxation can be solved using semidefinite programming in time $n^{O(d)}$.

H. Expansion and boundary expansion

To show security against SDP hierarchies, we will also require the notions of expansion and boundary expansion used in [23] and [25]. For a hypergraph $H = (V, E)$ and a set of hyperedges $S \subseteq E$, we will define $\Gamma(S)$ to be the set of all vertices contained in a hyperedge of S, i.e., $\Gamma(S) = \bigcup_{e\in S} e$. We will define $\partial S = \{v \in V : |\Gamma(v) \cap S| = 1\}$ to be the boundary vertices of S. Now we can define expansion and boundary expansion:

Definition II.5. A hypergraph is (r, e)-*expanding* if for any set of hyperedges S such that $|S| \leq r$, $|\Gamma(S)| \geq e|S|$. A hypergraph is (r, e)-*boundary expanding* if for any set of hyperedges S such that $|S| \leq r$, $|\partial S| \geq e|S|$.

It is well known that high expansion implies high boundary expansion (see e.g., [23], [25]):

Lemma II.6. *Let H be a k-uniform hypergraph. If H is $(r, k-d)$-expanding, then H is also $(r, k-2d)$ boundary expanding.*

In order to prove security against SDP hierarchies, we will need $H - S$ to have high expansion for sets S such that $|S| \leq r$ for some r. This is not true in general, but [26], [23], [25] give an algorithm for finding a superset $\bar{S}$ of S such that $\bar{S}$ is not too much bigger than S and $H - \bar{S}$ has high expansion:

Lemma II.7 ([23])**.** *If H is (r_1, e_1) expanding and S is a set of variables such that $|S| < (e_1 - e_2)r_1$ for some $e_2 \in (0, e_1)$, then there exists a set $\bar{S}$ such that $H - \bar{S}$ is (r_2, e_2) expanding, where $r_2 \geq r_1 - \frac{|S|}{e_1 - e_2}$ and $\bar{S} \leq \frac{k+2e_1-e_2}{e_1-e_2}|S|$.*

We will call $\bar{S}$ the closure of S.

I. Known limitations of the Goldreich generator

Herein we review the two known limitations of the Goldreich generator. (An overview for this material appears in Applebaum's survey from TCC 2013 [27].)

The first limitation is a simple one appearing in the work of Mossel, Shpilka, and Trevisan:

Proposition II.8. *([5].) If the predicate $P(x)$ has degree d as an $\mathbb{F}_2$-polynomial then Goldreich's generator is not secure unless $m \leq O(n^d)$.*

This is simply because there will be an $\mathbb{F}_2$-linear relation among the output bits (by a dimension argument) and thus the generator will be susceptible to an $\mathbb{F}_2$-linear attack.

The second limitation is somewhat more sophisticated:

Theorem II.9. *Let $t \geq 2$ and suppose the predicate P is* not *t-wise independent — i.e., P has nonzero correlation with some parity of at most t coordinates. Then if $f_{G,P} : \{0,1\}^n \to \{0,1\}^m$ is a random local function constructed from P with $m \geq Cn^{t/2} \log n$ for sufficiently large C, then with high probability $f_{G,P}$ can be efficiently inverted; i.e., $f_{G,P}$ is not even a OWF.*

This theorem is apparently recent "folklore", known to some experts [28]; however it does not appear to be universally known and has never appeared in print. Therefore we give a sketch of the proof below. Before doing so, we review some variants and consequences of this theorem.

The idea behind the theorem dates back to [5]; they showed the theorem with the weaker bound of $m \leq O(n^t)$, using an $\mathbb{F}_2$-linear attack. They further pointed out that if P has sufficiently large correlation with a size-2 parity then there is a "correlation attack" based on semidefinite programming which limits the stretch to $m = O(n)$. This idea was extended by Bogdanov and Qiao [29] who showed that if P has any nontrivial correlation with a size-2 parity then Goldreich's generator is not even a OWF unless $m \leq O(n)$. Mossel, Shpilka, and Trevisan also combined their weaker version of Theorem II.9 with Proposition II.8 and *Siegenthaler's Theorem* to deduce that the maximum secure stretch of *any* k-ary predicate P is at most $O(n^{\lceil k/2 \rceil})$. Siegenthaler's Theorem is the following:

Theorem II.10. *([30].) Suppose $P : \{0,1\}^k \to \{0,1\}$ has $\mathbb{F}_2$-degree at least $k - t > 1$. Then P is not t-wise independent.*

Using the stronger Theorem II.9 we can similarly deduce the following:

Theorem II.11. *Let $P : \{0,1\}^k \to \{0,1\}$, $k \geq 3$. Then Goldreich's generator is* not *a secure PRG once $m = \widetilde{\Theta}(n^{\frac{1}{2}\lfloor \frac{2}{3}k \rfloor})$. In particular, for $k = 5$ the upper bound is $\widetilde{\Theta}(n^{3/2})$.*

A classical example of a function showing Siegenthaler's Theorem is sharp is $P = \mathrm{XORAND}_{t,k-t}$. It's plausible that with $t = \lfloor \frac{2}{3}k \rfloor$ this function may reach the limit given in Theorem II.11; see Section VI for more discussion.

We conclude this section by sketching the proof of Theorem II.9.

Proof: (Sketch.) By assumption, $\widehat{P}(T) \neq 0$ for some $T \subseteq [k]$ with $|T| \leq t$; without loss of generality we may assume $|T| = t \geq 2$. Let $\epsilon = \widehat{P}(T) \neq 0$; we may assume that $\epsilon > 0$ without loss of generality (by negating P). In fact, since P is a function of k coordinates we must have $\epsilon > 2^{-k}$. By definition of $\widehat{P}(T) = \epsilon$ we have that for a randomly chosen $x \in \{0,1\}^n$,

$$\mathbf{Pr}[P(x) = \mathrm{XOR}_T(x)] = \tfrac{1}{2} + \tfrac{1}{2}\epsilon; \qquad (2)$$

here we are using the notation $\mathrm{XOR}_T(x) = \sum_{i \in T} x_i \pmod 2$.

Let $f(x)_j$ be the jth output bit of $f_{G,P}$ and let T_j be the ordered subset of S_j corresponding to T. We may now state the algorithm for inverting $f_{G,P}$; it is reminiscent both of the Bogdanov–Qiao algorithm and the Feige–Ofek noisy 3-LIN algorithm [31]:

1) Construct the following t-LIN CSP instance: For each output bit $f(x)_j$, include the equation "$\mathrm{XOR}_{T_j}(x_i) = f(x)_j$".
2) Find all pairs (j_1, j_2) such that the t-tuples T_{j_1} and T_{j_2} agree on their first $t-1$ coordinates. Call each such pair a *matched pair*.
3) Construct a 2-LIN CSP instance in the following way: For each matched pair (j_1, j_2), add the two corresponding t-LIN equations to get a 2-LIN equation: "$x_{T_{j_1,t}} + x_{T_{j_2,t}} = f(x)_{j_1} + f(x)_{j_2}$".
4) Solve the resulting 2-LIN instance to obtain an inverse for $f_{G,P}$.

We will elaborate on Step 4 shortly. We first observe that the t-LIN instance constructed in Step 1 can be thought of as a *δ-noisy random planted* t-LIN instance, wherein one first generates a planted, fully satisfiable t-LIN instance and then flips each "right-hand side" independently with probability $\delta = \frac{1}{2} - \frac{1}{2}\epsilon$ (cf. (2)).

Next, a simple estimate shows that the 2-LIN instance constructed in Steps 2 and 3 has many equations:

Claim II.12. *In a random t-LIN instance with $m = \Omega(n^{\frac{t}{2}} \log n)$ clauses, with high probability there are $\Omega(n \log n)$ pairs of constraints that share the same first $t-1$ coordinates.*

Proof: This is a simple extension of the proof of Lemma 3.1 in [31]. ∎

In addition, we can also think of this 2-LIN instance as being a δ'-noisy random planted instance with the right-hand side bits flipped independently. The right-hand side bit is flipped if exactly one of the right-hand side bits of the corresponding two t-LIN equations is flipped. This occurs with probability

$$\delta' = 2\delta(1-\delta) = \frac{1}{2} - \frac{1}{2}\epsilon^2 \leq \frac{1}{2} - \frac{1}{2}2^{-2k}.$$

Claim II.13. *The distribution of the 2-LIN equation left-hand sides produced by the algorithm is precisely independent and uniformly random.*

Proof: Imagine constructing a 2-LIN instance by the following equivalent process: Choose the first $t-1$

variables of every equation uniformly at random. For each matched pair, independently choose the last two variables of this pair. This process is equivalent to the one described above and demonstrates that the 2-LIN equations are chosen independently. By symmetry, these equations are chosen uniformly at random. Therefore, every equation is chosen independently and uniformly at random. ■

So we have a random δ'-noisy planted 2-LIN instance, where the noise δ' is bound away from $\frac{1}{2}$ by at least the constant $\frac{1}{2}2^{-2k}$. Håstad [32], [33] showed that in this setting for $O_k(n \log n)$ equations the planted solution and its complement are the only optimal solutions. This means that if we can solve the 2-LIN instance, we will get an inverse for $f_{G,P}$. Finally, we sketch two well-known algorithms for solving noisy 2-LIN instances:

1) We can apply the methods of Bogdanov and Qiao [29]: First use an efficient "2-LIN-Gain" SDP approximation algorithm (say, the one of [34]) to find a 2-LIN solution that is correlated with the planted solution. Then use their "local search" procedure to recover the planted solution.
2) Alternatively, it seems that a slight modification of the simpler spectral algorithm of Boppana [35] (for planted noisy bisection) can be used to directly find the planted 2-LIN solution.

■

III. Security against $\mathbb{F}_2$-linear attacks

In this section, we will prove the following theorem:

Theorem III.1. *Let $m \leq n^{3/2-\delta}$ for some $\delta > 0$. Then the following two statements hold:*

1) *If $H \sim H(n, m, 5)$, then $\mathrm{bias}(f_{H,\mathrm{TSA}}) \leq 2^{-\Omega(n^\delta)}$ with high probability.*
2) *If $H \sim H^{(5)}(n, m, 5)$, then $\mathrm{bias}(f_{H,\mathrm{TSA}}) \leq 2^{-\Omega(n^\delta)}$ with high probability.*

The proof is essentially the same in both the uniform and 5-partite cases. In the rest of this section, we will prove this theorem in the uniform case and indicate differences in the proof of the 5-partite case as they arise.

A. Outline of the proof

Think of a linear test L as a degree 2 polynomial over $\mathbb{F}_2$, the sum of its constituent TSA functions. We will refer to this polynomial as $L(x)$. Note that each TSA predicate is the mod 2 sum of an XOR part $(x_1+x_2+x_3)$ and an AND part $(x_4 \cdot x_5)$. $L(x)$ will have degree 1 terms corresponding to the sum of the XOR parts of its predicates and degree 2 terms corresponding to the AND parts.

Recall that the bias of a sum of constant-bias independent bits is exponentially small in the number of bits. This implies that if L can be divided up into enough "independent" pieces, it will have small bias. On the other hand, if L has degree-1 terms that are "independent" of the degree-2 terms, it will have bias 0.

There are two ways this can happen: Either the degree 2 terms of L can be broken up into a large number of mostly independent blocks or L has degree 1 terms that are independent of the degree 2 terms (or both).

This means that in order for L to have large bias *both* of the following conditions must hold:

1) The AND parts cannot be broken up into a large number of mostly independent blocks.
2) The XOR parts must be highly dependent on each other and the AND parts.

We show that it is very unlikely for both of these conditions to hold simultaneously, so L must have small bias with high probability. Previous analyses (e.g., [5], [9]) considered these conditions individually, showing that the AND parts are likely to consist of a large number of mostly independent blocks for large linear tests and that for small linear tests there are likely to be nonzero XOR terms that are independent of the AND terms. However, for higher values of m, it is likely that there will be medium size tests failing to meet either condition. We address this issue by showing that the probability that a linear test fails to meet both conditions simultaneously is low even though either condition individually might not hold.

The proof will have four sections. First, we will formalize conditions 1 and 2. Next, we will bound the probability that condition 1 is met, and then we will bound the probability that condition 2 is met. Using these bounds, we will show that the probability that condition 1 and condition 2 both occur is very small.

B. When do linear tests have large bias?

The starting point for our analysis is the work of [9], who proved the following theorem (a combination of Corollaries 3.3 and 3.7):

Theorem III.2. *Let $m = n^{3/2-\delta}$ for some $\delta > 0$. Then for all linear tests L such that $\mathrm{size}(L) \leq n^{2\delta}$ or $\mathrm{size}(L) \geq \frac{n}{4}$, the following statements hold:*

1) *If $H \sim H(n, m, 5)$, then $\mathrm{bias}(f_{H,\mathrm{TSA}}) \leq 2^{-\Omega(n^\delta)}$ with high probability.*
2) *If $H \sim H^{(5)}(n, m, 5)$, then $\mathrm{bias}(f_{H,\mathrm{TSA}}) \leq 2^{-\Omega(n^\delta)}$ with high probability.*

The proof follows the intuition described above. Actually, [9] only prove this theorem in the uniform case but their proof also works in the 5-partite case.

We therefore only need to show that $\mathrm{bias}(f_{H,\mathrm{TSA}}, L) \leq 2^{-\Omega(n^\delta)}$ for $\mathrm{size}(L) \in \left[n^{2\delta}, \frac{n}{4}\right]$. To do this, we will use the structure of the polynomial corresponding to the linear test. Specifically, we will need the following theorem. See, e.g, [36] for more details.

Theorem III.3 (Dickson's Theorem)**.** *Any polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most 2 can be expressed as*

$$p(x) = \ell_0(x) + \sum_{i=1}^{h} \ell_i(x)\ell_i'(x),$$

where ℓ_0 is an affine function and $\ell_1, \ell_1', \ldots, \ell_h, \ell_h'$ are linearly independent linear functions. h is called the rank of p. Then the following two statements hold:

1) *The bias of p is at most γ^h for some constant $\gamma \in [0, 1)$.*
2) *If ℓ_0 is not a constant and is linearly independent of the ℓ_i's and ℓ_i''s, then p is unbiased.*

Given a k-uniform hypergraph H, we can then write a linear test $L(f_{H,\mathrm{TSA}})$ in this representation: $L(x) = \ell_0(x) + \sum_{i=1}^{h} \ell_i(x)\ell_i'(x)$. We define $\mathrm{rk}_H(L)$ to be the rank of $L(f_{H,\mathrm{TSA}})$. We'll drop H and/or L when it's clear from the context. Note that in our case the rank only depends on the AND parts. Dickson's Theorem implies that if $\mathrm{bias}(f_{H,\mathrm{TSA}}, L) > 2^{-\Omega(n^\delta)}$, then both of the following two conditions hold:

1) $\mathrm{rk}_H(L) < n^\delta$.
2) ℓ_0 is linearly dependent on the ℓ_i's and ℓ_i''s.

Call L bad if $\mathrm{bias}(f_{H,\mathrm{TSA}}, L) > 2^{-\Omega(n^\delta)}$. Call k bad if there exists an L of size k such that $\mathrm{bias}(f_{H,\mathrm{TSA}}, L) > 2^{-\Omega(n^\delta)}$. Now define $\beta_k = \mathbf{Pr}_H[k \text{ bad}]$. Call H L-linearly dependent, or L-LD, if ℓ_0 is linearly dependent on the ℓ_i's and ℓ_i''s. Using the above conditions and the union bound, we can then write

$$\beta_k \leq \binom{m}{k} \mathop{\mathbf{Pr}}_H \left[\mathrm{rk}(L) < n^\delta\right] \cdot \mathop{\mathbf{Pr}}_H \left[H \ L\text{-LD} \mid \mathrm{rk}(L) < n^\delta\right]$$

for some $L = \ell_0 + \sum_{i=1}^{h} \ell_i\ell_i'$. Note that in the 5-partite case, we don't need to condition on $\mathrm{rk}(L) < n^\delta$ in the last term because the variables in the XOR and AND parts of the predicate are independent. We will show that $\beta_k \leq \frac{1}{n^2}$ for all k in $\left[n^{2\delta}, \frac{n}{4}\right]$. In the next two sections, we will upper bound $\mathbf{Pr}_H\left[\mathrm{rk}_H(L) < n^\delta\right]$ and $\mathbf{Pr}_H\left[H \ L\text{-LD} \mid \mathrm{rk}_H(L) < n^\delta\right]$.

C. Low rank is unlikely

Let H_{AND} be the graph on the variables of the instance constructed by putting an edge between two variables if and only if they both appear as AND variables in the same hyperedge. Let $H_{\mathrm{AND}}(L)$ be the subgraph of H_{AND} corresponding to the hyperedges of L. Note that each $\ell_i\ell_i'$ product is a complete bipartite subgraph of $H_{\mathrm{AND}}(L)$. If $\mathrm{rk}(L) < n^\delta$, there is a covering of $H_{\mathrm{AND}}(L)$ with at most n^δ complete bipartite graphs. We will show that this event is unlikely to happen. Ideally, we would simply show that $\mathbf{Pr}_H\left[\mathrm{rk}_H(L) < n^\delta\right]$ is small. However, we are first going to have to exclude some "bad" cases that would complicate the analysis.

Consider a complete bipartite graph $K_{s,t}$ such that $s \leq t$. First, we are going to show that it is very unlikely for H_{AND} to contain any $K_{s,t}$ with $s, t \geq 5$ as a subgraph. It is therefore also unlikely for any $H_{\mathrm{AND}}(L)$ to contain any $K_{s,t}$ with $s, t \geq 5$ as a subgraph. We will then restrict our attention to coverings of $H_{\mathrm{AND}}(L)$ with K_{s_i,t_i}'s such that $s_i \leq 4$ and bound the probability that there is a covering of this form of size at most n^δ. Note that this is the same as saying that we can write $L(x) = \ell_0(x) + \sum_{i=1}^{h} \ell_i(x)\ell_i'(x)$ such that all ℓ_i's (but not necessarily ℓ_i''s) have support at most 4 and $h < n^\delta$.

Write $G_1 \subseteq G_2$ if G_1 contains G_2 as a subgraph. We begin by showing that it is unlikely for H_{AND} to contain $K_{5,5}$, and therefore any $K_{s,t}$ with $s, t \geq 5$, as a subgraph.

Lemma III.4. *For $H \sim H(n, m, 5)$ and $H \sim H^{(5)}(n, m, 5)$,*

$$\mathop{\mathbf{Pr}}_H[K_{5,5} \subseteq H_{\mathrm{AND}}] \leq O\left(n^{-5/2}\right).$$

Proof: We will prove this lemma in the case of $H \sim H(n, m, k)$ and discuss how this differs from the $H \sim H^{(k)}(n, m, k)$ case at the end.

In the uniform case, H_{AND} is distributed like $G(n, m)$. Since $K_{5,5}$ contains 25 edges and H_{AND} contains m total edges, we can write

$$\mathop{\mathbf{Pr}}_H[K_{5,5} \subseteq H_{\mathrm{AND}}] \leq \binom{m}{25} \mathbf{Pr}[\text{edges of } T \text{ form } K_{5,5}]$$

for a particular set T containing exactly 25 edges. Each edge of T has $\binom{n(n-1)}{2}$ choices of endpoints, so there are $\binom{n(n-1)}{2}^{25}$ possible choices of endpoints for all edges of T. $K_{5,5}$ has 10 vertices. There are at most n^{10} choices for this set of 10 vertices and 25! orderings of its edges so the number of choices of endpoints resulting in a $K_{5,5}$ is at most $25! \cdot n^{10}$. Since H_{AND} is distributed uniformly over graphs with n vertices and m edges,

$$\begin{aligned}\mathbf{Pr}[\text{edges of } T \text{ form } K_{5,5}] &\leq \frac{25! \cdot n^{10}}{\binom{n(n-1)}{2}^{25}} \\ &\leq \frac{e \cdot 2^{25} \cdot 25!}{n^{40}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr_H[K_{5,5} \subseteq H_{\text{AND}}] &\leq \binom{m}{25} \frac{e \cdot 2^{25} \cdot 25!}{n^{40}} \\ &= O\left(n^{-5/2}\right),\end{aligned}$$

proving the lemma in the uniform case.

In the 5-partite case, H_{AND} is distributed as $B(n, m)$. Each edge can have n^2 possible endpoints. A set T of 25 edges therefore has n^{50} possible endpoints. There are still at most n^{10} possible sets of 5 left vertices and 5 right vertices and 25! possible orderings of the edges on these vertices, so there are at most $25! \cdot n^{10}$ possible choices of edges that result in a $K_{5,5}$. The rest of the proof then goes through as above. ∎

Next, we show that it is unlikely for $H_{\text{AND}}(L)$ to have a small covering of K_{s_i,t_i}'s with all $s_i \leq 4$ for all i.

Lemma III.5. *For any size k linear test L,*

1) *If $H \sim H(n, m, 5)$,*

$$\Pr_H\left[\text{rk} < n^\delta \mid \text{no } K_{5,5}\right] \leq e^2 8^k n^{4n^\delta - k(1-\delta)}.$$

2) *If $H \sim H^{(5)}(n, m, 5)$,*

$$\Pr_H\left[\text{rk} < n^\delta \mid \text{no } K_{5,5}\right] \leq e 2^{4n^\delta} 4^k n^{4n^\delta - k(1-\delta)}.$$

Proof: We begin with the uniform case. In order for $H_{\text{AND}}(L)$ to a have a small covering of K_{s_i,t_i}'s such that all $s_i \leq 4$, there must be a small set of vertices that has many edges adjacent to it. We will bound the probability that such a set exists.

Let C_i be the smaller of the two sets of vertices of K_{s_i,t_i}. We need only consider coverings of H_{AND} with at most n^δ K_{s_i,t_i}'s such that $|C_i| \leq 4$. Let $C = \bigcup_i C_i$. Every edge of $H_{\text{AND}}(L)$ must be adjacent to some vertex of C so $|C| \leq 4n^\delta$. We can then upper bound $\mathbf{Pr}_H\left[\text{rk}(L) < n^\delta \mid \text{no } K_{5,5}\right]$ by

$$\sum_{\substack{C \subseteq V \\ |C| \leq 4n^\delta}} \mathbf{Pr}[\text{all } H_{\text{AND}}(L) \text{ edges adjacent to } C].$$

The number of edges adjacent to some vertex in C is at most $4n^{1+\delta}$. The total number of possible edges is $\binom{n}{2}$ and each edge is chosen uniformly at random since H_{AND} is distributed like $G(n, m)$, so

$$\begin{aligned}\mathbf{Pr}[\text{all } H_{\text{AND}}(L) \text{ edges adjacent to } C] &\leq \left(\frac{4n^{1+\delta}}{\binom{n}{2}}\right)^k \\ &\leq e 8^k n^{-k(1-\delta)}.\end{aligned}$$

Summing over all possibles sets C, we have that

$$\begin{aligned}\Pr_H\left[\text{rk}(L) < n^\delta \mid \text{no } K_{5,5}\right] &\leq \sum_{\substack{C \subseteq V \\ |C| \leq 4n^\delta}} e 8^k n^{-k(1-\delta)} \\ &\leq e^2 8^k n^{4n^\delta - k(1-\delta)}.\end{aligned}$$

In the 5-partite case, H_{AND} has $2n$ vertices and the number of ways of choosing a subset of size at most $4n^\delta$ is then at most $(2n+1)^{4n^\delta}$. The probability that every edge touches at least one vertex of C is at most $\left(\frac{4n^{1+\delta}}{n^2}\right)^k = 4^k n^{-k(1-\delta)}$ since H_{AND} is distributed as $B(n, m)$. Plugging these values into the above proof gives the desired result. ∎

D. Linear dependence is unlikely

In this section, we will bound the probability that ℓ_0 is linearly dependent on the ℓ_i's and ℓ_i''s. Call H L-dense if $\text{rk}_H(L) < n^\delta$ and $K_{5,5} \not\subseteq H_{\text{AND}}$. Specifically, we will prove the following lemma:

Lemma III.6. *For any size k linear test L,*

1) *If $H \sim H(n, m, 5)$,*

$$\Pr_H[\text{LD} \mid \text{dense}] \leq \frac{3}{2} e^3 2^{2n^\delta} k (9e^2 k)^{3k/2} n^{-3k/2}.$$

2) *If $H \sim H^{(5)}(n, m, 5)$,*

$$\Pr_H[\text{LD} \mid \text{dense}] \leq \frac{3}{2} \cdot 2^{2n^\delta} k (9e^2 k)^{3k/2} n^{-3k/2}.$$

Proof: Let $L = \ell_0 + \sum_{i=1}^h \ell_i \ell_i'$ and $U = \bigcup_{i=1}^h \{\ell_i, \ell_i'\}$. Then by the union bound,

$$\Pr_H[\text{LD} \mid \text{dense}] \leq \sum_{T \subseteq U} \Pr_H\left[\ell_0 = \sum_{\ell \in T} \ell \;\middle|\; \text{dense}\right].$$

Since there are at most $2^{2h} \leq 2^{2n^\delta}$ subsets of U, it suffices to show that for any $u \in \{0,1\}^n$, the following hold:

1) If $H \sim H(n, m, 5)$,

$$\Pr_H[\ell_0 = u \mid \text{dense}] \leq \frac{3}{2} e^3 k (9e^2 k)^{3k/2} n^{-3k/2}.$$

2) If $H \sim H^{(5)}(n, m, 5)$,

$$\Pr_H[\ell_0 = u \mid \text{dense}] \leq \frac{3}{2} k (9e^2 k)^{3k/2} n^{-3k/2}.$$

To prove these statements, we need to determine how conditioning on the rank affects the XOR parts. We will start with the uniform case. Call $G \in G(n, m)$ L-dense if any 5-uniform hypergraph H such that $H_{\text{AND}} = G$ is L-dense. Then define

$$A_{L,h} = \{G \in G(n, m) \mid G \text{ } L\text{-dense}\},$$

We can then write Then $\mathbf{Pr}_H[\ell_0 = u \mid \text{dense}]$ is equal to

$$\begin{aligned}\Pr_H[\ell_0 = u \mid \text{dense}] = \sum_{G \in A_{L,h}} \Big(&\Pr_H[H_{\text{AND}} = G \mid \text{dense}] \\ &\cdot \Pr_H[\ell_0 = u \mid H_{\text{AND}} = G, \text{dense}]\Big).\end{aligned}$$

Now observe that the rank of L as well as whether or not H_{AND} contains $K_{5,5}$ as a subgraph is completely determined by H_{AND}, so this is equal to

$$\sum_{G \in A_{L,h}} \Big(\mathop{\mathbf{Pr}}_{H} [H_{\text{AND}} = G \mid \text{dense}] \cdot \mathop{\mathbf{Pr}}_{H} [\ell_0 = u \mid H_{\text{AND}} = G] \Big) \leq \max_{G \in A_{L,h}} \mathop{\mathbf{Pr}}_{H} [\ell_0 = u \mid H_{\text{AND}} = G] .$$

For the 5-partite case, we can replace $A_{L,h}$ with

$$A'_{L,h} = \{G \in B(n,m) \mid G \text{ } L\text{-dense}\}.$$

The same argument shows that $\mathbf{Pr}_H [\ell_0 = u \mid \text{dense}]$ is at most

$$\max_{G \in A'_{L,h}} \mathop{\mathbf{Pr}}_{H} [\ell_0 = u \mid H_{\text{AND}} = G]$$

for $H \sim H^{(5)}(n,m,5)$. As a result, it suffices to prove this claim:

Claim III.7. 1) *If $H \sim H(n,m,5)$ and $G \in G(m,n)$, then $\mathbf{Pr}_H [\ell_0 = u \mid H_{\text{AND}} = G]$ is upper bounded by*

$$\frac{3}{2} e^3 k (9e^2 k)^{3k/2} n^{-3k/2}.$$

2) *If $H \sim H^{(5)}(n,m,5)$ and $G \in B(m,n)$, then $\mathbf{Pr}_H [\ell_0 = u \mid H_{\text{AND}} = G]$ is upper bounded by*

$$\frac{3}{2} k (9e^2 k)^{3k/2} n^{-3k/2}.$$

We begin with the uniform case. It will be convenient for us to think of the process of drawing $H \sim H(n,m,k)$ in an equivalent sequential way: For i from 1 to m, pick an ordered subset of size k from $[n]$ uniformly at random to be the ith hyperedge of H. For each ordered hyperedge, we pick the first variable uniformly at random from $[n]$. We then pick the second variable uniformly at random from the remaining $n-1$ variables and continue in this manner for the remaining three variables.

Recall that ℓ_0 is the sum of the XOR parts of all hyperedges of L. Think of the process of constructing the XOR parts of L as filling in $3k$ blanks with variables. Let s be the size of the support of u. In order for $\ell_0 = u$, we need s of these $3k$ blanks to be the variables in the support of u. Define $\mathcal{E}_T$ to be the event that a specific set T of s blanks contains the variables of u. The remaining $3k - s$ blanks must be filled with pairs of variables that add to 0. Define the event $\mathcal{F}_T$ to be the event that all blanks not in T are filled with matching pairs of variables. We can then write

$$\mathop{\mathbf{Pr}}_{H} [\ell_0 = u \mid H_{\text{AND}} = G] \leq \sum_{\substack{T \subseteq [3k] \\ |T| = s}} \mathop{\mathbf{Pr}}_{H} [\mathcal{E}_T] \mathop{\mathbf{Pr}}_{H} [\mathcal{F}_T \mid \mathcal{E}_T].$$

First, we will bound $\mathbf{Pr}_H[\mathcal{E}_T]$. There are $s!$ possible orderings in which we could fill in the blanks of T with the elements of u. For each of these orderings, we need to bound the probability that all s blanks are assigned correctly. The probability that a single blank is filled in with a particular variable x is at most $\frac{1}{n-4}$: We have already set the two AND variables for this hyperedge and have set at most two XOR variables, so the blank is filled in by choosing a variable uniformly at random from at least $n-4$ unused choices. The probability that all s blanks are filled in correctly is then at most $\left(\frac{1}{n-4}\right)^s$. We have therefore shown that

$$\mathop{\mathbf{Pr}}_{H} [\mathcal{E}_T] \leq s! \left(\frac{1}{n-4}\right)^s .$$

To bound $\mathbf{Pr}_H[\mathcal{F}_T \mid \mathcal{E}_T]$, observe that all $3k-s$ blanks not in T must contain i variables for some $i \leq \frac{3k-s}{2}$. When we fill in each blank not in T, we then have i possible choices out of at least $n-4$ total choices. This implies that

$$\mathop{\mathbf{Pr}}_{H} [\mathcal{F}_T \mid \mathcal{E}_T] \leq \sum_{i=1}^{\frac{3k-s}{2}} \binom{n}{i} \left(\frac{i}{n-4}\right)^{3k-s} \leq \frac{3k}{2} \binom{n}{\frac{3k-s}{2}} \left(\frac{\frac{3k-s}{2}}{n-4}\right)^{3k-s} .$$

We have shown so far that $\mathbf{Pr}_H [\ell_0 = u \mid H_{\text{AND}}]$ is at most

$$\binom{3k}{s} s! \left(\frac{1}{n-4}\right)^s \cdot \frac{3k}{2} \binom{n}{\frac{3k-s}{2}} \left(\frac{\frac{3k-s}{2}}{n-4}\right)^{3k-s} .$$

Applying the bounds $\binom{n}{i} \leq \left(\frac{ne}{i}\right)^i$ and $n! \leq n^n$ and then combining like terms completes the proof in the uniform case. In the 5-partite case, observe that each variable of each hyperedge is selected independently. This means that the probability that we fill in a single blank in the XOR part with a particular variable is $\frac{1}{n}$.The rest of the proof follows as described above. ∎

E. Putting the pieces together

Now we will combine the results of the previous two sections to complete the proof of Theorem III.1. First, we will show that β_k, the probability that any linear test of size k fails, is at most $\frac{1}{n^2}$. Recall that we simplified our analysis of the low rank case by ruling out instances in which $K_{5,5}$ is a subgraph of H_{AND}. We can therefore write

$$\beta_k \leq \mathop{\mathbf{Pr}}_{H} [\text{no } K_{5,5}] \cdot \mathop{\mathbf{Pr}}_{H} [k \text{ bad} \mid \text{no } K_{5,5}] + \mathop{\mathbf{Pr}}_{H} [K_{5,5}] \leq \mathop{\mathbf{Pr}}_{H} [k \text{ bad} \mid \text{no } K_{5,5}] + \mathop{\mathbf{Pr}}_{H} [K_{5,5}]$$

where $\mathbf{Pr}_H[k \text{ bad} \mid \text{no } K_{5,5}]$ is at most

$$\binom{m}{k} \mathop{\mathbf{Pr}}_H \left[\mathrm{rk}(L) < n^\delta \mid \text{no } K_{5,5}\right] \cdot \mathop{\mathbf{Pr}}_H \left[L\text{-LD} \mid L\text{-dense}\right].$$

and L is some linear test of size k. In the uniform case, we can plug in the results of Lemmas III.4, III.5, and III.6 and then simplify using $\binom{m}{k} \leq \left(\frac{me}{k}\right)^k$ to get

$$\beta_k \leq \frac{3}{2} e^5 k (216e^4)^k 2^{2n^\gamma} n^{4n^\gamma - (1/2+\delta-\gamma)k} + O\left(n^{-5/2}\right).$$

Since $k = \Omega(n^{2\delta})$, we have that $\beta_k \leq \frac{1}{n^2}$ for large enough n. Plugging in the bounds for the 5-partite case and simplifying gives the same result.

Taking a union bound over all of the at most n possible values of $k \in \left[n^{2\delta}, \frac{n}{4}\right]$ gives us that

$$\mathop{\mathbf{Pr}}_H \left[\text{any } k \in \left[n^{2\delta}, \frac{n}{4}\right] \text{ bad}\right] \leq \frac{1}{n}.$$

Combining this result with Theorem III.2 using a union bound completes the proof of Theorem III.1.

IV. Security against SA$^+$ attacks

In this section, we will prove Theorem I.2:

Theorem I.2 restated.: For $t \geq 3$, fix a $(t-1)$-independent predicate P and any $\delta > 0$. Suppose we instantiate Goldreich's PRG with $m = O(n^{t/2-\delta})$ and with certain "bad tuples" removed. (With high probability there are only $o(m)$ such bad tuples, and they may be initially recognized and removed in polynomial time, once and for all.) Then the PRG is perfectly secure against the attack based on computing the SA$^+$ relaxation value, even for $n^{\Omega(\delta)}$ rounds.

A. Outline of the proof

Recall that in order to show that Goldreich's PRG is secure against attacks based on computing the SA$^+$ value, we need to prove that the SA$^+$ relaxation of the CSP $f_{H,P}(x) = y$ has a solution with value m for a random string in $y \in \{0,1\}^m$ with high probability over H. To do this, we need to give consistent local distributions on satisfying assignments and vectors whose dot products match the probabilities of pairs of assignments from these distributions. Note that in this section we will only consider H drawn from $H(n,m,k)$.

The first two parts of the proof very closely follow the analysis of [23], whose results we generalize for higher values of m and general t. We start by showing that random k-uniform hypergraphs have high expansion. We then show that high expansion suffices to guarantee the existance of locally-consistent distributions supported on satisfying assignments. In [23], the existance of SA$^+$ vectors relies on assignments to pairs of variables being uniformly distributed. This follows from the fact that the hypergraph corresponding to the instance still has high expansion, even when any two vertices are deleted. This property no longer holds for higher values of m; it is likely that there are small sets of hyperedges that do not have sufficiently high expansion. However, we can instead show that there are only $o(m)$ of these sets. We can then simply remove these hyperedges to get a hypergraph that does have the expansion property we want, allowing us to construct SA$^+$ vectors.

B. Random instances are well-behaved

First, we restate the well-known fact that random k-uniform hypergraphs have high expansion (e.g., [23], [25]).

Lemma IV.1. *Let $\delta > 0$. Consider $H \sim H(n,m,k)$ with $m = \Omega\left(n^{t/2-\delta}\right)$. With high probability, H is $\left(n^{\Omega_t(\delta)}, k - \frac{t}{2} + \frac{\delta}{2}\right)$-expanding.*

Proof: The proof is essentially the same as the proof of Lemma 4.1 in [23], except we consider higher values of m and general values of t. We also consider the edges of H to be ordered, as the predicates we consider may not be symmetric. We want to upper bound the probability that any set of $r \leq n^{\Omega_t(\delta)}$ edges contains less than $\left(k - \frac{t}{2} + \frac{\delta}{2}\right) r$ vertices.

We first consider the probability that any set of r tuples contains at most v vertices. Call such a set (r,v)-bad and set $\mathbf{Pr}_H[\text{any } (r,v)\text{-bad } S] = p_{r,v}$ By a union bound, we have

$$p_{r,v} \leq \binom{m}{r} \mathop{\mathbf{Pr}}_H \left[\Gamma(T) \leq v\right]$$

where T is a specific tuple of r edges. $\mathbf{Pr}_H[\Gamma(T) \leq v]$ is upper bounded by

$$\frac{(\#\text{ sets } S \text{ of } v \text{ vertices}) \cdot (\#\text{ sets of } r \text{ edges in } S)}{(\#\text{ ways of choosing } r \text{ edges})}.$$

We can then write

$$p_{r,v} \leq r! \binom{m}{r} \cdot \frac{\binom{n}{v}\binom{k!\binom{v}{k}}{r}}{\left(k!\binom{n}{k}\right)^r}.$$

By applying the inequalities $\left(\frac{n}{i}\right)^i \leq \binom{n}{i} \leq \left(\frac{ne}{i}\right)^i$ and $j! \leq j^j$ and then combining terms, we obtain

$$p_{r,v} \leq e^{(2+k)r+v} v^{kr-v} r^{-r} n^{v-kr} m^r.$$

We want to set $v = \left\lfloor \left(k - \frac{t}{2} + \frac{\delta}{2}\right) r \right\rfloor$. Note that all terms either don't contain v or are increasing in v except for v^{-v}. However, because we have an n^v term and $n \geq v$, we can conclude that this entire expression is increasing in v. It is therefore upper bounded by setting $v = \left(k - \frac{t}{2} + \frac{\delta}{2}\right) r$. Simplifying, we see that $p_{r, \left\lfloor \left(k - \frac{t}{2} + \frac{\delta}{2}\right) r \right\rfloor}$ is at most

$$\left(C(k,t) m n^{-(t/2-\delta/2)} r^{t/2-1-\delta/2}\right)^r.$$

where $C(k,t)$ is some constant depending on k and t.

For $m = n^{t/2-\delta}$, we can then upper bound the probability that a set of r edges has at most $\left(k - \frac{t}{2} + \frac{\delta}{2}\right) r$ variables for *any* $r \leq n^{\delta/(t-2)}$by

$$\sum_{r=1}^{\lfloor n^{\delta/(t-2)} \rfloor} p_{r,\lfloor (k-\frac{t}{2}+\frac{\delta}{2})r \rfloor}.$$

Plugging in the above bound gives that this is at most $n^{-\Omega(\delta)}$. Therefore, H has $(n^{\delta/(t-2)}, k - \frac{t}{2} + \frac{\delta}{2})$ expansion with high probability. ■

Note that this lemma with Lemma II.6 implies that $H \sim H(n, m, k)$ has $(n^{\delta/(t-2)}, k - t + \delta)$ boundary expansion.

C. Obtaining consistent distributions

In this section, we will use the high expansion of a random instance to construct a family of locally consistent distributions. We will then use these locally consistent distributions to show that the SA$^+$ relaxation has value m.

For a set of variables S, let $\mathcal{C}(S)$ be the set of hyperedges (constraints for the corresponding CSP) of H completely contained in S. As in [23] and [25], we define the distribution μ_S over assignments $\{0,1\}^S$ to be the uniform distribution over assignments satisfying of constraints in $\mathcal{C}(S)$. This means that $\mu_S(\alpha) > 0$ only if α satisfies all constraints in $\mathcal{C}(S)$. [23] and [25] actually define distributions over assignments to sets based on the existence for each constraint of a pairwise independent distribution over satisfying assignments to that constraint. Their distribution is equivalent to the uniform distribution over satisfying assignments in the case that for every constraint the uniform distribution over satisfying assigments to that constraint is pairwise independent. Our results can be extended in a similar manner to the case in which we have some other non-uniform t-wise independent distribution over satisfying assignments.

To show that the μ_S distributions are consistent, we require the following generalization of [23] Claim 3.3 to $(t-1)$-wise independent predicates.

Lemma IV.2. *Let $S_1 \subseteq S_2$ be two sets of variables such that both H and $H - S_1$ are $(r, k - t + \epsilon)$-boundary expanding for some $\epsilon > 0$ and $|\mathcal{C}(S_2)| \leq r$. Then there exists an ordering $C_{i_1}, \ldots, C_{i_\ell}$ of the constraints in $\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)$ and a partition of $S_2 \setminus S_1$ into sets of variables $F_1, \ldots, F_\ell, F_{\ell+1}$ such that for all $j \leq \ell$, $F_j \subseteq C_{i_j}$, $|F_j| \geq k - t + 1$, and $F_j \cap (\cup_{a>j} C_{i_a}) = \emptyset$.*

This follows by the same argument as in [23]. In [23], the authors assume $(r, k - 3 + \epsilon)$ boundary expansion and use the pigeonhole principle to select a sequence of constraints each contributing at least $k - 2$ variables to the boundary of $\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)$. Instead, we assume $(r, k-t+\epsilon)$ boundary expansion and use the pigeonhole principle in exactly the same manner to select a sequence of constraints each contributing at least $k-t+1$ variables to the boundary of $\mathcal{C}(S_2) \setminus \mathcal{C}(S_1)$.

Using this lemma, it is easy to show a generalization of [23] Lemma 3.2:

Lemma IV.3. *Let $S_1 \subseteq S_2$ be two sets of variables such that both G and $G - S_1$ are $(r, k-t+\epsilon)$-boundary expanding for some $\epsilon > 0$ and $|\mathcal{C}(S_2)| \leq r$. Then for any $\alpha_1 \in \{0,1\}^{S_1}$, the marginal of μ_{S_2} on S_1 is equal to μ_{S_1}, i.e.,*

$$\sum_{\substack{\alpha_2 \in \{0,1\}^{S_2} \\ \alpha_2(S_1) = \alpha_1}} \mu_{S_2}(\alpha_2) = \mu_{S_1}(\alpha_1).$$

This can be proved in exactly the same manner as [23] Lemma 3.2.

For any set S of size at most $O_{k,\delta}(n^{\delta/(t-2)})$, we can calculate a closure set $\bar{S}$ for which $G - \bar{S}$ is $\left(O_{k,\delta}(n^{\delta/(t-2)}), k - t + \frac{\delta}{2}\right)$-boundary expanding and $|\bar{S}| = O_{k,\delta}(|S|)$ by Theorem 3.1 of [23]. We then consider the family of distributions μ' defined so that μ'_S is the uniform distribution over satisfying assignments to the constraints in $\mathcal{C}(\bar{S})$. Using Lemma IV.3, it is again easy to show that the family of distributions $\{\mu'_S\}$ is s-locally consistent for $s = \Omega_{k,\delta}(n^{\delta/(t-2)})$ exactly as in [23]. This gives us an s-round Sherali-Adams solution.

D. Constructing SA$^+$ vectors

To obtain an SA$^+$ solution, it remains to show that we can construct vectors $\{v_{i,b}\}_{i \in [n], b \in \{0,1\}}$ such that $\langle v_{i,b}, v_{j,c} \rangle = \mathbf{Pr}_{\mu'_{ij}}[i = b \wedge j = c]$. In this section, we will give a sufficient condition from [23] for the existence of these vectors. Specifically, this paper shows that if μ distributions are uniform on assignments to every pair of variables, then we can find vectors v satisfying the above condition. If the hypergraph has high enough expansion, the μ distributions are uniform as desired.

We now fill in the details. Consider the matrix $M \in \mathbb{R}^{([n]\times\{0,1\})\times([n]\times\{0,1\})}$ indexed by variable-assignment pairs:

$$M_{(i,b),(j,c)} = \mathop{\mathbf{Pr}}_{\mu'_{ij}}[i = b \wedge j = c].$$

To obtain vectors satisfying the above condition, it suffices to show the M is positive semidefinite. The Cholesky decomposition of M then produces vectors $v_{i,b}$ satisfying $\langle v_{i,b}, v_{j,c} \rangle = \mathbf{Pr}_{\mu'_{ij}}[i = b \wedge j = c]$.

Now consider the following matrix M':

$$M'_{(i,b),(j,c)} = \begin{cases} \frac{1}{4} & \text{if } i \neq j \\ \frac{1}{2} & \text{if } i = j \text{ and } b = c \\ 0 & \text{if } i = j \text{ and } b \neq c. \end{cases}$$

In [23], the authors proved the following lemma:

Lemma IV.4 ([23] Lemma 4.4)**.** *M' is positive semidefinite.*

This means that if M were equal to M', we would be done. Unfortunately, this is not quite true. [23] then gives expansion conditions under which $M((i,b),(j,c)) = M'((i,b),(j,c))$. The following lemma is implicit in the proof of [23] Claim 3.4:

Lemma IV.5. *Let P be a $(t-1)$-wise independent arity k predicate and let H be a k-uniform hypergraph.*

1) *If H and $H-\{i\}$ are both $(r, k-t+\epsilon)$-boundary expanding for some $\epsilon > 0$, then for any set of variables S such that $|S| \leq r$, $\mathbf{Pr}_{\mu'_S}[i = b] = \frac{1}{2}$.*
2) *If H and $H-\{i,j\}$ are both $(r, k-t+\epsilon)$-boundary expanding for some $\epsilon > 0$, then for any set of variables S such that $|S| \leq r$, $\mathbf{Pr}_{\mu'_S}[i = b \wedge j = c] = \frac{1}{4}$.*

This follows immediately from Lemma IV.3 by observing that $\{i\}$ and $\{i,j\}$ cannot contain all variables of any constraint, so μ assigns i and i,j uniformly at random.

For the smaller values of m that [23] deal with, these expansion conditions hold. However, for higher values of m, it is likely that there do exist sets of vertices that violate this condition. We can still show that the conditions of Lemma IV.5 are met for all except constant size sets.

Lemma IV.6. *Let H be a $\left(r, k-t+\frac{\delta}{2}\right)$-boundary expanding k-regular hypergraph. For all i,j, the following holds: If S is a set of vertices in $H-\{i,j\}$ and $|S| \geq \frac{5}{\delta}$, then $|\partial S| \geq (k-t+\epsilon)|S|$ for some $\epsilon > 0$.*

Proof: Since H is $\left(r, k-t+\frac{\delta}{2}\right)$-boundary expanding,

$$|\partial S| \geq \left(k-t+\frac{\delta}{2}\right)|S| - 2 \geq (k-t+\epsilon)\,|S|$$

for $\epsilon = \frac{\delta}{10}$ and $|S| \geq \frac{5}{\delta}$. ∎

We now need to address the existence of constant size sets S that have too few boundary variables. Note that if for every i,j and every subset of vertices S of H such that $|S| \leq \frac{5}{\delta}$, it held that

$$|\partial S| \geq (k-t)|S| + 3 \tag{3}$$

then $H-\{i,j\}$ would be $(n^{\Omega(\delta)}, k-t+\epsilon)$-boundary expanding for some $\epsilon > 0$, $M = M'$ and we would be done. We will show that with high probability there are a sublinear number of hyperedges violating (3). We can then delete all of them to obtain a new hypergraph H' for which $M = M'$ and that still has $\Omega\left(n^{t/2-\delta}\right)$ hyperedges. The corresponding local function $f_{H',P}$ therefore is secure against SA$^+$ attacks.

E. There are few bad hyperedges

Define W to be

$$\#\left\{S \subseteq E_H \;\middle|\; |S| \leq \frac{5}{\delta},\ |\partial S| \leq (k-t)|S| + 2\right\}.$$

It then remains for us to show the following lemma:

Lemma IV.7. *With high probability, $W = o(m)$.*

Proof: Let $W = \#\left\{S \subseteq H \text{ such that } |S| \leq \frac{5}{\delta} \text{ and } |\partial S| \leq (k-t)|S| + 2\right\}$. We will first show that $\mathbf{E}[W] = O(n^{1-\delta})$ and then use Markov's Inequality to conclude that $W = o(m)$ with high probability.

To bound $\mathbf{E}[W]$, we first need to bound $\mathbf{Pr}[|\partial T| = s]$ for some particular set T of r hyperedges. We can think of selecting variables for these r hyperedges as filling rk blanks with variables. We know that s of these blanks must be the boundary vertices of T. Every variable that occurs in the remaining $kr-s$ blanks must occur at least twice. Otherwise, T would have more than s boundary vertices. This means that there are at most $\frac{kr-s}{2}$ variables in these $kr-s$ blanks. We can then write

$$\begin{aligned}\mathbf{Pr}[|\partial T| = s] &= \frac{\#\{S \subseteq E_H \mid |S| = r, |\partial S| = s\}}{\#\{S \subseteq E_H \mid |S| = r\}} \\ &\leq \frac{\binom{kr}{s} n^s n^{\frac{kr-s}{2}} \left(\frac{kr-s}{2}\right)^{kr-s}}{\left(\binom{n}{k} k!\right)^r}\end{aligned}$$

because there are $\binom{kr}{s}$ choices of blanks to fill with boundary vertices, at most n^s choices for the s boundary vertices, at most $n^{\frac{kr-s}{2}}$ choices of vertices for the remaining $kr-s$ blanks, and at most $\left(\frac{kr-s}{2}\right)^{kr-s}$ ways of filling these $kr-s$ blanks with the selected vertices. Since k, r, and s are constants, we get that

$$\underset{H}{\mathbf{Pr}}[|\partial T| = s] = O\left(n^{\frac{s-kr}{2}}\right).$$

Let $W_{r,s} = \#\{S \mid |S| = r, |\partial S| = s\}$. Since there are $\binom{O(n^{t/2-\delta})}{r}$ ways of choosing r hyperedges,

$$\begin{aligned}\mathbf{E}[W_{r,s}] &= \binom{O(n^{t/2-\delta})}{r} \cdot O\left(n^{\frac{s-kr}{2}}\right) \\ &= O\left(n^{\frac{s-r(k-t)}{2} - \delta r}\right).\end{aligned}$$

Now we can bound $\mathbf{E}[W]$:

$$\begin{aligned}\mathbf{E}[W] &= \sum_{r=1}^{5/\delta} \sum_{s=0}^{r(k-t)+2} \mathbf{E}[W_{r,s}] \\ &= O\left(n^{1-\delta}\right).\end{aligned}$$

Applying Markov's Inequality, we see that

$$\mathbf{Pr}\left[W \geq n^{3/2-2\delta}\right] \leq O\left(n^{-(1/2-\delta)}\right).$$

So $W = o(m)$ with high probability.

■

We therefore have $o(m)$ "bad" sets of hyperedges. Each such sets contains at most $\frac{5}{\delta}$ hyperedges, so the number of hyperedges participating in at least one of these small, low-expansion sets is $o(m)$. We can find and delete all such sets in time $n^{O(1/\delta)}$. This gives us a new hypergraph H' with $m' = \Omega(n^{t/2-\delta})$ hyperedges for which $f_{H',P}$ is secure against the attack based on computing the SA^+ relaxation value. This completes the proof of Theorem I.2.

Remark IV.8. Observe that these semirandom instances H' are also secure against linear tests when $P = \mathrm{TSA}$. Every linear test that can be applied to $f_{H',\mathrm{TSA}}$ can also be applied to $f_{H,\mathrm{TSA}}$, as we have only deleted hyperedges. Since every linear test for $f_{H,\mathrm{TSA}}$ is unbiased with high probability by the analysis of Section III, it must also be the case that every linear test for $f_{H',\mathrm{TSA}}$ is unbiased with high probability.

V. Security against Lasserre/Parrilo attacks

Herein we show that if Goldreich's generator is instantiated with $P = \mathrm{TSA}$ and a random 5-*partite* hypergraph with $m = O(n^{1.499})$ constraints then it is perfectly secure against $n^{\Omega(1)}$-round Lasserre/Parrilo attacks. More generally and more precisely, we prove the following:

Theorem V.1. *Fix integer constants $3 \le t \le k$ and let $P : \{0,1\}^k \to \{0,1\}$ be any predicate of the form $P(x) = x_1 + \cdots + x_t + Q(x_{t+1}, \ldots, x_k) \pmod 2$, where Q is any $(k-t)$-ary predicate. Fix also a constant $\delta > 0$. Suppose that we choose a list of $m = O(n^{t/2-\epsilon})$ tuples $S^1, \ldots, S^m \subset [n]^k$ independently and uniformly at random. Then except with probability at most $o_n(1)$ we have the following: For* every *choice of "right-hand sides" $b_1, \ldots, b_m \in \{0,1\}$, the constraint satisfaction problem on kn variables $x_1^1, \ldots, x_n^1, \ldots, x_1^k, \ldots, x_n^k$ in which the jth constraint ($j \in [m]$) is*

$$P(x^1_{S^j_1}, x^2_{S^j_2}, \ldots, x^k_{S^j_k}) = b_j$$

has a value-1 Lasserre/Parrilo SDP relaxation value for up to $r = \Omega(n^{(2/t)\delta})$ rounds.

In particular, when this CSP is viewed as a PRG mapping $\{0,1\}^{kn} \to \{0,1\}^m$, the r-round Lasserre/Parrilo SDP has 0 advantage in distinguishing the generator's output from a truly random m-bit string.

Proof: Note that Schoenebeck [37] has proven precisely this theorem in the case that $k = t$; i.e., when P is simply the t-ary XOR predicate.[1] In particular, one can view his proof as constructing an appropriate "pseudoexpectation" operator $\widetilde{\mathbf{E}}[\cdot]$ (see [22]) for degree-$2r$ polynomials under which all t-XOR constraints are satisfied with "pseudoprobability 1".

In our more general setting we can easily obtain the required pseudoexpectation operator by a black-box reduction. The pseudoexpectation operator can simply "deterministically commit" to an *arbitrary* fixed setting of the variables x_i^j for $t < j \le k$ — say, all-1's. This requires us to produce an appropriate pseudoexpectation operator for the resulting t-XOR system (with new jth right-hand side equal to $b'_j = b_j + Q(1, \ldots, 1) \pmod 2$); but Schoenebeck's theorem gives us one for any right-hand sides b'_j. ■

VI. Conclusions

It seems that that the most promising way to use Goldreich's generator to build a k-local PRG is to take a k-partite random hypergraph structure with the predicate $P = \mathrm{XORAND}_{t,k-t}$, where $t = \lfloor \frac{2}{3}k \rfloor$. One might conjecture that this PRG is secure with any stretch $m = o(n^{\frac{1}{2}\lfloor \frac{2}{3}k \rfloor})$ — in particular, stretch $n^{3/2}$ for $k = 5$, and stretch $n^{k/3}$ for k divisible by 3. If true, this would be optimal stretch in light of Theorem II.11.

In this work we have given two kinds of evidence supporting this conjecture for $k = 5$; we have shown security against all $\mathbb{F}_2$-linear attacks and against the Lasserre/Parrilo SDP hierarchy. We also extended the latter evidence to all larger values of k. A good open question that remains is to also extend the former evidence; i.e., to show that Goldreich's generator with $P = \mathrm{XORAND}_{t,k-t}$ is ϵ-biased for $\epsilon = 1/n^{\omega(1)}$ when $m = o(n^{t/2})$.

Acknowledgments

We would like to thank Benny Applebaum, Boaz Barak, Siu On Chan, and Li-Yang Tan for helpful discussions.

References

[1] J. Reif and J. Tygar, "Efficient parallel pseudo-random number generation," in *Proceedings of the 5th annual Advances in Cryptology (CRYPTO)*, 1985, pp. 433–446. I

[1] Actually, in his theorem the constraints are not chosen with a t-partite structure, but rather simply as a random t-uniform hypergraph. However all that his Lasserre/Parrilo lower bound needs is that the hypergraph have a certain strong expansion property; he verifies the property holds with probability $1 - o_n(1)$ for a random t-uniform hypergraph, but it's easy to see the appropriate calculation also holds for a random t-partite hypergraph (assuming t is a constant). Alternatively, in our theorem instead of a k-partite structure we could have just a bipartite random hypergraph structure, with the "XOR variables" segregated from the Q-variables.

[2] O. Goldreich, "Candidate one-way functions based on expander graphs," Electronic Colloquium on Computational Complexity (ECCC), Tech. Rep. 90, 2000. I, I, II-B

[3] M. Cryan and P. B. Miltersen, "On pseudorandom generators in NC^0," in *Proceedings of the 26th annual International Symposium on Mathematical Foundations of Computer Science*, 2001, pp. 272–284. I, II-F

[4] B. Applebaum, Y. Ishai, and E. Kushilevitz, "Cryptography in NC^0," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 845–888, 2006. I

[5] E. Mossel, A. Shpilka, and L. Trevisan, "On ϵ-biased generators in NC^0," in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 136–145. I, I, I-A, II-F, II.8, II-I, III-A

[6] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptography with constant computational overhead," in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, 2008, pp. 433–442. I

[7] B. Applebaum, B. Barak, and A. Wigderson, "Public-key cryptography from different assumptions," in *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, 2010, pp. 171–180. I

[8] J. Håstad and S. Khot, "Query efficient PCPs with perfect completeness," in *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, 2001, pp. 610–619. I

[9] B. Applebaum, A. Bogdanov, and A. Rosen, "A dichotomy for local small-bias generators," in *Proceedings of the 9th International Conference on Theory of Cryptography*, 2012, pp. 600–617. I, I-A, II-F, III-A, III-B, III-B

[10] J. Cook, O. Etesami, R. Miller, and L. Trevisan, "Goldreich's one-way function candidate and myopic backtracking algorithms," in *Proceedings of the 6th annual Theory of Cryptography Conference (TCC)*, 2009, pp. 521–538. I

[11] M. Alekhnovich, E. Hirsch, and D. Itsykson, "Exponential lower bounds for the running time of DPLL algorithms on satisfiable formulas," *Journal of Automated Reasoning*, vol. 35, no. 1-3, pp. 51–72, 2005. I

[12] B. Applebaum, "Pseudorandom generators with long stretch and low locality from random local one-way functions," in *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, 2012, pp. 805–816. I

[13] E. Chlamtac and G. Singh, "Improved approximation guarantees through higher levels of SDP hierarchies," in *Proceedings of the 11th Annual International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, 2008, pp. 49–62. I-A

[14] S. Khot and R. Saket, "SDP integrality gaps with local ℓ_1-embeddability," in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009, pp. 565–574. I-A

[15] P. Raghavendra and D. Steurer, "Integrality gaps for strong SDP relaxations of Unique Games," in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009, pp. 575–585. I-A

[16] J. Lasserre, "Optimisation globale et théorie des moments," *Comptes Rendus de l'Académie des Sciences*, vol. 331, no. 11, pp. 929–934, 2000. I-A

[17] P. Parrilo, "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization," Ph.D. dissertation, California Institute of Technology, 2000. I-A

[18] L. Barto and M. Kozik, "Constraint satisfaction problems of bounded width," in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009, pp. 595–603. I-A

[19] ——, "Robust satisfiability of constraint satisfaction problems," in *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, 2012, pp. 931–940. I-A

[20] M. Wainwright and M. Jordan, "Treewidth-based conditions for exactness of the Sherali–Adams and Lasserre relaxations," University of California, Berkeley, Tech. Rep. 671, 2004. I-A

[21] P. Raghavendra, "Approximating NP-hard problems: efficient algorithms and their limits," Ph.D. dissertation, University of Washington, 2009. I-A

[22] B. Barak, F. Brandão, A. Harrow, J. Kelner, D. Steurer, and Y. Zhou, "Hypercontractivity, sum-of-squares proofs, and their applications," in *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, 2012, pp. 307–326. I-A, II-G2, V

[23] S. Benabbas, K. Georgiou, A. Magen, and M. Tulsiani, "SDP gaps from pairwise independence," *Theory of Computing*, vol. 8, pp. 269–289, 2012. I-A, II-G, II-H, II-H, II-H, II.7, IV-A, IV-B, IV-B, IV-C, IV-C, IV-C, IV-D, IV.4, IV-D, IV-D

[24] R. O'Donnell and Y. Zhou, "Approximability and proof complexity," *arXiv preprint arXiv:1211.1958*, 2012. II-G2

[25] M. Tulsiani and P. Worah, "LS_+ lower bounds from pairwise independence," in *Proceedings of the 27th Annual IEEE Conference on Computational Complexity*, 2013, pp. 121–132. II-H, II-H, II-H, IV-B, IV-C

[26] M. Alekhnovich, S. Arora, and I. Tourlakis, "Towards strong nonapproximability results in the Lovász–Schrijver hierarchy," in *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 2005, pp. 294–303. II-H

[27] B. Applebaum, "Cryptographic hardness of random local functions (survey)," 2013, survey talk at TCC 2013; slides available from the author. II-I

[28] ——, Personal Communication, 2013. II-I

[29] A. Bogdanov and Y. Qiao, "On the security of Goldreich's one-way function," *Computational Complexity*, vol. 21, no. 1, pp. 83–127, 2012. II-I, 1

[30] T. Siegenthaler, "Correlation-immunity of nonlinear combining functions for cryptographic applications," *IEEE Transactions on Information Theory*, vol. 30, no. 5, pp. 776–780, 1984. II.10

[31] U. Feige and E. Ofek, "Easily refutable subformulas of large random 3CNF formulas," *Theory of Computing*, vol. 3, pp. 25–43, 2007. [Online]. Available: http://dx.doi.org/10.4086/toc.2007.v003a002 II-I, II-I

[32] J. Håstad, "An NP-complete problem — some aspects of its solution and some possible applications," Institut Mittag-Leffler, Tech. Rep. 16, 1984. II-I

[33] ——, Personal communication, 2013. II-I

[34] M. Charikar and A. Wirth, "Maximizing quadratic programs: Extending Grothendieck's Inequality," in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004, pp. 54–60. 1

[35] R. Boppana, "Eigenvalues and graph bisection: An average-case analysis," in *Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science*, 1987, pp. 280–285. 2

[36] F. J. MacWilliams and N. Sloane, *The theory of error-correcting codes.* Amsterdam: North-Holland Publishing Co., 1977. III-B

[37] G. Schoenebeck, "Linear level Lasserre lower bounds for certain k-CSPs," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 593–602. V