Log-Seed Pseudorandom Generators via Iterated Restrictions

Dean Doron
Department of Computer Science, Stanford University, CA, USA
https://cs.stanford.edu/~ddoron/
ddoron@stanford.edu

Pooya Hatami
Department of Computer Science & Engineering, Ohio State University, Columbus, OH, USA
https://pooyahatami.org/
pooyahat@gmail.com

William M. Hoza
Department of Computer Science, University of Texas at Austin, TX, USA
https://williamhoza.com/
whoza@utexas.edu

Abstract

There are only a few known general approaches for constructing explicit pseudorandom generators (PRGs). The "iterated restrictions" approach, pioneered by Ajtai and Wigderson [2], has provided PRGs with seed length polylog n or even $\widetilde{O}(\log n)$ for several restricted models of computation. Can this approach ever achieve the optimal seed length of $O(\log n)$?

In this work, we answer this question in the affirmative. Using the iterated restrictions approach, we construct an explicit PRG for *read-once depth-2* $\mathbf{AC}^0[\oplus]$ *formulas* with seed length

$$O(\log n) + \widetilde{O}(\log(1/\varepsilon)).$$

In particular, we achieve optimal seed length $O(\log n)$ with near-optimal error $\varepsilon = \exp(-\widetilde{\Omega}(\log n))$. Even for constant error, the best prior PRG for this model (which includes read-once CNFs and read-once $\mathbb{F}_2$-polynomials) has seed length $\Theta(\log n \cdot (\log\log n)^2)$ [22].

A key step in the analysis of our PRG is a tail bound for *subset-wise symmetric polynomials*, a generalization of elementary symmetric polynomials. Like elementary symmetric polynomials, subset-wise symmetric polynomials provide a way to organize the expansion of $\prod_{i=1}^{m}(1+y_i)$. Elementary symmetric polynomials simply organize the terms by *degree*, i.e., they keep track of the number of variables participating in each monomial. Subset-wise symmetric polynomials keep track of more data: for a fixed partition of $[m]$, they keep track of the number of variables *from each subset* participating in each monomial. Our tail bound extends prior work by Gopalan and Yehudayoff [17] on elementary symmetric polynomials.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Pseudorandom generators, Pseudorandom restrictions, Read-once depth-2 formulas, Parity gates

Digital Object Identifier 10.4230/LIPIcs.CCC.2020.6

Funding *Dean Doron*: Supported by NSF grant CCF-1763311. Part of this work was done while at UT Austin and supported by NSF grant CCF-1705028.
Pooya Hatami: Supported by NSF grant CCF-1947546. Part of this work was done while at UT Austin and supported by a Simons Investigator Award (#409864, David Zuckerman).
William M. Hoza: Supported by the NSF GRFP under Grant DGE-1610403 and by a Harrington Fellowship from UT Austin.

Acknowledgements We thank David Zuckerman for very helpful discussions.

35th Computational Complexity Conference (CCC 2020).
Editor: Shubhangi Saraf; Article No. 6; pp. 6:1–6:36

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The famous "**L** vs. **BPL**" problem asks whether randomness is ever truly necessary for space-efficient computation. To prove $\mathbf{L} = \mathbf{BPL}$, it suffices to design a suitable *pseudorandom generator* (PRG), i.e., an efficient algorithm that stretches a short truly random seed to a long bitstring that "looks random". To be more specific, the action of a small-space algorithm on its random bits can be modeled by a *read-once branching program* (ROBP). Therefore, to prove $\mathbf{L} = \mathbf{BPL}$, it suffices to design an efficient PRG with seed length $O(\log n)$ that fools polynomial-width ROBPs.

A large and growing body of work has made significant progress toward this ambitious goal. Most work on **L** vs. **BPL** can be broadly divided into two main approaches.

1.1 The "Seed Recycling" Approach

The "classical" approach to **L** vs. **BPL** is based on the observation that there is limited communication between the first half of an ROBP and its second half. Therefore, after using a few truly random bits to generate the first half of a pseudorandom string, the truly random bits can be efficiently *recycled* to generate the second half of the pseudorandom string. This insight is essentially due to Nisan [26].

Of the line of work that uses this approach, some highlights include PRGs for polynomial-width ROBPs with seed length $O(\log^2 n)$ [26, 20, 15]; PRGs for constant-width "regular" ROBPs with seed length $\widetilde{O}(\log n)$ [7, 11, 21, 32, 6]; and derandomization techniques that go beyond the construction of PRGs [27, 31]. More recently, this "seed recycling" approach has been used to obtain improved generators for polynomial-width ROBPs when the error parameter ε is very small [5, 19].

1.2 The "Iterated Restrictions" Approach

The more "modern" approach to **L** vs. **BPL** is to design a pseudorandom generator by *iterated pseudorandom restrictions*. That is, we pseudorandomly assign values to a pseudorandomly chosen subset of the variables, and then repeat the process to assign values to all variables. Intuitively, designing a pseudorandom restriction for some function f is easier than fooling f outright, because designing a pseudorandom restriction amounts to fooling a "smoothed out" version of f [16], or equivalently, designing a PRG that *would* fool f if some *noise* were added [18]. This "iterated restrictions" approach goes back to early work by Ajtai and Wigderson [2], but its modern incarnation is largely due to Gopalan et al. [16].

Of the line of work that takes this approach, some highlights include PRGs for *arbitrarily-ordered* ROBPs with seed length $\operatorname{polylog} n$ [33, 9, 14]; PRGs for width-3 ROBPs with seed length $\widetilde{O}(\log n)$ [16, 33, 24]; PRGs for bounded-depth read-once formulas with seed length $\widetilde{O}(\log n)$ [16, 10, 13]; and near-optimal PRGs for arbitrary-order product tests [18, 22].

1.3 Log-Seed PRGs and Our Main Result

At two extremes, one can either try to derandomize *all* of **BPL** as *efficiently as possible* (e.g. [26, 31]), or else one can try to *optimally* derandomize *as much* of **BPL** as possible (e.g. [28, 29]). Let us adopt the second goal.

In some cases, the "seed recycling" approach has indeed yielded PRGs with truly optimal seed length, at least for moderate error. For example, PRGs are known with seed length $O(\log n)$ that fool all $O(\log n)$-space algorithms that use only $\operatorname{polylog}(n)$ random bits in the first place [1, 28, 19]. For another example, PRGs for constant-width "permutation" ROBPs are known with seed length $O(\log n)$ [11, 21, 32].

The present work considers the question of whether the "iterated restrictions" approach can also yield a PRG with seed length $O(\log n)$ for some interesting class of tests. At first glance, this might seem doubtful, since after all we must pay for many pseudorandom restrictions. Nevertheless, we answer in the affirmative, proving the following theorem.

▶ **Theorem 1.** *For all $n \in \mathbb{N}$ and $\varepsilon > 0$, there is an explicit ε-PRG for read-once depth-2 $\mathbf{AC}^0[\oplus]$ formulas on n input bits with seed length*

$$O(\log n) + \widetilde{O}(\log(1/\varepsilon)).$$

Specifically, the seed length of our PRG is $O\left(\log n + \log(1/\varepsilon) \cdot (\log\log(1/\varepsilon))^5\right)$. One can prove a lower bound of $\Omega(\log n + \log(1/\varepsilon))$ on the seed length of any PRG for this model.[1]

1.4 Read-Once Depth-2 $\mathbf{AC}^0[\oplus]$ Formulas

The class of functions that is fooled by our PRG (read-once depth-2 formulas over the basis $\{\wedge, \vee, \oplus\}$, with negations allowed at the inputs for free) is certainly of interest. It includes *read-once CNFs* and *read-once $\mathbb{F}_2$-polynomials* as special cases. The problems of fooling these classes have both received a lot of attention [12, 16, 4, 23, 24, 22]. Previously, even for read-once CNFs, PRGs with seed length $O(\log n)$ were only known for *constant* error [8, 12], whereas our PRG maintains seed length $O(\log n)$ with *near-optimal* error $\varepsilon = \exp(-\widetilde{\Omega}(\log n))$. Meanwhile, for read-once $\mathbb{F}_2$-polynomials, no PRGs with seed length $O(\log n)$ were known at all prior to our work.

Gopalan et al. did give a PRG with *near*-optimal seed length $\widetilde{O}(\log(n/\varepsilon))$ for read-once CNFs, and more generally for read-once depth-2 $\mathbf{AC}^0[\oplus]$ formulas with the property that the output gate is not $\oplus$ [16]. They used their PRG to construct a near-optimal hitting set for width-3 ROBPs [16]. A subsequent line of work provided near-optimal PRGs for *all* read-once depth-2 $\mathbf{AC}^0[\oplus]$ formulas [23, 24, 22].[2]

Conversely, a read-once depth-2 $\mathbf{AC}^0[\oplus]$ formula can be simulated by a width-4 ROBP (after suitably permuting the variables). The problems of designing improved PRGs for width-4 ROBPs and for read-once $\mathbf{AC}^0[\oplus]$ formulas of any constant depth are two major frontiers in unconditional pseudorandomness [24, 13]. The model we study in this paper is an interesting special case.

1.5 Overview of Our Approach

Let us focus on the problem of designing a PRG with seed length $O(\log n)$, with ε as small as possible. For simplicity, assume the test function is a read-once $\mathbb{F}_2$-polynomial $f = f_1 \oplus \cdots \oplus f_m$.

1.5.1 One Restriction

Ultimately, we wish to design a full PRG via *iterated* pseudorandom restrictions. To begin, we will explain how to construct just *one* pseudorandom restriction that assigns values to a constant fraction of the inputs. We use almost $O(\log n)$-wise independence to select the subset of inputs to keep "alive" for each coordinate, where the probability of staying alive is a constant $p \approx 1$. We use a small-bias distribution to assign values to the remaining inputs. Sampling this pseudorandom restriction only costs $O(\log n)$ truly random bits.

[1] This lower bound holds already for fooling parity functions.

[2] The PRGs we are referring to were designed to fool read-once $\mathbb{F}_2$-polynomials, but in fact they fool all of read-once depth-2 $\mathbf{AC}^0[\oplus]$.

We must show that our pseudorandom restriction X is correct. That is, we need to show that

$$\left| \mathop{\mathbb{E}}_{X,U} [f|_X(U)] - \mathbb{E}[f] \right| \leq \varepsilon,$$

where U is a uniform random variable over $\{0,1\}^n$.

We will outline three different arguments for proving correctness, each of which works under certain assumptions about f. We defer to the full proof to explain how to stitch these three arguments together to get a general proof of correctness for any f.

1.5.1.1 Argument 1: Keeping Many Terms Alive

Assume f is a *homogeneous* $\mathbb{F}_2$-polynomial of degree $w \gg \log\log n$, and assume there are many terms, $m \geq 3^w$. (For simplicity, in this informal discussion, we are making stronger assumptions than necessary.) Since f is the *parity* of all these terms, one can show from these assumptions that f is approximately *balanced*, i.e., $\mathbb{E}[f] \approx \frac{1}{2}$. Under a *truly* random restriction, for each term, the probability that all variables in the term remain alive would be p^w, so with high probability, the number of nonconstant terms after the restriction would be at least $m \cdot p^w \geq (3p)^w$. Standard techniques suffice to derandomize this calculation, so after our pseudorandom restriction, with high probability, there are still many terms alive – enough that the restricted function is still approximately balanced.

1.5.1.2 Argument 2: The Forbes-Kelley Approach [14]

Building on prior work [30, 18, 9], Forbes and Kelley showed that a restriction based on δ-biased distributions preserves the expectation of any arbitrary-order constant-width ROBP to within error $1/n$, where $\log(1/\delta) = O(\log n \log\log n)$ [14]. Our test function f can be simulated by a width-4 ROBP under some variable order. Unfortunately, given our budget of $O(\log n)$ truly random bits, we can only afford to sample from a $(1/\operatorname{poly}(n))$-biased distribution.

To move forward, let us turn things around a little: the analysis of Forbes and Kelley shows that a restriction based on δ-biased distributions preserves the expectation to within error ε, where $\varepsilon = \exp(-\Omega(\log(1/\delta)/\log\log(1/\delta)))$. The point is that this latter statement holds even for a relatively large δ, *assuming* the ROBP reads at most $1/\varepsilon$ variables. Therefore, if we assume that our test function f only reads a few variables (say, $\operatorname{polylog} n$ many), then the Forbes-Kelley approach shows that our pseudorandom restriction preserves the expectation of f to within error $\varepsilon = \exp(-\Omega(\log n/\log\log n))$.

1.5.1.3 Argument 3: Subset-Wise Symmetric Polynomials

Assume this time that the degree of every term of f is in the interval $[C\log\log n, C\log n]$ for some appropriate constant C. Assume also that for every w, there are at most 3^w terms of degree w. For this case, we return to an older approach based on symmetric polynomials [16, 17, 24], introduced by Gopalan et al. [16]. The idea is as follows. Let $Z \in \{0,1\}^n$ indicate which variables will remain alive. For convenience, for any $\{0,1\}$-valued function f, let $\overline{f} = (-1)^f$. Having already sampled Z, our remaining task is to argue that the small-bias distribution Y fools the "bias function" defined by

$$\widetilde{f}(x) = \mathop{\mathbb{E}}_{U}[\overline{f}(x + Z \wedge U)].$$

Translating $\{0,1\}$ to $\{\pm 1\}$, the $\oplus$ operation becomes multiplication, i.e., $\overline{f} = \prod_i \overline{f_i}$. For independent random variables, product and expectation can be interchanged, so the bias function of f is the product of the bias functions of the f_i-s. Define $\breve{f}_i$ so that the bias function of f_i is $\mathbb{E}[\overline{f_i}] \cdot (1 + \breve{f}_i)$. That way,

$$\widetilde{f} = \mathbb{E}[\overline{f}] \cdot \prod_{i=1}^{m} (1 + \breve{f}_i). \tag{1}$$

The approach used in prior work [16, 17, 24] is to expand Equation (1) in terms of *elementary symmetric polynomials*. Recall that for $y \in \mathbb{R}^m$, the k-th elementary symmetric polynomial $S_k(y)$ is defined by

$$S_k(y) = \sum_{\substack{I \subseteq [m] \\ |I| = k}} \prod_{i \in I} y_i.$$

We can expand Equation (1) as

$$\widetilde{f} = \mathbb{E}[\overline{f}] \cdot \sum_{k=0}^{m} S_k(\breve{f}_1, \ldots, \widecheck{f_m}). \tag{2}$$

Therefore, the error of our pseudorandom restriction is captured by $\sum_{k=1}^{m} S_k(\breve{f}_1, \ldots, \widecheck{f_m})$. Now we can reason as follows. Pick a cutoff point k_0.

- For $k \leq k_0$, we do a Fourier L_1 calculation to show that $S_k(\breve{f}_1, \ldots, \widecheck{f_m})$ has near-zero expectation even under the small-bias distribution Y.
- For $k \approx k_0$, we do a variance calculation to show that $S_k(\breve{f}_1, \ldots, \widecheck{f_m})$ is small *with high probability* under the uniform distribution, hence also under Y by the previous L_1 calculation.
- Finally we invoke a *tail bound* [17], which says that if S_{k_0} and S_{k_0+1} are both small, then the sum of all subsequent values is also small.

How should we choose the cutoff point k_0? If f is a *homogeneous* $\mathbb{F}_2$-polynomial of degree w, then we should pick $k_0 = \Theta(\frac{\log n}{w})$. That way, k_0 is small enough for the L_1 calculation to work out, because the number of monomials in $S_{k_0}(y_1, \ldots, y_m)$ is

$$\binom{m}{k_0} \leq m^{k_0} \leq 3^{wk_0} \leq \operatorname{poly}(n).$$

But at the same time, k_0 is large enough to sufficiently dampen $S_k(\breve{f}_1, \ldots, \widecheck{f_m})$ for $k \approx k_0$. In fact, one can show that

$$\mathbb{E}[S_k^2(\breve{f}_1(Y), \ldots, \widecheck{f_m}(Y))] \leq \frac{\exp(-\Omega(wk))}{k!},$$

which for $k \approx k_0$ is $\frac{1}{\operatorname{poly}(n) \cdot k!}$. This is small enough for the tail bound to give an overall error of $1/\operatorname{poly}(n)$.

The difficulty, of course, is that f is not necessarily homogeneous, i.e., the terms of f do not necessarily all have the same degree. To address this difficulty, following prior work, let us partition the terms of f into $Q = O(\log \log n)$ *buckets* based on degree, say $f = F_1 \oplus F_2 \oplus \cdots \oplus F_Q$. For each bucket $q \in [Q]$, there is a suitable cutoff point k_0, so our restriction preserves the expectation of F_q.

At this point, the approach taken by prior work has been to invoke a generic *XOR lemma* (see Lemma 6) to argue that our restriction must also preserve the expectation of the parity of the F_q's, i.e., our test function f. This XOR lemma is a suitable generalization of the fact that the Fourier L_1 norm is submultiplicative. Unfortunately, invoking the XOR lemma would require us to start with a smaller-bias distribution Y. Effectively, to invoke the XOR lemma, we would have to pay a factor of Q in the seed length, which we cannot afford.

Therefore, we take a different approach. Our observation is that ideally, the cutoff point k_0 should guarantee that every product $\prod_{i\in I} \widecheck{f}_i$ appearing in $S_{k_0}(\widecheck{f}_1, \ldots, \widecheck{f_m})$ involves $\Theta(\log n)$ of the *input* variables $x_1, \ldots, x_n$. Intuitively, that's why the right choice is $k_0 = \Theta(\frac{\log n}{w})$ for degree w. When the terms of f do not all have the same degree, the products $\prod_{i\in I} \widecheck{f}_i$ appearing in $S_k(\widecheck{f}_1, \ldots, \widecheck{f_m})$ do not all involve the same number of input variables $x_1, \ldots, x_n$, hence there isn't a well-defined correct choice of k_0. This suggests that Equation (2) is simply *not the best expansion* of Equation (1).

These observations motivate the definition of *subset-wise symmetric polynomials*. We defer to Section 2 for the precise definition, but the point is that they allow us to give a more refined expansion of Equation (1), where instead of just keeping track of k (the number of f_i-s participating in each monomial of S_k) we keep track of a whole vector $\vec{k}$ giving the numbers of f_i-s *from each bucket* participating in each monomial of $S_{\vec{k}}$. This allows us to define a norm $\|\vec{k}\|$ that measures the number of *input* variables $x_1, \ldots, x_n$ that participate in each monomial of $S_{\vec{k}}(\widecheck{f}_1, \ldots, \widecheck{f_m})$.

We expand Equation (1) in terms of subset-wise symmetric polynomials by summing over all vectors $\vec{k}$:

$$\widetilde{f} = \mathbb{E}[f] \cdot \sum_{\vec{k}\in\mathbb{N}^Q} S_{\vec{k}}(\widecheck{f}_1, \ldots, \widecheck{f_m}).$$

Now we can cut off this sum at $\|\vec{k}\| = \Theta(\log n)$. To complete the argument, we extend known tail bounds for elementary symmetric polynomials [17] to the case of subset-wise symmetric polynomials.

1.5.2 Iterating the Restriction to Get a Full PRG

So far, we have outlined the proof that our pseudorandom restriction preserves the expectation of the test function f. Our pseudorandom restriction costs $O(\log n)$ truly random bits. But our goal is to design a full PRG with seed length $O(\log n)$. It seems that one restriction already uses up our entire budget of truly random bits, so how can we afford to iterate the process?

A key insight is that if f only reads n' variables ($n' \le n$), then a pseudorandom restriction for f ought to only cost $O(\log n')$ truly random bits rather than $O(\log n)$. This intuition can be justified using standard constructions of n'-wise small-bias distributions [25, 3], provided $n' \ge \log n$. (A similar insight was used previously by Lee and Viola [23].) Let C be a constant such that one pseudorandom restriction costs $C \log n'$ truly random bits.

To simplify the discussion, assume f is homogeneous of degree $w = \Theta(\log n)$. Each restriction keeps approximately a p-fraction of variables alive. For simplicity, assume that in each term, *exactly* a p-fraction of variables remain alive, i.e., assume that after i pseudorandom restrictions, the restricted $\mathbb{F}_2$-polynomial is homogeneous of degree $p^i w$.

We divide into two cases. For the first case, suppose that the number of terms is always at most exponential in the degree. Specifically, suppose the number of terms is at most $16^{w'}$, where w' is the degree at that stage. In this case, our pseudorandom restrictions get cheaper and cheaper as we go. Quantitatively, after i restrictions, the restricted polynomial reads only n' variables, where $n' = p^i w \cdot 16^{p^i w}$. Therefore, the cost of restriction $i+1$ is only

$$C \log\left(p^i w \cdot 16^{p^i w}\right) \leq 5C \cdot p^i w.$$

Therefore, if we do a total of t pseudorandom restrictions, the total cost is bounded by

$$\sum_{i=0}^{t-1} 5Cp^i w.$$

This geometric sum is bounded by $O(w) = O(\log n)$, *regardless* of t. To optimize the error of our PRG, we choose $t = O(\log \log \log n)$; after this many restrictions, the number of living variables is small enough that we can stop the iteration and apply a prior *near*-optimal PRG by Lee [22] to finish the job.

For the second case, suppose that at some stage the number of terms is enormous compared to the degree: the degree is w' and the number of terms is more than $16^{w'}$. This setting was studied previously by Meka, Reingold, and Tal [24], who gave an *optimal* PRG for any function that can be written as a parity of an enormous number of functions on small disjoint variable sets. Therefore, in this case, we can stop doing pseudorandom restrictions, and instead fool the function outright using the PRG by Meka et al. [24].

Of course we do not know in advance which case we are in, but this difficulty can be resolved by straightforward XORing.

2 Subset-Wise Symmetric Polynomials

In this section, we will formally define subset-wise symmetric polynomials and prove suitable tail bounds for them. This section can be read on its own, independent of the application to PRGs. We start by recalling known tail bounds for elementary symmetric polynomials.

2.1 Gopalan and Yehudayoff's Bounds for Symmetric Polynomials

As a reminder, the k-th elementary symmetric polynomial is defined by

$$S_k(y) = \sum_{\substack{I \subseteq [m], \\ |I|=k}} \prod_{i \in I} y_i.$$

We rely on the following tail bound by Gopalan and Yehudayoff [17]. As discussed in Section 1.5.1, the bound says that if two S_k-s in a row are small, then all subsequent S_k-s are small.

▶ **Theorem 2** ([17]). *Let $y \in \mathbb{R}^m$, $\theta > 0$, and $\ell \in \mathbb{N}$ satisfy $S_\ell^2(y) \leq \frac{\theta^\ell}{\ell!}$ and $S_{\ell+1}^2(y) \leq \frac{\theta^{\ell+1}}{(\ell+1)!}$. Then, for every $k \geq \ell$,*

$$|S_k(y)| \leq \left(\frac{64e^2\theta\ell}{k}\right)^{k/2}.$$

The exact statement of Theorem 2 does not appear in Gopalan and Yehudayoff's work [17], but it follows readily from their analysis, and it was used previously by Meka et al. [24, Theorem 5.2].[3]

2.2 Our Tail Bounds for Subset-Wise Symmetric Polynomials

Let $\mathcal{B} = (B_1, \ldots, B_Q)$ be a partition of $[m]$, namely $[m] = B_1 \sqcup \cdots \sqcup B_Q$. (The sets $B_1, \ldots, B_Q$ correspond to the "buckets" discussed in Section 1.5.1.) Throughout this paper, let $\mathbb{N}$ denote the set of *nonnegative* integers, $\mathbb{N} = \{0, 1, 2, \ldots\}$. For a vector $\vec{k} = (\vec{k}[1], \ldots \vec{k}[Q]) \in \mathbb{N}^Q$ and $y \in \mathbb{R}^m$, we define the following polynomial:

$$S_{\vec{k},\mathcal{B}}(y) = \sum_{\substack{I \subseteq [m], \\ \forall q, |B_q \cap I| = \vec{k}[q]}} \prod_{i \in I} y_i.$$

We name these polynomials as *subset-wise symmetric polynomials*, since for every $q \in [Q]$, $S_{\vec{k}}(y)$ when restricted to the B_q variables is a degree $\vec{k}[q]$ symmetric polynomial.

Throughout this section we fix $\mathcal{B} = (B_1, ..., B_Q)$ to be a partition of $[m]$. When the partition $\mathcal{B}$ is clear from the context, we will simply write $S_{\vec{k}}$ instead of $S_{\vec{k},\mathcal{B}}$. To formulate our tails bounds for the subset-wise symmetric polynomials, we will need the following auxiliary polynomials:

$$R_{\vec{k}}(y) \stackrel{\text{def}}{=} S^2_{\vec{k}}(y) \cdot \prod_{q=1}^{Q} \vec{k}[q]!.$$

Given $c > 1$, we will assign each vector $\vec{k} \in \mathbb{N}^Q$ a weight, defined as

$$\|\vec{k}\|_{(c)} = \sum_{q=1}^{Q} c^q \vec{k}[q].$$

(In our PRG application, B_q will be the set of terms with approximately c^q input variables, so $\|\vec{k}\|_{(c)}$ will be approximately the number of input variables participating in each monomial of $S_{\vec{k}}$, as outlined in Section 1.5.1.) It is easy to verify that the above weight function is indeed a norm; however, we will not be using this observation.

The main result of this section is a tail-bound for subset-wise symmetric polynomials. In Lemma 3, the parameter A is analogous to the "cutoff point" k_0 discussed in Section 1.5.1.

▶ **Lemma 3.** *Suppose $c > 1$ and $Q, A \in \mathbb{N}$ satisfy $A > \max\left\{\left(\frac{10^6 c}{c-1}\right) \cdot c^Q, 2^{60} Q^2\right\}$. Let Y be a random variable taking values in $\mathbb{R}^m$. Moreover, suppose for every $\vec{k} \in \mathbb{N}^Q$ with $\|\vec{k}\|_{(c)} \le A$,*

$$\mathop{\mathbb{E}}_{Y}\left[R_{\vec{k}}(Y)\right] \le 2^{-\frac{1}{8}\|\vec{k}\|_{(c)}}.$$

Then, except with probability $2^{-A/2^{23}}$ over $y \sim Y$,

$$\sum_{\substack{\vec{k} \in \mathbb{N}^Q, \\ \|\vec{k}\|_{(c)} > A}} |S_{\vec{k}}(y)| \le 2^{-\frac{A}{1024}}.$$

[3] The careful reader will notice a slight discrepancy between the exact constants of Theorem 2 on the one hand and the statements by Gopalan and Yehudayoff [17] and Meka et al. [24] on the other. This discrepancy reflects a minor mistake in the original paper by Gopalan and Yehudayoff [17] that we have here corrected.

Lemma 3 is similar in spirit to Theorem 2: it says that if the "early" subset-wise symmetric polynomials are small (with high probability), then the "late" subset-wise symmetric polynomials are all small (with high probability).

2.3 Non-probabilistic Tail Bound

Before moving to the proof of Lemma 3 in the next subsection, here we first give a tail-bound in the case when the input y satisfies some useful properties. We will later prove Lemma 3, by showing that a random Y satisfies these properties with high probability. Given a vector $\vec{k} \in \mathbb{N}^Q$, we define the restriction of $\vec{k}$ to a set $\mathcal{Q} \subseteq [Q]$ by

$$\vec{k}|_{\mathcal{Q}}[q] = \begin{cases} \vec{k}[q] & \text{if } q \in \mathcal{Q}, \\ 0 & \text{if } q \notin \mathcal{Q}. \end{cases}$$

Our non-probabilistic tail bound goes as follows.

▶ **Lemma 4.** *Suppose $c > 1$ and $Q, A \in \mathbb{N}$ satisfy $A > \max\left\{\left(\frac{10^6 c}{c-1}\right) \cdot c^Q, 2^{60}Q^2\right\}$. Let $y \in \mathbb{R}^m$ be a fixed vector. Suppose that for every $\vec{k} \in \mathbb{N}^Q$, with $A/10^5 \leq \|\vec{k}\|_{(c)} \leq A$, and for every pair of disjoint sets $\mathcal{Q}_1, \mathcal{Q}_2 \subseteq [Q]$ satisfying $\{q : \vec{k}[q] > 1\} \subseteq \mathcal{Q}_1 \cup \mathcal{Q}_2$, we have*

$$R_{(\vec{k}|_{\mathcal{Q}_1})}(y) \cdot R_{(\vec{k}|_{\mathcal{Q}_2})}(y)^4 \leq 2^{-\frac{1}{32} \cdot \|\vec{k}\|_{(c)}}.$$

Then,

$$\sum_{\substack{\vec{k} \in \mathbb{N}^Q, \\ \|\vec{k}\|_{(c)} > A}} \left|S_{\vec{k}}(y)\right| \leq 2^{-\frac{A}{1024}}.$$

Proof. For a fixed $\ell \in \mathbb{N}$ and $q \in [Q]$, define

$$S_{\ell,q} = \sum_{I \subseteq B_q, |I| = \ell} \prod_{i \in I} y_i,$$

which is the ℓ-th elementary symmetric polynomial applied to $(y_i)_{i \in B_q}$. Similarly, define

$$R_{\ell,q} = S_{\ell,q}^2(y) \cdot \ell!.$$

Fix $\vec{k}$ with $\|\vec{k}\|_{(c)} > A$, let $\lambda = 10^5 \cdot \|\vec{k}\|_{(c)}/A$ and let $\vec{k}' \in \mathbb{N}^Q$ be such that $\vec{k}'[q] = \lceil \vec{k}[q]/\lambda \rceil$. Thus, $A/10^5 \leq \|\vec{k}'\|_{(c)} \leq A/2$. Let $\mathcal{Q} := \{q \in Q : \vec{k}[q] \geq 1\}$, and for each $q \in \mathcal{Q}$, let $\theta_q > 0$ be the smallest[4] number satisfying

$$R_{\vec{k}'[q],q} \leq \theta_q^{\vec{k}'[q]} \quad \text{and} \quad R_{\vec{k}'[q]+1,q} \leq \theta_q^{\vec{k}'[q]+1}. \tag{3}$$

By Theorem 2,

$$\left|S_{\vec{k}[q],q}\right| \leq \left(\frac{64e^2\theta_q \vec{k}'[q]}{\vec{k}[q]}\right)^{\vec{k}[q]/2}.$$

[4] It is possible that $\theta_q = 0$ satisfies Equation (3). In this degenerate case, we must have $S_{\vec{k}[q],q} = 0$. This implies $S_{\vec{k}}(y) = 0$, hence Equation (4) trivially holds.

Subset-wise symmetric polynomials by design can be expressed as a product of elementary symmetric polynomials, hence

$$|S_{\vec{k}}(y)| = \prod_{q=1}^{Q} \left|S_{\vec{k}[q],q}(y)\right| \le \prod_{q\in\mathcal{Q}} \left(\frac{64e^2\theta_q \vec{k'}[q]}{\vec{k}[q]}\right)^{\vec{k}[q]/2}$$
$$= \left(\prod_{q\in\mathcal{Q}} \theta_q^{\vec{k}[q]/\lambda}\right)^{\lambda/2} \cdot \prod_{q\in\mathcal{Q}} \left(8e\sqrt{\vec{k'}[q]/\vec{k}[q]}\right)^{\vec{k}[q]}.$$

By our choice of θ_q,

$$\theta_q^{\vec{k'}[q]} = \max\left\{R_{\vec{k'}[q],q}(y), R_{\vec{k'}[q]+1,q}(y)^{\vec{k'}[q]/(\vec{k'}[q]+1)}\right\}$$
$$\le \max\left\{R_{\vec{k'}[q],q}(y), R_{\vec{k'}[q]+1,q}(y), \sqrt{R_{\vec{k'}[q]+1,q}(y)}\right\}.$$

Observe that $\vec{k}[q]/\lambda \in \left[\vec{k'}[q]-1, \vec{k'}[q]\right]$, and thus $\theta_q^{\vec{k}[q]/\lambda}$ is between $\theta_q^{\vec{k'}[q]-1}$ and $\theta_q^{\vec{k'}[q]}$. If $\vec{k'}[q]=1$, then $\theta_q^{k'_q-1}=1$, and otherwise $\theta_q^{\vec{k'}[q]-1}$ is between $\theta_q^{\vec{k'}[q]}$ and $\sqrt{\theta_q^{\vec{k'}[q]}}$. Therefore,

$$\theta_q^{\vec{k}[q]/\lambda} \le \max\left\{\theta_q^{\vec{k'}[q]}, \sqrt{\theta_q^{\vec{k'}[q]}}, \mathbb{1}_{\vec{k'}[q]=1}\right\}$$
$$\le \max\left\{R_{\vec{k'}[q],q}(y), R_{\vec{k'}[q],q}(y)^{1/4}, R_{\vec{k'}[q]+1,q}(y), R_{\vec{k'}[q]+1,q}(y)^{1/4}, \mathbb{1}_{\vec{k'}[q]=1}\right\}.$$

For every q, choose $\vec{k''}[q] \in \{\vec{k'}[q], \vec{k'}[q]+1\}$ such that

$$\theta_q^{\vec{k}[q]/\lambda} \le \max\left\{R_{\vec{k''}[q],q}(y), R_{\vec{k''}[q],q}(y)^{1/4}, \mathbb{1}_{\vec{k''}[q]=1}\right\}.$$

Note that $\|\vec{k''}\|_{(c)} \ge \|\vec{k'}\|_{(c)}$ and

$$\|\vec{k''}\|_{(c)} \le \|\vec{k'}\|_{(c)} + \sum_{q=1}^{Q} c^q \le \|\vec{k'}\|_{(c)} + \frac{A}{10^6} < A.$$

Therefore, there exist disjoint sets $\mathcal{Q}_1, \mathcal{Q}_2 \subseteq [Q]$ such that $\{q : \vec{k''}[q] > 1\} \subseteq \mathcal{Q}_1 \cup \mathcal{Q}_2$, and that for every $q \in \mathcal{Q}$,

$$\theta_q^{\vec{k}[q]/\lambda} \le \begin{cases} R_{\vec{k''}[q],q}(y) & \text{if } q \in \mathcal{Q}_1, \\ R_{\vec{k''}[q],q}(y)^{1/4} & \text{if } q \in \mathcal{Q}_2, \\ 1 & \text{otherwise.} \end{cases}$$

Multiplying over $q \in \mathcal{Q}$, we get

$$\prod_{q\in\mathcal{Q}} \theta_q^{\vec{k}[q]/\lambda} \le \prod_{q\in\mathcal{Q}_1} R_{\vec{k''}[q],q}(y) \cdot \prod_{q\in\mathcal{Q}_2} \mathbb{R}_{\vec{k''}[q],q}(y)^{1/4}$$
$$= \left(R_{(\vec{k''}|_{\mathcal{Q}_1})}(y)^4 \cdot R_{(\vec{k''}|_{\mathcal{Q}_2})}(y)\right)^{1/4} \le 2^{-\frac{1}{128}\cdot\|\vec{k''}\|_{(c)}} \le 2^{-\frac{1}{128}\cdot\|\vec{k'}\|_{(c)}}.$$

As a result,

$$\begin{aligned}|S_{\vec{k}}(y)| &\le 2^{-\frac{\|\vec{k'}\|_{(c)}}{128}\cdot\frac{\lambda}{2}}\cdot\prod_{q\in\mathcal{Q}}\left(8e\sqrt{\vec{k'}[q]/\vec{k}[q]}\right)^{\vec{k}[q]}\\ &\le 2^{-\frac{\|\vec{k'}\|_{(c)}}{256}\cdot\lambda}\cdot\prod_{q\in\mathcal{Q}}\left(8e\sqrt{2/10^5}\right)^{\vec{k}[q]}\cdot\left(\sqrt{10^5/2}\right)^{\lambda}\\ &\le 2^{-\frac{\|\vec{k'}\|_{(c)}}{256}\cdot\lambda}\cdot 2^{8Q\cdot\lambda}\cdot 4^{-\|\vec{k}\|_1}\le 2^{-\frac{\|\vec{k'}\|_{(c)}}{512}\cdot\lambda}\cdot 4^{-\|\vec{k}\|_1}\le 2^{-\frac{\|\vec{k}\|_{(c)}}{512}}\cdot 4^{-\|\vec{k}\|_1}. \qquad (4)\end{aligned}$$

To see the second inequality, observe that when $\vec{k}[q]>\lambda$, then $\left(8e\sqrt{\vec{k'}[q]/\vec{k}[q]}\right)^{\vec{k}[q]}\le(8e\sqrt{2/10^5})^{\vec{k}[q]}$, and otherwise $\left(8e\sqrt{\vec{k'}[q]/\vec{k}[q]}\right)^{\vec{k}[q]}\le(8e)^{\lambda}$. Summing up over all choices of $\vec{k}$ we get,

$$\begin{aligned}\sum_{\vec{k}\in\mathbb{N}^Q,\|\vec{k}\|_{(c)}>A}|S_{\vec{k}}(y)| &\le \sum_{L=1}^{m}\sum_{\substack{\vec{k}\in\mathbb{N}^Q,\\ \|\vec{k}\|_{(c)}>A,\|\vec{k}\|_1=L}}2^{-\frac{\|\vec{k}\|_{(c)}}{512}}\cdot 4^{-L}\\ &\le 2^{-\frac{A}{512}}\cdot\sum_{L=1}^{m}4^{-L}\cdot\left|\left\{\vec{k}\in\mathbb{N}^q:\|\vec{k}\|_1=L\right\}\right|\\ &= 2^{-\frac{A}{512}}\cdot\sum_{L=1}^{m}4^{-L}\cdot\binom{Q-1+L}{Q-1}\\ &\le 2^{-\frac{A}{512}}\cdot\sum_{L=1}^{m}4^{-L}\cdot 2^{Q-1+L}\le 2^{-\frac{A}{512}}\cdot 2^{Q-1}\cdot\sum_{L=1}^{m}2^{-L}\le 2^{-\frac{A}{1024}}.\end{aligned}$$

◀

2.4 Probabilistic Tail Bound: Proof of Lemma 3

Proof. Let $\vec{k}$, $\mathcal{Q}_1$, and $\mathcal{Q}_2$ be as in the statement of Lemma 4. Using the Cauchy-Schwarz inequality and the concavity of $(\cdot)^{1/4}$, we get

$$\begin{aligned}\mathbb{E}\left[\left(R_{\vec{k}|_{\mathcal{Q}_1}}(Y)\right)^{1/8}\cdot\left(R_{\vec{k}|_{\mathcal{Q}_2}}(Y)\right)^{1/2}\right] &\le \left(\mathbb{E}\left[\left(R_{\vec{k}|_{\mathcal{Q}_1}}(Y)\right)^{1/4}\right]\cdot\mathbb{E}\left[R_{\vec{k}|_{\mathcal{Q}_2}}(Y)\right]\right)^{1/2}\\ &\le \left(\mathbb{E}\left[R_{\vec{k}|_{\mathcal{Q}_1}}(Y)\right]^{1/4}\cdot\mathbb{E}\left[R_{\vec{k}|_{\mathcal{Q}_2}}(Y)\right]\right)^{1/2}\\ &\le \left(2^{-\frac{1}{32}\cdot\|\vec{k}|_{\mathcal{Q}_1}\|_{(c)}}\cdot 2^{-\frac{1}{8}\cdot\|\vec{k}|_{\mathcal{Q}_2}\|_{(c)}}\right)^{1/2}\\ &\le 2^{-\frac{1}{64}\cdot\|\vec{k}|_{\mathcal{Q}_1\cup\mathcal{Q}_2}\|_{(c)}}\\ &\le 2^{-\frac{1}{64}\cdot\left(\|\vec{k}\|_{(c)}-\left(\frac{c}{c-1}\right)\cdot c^Q\right)}\\ &\le 2^{-\frac{1}{64}\cdot\left(\|\vec{k}\|_{(c)}-\frac{A}{20000}\right)}\\ &\le 2^{-\frac{1}{128}\cdot\|\vec{k}\|_{(c)}}.\end{aligned}$$

Therefore, by Markov's inequality, except with probability at most $2^{-\|\vec{k}\|_{(c)}/256}\le 2^{-A/2560000}$, we have

$$\left(R_{\vec{k}|_{\mathcal{Q}_1}}(Y)\right)\cdot\left(R_{\vec{k}|_{\mathcal{Q}_2}}(Y)\right)^4\le 2^{-\frac{\|\vec{k}\|_{(c)}}{32}}.$$

The above analysis was done for a fixed choice of $\vec{k}$, $\mathcal{Q}_1$, and $\mathcal{Q}_2$. The number of choices for such $\vec{k}$ is A^Q (which is subexponential in A), and the number of such $\mathcal{Q}_1$, $\mathcal{Q}_2$ is at most 3^Q (which is a polynomial in A), thus Lemma 3 follows by a union bound. More precisely, one can check that since $A \geq 2^{60}Q^2$, then $(3A)^Q \cdot 2^{-A/2560000} \leq 2^{-A/2^{23}}$. ◀

3 Pseudorandomness Preliminaries

Having completed our analysis of subset-wise symmetric polynomials, we now move on to setting the groundwork for our PRG construction and analysis.

3.1 Probability Basics

Let U_n denote the uniform distribution over $\{0,1\}^n$. We will simply write U if n is clear from context. For $f\colon \{0,1\}^n \to \mathbb{R}$, as a shorthand, we write $\mathbb{E}[f]$ to denote $\mathbb{E}[f(U)]$ and $\mathrm{Var}[f]$ to denote $\mathrm{Var}[f(U)]$. If X is a distribution over $\{0,1\}^n$, we say that X *ε-fools* f, or X *fools* f *with error* ε, if

$$|\mathbb{E}[f(X)] - \mathbb{E}[f]| \leq \varepsilon.$$

We say that X ε-fools a family $\mathcal{F}$ of functions, if it ε-fools every $f \in \mathcal{F}$.

3.2 Small Bias

A *parity function* is a function of the form $f(x) = \bigoplus_{i \in I} x_i$ for some set $I \subseteq [n]$. We say that a random variable $Y \in \{0,1\}^n$ is *δ-biased* if it δ-fools all parity functions. We say that Y is *n'-wise δ-biased* if it δ-fools all parity functions on at most n' bits, i.e., all parity functions with $|I| \leq n'$. There are explicit constructions of n'-wise δ-biased distributions that can be sampled with $O(\log(n'/\delta) + \log\log n)$ truly random bits [25, 3].

Recall that for a function $f\colon \{0,1\}^n \to \mathbb{R}$ with Fourier expansion $f = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \chi_S$, the L_1 norm of f is defined by

$$L_1(f) = \sum_{S \subseteq [n]} |\widehat{f}(S)|.$$

This norm is subadditive ($L_1(f+g) \leq L_1(f) + L_1(g)$) and submultiplicative ($L_1(f \cdot g) \leq L_1(f) \cdot L_1(g)$). Functions with bounded L_1 norm are fooled by small-bias distributions:

▷ Claim 5. If $f\colon \{0,1\}^n \to \mathbb{R}$ and Y is δ-biased, then Y fools f with error $2\delta \cdot L_1(f)$.

We will also rely on the following "XOR lemma" for small-bias distributions.

▶ **Lemma 6** ([16, 24]). *Let $0 < \delta < \varepsilon \leq 1$. Let $f_1, \ldots, f_k\colon \{0,1\}^n \to [-1,1]$ depend on disjoint variable sets, and define*

$$f(x) = \prod_{i=1}^{k} f_i(x).$$

If every δ-biased distribution ε-fools every f_i, then every δ^k-biased distribution fools f with error $16^k \cdot 2\varepsilon$.

3.3 Limited Independence

For $p \in [0,1]$, let Bernoulli$(p)^{\otimes n}$ denote the distribution over $\{0,1\}^n$ where the bits are i.i.d. and each bit has expectation p. For example, $U_n = \text{Bernoulli}(1/2)^{\otimes n}$. For a set $I = \{i_1 < i_2 < \cdots < i_\ell\} \subseteq [n]$ and a string $z \in \{0,1\}^n$, we let $z|_I = z_{i_1} z_{i_2} \dots z_{i_\ell} \in \{0,1\}^\ell$. We say that $Z \in \{0,1\}^n$ is *γ-almost k-wise independent with marginals p* if for every set $I \subseteq [n]$ with $|I| \le k$, the total variation distance between $Z|_I$ and Bernoulli$(p)^{\otimes |I|}$ is at most γ.

▷ Claim 7. For every $n, k, C \in \mathbb{N}$ and $\gamma > 0$, there is an explicit γ-almost k-wise independent distribution with marginals $p = 1 - 2^{-C}$ that can be sampled with $O(Ck + \log(1/\gamma) + \log\log n)$ truly random bits.

Proof. Sample $Y \in \{0,1\}^{Cn}$ from a (Ck)-wise $(2^{-Ck/2-1}\gamma)$-biased distribution. Note that as discussed above Y can be sampled using

$$O\left(\log(2^{Ck}/\gamma) + \log\log n\right) = O\left(Ck + \log(1/\gamma) + \log\log n\right)$$

truly random bits. Divide Y into n blocks $Y^{(1)}, \dots, Y^{(n)} \in \{0,1\}^C$, and set

$$Z_i = 0 \iff Y^{(i)} = 1^C.$$

The desired distribution is $Z \in \{0,1\}^n$.

To prove correctness, let $f\colon \{0,1\}^n \to \{0,1\}$ be any test function depending on only k variables. There is a function $g\colon \{0,1\}^{Cn} \to \{0,1\}$ depending on only Ck variables such that $f(Z) = g(Y)$. By Claim 5,

$$\begin{aligned} |\,\mathbb{E}[f(Z)] - \mathbb{E}[f(\text{Bernoulli}(p)^{\otimes n})]| &= |\,\mathbb{E}[g(Y)] - \mathbb{E}[g]| \\ &\le 2^{-Ck/2-1} \cdot 2\gamma \cdot L_1(g) \le \gamma. \end{aligned}$$ ◁

The expectation parameter p can be "amplified" by drawing independent samples and combining with a coordinate-wise conjunction:

▷ Claim 8. Let Z be γ-almost k-wise independent with marginals p. Draw t independent samples $z^{(1)}, \dots, z^{(t)} \sim Z$, and let $Z' = z^{(1)} \wedge \cdots \wedge z^{(t)}$. Then Z' is $(t\gamma)$-almost k-wise independent with marginals p^t.

Proof sketch. The proof is a simple hybrid argument. Draw t independent samples $r^{(1)}, \dots, r^{(t)} \sim \text{Bernoulli}(p)^{\otimes n}$, and let

$$Z^{(i)} = z^{(1)} \wedge \cdots \wedge z^{(i)} \wedge r^{(i+1)} \wedge \cdots \wedge r^{(t)}.$$

One can show by induction on i that $Z^{(i)}$ is $(i\gamma)$-almost k-wise independent with marginals p^t. ◁

3.4 PARITY ∘ AND Formulas

Recall that our main result (Theorem 1) is a PRG for read-once depth-2 $\mathbf{AC}^0[\oplus]$. For most of the paper, we will focus on the special case that the root gate is $\oplus$ and its immediate children are $\wedge$ gates. That is, define a PARITY ∘ AND *formula* to be a function of the form

$$f(x) = \bigoplus_{i=1}^{m} f_i(x),$$

where each f_i is a conjunction of literals, i.e., variables or their negations. We refer to $f_1, \ldots, f_m$ as the *terms* of f. We say that the formula is *read-once* if each variable x_i appears in at most one term. Most of our effort will be spent fooling read-once PARITY $\circ$ AND formulas. Note that this is a slight generalization of read-once $\mathbb{F}_2$-polynomials due to the availability of $\neg$ gates. We will explain in Section 5.6 why it is sufficient to focus on this special case.

The *width* of a term is the number of variables in the term; the width of f is the maximum width of its terms. The *length* of f is m, the number of its terms.

For convenience, if f is a function taking values in $\{0,1\}$, we let $\overline{f} = (-1)^f$. That way, if f is a PARITY $\circ$ AND formula,

$$\overline{f} = \prod_{i=1}^{m} \overline{f_i}.$$

3.5 Restrictions

A *restriction* is a string $x \in \{0,1,\star\}^n$; intuitively, $x_i = \star$ means that x_i has still not been assigned a value. We define an associative *composition* operation on restrictions by the formula

$$(x \circ x')_i = \begin{cases} x_i & \text{if } x_i \neq \star, \\ x'_i & \text{otherwise.} \end{cases}$$

For a function f on $\{0,1\}^n$, the *restricted* function $f|_x$ on $\{0,1\}^n$ is defined by

$$f|_x(x') = f(x \circ x').$$

A restriction x can be specified by two strings $y, z \in \{0,1\}^n$ using the following notation[5]. Define $\mathrm{Res}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1,\star\}^n$ by

$$(\mathrm{Res}(y,z))_i = \begin{cases} \star & \text{if } z_i = 1, \\ y_i & \text{if } z_i = 0. \end{cases}$$

In words, z indicates the $\star$ positions, and y provides the bits in the non-$\star$ positions.

3.6 Pseudorandom Restrictions

Let Y, Z be distributions over $\{0,1\}^n$, and let $X = \mathrm{Res}(Y,Z)$. For a function $f\colon \{0,1\}^n \to \mathbb{R}$, we say that the distribution X *preserves the expectation of* f *with error* ε if

$$|\mathbb{E}[f|_X(U)] - \mathbb{E}[f]| \leq \varepsilon.$$

An equivalent condition is that $|\mathbb{E}[f(Y + Z \wedge U)] - \mathbb{E}[f]| \leq \varepsilon$, where $+$ denotes addition over $\mathbb{F}_2^n$ and $\wedge$ denotes coordinate-wise conjunction. This second condition is the "pseudorandomness plus noise" perspective [18] (the string $Z \wedge U$ can be thought of as a noise vector.)

[5] With apologies, we here flip the order of the arguments to Res compared to the notation used in the authors' prior work [13].

If f takes on values in $\{0,1\}$, for each particular value z that Z might take on, we define the *bias function* [16] $\widetilde{f_z}\colon \{0,1\}^n \to [-1,1]$ by

$$\widetilde{f_z}(x) = \mathbb{E}\left[\overline{f}(x + z \wedge U)\right].$$

(We use $\overline{f}$ rather than f simply for convenience.) The statement that X preserves the expectation of f with error ε is also equivalent to the condition

$$\left|\mathop{\mathbb{E}}_{Z}\left[\mathop{\mathbb{E}}_{Y}[\widetilde{f_Z}(Y)] - \mathbb{E}\left[\overline{f}\right]\right]\right| \leq 2\varepsilon.$$

When z is clear from context, we will just write $\widetilde{f}$ instead of $\widetilde{f_z}$.

If X is a distribution over $\{0,1,\star\}^n$ and $t \in \mathbb{N}$, let $X^{\circ t}$ denote the distribution over $x \in \{0,1,\star\}^n$ obtained by drawing independent samples $x^{(1)}, \ldots, x^{(t)} \sim X$ and composing them, $x = x^{(1)} \circ \cdots \circ x^{(t)}$.

Suppose $\mathcal{F}$ is a class of Boolean functions that is closed under restriction. If X preserves the expectation of every $f \in \mathcal{F}$ with error ε, then $X^{\circ t}$ preserves the expectation of every $f \in \mathcal{F}$ with error $t\varepsilon$. Furthermore, informally, if X "has $\star$-probability p", then $X^{\circ t}$ "has $\star$-probability p^t". To be precise, we can consider the case $X = \mathrm{Res}(Y, Z)$ where Z is γ-almost k-wise independent with marginals p. Then the distribution of $\star$ positions in $X^{\circ t}$ is described by Claim 8.

4 Applying a Single Restriction

In this section, we prove that the expectation of a PARITY $\circ$ AND formula is preserved under a suitable pseudorandom restriction. The cost of the restriction is only $O(\log n)$ truly random bits, the error is $\exp(-\widetilde{\Omega}(\log n))$ (near-optimal), and the restriction assigns values to a constant fraction of the inputs.

4.1 Restriction Construction

Set $C = 500$, $\overline{C} = 2000C$, $c = 1.1$, and $\beta = 0.95$, and consider the following two distributions.

- Let Y be a δ^3-biased distribution over $\{0,1\}^n$ for $\delta = \min\left\{n^{-12\overline{C}}, \frac{1}{2}n^{-\frac{5c}{c-1}-1}\right\} = n^{-12{,}000{,}000}$.[6]
- Let Z be a γ-almost k-wise independent distribution over $\{0,1\}^n$ with marginals $p = 1 - 2^{-C}$, for $k = 6\log n$ and $\gamma = n^{-9}$.

Our restriction is $\mathrm{Res}(Y, Z)$, i.e., Z indicates where to put $\star$ and Y fills in the non-$\star$ bits.

▶ **Lemma 9.** *Let f be a read-once* PARITY $\circ$ AND *formula over n variables of width at most $C\log n$. Then, $\mathrm{Res}(Y, Z)$ preserves the expectation of f to within error $2^{-C\frac{\log n}{\log\log n}}$, i.e.,*

$$|\mathbb{E}[f(Y + Z \wedge U)] - \mathbb{E}[f]| \leq 2^{-C\frac{\log n}{\log\log n}}.$$

4.2 Buckets

Toward proving Lemma 9, we first set some preliminary notations. Recall that f is of the form

$$f = \bigoplus_{i=1}^{m} f_i = \bigoplus_{i=1}^{m} \bigwedge_{j=1}^{w_i} \ell_{ij},$$

where every literal ℓ_{ij} is either some variable in $\{x_1, \ldots, x_n\}$ or its negation.

[6] No attempt was made to optimize the constants.

Set $Q = \lceil \log_c(C \log n) \rceil = O(\log \log n)$. We partition the terms of f into Q buckets according to their *width*. Namely, for each $q \in [Q]$ we define the interval $I_q = [c^{q-1}, c^q)$ and define $B_q \subseteq [m]$ to be the set of indices i such that $w_i \in I_q$. Also, for $q \in [Q]$ we define

$$F_q = \bigoplus_{i \in B_q} f_i,$$

so $f = \bigoplus_{q=1}^{Q} F_q$. For every $q \in [Q]$ we further denote $m_q = |B_q|$.

We divide into two cases (Section 4.3 and Section 4.4) depending on whether there exists a bucket with substantially many terms. Lemma 9 will follow immediately from Lemma 10 and Lemma 15, which cover these two cases respectively.

4.3 Case I – There Exists a Heavy Bucket

Say that bucket $q \in [Q]$ is *heavy* if both $m_q > 3^{c^q}$ and $m_q > \log^C n$. The first case is that there exists a heavy bucket (i.e., there are many terms of roughly the same width, even relative to q). In this case, we will argue that f itself is balanced and also that it stays balanced, w.h.p., after a pseudorandom restriction.

▶ **Lemma 10.** *Let f be a read-once* PARITY $\circ$ AND *formula over n variables of width at most $C \log n$. Suppose there exists a heavy bucket as defined above. Then, with probability at least $1 - \frac{1}{n}$ over $(y, z) \sim Y \times Z$,*

$$\left| \mathbb{E}[\overline{f}|_{\mathrm{Res}(y,z)}] - \mathbb{E}[\overline{f}] \right| \leq \frac{1}{n}.$$

Toward proving Lemma 10, let us define a few more auxiliary notations. Write

$$f = f_{\mathsf{rest}} \oplus F_q,$$

where q is a heavy bucket.

▷ Claim 11. It holds that $\left| \mathbb{E}[\overline{f}] \right| \leq \frac{1}{4n}$.

Proof. By the read-once property and the fact that $\overline{f_{\mathsf{rest}}}$ is bounded,

$$\left| \mathbb{E}[\overline{f}] \right| = \left| \mathbb{E}[\overline{f_{\mathsf{rest}}}] \, \mathbb{E}[\overline{F_q}] \right| \leq \left| \mathbb{E}[\overline{F_q}] \right| = \prod_{i \in B_q} \left| \mathbb{E}[\overline{f_i}] \right|.$$

Each term in F_q has width at least c^{q-1}, so

$$\left| \mathbb{E}[\overline{f}] \right| \leq \left(1 - 2 \cdot 2^{-c^{q-1}}\right)^{m_q} \leq e^{-2 \cdot 2^{-c^{q-1}} \cdot m_q}.$$

Recalling that $m_q \geq 3^{c^q}$, we have $2^{-c^{q-1}} \geq m_q^{\gamma}$ for $\gamma = \log_3 2^{c^{-1}} < \frac{3}{4}$. Thus, using that fact that $m_q \geq \log^C n$,

$$\left| \mathbb{E}[\overline{f}] \right| \leq e^{-2 m_q^{1-\gamma}} \leq e^{-2 \log^{(1-\gamma)C} n} \leq 2^{-\log^{100} n}. \qquad \triangleleft$$

Next, we must analyze the bias of f after the pseudorandom restriction. Let n_q be the number of variables read by F_q. Let $b = \lceil \log_3 n_q \rceil$. We will group the terms of F_q into *blocks*, each of which reads roughly b variables. To define this grouping, first observe that

$b \geq \log_3 m_q$, as each term reads at least one variable. Recalling that $c^q < \log_3 m_q$, we know that $b > c^q$. Therefore, since each term in F_q has width at most c^q, we can write

$$F_q = \bigoplus_{i=1}^{B} g_i,$$

where each block g_i reads b_i variables for $b_i \in \left[b - \frac{1}{2}c^q, b + \frac{1}{2}c^q\right]$.

Let us now estimate B, the number of blocks. Since $b > c^q$, $b_i \in \left[\frac{b}{2}, \frac{3b}{2}\right]$. Also, $m_q > \log^C n$ so $b > \frac{C}{2}\log\log n$. Thus, on the one hand,

$$B \geq \frac{2n_q}{3b} \geq \frac{2 \cdot 3^b}{9b},$$

and on the other hand, $B \leq n_q \leq 3^b$.

Toward arguing that f is balanced after pseudorandom restrictions, we wish to show that with high probability, $z \sim Z$ keeps many variables in many terms alive.

▶ **Definition 12.** *For $z \in \{0,1\}^n$ and a formula f, we say f is* good under z *if z assigns* 1 *to at least a $(1-\beta)$-fraction of the variables f reads.*

▷ Claim 13. For a fixed $z \in \{0,1\}^n$, let $\mathbf{X}_z \subseteq [B]$ be the set of blocks g_i that are not good under z. Then, with probability at least $1 - \frac{1}{2n}$ over $z \sim Z$,

$$|\mathbf{X}_z| \leq \left\lceil \frac{4\log n}{b} \right\rceil.$$

Proof. Set $k_0 = \left\lceil \frac{4\log n}{b} \right\rceil$. Let $S \subseteq [B]$ be some subset of cardinality k_0. We first bound the probability p that every block g_i for $i \in S$ is bad under $z \sim Z$. For a *truly random* $z \sim \text{Bernoulli}(1-2^{-C})^{\otimes n}$, the above probability is bounded by

$$\prod_{i \in S} \binom{b_i}{\beta b_i} 2^{-C\beta b_i} \leq \prod_{i \in S} 2^{b_i} 2^{-5b_i} \leq \left(2^{-4 \cdot \frac{b}{2}}\right)^{|S|} \leq n^{-8}.$$

Now, for every $i \in [B]$, $k \geq k_0 b_i$ so for $z \sim Z$, we get that $p \leq n^{-8} + \gamma \leq 2n^{-8}$. Thus, by the union bound, with probability at most

$$\binom{B}{k_0} p \leq 2B^{k_0} n^{-8} \leq 2\left(3^b\right)^{\frac{4\log n}{b}} n^{-8} \leq 2n^{-(2-\log 3)4} \leq n^{-\frac{4}{3}} < \frac{1}{2n}$$

there will be some S whose all blocks are bad. Taking the contrapositive, we infer that with probability at least $1 - \frac{1}{2n}$ over $z \sim Z$, at most k_0 of the g_i-s are bad under z. ◁

▶ **Lemma 14.** *With probability at least $1 - \frac{1}{n}$ over $(y,z) \sim Y \times Z$, it holds that*

$$\left|\mathbb{E}\left[\overline{f|_{\text{Res}(y,z)}}\right]\right| \leq \frac{1}{2n}.$$

Proof. Fix a *good* z, for which at most $\frac{4\log n}{b}$ of the g_i-s are not good under it. By Claim 13, z is good with probability at least $1 - \frac{1}{2n}$. Let $B^{\mathsf{alive}} = [B] \setminus \mathbf{X}_z$, so

$$f = f_{\mathsf{rest}} \oplus \left(\bigoplus_{i \in B^{\mathsf{alive}}} g_i\right) \oplus \left(\bigoplus_{i \in [B] \setminus B^{\mathsf{alive}}} g_i\right).$$

For every $i \in B^{\mathsf{alive}}$, set the following notations.

- For every $y \in \{0,1\}^n$, let g_i^y denote the function $g_i|_{\mathrm{Res}(y,z)}$.
- Let $I_i^{\mathsf{dead}} \subseteq [n]$ be the literals read by g_i for which $z = 0$. As $i \in B^{\mathsf{alive}}$, $\left|I_i^{\mathsf{dead}}\right| \leq \beta b_i \leq \frac{3\beta}{2} b$. Note that each literal $j \in I_i^{\mathsf{dead}}$ is set by $y \sim Y$.
- Let $I_i^{\mathsf{alive}} \subseteq [n]$ be the literals read by g_i for which $z = 1$. As $i \in B^{\mathsf{alive}}$, $\left|I_i^{\mathsf{alive}}\right| \geq (1-\beta) b_i \geq \frac{1-\beta}{2} b$.

Define the function h_i so that $h_i(y) = 1$ if g_i^y is a nonconstant function, and 0 otherwise. Namely,

$$h_i(y) = \bigwedge_{j \in I_i^{\mathsf{dead}}} y'_j,$$

where y'_j is either y_j or $\neg y_j$ depending on whether y_j appears positively or negatively in g_i. Also, define

$$S(y) = \sum_{i \in B^{\mathsf{alive}}} h_i(y),$$

where the sum is over the reals. Denote

$$\mu = \mathbb{E}[S(U)] = \sum_{i \in B^{\mathsf{alive}}} 2^{-\left|I_i^{\mathsf{dead}}\right|},$$

and note that $\mu \geq |B^{\mathsf{alive}}| \cdot 2^{-\frac{3\beta}{2} b}$. Set $\Delta S = S - \mu$. The spectral norm of the AND function is 1, and so by the sub-additivity we get that $L_1(\Delta S) \leq 2|B^{\mathsf{alive}}|$. Set $\ell = 2\left\lceil \frac{C \log n}{2 \log(2|B^{\mathsf{alive}}|)} \right\rceil$. By the sub-multiplicativity of the spectral norm we have that

$$L_1\left(\Delta S^\ell\right) \leq \left(2\left|B^{\mathsf{alive}}\right|\right)^\ell \leq n^C.$$

For $\varepsilon = \frac{1}{2}$, note that $\delta \leq \frac{\varepsilon}{2} \cdot L_1\left(\Delta S^\ell\right)^{-1}$. By Claim 5, Y ε-fools the function ΔS^ℓ, so

$$\left|\mathbb{E}\left[(S(Y) - \mu)^\ell\right] - \mathbb{E}\left[(S(U) - \mu)^\ell\right]\right| \leq \varepsilon. \tag{5}$$

Next, observe that $\Delta S(U)$ is the sum of zero-mean *independent* random variables, as the h_i-s are supported over disjoint set of variables. Set $A = |B^{\mathsf{alive}}| \cdot 2^{-4\beta b}$. By the Chernoff bound,

$$\begin{aligned}\mathbb{E}\left[\Delta S(U)^\ell\right] &\leq \mathbb{E}\left[\Delta S(U)^\ell \mid \Delta S(U)^\ell \leq A^\ell\right] \\ &\qquad + \mathbb{E}\left[\Delta S(U)^\ell \mid \Delta S(U)^\ell \geq A^\ell\right] \cdot \Pr\left[\Delta S(U)^\ell \geq A^\ell\right] \\ &\leq A^\ell + \left|B^{\mathsf{alive}}\right|^\ell \cdot \Pr\left[\Delta S(U) \geq A\right] \leq \left|B^{\mathsf{alive}}\right|^\ell \cdot \left(2^{-4\beta b\ell} + e^{-\frac{2A^2}{|B^{\mathsf{alive}}|}}\right).\end{aligned}$$

Recall that $b > \frac{C}{2} \log\log n$, so $3^b \geq 36 \log n$ for a large enough n, and since $B \geq \frac{2}{9b} 3^b$ we get that $B \geq \frac{8 \log n}{b}$ and $|B^{\mathsf{alive}}| \geq B - \frac{4 \log n}{b} \geq \frac{B}{2}$. Next, we observe that

$$\frac{2A^2}{|B^{\mathsf{alive}}|} = 2\left|B^{\mathsf{alive}}\right| 2^{-8\beta b} \geq B \cdot 2^{-8\beta b} \geq \frac{2}{9b} 2^{(\log 3 - 8\beta) b} \geq 2^b.$$

As $b\ell \leq \frac{Cb \log n}{\log B} \leq C \log n$, we can conclude that $2^b \geq 4\beta b\ell$ and so $e^{-\frac{2A^2}{|B^{\mathsf{alive}}|}} \leq 2^{-4\beta b\ell}$, which implies that $\mathbb{E}\left[\Delta S(U)^\ell\right] \leq 2|B^{\mathsf{alive}}|^\ell \cdot 2^{-4\beta b\ell}$.

Using Equation (5) and the above bound yields a bound on $\mathbb{E}\left[\Delta S(Y)^\ell\right]$. By Markov's inequality,

$$\Pr\left[S(Y) < \frac{\mu}{2}\right] \leq \frac{\mathbb{E}\left[(S(Y)-\mu)^\ell\right]}{(\mu/2)^\ell} \leq \frac{\varepsilon + 2|B^{\mathsf{alive}}|^\ell \cdot 2^{-4\beta b\ell}}{(\mu/2)^\ell} \leq \left(\frac{8\left|B^{\mathsf{alive}}\right| 2^{-4\beta b}}{\mu}\right)^\ell. \quad (6)$$

Recalling that $\mu \geq |B^{\mathsf{alive}}| \cdot 2^{-\frac{3\beta}{2}b}$, Equation (6) becomes

$$\Pr\left[S(Y) < \frac{\mu}{2}\right] \leq \left(8 \cdot 2^{\left(-4\beta+\frac{3\beta}{2}\right)b}\right)^\ell < 2^{-2\beta b\ell} \leq 2^{-\frac{1}{2}\beta C \log n} \leq \frac{1}{2n},$$

where we have used the fact that $b\ell \geq \frac{C\log n}{4}$.

Overall, with probability at least $1 - \frac{1}{2n}$ over $y \sim Y$, g_i^y is nonconstant for at least $\frac{\mu}{2}$ of the i-s, and recall that each such g_i^y is over at least $(1-\beta)b_i$ variables. Fix such a good y, and let $\mathbf{G} \subseteq [B^{\mathsf{alive}}]$ be the set of nonconstant g_i^y-s. Again, we can write

$$\bigoplus_{i \in B^{\mathsf{alive}}} g_i^y = \left(\bigoplus_{i \in \mathbf{G}} g_i^y\right) \oplus \left(\bigoplus_{i \in B^{\mathsf{alive}} \setminus \mathbf{G}} g_i^y\right) \triangleq t_1 \oplus t_2.$$

Similarly to Claim 11, in order to bound the bias of $\overline{f|_{\mathrm{Res}(Y,Z)}}$ it is sufficient to bound the bias of $\overline{t_1}$, and so

$$\mathbb{E}[\overline{t_1}] \leq \left(1 - 2^{-\frac{3b}{2}}\right)^{\frac{\mu}{2}}.$$

Using the fact that $\mu \geq \frac{1}{2}B \cdot 2^{-\frac{3\beta}{2}b} \geq \frac{1}{9b}2^{\left(\log 3 - \frac{3\beta}{2}\right)b} > 2^{\frac{301}{200}b}$, we get

$$\mathbb{E}[\overline{t_1}] \leq e^{-2^{-\frac{3b}{2}} 2^{\frac{301b}{200}}} \leq e^{-\log^{\frac{C}{400}} n} \leq \frac{1}{2n}. \quad \blacktriangleleft$$

Proof of Lemma 10. Finally, the fact that with probability at least $1 - \frac{1}{n}$ over $(y,z) \sim Y \times Z$, $\left|\widetilde{f}_z(y) - \mathbb{E}[\overline{f}]\right| \leq \frac{1}{n}$, follows immediately from Claim 11 and Lemma 14. $\blacktriangleleft$

4.4 Case II – There Are No Heavy Buckets

In this subsection, we prove that a single pseudorandom restriction preserves the expectation in the case where there is no such a heavy B_q. Namely, for every $q \in [Q]$, either $m_q \leq 3^{c^q}$ or $m_q \leq \log^C n$ (or both).

▶ **Lemma 15.** *Let f be a read-once* PARITY $\circ$ AND *formula over n variables in which the width of every term is at most $C \log n$, and in which there are no heavy buckets as described above. Then, with probability at least $1 - \frac{1}{2} \cdot 2^{-C\frac{\log n}{\log\log n}}$ over $z \sim Z$ it holds that*

$$\left|\mathbb{E}\left[\widetilde{f}_z(Y)\right] - \mathbb{E}[\overline{f}]\right| \leq \frac{1}{2} \cdot 2^{-C\frac{\log n}{\log\log n}}.$$

Toward proving Lemma 15, we partition the Q buckets into two sets and treat terms that fall into each set of buckets separately. Namely, define the two sets as follows.

- $\mathcal{A} = \left\{q \in [Q] : m_q \leq \log^{2C} n\right\}$. We refer to these buckets as the *sparse buckets*.
- $\mathcal{B} = [Q] \setminus \mathcal{A}$. We refer to these buckets as the *well-behaved buckets*.

For each set $\mathcal{T} \in \{\mathcal{A}, \mathcal{B}\}$ we denote

$$f_{\mathcal{T}} = \bigoplus_{i \in \mathcal{T}} F_i,$$

and so $f = f_{\mathcal{A}} \oplus f_{\mathcal{B}}$. The next two subsections will be devoted to proving that the expectation of each $f_{\mathcal{T}}$ is preserved after a single pseudorandom restriction. In Section 4.4.3 we will combine the two results using the XOR lemma for small-bias distributions (Lemma 6) to prove Lemma 15.

4.4.1 Handling Sparse Buckets

For the sparse buckets, we will follow the Forbes-Kelley approach [14] to prove the following.

▶ **Lemma 16.** *With probability at least* $1 - \frac{1}{4} \cdot 2^{-C\frac{\log n}{\log\log n}}$ *over* $z \sim Z$*, it holds that*

$$\left| \mathbb{E}\left[\left(\widetilde{f_{\mathcal{A}}}\right)_z (Y) \right] - \mathbb{E}[\widetilde{f_{\mathcal{A}}}] \right| \leq \frac{1}{4} \cdot 2^{-C\frac{\log n}{\log\log n}}.$$

As outlined in Section 1.5.1, Lemma 16 follows readily from the work by Forbes and Kelley [14]. We require our restriction to work *with high probability* over $z \sim Z$, not merely in expectation, so we must redo some of Forbes and Kelley's analysis. (No substantial modification is needed.) The details follow.

Proof of Lemma 16. First, recall that each term in $f_{\mathcal{A}}$ is of width at most $C \log n$. There are at most $\log^{2C} n$ terms in each bucket, and at most $Q = O(\log\log n)$ such buckets, so overall $f_{\mathcal{A}}$ reads at most $n' = \log^{2C+2} n$ variables.

Note that $f_{\mathcal{A}}$ can be computed by a width-4 ROBP of length n'. We follow [14] and let $G\colon \{0,1\}^{n'} \to \mathbb{R}^{4\times 4}$ encode the transition of the branching program. Namely, perhaps after renumbering the variables, we have $G(x) = G_1(x_1) \cdot \ldots \cdot G_{n'}(x_{n'})$ where $G_i(x_i) = A_{i,x_i}$ for $A_{i,b}$ being the transition matrix that corresponds to taking the bit b while at layer i. Set $k_0 = \frac{8\log n}{\log\log n}$, and note that $k_0 \leq k$. By [14, Lemma 4.1], G can be written as

$$G = \mathbb{E}[G] + L + \sum_{i=1}^{n'} H_i \cdot G^{>i},$$

where L has degree[7] less than k_0, H_i is of degree exactly k_0, $G^{>i}$ is a width-4 ROBP, and H_i and $G^{>i}$ are on disjoint set of variables. More specifically,

$$L = \sum_{\alpha \in \mathbb{F}_2^{n'}, 0<|\alpha|<k_0} \widehat{G}_\alpha \chi_\alpha$$

is the truncated Fourier expansion of G, $G^{>i}(x_{i+1}, \ldots, x_n) = G_{i+1}(x_{i+1}) \cdot \ldots \cdot G_{n'}(x_{n'})$, and

$$H_i = \sum_{\alpha \in \mathbb{F}_2^{n'}, |\alpha|=k_0, \alpha_i=1} \widehat{G^{\leq i}}_\alpha \chi_\alpha,$$

where $G^{\leq i}(x_1, \ldots, x_i) = G_1(x_1) \cdot \ldots \cdot G_i(x_i)$. Let $\|\cdot\|$ be the Frobenius norm. By sub-additivity, we have

[7] We say a function $H\colon \{0,1\}^n \to \mathbb{R}^{w\times w}$ having Fourier expansion $\sum_{\alpha\in\mathbb{F}_2^n} \widehat{H}_\alpha \chi_\alpha$ has degree d if $\widehat{H}_\alpha$ is the zero matrix for every α with Hamming weight larger than d.

$$\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}[G(Y+Z\wedge U)]-\mathbb{E}[G]\right\|\right]\leq\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}[L(Y+Z\wedge U)]\right\|\right]+\sum_{i=1}^{n'}\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}\left[(H_i\cdot G^{>i})(Y+Z\wedge U)\right]\right\|\right]. \quad (7)$$

Just as in [14], the low-degree term L is dealt with a δ-biased distribution. From the work of Chattopadhyay, Hatami, Reingold, and Tal [9] we know that

$$L_1(L)=\sum_{k'=1}^{k_0}(c_{\mathsf{CHRT}}\log n')^{4k'}\leq 2(c_{\mathsf{CHRT}}\log n')^{4k_0}$$

for some universal constant $c_{\mathsf{CHRT}}\geq 1$. Thus, by Claim 5, we get that the first term of Equation (7) is bounded by

$$2\delta\cdot 2(c_{\mathsf{CHRT}}\log n')^{4k_0}\leq 2^{-C\log n}\cdot 2^{8k_0\log\log\log n}\leq n^{-\frac{C}{2}},$$

taking into account the fact that $\mathbb{E}[L(U)]=0$.

For each i of the second term of Equation (7), we use sub-multiplicativity and the fact that H_i and $G^{>i}$ are on disjoint set of variables to get

$$\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}\left[(H_i\cdot G^{>i})(Y+Z\wedge U)\right]\right\|\right]\leq\mathop{\mathbb{E}}_{Y,Z}\left[\left\|\mathop{\mathbb{E}}_{U}[H_i(Y+Z\wedge U)]\right\|\cdot\left\|\mathop{\mathbb{E}}_{U}[G^{>i}(Y+Z\wedge U)]\right\|\right].$$

As $G^{>i}$ is a width-4 ROBP, $\left\|\mathbb{E}_U[G^{>i}(y+z\wedge U)]\right\|\leq 2$ for all $y\sim Y$ and $z\sim Z$. Continuing the above bound, by Cauchy-Schwarz we get

$$\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}(H_i\cdot G^{>i})[Y+Z\wedge U]\right\|\right]\leq 2\sqrt{\mathop{\mathbb{E}}_{Y,Z}\left[\left\|\mathop{\mathbb{E}}_{U}[H_i(Y+Z\wedge U)]\right\|^2\right]}.$$

Following [14, Lemma 7.1][8], using the bound by Chattopadhyay et al. [9] and Parseval's identity [14, Proposition 3.1], we get

$$\begin{aligned}\mathop{\mathbb{E}}_{Y,Z}\left[\left\|\mathop{\mathbb{E}}_{U}[H_i(Y+Z\wedge U)]\right\|^2\right]&\leq\left(2^{-Ck_0}+\gamma\right)\cdot\left(\delta\left(\sum_{\alpha\in\mathbb{F}_2^{n'}}\left\|(\widehat{H_i})_\alpha\right\|\right)^2+\sum_{\alpha\in\mathbb{F}_2^{n'}}\left\|(\widehat{H_i})_\alpha\right\|^2\right)\\&\leq\left(2^{-Ck_0}+\gamma\right)\cdot\left(\delta\cdot L_1^2\left(G^{\leq i}\right)+\mathbb{E}\left[\left\|G^{\leq i}(U)\right\|^2\right]\right)\\&\leq 8\cdot 2^{-Ck_0}.\end{aligned}$$

Overall, we get that

$$\mathop{\mathbb{E}}_{Z}\left[\left\|\mathop{\mathbb{E}}_{Y,U}[G(Y+Z\wedge U)]-\mathbb{E}[G]\right\|\right]\leq n^{-\frac{C}{2}}+2n'\sqrt{8\cdot 2^{-Ck_0}}\leq\frac{1}{16}\cdot 2^{-\frac{C}{4}k_0}=\frac{1}{16}\cdot 2^{-\frac{2C\log}{\log\log n}},$$

and we can choose the encoding G so that $f_{\mathcal{A}}(x)=G(x)_{1,1}$. Markov's inequality completes the proof. ◀

[8] Forbes and Kelley [14] take the bits of Z to have marginals $p=\frac{1}{2}$, but one can extend the lemma easily for the case of a general p.

4.4.2 Handling Well-Behaved Buckets

We will use our tail bounds for subset-wise symmetric polynomials to prove the following lemma.

▶ **Lemma 17.** *With probability at least $1-\frac{1}{2n}$ over $z \sim Z$, $f_{\mathcal{B}}$ can be written as $f_{\mathcal{B}} = f'_{\mathcal{B}} \oplus f''_{\mathcal{B}}$, where $f'_{\mathcal{B}}$ and $f''_{\mathcal{B}}$ are over disjoint set of variables, and for every $g \in \{f'_{\mathcal{B}}, f''_{\mathcal{B}}\}$ it holds that*

$$|\mathbb{E}\left[\widetilde{g}_z(Y)\right] - \mathbb{E}\left[\overline{g}\right]| \leq \frac{1}{n}.$$

The proof of Lemma 17 will follow immediately from Claim 20 and Lemma 21. Toward proving the above lemma, let us set some preliminaries.

▷ Claim 18. If $q \in \mathcal{B}$ then $c^q \in [C \log\log n, C \log n]$ and $m_q \leq 3^{c^q}$.

Proof. The upper bound on c^q follows immediately from the assumption in Lemma 15 that every term has width at most $C \log n$. Also, $m_q > \log^{2C} n$ since $q \notin \mathcal{A}$. Since we are at Case II, $m_q > \log^{2C} n$ implies that $m_q \leq 3^{c^q}$. From the fact that $\log^{2C} n < 3^{c^q}$ we get $c^q > \log_3(\log^{2C} n) > C \log\log n$. ◁

Recall that a term f_i is *good* under z if the variables read by f_i intersects with z in at least $1-\beta$ fraction.

▷ Claim 19. For a fixed $z \in \{0,1\}^n$, let $\mathbf{X}_z \subseteq [m]$ be the set of terms in $f_{\mathcal{B}}$ that are not good under z. Then, with probability at least $1-\frac{1}{2n}$ over $z \sim Z$,

$$|\mathbf{X}_z| \leq \frac{3c}{c-1} \log n.$$

Proof. The proof is very similar to Claim 13. Fix a bucket $q \in \mathcal{B}$, set $k_q = \frac{3 \log n}{c^q}$ and observe that $k \geq k_q$. Let $S \subseteq B_q$ be some subset of cardinally k_q. We first bound the probability p that every term f_i for $i \in S$ is bad under $z \sim Z$.

For a *truly random* $z \sim \text{Bernoulli}(1-2^{-C})^{\otimes n}$, the above probability is bounded by

$$\prod_{i \in S} \binom{w_i}{\beta w_i} 2^{-\beta C w_i} \leq \prod_{i \in S} 2^{w_i} 2^{-5 w_i} \leq \left(2^{-4 \cdot c^{q-1}}\right)^{k_q} \leq 2^{-3 k_q c^q} \leq n^{-9}.$$

For $z \sim Z$, we get that $p \leq n^{-9} + \gamma \leq 2n^{-9}$. Thus, with probability at most $\binom{m_q}{k_q} p$ over $z \sim Z$ there exists a set of k_q terms in B_q whose all terms are bad under z. By using Claim 18, we get

$$\binom{m_q}{k_q} p \leq m_q^{k_q} \cdot 2^{-9 \log n + 1} \leq 3^{k_q c^q + \log_3 2 \cdot (-9 \log n + 1)} \leq 3^{-\frac{9}{4} \log n} \leq n^{-3}.$$

Moreover, with probability at most $|\mathcal{B}| n^{-3} \leq n^{-2}$ over $z \sim Z$ there exists a $q \in \mathcal{B}$ and a set of k_q terms in B_q whose all terms are bad under z. Taking the contrapositive, we infer that with probability at least $1 - n^{-2} \geq 1 - \frac{1}{2n}$ over $z \sim Z$, we have at most

$$\sum_{q \in \mathcal{B}} k_q \leq \sum_{q=1}^{Q} \frac{3 \log n}{c^q} \leq \frac{3c}{c-1} \log n.$$

terms that are bad for z. ◁

From here onwards, we fix a z satisfying $|\mathbf{X}_z| \leq \frac{3c}{c-1}\log n$. Write

$$f_{\mathcal{B}} = \bigoplus_{i \in \mathbf{C} \setminus \mathbf{X}_z} f_i \oplus \bigoplus_{i \in \mathbf{X}_z} f_i \triangleq f'_{\mathcal{B}} \oplus f''_{\mathcal{B}},$$

where $\mathbf{C} = \bigcup_{q \in \mathcal{B}} B_q \subseteq [m]$ is the set of all terms that belong to $\mathcal{B}$'s buckets. Simply put, we divide $f_{\mathcal{B}}$ to the parity of exceptional terms $f''_{\mathcal{B}}$ and non-exceptional terms $f'_{\mathcal{B}}$ for whom we will refer to as *good terms*. We stress that both $f'_{\mathcal{B}}$ and $f''_{\mathcal{B}}$ depend on z.

▷ Claim 20 (Exceptional terms).

$$\left| \mathbb{E}\left[\left(\widetilde{f''_{\mathcal{B}}}\right)_z (Y) \right] - \mathbb{E}[\overline{f''_{\mathcal{B}}}] \right| \leq \frac{1}{n}.$$

Proof. For brevity, let $g = f''_{\mathcal{B}}$. For a fixed $w \in \{0,1\}^n$, let $g^w(x) = g(x+w)$. The proof will follow from bounding the spectral norm of $\overline{g^w}$. Indeed, $\overline{g^w}$ is a multiplication of at most $\frac{3c}{c-1}\log n$ terms, each of which has spectral norm at most 3. By sub-multiplicativity,

$$L_1\left(\overline{g^w}\right) \leq 3^{\frac{3c}{c-1}\log n} \leq n^{\frac{5c}{c-1}}.$$

Now, $\delta \leq \frac{1}{2} n^{-\frac{5c}{c-1}-1}$, so by Claim 5 we get that $|\mathbb{E}[\overline{g^w}(Y)] - \mathbb{E}[\overline{g^w}]| \leq \frac{1}{n}$ for every $w \in \{0,1\}^n$. Fooling $\overline{g^w}$ is sufficient to fool $\widetilde{g}_z$. To see this, note that

$$\begin{aligned} |\mathbb{E}\left[\widetilde{g}_z(Y)\right] - \mathbb{E}[\overline{g}]| &= |\mathbb{E}[\overline{g}(Y + z \wedge U)] - \mathbb{E}[\overline{g}(U + z \wedge U')]| \\ &= \left| \mathop{\mathbb{E}}_{w \sim U} \left[\mathbb{E}\left[\overline{g^w}(Y)\right] - \mathbb{E}\left[\overline{g^w}\right] \right] \right| \leq \frac{1}{n}, \end{aligned}$$

where U' is an independent copy of U. ◁

Next, we prove:

▶ **Lemma 21** (Good terms)**.**

$$\left| \mathbb{E}\left[\left(\widetilde{f'_{\mathcal{B}}}\right)_z (Y) \right] - \mathbb{E}[\overline{f'_{\mathcal{B}}}] \right| \leq \frac{1}{n}.$$

Proof. For brevity, let $g = f'_{\mathcal{B}}$ and recall that its set of terms is given by $\mathbf{C} \setminus \mathbf{X}_z$. Shifting the bias function $\widetilde{g} = \widetilde{g}_z$ to mean zero, recall that we define

$$\check{g}(x) = \frac{\widetilde{g}(x)}{\mathbb{E}[\widetilde{g}]} - 1.$$

Thus, we can write

$$\widetilde{g} = \mathbb{E}[g] \prod_{i \in \mathbf{C} \setminus \mathbf{X}_z} (1 + \check{g}_i) = \mathbb{E}[g] \sum_{I \subseteq \mathbf{C} \setminus \mathbf{X}_z} \prod_{i \in I} \check{g}_i = \mathbb{E}[g] \sum_{\vec{k} \in \mathbb{N}^Q} \sum_{I \subseteq \mathbf{C} \setminus \mathbf{X}_z, K(I) = \vec{k}} \prod_{i \in I} \check{g}_i,$$

where by $K(I) = \vec{k}$ we mean that for every $q \in [Q]$, there are $\vec{k}[q]$ terms in I that belong to the q-th bucket, i.e., $|I \cap B_q| = \vec{k}[q]$. For simplicity, we reorder the terms of g and write $g = \bigoplus_{i \in [m']} g_i$ for $m' = |\mathbf{C} \setminus \mathbf{X}_z|$, and for $q \in [Q]$, $B_q \subseteq [m']$ is the set of terms in g that belong to the q-th bucket. We abbreviate $\vec{g} = (\check{g_1}, \ldots, \widetilde{g_{m'}})$, and write

$$S_{\vec{k}}(\vec{g}) = \sum_{I \subseteq [m'], K(I) = \vec{k}} \prod_{i \in I} \check{g}_i.$$

Under these notations, $\widetilde{g} = \mathbb{E}[g] \sum_{\vec{k} \in \mathbb{N}^Q} S_{\vec{k}}(\vec{g})$.

Let $I_g(x)$ be the Boolean-valued function which is 1 if and only if

$$\sum_{\vec{k}\in\mathbb{N}^Q, \|\vec{k}\|_{(c)}>A} \left|S_{\vec{k}}\left(\vec{g}(x)\right)\right| \leq 2^{-\frac{A}{1024}},$$

where $A = \overline{C}\log n$ and $\|\vec{k}\|_{(c)} = \sum_{q=1}^{Q} c^q \cdot \vec{k}[q]$. Section 4.4.4 will be devoted to showing that $\mathbb{E}[I_g(Y)]$ is very close to 1. Namely,

▶ **Lemma 22.** *The following two inequalities hold.*

1. $\mathbb{E}[I_g(Y)] \geq 1 - e^{-c_I A}$ *for* $c_I = \frac{\ln 2}{2^{23}}$.
2. $\mathbb{E}\left[S^2_{\vec{k}}(\vec{g}(Y))\right] \leq 2^{-\frac{1}{8}\|\vec{k}\|_{(c)}}$.

For now, let us take Lemma 22 as given and continue with the proof of Lemma 21. We proceed by writing

$$|\mathbb{E}[\widetilde{g}(Y)] - \mathbb{E}[\widetilde{g}]| \leq |\mathbb{E}\left[\widetilde{g}(Y) \mid I_g(Y)=1\right] - \mathbb{E}[\widetilde{g}]| + 2\Pr\left[I_g(Y)=0\right]. \tag{8}$$

By Lemma 22, we have that $\Pr[I_g(Y)=0] \leq e^{-c_I A}$. Next, observe that

$$|\mathbb{E}\left[\widetilde{g}(Y) \mid I_g(Y)=1\right] - \mathbb{E}[\widetilde{g}]| = \left|\mathbb{E}[g] \sum_{\vec{k}\in\mathbb{N}^Q, \|\vec{k}\|_{(c)}>0} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \mid I_g(Y)=1\right]\right|,$$

and set

$$\boldsymbol{\Delta} = \left|\sum_{\vec{k}\in\mathbb{N}^Q, \|\vec{k}\|_{(c)}>0} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \mid I_g(Y)=1\right]\right|,$$

so Equation (8) gives us

$$|\mathbb{E}[\widetilde{g}(Y)] - \mathbb{E}[\widetilde{g}]| \leq \boldsymbol{\Delta} + 2e^{-c_I A}. \tag{9}$$

We bound $\boldsymbol{\Delta}$ as follows.

$$\boldsymbol{\Delta} \leq \left|\sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \mid I_g(Y)=1\right]\right| + \max_{y\in\{0,1\}^n, I_g(y)=1} \sum_{\vec{k}\in\mathbb{N}^Q, \|\vec{k}\|_{(c)}>A} \left|S_{\vec{k}}(\vec{g}(y))\right|.$$

By definition, the second term is at most $2^{-\frac{A}{1024}}$. The first term, call it $\boldsymbol{\Delta}_1$, can be split into two terms as follows.

$$\begin{aligned}
\boldsymbol{\Delta}_1 &= \frac{1}{\Pr[I_g(Y)=1]} \left|\sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \cdot I_g(Y)\right]\right| \\
&\leq 2\left|\sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \cdot I_g(Y)\right]\right| \qquad (10) \\
&\leq 2\left|\sum_{\substack{\vec{k}\in\mathbb{N}^Q, \\ 0<\|\vec{k}\|_{(c)}\leq A}} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y))\right]\right| + 2\left|\sum_{\substack{\vec{k}\in\mathbb{N}^Q, \\ 0<\|\vec{k}\|_{(c)}\leq A}} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y)) \cdot (1-I_g(Y))\right]\right|
\end{aligned}$$

$$\leq 2\left|\sum_{\substack{\vec{k}\in\mathbb{N}^Q,\\ 0<\|\vec{k}\|_{(c)}\leq A}} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y))\right]\right| + 2\sqrt{\mathbb{E}[1-I_g(Y)]}\cdot \sum_{\substack{\vec{k}\in\mathbb{N}^Q,\\ 0<\|\vec{k}\|_{(c)}\leq A}} \sqrt{\mathbb{E}\left[S^2_{\vec{k}}(\vec{g}(Y))\right]}, \tag{11}$$

where the last inequality follows from the triangle inequality followed by Cauchy-Schwarz. By Lemma 22, the second term of Equation (11), $\mathbf{\Delta}_{1,2}$, is at most

$$\begin{aligned}
\mathbf{\Delta}_{1,2} &\leq 2\cdot e^{-c_I A}\cdot \sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \sqrt{2^{-\frac{1}{8}\|\vec{k}\|_{(c)}}} \\
&\leq 2\cdot e^{-c_I A}\cdot \sum_{w=1}^{A-1} \left|\left\{\vec{k}\in\mathbb{N}^Q : w<\|\vec{k}\|_{(c)}\leq w+1\right\}\right| 2^{-\frac{1}{\sqrt{8}}w} \\
&\leq 2\cdot e^{-c_I A}(A+1)^Q \sum_{w=1}^{A} 2^{-\frac{1}{\sqrt{8}}w} \leq 8(A+1)^Q e^{-c_I A} \leq 2^{\frac{2}{\log c}(\log\log n)^2} e^{-c_I A} \leq \frac{1}{8n}.
\end{aligned}$$

To finish bounding $\mathbf{\Delta}_1$, it is left to bound the first term of Equation (11), denoted by $\mathbf{\Delta}_{1,1}$.

▷ Claim 23. $\mathbf{\Delta}_{1,1} = 2\left|\sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y))\right]\right| \leq \frac{1}{8n}$.

Proof. The proof goes by bounding the spectral norm of the function $S_{\vec{k}}(\vec{g}(x))$. As for every $\vec{k}\in\mathbb{N}^Q$ with $\|\vec{k}\|_{(c)}\neq 0$, $\mathbb{E}[S_{\vec{k}}(\vec{g}(U))]=0$, the claim will follow by using Claim 5, together with sub-additivity and sub-multiplicativity. First, note that:

▷ Claim 24. For every $i\in[m]$, $L_1(\breve{g}_i)\leq 4$.

Proof. Consider the function $h_i = 1-g_i$, so $L_1(\widetilde{g}_i)\leq L_1(\widetilde{h}_i)+1$ and $\mathbb{E}[\widetilde{h}_i]=\mathbb{E}[h_i]=1-\mathbb{E}[g_i]\geq\frac{1}{2}$. Now, $L_1(\breve{h}_i)\leq\frac{1}{\mathbb{E}[h_i]}L_1(\widetilde{h}_i)+1\leq 2L_1(\widetilde{h}_i)+1$. Recalling that $\widetilde{h}_i(x)=\mathbb{E}[h_i(x+z\wedge U)]$, we get $L_1(\widetilde{h}_i)\leq 1$ as every shift of h_i is a negated conjunction of literals. Thus, $L_1(\breve{h}_i)\leq 3$ and $L_1(\breve{g}_i)\leq 4$. ◁

Then, for every such $\vec{k}\in\mathbb{N}^Q$,

$$\begin{aligned}
L_1\left(S_{\vec{k}}(\vec{g})\right) &\leq \sum_{I\subseteq\mathbf{C}\setminus\mathbf{X}_T, K(I)=\vec{k}} \prod_{i\in I} L_1\left(\breve{g}_i\right) \leq \sum_{I\subseteq\mathbf{C}\setminus\mathbf{X}_T, K(I)=\vec{k}} 4^{|I|} \\
&= \sum_{I\subseteq\mathbf{C}\setminus\mathbf{X}_T, K(I)=\vec{k}} \prod_{q\in[Q]} 4^{\vec{k}[q]} = \prod_{q\in[Q]} \binom{\vec{k}[q]}{m_q} 4^{\vec{k}[q]}.
\end{aligned}$$

Recall that Claim 18 tells us that $m_q\leq 3^{c^q}$, so

$$L_1\left(S_{\vec{k}}(\vec{g})\right) \leq \prod_{q\in[Q]} 3^{c^q(1+\log_3 4)\vec{k}[q]} \leq 12^{\|\vec{k}\|_{(c)}}. \tag{12}$$

Finally,

$$L_1\left(\sum_{\vec{k}\in\mathbb{N}^Q, 0<\|\vec{k}\|_{(c)}\leq A} \mathbb{E}\left[S_{\vec{k}}(\vec{g}(Y))\right]\right) \leq (A+1)^Q\cdot 12^A \leq 2^{6A} \leq n^{6\overline{C}},$$

and the claim follows by observing that $\delta\leq\frac{1}{32n}n^{-6\overline{C}}$. ◁

Incorporating the above claim, we get that $\mathbf{\Delta}_1 = \mathbf{\Delta}_{1,1} + \mathbf{\Delta}_{1,2} \leq \frac{1}{8n} + \frac{1}{8n} \leq \frac{1}{4n}$, which readily gives $\mathbf{\Delta} \leq \frac{1}{4n} + 2^{-\frac{A}{1024}} \leq \frac{1}{2n}$. Plugging-it in Equation (9), we finally get

$$|\mathbb{E}[\widetilde{g}(Y)] - \mathbb{E}[\widetilde{g}]| \leq \frac{1}{2n} + 2e^{-c_I A} \leq \frac{1}{n}$$

and the desired result. ◀

4.4.3 Putting It Together

Here we finally incorporate Lemma 16 and Lemma 17.

Proof of Lemma 15. By Lemma 16 and Lemma 17, with probability at least $1 - \frac{1}{4} \cdot 2^{-C\frac{\log n}{\log\log n}} - \frac{1}{n} \geq 1 - \frac{1}{2} \cdot 2^{-C\frac{\log n}{\log\log n}}$ over $z \sim Z$, we can write

$$f = f_{\mathcal{A}} \oplus f'_{\mathcal{B}} \oplus f''_{\mathcal{B}},$$

where the three functions are over disjoint set of variables, and it holds that for each $\mathcal{T} \in \{\mathcal{A}, \mathcal{B}, \mathcal{B}'\}$,

$$\left|\left(\widetilde{f_{\mathcal{T}}}\right)_z (Y') - \mathbb{E}\left[\overline{f_{\mathcal{T}}}\right]\right| \leq \frac{1}{4} \cdot 2^{-C\frac{\log n}{\log\log n}}$$

for any δ-biased distribution Y'. Using the XOR lemma for small-biased spaces (see Lemma 6), taking into account that our distribution Y is in fact δ^3-biased, we conclude that

$$\left|\mathbb{E}[\widetilde{f}_z(Y)] - \mathbb{E}[\overline{f}]\right| \leq 16^3 \cdot 2 \cdot \frac{1}{4} \cdot 2^{-C\frac{\log n}{\log\log n}} \leq \frac{1}{2} \cdot 2^{-C\frac{\log n}{\log\log n}},$$

and the lemma follows. ◀

4.4.4 I_g Almost Always Happens

We keep using the notations of Section 4.4.2. Specifically, recall that $g = f'_{\mathcal{B}} = \bigoplus_{i\in[m']} g_i$ for $m' = |\mathbf{C} \setminus \mathbf{X}_z|$, and for $q \in [Q]$, $B_q \subseteq [m']$ is the set of terms in g that belong to the q-th bucket. Also, for $\vec{g} = (\widetilde{g_1}, \ldots, \widetilde{g_{m'}})$,

$$S_{\vec{k}}(\vec{g}) = \sum_{I\subseteq[m'], K(I)=\vec{k}} \prod_{i\in I} \widetilde{g}_i.$$

Recall that $I_g(x) \in \{0,1\}$ is 1 if and only if

$$\sum_{\vec{k}\in\mathbb{N}^Q, \|\vec{k}\|_{(c)}>A} \left|S_{\vec{k}}\left(\vec{g}(x)\right)\right| \leq 2^{-\frac{A}{1024}},$$

where $A = \overline{C}\log n$ and $\|\vec{k}\|_{(c)} = \sum_{q\in[Q]} c^q \cdot \vec{k}[q]$.

Proof of Lemma 22. As in Section 2, we define

$$R_{\vec{k}}(\vec{g}) = S^2_{\vec{k}}(\vec{g}) \cdot \prod_{q\in[Q]} \vec{k}[q]!.$$

By Lemma 3, to prove the bound on $\Pr[I_g(Y) = 0]$ it is sufficient to prove that for every $\vec{k} \in \mathbb{N}^Q$ with $\|\vec{k}\|_{(c)} \leq A$ we have that

$$\mathbb{E}\left[R_{\vec{k}}(\vec{g}(Y))\right] \leq 2^{-\frac{1}{8}\|\vec{k}\|_{(c)}}.$$

By now a standard course of action, we aim at bounding the spectral norm of the function $R_{\vec{k}}(\vec{g})$, together with its expectation under the uniform distribution. To this end, let us define, for $q \in [Q]$ and an integer ℓ,

$$\check{S}_{\ell,q} = \sum_{I \subseteq B_q, |I|=\ell} \prod_{i \in I} \breve{g}_i,$$

so $R_{\vec{k}}(\vec{g}) = \prod_{q \in [Q]} \check{S}^2_{\vec{k}[q],q} \vec{k}[q]!$. First, we record that:

▷ Claim 25. For every $i \in [m']$, $\mathbb{E}\left[\breve{g}_i^{\,2}\right] \leq 2^{-(2-2\beta)w_i}$.

Proof. Let $V_i \subseteq [n]$ be the set of variables read by g_i, of cardinality w_i, and let $\ell_i = |V_i \cap \{j \in [n] : z_j = 1\}|$ be the number of *live* variables read by g_i. Note that

$$\widetilde{g}_i(x) = \mathbb{E}[g_i(x + z \wedge U)] = \begin{cases} 0 & \text{if there exists } j \in V_i \text{ such that } x_j = z_j = 0, \\ 2^{-\ell_i} & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} \mathbb{E}\left[\widetilde{g}_i^{\,2}\right] &= 2^{-2\ell_i} \Pr_{x \sim U}[\text{for every} j \in V_i \text{ s.t. } z_j = 0 \text{ it holds that } x_j = 1] \\ &= 2^{-2\ell_i} 2^{-(w_i - \ell_i)} = 2^{-w_i - \ell_i}. \end{aligned}$$

Recalling that $\ell_i \geq (1-\beta)w_i$ (g_i is good under z), we have $\mathbb{E}[\widetilde{g}_i^{\,2}] \leq 2^{-(2-\beta)w_i}$. Let $h_i = 1 - g_i$, and note that

$$\mathbb{E}\left[\breve{h}_i^{\,2}\right] = \mathrm{Var}\left[\breve{h}_i\right] = \frac{\mathrm{Var}[\widetilde{g}_i]}{\mathbb{E}^2[h_i]} \leq 4 \cdot \mathbb{E}\left[\widetilde{g}_i^{\,2}\right] \leq 4 \cdot 2^{-(2-\beta)w_i} \leq 2^{-(2-2\beta)w_i}.$$

The fact that $\mathrm{Var}[\breve{h}_i] = \mathrm{Var}[\breve{g}_i] = \mathbb{E}[\breve{g}_i^{\,2}]$ finishes the proof. ◁

Now,

$$\begin{aligned} \mathbb{E}\left[\check{S}^2_{\ell,q}\right] &= \sum_{I \subseteq B_q, |I|=\ell} \prod_{i \in I} \mathbb{E}\left[\breve{g}_i^{\,2}\right] \leq \sum_{I \subseteq B_q, |I|=\ell} \prod_{i \in I} 2^{-(2-2\beta)w_i} \\ &\leq \sum_{I \subseteq B_q, |I|=\ell} 2^{-(2-2\beta)c^q \ell} \leq \binom{m_q}{\ell} 2^{-(2-2\beta)c^q \ell} \leq \frac{3^{c^q \ell} e^\ell 2^{-(2-2\beta)c^q \ell}}{\ell!} \leq \frac{1}{\ell!} 2^{-\frac{c^q \ell}{4}}. \end{aligned}$$

Plugging it in our expression for $R_{\vec{k}}$, we get

$$\mathbb{E}\left[R_{\vec{k}}(\vec{g})\right] = \prod_{q \in [Q]} \mathbb{E}\left[\check{S}_{\vec{k}[q],q} \vec{k}[q]!\right] \leq \prod_{q \in [Q]} 2^{-\frac{c^q \vec{k}[q]}{4}} = 2^{-\frac{1}{4}\|\vec{k}\|_{(c)}}. \tag{13}$$

Finally, let us bound $L_1(R_{\vec{k}}(\vec{g}))$. In Equation (12) we established the fact that $L_1(S_{\vec{k}}(\vec{g})) \leq 12^{\|\vec{k}\|_{(c)}} \leq 12^A$. Thus,

$$L_1\left(R_{\vec{k}}(\vec{g})\right) \leq 12^{2A} \prod_{q \in [Q]} \vec{k}[q]! \leq 12^{2A} e^{\sum_{q \in [Q]} \vec{k}[q] \ln \vec{k}[q]} \leq 12^{2A} e^{(\ln A) \sum_{q \in [Q]} \vec{k}[q]}.$$

As $\|\vec{k}\|_{(c)} = \sum_{q \in [Q]} c^q \vec{k}[q] \leq A$ and $c^q \geq C \log\log n$ (see Claim 18), $\sum_{q \in [Q]} \vec{k}[q] \leq \frac{A}{C \log\log n}$ and we get

$$L_1\left(R_{\vec{k}}(\vec{g})\right) \leq 12^{2A} e^{\ln A \frac{A}{C \log\log n}} \leq 12^{2A} 2^{\frac{\overline{C}}{C} \log n} \leq n^{10\overline{C}}.$$

Note that $\delta \leq \frac{1}{32} n^{-10\overline{C}} 2^{-\frac{A}{4}}$. Thus, by Claim 5,

$$\mathbb{E}\left[R_{\vec{k}}(\vec{g}(Y))\right] \leq 2^{-\frac{1}{4}\|\vec{k}\|_{(c)}} + \delta \cdot n^{10\overline{C}} \leq 2^{-\frac{1}{8}\|\vec{k}\|_{(c)}},$$

and we are done with bounding $\Pr[I_g(Y) = 0]$. For the bound on $\mathbb{E}\left[S^2_{\vec{k}}(\vec{g}(Y))\right]$, simply observe that $\mathbb{E}\left[S^2_{\vec{k}}(\vec{g}(Y))\right] < \mathbb{E}\left[R_{\vec{k}}(\vec{g}(Y))\right]$. ◀

5 Full PRG via Iterated Restrictions

So far, we have shown how to pseudorandomly assign values to a *constant fraction* of the inputs of any read-once PARITY ∘ AND formula using $O(\log n)$ truly random bits, preserving the expectation of the formula to within near-optimal error. In this section, to complete the proof of Theorem 1, we show how to pseudorandomly assign values to *all* the inputs, i.e., we give a genuine PRG.

For convenience, we make the following definitions.

▶ **Definition 26.** *Let $w > 0$. A* w-proper formula *is a read-once* PARITY ∘ AND *formula of width at most w and length most 2^{8w}. We say that such a formula is* short *if its length is at most 2^{4w}; otherwise, we say that the formula is* long.

Our main goal is to fool $(C \log n)$-proper formulas, but along the way, we will obtain a PRG for w-proper formulas with seed length $O(w)$ and error $\exp(-\widetilde{\Omega}(w))$, even for w substantially smaller than $\log n$.

5.1 Restrictions for Proper Formulas

Recall that Lemma 9 provides a pseudorandom restriction that uses only $O(\log n)$ truly random bits. We now generalize this fact in two respects. First, in the case of w-proper formulas ($\log \log n \leq w \leq C \log n$), we improve the seed length to $O(w)$. Second, in the case of *short* w-proper formulas, we argue that the restriction *simplifies* the formula, in the sense that it transforms it into a $(w/2)$-proper formula.

▶ **Lemma 27.** *For every $w, n \in \mathbb{N}$ with $w \leq C \log n$, there is a distribution X over $\{0,1,\star\}^n$ with the following properties.*

1. *(Seed length) There is an explicit algorithm to sample from X using just $O(w + \log \log n)$ truly random bits.*
2. *(Expectation preservation) If f is a w-proper formula, then X preserves the expectation of f with error $\exp(-\Omega(w/\log w))$.*
3. *(Simplification) If f is a* short *w-proper formula, then*

$$\Pr[f|_X \text{ is a } (w/2)\text{-proper formula}] \geq 1 - 2^{-w}.$$

Proof. Let $n' = 2^{8w} \cdot w$. Let Y be an n'-wise δ^3-biased distribution where $\delta = (n')^{-12\overline{C}}$, and let Z be γ-almost k-wise independent with marginals $1 - 2^{-C}$, where $k = 6 \log n'$ and $\gamma = (n')^{-9}$. Our restriction is

$$X = \mathrm{Res}(Y, Z)^{\circ 2^{C+4}}.$$

By standard constructions [25, 3] and Claim 7, X can be explicitly sampled using $O(w + \log \log n)$ truly random bits.

Now, to prove expectation preservation, let f be a w-proper formula. By w-properness, there is some set of indices $I \subseteq [n]$, $|I| \leq n'$, such that $f(x)$ only depends on $x|_I$. Let $g\colon \{0,1\}^{|I|} \to \{0,1\}$ be the w-proper formula such that $f(x) = g(x|_I)$. Since $Y|_I$ is δ^3-biased and $Z|_I$ is γ-almost k-wise independent with marginals $1 - 2^{-C}$, Lemma 9 implies that $\mathrm{Res}(Y|_I, Z|_I)$ preserves the expectation of g with error $\exp(-\Omega(\frac{\log n'}{\log\log n'}))$, which is $\exp(-\Omega(w/\log w))$. It follows that $\mathrm{Res}(Y, Z)$ preserves the expectation of f with the same error. The error of X is only larger by a constant factor 2^{C+4}, because any restriction of a w-proper formula is trivially another w-proper formula.

Finally, to prove simplification, let f be a *short* w-proper formula, and let f_i be a term. Since $k > w/2$, by Claim 8, the probability that more than $w/2$ variables from f_i are assigned $\star$ by X is bounded by

$$\binom{w}{w/2} \cdot \left((1-2^{-C})^{2^{C+4}\cdot w/2} + 2^{C+4}\gamma\right) \leq 2^w \cdot \left(e^{-2^{-C}\cdot 2^{C+3}w} + 2^{C+4}\cdot (n')^{-9}\right) < 2^{-5w}.$$

The number of terms in f is at most 2^{4w}, so by the union bound, except with probability 2^{-w}, $f|_X$ has maximum width at most $w/2$. Furthermore, restricting cannot *increase* the number of terms, so the number of terms is still bounded by $2^{4w} = 2^{8(w/2)}$. Therefore, in this case, $f|_X$ is $(w/2)$-proper. ◀

5.2 Full PRGs for Long Proper Formulas [24]

The simplification clause of Lemma 27 only applies if f is *short*. If f is *long*, we will therefore need a different approach. We will take a similar approach as Meka, Reingold, and Tal [24]. A *full PRG* for long w-proper formulas follows readily from their work.

▶ **Lemma 28** ([24]). *For every $w, n \in \mathbb{N}$, there is an explicit (2^{-w})-PRG for* long *w-proper formulas with seed length*

$$O(w + \log\log n).$$

Proof sketch. In short, the PRG is one of the PRGs by Meka et al. [24, full version, Lemma 6.2], except we replace every δ-biased distribution with a $(\cdot)$-wise δ-biased distribution to optimize the seed length.

In more detail, let $n' = 2^{8w}\cdot w$. Sample $v \in \{0,1\}^{wn}$ from an $(n'w)$-wise (c_{MRT}^{-w})-biased distribution, where c_{MRT} is a suitable constant. Think of v as n blocks of w bits. Define a set $I \subseteq [n]$ as follows: include i in I if and only if the i-th block of v is 1^w.

Sample $x^{(0)}, x^{(1)}, \ldots, x^{(16)} \in \{0,1\}^n$ independently from an (n')-wise (c_{MRT}^{-w})-biased distribution. The PRG outputs the string x defined by

$$x_i = \begin{cases} x_i^{(0)} & \text{if } i \notin I \\ \bigoplus_{j=1}^{16} x_i^{(j)} & \text{if } i \in I. \end{cases}$$

By standard constructions [25, 3], the seed length of this PRG is

$$O(\log n' + w + \log\log n) = O(w + \log\log n).$$

As for correctness, let f be a long w-proper formula. Let $J \subseteq [n]$ be the set of indices of variables that f reads, so there is some long w-proper formula g on $|J|$ input bits such that $f(x) = g(x|_J)$. Let X be the distribution output by the PRG. Since $|J| \leq n'$, the

distribution $X|_J$ is exactly the pseudorandom distribution designed by Meka et al. [24, full version, Lemma 6.2]. Furthermore, since f is long, $|J| > 2^{4w}$. It follows that g is in the class of functions fooled by Meka et al.'s pseudorandom distribution: g is an XOR of m non-constant Boolean functions on disjoint variables, where each function is on at most w variables, with $16^w < m \le 16^{2w}$ and $\log\log(|J|/2^w) \ll w \le \log|J|$. Therefore, $X|_J$ fools g with error 2^{-w}, and hence X fools f with error 2^{-w}. ◀

5.3 Full PRGs for Width-$O(\log n)$ Formulas

For *short* proper w-formulas, to get a full PRG, we will iterate the restriction of Lemma 27 several times, assigning values to more and more variables. Eventually, we'll stop this recursive process and use a different PRG. Specifically, for the "base case," we'll use a PRG by Lee [22] with minor modifications:

▶ **Lemma 29** ([22]). *For every $w, n \in \mathbb{N}$ and every $\varepsilon > 0$, there is an explicit ε-PRG for w-proper formulas with seed length*

$$O((w + \log(1/\varepsilon)) \cdot (\log w + \log\log(1/\varepsilon))^2) + \mathrm{poly}(\log\log(n/\varepsilon)).$$

Proof sketch. In short, the PRG is one of the PRGs by Lee [22, Theorem 6], except we replace every δ-biased distribution with a $(\cdot)$-wise δ-biased distribution to optimize the seed length, just like the proofs of Lemma 27 and Lemma 28.

To give a little more detail, let $n' = 2^{8w} \cdot w$; a w-proper formula only reads n' variables. Lee's PRG [22, Theorem 6] is designed to fool *arbitrary-order combinatorial checkerboards*, i.e., parities of functions on disjoint variable sets of size at most w. This class includes w-proper formulas as a special case. Lee's original PRG has seed length

$$O((w + \log(n/\varepsilon)) \cdot (\log w + \log\log(n/\varepsilon))^2).$$

After making suitable replacements, one can show that the seed length is reduced to

$$O((w + \log(n'/\varepsilon)) \cdot (\log w + \log\log(n'/\varepsilon))^2) + \mathrm{poly}(\log\log(n/\varepsilon)).$$

(We omit the full proof, since it repeats much of Lee's analysis [22].) Plugging in the value of n', we get the claimed seed length. ◀

We now give our full PRG for general formulas of width at most $C \log n$. The PRG follows a similar approach to one of the PRGs by Meka et al. [24, full version, Algorithm 3]: iteratively apply the restriction of Lemma 27, but at each step, XOR with the PRG of Lemma 28 in case the formula is long.

▶ **Lemma 30.** *For every $n \in \mathbb{N}$, there is an explicit PRG for read-once* PARITY $\circ$ AND *formulas of width at most $C \log n$ with seed length $O(\log n)$ and error*

$$2^{-\Omega\left(\frac{\log n}{(\log\log n)^3}\right)}.$$

Proof. Define

$$w_0 = \frac{\log n}{(\log\log n)^2}.$$

We recursively define a PRG G_w for w-proper formulas, $w_0 \le w \le C \log n$, as follows.

- (Base case) If $w \leq 2w_0$, then G_w is the (2^{-w_0})-PRG of Lemma 29 based on Lee's work [22].
- (Recursive case) If $w > 2w_0$, sample $X \in \{0,1,\star\}^n$ from the distribution guaranteed by Lemma 27 based on the work in Section 4. Sample $Y \in \{0,1\}^n$ using the PRG of Lemma 28 based on Meka et al.'s work [24]. Recursively sample $G_{\lceil w/2 \rceil}$, and set

$$G_w = Y \oplus (X \circ G_{\lceil w/2 \rceil}).$$

For the analysis, observe first that in the base case $w \leq 2w_0$, G_w fools w-proper formulas with error 2^{-w_0}. Now, for the inductive step, consider some $w > 2w_0$. Assume $G_{\lceil w/2 \rceil}$ fools $\lceil w/2 \rceil$-proper formulas with error $\varepsilon_{\lceil w/2 \rceil}$; we will show that G_w fools w-proper formulas with error ε_w, where

$$\varepsilon_w = \varepsilon_{\lceil w/2 \rceil} + 2^{-\Omega(w/\log w)}.$$

Let f be a w-proper formula, and for brevity, let $G = G_{\lceil w/2 \rceil}$. For the first case, suppose f is long. Any shift of f is also a long w-proper formula, so

$$\begin{aligned}
|\mathbb{E}[f(G_w)] - \mathbb{E}[f]| &= \left| \mathop{\mathbb{E}}_{X,G} \left[\mathop{\mathbb{E}}_{Y} [f(Y \oplus (X \circ G))] \right] - \mathbb{E}[f] \right| \\
&\leq \mathop{\mathbb{E}}_{X,G} \left[\left| \mathop{\mathbb{E}}_{Y} [f(Y \oplus (X \circ G))] - \mathbb{E}[f] \right| \right] \\
&= \mathop{\mathbb{E}}_{X,G} \left[\left| \mathop{\mathbb{E}}_{Y} [f(Y \oplus (X \circ G))] - \mathop{\mathbb{E}}_{U} [f(U \oplus (X \circ G))] \right| \right] \\
&\leq 2^{-w}.
\end{aligned}$$

For the second case, suppose f is short. For each $y \in \{0,1\}^n$, define $f_y(x) = f(y \oplus x)$, another short w-proper formula. Fix $y \sim Y$, and let E be the event that $f_y|_X$ is $(w/2)$-proper, so whether E occurs depends only on X. Then

$$\begin{aligned}
&|\mathbb{E}[(f_y|_X)(G)] - \mathbb{E}[f]| \\
&\quad\leq \left| \mathop{\mathbb{E}}_{X} \left[\mathop{\mathbb{E}}_{G} [(f_y|_X)(G)] \;\middle|\; E \right] - \mathbb{E}[f] \right| + \Pr[\neg E] \\
&\quad\leq \left| \mathop{\mathbb{E}}_{X} \left[\mathop{\mathbb{E}}_{U} [(f_y|_X)(U)] \;\middle|\; E \right] - \mathbb{E}[f] \right| + \varepsilon_{\lceil w/2 \rceil} + \Pr[\neg E] && \text{(Induction)} \\
&\quad\leq \left| \mathop{\mathbb{E}}_{X} \left[\mathop{\mathbb{E}}_{U} [(f_y|_X)(U)] \right] - \mathbb{E}[f] \right| + \varepsilon_{\lceil w/2 \rceil} + 2\Pr[\neg E] \\
&\quad\leq 2^{-\Omega(w/\log w)} + \varepsilon_{\lceil w/2 \rceil} + 2\Pr[\neg E] && \text{(Item 2 of Lemma 27)} \\
&\quad\leq 2^{-\Omega(w/\log w)} + \varepsilon_{\lceil w/2 \rceil} + 2 \cdot 2^{-w} && \text{(Item 3 of Lemma 27).}
\end{aligned}$$

Let ε_w be the final right-hand side, so indeed $\varepsilon_w = \varepsilon_{\lceil w/2 \rceil} + \exp(-\Omega(w/\log w))$. Then

$$\begin{aligned}
|\mathbb{E}[f(G_w)] - \mathbb{E}[f]| &\leq \mathop{\mathbb{E}}_{Y} \left[\left| \mathop{\mathbb{E}}_{X,G} [(f_Y|_X)(G)] - \mathbb{E}[f] \right| \right] \\
&\leq \varepsilon_w.
\end{aligned}$$

Now, let us add up all these errors. Since $w \geq w_0$ always holds, we have $\varepsilon_w \leq \varepsilon_{\lceil w/2 \rceil} + \exp(-\Omega(w_0/\log w_0))$. Starting at $w = C \log n$, we only need to halve w a total of $O(\log\log\log n)$ times to reach the base case $w \leq 2w_0$. Therefore, the total error of $G_{C \log n}$ is bounded by

$$2^{-w_0} + 2^{-\Omega(w_0/\log w_0)} \cdot O(\log\log\log n) = 2^{-\Omega\left(\frac{\log n}{(\log\log n)^3}\right)}.$$

Finally, let us bound the seed length of G_w. In the base case $w \le 2w_0$, by our choice of w_0, the seed length s_w of G_w is bounded by some value $s_{\mathsf{base}} \le O(\log n)$. In the recursive case $w > 2w_0$, the seed length s_w of G_w is bounded by

$$s_w = s_{\lceil w/2 \rceil} + O(w + \log\log n) = s_{\lceil w/2 \rceil} + O(w).$$

The point is that this is essentially a geometric series. More precisely, let c_{seed} be a constant such that $s_w \le s_{\lceil w/2 \rceil} + c_{\mathsf{seed}} \cdot w$ for all $w > 2w_0$. Then by induction, for all $w \ge w_0$, we have

$$s_w \le s_{\mathsf{base}} + 3c_{\mathsf{seed}} w,$$

because

$$\begin{aligned} s_w &\le s_{\lceil w/2 \rceil} + c_{\mathsf{seed}} w \\ &\le s_{\mathsf{base}} + 3c_{\mathsf{seed}} \lceil w/2 \rceil + c_{\mathsf{seed}} w && \text{(Induction)} \\ &< s_{\mathsf{base}} + 3c_{\mathsf{seed}} w. \end{aligned}$$

Therefore, we can take the desired PRG to be $G_{C \log n}$, because $s_{C \log n} \le O(\log n)$, and any read-once PARITY $\circ$ AND formula of width at most $C \log n$ is $(C \log n)$-proper. ◀

5.4 Arbitrary-Error PRGs for Width-$O(\log(n/\varepsilon))$ Formulas

At this point, the main work of proving Theorem 1 is complete. We just need to address three minor issues: small ε, large width, and formulas not of the form PARITY $\circ$ AND. We begin by addressing the case of small ε. Recall that we wish to achieve seed length $O(\log n) + \widetilde{O}(\log(1/\varepsilon))$ for an *arbitrary* error ε. This follows readily by combining the PRG of Lemma 30 with Lee's PRG (Lemma 29).

▶ **Lemma 31.** *For any $n \in \mathbb{N}, \varepsilon > 0$, there is an explict ε-PRG for read-once* PARITY $\circ$ AND *formulas of width at most $\frac{C}{2}\log(n/\varepsilon)$ with seed length*

$$O(\log n + \log(1/\varepsilon) \cdot (\log\log(1/\varepsilon))^5).$$

Proof. Let ε_0 be the error parameter in Lemma 30, so $\varepsilon_0 = \exp(-\Omega(\frac{\log n}{(\log\log n)^3}))$. If $\varepsilon \ge \varepsilon_0$, the PRG of Lemma 30 works, because $\frac{C}{2}\log(n/\varepsilon) < C \log n$. If $\varepsilon < \varepsilon_0$, use Lee's PRG [22], i.e., the ε-PRG of Lemma 29 for $(\frac{C}{2}\log(n/\varepsilon))$-proper formulas, which has seed length

$$O(\log(n/\varepsilon) \cdot (\log\log(n/\varepsilon))^2) \le O(\log(1/\varepsilon) \cdot (\log\log(1/\varepsilon))^5).$$ ◀

(In the proof of Lemma 31, we could just as well have used Lee's original PRG [22, Theorem 6] instead of the slightly modified version given by Lemma 29.)

5.5 PRGs for Any Width

In this section, we eliminate the assumption that the maximum width is bounded.

▶ **Lemma 32.** *For all $n \in \mathbb{N}$ and $\varepsilon > 0$, there is an explicit ε-PRG for read-once* PARITY $\circ$ AND *formulas on n input bits with seed length*

$$O\left(\log n + \log(1/\varepsilon) \cdot (\log\log(1/\varepsilon))^5\right).$$

Proof. Sample G from the $(\varepsilon/3)$-PRG for formulas of width $\frac{C}{2}\log(3n/\varepsilon)$ guaranteed by Lemma 31. Sample Y from an $(\frac{\varepsilon}{6n})$-biased distribution. Our final PRG outputs

$$H \stackrel{\text{def}}{=} G \oplus Y.$$

To prove that this works, let f be a read-once PARITY $\circ$ AND formula. Write $f = f' \oplus f''$, where every term in f' has width at most $\frac{C}{2}\log(3n/\varepsilon)$ and every term in f'' has width greater than $\frac{C}{2}\log(3n/\varepsilon)$.

Since any shift of a width-w read-once PARITY $\circ$ AND formula is another width-w read-once PARITY $\circ$ AND formula, H fools f' with error $\varepsilon/3$. Meanwhile, since each term f''_i of f'' is a conjunction of more than $\frac{C}{2}\log(3n/\varepsilon)$ literals,

$$\mathbb{E}[f''_i] \leq \left(\frac{\varepsilon}{3n}\right)^{C/2} < \frac{\varepsilon}{6n}.$$

Furthermore, the L_1 norm of any conjunction of literals is 1, and H is $(\frac{\varepsilon}{6n})$-biased, so by Claim 5, $\mathbb{E}[f''_i(H)] < \frac{\varepsilon}{3n}$. Therefore, by the union bound, for either distribution $X \in \{H, U\}$,

$$\mathbb{E}[f''(X)] < \varepsilon/3.$$

This allows us to bound the error of the final PRG as follows:

$$\begin{aligned}
|\mathbb{E}[f(H)] - \mathbb{E}[f]| &\leq |\mathbb{E}[f(H)] - \mathbb{E}[f'(H)]| + |\mathbb{E}[f'(H)] - \mathbb{E}[f']| + |\mathbb{E}[f'] - \mathbb{E}[f]| \\
&\leq \mathbb{E}[|f(H) - f'(H)|] + |\mathbb{E}[f'(H)] - \mathbb{E}[f']| + \mathbb{E}[|f' - f|] \\
&= \mathbb{E}[f''(H)] + |\mathbb{E}[f'(H)] - \mathbb{E}[f']| + \mathbb{E}[f''] \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\end{aligned}$$

◀

5.6 Proof of Theorem 1

In this section, we finally complete the proof of Theorem 1 by showing that fooling read-once PARITY $\circ$ AND formulas is sufficient for fooling read-once depth-2 $\mathbf{AC}^0[\oplus]$:

▶ **Lemma 33.** *Let X be a distribution over $\{0,1\}^n$, and let $\varepsilon > 0$. If X fools all read-once* PARITY $\circ$ AND *formulas with error ε, then X fools all read-once depth-2 $\mathbf{AC}^0[\oplus]$ formulas with error 2ε.*

Proof. Let f be a read-once depth-2 $\mathbf{AC}^0[\oplus]$ formula.

For the first case, suppose the output gate of f is $\oplus$. By merging the output gate with any $\oplus$ children and introducing trivial $\wedge$ gates with fan-in 1 as necessary, we see that without loss of generality, every child of the output gate is either $\wedge$ or $\vee$. By de Morgan's laws, it follows that either f or $\neg f$ can be computed by a read-once PARITY $\circ$ AND formula. Either way, this implies that X ε-fools f.

For the second case, suppose the output gate of f is $\wedge$, say $f = \bigwedge_{i=1}^m f_i$. Using the Fourier expansion of the m-input AND function, we get

$$\begin{aligned}
f &= \sum_{I \subseteq [m]} \frac{(-1)^{|I|}}{2^m} \cdot \prod_{i \in I} (-1)^{f_i} \\
&= \sum_{I \subseteq [m]} \frac{(-1)^{|I|}}{2^m} \cdot \left(1 - 2 \cdot \bigoplus_{i \in I} f_i\right).
\end{aligned}$$

By our analysis for the first case, X fools $\bigoplus_{i \in I} f_i$ with error ε. Therefore, by the triangle inequality,

$$\begin{aligned} |\mathbb{E}[f(X)] - \mathbb{E}[f]| &\leq \sum_{I \subseteq [m]} \left| \frac{(-1)^{|I|} \cdot (-2)}{2^m} \right| \cdot \left| \mathbb{E}\left[\left(\bigoplus_{i \in I} f_i \right)(X) \right] - \mathbb{E}\left[\bigoplus_{i \in I} f_i \right] \right| \\ &\leq \sum_{I \subseteq [m]} \frac{2}{2^m} \cdot \varepsilon = 2\varepsilon. \end{aligned}$$

For the final case, suppose the output gate of f is $\vee$. By de Morgan's laws, $\neg f$ can be computed by a read-once depth-2 $\mathbf{AC}^0[\oplus]$ formula with output gate $\wedge$. By our analysis for the second case, X fools $\neg f$ with error 2ε, hence X fools f with the same error. ◀

6 Directions for Further Work

Is there any setting where the iterated restrictions approach (with $\omega(1)$ iterations) can give a pseudorandom generator (or even a hitting set generator) with truly *optimal* seed length $O(\log(n/\varepsilon))$?

Suppose X, X', X'' are three independent small-bias distributions. Does $X + X' \wedge X''$ fool read-once CNFs with optimal seed length $O(\log(n/\varepsilon))$?

References

1. M. Ajtai, J. Komlos, and E. Szemeredi. Deterministic simulation in LOGSPACE. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing (STOC)*, pages 132–140, New York, NY, USA, 1987. ACM. doi:10.1145/28395.28410.
2. Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. *Advances in Computing Research*, 5(199-222):1, 1989.
3. Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.
4. Louay Bazzi and Nagi Nahas. Small-bias is not enough to hit read-once CNF. *Theory of Computing Systems*, 60(2):324–345, February 2017. doi:10.1007/s00224-016-9680-6.
5. Mark Braverman, Gil Cohen, and Sumegha Garg. Pseudorandom pseudo-distributions with near-optimal error for read-once branching programs. *SIAM Journal on Computing*, 0(0):STOC18–242–STOC18–299, 2020. doi:10.1137/18M1197734.
6. Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. *SIAM Journal on Computing*, 43(3):973–986, 2014.
7. J. Brody and E. Verbin. The coin problem and pseudorandomness for branching programs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 30–39, October 2010. doi:10.1109/FOCS.2010.10.
8. Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *Journal of Computer and System Sciences*, 61(1):81–107, 2000. doi:10.1006/jcss.1999.1695.
9. Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, pages 363–375, New York, NY, USA, 2018. ACM. doi:10.1145/3188745.3188800.
10. Sitan Chen, Thomas Steinke, and Salil Vadhan. Pseudorandomness for read-once, constant-depth circuits. *arXiv preprint*, 2015. arXiv:1504.04675.
11. Anindya De. Pseudorandomness for permutation and regular branching programs. In *Proceedings of the 26th Annual IEEE 26th Annual Conference on Computational Complexity (CCC)*, pages 221–231. IEEE, 2011.

12 Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Computer Science*, pages 504–517. Springer, Berlin, 2010. doi:10.1007/978-3-642-15369-3_38.

13 Dean Doron, Pooya Hatami, and William M Hoza. Near-optimal pseudorandom generators for constant-depth read-once formulas. In *34th Computational Complexity Conference (CCC)*, 2019.

14 Michael A. Forbes and Zander Kelley. Pseudorandom generators for read-once branching programs, in any order. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2018.

15 Anat Ganor and Ran Raz. Space pseudorandom generators by communication complexity lower bounds. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 28. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

16 Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 120–129. IEEE, 2012.

17 Parikshit Gopalan and Amir Yehudayoff. Inequalities and tail bounds for elementary symmetric polynomial with applications. *arXiv preprint*, 2014. arXiv:1402.3543.

18 Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. *SIAM Journal on Computing*, 47(2):493–523, 2018. doi:10.1137/17M1129088.

19 William M. Hoza and David Zuckerman. Simple optimal hitting sets for small-success RL. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2018.

20 Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing (STOC)*, pages 356–364. ACM, 1994.

21 Michal Koucký, Prajakta Nimbhorkar, and Pavel Pudlák. Pseudorandom generators for group products. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 263–272. ACM, New York, 2011. doi:10.1145/1993636.1993672.

22 Chin Ho Lee. Fourier bounds and pseudorandom generators for product tests. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:25, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2019.7.

23 Chin Ho Lee and Emanuele Viola. More on bounded independence plus noise: Pseudorandom generators for read-once polynomials. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 24, page 167, 2017.

24 Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC)*, pages 626–637. ACM, New York, 2019.

25 Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

26 Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

27 Noam Nisan. RL $\subseteq$ SC. *computational complexity*, 4(1):1–11, March 1994. doi:10.1007/BF01205052.

28 Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996. doi:10.1006/jcss.1996.0004.

29 Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM*, 55(4):Art. 17, 24, 2008. doi:10.1145/1391289.1391291.

30 Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 655–670. Springer, 2013.

31 Michael E. Saks and Shiyu Zhou. $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$. *Journal of Computer and System Sciences*, 58(2):376–403, 1999.

32 Thomas Steinke. Pseudorandomness for permutation branching programs without the group theory. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2012.

33 Thomas Steinke, Salil Vadhan, and Andrew Wan. Pseudorandomness and Fourier-growth bounds for width-3 branching programs. *Theory of Computing*, 13(12):1–50, 2017. doi:10.4086/toc.2017.v013a012.