Pseudorandom Generators for Read-Once ACC^0

Dmitry Gavinsky
NEC Laboratories America, Inc.,
Princeton, NJ 08540, U.S.A.
Email: dmitry.gavinsky@gmail.com

Shachar Lovett
Institute for Advanced Study,
Princeton, NJ 08540, U.S.A.
Email: shachar@math.ias.edu

Srikanth Srinivasan
DIMACS, Rutgers University,
New Brunswick NJ, U.S.A.
Email: srikanth@dimacs.rutgers.edu

***Abstract*—We consider the problem of constructing pseudorandom generators for read-once circuits. We give an explicit construction of a pseudorandom generator for the class of read-once constant depth circuits with unbounded fan-in AND, OR, NOT and generalized modulo m gates, where m is an arbitrary fixed constant. The seed length of our generator is poly-logarithmic in the number of variables and the error.**

***Keywords*-Pseudorandom generators; Derandomization; Random restrictions**

I. Introduction

The quest for circuit lower bounds originated in the 1980s as a combinatorial approach to the P vs NP problem, and has proved to be one the harder challenges in computational complexity. The term 'circuit lower bounds' can be interpreted in several ways. The weakest form is *worst-case hardness*, where one needs to exhibit a function which cannot be computed exactly by the given class of circuits. A stronger notion is *average-case hardness*, where the requirement is strengthened so that this function can not even be approximated by the given class of circuits. The strongest notion is that of exhibiting a *pseudorandom generator* for the class of circuits. In many cases, average case hardness can be used to construct pseudorandom generators [1], [2]. However, we note this is not always the case, in particular when the class of circuits for which one can prove average case hardness is (in a certain sense) too weak.

Formally, a pseudorandom generator (PRG for short) for a class $\mathcal{C}$ of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ is an explicit map $G : \{0,1\}^s \to \{0,1\}^n$, such that no function in $\mathcal{C}$ can distinguish a uniform output of G from a uniform string in $\{0,1\}^n$. We say G fools the class $\mathcal{C}$ with error ε if for any $f \in \mathcal{C}$, $\left|\Pr_{x\in\{0,1\}^n}[f(x)=1] - \Pr_{y\in\{0,1\}^s}[f(G(y))=1]\right| \leq \varepsilon$.

The main challenge when constructing PRGs is to minimize the seed length s and the error ε. We usually consider a pseudorandom generator as good if its seed length is logarithmic in n, since in such a case it can be derandomized in polynomial time, by enumerating all possible seeds. Pseudorandom generators have been a major object of study in theoretical computer science for several decades, and have found applications in the area of computational complexity, cryptography, algorithms design and more. For more details, we refer the reader to the excellent books [3], [4].

There is essentially just one general class of circuits where strong average-case lower bounds are known: AC^0, the class of bounded-depth circuits with unbounded fan-in $\mathrm{AND}, \mathrm{OR}$ and NOT gates. This class represents parallel computation with a bounded number of rounds, where basic computations correspond to gates, and only trivial basic computations are allowed (corresponding to $\mathrm{AND}, \mathrm{OR}$ and NOT gates). A sequence of works, culminating with the celebrated result of Håstad [5], showed that AC^0 circuits of sub-exponential size cannot predict the parity of n bits with better than exponentially small advantage. Nisan [1] used this average-case hardness to construct pseudorandom generators against AC^0 with poly-logarithmic seed length.

Obviously, one would like to prove strong lower bounds on more realistic models of parallel computation, where the main challenge is to allow more general local computations, for example, general symmetric gates. This amounts to constant depth circuits with arbitrary symmetric gates; which is known to also be equivalent to TC^0, where arbitrary threshold gates are allowed [6], [7]. Research on these problems has been extensive and lead to many ingenious ideas, but no super-polynomial lower bounds for these classes is known to date, and also no pseudorandom generator is known.

Faced with this adversity, research has turned towards studying more restricted models, with the goal of coming up with new ideas for hard functions and pseudorandom generators, and even more importantly, better proof techniques. One line of research has been to allow arbitrary constant depth, but limit the symmetric basic gates allowed. This approach was essentially successful in only one model of computation: $\mathrm{ACC}^0[p]$, where in addition to $\mathrm{AND}, \mathrm{OR}, \mathrm{NOT}$ gates also counting gates modulo p are allowed. Razborov [8] and Smolensky [9] showed a (weak form) of an average-case lower bound when p is prime or prime power. Explicitly, they showed that such circuits require exponential size to compute the sum of the bits modulo q, where q is any fixed number co-prime to p. In fact, they showed that such circuits cannot even approximate the sum modulo q with very good accuracy. However, these average-case lower bounds are too weak to produce pseudorandom generators for $\mathrm{ACC}^0[p]$, and this remains an important open problem.

If one allows modular gates with non-prime modulus, then all the previous techniques break down. While it is widely believed that constant depth circuits with counting gates cannot compute very simple functions (for example majority), the best result to date is a relatively recent breakthrough of Williams [10], who showed that such circuits cannot compute all nondeterministic exponential time algorithms (NEXP). Clearly, this state of affairs is far from optimal, and in particular no average-case lower bounds or efficient pseudorandom generators are known when general counting gates are allowed.

A. Our results

In this work we restrict our attention to *read-once* circuits, which is a limited model of computation where each gate in the circuit has fan-out one. The study of read-once models has been extensively studied in the context of branching programs, but not as much in the context of circuits. Our main result is an explicit pseudorandom generator which fools read-once $\mathrm{ACC}^0[m]$ circuits for any fixed m (not necessarily prime).

We first define the model of computation exactly. A Boolean circuit is represented by a directed acyclic graph; the inputs are placed on the input nodes (nodes with in-degree zero); the output on the output node (node with out-degree zero); and basic gates are placed on non-input nodes which define the function that the circuit computes. A circuit is *read-once* if the out-degree of each node is at most one. The depth of a circuit is the maximal length of a path between inputs and output.

In our case, a read-once $\mathrm{ACC}^0[m]$ circuit, where $m \geq 1$ is a fixed integer, is a read-once circuit with several types of basic gates: the standard $\mathrm{AND}, \mathrm{OR}, \mathrm{NOT}$ gates and also MOD_m gates. A MOD_m gate computes some linear combination of its inputs *modulo* m, and its output depends only on the outcome of this linear combination in $\mathbb{Z}_m$ (in technical terms, these are commonly called generalized MOD_m gates). That is, a function $g : \{0,1\}^t \to \{0,1\}$ is a MOD_m gate if $g(x) = \mathbb{1}_{\langle a,x\rangle \bmod m \in A}$, where $a \in \mathbb{Z}_m^t$ is a linear combination; $\langle a, x\rangle = \sum_{i=1}^t a_i x_i$ is the inner product; and $A \subseteq \mathbb{Z}_m$ is an accepting set. When we need to specify the linear combination and the accepting set, we denote this by $g = \mathrm{MOD}_{\mathrm{m}}^{\mathrm{a,A}}$.

Theorem 1 (Main theorem)**.** *Let m denote the modulus, d the depth of a circuit, n the number of variables and ε the required error. There exists an explicit generator $G : \{0,1\}^s \to \{0,1\}^n$ where $s = 2^{O(d^2)} \cdot (m \log n)^{O(d)} \cdot \log(1/\varepsilon)^{O(1)}$ such that the following holds. If $C : \{0,1\}^n \to \{0,1\}$ is a depth d read-once $\mathrm{ACC}^0[m]$ circuit then*

$$\left| \Pr_{x \in \{0,1\}^n}[C(x) = 1] - \Pr_{y \in \{0,1\}^s}[C(G(y)) = 1] \right| \leq \varepsilon.$$

B. Related results

As we stated already, the study of read-once models is common in the context of branching programs. A branching program is a combinatorial model for an algorithm with small space (memory) usage. The main question in this area is the RL vs L problem, which asks whether randomization help in computation in the context of small space algorithms. This area has been very successful, and pseudorandom generators with poly-logarithmic seed length are known for small space branching programs [11], [12], [13], [14].

The relation to read-once circuits is simple. It is easy to see that any read-once ACC^0 circuit can be converted to a read-once branching program which uses only a constant amount of space, by evaluating the gates in a depth-first search. Thus, one may hope to use the previous results for small space branching programs in order to construct pseudorandom generators for read-once circuits. The main obstacle is that the pseudorandom generators for branching programs mentioned above, crucially require the input bits to be read in a prescribed known order; while a read-once circuit may use the bits in any order. Thus, a more fitting model for comparison is that of branching programs where each bit is read once, but in an arbitrary and unknown order. This more general problem was studied by [15] who gave a pseudorandom generator whose seed length is cn for some $1/2 < c < 1$. Note that this generator saves only a constant factor in the amount of randomness required; contrast this with the previous scenario, when the order of bits read is known, where there exist explicit pseudorandom generators with just poly-logarithmic seed length.

Some special cases of read-once $\mathrm{ACC}^0[m]$ circuits were studied in previous works. The works of [16], [17] studied pseudorandom generators which fool the sum of random bits modulo m; this corresponds to a read-once depth-one $\mathrm{ACC}^0[m]$ circuit with a single MOD_m gate. The work of [18] studied read-once DNFs; this corresponds to a read-once depth-two AC^0 circuit. In both cases the authors gave constructions with logarithmic seed length.

Paper organization: We start with some preliminaries in Section II and then define our PRG formally in Section III. We give an overview of the analysis in Section IV. The detailed analysis follows in subsequent sections, with the main result appearing in Section VII. For lack of space, we omit many proofs.

II. Definitions and Preliminaries

Given two distributions μ and ν defined on a finite set X, we denote the *statistical distance* between μ and ν by $d_{TV}(\mu, \nu)$. For $\varepsilon > 0$, we say that μ and ν are ε-close if $d_{TV}(\mu, \nu) \leq \varepsilon$. Given random variables Y and Z taking values over the same finite set X, we define the statistical distance between Y and Z to be the statistical distance between their distributions. Also, for $\varepsilon > 0$ we say that Y and Z are ε-close if their distributions are ε-close.

Definition 2 (Fooling)**.** *Let R be a fixed finite set and $n \in \mathbb{N}, \varepsilon > 0$ be parameters. Given a function $f : \{0,1\}^n \to R$ and a distribution μ over $\{0,1\}^n$. We say that μ* ε-fools *f if the random variables $f(Y)$ and $f(Z)$ are ε-close, where Y is a uniformly distributed element of $\{0,1\}^n$ and Z is a random distribution drawn according to distribution μ. Given a tuple of functions $\overline{f} = (f_1, \ldots, f_\ell)$, we say that μ ε-fools $\overline{f}$ if it fools the function $F : \{0,1\}^n \to R^\ell$ defined as $F(a) = (f_1(a), \cdots, f_\ell(a))$. For a family of functions $\mathcal{F}$ mapping $\{0,1\}^n$ to R, we say that μ* ε-fools *$\mathcal{F}$ if μ ε-fools f for each $f \in \mathcal{F}$. Finally, given a function $g : \{0,1\}^n \to \mathbb{C}$, we say that μ* ε-fools *g* in expectation *if $\left|\boldsymbol{E}_{Y \sim \{0,1\}^n}[g(Y)] - \boldsymbol{E}_{Z \sim \mu}[g(Z)]\right| \leq \varepsilon$.*

Fact 3. *Fix $\varepsilon > 0$. Let $F : \{0,1\}^n \to R$ where R is a finite set and let $g : R \to \mathbb{C}$ s.t. $|g(a)| \leq 1$ for each $a \in R$. If a distribution μ over $\{0,1\}^n$ ε-fools f, then it (2ε)-fools $g \circ f$ in expectation.*

Definition 4 (Pseudorandom Generators (PRGs))**.** *Let R be a fixed finite set and $n \in \mathbb{N}, \varepsilon > 0$ be parameters. Fix a family of functions $\mathcal{F}$ mapping $\{0,1\}^n$ to R. A function $G : \{0,1\}^s \to \{0,1\}^n$ is said to be an* ε-Pseudorandom Generator (PRG) for *$\mathcal{F}$ if μ_G ε-fools $\mathcal{F}$, where μ_G is the distribution of the random variable $G(Y')$, where Y' is a uniformly random element of $\{0,1\}^s$. The quantity s is called the* seedlength of *G.*

Definition 5 (Almost k-wise indistinguishability)**.** *Fix a finite set X and parameters $t, k \in \mathbb{N}$ and $\delta > 0$. Given two distributions μ, ν over the product set X^t, we say that μ and ν are* δ-close to k-wise indistinguishable*, if for every subset $S \subseteq [t]$ of size at most k, the marginals $\mu \mid_S$ and $\nu \mid_S$ are δ-close.*

Quite often, we will show that to fool functions of a certain form, it suffices to fool related functions that are easier to analyze. We state a few such reductions below. The proofs are omitted.

Fact 6. *Let $f : \{0,1\}^n \to \mathbb{Z}_m$ and $\omega = e^{2\pi i/m}$. We define $\omega^f : \{0,1\}^n \to \mathbb{C}$ as follows: $\omega^f(a) = \omega^{f(a)}$. Then, a distribution μ ε-fools f iff μ ε'-fools the function $\omega^{\alpha \cdot f}$ in expectation for each $\alpha \in \mathbb{Z}_m$, where $\varepsilon' = \varepsilon/2m$.*

Lemma 7. *Let $C_1, \ldots, C_k$ be arbitrary Boolean functions. Then any distribution that ε-fools all functions of the form $\wedge(C_j : j \in S)$ for $S \subseteq [k]$ also $(3^k \cdot \varepsilon)$-fools the tuple $(C_1, \ldots, C_k)$.*

Lemma 8. *Let $C_1, \ldots, C_k$ be arbitrary functions taking values in $\mathbb{Z}_m$. Then any distribution that ε-fools all functions of the form $\sum_i \alpha_i \cdot C_i$ for $(\alpha_1, \ldots, \alpha_k) \in \mathbb{Z}_m^k$ also $(m^{k/2} \cdot \varepsilon)$-fools the tuple $(C_1, \ldots, C_k)$.*

A. Deviation Inequalities

We need the following form of the Chernoff-Hoeffding bound, which follows from [19, Theorem 1.2].

Fact 9. *Let $Z_1, \ldots, Z_t$ be independent $[0,1]$-valued random variables and $Z = \sum_{i=1}^t Z_i$. Assume that $\boldsymbol{E}[Z] = M$. Then, $\Pr[Z > M + A], \Pr[Z < M - A] \leq \exp\left\{-\frac{A^2}{2(M+A)}\right\}$.*

The following lemma was proved in [16].[1]

Lemma 10. *Fix $\delta > 0$ and $k \in \mathbb{N}$ s.t. k is even. Let $Y_1, \ldots, Y_t$ be a collection of independent $\{0,1\}$-valued random variables and let $Y_1', \ldots, Y_t'$ be a collection of $\{0,1\}$-valued random variables s.t. given any subset $S \subseteq [t]$ of size at most k, $|\boldsymbol{E}[\prod_{i \in S} Y_i] - \boldsymbol{E}[\prod_{i \in S} Y_i']| \leq \delta$. Let $Y = \sum_{i=1}^t Y_i$ and $Y' = \sum_{i=1}^t Y_i'$. Assume, moreover that $\boldsymbol{E}[Y] = M$. Then, for any $A > 0$ we have, $\Pr[|Y' - M| \geq A] \leq 8\left(\frac{kM+k^2}{A^2}\right)^{k/2} + (t+M)^k\delta$.*

The above also implies similar bounds for $[0,1]$-valued random variables, the proof of which we omit.

Corollary 11. *Fix $\delta > 0$ and $k \in \mathbb{N}$ s.t. k is even. Let $Y_1, \ldots, Y_t$ be a collection of independent $[0,1]$-valued random variables and let $Y_1', \ldots, Y_t'$ be a collection of $[0,1]$-valued random variables s.t. given any subset $S \subseteq [t]$ of size at most k, $|\boldsymbol{E}[\prod_{i \in S} Y_i] - \boldsymbol{E}[\prod_{i \in S} Y_i']| \leq \delta$. Let $Y = \sum_{i=1}^t Y_i$ and $Y' = \sum_{i=1}^t Y_i'$. Assume, moreover, that $\boldsymbol{E}[Y] = M$. Then, for any $A > 0$ we have, $\Pr[|Y' - M| \geq A] \leq 8\left(\frac{4(kM+k^2)}{A^2}\right)^{k/2} + (t+M)^k\delta + \exp\left\{-\frac{A^2}{8(M+2A)}\right\}$.*

B. The class of functions we will fool

The class of functions we consider are those computed by a class of circuits defined on n boolean variables. These circuits are made up of AND and MOD_m gates, where m is some fixed constant. In general, a MOD_m gate applied to inputs $y_1, \ldots, y_t$ accepts iff $\sum_{i=1}^t \alpha_i y_i \in A$ for some fixed $\alpha_1, \ldots, \alpha_t \in \mathbb{Z}_m$ and $A \subseteq \mathbb{Z}_m$; we call such gates *Boolean* MOD_m gates. We also consider MOD_m gates that just output a $\mathbb{Z}_m$-linear combination of their inputs; such gates are called $\mathbb{Z}_m$*-valued* MOD_m gates.

For any constant $d \in \mathbb{N}$ and $n \in \mathbb{N}$, we denote by $\mathcal{C}_d$ the class of read-once depth d-circuits made up of alternating layers of AND and MOD_m gates where the intermediate MOD_m gates in the circuit are boolean MOD_m gates, and the output gate is either an AND gate or a $\mathbb{Z}_m$-valued MOD_m gate. Thus, we allow the circuit to output an arbitrary element of $\mathbb{Z}_m$. We allow the variables to be negated at the input. Given an arbitrary read-once $\mathrm{ACC}^0[m]$ circuit C of depth d, there is a $C' \in \mathcal{C}_{2d}$ s.t. any distribution that ε-fools C' also ε-fools C (for any $\varepsilon > 0$).

Given $C \in \mathcal{C}_d$, we say that a circuit $C = \wedge(C_1, \ldots, C_t)$ if C is an AND of subcircuits $C_1, \ldots, C_t$. We also say that $C = \mathrm{MOD}_m(C_1, \ldots, C_t)$ if C is a (boolean or $\mathbb{Z}_m$-valued) MOD_m gate applied to subcircuits $C_1, \ldots, C_t$. For $a \in \mathbb{Z}_m^t$,

[1] Strictly speaking, the proof in [16] also assumes that the marginal of Y_i is the same as that of Y_i' for each i. However, it is easy to check that their proof works in the above, more general, scenario.

we say $C = \mathrm{MOD}_m^a(C_1, \ldots, C_t)$ for $a \in \mathbb{Z}_m^t$ if C is a $\mathbb{Z}_m$-valued MOD_m gate applied to $C_1, \ldots, C_t$ with coefficients $a_1, \ldots, a_t$; similarly, $C = \mathrm{MOD}_m^{a,A}(C_1, \ldots, C_t)$ for $a \in \mathbb{Z}_m^t$ and $A \subseteq \mathbb{Z}_m$ if C is a boolean MOD_m gate applied to $C_1, \ldots, C_t$ with coefficients $a_1, \ldots, a_t$ and accepting set A.

Given circuits $C, C' \in \mathcal{C}_d$, we say that C' is a *projection* of C if C' is obtained from C by possibly setting some input variables to 0 or 1 and possibly negating some input variables. Given circuits $C, C' \in \mathcal{C}_d$, we call C' a *modification* of the circuit C if C' is obtained from C in the following way: If $C = \wedge(C_1, \ldots, C_t)$, then $C' = \wedge(C'_i : i \in S \subseteq [t])$, where each C'_i is a projection of C_i. If $C = \mathrm{MOD}_m(C_1, \ldots, C_t)$, then $C' = \mathrm{MOD}_m(C'_1, \ldots, C'_t)$, where each C'_i is a projection of C_i, and the coefficients corresponding to the output MOD_m-gates of C and C' can be different.

We now define the class of circuits we are going to fool. Let n be a growing parameter and $d, k \in \mathbb{N}$ be constants. We denote by $\mathcal{C}_{d,k}$ the set of all k-tuples of circuits $(C^1, \ldots, C^k)$ s.t. there is a circuit $C \in \mathcal{C}_d$ with the property that for each $j \in [k]$, C^j is a modification of C. Clearly, each tuple of circuits $(C^1, \ldots, C^k)$ gives us a tuple of functions mapping $\{0,1\}^n$ to $\{0,1\}$ or $\mathbb{Z}_m$.

We now state the more technical main result, which implies Theorem 1.

Theorem 12. *Fix constants $k, d \in \mathbb{N}$. For any $n \in \mathbb{N}$ and $0 < \varepsilon \leq 1/n$, there is an explcit PRG $G_{d,k,\varepsilon} : \{0,1\}^s \to \{0,1\}^n$ that ε-fools $\mathcal{C}_{d,k}$ with error at most ε and has seedlength $s = 2^{O(d^2)} \cdot (km \log n)^{O(d)} \cdot (\log(1/\varepsilon))^{O(1)}$.*

C. The case $d = 1$

The proof of Theorem 12 is by induction on the depth d of the circuits considered. The base case of the induction, $d = 1$, follows from the result of [16] stated below. (We could also use an older result of Nisan [13] for the parameters that we are interested in.)

Theorem 13. *For any $n \in \mathbb{N}$ and $\varepsilon > 0$, there is an explicit PRG $G^1 : \{0,1\}^{s_1} \to \{0,1\}^n$ s.t. G^1 ε-fools any function $f : \{0,1\}^n \to \mathbb{Z}_m$, where $f(x) = \sum_i \alpha_i x_i$, with $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}_m$. Furthermore, the seedlength of G^1 is $O(\log n + \log(1/\varepsilon) \log\log(1/\varepsilon))$.*

It is not hard to see that the generator given by Theorem 13 in fact fools a small number of modifications of the same circuit. If the top gate is a MOD_m gate it follows from Lemma 8, and if the top gate is an AND gate it follows from Lemma 7. We omit the detailed proof.

Corollary 14. *Fix constant $k \in \mathbb{N}$. For any $n \in \mathbb{N}$ and $0 < \varepsilon < 1/n$, there is an explicit PRG $G_{1,k,\varepsilon}$ that ε-fools $\mathcal{C}_{1,k}$ with error at most ε. The seedlength of $G_{1,k,\varepsilon}$ is at most $(km)^2 \log n \cdot (\log(1/\varepsilon))^2$.*

III. The construction of the PRG

In this section, we present the formal construction of the PRG $G_{d,k,\varepsilon}$ and analyze its seedlength. The construction will be fully analyzed in Section VII.

The construction is inductive based on the depth d of the circuits. We will define, by induction on d, a PRG $G_{d,k,\varepsilon}$ with seedlength at most $2^{100d^2}(km \log n)^{10d}(\log(1/\varepsilon))^2$. For the base case ($d = 1$), we use the PRG from Corollary 14. Clearly, $G_{1,k,\varepsilon}$ has the required seedlength.

Now, fix $d > 1$ and assume that for each constant k and $\varepsilon \leq 1/n$, we have defined the PRG $G_{d-1,k,\varepsilon}$ (for every k) with seedlength at most $2^{100(d-1)^2}(km \log n)^{10(d-1)} \cdot (\log(1/\varepsilon))^2$. Let $\delta = 1/n^{c'(km)^5 \log(1/\varepsilon)}$ where $c' = 10^6$. The PRG $G_{d,k,\varepsilon}$ is obtained by combining the outputs of several PRGs $G^i_{d,k,\varepsilon}$ ($i \in \{0, \ldots, \log n\}$) which we now define.

For $i \in \{0, \ldots, \log n\}$, the PRG $G^i_{d,k,\varepsilon}$ uses as random seed mutually independent strings $y_0, \ldots, y_{i+1}$, where each y_j ($j \in \{0, \ldots, i+1\}$) is a seed to the PRG $G_{d-1,2k,\delta}$. Let a'_j denote $G_{d-1,2k,\delta}(y_j)$ for $j \in \{0, \ldots, i+1\}$. Moreover, let $I \subseteq [n]$ be the set whose characteristic vector is $\bigwedge_{1 \leq j \leq i} a'_j$ (if $i = 0$, we assume $I = [n]$). Then, $G^i_{d,k,\varepsilon}(y_0, \ldots, y_{i+1})$ is defined to be z, where $z|_{\overline{I}} = a'_0$ and $z|_I = a'_{i+1}$. That is, we ensure that the output of $G^i_{d,k,\varepsilon}$ on the coordinates inside and outside I are projections of outputs of the PRG $G_{d-1,2k,\delta}$ on the *independent* random seeds y_0, y_{i+1} to these coordinates. Using the inductive hypothesis, we see that the seedlength of $G_{d-1,2k,\delta}$ is at most $2^{100d^2-100d}(2km \log n)^{10d-10} \cdot (\log(1/\delta))^2 = O(2^{100d^2-100d} 2^{10d}(km \log n)^{10d-10} \cdot ((km)^5 \log n \log(1/\varepsilon))^2) = O(2^{100d^2}(km)^{10d}(\log n)^{10d-8}(\log(1/\varepsilon))^2)$. Thus, the seedlength of $G^i_{d,k,\varepsilon}$ is at most $(\log n + 2) \cdot O(2^{100d^2}(km)^{10d}(\log n)^{10d-8}(\log(1/\varepsilon))^2) = O(2^{100d^2}(km)^{10d}(\log n)^{10d-7}(\log(1/\varepsilon))^2)$.

The PRG $G_{d,k,\varepsilon}$ takes as input mutually independent strings $y'_0, \ldots, y'_{\log n}$ where for each $i \in \{0, \ldots, \log n\}$, the string y'_i is a seed to $G^i_{d,k,\varepsilon}$. Then, we set $G_{d,k,\varepsilon}(y'_0, \ldots, y'_{\log n})$ to be $\bigoplus_{i=0}^{\log n} G^i_{d,k,\varepsilon}(y'_i)$. The seedlength of $G_{d,k,\varepsilon}$ is at most $(\log n + 1) \cdot O(2^{100d^2}(km)^{10d}(\log n)^{10d-7}(\log(1/\varepsilon))^2) \leq 2^{100d^2}(km \log n)^{10d} \cdot (\log(1/\varepsilon))^2$, and hence the seedlength of $G_{d,k,\varepsilon}$ still obeys the inductive claim.

IV. Proof overview

We now describe the main ideas that come into the proof of Theorem 12. The proof is by an induction on the depth of the circuit. Recall that we need to construct a pseudorandom generator which fools a k-tuple of modifications of a read-once circuit. For the sake of clearness, we first describe our approach in the case of $k = 1$ which corresponds to a single read-once circuit. We then explain how these ideas can be expanded to allow for a few simultaneous modifications

of the same read-once circuit (which as stated already, is needed for our inductive step). We will also hide the exact dependency of the parameters on m in this proof overview.

Let $C = g(C_1(x_1), \ldots, C_t(x_t))$ be a read-once circuit, where $g \in \{\wedge, \mathrm{MOD}_m\}$. Let us first consider the (more interesting) case when $g = \mathrm{MOD}_m$. Assume that $C(x) = \sum_{i=1}^{t} \alpha_i C_i(x)$. By Fact 6, to fool C, it suffices to fool the function $\omega^{\alpha \cdot C(x)} = \omega^{\sum_i \cdot \sum_i \alpha\alpha_i C_i(x)}$ for each $\alpha \in \mathbb{Z}_m$. Fix an $\alpha \in \mathbb{Z}_m$. Let $F(x)$ denote $\omega^{\alpha C(x)}$ and let $C_i'(x), F_i(x)$ denote $\alpha\alpha_i C_i(x), \omega^{\alpha\alpha_i C_i(x)}$ respectively. We define the *weight of* F to be $\mathrm{wt}(F) := \sum_i \mathrm{Var}(F_i)$. A simple but crucial observation we use is that when $\mathrm{wt}(F)$ is large, then F is unbiased: more formally, $|\mathbf{E}_x[F(x)]| \leq \exp\{-\Omega(\mathrm{wt}(F))\}$.

Our PRG construction (and analysis) are naturally partitioned into two parts depending whether $\mathrm{wt}(F)$ is small (at most $c\log(1/\varepsilon)$) or large (at least $c\log(1/\varepsilon)$), where $c > 0$ is an appropriately chosen constant. The generators for the low-weight and high-weight cases are then combined to give a single generator which handles both cases simultaneously. This approach has been used successfully before in several contexts, for example in [16], [17], [18], and is also instrumental in our work.

Low weight case: When the weight of C is small, (at most $c\log(1/\varepsilon)$), then we show that a PRG G' for depth $d-1$ (with somewhat smaller error δ) already ε-fools C. Intuitively, this is because if F has low weight, then there is a fixed $z \in \mathbb{Z}_m^t$ s.t. for most inputs x, the vector $(C_1'(x), \ldots, C_t'(x))$ is close to z in Hamming distance (at distance $O(\log(1/\varepsilon))$). We can use this to approximate F by low-degree polynomials in the F_i (which are in turn low-degree polynomials in C_i'). Now, since the generator G' fools ANDs of the C_i', it also fools these low-degree polynomials and hence the function F. The formal proof proceeds by construction of sandwiching polynomials, similar to the work of [20], [18].

High weight case: Assume now that $\mathrm{wt}(F)$ is large (at least $c\log(1/\varepsilon)$). Note that in this case $|\mathbf{E}[F(x)]| \ll \varepsilon$ and it is sufficient to show that this holds under the output of our PRG as well. Moreover, also note that for any function F' such that $\mathrm{wt}(F') \geq 2\log(1/\varepsilon)$ (say), $|\mathbf{E}[F'(x)]|$ is ε-close to $|\mathbf{E}[F(x)]|$.

The idea is to show that if we randomly restrict F suitably, then w.h.p. we obtain a function F' s.t. $\mathrm{wt}(F')$ is roughly a large constant multiple of $\log(1/\varepsilon)$. We then fool F' (on the unset input bits) using the PRG for the low-weight case. Our final PRG construction is a derandomization of this random restriction argument.

Consider now a random restriction, where we set each bit of x to 0 with probability $1/2-p$, to 1 with probability $1/2-p$, or keep it alive with probability p. It is not hard to see that random restrictions decrease $\mathrm{wt}(F)$ (since they decrease the variance of each F_i), and hence there exists a probability $1/n < p < 1$ for which the restricted circuit has on average low weight (say, between $c\log(1/\varepsilon)/4$ and $c\log(1/\varepsilon)/2$). We can even assume that $p = 2^{-\ell}$ for some $0 \leq \ell \leq \log(n)$. Moreover, we can ensure that the weight falls in this range with probability $1-\varepsilon$ by a Chernoff argument. Thus, if G' is the generator for the low-weight case, it will fool F'; and since F' has weight above $c\log(1/\varepsilon)/4$, it will have similar distribution to that of C (for large enough c).

The main challenge is to how to derandomize the random restriction argument. Assume first we know the correct value of p. Let ρ be a random restriction. It will be useful to think of ρ as composed from two parts: the set of live variables $I \in \{0,1\}^n$, where $\Pr[I_i = 1] = p$; and an assignment to the variables $a \in \{0,1\}^n$ which will be used for the fixed variables. Let $I \cdot x$ be defined as the coordinate-wise AND of v and x. Then the value of $x \in \{0,1\}^n$ under the restriction $\rho = (I, a)$ is given by

$$x_\rho = I \cdot x + (\neg I) \cdot a.$$

The function F under the restriction ρ is given by $F_\rho(x) = F(x_\rho)$. Now, the average over restrictions ρ of the weight of F_ρ is

$$\mathop{\mathbf{E}}_{\rho}[\mathrm{wt}(F_\rho)] = \sum_{i=1}^{t} \mathop{\mathbf{E}}_{\rho}[\mathrm{Var}((F_i)_\rho)]$$

Crucially, the average variance of $(F_i)_\rho$ in the above formula can be expressed as

$$\begin{aligned}\mathop{\mathbf{E}}_{\rho}[\mathrm{Var}((F_i)_\rho)] &= \mathop{\mathbf{E}}_{(I,a)}[\mathop{\mathbf{E}}_{x}[(F_i)_\rho(x)] \cdot \mathop{\mathbf{E}}_{y}[\overline{(F_i)_\rho(x)}]] \\ &= 1 - \mathop{\mathbf{E}}_{x,y}[\mathop{\mathbf{E}}_{\rho}[(F_i)_\rho(x) \cdot \overline{(F_i)_\rho(x)}]]\end{aligned}$$

Fix any x, y. We would like to replace the random choice of I, a by a pseudorandom choice so that the above expression remains the same. Assume for now I is fixed, and we just wish to derandomize the choice of a. The key observation is that to preserve $\mathbf{E}_a[(F_i)_\rho(x) \cdot \overline{(F_i)_\rho(x)}]$, we only need to fool the output distribution of of *two* modifications of a read-once circuit C_i of depth $d-1$ (this is precisely why we need to consider tuples of circuits in general). Thus, we can use our generator for depth $d-1$ and $k = 2$ to generate a, without changing $\mathbf{E}_\rho[\mathrm{Var}(F_i)]$ by much. In order to choose I pseudo-randomly, we write $I = I_1 \cdot \ldots \cdot I_\ell$ where $I_1, \ldots, I_\ell \in \{0,1\}^n$ are uniform, and replace each one by an independent output of the same generator for depth $d-1$ and $k = 2$.

The above approach suffices if we only needed a pseudorandom distribution which gives a suitable low weight *on average*. However, to make the argument work we need the stronger property that the weight is of the order of $\log(1/\varepsilon)$ with probability $1-\varepsilon$. This is achieved by a similar argument which considers the joint distribution of the variance of a small number of F_i, and is similar in spirit to concentration bounds derived from bounded moments.

The above argument assumed that the correct probability of restriction $p = 2^{-\ell}$ is known. We wish, however, to construct a pseudorandom generator which would work for any value of p. To do so, we construct a pseudorandom generator for all choices of $0 \leq \ell \leq \log(n)$, each using an independent seed, and the combine them together (by xoring them) in order to create a single generator which works for all values of p. This multiplies the seed length by an additional $\log(n)$ factor for each level of the circuit.

This finishes the argument when the top-level gate g of C is a MOD_m gate. When it is an AND gate, things are much simpler. If $C = \wedge(C_1, \ldots, C_t)$ is read-once and $\mathrm{wt}(C) := \sum_i \mathrm{Var}(C_i)$ is large, then it is easy to see that C almost always takes the value 0. Hence, in this case, only the low-weight case is relevant and this is handled analogously to the low-weight MOD_m case.

V. Low weight cases

In this section, we prove a few results that will allow us to show that the PRG $G_{d,k,\varepsilon}$ from Section V-B fools *low weight* functions. We first need a basic lemma.

A. A low weight lemma

For this section, fix a constant-sized set X. We will consider functions defined on domain X^t where t is a growing parameter. Given a distribution ν over a set X, we define $\mathbb{V}(\nu) = 1 - \max_{x \in X} \nu(x)$. Let m_1 denote $|X|$. The following is easy to check.

Fact 15. *Let X be a random variable taking values in $\mathbb{Z}_m$. Then $\mathbb{V}(X) \leq m^2 \cdot \mathrm{Var}(e^{\frac{2\pi i}{m} \cdot X})$.*

Let $\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_t$ be a product distribution on X^t. The lemma below states that if $\mathbb{V}(\mu)$ is not too large, and μ' is a distribution that is δ-close to k-wise indistinguishable from μ for reasonably large k and small δ, then μ and μ' are in fact statistically indistinguishable. Formally,

Lemma 16. *Let $F : X^t \to \mathbb{C}$ be s.t. $|F(x_1, \ldots, x_t)| \leq 1$ for all $x_1, \ldots, x_t$. Assume the product distribution $\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_t$ satisfies $\sum_i \mathbb{V}(\mu_i) \leq c \log(1/\varepsilon_1)$ for some $c \geq 1$ and $\varepsilon_1 > 0$. Let μ' be any distribution on X^t s.t. μ and μ' are ε_2-close to $(c_1 \log \frac{1}{\varepsilon_1})$-wise indistinguishable, where $\varepsilon_2 = t^{-c_1 \log \frac{1}{\varepsilon_1}}$, where $c_1 = 500 m_1^2 c$. Then we have $|\boldsymbol{E}_{\mu'}[F] - \boldsymbol{E}_{\mu}[F]| \leq \varepsilon_1$.*

The proof of the above lemma is through a modification of the standard sandwiching polynomials technique ([20], [18]). We omit the proof in this extended abstract.

B. PRG for depth $d-1$ fools low-weight circuits

We are ready to prove the following theorem, which deals with the case when the output gate of the circuit we are trying to fool is an AND. We state a more general result. For lack of space, we omit the proof.

Theorem 17. *For all $p \in [t]$, let $C_p = \mathrm{MOD}_m^{\beta_p}(C_{p,1}, \ldots, C_{p,s_p})$, where $s_p \in \mathbb{N}$, $\beta_p \in \mathbb{Z}_m^{s_p}$ and $C_{p,q}$'s are arbitrary functions that depend on mutually disjoint sets of variables. Let $C = \wedge(C_1, \ldots, C_t)$. For all $j \in [k]$, let $C^j = \wedge(C_p^j : p \in T_j)$, where $T_j \subseteq [t]$ and every C_p^j is a projection of C_p (that is, C^j is a projection of C). Then for any $\varepsilon > 0$ and*

$$\delta = \left(\frac{\varepsilon^{k \ln m + 2 \ln t}}{t^{2k} \cdot m^{k^2}} \right)^{10000}$$

the following holds: If μ is a distribution that δ-fools all tuples of the form $(\tilde{C}^1, \ldots, \tilde{C}^k)$, where each $\tilde{C}^j$ is a modification of $\mathrm{MOD}_m(C_{p,q} : p \in [t], q \in [s_p])$, then μ also ε-fools the tuple $(C^1, \ldots, C^k)$.

When the output gate is MOD_m the analysis is more complicated, and we will have to treat separately the low-weight and the high-weight cases. The following is a statement for the (simpler) low-weight case.

Theorem 18. *For all $p \in [t]$, let $C_p = \wedge(C_{p,1}, \ldots, C_{p,s_p})$, where $s_p \in \mathbb{N}$ and $C_{p,q}$'s are arbitrary functions that depend on mutually disjoint sets of variables. Let $C = \mathrm{MOD}_m(C_1, \ldots, C_t)$. For all $j \in [k]$, let $T_j \subseteq [t]$; for every $p \in T_j$, let C_p^j be a projection of C_p. For all $p \in [t]$, let*

$$f_p = \sum_{j : T_j \ni p} \beta_p^j \cdot C_p^j,$$

where $\beta_p^j \in \mathbb{Z}_m$. Then for any $\varepsilon > 0$, $c \geq 1$ and $\delta = \varepsilon^{800 c m^4 (k + \log t)}$ the following holds: If

$$\sum_{p=1}^{t} \mathrm{Var}(e^{\frac{2\pi i}{m} f_p}) \leq c \log(1/\varepsilon)$$

and μ is a distribution that δ-fools all tuples of the form $(\tilde{C}^1, \ldots, \tilde{C}^k)$, where each $\tilde{C}^j$ is a projection of $\wedge(C_{p,q} : p \in [t], q \in [s_p])$, then μ also ε-fools in expectation the function $e^{\frac{2\pi i}{m} \sum_{p=1}^{t} f_p}$.

Proof: We will derive a chain of sufficient conditions to guarantee ε-fooling in expectation of $e^{\frac{2\pi i}{m} \sum_{p=1}^{t} f_p}$.

Let the input distribution be uniform, and view f_p's as random variables taking values in $\mathbb{Z}_m$. Then by Fact 15, $\sum_{p=1}^{t} \mathbb{V}(f_p) \leq c m^2 \log(1/\varepsilon)$, and Lemma 16 implies that for $\varepsilon_1 = t^{-500 c m^4 \log \frac{1}{\varepsilon}}$ and $w = 500 c m^4 \log \frac{1}{\varepsilon}$ it holds that any distribution that ε_1-fools any tuple $(f_p : p \in U)$ for $U \subseteq [t]$, $|U| \leq w$, must also ε-fool in expectation $e^{\frac{2\pi i}{m} \sum_{p=1}^{t} f_p}$. From the definition of f_p's, the latter is also guaranteed by ε_1-fooling of the tuples $(C_p^j : p \in U, j \in [k])$.

Now we apply Lemma 7, concluding that for $\varepsilon_2 = \varepsilon_1 / 3^{kw}$, any distribution that ε_2-fools all functions of the form $\wedge(C_p^j : (p, j) \in S)$ for $S \subseteq U \times [k] \subseteq [t] \times [k]$ also ε-fools in expectation $e^{\frac{2\pi i}{m} \sum_{p=1}^{t} f_p}$.

Note that any function like $\wedge(C_p^j : (p, j) \in S)$ for $S \subseteq [t] \times [k]$ can be written as $\wedge(\tilde{C}^j : j \in [k])$, where each

$\tilde{C}^j$ is a projection of $\wedge(C_{p,q} : p \in [t], q \in [s_p])$. Hence, any distribution that ε_2-fools the tuples $(\tilde{C}^1, \ldots, \tilde{C}^k)$ must ε-fools in expectation $e^{\frac{2\pi i}{m}\sum_{p=1}^{t} f_p}$. The result follows by setting $\varepsilon_2 := \delta$. ∎

VI. The High-weight case

A. Random restrictions

Definition 19 (Restrictions)**.** *A* restriction *on the set of variables $X = \{x_1, \ldots, x_n\}$ is a pair $\rho = (I, w)$, where $I, w \in \{0,1\}^n$. We will think of I as a subset of X. The set of all restrictions on the set of variables X is denoted $R(X)$. A* random restriction *is a distribution $\mathcal{R}$ over the set $R(X)$.*

Given a restriction $\rho = (I, a)$ over the set X, we define the function $f|_\rho$ as follows. Given $b \in \{0,1\}^n$, define input $b^\rho \in \{0,1\}^n$ so that for each $i \in [n]$, $b^\rho_i = b_i$ if $i \in I$ and a_i otherwise. Given function $f : \{0,1\}^n \to A$, for any set A, we define $f|_\rho : \{0,1\}^n \to A$ so that for any $b \in \{0,1\}^n$, $f|_\rho(b) := f(b^\rho)$. We point out that in the literature, $f|_\rho$ is often thought of as a function defined on the variables $\{x_i \mid i \in I\}$ only: our definition is somewhat different, and this will help in some technical matters later on.

Fixing parameter $r \in [0,1]$, we define the random restriction $\mathcal{R}_r$ as follows. To sample $\rho \sim \mathcal{R}_r$, we pick $I \in \{0,1\}^n$ by set each $I_i = 1$ independently with probability r and 0 with probability $1 - r$; independently, we pick $a \in \{0,1\}^n$ uniformly at random; the random restriction sampled is $\rho = (I, a)$.

Fix a function $F : \{0,1\}^n \to \mathbb{C}$. We need to understand the behavior of the variance of $F|_\rho$ where $\rho \sim \mathcal{R}_r$. The following lemma essentially follows from the proof of [21, Lemma 6]. For background on Fourier Analysis over $\{0,1\}^n$, we refer the reader to Ryan O'Donnell's lecture notes [22].

Lemma 20. *Fix $r \in [0,1]$ and $F : \{0,1\}^n \to \mathbb{C}$. We have $\boldsymbol{E}_{\rho\sim\mathcal{R}_r}[\mathrm{Var}(F|_\rho)] = \sum_{\emptyset \neq S \subseteq [n]} |\hat{F}(S)|^2(1 - (1-r)^{|S|})$.*

B. The high-weight lemma

In this section, we prove an integral lemma in the proof of the main theorem in Section VII. We will assume the following notation throughout the rest of this section. Say we have $(C^1, \ldots, C^k) \in \mathcal{C}_{d,k}$ s.t. each C^j ($j \in [k]$) is a modification of a circuit $C \in \mathcal{C}_d$ where $C = \mathrm{MOD}_m(C_1, \ldots, C_t)$. For each $p \in [t]$, let $C_p = \wedge(C_{p,q} : q \in [s_p])$, where $s_p \in \mathbb{N}$. Similarly, for each $j \in [k]$, let $C^j = \mathrm{MOD}_m(C^j_1, \ldots, C^j_t)$ and for each $p \in [t]$, $C^j_p = \wedge(C^j_{p,q} : q \in [s_p])$.

For a fixed choice of $\alpha_1, \ldots, \alpha_k \in \mathbb{Z}_m$ define $f := \sum_{j=1}^{k} \alpha_j C^j$. Since $C^j = \mathrm{MOD}_m(C^j_1, \ldots, C^j_t)$, we may write $C^j = \sum_{p=1}^{t} \gamma^j_p C^j_p$ for $\gamma^j_1, \ldots, \gamma^j_t \in \mathbb{Z}_m$. Substituting in the definition of f above and reordering summations, we see that $f = \sum_{p=1}^{t}\sum_{j=1}^{k} \beta^j_p C^j_p$ for some $\beta^j_p \in \mathbb{Z}_m$ where $j \in [k]$ and $p \in [t]$. Define, for each $p \in [t]$, define $f_p : \{0,1\}^n \to \mathbb{Z}_m$ by $f_p = \sum_{j=1}^{k} \beta^j_p C^j_p$. Note that the functions f_p depend on pairwise disjoint sets of variables. Finally, fix an $\alpha \in \mathbb{Z}_m$ and consider $F = \omega^{\alpha\sum_p f_p}$. Let F_p ($p \in [t]$) denote the function $\omega^{\alpha \cdot f_p}$.

Lemma 21. *Assume we have $(C^1, \ldots, C^k) \in \mathcal{C}_{d,k}$ and f_p, F_p ($p \in [t]$) and f, F as above. Furthermore, say that $\mathrm{wt}(F) \geq c\log(1/\varepsilon_2)$, where $c = 1000$. Also, assume that for any $k \in \mathbb{N}$ and $\delta > 0$, the PRG $G_{d-1,k,\delta}$ δ-fools $\mathcal{C}_{d-1,k}$. Then, there is an $i \in [\log n]$ s.t. for each $a \in \{0,1\}^n$, the PRG $G^i_{d,k,\varepsilon}$ ε_2-fools $F(x \oplus a)$ in expectation, where $\varepsilon_2 = \varepsilon/2m^{k/2+1}$.*

Proof: We first note that when $\mathrm{wt}(F)$ is quite large, $|\mathbf{E}_b[F(b)]|$ is small, where b is chosen uniformly at random. This follows from the following claim, whose proof is omitted.

Claim 22. *Assume we have a function $G : \{0,1\}^n \to \mathbb{C}$ s.t. $G = \sum_{p=1}^{s} G_p$, where $G_p = \omega^{g_p}$ amd $g_p : \{0,1\}^n \to \mathbb{Z}_m$ for $p \in [s]$. Moreover, assume that for $p_1 \neq p_2$, the functions g_{p_1} and g_{p_2} depend on disjoint sets of variables. Then, we have $|\boldsymbol{E}_b[G(b)]| \leq e^{-\frac{1}{2}(\sum_p \mathrm{Var}(G_p))}$.*

Fix $a \in \{0,1\}^n$. For any function H defined on $\{0,1\}^n$, we denote by H^a the function $H(x \oplus a)$. Note that for any $a \in \{0,1\}^n$, we have $\mathrm{wt}(F^a(x)) = \sum_{p\in[t]} \mathrm{Var}(F^a_p(x)) = \sum_{p\in[t]} \mathrm{Var}(F_p) = \mathrm{wt}(F) \geq c\log(1/\varepsilon_2)$. In particular, Claim 22 implies that for every $a \in \{0,1\}^n$, $|\mathbf{E}_b[F^a(b]| \leq e^{-\frac{1}{2}\sum_{p\in[t]} \mathrm{Var}(F_p)} = e^{-\mathrm{wt}(F)/2} < \varepsilon_2/2$. Recall that we need to show that for some $i \in [\log n]$, we need to show that $G^i_{d,k,\varepsilon}$ ε_2-fools F^a in expectation for every $a \in \{0,1\}^n$. Since Claim 22 tells us that $|\mathbf{E}[F^a(x)]| \leq \varepsilon_2/2$, to show that $G^i_{d,k,\varepsilon}$ ε_2-fools the function $F^a(x)$ in expectation, it suffices to show that for a random input seed y'_i to PRG $G^i_{d,k,\varepsilon}$, $|\mathbf{E}_{y'_i}[F^a(G^i_{d,k,\varepsilon}(y'_i))]| \leq \varepsilon_2/2$.

We therefore try to prove the above. Recall the definition of $G^i_{d,k,\varepsilon}$: the seed y'_i of G^i consists of independent seeds $y_0, \ldots, y_{i+1}$ for PRG $G_{d-1,2k,\delta}$, where $\delta = 1/n^{c'(km)^5 \log(1/\varepsilon)}$ and c' is a constant. For $j \in [i+1] \cup \{0\}$, let a'_j denote $G_{d-1,2k,\delta}(y_j)$ and let $I \subseteq [n]$ be the set whose characteristic vector is $a'_1 \wedge \cdots \wedge a'_i$. The output of the PRG $G^i_{d,k,\varepsilon}(y'_i) = z$ where $z|_{\overline{I}} = a'_0$ and $z|_I = a'_{i+2}$.

We will take a slightly different view of $G^i_{d,2k,\delta}$. Given $a_0, \ldots, a_i \in \{0,1\}^n$, we define the restriction $\rho(a_0, \ldots, a_i) \in R(X)$ as follows: let $J \in \{0,1\}^n$ be $a_1 \wedge \cdots \wedge a_i$; we define $\rho(a_0, \ldots, a_i)$ to be (J, a_0). It is easy to check that for any input seed y'_i to PRG $G^i_{d,k,\varepsilon}$ and any function f defined on $\{0,1\}^n$, we have $f(G^i_{d,k,\varepsilon}(y'_i)) = f|_{\rho'_i}(a'_{i+1})$, where $\rho'_i = \rho(a'_0, \ldots, a'_i)$ and $a'_0, \ldots, a'_{i+1}$ are as defined above. Note that ρ'_i and a'_{i+1} are independent.

Hence, we need to prove that $|\mathbf{E}_{\rho'_i, a'_{i+1}}[F^a|_{\rho'_i}(a'_{i+1})]| \leq \varepsilon_2/2$. $(*)$

We will prove $(*)$ in two steps: first, we show that there is an $i \in [\log n]$ such that w.h.p. over the choice of ρ' as above, the quantity $\sum_p \mathrm{Var}(F^a_p|_{\rho'})$ is a large, but constant, multiple

of $\log(1/\varepsilon_2)$; we then argue that if this occurs, then the left hand side of $(*)$ is very small. The first of these statements is captured in the following lemma, which is proved in the next section.

Lemma 23. *Let F_p ($p \in [t]$) be as defined above. For $i \in [\log n]$, let $\mathcal{R}$ be the random restriction that samples the outputs $a'_0, \ldots, a'_i$ of the PRG $G_{d-1,2k,\delta}$ on independent random seeds and outputs $\rho'_i = \rho(a'_0, \ldots, a'_i)$, where $\delta \leq 1/n^{10\log(1/\varepsilon_2)}$. Then, there exists an $i \in [\log n]$ such that for any $a \in \{0,1\}^n$, $\Pr_{\rho'_i \sim \mathcal{R}}[\sum_{p\in[t]} \mathrm{Var}(F^a_p|_{\rho'_i}) \notin [250\log(1/\varepsilon_2), 2000\log(1/\varepsilon_2)]] \leq \varepsilon_2^2$.*

Call a restriction ρ' *regular* if $\sum_{p\in[t]} \mathrm{Var}(F^a_p|_{\rho'}) \in [250\log(1/\varepsilon_2), 2000\log(1/\varepsilon_2)]$. Fix any regular ρ'. We will show that $|\mathbf{E}_{a'_{i+1}}[F^a|_{\rho'}(a'_{i+1})]|$ is small. Thus, we know, by Claim 22, that for a uniformly random $b \in \{0,1\}^n$, $|\mathbf{E}_b[F^a|_{\rho'}(b)]| \leq e^{-\frac{1}{2}(\sum_p \mathrm{Var}(F^a_p|_{\rho'}))} < \varepsilon_2^2$.

Consider the tuple $(C^{1,a}|_{\rho'}, \ldots, C^{k,a}|_{\rho'})$ where $C^{j,a}(x) = C^j(x \oplus a)$. Since each $C^{j,a}|_{\rho'}$ is a modification of the circuit $C^a|_{\rho'} \in \mathcal{C}_d$, we see that $(C^{1,a}|_{\rho'}, \ldots, C^{k,a}|_{\rho'}) \in \mathcal{C}_{d,k}$. Hence, by applying Theorem 18, we see that $G^i_{d-1,2k,\delta}$ $\varepsilon_2/8$-fools the tuple $(C^{1,a}|_{\rho'}, \ldots, C^{k,a}|_{\rho'})$ and hence, by Fact 3, we see that $G_{d-1,2k,\delta}$ $\varepsilon_2/4$-fools $F^a|_{\rho'}$ in expectation as well. In particular, since a'_{i+1} is the output of the PRG $G_{d-1,2k,\delta}$ on a random seed, we see that

$$|\mathop{\mathbf{E}}_{a'_{i+1}}[F^a|_{\rho'}(a'_{i+1})]| \leq |\mathop{\mathbf{E}}_{b}[F^a|_{\rho'}(b)]| + \varepsilon_2/4 \leq \varepsilon_2^2 + \varepsilon_2/4 \quad (1)$$

where the second inequality is a consequence of the regularity of ρ'. Now, we are ready to prove $(*)$. Fix i as guaranteed by Lemma 23. We have $|\mathbf{E}_{\rho'_i, a'_{i+1}}[F^a|_{\rho'_i}(a'_{i+1})]| \leq \Pr_{\rho'_i}[\rho'_i \text{ not regular}] + \mathbf{E}_{\rho'_i, a'_{i+1}}[F^a|_{\rho'_i}(a'_{i+1}) \mid \rho'_i \text{ regular}] \leq \varepsilon_2^2 + \varepsilon_2^2 + \varepsilon_2/4 \leq \varepsilon_2/2$ where the first inequality follows from the choice of i as given by Lemma 23 and using (1). This concludes the proof of $(*)$ and hence the proof of Lemma 21. ∎

C. Constructing pseudorandom restrictions: Proof of Lemma 23

Say $i \in [\log n]$ and let $r = 1/2^i$. In this section, we will assume that $\rho \sim \mathcal{R}_r$ is sampled in the following equivalent way: we choose $a_0, \ldots, a_i \in \{0,1\}^n$ independently and uniformly at random, and set $\rho = \rho(a_0, \ldots, a_i)$. Define the random restriction $\mathcal{R}'_r$ as follows: for each $j \in [i] \cup \{0\}$, sample $a'_i \sim \mu_i$ and set the sampled restriction ρ' to be $\rho(a'_0, \ldots, a'_i)$, where for $j \in [i] \cup \{0\}$, the strings a'_j are chosen to be the output of $G_{d-1,2k,\delta}$ on mutually independent random seeds, where $\delta \leq 1/n^{10\log(1/\varepsilon_2)}$.

We now restate Lemma 23 in an equivalent manner using the above notation.

Lemma 23 (Restated from Section VI-B)**.** *There exists an $i \in [\log n]$ such that for $r = 1/2^i$ and any $a \in \{0,1\}^n$,*

$$\Pr_{\rho' \sim \mathcal{R}'_r}[\sum_{p\in[t]} \mathrm{Var}(F^a_p|_{\rho'}) \notin [250, 2000]\cdot\log(1/\varepsilon_2)] \leq \varepsilon_2^2$$

We prove this lemma in two steps. To show that there exists an $r = 1/2^i$ s.t. $\mathcal{R}'_r$ has the property stated in the lemma, we first show that the above property holds for the somewhat similar random restriction $\mathcal{R}_r$ for some such r. We then use this to argue that for the same r, $\mathcal{R}'_r$ continues to have this property. The first step follows as a corollary to Lemma 20.

Corollary 24. *Let F_p ($p \in [t]$) be as defined in the statement of Lemma 21. Then, there exists an $i \in [\log n]$ s.t. for $r = 1/2^i$ and any $a \in \{0,1\}^n$*

$$\mathop{\boldsymbol{E}}_{\rho\sim\mathcal{R}_r}[\sum_{p\in[t]} \mathrm{Var}(F^a_p|_\rho)] \in [500, 1000]\cdot\log(1/\varepsilon_2)$$

$$\Pr_{\rho\sim\mathcal{R}_r}[\sum_{p\in[t]} \mathrm{Var}(F^a_p|_\rho) \notin [250, 2000]\cdot\log(1/\varepsilon_2)] \leq \varepsilon_2^2$$

Proof Sketch: Initially, assume $a = 0$. For $i \in [\log n]$, let v_i denote $\mathbf{E}_{\rho\sim\mathcal{R}_{1/2^i}}[\sum_{p\in[t]} \mathrm{Var}(F_p|_\rho)]$. By Lemma 20, it directly follows that for each $i < \log n$, we have $v_i \geq v_{i+1} \geq v_i/2$. Thus, for some i, v_i lies in the required range. The second claim, regarding the concentration of $\sum_{p\in[t]} \mathrm{Var}(F_p|_\rho)$ follows from a standard application of the Chernoff bound. The same argument works for any $a \in \{0,1\}^n$ since $|\hat{F}^a_p(S)| = |\hat{F}_p(S)|$. ∎

We now argue that a concentration bound similar to that in Corollary 24 holds for the restriction $\mathcal{R}'_r$ for the *same* $r = 1/2^i$, which is, in particular, independent of the choice of a. We will show this by showing that for $\rho \sim \mathcal{R}_r$ and $\rho' \sim \mathcal{R}'_r$, the random variables $(\mathrm{Var}(F^a_p|_\rho) : p \in [t])$ and $(\mathrm{Var}(F^a_p|_{\rho'}) : p \in [t])$ satisfy the assumptions of Corollary 11, which allows to conclude a concentration bound for $\sum_p \mathrm{Var}(F^a_p|_{\rho'})$. Though the argument below works for any $a \in \{0,1\}^n$, but for simplicity of notation, we assume $a = 0$. In order to show that Corollary 11 is applicable, we need the following lemma.

Lemma 25. *Fix any $i \in [\log n]$. Then, for any $S \subseteq [t]$ s.t. $|S| \leq \ell = 4\log(1/\varepsilon_2)$, we have $\left|\boldsymbol{E}_{\rho\sim\mathcal{R}_r}[\prod_{p\in S} \mathrm{Var}(F_p|_\rho)] - \boldsymbol{E}_{\rho'\sim\mathcal{R}'_r}[\prod_{p\in S} \mathrm{Var}(F_p|_{\rho'})]\right| \leq \delta'$, where $\delta' = \delta \cdot \left(\frac{1}{\varepsilon_2}\right)^{O(1)}$.*

The following technical statement captures most of the content of the above lemma.

Lemma 26. *Fix any $i \in [\log n]$. Then, for any $S \subseteq [t]$ s.t. $|S| \leq \ell = 4\log(1/\varepsilon_2)$, we have $|\boldsymbol{E}_{\rho\sim\mathcal{R}_r}[\prod_{p\in S} \boldsymbol{E}[(F_p|_\rho)] \cdot \boldsymbol{E}[(\overline{F_p|_\rho})]] - \boldsymbol{E}_{\rho'\sim\mathcal{R}'_r}[\prod_{p\in S} \boldsymbol{E}[(F_p|_{\rho'})] \cdot \boldsymbol{E}[(\overline{F_p|_{\rho'}})]]| \leq \delta'$, where $\delta' = \delta \cdot \left(\frac{1}{\varepsilon_2}\right)^{O(1)}$ and $\overline{F_p|_\rho}$ represents the complex conjugate of the function $F_p|_\rho$.*

Assuming Lemma 26, let the proof of Lemma 25 is immediate. We omit the details.

Lemma 25 straightaway gives us the main result of this section, Lemma 23.

Proof sketch of Lemma 23: We will use the deviation inequality from Corollary 11. By Corollary 24, there exists an $i \in [\log n]$ s.t. $\mathbf{E}_{\rho\sim\mathcal{R}_r}[\sum_{p\in[t]} \mathrm{Var}(F_p|_\rho)] = M \in [500\log(1/\varepsilon_2), 1000\log(1/\varepsilon_2)]$, where $r = 1/2^i$. Fix this r. Note that for $\rho \sim \mathcal{R}_r$, the random variables $\mathrm{Var}(F_p|_\rho)$ ($p \in [t]$) are independent random variables taking values in $[0,1]$. Now consider $\rho' \sim \mathcal{R}'_r$. By Lemma 25, for every $S \subseteq [t]$ s.t. $|S| \leq 4\log(1/\varepsilon_2)$, we have $\left|\mathbf{E}_{\rho\sim\mathcal{R}_r}[\prod_{p\in S}\mathrm{Var}(F_p|_\rho)] - \mathbf{E}_{\rho'\sim\mathcal{R}'_r}[\prod_{p\in S}\mathrm{Var}(F_p|_{\rho'})]\right| \leq \delta'$, where $\delta' = \delta\cdot(1/\varepsilon_2)^{O(1)}$.

Hence, we see that the random variables $\mathrm{Var}(F_p|_\rho)$ ($p\in[t]$) and $\mathrm{Var}(F_p|_{\rho'})$ ($p\in[t]$) satisfy the hypotheses of Corollary 11 with $k = 4\log(1/\varepsilon_2)$. Thus, by Corollary 11, we see that $\Pr_{\rho'\sim\mathcal{R}'_r}[|\sum_{p\in[t]}\mathrm{Var}(F_p|_{\rho'}) - M| > M/2] < \varepsilon_2^2$ for our parameters k, M, and δ. This finishes the proof. ■

Proof of Lemma 26: We assume w.l.o.g. that $S = [\ell']$ for some $\ell' \leq \ell$. For now, we will keep δ' a parameter and fix $\delta' = \delta\cdot\left(\frac{1}{\varepsilon_2}\right)^{c_1}$ for a large constant c_1 later on in the proof.

We first introduce a larger family of random restrictions $\mathcal{R}_j$ ($j\in\{0,\ldots,i\}$) as follows. We pick $a_0,\ldots,a_i$ independently and uniformly at random from $\{0,1\}^n$ and $a'_0,\ldots,a'_i$ independently s.t. $a'_j \sim \mu_j$ where the distributions μ_j are as defined above. Then, for $j \in \{0,\ldots,i+1\}$, to sample $\rho\sim\mathcal{R}_j$, we simply output $\rho(a_0,\ldots,a_{i-j},a'_{i-j+1},\ldots,a'_i)$. In particular, $\mathcal{R}_0 = \mathcal{R}_r$ and $\mathcal{R}_{i+1} = \mathcal{R}'_r$. Thus, restating the claim of the lemma, we need to show that $|\mathbf{E}_{\rho\sim\mathcal{R}_0}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_\rho)]\cdot\mathbf{E}[(\overline{F_p|_\rho})]] - \mathbf{E}_{\rho'\sim\mathcal{R}_{i+1}}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_{\rho'})]\cdot\mathbf{E}[(\overline{F_p|_{\rho'}})]]| \leq \delta'$.

We will now simplify the above statement in a sequence of steps. By a simple hybrid argument, it suffices to show that for each $j\leq i$, $|\mathbf{E}_{\rho\sim\mathcal{R}_j}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_\rho)]\cdot\mathbf{E}[(\overline{F_p|_\rho})]] - \mathbf{E}_{\rho'\sim\mathcal{R}_{j+1}}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_{\rho'})]\cdot\mathbf{E}[(\overline{F_p|_{\rho'}})]]| \leq \delta''$ (∗) for any δ'' s.t. $\delta'' \leq \delta'/(\log n+1)$. In particular, we choose $\delta'' = \delta'\cdot\varepsilon_2$ which is at most $\delta'/(\log n+1)$ since $\varepsilon_2 \leq 1/n$.

We therefore try to prove (∗). Recall that $\rho = \rho(a_0,\ldots,a_{i-j},a'_{i-j+1},\ldots,a'_i)$ and $\rho' = \rho(a_0,\ldots,a_{i-j-1},a'_{i-j},\ldots,a'_i)$. Fix any particular choice of $a_0,\ldots,a_{i-j-1},a'_{i-j+1},\ldots,a'_i$ and let $\mathcal{R}'_j$ and $\mathcal{R}'_{j+1}$ are the distributions $\mathcal{R}_j$ and $\mathcal{R}_{j+1}$ conditioned on this fixing. It suffices to show for each such fixing that $|\mathbf{E}_{\rho\sim\mathcal{R}'_j}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_\rho)]\cdot\mathbf{E}[(\overline{F_p|_\rho})]] - \mathbf{E}_{\rho'\sim\mathcal{R}'_{j+1}}[\prod_{p\in[\ell']}\mathbf{E}[(F_p|_{\rho'})]\cdot\mathbf{E}[(\overline{F_p|_{\rho'}})]]| \leq \delta''$. (†)

We therefore try to prove (†). Now, since $\mathrm{Vbl}(F_p)\cap\mathrm{Vbl}(F_{p'}) = \emptyset$ for any distinct p,p', we see that $\prod_{p\in[\ell']}\mathbf{E}[F_p|_\rho] = \mathbf{E}[\prod_{p\in[\ell']}F_p|_\rho]$ for any restriction ρ and similarly, $\prod_{p\in[\ell']}\mathbf{E}[\overline{F_p|_\rho}] = \mathbf{E}[\prod_{p\in[\ell']}\overline{F_p|_\rho}]$. Thus, we have $\prod_{p\in[\ell']}\mathbf{E}[F_p|_\rho]\cdot\mathbf{E}[\overline{F_p|_\rho}] = \mathbf{E}[\prod_{p\in[\ell']}(F_p|_\rho)]\,\mathbf{E}[\prod_{p\in[\ell']}(\overline{F_p|_\rho})] = \mathbf{E}_{b',b''}[\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')]$, where b',b'' are chosen independently and uniformly at random from $\{0,1\}^n$. Substituting into (†) and changing the order of the expectations, we see that it suffices to show that $|\mathbf{E}_{b',b''}[\mathbf{E}_{\rho\sim\mathcal{R}'_j}[\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')] - \mathbf{E}_{\rho'\sim\mathcal{R}'_{j+1}}[\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')]]| \leq \delta''$.

By the triangle inequality, to show the above, it suffices to show that for any fixed $b',b''\in\{0,1\}^n$, we have $|\mathbf{E}_{\rho\sim\mathcal{R}'_j}[\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')] - \mathbf{E}_{\rho'\sim\mathcal{R}'_{j+1}}[\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')]| \leq \delta''$. (‡)

Since we have conditioned on a choice of $a_1,\ldots,a_{i-j-1},a'_{i-j+1},\ldots,a'_i$, ρ is a function of just a_{i-j} and ρ' is a function of just a'_{i-j}. Let G denote $\prod_{p\in[\ell']}F_p|_\rho(b')\cdot\overline{F_p|_\rho}(b'')$. We think of G as a function applied to the random string a_{i-j}, and to show (‡), we need to show that μ_{i-j} δ''-fools G in expectation. Let us now analyze the structure of the statistical test G and show that to δ''-fool G it suffices to fool $\mathcal{C}_{d-1,2k}$ with comparable error.

Say $\rho = \rho(a_1,\ldots,a_{i-j-1},y,a'_{i-j+1},\ldots,a'_i)$ for some $y\in\{0,1\}^n$. Consider $F_p|_\rho(b')$ for any $p\in[t]$ as a function of y. We analyze the complexity of this function. Note that $F_p = \omega^{f_p}$ and hence $F_p|_\rho(b') = \omega^{f_p|_\rho(b')}$. Moreover, for any $y\in\{0,1\}^n$, $f_p|_\rho(b') = f_p(z)$, where z is defined as follows:

- If $j = i$, then for each $s\in[n]$, $z_s = b'_s$ if $a'_{1,s} = \cdots = a'_{i,s} = 1$ and $z_s = y_s$ otherwise.
- If $j < i$, then for each $s\in[n]$,

$$z_s = \begin{cases} a_{0,s} & \text{if } \exists j' \text{ s.t. } 0<j'<j \text{ and } a_{j',s}=0,\\ a_{0,s} & \text{if } \exists j' \text{ s.t. } j<j'\leq i \text{ and } a'_{j',s}=0,\\ a_{0,s} & \text{if } a_{0,s} = b'_s,\\ y_s & \text{if } b'_s = 1 \text{ and } a_{0,s}=0,\\ \neg y_s & \text{if } b'_s = 0 \text{ and } a_{0,s}=1. \end{cases}$$

Thus, we see that $f_p|_\rho(b')$, when considered as a function of y, is simply a projection of the function f_p. Thus $f_p|_\rho(b') = \sum_{j,p}\beta^j_p\tilde{C}^j_p$ where $\tilde{C}^j_p$ is a projection of C^j_p, and hence of C_p, for each $j\in[k]$ and $p\in[\ell']$. Similarly, $\overline{F_p|_\rho}(b'') = \omega^{-f_p|_\rho(b'')}$ and $f_p|_\rho(b'') = \sum_{j,p}\beta^j_p\tilde{\tilde{C}}^j_p$, where $\tilde{\tilde{C}}^j_p$ is a projection of C^j_p, and hence of C_p, for any $j\in[k]$ and $p\in[\ell']$.

As a result, using Fact 3, we see that to δ''-fool G, it suffices to $(\delta''/2)$-fool $(\tilde{C}^j_p,\tilde{\tilde{C}}^j_p : j\in[k],p\in[\ell'])$. However, to fool this tuple, by Lemma 7, it suffices to δ'''-fool all functions of the form $\bigwedge_{(j,p)\in U}\tilde{C}^j_p\wedge\bigwedge_{(j,p)\in V}\tilde{\tilde{C}}^j_p$ for all $U,V\subseteq[k]\times[\ell']$; here, $\delta''' = \delta''/(2\cdot3^{k\ell'}) = \delta''\cdot(\varepsilon_2)^{O(1)} = \delta'\cdot(\varepsilon_2)^{O(1)}$.

Fix any $U,V\subseteq[k]\times[\ell']$ and consider the function $\tilde{C} = \bigwedge_{(j,p)\in U}\tilde{C}^j_p\wedge\bigwedge_{(j,p)\in V}\tilde{\tilde{C}}^j_p$. Since for each j and p, we have $C^j_p = \wedge(C^j_{p,q} : q\in[s_p])$, we

have $\tilde{C} = \bigwedge_{(j,p)\in U} \bigwedge_{q\in[s_p]} \tilde{C}^j_{p,q} \wedge \bigwedge_{(j,p)\in V} \bigwedge_{q\in[s_p]} \tilde{\tilde{C}}^j_{p,q} = \bigwedge_{j\in[k]} \left(\bigwedge_{p\in U_j} \bigwedge_{q\in[s_p]} \tilde{C}^j_{p,q} \wedge \bigwedge_{p\in V_j} \bigwedge_{q\in[s_p]} \tilde{\tilde{C}}^j_{p,q} \right)$ where for each j and p, $\tilde{C}^j_{p,q}$ and $\tilde{\tilde{C}}^j_{p,q}$ are projections of $C^j_{p,q}$ and hence of $C_{p,q}$, and $U_j = \{p \in [\ell'] \mid (j,p) \in U\}$ and $V_j = \{p \in [\ell'] \mid (j,p) \in V\}$.

Clearly, to fool the circuit $\tilde{C}$, it suffices to fool the $2k$-tuple $(\bigwedge_{p\in U_j, q\in[s_p]} \tilde{C}^j_{p,q}, \bigwedge_{p\in V_j, q\in[s_p]} \tilde{C}^j_{p,q} \;:\; j \in [k])$. Since each of the circuits in the tuple are modifications of the depth-$d-1$ circuit $C' = \bigwedge_{p\in[\ell'],q\in[s_p]} C_{p,q}$, we see that to show that μ_{i-j} δ''-fools G in expectation, it suffices to show that μ_{i-j} δ'''-fools $\mathcal{C}_{d-1,2k}$. However, by the choice of μ_{i-j}, we know that μ_{i-j} δ-fools $\mathcal{C}_{d-1,2k}$. Hence, we are done as long as $\delta \leq \delta''' = \delta' \cdot (\varepsilon_2)^{O(1)}$, which is true if δ' is chosen to be $\delta \cdot \left(\frac{1}{\varepsilon_2}\right)^c$ for a large enough constant $c > 0$. This ends the proof of the lemma. ∎

VII. Main proof

In this section, we prove Theorem 12, which implies Theorem 1.

Proof of Theorem 12: We present here the proof of the above theorem using the results from Sections V-B and VI-B. Throughout, we assume w.l.o.g. that $\log(1/\varepsilon)$ and $\log n$ are integers and also that $\varepsilon < 1/n$. (If $\varepsilon > 1/n$, we may set $\varepsilon = 1/n$ and this does not affect the parameters of our PRG significantly.)

We now analyze the above construction. For a fixed $a \in \{0,1\}^n$ and function f defined on $\{0,1\}^n$, we use $f(x \oplus a)$ to denote the function that on input $b \in \{0,1\}^n$ outputs $f(b \oplus a)$. Fix $i \in \{0,\ldots,\log n\}$ and $\eta > 0$. Given function $g_1 : \{0,1\}^n \to R$, where R is a finite set, we say that i is η-good for g_1 if $G^i_{d,k,\varepsilon}$ η-fools $g_1(x \oplus a)$ for each $a \in \{0,1\}^n$. Similarly, if $g_2 : \{0,1\}^n \to \mathbb{C}$, we say that i is η-good for g_2 if $G^i_{d,k,\varepsilon}$ η-fools $g_2(x \oplus a)$ in expectation for each $a \in \{0,1\}^n$. The proof is omitted.

Claim 27. *Assume $g_1 : \{0,1\}^n \to R$ where R is a finite set and $g_2 : \{0,1\}^n \to \mathbb{C}$. Then, for any $\eta > 0$, we have the following: If there is an $i \in \{0,\ldots,\log n\}$ that is η-good for g_1, then $G_{d,k,\varepsilon}$ η-fools g_1. If there is an $i \in \{0,\ldots,\log n\}$ that is η-good for g_2, then $G_{d,k,\varepsilon}$ η-fools g_2 in expectation.*

We now show that the above claim allows us to show that $G_{d,k,\varepsilon}$ ε-fools $\mathcal{C}_{d,k}$ and complete the proof of the induction step. Let $(C^1,\ldots,C^k) \in \mathcal{C}_{d,k}$. Then, there is a circuit $C \in \mathcal{C}_d$ s.t. for each $j \in [k]$, C^j is a modification of C. We proceed by case analysis based on the output gate of C, which is either an AND gate or a $\mathbb{Z}_m$-valued MOD_m gate.

AND case: We first consider the case when the output gate of C is an AND gate. In this case, we show that 0 is ε-good for $(C^1,\ldots,C^k)$, which by Claim 27, would prove what we wanted. Thus, we need to show that $G^0_{d,k,\varepsilon}$ ε-fools $(C^1(x \oplus a),\ldots,C^k(x \oplus a))$ for *any* $a \in \{0,1\}^n$. The following claim that follows from Theorem 17 in Section V-B (via noting that $t \leq n$, since C is read-once), straightaway implies this for $a = 0$.

Claim 28. *If μ is any distribution that δ-fools $\mathcal{C}_{d-1,k}$ for $\delta = 1/n^{c'(km)^5 \log(1/\varepsilon)}$, then μ ε-fools $(C^1,\ldots,C^k)$ where C^j ($j \in [k]$) is a modification of $C \in \mathcal{C}_d$, where the output gate of C is an* AND *gate.*

Let us now see that Claim 28 in fact works for any $a \in \{0,1\}^n$. Fix $a \in \{0,1\}^n$ and define $\tilde{C}^j = C^j(x \oplus a)$ for $j \in [k]$. Note that $\tilde{C}^j$ is a modification of the circuit $\tilde{C} = C(x \oplus a)$ that lies in $\mathcal{C}_d$. Claim 28 implies that $G^0_{d,k,\varepsilon}$ ε-fools $(\tilde{C}^1,\ldots,\tilde{C}^k)$ as well. This shows that 0 is in fact ε-good for $(C^1,\ldots,C^k)$.

MOD_m case: We now consider the case when the output gate of C is a $\mathbb{Z}_m$-valued MOD_m gate. In this case, we assume that $C = \mathrm{MOD}_m(C_1,\ldots,C_t)$ for each $p \in [t]$. Since C^j ($j \in [k]$) is a modification of C, we have $C^j = \mathrm{MOD}_m(C^j_1,\ldots,C^j_t)$, where C^j_p is a projection of C_p for each $p \in [t]$.

In order to show that $G_{d,k,\varepsilon}$ ε-fools $(C^1,\ldots,C^k)$, we use Lemma 8. By Lemma 8, it suffices to show that $G_{d,k,\varepsilon}$ ε_1-fools $f = \sum_{j=1}^k \alpha_j C^j$ for any $\alpha_1,\ldots,\alpha_k \in \mathbb{Z}_m$ and $\varepsilon_1 = \varepsilon/m^{k/2}$. Since $C^j = \mathrm{MOD}_m(C^j_1,\ldots,C^j_t)$, we may write $C^j = \sum_{p=1}^t \gamma^j_p C^j_p$ for $\gamma^j_1,\ldots,\gamma^j_p \in \mathbb{Z}_m$. Substituting in the definition of f above and reordering summations, we see that $f = \sum_{p=1}^t \sum_{j=1}^k \beta^j_p C^j_p$ for some $\beta^j_p \in \mathbb{Z}_m$ where $j \in [k]$ and $p \in [t]$. Define, for each $p \in [t]$, the function $f_p : \{0,1\}^n \to \mathbb{Z}_m$ as follows: $f_p = \sum_{j=1}^k \beta^j_p C^j_p$. Note that the functions f_p depend on pairwise disjoint sets of variables.

By Fact 6, to show that $G_{d,k,\varepsilon}$ ε_1-fools $f = \sum_p f_p$, it suffices to show that for each $\alpha \in \mathbb{Z}_m$, $G_{d,k,\varepsilon}$ ε_2-fools the function $\omega^{\alpha \cdot \sum_{p=1}^t f_p}$ in expectation, where $\varepsilon_2 = \varepsilon_1/2m = \varepsilon/2m^{k/2+1}$. Fix an α and consider $F = \omega^{\alpha \sum_p f_p}$. Let F_p ($p \in [t]$) denote the function $\omega^{\alpha \cdot f_p}$. By Claim 27, to show that $G_{k,d,\varepsilon}$ ε_2-fools F in expectation, it suffices to show that there is some $i \in \{0,\ldots,\log n\}$ that is ε_2-good for F. We show this below.

First, we need the notion of the *weight* of the function F (denoted $\mathrm{wt}(F)$) which is defined as follows: $\mathrm{wt}(F) = \sum_{p=1}^t \mathrm{Var}(F_p)$. Note that for any $a \in \{0,1\}^n$, we have $\mathrm{wt}(F(x \oplus a)) = \mathrm{wt}(F(x \oplus a))$. We proceed to show that there is some i that is ε_2-good for F using one of two separate arguments, based on the value of $\mathrm{wt}(F)$.

Low weight case: The first case is when $\mathrm{wt}(F)$ is somewhat small. Consider the case when $\mathrm{wt}(F) \leq c \log(1/\varepsilon_2)$, where $c = 1000$. In this case, we proceed as in the case of the AND gate and show that 0 is ε_2-good for F. We use the following claim, which readily follows from the above discussion and Theorem 18, proved in Section V-B.

Claim 29. *Assume we have $(C^1,\ldots,C^k) \in \mathcal{C}_{d,k}$ as above, and assume F is obtained from $(C^1,\ldots,C^k)$ also as above. Furthermore, say that $\mathrm{wt}(F) \leq c \log(1/\varepsilon_2)$. Then any*

distribution that δ'-fools $\mathcal{C}_{d-1,k}$ must also ε_2-fool F in expectation, where $\delta' = 1/n^{c'km^4 \log \frac{1}{\varepsilon_2}}$.

Since $G^0_{d,k,\varepsilon}$ fools $\mathcal{C}_{d-1,k}$ with error at most $1/n^{c'(km)^5 \log(1/\varepsilon)} < \delta'$ where δ' is as in the statement of the above lemma, we straightaway know that for $a = 0$, $G^0_{d,k,\varepsilon}$ ε_2-fools $F(x \oplus a)$. However, as in the case of the AND gate, we show that the above lemma actually applies to $F(x \oplus a)$ for all $a \in \{0,1\}^n$. Fix any $a \in \{0,1\}^n$. Then $F(x \oplus a) = \omega^{\alpha \sum_p f_p(x \oplus a)}$ where $f_p = \sum_{j=1}^{k} \beta_p^j C_p^j(x \oplus a)$. Let $\tilde{C}^j = C^j(x \oplus a)$ for $j \in [k]$ and $\tilde{F} = F(x \oplus a)$. Note that for each $j \in [k]$, $\tilde{C}^j$ is a modification of the circuit $\tilde{C} = C(x \oplus a)$ that lies in $\mathcal{C}_d$ and moreover, $\tilde{F}$ is obtained from $(\tilde{C}^1, \ldots, \tilde{C}^k)$ in the same way that F is obtained from $(C^1, \ldots, C^k)$. Finally, note that $\mathrm{wt}(\tilde{F}) = \sum_p \mathrm{Var}(f_p(x \oplus a)) = \sum_p \mathrm{Var}(f_p) = \mathrm{wt}(F) < c \log(1/\varepsilon_2)$. Thus, applying Claim 29 to $(\tilde{C}^1, \ldots, \tilde{C}^k)$ and $\tilde{F}$, we see that $G^0_{d,k,\varepsilon}$ ε_2-fools $\tilde{F}$. Since $a \in \{0,1\}^n$ was arbitrary, we have shown that 0 is in fact ε_2-good for F. This finishes the proof in the case that $\mathrm{wt}(F) \leq c \log(1/\varepsilon_2)$.

High weight case: We now consider the case when $\mathrm{wt}(F)$ is somewhat high. Specifically, assume that $\mathrm{wt}(F) \geq c \log(1/\varepsilon_2)$. We handled this case entirely in Section VI-B. By Lemma 21, it directly follows that there is an i s.t. i is ε_2-good for F. This concludes the inductive step. ■

Acknowledgments

DG is grateful to Pavel Pudlák for helpful discussions. DG acknowledges support by ARO/NSA under grant W911NF-09-1-0569. SL acknowledges support from NSF grant DMS-0835373.

REFERENCES

[1] N. Nisan, "The communication complexity of threshold gates," in *In Proceedings of "Combinatorics, Paul Erdos is Eighty"*, 1994, pp. 301–315.

[2] N. Nisan and A. Wigderson, "Hardness vs. randomness," *J. Comput. Syst. Sci.*, vol. 49, no. 2, pp. 149–167, 1994.

[3] O. Goldreich, *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, 2008.

[4] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[5] J. Hastad, "Almost optimal lower bounds for small depth circuits," in *Proceedings of the 18^{th} annual ACM symposium on Theory of computing (STOC '86)*. New York, NY, USA: ACM, 1986, pp. 6–20.

[6] K. Siu and J. Bruck, "On the power of threshold circuits with small weights," *SIAM Journal on Discrete Mathematics*, vol. 4, no. 3, pp. 423–435, 1991.

[7] M. Goldmann, J. Hastand, and A. Razborov, "Majority gates vs. general weighted threshold gates," in *Proceedings of the 7th Structure in Complexity Theory annual conference*, 1992.

[8] A. A. Razborov, "Lower bounds for the size of circuits of bounded depth with basis $\{\&, \oplus\}$," *Math. Notes Acad. Sci. USSR*, vol. 41, no. 4, pp. 333–338, 1987.

[9] R. Smolensky, "Algebraic methods in the theory of lower bounds for Boolean circuit complexity," in *Proceedings of the 19^{th} annual ACM symposium on Theory of computing (STOC '87)*. New York, NY, USA: ACM, 1987, pp. 77–82.

[10] R. Williams, "Non-uniform ACC circuit lower bounds," in *Conference on Computational Complexity*, 2011.

[11] M. Ajtai, J. Komlós, and E. Szemerédi, "Deterministic simulation in LOGSPACE," in *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, New York City, 25–27 May 1987, pp. 132–140.

[12] L. Babai, N. Nisan, and M. Szegedy, "Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs," *Journal of Computer and System Sciences*, pp. 204–232, 15–17 May 1989.

[13] N. Nisan, "Pseudorandom generators for space-bounded computation," *Combinatorica*, vol. 12, no. 4, pp. 449–461, 1992.

[14] R. Impagliazzo, N. Nisan, and A. Wigderson, "Pseudorandomness for network algorithms," in *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, Montréal, Québec, Canada, 23–25 May 1994, pp. 356–364.

[15] A. Bogdanov, P. Papakonstantinou, and A. Wan, "Pseudorandomness for read-once formulas," in *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, 2011.

[16] S. Lovett, O. Reingold, L. Trevisan, and S. Vadhan, "Pseudorandom bit generators fooling modular sums," in *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*, ser. Lecture Notes in Computer Science. Springer-Verlag, 2009, pp. 615–630.

[17] R. Meka and D. Zuckerman, "Small-bias spaces for group products," in *APPROX-RANDOM*, 2009, pp. 658–672.

[18] A. De, O. Etesami, L. Trevisan, and M. Tulsiani, "Improved pseudorandom generators for depth 2 circuits," 2009, preprint.

[19] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomised Algorithms*. Cambridge University Press, 2009.

[20] L. M. J. Bazzi, "Polylogarithmic independence can fool dnf formulas," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 63–73. [Online]. Available: http://dl.acm.org/citation.cfm?id=1333875.1334182

[21] N. Linial, Y. Mansour, and N. Nisan, "Constant depth circuits, fourier transform, and learnability," *Journal of the ACM*, vol. 40, no. 3, pp. 607–620, 1993.

[22] R. O'Donnell, "Lecture notes on Analysis of Boolean Functions." [Online]. Available: http://www.cs.cmu.edu/ odonnell/boolean-analysis/