New Pseudorandom Generators and Correlation Bounds Using Extractors

Vinayak M. Kumar
University of Texas at Austin, TX, USA

Abstract

We establish new correlation bounds and pseudorandom generators for a collection of computation models. These models are all natural generalization of structured low-degree $\mathbb{F}_2$-polynomials that we did not have correlation bounds for before. In particular:

- We construct a PRG for width-2 $\mathsf{poly}(n)$-length branching programs which read d bits at a time with seed length $2^{O(\sqrt{\log n})} \cdot d^2 \log^2(1/\varepsilon)$. This comes quadratically close to optimal dependence in d and $\log(1/\varepsilon)$. Improving the dependence on n would imply nontrivial PRGs for $\log n$-degree $\mathbb{F}_2$-polynomials. The previous PRG by Bogdanov, Dvir, Verbin, and Yehudayoff had an exponentially worse dependence on d with seed length of $O(d \log n + d2^d \log(1/\varepsilon))$.
- We provide the first nontrivial (and nearly optimal) correlation bounds and PRGs against size-$n^{\Omega(\log n)}$ AC^0 circuits with either $n^{.99}$ SYM gates (computing an arbitrary symmetric function) or $n^{.49}$ THR gates (computing an arbitrary linear threshold function). This is a generalization of sparse $\mathbb{F}_2$-polynomials, which can be simulated by an AC^0 circuit with one parity gate at the top. Previous work of Servedio and Tan only handled $n^{.49}$ SYM gates or $n^{.24}$ THR gates, and previous work of Lovett and Srinivasan only handled polynomial-size circuits.
- We give exponentially small correlation bounds against degree-$n^{O(1)}$ $\mathbb{F}_2$-polynomials which are set-multilinear over *some* arbitrary partition of the input into $n^{1-O(1)}$ parts (noting that at n parts, we recover *all* low degree polynomials). This vastly generalizes correlation bounds against degree-d polynomials which are set-multilinear over a fixed partition into d blocks, which were established by Bhrushundi, Harsha, Hatami, Kopparty, and Kumar.

The common technique behind all of these results is to fortify a hard function with the right type of extractor to obtain stronger correlation bounds for more general models of computation. Although this technique has been used in previous work, they rely on the model simplifying drastically under random restrictions. We view our results as a proof of concept that such fortification can be done even for classes that do not enjoy such behavior.

2012 ACM Subject Classification Theory of computation → Circuit complexity; Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Pseudorandom Generators, Correlation Bounds, Constant-Depth Circuits

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.68

Related Version *Full Version*: `https://eccc.weizmann.ac.il/report/2025/002/`

Funding *Vinayak M. Kumar*: Supported by NSF Grant CCF-2008076, CCF-2312573, and a Simons Investigator Award (#409864, David Zuckerman).

Acknowledgements We thank David Zuckerman for helpful discussions. We also thank anonymous reviewers for helpful comments. We thank Jeffrey Champion, Chin Ho Lee, and Geoffrey Mon for comments on an earlier draft of the paper.

1 Introduction/Outline of Results

Many central questions in complexity theory revolve around proving limitations of various computational models. For example, there are research programs which seek lower bounds against constant depth circuits, low degree polynomials over $\mathbb{F}_2$, and perhaps most famously the complexity class P.

Usually, lower bounds against a simple class of n-bit Boolean functions $\mathcal{C}$ is established by demonstrating an explicit function f such that no $g \in \mathcal{C}$ can compute f on every input. This is referred to as *worst-case hardness*. However, we may not be satisfied with this in practice and stipulate that no $g \in \mathcal{C}$ can even *approximate* f. After all, if there exists a g that agrees with f on all but one point, the difference may be impossible to detect in practice. Furthermore, establishing average case hardness against $\mathcal{C}$ can allow us to create PRGs against $\mathcal{C}$ via the "hardness to randomness" framework introduced by Nisan and Wigderson [24], as well as show hardness results against related function classes, like the majority of functions in $\mathcal{C}$. This *average-case hardness* statement is exactly what the study of correlation bounds capture.

To formally define this, let D a distribution over $\{0,1\}^n$. Define the *correlation* of two Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$ *over* D to be

$$\mathsf{corr}_D(f,g) = |\mathsf{E}_{x\sim D}[(-1)^{f(x)+g(x)}]|.$$

We will usually be concerned with $D = U_n$, the uniform distribution, and should be assumed so if no distribution D is specified. Notice that this quantity is a real number in $[0,1]$. For intuition, note that if $f = g$ or $f = 1 - g$, the correlation is 1, whereas if f and g only match on about half the inputs, the correlation becomes small. This fact allows us to observe correlation is the right notion, as $\mathsf{corr}(f,g)$ being small implies that g cannot predict f much better than a coin flip. For a function f and a function class $\mathcal{C}$, we can define $\mathsf{corr}(f,\mathcal{C}) = \max_{g\in\mathcal{C}} \mathsf{corr}(f,g)$. Hence the notion of f being average-case hard for $\mathcal{C}$ is captured by $\mathsf{corr}(f,\mathcal{C})$ being small.

In this paper, we are most interested in the case $\mathcal{C}$ is the class of low degree $\mathbb{F}_2[x_1,\ldots,x_n]$ polynomials. Establishing correlation bounds against low degree $\mathbb{F}_2$ polynomials is an extremely interesting and central question in complexity theory, as it is either necessary or sufficient to understand a plethora of other problems, some of which concern communication protocols, matrix rigidity, and PRGs for circuits. See Viola's survey [30] for a detailed exposition on this rich program.

Unfortunately, there is a "$\log n$-degree barrier" for PRGs and correlation bounds against low degree polynomials. Current PRGs and correlation bounds are asymptotically tight for constant degree polynomials, but become *trivial* at degree $\log n$ [29]. Getting nontrivial PRGs (or even correlation bounds) against $\log n$-degree polynomials has been a tantalizing and longstanding open problem.

Towards breaking this barrier, researchers have shown strong correlation bounds for *structured* subsets of low degree $\mathbb{F}_2$-polynomials (such as sparse polynomials [20, 26], tensors [3], small-read polynomials, and symmetric polynomials [4]) with the hope of being able to generalize them. In this work, we establish new correlation bounds and PRGs for computation models generalizing some of these polynomials, namely width-2 branching programs reading d bits at a time, AC^0 containing a small number of arbitrary symmetric or linear threshold gates, and set-multilinear polynomials.

Interestingly, all of these correlation bounds are obtained by taking a function hard for a more specific class of polynomials, and then *fortifying* it with a well suited extractor. Although such a fortification technique is not new and has been used for establishing stronger lower bounds for formulas [18, 8], they usually rely on the fact that upon randomly fixing a subset of variables of a formula, there are extremely few possibilities for the resulting function. Our work shows that extractor fortification is a much broader technique that can strengthen lower bounds against function classes even if they do not simplify greatly under a random restriction. In particular, our correlation bounds demonstrate extractor fortification can work if the function class, after a random restriction, has low communication complexity or good algebraic structure.

Inspired by this, we would like to show that extractors will always strengthen correlation bounds, no matter what the proof of the bound is. At a first glance, this may feel intuitive. However, due to technical reasons, this seems challenging to establish.

The remainder of this section is devoted to introducing and motivating each computational model studied, surveying prior work in the topic, and stating all key results proven.

1.1 Better Bounds and PRGs Against AC^0 with More $\{\mathsf{SYM}, \mathsf{THR}\}$ Gates

Our knowledge of hardness and PRG results for AC^0 is far more developed than that of TC^0. Our state of the art PRGs for AC^0 is Lyu's construction [22], which ε-fools polysize AC^0 circuits with seed length $\tilde{O}(\log^{d-1}(n)\log(n/\varepsilon))$, whereas the current best PRG of Hatami, Hoza, Tal, and Tell which (2^{-n^δ})-fools size- $O(n^{1+\delta})$ TC^0 circuits have seed length $O(n^{1-\delta})$ [15]. Due to this stark contrast in parameters, it is natural to gradually work upward from AC^0 by allotting a budget of SYM (calculates an arbitrary symmetric function) or THR (calculates an arbitrary linear threshold function) gates in the circuit. This approach has been explored for more than a decade [28, 20, 26], building upon the study of PRGs for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits pioneered by Luby, Velicković, and Wigderson [21]. This context explains why this circuit class a compelling generalization of sparse polynomials (which can be written as a small-size parity of ands). All the mentioned works use the following function introduced by Razborov and Wigderson in 1993 [25] (all arithmetic is over $\mathbb{F}_2$).

$$\mathsf{RW}_{m,k,r}(x) = \sum_{i=1}^{m} \prod_{j=1}^{k} \sum_{\ell=1}^{r} x_{ij\ell} \tag{1}$$

Most recently, Servedio and Tan [26] use $\mathsf{RW}_{m,k,r}$ to uncorrelate against constant-depth size-$n^{O(\log n)}$ AC^0 circuits whose top gate is $\{\mathsf{SYM}, \mathsf{THR}\}$ (denoted as $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$). Their explicit bound is

$$\mathsf{corr}\left(\mathsf{RW}_{\sqrt{\frac{n}{\log n}}, \log n, \sqrt{\frac{n}{\log n}}},\ \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0\right) \leq 2^{-\Omega(n^{.499})}.$$

Via techniques used in [20], this can be translated to correlation bounds against AC^0 circuits with up to $n^{.499}$ SYM gates or $n^{.249}$ THR gates. As can be surmised by the repeated occurrences of $n^{.499}$, the strength of the correlation bound dictates how many $\{\mathsf{SYM}, \mathsf{THR}\}$ gates we can afford in our budget.

We show that RW is just one of many functions from a general class of hard functions with small correlation against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits. For functions $f : (\{0,1\}^r)^k \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}^r$, denote $f \circ g^k(x_1 \ldots, x_k) := f(g(x_1), \ldots g(x_k))$.

▶ **Theorem 1** (informal). *Let g be computable by a size $n^{O(\log n)}$ $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit. Let f be average-case hard against multiparty protocols*[1]*, and let* Ext *be a suitable extractor. Then*

$$\mathsf{corr}(f \circ \mathsf{Ext}^{\cdot 01 \log n}, g) \leq 2^{-\Omega(n^{.999})}.$$

To our knowledge, this theorem gives the first context where generically precomposing with an extractor boosts correlation bounds whose proof does not rely on simplification under random restriction (indeed parity does not simplify under restriction and is contained in

[1] the formal condition is any function with small "k-party norm" or "cube norm", but this is currently the only technique we know that establishes average case hardness against multiparty protocols.

Table 1 Correlation bounds against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0_d$ circuits and the PRGs that follow via the [24] framework. In all previous work, the "hard" function used was the RW function, which was first considered by Razborov and Wigderson [25]. Our work uses a better suited function. This table is an extension of the one found in [26].

	Circuit type	Circuit size S	Correlation bound	PRG seed length
[28]	$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$	$n^{c \log n}$	$n^{-c_d \log n}$	$2^{O(\sqrt{\log(S/\varepsilon)})}$
[20]	$\mathsf{SYM} \circ \mathsf{AC}^0$	$n^{c \log\log n}$	$\exp(-n^{0.999})$	$2^{O\left(\frac{\log S}{\log\log S}\right)} + (\log(1/\varepsilon))^{2.01}$
[20]	$\mathsf{THR} \circ \mathsf{AC}^0$	$n^{c \log\log n}$	$\exp(-n^{0.499})$	$2^{O\left(\frac{\log S}{\log\log S}\right)} + (\log(1/\varepsilon))^{4.01}$
[26]	$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0_d$	$n^{c \log n}$	$\exp(-\Omega(n^{0.499}))$	$2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{4.01}$
This work	$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$	$n^{c \log n}$	$\exp(-\Omega(n^{0.999}))$	$2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{2.01}$

$\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0\}$). Previously, extractors have only been used to boost correlation bounds for classes that heavily simplify under random restriction [18, 8].[2] Our theorem states that extractors can still boost correlation bounds, even if they were proven using communication complexity rather than random restrictions.

Furthermore, our theorem distills out the reason why RW was so effective as a hard function. Quantitatively, we can instantiate the template with a suitable extractor to obtain a new hard function with nearly-optimal correlation bounds.

Due to our strengthened correlation bounds, we can now get correlation bounds and PRGs against size-$n^{O(\log n)}$ AC^0 circuits with up to $n^{.999}$ SYM gates or $n^{.499}$ THR gates. Prior to this, no nontrivial correlation bound or PRG was known to handle such large size *and* number of $\{\mathsf{SYM}, \mathsf{THR}\}$ gates ([20] could handle the same number of $\{\mathsf{SYM}, \mathsf{THR}\}$ gates but only for $n^{O(\log\log n)}$-size circuits, and [26] could handle the same size circuits, but only $n^{.499}$ SYM or $n^{.249}$ THR gates).

Even for $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits which have only one $\{\mathsf{SYM}, \mathsf{THR}\}$ gate, our correlation bounds yields improved PRGs whose seed length is $2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{2.01}$, which has a better dependence on ε, than previous work (see Table 1). In fact, since the best correlation bound one can hope for is $2^{-\Omega(n)}$, this dependence is almost optimal under the Nisan-Wigderson framework, and an alternative approach is needed to reach the optimal dependence of $\log(1/\varepsilon)$. Since any $\log n$-degree $\mathbb{F}_2$ polynomial can be expressed as a $\mathsf{SYM} \circ \mathsf{AND}_{\log n}$ circuit of size $n^{\log n}$, any improvement of the dependence of the seed length on S would give nontrivial PRGs for $\log n$-degree polynomials, a breakthrough result.

1.2 Much Better PRGs Against Width-2 Branching Programs Reading d Bits at a Time

Usually, one constructs PRGs for natural *computational* models, with the idea that we can drastically reduce the randomness we use if the randomized algorithm we are running can be simulated by such a model. Low degree polynomials is an extremely natural mathematical model with applications to circuit complexity, but some may not believe it is well grounded as

[2] There have been uses of extractors as a hard function against classes that do not simplify under restriction, like DNFs of Parities [9] and strongly read-once linear branching programs [12, 19, 7]. However they directly establish a correlation bound against the extractor rather than amplify a weaker hard function by precomposing with an extractor.

a computational one and thus not worth finding a PRG for. However, the work of Bogdanov, Dvir, Verbin, and Yehudayoff [5] showed the beautiful connection that PRGs for degree d polynomials are also PRGs against a particular model described as *width-2 length-*$\mathsf{poly}(n)$ *branching programs which read* d *bits at a time.*

▶ **Definition 2** ((d, ℓ, n)-2BP ([5], adapted)). *A (d, ℓ, n)-2BP (or more colloquially a width-2 length-ℓ branching program over n bits which reads d bits at a time) is a layered directed acyclic graph, where there are ℓ layers and each layer contains two nodes, which we label by 0 and 1. Each vertex in each layer j is associated with an arbitrary d-bit substring $x|_v$ of the input x. Each node in layer j has 2^d outgoing edges to layer $j+1$ that are labelled by all possible values in $\{0,1\}^d$. On input x, the computation starts with the first node v_{start} in the first layer, then follows the edge labelled by $x|_{v_{start}}$ onto the second layer, and so on until a node in the last layer is reached. The identity of this last node is the outcome of the computation.*

Such branching programs are a well motivated computation model which cover computation with only one bit of usable memory, low degree polynomials, and small width DNFs. The survey of unconditional PRGs by Hatami and Hoza refer to this model as a compelling computational model that places low degree polynomials in the computational landscape [14].

Unfortunately, there is a "$\log n$-degree barrier" for PRGs and correlation bounds against low degree polynomials. Current PRGs and correlation bounds are asymptotically tight for constant degree polynomials, but become *trivial* at degree $\log n$, as can be seen by the current best known PRG for degree-d polynomials by Viola which has seed length $O(d \log n + d2^d \log(n/\varepsilon))$ [29]. Getting nontrivial PRGs (or even correlation bounds) against $\log n$-degree polynomials has been a tantalizing and longstanding open problem, and thus PRGs for $(d, \mathsf{poly}(n), n)$-2BPs also seemingly appeared to inherit this "$d = \log n$ barrier" due to the reduction result of [5].

In this work, we construct PRGs against $(d, \mathsf{poly}(n), n)$-2BPs with exponentially better seed length, thereby giving nontrivial PRGs even in the regime $d = n^{1-o(1)}$. Define a d-junta to be a function $\phi : \{0,1\}^n \to \{0,1\}$ which is solely dependent on d input bits (i.e. can be written as $\phi'(x_i)_{i \in S}$ for some subset $S \subset [n]$ of size d). To get our shortened seed length, we evade the $\log n$-degree barrier by instead showing the equivalence between PRGs for $(d, \mathsf{poly}(n), n)$-2BPs and PRGs for the XOR of $\mathsf{poly}(n)$ many d-juntas (denoted as $\mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}$). This class is already interesting in its own right, as it can be seen as a generalization of sparse $\mathbb{F}_2$-polynomials and combinatorial checkerboards (defined by Watson [32] and also studied by Gopalan, Meka, Reingold and Zuckerman [11]), as well as a specific class bounded collusion protocols studied by Chattopadhyay et al. [6]. However, we are not aware of any literature studying $\mathsf{JUNTA}_{n,d}^{\oplus m}$ specifically.

Our main technical contribution is strong correlation bounds for $\mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}$. In particular, we show the following.

▶ **Theorem 3.** *There exists an explicit function f such that*

$$\mathsf{corr}(f, \mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}) \leq \exp\left(-\frac{n}{d2^{O(\sqrt{\log n})}}\right)$$

By combining this with the "hardness-to-randomness" framework of Nisan and Wigderson [24], we construct a PRG of seed length $2^{O(\sqrt{\log n})} d^2 \log^2(1/\varepsilon)$. This is only a quadratic factor away from optimal dependence on d and ε. Improving the dependence on n would be a breakthrough, since if we set $n' = 2^{\sqrt{\log n}}$, a $(d, n, \mathsf{poly}(n))$-2BP can simulate any $\log n'$-degree polynomial over $x_1, \ldots x_{n'}$, and so having seed length $o(n')$ would effectively be breaking the $\log n$-degree barrier for $\mathbb{F}_2$-polynomial PRGs.

Interestingly enough, by combining an "simplification under restriction" approach pioneered by Ajtai and Wigderson [1] with a PRG for sparse $\mathbb{F}_2$-polynomials by Servedio and Tan [27], we are able to construct a PRG against $\mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}$, and thus $(d, \mathsf{poly}(n), n)$-2BPs, with seed length $d2^{O(\sqrt{\log(n/\varepsilon)})}$. This gives us an optimal dependence on d, but an exponentially worse dependence on ε. This suggests perhaps with a combination of these two approaches, one might be able to achieve seed length $2^{O(\sqrt{\log n})} d \log(1/\varepsilon)$.

1.3 Near-Optimal Bounds Against High Degree Set-Multilinear Polynomials

As explained earlier, a central open question in complexity theory is to establish better-than-$O(1/\sqrt{n})$ nontrivial correlation bounds against $\Omega(\log n)$-degree polynomials. In order to make progress on this question, it is natural to consider structured low-degree $\mathbb{F}_2$-polynomials. This is what the work of Bhrushundi et al. does [3].

Define a polynomial $p : \{0,1\}^n \to \{0,1\}$ to be set-multilinear over a partition $X = (X_1, \ldots, X_d)$ of the input bits if every monomial contains *at most one* variable from each X_i (this is slightly more general than the usual definition of *exactly one*). The work of Bhrushundi et al. [3] prove that a random degree d set-multilinear tensor has exponentially small correlation against generic degree $d/2$ $\mathbb{F}_2$-polynomials for $d = \Omega(n)$. Towards making this correlation bound explicit, they defined $\mathsf{FFM}(X_1, \ldots, X_d) = \mathsf{lsb}(X_1 \cdot X_2 \cdots X_d)$, where multiplication is done by treating the X_i as field elements, and lsb outputs the least significant bit of the string. Bhushrundi et al. were able to give exponentially small correlation bounds against polynomials up to degree $o(n/\log n)$ which are set-multilinear over the fixed partition $(X_1, \ldots, X_d)$. However, this leaves more to be wanted. The partition with respect to which the polynomial is set-multilinear over needing to be fixed and dependent on FFM_d feels like an extremely strong and asymmetric condition. Can we uncorrelate against degree $< d$ polynomials set-multilinear over *any* equipartition of X into d parts? Can the parts be unequal? Can we have more than d of them?

We show the affirmative to all the above questions. If we take $\delta > 0$ to be an arbitrarily small constant, we can obtain exponentially small correlation against degree $< n^\delta$ polynomial for which there exists *some* partition of X into up to $n^{1-\delta}$ (not necessarily equal) parts such that p is set-multilinear over it. Notice improving $n^{1-\delta}$ parts to n would be a breakthrough, since all polynomials are set-multilinear over the n-partition of $X = (x_1, \ldots, x_n)$.

To do so, we fortify the hard function FFM with an extractor. Let $\mathsf{Ext}(X, W)$ be a strong linear seeded extractor (for each fixing of W, $\mathsf{Ext}(\cdot, W)$ is linear). For some parameter d, define the function

$$\mathsf{ExtFFM}_d(X_1, \ldots, X_d, W) := \mathsf{lsb}(\mathsf{Ext}(X_1, W) \cdot \mathsf{Ext}(X_2, W) \cdot \ldots \cdot \mathsf{Ext}(X_d, W)),$$

where multiplication is done over a finite field, and lsb outputs the least significant bit of the string. First note that $\mathsf{ExtFFM}_d(X_1, \ldots, X_d, W)$, for a fixed W, is set-multilinear over $X_1, \ldots, X_d$. Hence our intuition that set-multilinear polynomials might correlate the most with the hard function is preserved in ExtFFM as well. Using ExtFFM, we are able to obtain correlation bounds against the more intuitive notion of set-multilinear polynomials, where the structure of the partition does not matter. This gives more leeway since now if we want to implement this approach towards correlation bounds against low-degree polynomials, there is a larger class of set-multilinear polynomials that we can reduce generic polynomials to.

2 Technical Overview Of the Results

In this section, we give the overview of the proofs of the main results we covered above.

2.1 Stronger Correlation Bounds Against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$

We focus on showing stronger correlation bounds against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$, since the subsequent arguments turning this into PRGs against AC^0 with a few $\{\mathsf{SYM}, \mathsf{THR}\}$ gates are standard. The blueprint behind this argument follows the "simplification under restrictions" approach of previous works, but most similarly of Tan and Servedio [26]. A random restriction is a random partial assignment where for each variable, it is left unfixed (or "alive") with probability p, and is otherwise set to a uniform bit. [26] shows that under a random restriction, the hard function $\mathsf{RW}_{m,k,r}$ maintains integrity and uncorrelates against multiparty protocols, while the $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ simplifies to a short multiparty protocol. However, the roadblock met in [26] preventing a correlation bound of $2^{-\Omega(n)}$ and only giving one of size $2^{-\Omega(n^{.499})}$ is due to parameters in $\mathsf{RW}_{m,k,r}$ being in contention with each other. To elucidate, if n is the input size, then we must have $mkr = n$. Via the analysis done in [26], the correlation bound ends up being in the form of $2^{-\Omega(m)} + 2^{-\tilde{\Omega}(r)}$, which forces any established correlation bound to be at best $2^{-\Omega(\sqrt{n})}$.

To understand why both conflicting terms show up, we give a quick overview of the argument of [26]. First, $\mathsf{RW}_{m,k,r}$ (as defined in Equation (1)) can be thought of as a fortified version of the *generalized inner product*, $\mathsf{GIP}_{m,k}(x_1, \ldots, x_k) := \sum_{i=1}^{m} \prod_{j=1}^{k} x_{ij}$, where each variable is now replaced by the parity of r new variables. This is effective against random restrictions, since as long as one of the r copies $x_{ij1}, \ldots, x_{ijr}$ survive the restriction, the corresponding term x_{ij} in GIP will survive. They argue that after a random restriction ρ is applied, the $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit simplifies to a short multiparty protocol, while $\mathsf{RW}_{m,k,r}|_\rho$ is still capable of computing $\mathsf{GIP}_{m/2,k}$ with high probability. Conditioning upon this, previous results of Babai, Nisan, and Szegedy [2] show that $\mathsf{GIP}_{m/2,k}$ has $2^{-\Omega(m/2^k)}$ correlation against these multiparty protocols, explaining the emergence of the $2^{-\Omega(m)}$ term in the correlation. Conditioning on $\mathsf{RW}_{m,k,r}|_\rho$ being able to compute $\mathsf{GIP}_{m/2,k}$ introduces an additive error to the correlation corresponding to the probability $\mathsf{RW}_{m,k,r}|_\rho$ *fails* to simplify. [26] bounds this by the chance all r copies of some variable x_{ij} becomes fixed by a restriction, which will be $(1-p)^r \approx \exp(-pr)$, explaining the occurrence of the $2^{-\Omega(r)}$ term in the correlation.

In summary, the argument of [26] requires r needs to be large to strongly fortify the hard function against random restrictions, while m needs to be large to have a stronger correlation bound against multiparty protocols. However, with the constraint $mr \leq n$, we are forced to compromise and reach the setting $m = r \approx \sqrt{n}$.

We now propose an abstraction of the hard function, which naturally yields a stronger correlation bound. If we define $\oplus_{m,r} : (\{0,1\}^r)^m \to \{0,1\}^m$ to be

$$\oplus_{m,r}(x_1, \ldots, x_m) = \left(\sum_{i=1}^{r} x_{1i}, \ldots, \sum_{i=1}^{r} x_{mi} \right),$$

we observe $\mathsf{RW}_{m,k,r} = \mathsf{GIP}_{m,k} \circ \oplus_{m,r}^k(x_1, \ldots, x_k) := \mathsf{GIP}_{m,k}(\oplus_{m,r}(x_1), \ldots, \oplus_{m,r}(x_k))$. The key insight is that our argument can be generalized to not just RW, but any function

$$f \circ \mathsf{Ext}^k := f(\mathsf{Ext}(x_1), \ldots, \mathsf{Ext}(x_k))$$

where f is average-case hard for multiparty protocols, and Ext is an *oblivious bit-fixing source extractor* (OBF extractor). Informally, an oblivious bit-fixing source extractor for min-entropy k is a function Ext such that if $\mathbf{X}$ is uniform over $\{0,1\}^n$ and ρ is a restriction which leaves $\geq k$ bits alive, the output $\mathsf{Ext}(\mathbf{X}|_\rho)$ is close to uniform. Recall our approach first applies a random restriction to simplify our circuit to a small multiparty protocol, which we then deal with using GIP. If the random restriction leaves sufficiently many variables alive with high probability, then $f \circ \mathsf{Ext}^k$ should still behave like f due to Ext being an OBF extractor. Since the circuit is now a multiparty protocol, the average-case hardness of f gives us a correlation bound.

Notice in the RW construction and the setting of parameters $m = r \approx \sqrt{n}$, $\oplus_{m,r}$ is an OBF extractor which maps n bits to $\sqrt{n}$ bits. But this means the input to the outer GIP function will only have $\approx \sqrt{n}$ bits, and so the best correlation bound we can hope to achieve is $\exp(-\Omega(\sqrt{n}))$. The restrictions used in the proof leave $n^{.99}$ variables alive with high probability, so intuitively we could hope that all these $n^{.99}$ "bits of randomness" could be preserved for GIP (or in general any f) rather than only $\sqrt{n}$, potetially resulting in a $\exp(-\Omega(n^{.99}))$ correlation bound alive. We do just this by using a much better OBF extractor of Kamp and Zuckerman [17]. By making this intuition more formal using techniques developed by Viola and Wigderson [31], we obtain $2^{-\Omega(n^{1-O(1)})}$ correlation bound. The idea of replacing parities with better suited extractors has also appeared in previous work [18, 8].

2.2 PRGs for $\mathsf{JUNTA}^{\oplus t}_{n,d}$ and (d,t,n)-2BPs

Our PRG construction blueprint can be briefly described as follows. We first establish correlation bounds against $\mathsf{JUNTA}^{\oplus t}_{n,d}$. We then put this through the Nisan-Wigderson "hardness vs. randomness" framework to create a PRG against $\mathsf{JUNTA}^{\oplus t}_{n,d}$. We then show that PRGs which fool $\mathsf{JUNTA}^{\oplus t}_{n,d}$ actually fool (d,t,n)-2BPs, making the $\mathsf{JUNTA}^{\oplus t}_{n,d}$ PRG our final construction. We first discuss why PRGs for $\mathsf{JUNTA}^{\oplus t}_{n,d}$ imply PRGs for (d,t,n)-2BPs, and then discuss the techniques needed to show strong correlation bounds against $\mathsf{JUNTA}^{\oplus t}_{n,d}$.

2.2.1 PRGs for $\mathsf{JUNTA}^{\oplus t}_{n,d} \implies$ PRGs for (d,t,n)-2BPs

Adopting the exposition in [14], the previous work of [5] can be outlined as follows. Consider a (d,t,n)-2BP B. By noticing that all transition functions in B are d-juntas, one can derive that $B(x) = B'(\phi_1, \ldots, \phi_{2t}(x))$, where B' is a $(1,t,2t)$ branching program. By Fourier expanding B', this can be decomposed as

$$B(x) = \sum_{S \subset [2t]} \widehat{B}(S)(-1)^{\sum_{i \in S} \phi_i(x)}.$$

[5] shows that $\sum_{S \subset [t]} |\widehat{B}(S)|$ is bounded , so by linearity of expectation and the Triangle Inequality, it suffices to fool the terms $(-1)^{\sum_{i \in S} \phi_i(x)}$. The approach in [5] makes the observation that each ϕ_i, by virtue of being a d-junta, can be written as a degree d polynomial. Consequently, a PRG for degree d polynomials will fool (d,t,n)-2BPs with seed length $O(d \log n + d2^d \log(n/\varepsilon))$. The issue here is that at $d = \log n$, the seed length becomes trivial.

However, we can notice that the $\mathbb{F}_2$-polynomial $p(x) \coloneqq \sum_{i \in S} \phi_i(x)$ has some additional structure. If $t = \mathsf{poly}(n)$, p is the sum of only a polynomial number of d-juntas. If there was a way to leverage this, and get a better PRG that fools $\mathsf{JUNTA}^{\oplus t}_{n,d}$, then we might hope to get nontrivial PRGs even in the regime $d = \Omega(\log n)$.

This observation already yields nontrivial PRGs for $d = \omega(\log n)$. Servedio and Tan [26] provide a PRG fooling $\mathbb{F}_2$-polynomials with S terms with seed length $2^{O(\sqrt{\log S}}\log(1/\varepsilon))$. Since each junta can be written as a polynomial with up to 2^d terms, each $g \in \mathsf{JUNTA}_{n,d}^{\mathsf{poly}(n)}$ can be written as a polynomial with $S = 2^d\mathsf{poly}(n)$ terms, yielding a PRG with seed length $(2^{O(\sqrt{d})}+O(\log n))\log(1/\varepsilon)$. Hence we get nontrivial seed length for $d = o(\log^2 n)$. [3] However, we proceed alternatively to get an exponentially better seed length.

2.3 The Nisan-Wigderson Framework and Correlation Bounds for $\mathsf{JUNTA}_{n,d}^{\oplus\mathsf{poly}(n)}$

We will once again use $f \circ \mathsf{Ext}^k$ as our hard function[4] to establish exponentially small correlation bounds against the class, and then apply the Nisan-Wigderson [24] framework to construct the PRG. The latter portion is straightforward, so we focus on establishing the correlation bounds.

Let $g \in \mathsf{JUNTA}_{n,d}^{\oplus\mathsf{poly}(n)}$. We first show that there exist a subset of variables, S, such that upon arbitrarily fixing bits outside of this set, g can be expressed as a sparse $\mathbb{F}_2$ polynomial, whereas each input block of $f \circ \mathsf{Ext}^k$ heavily intersect S. Hence if we fix $X_{\bar{S}}$ and take the correlation over S, each input block still maintains high min-entropy while g becomes a sparse polynomial, which is a small $\mathsf{SYM} \circ \mathsf{AC}^0$ circuit. Since the hard function is also the same, we can then apply techniques in the previous section to conclude.

2.4 Correlation Bounds against Set-Multilinear Polynomials

Recall that [3] has shown FFM_d uncorrelates against any lower degree polynomial which is set-multilinear over $(X_1, \ldots, X_d)$. The key ingredient behind proving strong correlation bounds against set-multilinear polynomials over arbitrary parititons is to first fortify each input block with extractors, and instead consider ExtFFM_d. This allows us to establish the following structural lemma, which intuitively states that even if you do not start out with a polynomial that is set-multilinear over $(X_1, \ldots, X_d)$, if not too many bits in each input block can be restricted to 1s such that the resulting function is set-multilinear over $(X_1, \ldots, X_d)$ induced by the live variables in each block, exponential correlation bounds can still be obtained.

▶ **Theorem 4.** *Let g be a polynomial of degree $< d$. Let $S_1, \ldots, S_d \subset [n/d]$ be subsets, and let ρ denote the restriction created by fixing the bits in X_i whose index is outside S_i to 1 for each $i \in [d]$. If the restricted function $g|_\rho(X_1, \ldots, X_d)$ becomes set-multilinear in $(X_1, \ldots, X_d)$, then have*

$$\mathsf{corr}(\mathsf{ExtFFM}_d, g) \leq 2^{-\Omega(\frac{n}{cd})}.$$

To explain the proof at a high level, if the sets S_i we leave alive aren't too small, then our strong extractor (conditioned on a good seed) will keep each block $\mathsf{Ext}(X_i, W)$ approximately uniform, and since the restricted function $g|_\rho$ is now set-multilinear over $(X_1, \ldots, X_d)$ we may use a similar approach as [3] to prove the theorem.

It turns out that via a combinatorial argument, one can show that polynomials which are set-multilinear over a large number of blocks can be turned into polynomials set-multilinear over $(X_1, \ldots, X_d)$ by fixing not too many bits per input block X_i. The correlation bounds then follow by the structural lemma.

[3] it is actually the case that a PRG from [21] already gets nontrivial seed length in the same regime, albeit with exponentially worse dependence in ε

[4] we also precompose with parities in the formal argument

3 Preliminaries

For positive integer n, $[n] := \{1, \ldots, n\}$ and $\binom{[n]}{s}$ is the set of all subsets of $[n]$ with $|S| = s$. We denote $e(x) := (-1)^x$.

3.1 Convention About Input Blocks

We will canonically fix a partition of bit strings into d contiguous blocks, each with n/d bits. In particular, any $X \in \{0,1\}^n$ can be written as $X = (X_1, \ldots, X_d)$ where each X_i is the n/d-bit substring. If a string $Y \in \{0,1\}$ is defined, Y_i will be assumed to mean the length n/d substring of Y contained in the ith input block, defined with respect to the canonical partition. Also, we will denote $X_{-i} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$ to be the input with the ith block removed.

For a string $X \in \{0,1\}$, we may sometimes identify the n/d bit string X_i as an n-bit string in the following way: the ith block is filled with X, and all other blocks are filled with 0s. Hence, if we interpret bit strings as elements of $\mathbb{F}_2^n$, and we have $X, Y \in \mathbb{F}_2^n$, the expression $X + Y_i$ is well defined.

For parameters $k, d \leq n$ and two functions $f : (\{0,1\}^m)^k \to \{0,1\}$ and $g : \{0,1\}^{n/k} \to \{0,1\}^d$, we will define

$$f \circ g^k = f(g(X_1), \ldots, g(X_d)).$$

3.2 Finite Fields

We will be working with finite fields of characteristic 2. For the finite field over 2^n elements, $\mathbb{F}_{2^n}$, we can naturally identify each element with an n-bit string.

▶ **Definition 5** (character). *A map $\chi : \mathbb{F}_{2^n} \to \mathbb{F}_2$ is called an* additive character *if for all $x, y \in \mathbb{F}_{2^n}$, $\chi(x + y) = \chi(x) + \chi(y)$. It is nontrivial if it is not the zero function.*

Since $\mathbb{F}_{2^n}$ is an n-dimensional vector space, we see the valuations on n basis vectors uniquely define the character. Consequently there are 2^n such characters. Notice we can conveniently characterize all characters either by $\chi_c(x) = \langle x, c \rangle$, or by fixing some character χ, and then defining $\chi_c(x) := \chi(c \cdot x)$. This can be seen by verifying these maps are characters, are distinct, and that there are 2^n of them (the latter is obvious since there are 2^n values of c).

3.3 Models of Computation

▶ **Definition 6** ($\mathbb{F}_2$-polynomials). *An $\mathbb{F}_2$-polynomial (or polynomial for short) is a function of the form $p(x) := \sum_{S \subset [n]} c_S \prod_{i \in S} x_i$ for some $c_i \in \mathbb{F}_2$ (all arithmetic here are over $\mathbb{F}_2$).*

▶ **Definition 7** (set-multilinearity). *An $\mathbb{F}_2$-polynomial p is* set-multilinear *over a partition $(X_1, \ldots, X_d)$ of variables if every monomial of p contains at most one variable from each X_i. Notice that all polynomials are trivially set-multililinear over $(x_1, \ldots, x_n)$.*

▶ **Definition 8** (junta). *Define the class $\mathsf{JUNTA}_{n,k}$ to be a function $\phi : \{0,1\}^n \to \{0,1\}$ which is solely dependent on k input bits (i.e. can be written as $\phi'(x_i)_{i \in S}$ for some subset $S \subset [n]$ of size k). Define $\mathsf{JUNTA}_{n,k}^{\oplus t}$ to be the class of functions which is the parity of t k-juntas.*

▶ **Definition 9** (k-party NOF protocol). *A boolean function $f : (\{0,1\}^{n/d})^d$ can be computed by a k-party NOF protocol with c bits of communication if on input $X = (X_1, \ldots, X_d)$, d players, can take turns writing a bit on the board, where player i's bit can only depend on X_{-d} and the other bits on the board, and the cth bit written is $f(X)$. We denote this class of functions to be Π_k^c.*

Circuits

We measure the size of a circuit by the total number of wires (including input wires) in it. AC^0_d are depth d circuits with unbounded fan-in whose gate set is $\{\mathsf{AND}, \mathsf{OR}, \mathsf{NOT}\}$. SYM is a gate which computes an arbitrary symmetric function, and THR is a gate which computes an arbitrary linear threshold function. In general, if we have a gate G, a subscript G_k will refer to its fan-in (in this case, G is fixed to have fan-in k).

▶ **Definition 10** ($(d, \mathcal{C})$-tree). *Let d be an integer and $\mathcal{C}$ a computational model (e.g. a circuit class). A function is computable by a $(d, \mathcal{C})$-tree if it is computable by a depth t decision tree with $\mathcal{C}$ functions as its leaves. That is, there exists a depth d decision tree T such that for every path π in T, $F|_\pi \in \mathcal{C}$.*

3.4 Probability

We will denote U_m to be the uniform distribution over the finite set $\{0,1\}^m$. We will also denote $S \subset_p T$ to be a random subset of T where each $t \in T$ is added to S independently with probability p.

▶ **Definition 11** (k-wise uniform). *Consider a distribution D over $(\{0,1\}^{n/d})^d$. We say that D is k-wise uniform if for all subsets $S = \{i_1, \ldots, i_k\} \subset [d]$ and all strings $y_1, \ldots, y_k \in \{0,1\}^{n/d}$,*

$$\Pr_{X \sim D}[\forall j, X_{i_j} = y_j] = 2^{-kn/d}.$$

▶ **Definition 12** (ε-close in distribution). *Let D_1 and D_2 be distributions over $\{0,1\}^n$. We say $D_1 \approx_\varepsilon D_2$, or equivalently D_1 is ε-close to D_2, if for all $S \subset \{0,1\}^n$,*

$$|\Pr_{x \sim D_1}[x \in S] - \Pr_{x \sim D_2}[x \in S]| \leq \varepsilon.$$

3.5 Random Restrictions and Partial Assignments

A *partial assignment* or *restriction* is a string $\rho \in \{0,1,\star\}^n$. Intuitively, a $\star$ represents an index that is still "alive" and hasn't been fixed to a value yet.

We also define a composition operation on partial assignments. For two restrictions ρ^1, ρ^2, define $\rho^1 \circ \rho^2$ so that

$$(\rho^1 \circ \rho^2)_i = \begin{cases} \rho^1_i & \rho^1_i \neq \star \\ \rho^2_i & \rho^1_i = \star. \end{cases}$$

Intuitively, one can see this as fixing bits determined by ρ^1 first, and then out of the remaining alive positions, fix them according to ρ^2.

A *random restriction* is simply a distribution over restrictions. A common random restriction we will use is R_p, the distribution where each index will be assigned $\star$ with probability p, and $0, 1$ each with probability $\frac{1-p}{2}$.

The main reason for defining restrictions is to observe their action on functions. Given a restriction ρ and function $f : \{0,1\}^n \to \{0,1\}$, we define $f|_\rho : \{0,1\}^n \to \{0,1\}$ to be the restricted function mapping $f|_\rho(x) := f(\rho \circ x)$.

3.6 Pseudorandomness

Our work will involve working with pseudorandomness primitives, like pseudorandom generators (PRGs) and randomness extractors (or simply extractors).

▶ **Definition 13** (ε-PRG). *A polytime computable function $G : \{0,1\}^s \to \{0,1\}$ is an ε-PRG for a subset $\mathcal{F}$ of functions $\{0,1\}^n \to \{0,1\}$ if for all $f \in \mathcal{F}$,*

$$|\mathsf{E}_{x\sim U_n}[(-1)^{f(x)}] - \mathsf{E}_{s\sim U_s}[(-1)^{f(G(s))}]| \leq \varepsilon.$$

We also say that G ε-fools $\mathcal{F}$. The parameter s is the seed length. In this paper, we will use a PRG of [27] which ε-fools $\mathbb{F}_2$ polynomials with $\leq S$ terms with seed length $2^{O(\sqrt{\log S})}\log(1/eps)$.

▶ **Definition 14** (min-entropy). *Let D be a distribution over $\{0,1\}^n$, and define $supp(D) = \{y \in \{0,1\}^n : \Pr_{x\sim D}[x = y] > 0\}$. Define the min-entropy of D to be the quantity*

$$-\log\left(\max_{x\in\{0,1\}^n} \Pr_{y\sim D}[y = x]\right).$$

It is helpful to note that if for a particular k and all $y \in \{0,1\}^n$, all probabilities $\Pr_{x\sim D}[x = y] \leq 2^{-k}$, then we know D has min-entropy $\geq k$.

▶ **Definition 15** (Strong/Linear/Seeded Extractors). *A (k,ε)-seeded extractor is a function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ such that for any D with min-entropy $\geq k$, we have for $\boldsymbol{X} \sim D$ and $\boldsymbol{W} \sim U_d$ the following*

$$\mathsf{Ext}(\boldsymbol{X}, \boldsymbol{W}) \approx_\varepsilon U_m.$$

Ext is a strong seeded extractor if we also have

$$\Pr_{w\sim U_d}[\mathsf{Ext}(\boldsymbol{X}, w) \approx_\varepsilon U_m] \geq 1 - \varepsilon$$

Ext is a linear seeded extractor if for every fixed W, $\mathsf{Ext}(\cdot, W)$ is linear over $\mathbb{F}_2$. The Leftover Hash Lemma [16] allows us to construct a strong seeded (k,ε) extractor with seed length $2n$, $\mathsf{Ext} : \{0,1\}^n \cdot \{0,1\}^{2n} \to \{0,1\}^{k-2\log(1/\varepsilon)}$.

▶ **Definition 16** (Oblivious Bit-Fixing Source Extractors). *An (n,k) oblivious bit-fixing source (or OBF) is a distribution D over $\{0,1\}^n$ created by fixing some $n-k$ of the bits, and then filling in the remaining k indices with uniform and independent bits. An (k,ε) oblivious bit-fixing source extractor (or OBF extractor) is a function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ such that for every (n,k) OBF D, we have that for $\boldsymbol{X} \sim D$,*

$$\mathsf{Ext}(\boldsymbol{X}) \approx_\varepsilon U_m.$$

For any $k > \sqrt{n}$, Kamp and Zuckerman [17] allows us to construct $(k, 2^{-\Omega(k^2/n)})$ OBF extractors $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^{\Omega(k^2/n)}$.

3.7 Correlation Bounds

We will need some tools and definitions from the literature of correlation bounds. We first give a formal definition of correlation.

▶ **Definition 17** (correlation). *For two Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$, and a distribution D over $\{0,1\}^n$, define the correlation of f and g over D to be*

$$\mathsf{corr}_D(f,g) = |\mathsf{E}_{x\sim D}(-1)^{f(x)+g(x)}|.$$

If no distribution is mentioned, we always assume $D = U_n$. Furthermore, for a subset of functions $\mathcal{C}$, we define

$$\mathsf{corr}_D(f,\mathcal{C}) = \max_{g\in\mathcal{C}} \mathsf{corr}_D(f,g).$$

Viola and Wigderson defined a convenient quantity R_k, which is very useful in bounding correlations against NOF protocols.

▶ **Definition 18** (k-party Norm). *For a function $f : (\{0,1\}^{n/k})^k \to \{0,1\}$, define the k-party norm of f to be*

$$R_k(f) := \mathsf{E}_{X_1^{(0)},\ldots,X_k^{(0)},X_1^{(1)},\ldots,X_k^{(1)} \sim U_{n/k}} e\left(\sum_{\delta \in \{0,1\}^k} f(X_1^{(\delta_1)},\ldots,X_k^{(\delta_k)})\right).$$

This norm is useful due to the following theorem.

▶ **Theorem 19** ([31]). *Let $f : \{0,1\}^n \to \{0,1\}$ be arbitrary, and let g be computable by a d-party NOF protocol exchanging c bits. Then*

$$R_d(f) \leq \mathsf{corr}(f,g) \leq 2^c R_d(f)^{1/2^d}.$$

We will also use the following theorem of Nisan and Wigderson, which allow us to translate correlation bounds into PRGs.This version is seen in the survey of Hatami and Hoza [14]

▶ **Theorem 20** ([24], [14, Theorem 4.2.2]). *Let $f : \{0,1\}^n \to \{0,1\}$. Suppose $h : \{0,1\}^r \to \{0,1\}$ is ε-hard for $f \circ \mathsf{JUNTA}_{r,k}$ with respect to the uniform distribution. Then there exists a PRG for f with seed length $s = O(n^{\frac{1}{k+1}} \cdot r^2/k)$ and error εn.*

4 Nearly Optimal Correlation Bounds against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$

We strictly improve upon the result [26] by proving a stronger correlation bound against $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits. This immediately gives PRGs against this class with improved seed length via the "hardness vs. randomness" framework [24] All previous work [28, 20, 26] looked at the function introduced in [25] created by taking the generalized inner product of parities. We present a new function comprised of field multiplication of extractors in order to prove stronger correlation bounds. Let m, n be parameters, and define $k := n/d$. We now prove the following result:

▶ **Theorem 21.** *Let $\mathsf{Ext} : \{0,1\}^k \to \{0,1\}^{.2k^{.996}}$ be a $(k^{.998}.2^{-.4k^{.996}})$ OBF-source extractor (explicit ones exist due to [17]). Let $f : (\{0,1\}^{.2k^{.996}})^d \to \{0,1\}$ be any function such that $\mathsf{corr}(f, \Pi_d^d) \leq 2^{-\Omega(k^{.996}/2^d)}$. Define $f \circ \mathsf{Ext}^d : (\{0,1\}^k)^d \to \{0,1\}$ to be the function*

$$f \circ \mathsf{Ext}^d(X) := f(\mathsf{Ext}(X_1),\ldots,\mathsf{Ext}(X_d)).$$

Let g be any function implementable by a $n^{O(\log n)}$-size $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit, and let $m = .0005 \log n$. Then

$$\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}, g) \leq 2^{-\Omega(n^{.995})}.$$

In particular, by instantiating this template, say, with Ext being the extractor of [17] and f being either GIP [2] or FFM [10], we get explicit $f \circ \mathsf{Ext}^{m+1}$. We also note by simple adjusting of constants, we can get any $2^{-\Omega(n^{1-\varepsilon})}$ for constant $\varepsilon > 0$. This gives an improvement of the correlation bound given in [26] of $2^{-\Omega(n^{.499})}$.

Proof. We follow the same approach as done in [26]. The uniform distribution can be expressed as applying a random restriction, and then filling in the remaining bits uniformly. For good random restrictions, we argue that g simplifies to a $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_m$ circuit.

We then argue that even after the random restriction, $f \circ \mathsf{Ext}^{m+1}$ maintains its structural integrity due to the extractor. We then finish the argument by using Hastad and Goldmann's connection between $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AND}_m$ and NOF protocols, and the fact that f has small correlation with $(m+1)$-party protocols.

The proof for the simplification of g is the same as seen in [26] so we merely cite it here. The only change is the tuning of parameters. Here is the lemma restated for our use.

▶ **Lemma 22.** *Let $g \in \{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0_d$ with circuit size $s = n^{\tau \log n}$. Then for $p = \frac{1}{48}(48 \log s)^{-(d-1)}$*

$$\begin{aligned}\Pr_{\rho \leftarrow \mathcal{R}_p}[g|\rho \text{ is not computed by } (.001pk, \{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_{\log s})\text{-tree}] \\ &\leq s \cdot 2^{-.001pk/2^d} \\ &= 2^{-\Omega_d(pk)}\end{aligned}$$

Notice that for constant d this gives a bound of $2^{-\Omega(n/\mathsf{polylog}(n))}$, versus its use in [26] in which a $2^{-\Omega(\sqrt{n/\log n})}$ error was gained. We will see later that we can liberally set parameters here because our hard function maintains integrity even after traversing down a path of size $n/\mathsf{polylog}(n)$ (equivalent to randomly fixing $n/\mathsf{polylog}(n)$ bits), whereas the previous GIP function could only withstand $\sqrt{n}$ bits. This is result of using an OBF extractor with much better parameters than simply taking the XOR of many copies.

The leaves of our tree is now much simpler class of circuits, but it is not simple enough. Our correlation bounds can only handle circuits with fan in $m = O(\log n)$, but we currently have fan in $\log s = O(\log^2 n)$. Fix a leaf ℓ of the tree, and let $\{C_1, \ldots, C_{s^2}\}$ be a collection of subsets of $[n]$ where C_i contains the $\leq \log s$ indices of the variables that feed into the ith $\mathsf{AND}_{\log s}$ gate in the bottom layer. We now use the following basic fact, as in [20] and [26], that there is a large subset of variables that minimally intersect with each C_i.

▷ Claim 23. A random $\mathbf{L} \subset_q [n]$ (add each element to $\mathbf{L}$ with probability q) satisfies

$$\Pr[\exists i \in [s^2] \text{ such that } |C_i \cap \mathbf{L}| > m] \leq s^2 \binom{w}{m} q^m.$$

Instantiating this claim with our parameter setting of m and s, and setting $q = \Theta(n^{-.001})$ tells us

$$\Pr[\exists i \in [s^2] \text{ such that } |C_i \cap \mathbf{L}| > m] \leq \frac{1}{s}.$$

Hence there exists such an $L = L(\ell)$ such that restricting all bits outside L makes only $\leq m$ variables feed into each AND gate as desired.

To summarize, our restriction ρ is sampled by a distribution D specified by these three steps.

1. We first perform restriction $\mathcal{R}_p$,
2. and then randomly restrict $\leq .001pk$ while walking down the depth-$.001pk$ tree to a leaf ℓ,
3. and then randomly restrict all the variables alive in this leaf that is *not* in the $L(\ell)$ set that we showed existed

At the end of this process, we have by the union bound that with all but $2^{-\Omega(-pk)}$ probability, $g|_\rho$ becomes a $\{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_m$ circuit.

We now observe what happens to $f \circ \mathsf{Ext}^{m+1}$ under this restriction ρ. We claim $f \circ \mathsf{Ext}^{m+1}$ retains its structure. Our wish is for at least $k^{.998}$ bits in each block to survive. That way, we will have a high entropy oblivious bit-fixing source fed into each extractor, and the function will be able to continue to strongly uncorrelate with m-party protocols. In Step 1, we draw a restriction from $\mathcal{R}_p$. Notice the live variables are distributed like a set $S \subset_p [n]$. We see that by a simple Chernoff and union bound,

$$\Pr_{\mathbf{S} \leftarrow \mathcal{R}_p} \left[\exists i \in [m+1] \text{ such that } |X_i \cap \mathbf{S}| < \frac{pk}{2} \right] \leq (m+1)2^{-\Omega(pk)}$$

Hence except for probability $m2^{-\Omega(pk)} = 2^{-\Omega(n^{1-o(1)})}$, each block X_i will have $\geq pk/2$ live variables. Conditioned on this, when we follow Step 2 and perform a random walk down the decision tree to a leaf, we will assign at most $.001pk$ bits, so we are guaranteed that each block X_i will contain at least $.499pk$ live variables. Step 3 is to take set $L(\ell)$ and arbitrarily restrict variables outside of it. We showed there exists an $L(\ell)$ which minimally overlaps with the input variables to the $\mathsf{AND}_{\log s}$ gates, but we want it to simultaneously overlap heavily with each block. That way most of the X_i will stay alive after restricting the bits outside of $L(\ell)$ The existence of such an $L(\ell)$ can be established by "completing the probabilistic method" started a few paragraphs above. Conditioning on good restrictions so far, let Y_i denote the variables that survived in X_i (hence $|Y_i| \geq .499pk$). We see that

$$\Pr_{\mathbf{L} \subset_q [n]} \left[\exists i \in [m+1] \text{ such that } |Y_i \cap \mathbf{L}| < \frac{.499pqk}{2} \right] \leq (m+1)2^{-\Omega(pqk)}.$$

Hence, the probability that $\mathbf{L}$ either intersects some C_i too much or some Y_i too little will happen with probability $\leq \frac{1}{s} + (m+1)2^{-\Omega(pqk)} \ll 1$. Thus there exists an $L(\ell)$ such that restricting all variables outside of it will simultaneously simplify g to a $\{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_m$ and also leave at least $\frac{.499pqk}{2} \geq .249k^{.999}/\mathsf{polylog}(n) \gg k^{.998}$ variables alive. Stringing all three steps together, we know that except with probability $2^{-\Omega(-pk)}$, our random restriction ρ reduces g to $\{\mathsf{SYM}_{s^2}, \mathsf{THR}_{s^2}\} \circ \mathsf{AND}_m$, while simultaneously keeping $\geq k^{.998}$ variables in each X_i block alive.

We are now in the final phase of the argument where we now directly bound the correlation against the simplified circuit. We first state the results that will convert our circuits to NOF protocols.

▶ **Theorem 24** ([13]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a size-s $\mathsf{SYM} \circ \mathsf{AND}_m$ circuit. Then for any partition of the n inputs of f into $m+1$ blocks, there is a deterministic NOF $(m+1)$-party communication protocol that computes f using $O(m \log s)$ bits of communication.*

▶ **Theorem 25** ([23]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by a $\mathsf{THR} \circ \mathsf{AND}_m$ circuit. Then for any partition of the n inputs of f into $m+1$ blocks, there is a randomized NOF $(m+1)$-party communication protocol that computes f with error γ_{err} using $O(m^3 \log n \log(n/\gamma_{err}))$ bits of communication.*

We now need to show an average-case hardness result for $f \circ \mathsf{Ext}^{m+1}|_\rho$ against NOF protocols. To do so, we will first calculate the k-party norm of $f \circ \mathsf{Ext}^{m+1}|_\rho$.

▶ **Lemma 26.** *Let ρ be a restriction which keeps $\geq k^{.998}$ variables in each X_i alive. Then $R_{m+1}(f \circ \mathsf{Ext}^{m+1}|\rho) \leq R_{m+1}(f) + 4(m+1) \cdot 2^{-4k^{.996}}$*

Proof. Now notice that

$$R_{m+1}(f \circ \mathsf{Ext}^{m+1}|_\rho) = \mathsf{E}_{X^{(0)},X^{(1)}} e\left(\sum_{\delta\in\{0,1\}^{m+1}} f(\mathsf{Ext}(X_1^{(\delta_1)}|_\rho),\ldots,\mathsf{Ext}(X_{m+1}^{(\delta_{m+1})}|_\rho))\right) \quad (2)$$

By assumption of ρ, each $X_i^{(\delta_i)}|\rho$ over uniform X_i is an OBF source with min-entropy $k^{.998}$, and so each $\mathsf{Ext}|_\rho(X_i) \approx_{2^{-4k^{.996}}} U_{.2k^{.996}}$. Since all $X_i^{(b)}$ for $i \in [m+1], b \in \{0,1\}$ are mutually independent, it follows by a hybrid argument that

$$(\mathsf{Ext}|_\rho(X_i^{(b)}|_\rho)_{i\in[m+1],b\in\{0,1\}} \approx_{2(m+1)2^{-4k^{.996}}} (U_{.2k^{.996}})_{i\in[m+1],b\in\{0,1\}}.$$

Therefore, we can upper bound Equation 2 by

$$\mathsf{E}_{(Y_i^{(b)})_{i\in[m],b\in\{0,1\}}} e\left(\sum_{\delta\in\{0,1\}^{m+1}} f(Y_1^{(\delta_1)}),\ldots,Y_{m+1}^{(\delta_{m+1})})\right) + 4(m+1)2^{-4k^{.996}}$$
$$\leq R_{m+1}(f) + 4(m+1)2^{-4k^{.996}}$$

as desired. ◀

With this, we can show that $f \circ \mathsf{Ext}^{m+1}|_\rho$ uncorrelates against randomized multiparty protocols.

▶ **Theorem 27.** *Let $g : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and let ρ be a restriction such that $X_i|_\rho$ has $\geq k^{.998}$ live variables for each i, and $g|_\rho$ can be computed by an $(m+1)$-party NOF randomized protocol with with $\leq c$ bits and with error γ. Then*

$$\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}|_\rho, g|_\rho) \leq 2\gamma + 2^{c-\Omega(k^{.996}/2^m)}.$$

This proof is deferred to the full version.

We now have all the ingredients to finish. Say ρ is good if ρ keeps $\geq k^{.998}$ variables alive in each block X_i *and* $g|_\rho$ is computable by $\{\mathsf{SYM},\mathsf{THR}\} \circ \mathsf{AND}_m$. We have shown for $\rho \sim D$, this doesn't happen only with probability $2^{-\Omega(pk)}$. If $g|_\rho$ has a SYM gate at the top, then Theorem 24 says the $\mathsf{SYM} \circ \mathsf{AND}_m$ circuit can be computed by a deterministic NOF protocol over $X_1,\ldots,X_{m+1}$ using $O(m \log s)$ bits. Plugging this in to Theorem 27 tells us

$$\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}|_\rho, g|_\rho) \leq 2^{m\log s - \Omega(k^{.996}/2^m)} \leq 2^{-\Omega(n^{.995})}.$$

If the top gate is a THR, use Theorem 25 with $\gamma_{err} = 2^{-n^{.997}}$ to get that the circuit is a randomized NOF protocol over $X_1,\ldots,X_{m+1}$ using $O(m^3 \log n \log(n/\gamma_{err})) = O(n^{.995})$ bits. Plugging this into Theorem 27 gives us a correlation bound of

$$\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}|_\rho, g|_\rho) \leq 2^{n^{.995} - \Omega(k^{.996}/2^m)} \leq 2^{-\Omega(n^{.996})}.$$

In either case we get the same bound, so we can bound

$$\begin{aligned}\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}, g) &= |\mathsf{E}_{\rho\sim D}\mathsf{E}_X(-1)^{f\circ\mathsf{Ext}^{m+1}|_\rho(X)+g|_\rho(X)}| \\ &\leq 2^{-\Omega(pk)} + \mathsf{E}_{\rho\sim D}[|\mathsf{E}_X(-1)^{f\circ\mathsf{Ext}^{m+1}|_\rho(X)+g|_\rho(X)}||\rho \text{ is good}] \\ &\leq 2^{-\Omega(pk)} + 2^{-\Omega(n^{.995})} = 2^{-\Omega(n^{.995})}.\end{aligned}$$

The theorem is proved. ◀

▶ Remark 28. We note that the original RW function instantiated with different parameters can also get the same strengthened correlation bound. This requires a more nuanced analysis than present in [26], and does not extend to general functions of the form $f \circ \mathsf{Ext}^{m+1}$ as it relies on the specific structure of GIP and $\bigoplus$.

To recap the argument for a size s circuit, we first use the multi-switching lemma to reduce to a depth-2 circuit of fan-in $\log s$. We then restrict more variables so that the fan-in reduces to $\sqrt{\log s}$. We then apply correlation bounds for $\sqrt{\log s}$-party protocols to get an error of $\exp(-n/2^{\sqrt{\log s}})$. If one trusts that this error is the bottleneck in the argument, one can imagine running through the above argument again with $s = n^{\Theta(1)}$ to get a better error.

▶ **Corollary 29.** *Let $g(X)$ be a function implementable by a size $s = n^{O(1)}$-size $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuit, and let $m = 2^{\sqrt{\log n}}$. Define $k := n/(m+1)$, and let $\mathsf{Ext} : \{0,1\}^k \to \{0,1\}^{k/2^{O(\sqrt{\log n})}}$ be a $(k/2^{O(\sqrt{\log n})}, 2^{-k/2^{O(\sqrt{\log n})}})$-extractor constructed from [17]. Then*

$$\mathsf{corr}(f \circ \mathsf{Ext}^{m+1}, g) \leq 2^{-(n/2^{O(\sqrt{\log s})})}.$$

This refinement will be useful for our correlation bounds against branching programs in the next section. As the proof is extremely similar to the above, we defer the sketch to the full version.

From Theorem 21, we derive the following two theorems as well.

▶ **Theorem 30.** *There exists an ε-PRG against size-S $\{\mathsf{SYM}, \mathsf{THR}\} \circ \mathsf{AC}^0$ circuits with seed length $s = 2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{2.01}$.*

▶ **Theorem 31.** *There is an efficient ε-PRG which fools $\mathsf{AC}^0[\mathsf{SYM}, n^{.999}, S]$ with seed length $2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{2.01}$ and an ε-PRG which fools $\mathsf{AC}^0[\mathsf{THR}, n^{.499}, S]$ with seed length $2^{O(\sqrt{\log S})} + (\log(1/\varepsilon))^{4.01}$.*

The proofs of these theorems follow by applying the Nisan-Wigderson hardness to randomness approach, as well as the decision tree bootstrapping idea of [20]. The details are deferred to the full version of the paper.

5 PRGs against $(d, \mathsf{poly}(n), n)$-2BPs

In this section, we use fortified hard functions to establish strong correlation bounds against the XOR of juntas, $\mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}$. These are then pushed through the Nisan-Wigderson "hardness vs. randomness" framework to create PRGs which can fool $(d, \mathsf{poly}(n), n)$-2BPs. We first establish the correlation bounds, and then we show that this implies our desired PRG.

5.1 Correlation Bounds Against $\mathsf{JUNTA}_{n,d}^{\oplus \mathsf{poly}(n)}$

This subsection is devoted to proving the following result.

▶ **Theorem 32.** *Let $m = d \log n$, let h be the hard function in Corollary 29 instantiated on $k := n/m$ bits, and let $\oplus_m : \{0,1\}^m \to \{0,1\}$ be the parity function on m bits. We then have*

$$\mathsf{corr}(h \circ \oplus_m^k, \mathsf{JUNTA}_{n,d}^{\oplus n^c}) \leq \exp\left(-\frac{n}{d2^{O(\sqrt{\log n})}}\right)$$

Proof. Consider arbitrary $g \in \mathsf{JUNTA}_{n,d}^{\oplus n^c}$. We will show that there exists a subset $T \subset [n]$ of variables such that upon fixing all variables outside T, g simplifies to a sparse polynomial, while at least one input variable in each $\oplus_m$ stays alive. Write $f = \sum_{i=1}^{n^c} \phi_i$, where each ϕ_i is a d-junta. Let $S_i \subset [n]$ be the indices of the variables that ϕ_i depends on. Pick $T \subset_{1/d} [n]$. For a fixed i, we can bound

$$\Pr_T[|T \cap S_i| \geq \ell] \leq \sum_{\substack{S \subset S_i \\ |S|=\ell}} \Pr_T[S \subset T] = \binom{d}{\ell} \left(\frac{1}{d}\right)^\ell \leq \exp(-\Omega(\ell \log \ell)) \leq 0.1 n^{-c}.$$

for $\ell = \Theta(\log n)$. Union bounding over all i, it follows that

$$\Pr_{\rho \sim \mathcal{R}_{1/d}}[\exists i, |T \cap S_i| \geq \ell] < 0.1. \tag{3}$$

Let $X_1, \ldots, X_k$ be the input blocks of size m feeding into h. We can easily calculate

$$\Pr_T[\exists i, X_i \cap T = \emptyset] \leq k(1-1/d)^m \leq k \exp(-m/d) = 1/m = o(1). \tag{4}$$

Union bounding Equation (3) and Equation (4), it follows that there exists a subset $T \subset [n]$ that simultaneously intersects at most ℓ variables alive in each junta ϕ_i, and intersects at least one variable in each X_i. By pruning out elements, we can assume WLOG that there is exactly one variable in each X_i.

Since a function over b bits can be written as an $\mathbb{F}_2$-polynomial with up to 2^b terms, it follows for any restriction ρ with $\rho^{-1}(\star) = T$, $\phi_i|_\rho$ is a polynomial with $2^\ell = n^{\Theta(1)}$ terms. Therefore, $f|_\rho$ is a polynomial with $n^{\Theta(1)}$ terms as well, which can be written as a $n^{\Theta(1)}$-sized $\mathsf{PAR} \circ \mathsf{AND}$ circuit. Furthermore, we know that $h \circ \oplus_m^k|_\rho$ is equivalent to h up to negations of the inputs. As $\mathsf{SYM} \circ \mathsf{AC}^0$ is invariant under shifts of the input, we can appeal to Corollary 29 and observe

$$\begin{aligned}\mathsf{corr}(h \circ \oplus_m^k, g) &= |\mathsf{E}_X(-1)^{h \circ \oplus_m^k(X) + g(X)}| \\ &\leq \mathsf{E}_{X_{\overline{T}}} |\mathsf{E}_{X_T}(-1)^{h \circ \oplus_m^k(X_T, X_{\overline{T}}) + g(X_T, X_{\overline{T}})}| \leq \exp\left(-(n/d)/2^{O(\sqrt{\log n})}\right)\end{aligned}$$ ◀

5.2 Constructing and Analyzing the PRG

With this correlation bound in hand, we can construct good PRGs against the XOR of juntas using the Nisan-Wigderson framework.

▶ **Corollary 33.** *There is an ε-PRG for $\mathsf{JUNTA}_{n,d}^{\oplus n^{\Theta(1)}}$ with seed length $s = 2^{O(\sqrt{\log n})} d^2 \log^2(1/\varepsilon)$*

The proof is a straightforward application of the Nisan-Wigderson framework that we defer to the full version.

Fooling the parity of juntas actually allow us to fool arbitrary functions of juntas as long as the function has low Fourier L_1 norm.

▶ **Theorem 34.** *Let G be an ε-PRG for $\mathsf{JUNTA}_{n,d}^{\oplus m}$, and let $f : \{0,1\}^m \to \{0,1\}$. Then G is an $\varepsilon \cdot L_1(f)$-PRG for $f \circ \mathsf{JUNTA}_{n,d}$.*

We also defer this proof to the full version.

Finally, as an application, we show PRGs against (d, t, n)-2BPs, branching programs over n bits with width 2, length t, and reads d bits at a time. We will use the fact that width-2 branching programs which read one bit at a time have low Fourier L_1 norm (a proof can be found in [14]).

▶ **Lemma 35.** *If f is a $(1,t,n)$-2BP, then $L_1(f) \leq (t+1)/2$.*

We now use the fact that a (d,t,n)-2BP can be represented by a normal width-2 branching program acting on juntas to prove that the PRG from Corollary 33 fools (d,t,n)-2BPs.

▶ **Theorem 36.** *There exists an ε-PRG for (d,n^c,n)-2BPs with seed length $s = 2^{O(\sqrt{\log n})} \cdot d^2 \log^2(n/\varepsilon)$.*

Proof. Given a (d,n^c,n)-2BP B, we note that at each vertex $v \in [2n^c]$ of B, the transition function is some d-junta ϕ_v which will map the d bits read at that vertex to the next vertex to move to. Now consider the $(1,n^c,2n^c)$-2BP B' defined with the same vertex set as B, and define the transition function for $v \in [2n^c]$ in B' to read the vth bit of the input, and then map to the node in the next layer labeled by that bit. It is easy to see by construction that $B(x) = B'(\phi_1(x),\ldots,\phi_{2n^c}(x))$, which is a function in $B' \circ \mathsf{JUNTA}_{n,d}$. By Theorem 34, this can be ε-fooled by an $(\varepsilon/L_1(B'))$-PRG for $\mathsf{JUNTA}_{n,d}^{\oplus 2n^c}$. Using the L_1 bound from Lemma 35 and the construction from Corollary 33, we see that such a PRG has seed length $2^{O(\sqrt{\log n})} d^2 \log^2(1/\varepsilon)$. ◀

▶ Remark 37. There is an alternative PRG construction using the Ajtai-Wigderson framework [1] which gives optimal dependence on d, but exponentially worse dependence on ε. This is presented in the full version of the paper.

6 Correlation Bounds Against Set-Multilinear Polynomials

Our correlation bound for set-multilinear polynomials follows from an instantiation of the following theorem.

▶ **Theorem 38.** *Let $d \leq n$ be an integer. Let $\mathsf{Ext} : \{0,1\}^{n/d} \times \{0,1\}^{2n/d} \to \{0,1\}^{k-2\log(1/\varepsilon)}$ be a strong linear seeded (k,ε)-extractor with seed length $2n/d$ created from the Leftover Hash Lemma [16], and let χ some nontrivial additive character of $\mathbb{F}_{2^{n/d}}$. Define $\mathsf{ExtFFM}_d : \{0,1\}^{n+2n/d} \to \{0,1\}$ to be*

$$\mathsf{ExtFFM}_d(X,W) = \chi\left(\prod_{i=1}^{d} \mathsf{Ext}(X_i,W)\right).$$

Let $g : \{0,1\} \to \{0,1\}^n$ be a function, and let $S_1,\ldots,S_d \subset [n/d]$ be subsets of size $\geq k$ such that for any restriction ρ created by arbitrarily fixing all bits in W and outside S_i in X_i for each i, $g|_\rho$ always becomes set multilinear in $X_1,\ldots,X_d$. We then have

$$\mathsf{corr}(\mathsf{ExtFFM}_d, g) \leq d\varepsilon + (d-1)\left(\frac{1}{2^k\varepsilon^2} + \varepsilon\right).$$

Proof. For brevity, we let $f := \mathsf{ExtFFM}_d$ in this proof. We will first split the correlation expectation into first randomizing over all restrictions ρ of the bits in X outside of $S_1,\ldots,S_d$, then over the seed W, and then over the remaining live variables denoted by the S_i, which we denote $X_1|_\rho,\ldots,X_d|_\rho$. Now let W_ρ be the set of seeds w such that $\mathsf{Ext}(X_i|_\rho,w) \approx_\varepsilon U_k$ for all i. As Ext is strong-seeded, it follows by a union bound that W_ρ cover all but a $d\varepsilon$ fraction of seeds. Thus one can write

$$\begin{aligned}
\mathsf{corr}(f,g) &= |\mathsf{E}_X(-1)^{f'(X)+g(X)}| \\
&\leq \mathsf{E}_{W,\rho}\left|\mathsf{E}_X(-1)^{f|_\rho(X,W)+g|_\rho(X,W)}\right| \\
&\leq d\varepsilon + \mathsf{E}_\rho \mathsf{E}_{w\in W_\rho} |\mathsf{E}_X(-1)^{f|_\rho(X,w)+g|_\rho(X,w)}| \qquad (5)
\end{aligned}$$

Now fix a partial assignment ρ and seed $w \in W_\rho$. For brevity, let $f(\cdot) := f|_\rho(\cdot, w)$, and similarly for g'. By assumption, g' is set-multilinear over X We now apply a similar argument showing up in [3]. Let α be a map taking linear forms $\sum_{i\in[n/d]} c_i X_{d,i}$ in X_d to its vector of coefficients $(c_i) \in \mathbb{F}_2^{n/d}$. Note that by this definition, for any linear form $\ell(X_d)$, $\langle \ell(X_d), X_d\rangle = \ell(X_d)$. Letting $e(x) = (-1)^x$. We then see

$$\begin{aligned}\left|\mathsf{E}_X(-1)^{f'(X)+g'(X)}\right| &= \left|\mathsf{E}_X e\left(f(X_i) + \sum_{i\in[d-1]} g_i(X_{-i}) + g_d(X_d)\right)\right| \\ &\le \mathsf{E}_{X_{[d-1]}}\left|\mathsf{E}_{X_d} e\left(\langle \alpha(f(X_i) + \sum_{i\in[d-1]} g_i(X_{-i})), X_d\rangle + g_d(X_{-d})\right)\right| \\ &\le \Pr_{X_{[d-1]}}\left[\alpha(f'(X) + \sum_{i\in[d-1]} g_i(X_{-i})) = 0\right] \qquad (6)\end{aligned}$$

where we used the facts that f' is linear in X_d (as Ext here is a linear seeded extractor), $g_d(X_{-d})$ is independent of X_d, and linear forms are perfectly unbiased if their coefficient vector is nonzero. We now repeatedly use the simple inequality that for a linear map $h : \mathbb{F}_2^m \to \mathbb{F}_2^k$ and $a \in \mathbb{F}_2^k$, $\Pr_x[h(x) = a] \le \Pr_x[h(x) = 0]$ as follows.

$$\begin{aligned}&\Pr_{X_{[d-1]}}\left[\alpha(f'(X) + \sum_{i\in[d-1]} g_i(X_{-i})) = 0\right] \qquad (7)\\ &= \mathsf{E}_{X_{[d-2]}} \Pr_{X_{d-1}}\left[\alpha(f'(X) + \sum_{i=1}^{d-2} g_i(X_{-i}))) = \alpha(g_{d-1}(X_{-(d-1)}))\right] \\ &\le \Pr_{X_{[d-1]}}\left[\alpha\left(f'(X) + \sum_{i=1}^{d-2} g_i(X_{-i}))\right) = 0\right] \\ &\le \cdots \\ &\le \Pr_{X_{[d-1]}}[\alpha(f'(X)) = 0] \qquad (8)\end{aligned}$$

To analyze this probability, we state a lemma whose proof is deferred to the full version.

▶ **Lemma 39.** *For a linear form $\ell(X_d)$, $\alpha(\ell(X_d)) = 0$ if and only if $\ell(X_d) = 0$ for all X_d.*

Therefore, by Lemma 39,

$$\Pr_{X_{[d-1]}}[\alpha(f'(X)) = 0] = \Pr_{X_{[d-1]}}\left[\forall X_d, \chi\left(\prod_{i=1}^{d} \mathsf{Ext}(X_i|_\rho, w)\right) = 0\right].$$

Clearly if $\prod_{i=1}^{d-1} \mathsf{Ext}(X_i|_\rho, W) = 0$, f' becomes identically zero. When this doesn't happen, the function becomes of the form $\chi(c \cdot \mathsf{Ext}(X_d|_\rho, w))$ for some nonzero $c \in \mathbb{F}_{2^{n/d}}$. We now claim that there must exist some $X_d|_\rho$ such that $\chi(c \cdot \mathsf{Ext}(X_d|_\rho, w))$. Notice that for exactly $2^{n/d-1}$ values of Y, $\chi(cY) = 0$. As $w \in W_\rho$, the probability that a random $X_d|_\rho$ has $\mathsf{Ext}(X_d|_\rho, w)$ hit one of these values must be $\ge 1/2 - \varepsilon > 0$, proving the claim. Therefore, in order for $\alpha(f'(X)) = 0$, it is necessary that $\prod_{i=1}^{d-1} \mathsf{Ext}(X_i|_\rho, W) = 0$. Therefore,

$$\begin{aligned}\Pr_{X_{[d-1]}}[\alpha(f'(X)) = 0] &\leq \Pr_{X_{[d-1]}}\left[\prod_{i=1}^{d-1} \mathsf{Ext}(X_i|_\rho, w) = 0\right] \\ &\leq \sum_{i=1}^{d-1} \Pr_{X_i}[\mathsf{Ext}(X_i|_\rho, w) = 0] \\ &\leq (d-1)\left(\frac{1}{2^{k-2\log(1/\varepsilon)}} + \varepsilon\right)\end{aligned}$$

Stringing the above with inequalities (5), (6), and (8), we find

$$\mathsf{corr}(\mathsf{ExtFFM}_d, g) \leq d\varepsilon + (d-1)\left(\frac{1}{2^k\varepsilon^2} + \varepsilon\right). \qquad \blacktriangleleft$$

As a very nice application of this structural theorem, we show that we can achieve exponentially small correlation against $n^{O(1)}$-degree polynomials which are set-multilinear over some partition of the input into up to $n^{1-O(1)}$ parts.

▶ **Corollary 40.** *Let g be a degree $< d$ polynomial which is set-multilinear over an arbitrary partition $(A_1, \ldots, A_c)$ of X into c parts. Then*

$$\mathsf{corr}(\mathsf{ExtFFM}_d, g) \leq 2^{-\Omega(n/cd)}.$$

Proof. For each $i \in [n/d]$, define S_i to be the largest set among $\{X_i \cap A_1, \ldots, X_i \cap A_c\}$ (arbitrarily pick one if there are ties). Notice that the sets $\{X_i \cap A_j\}_{j\in[c]}$ partition X_i, and $|X_i| = n/d$. Therefore, we know that each $|S_i| \geq \frac{n/d}{c} = \frac{n}{cd}$. We now claim that any restriction ρ formed by arbitrarily fixing all the bits in X_i which are outside S_i, for each i, will make $g|_\rho$ set-multilinear over $(X_1, \ldots, X_d)$. Assume for the sake of contradiction there existed some monomial in $g|_\rho(X)$ that contained 2 variables from some X_i. Since $S_i \subset X_i$ and $S_j \cap X_i = \emptyset$ for $j \neq i$, both of these variables had to have come from S_i. But note that $S_i = X_i \cap A_\ell \subset A_\ell$ for some ℓ, and we know no monomial has 2 terms from the same A_i by our assumption of g. This yields our desired contradiction.

Therefore, we can apply Theorem 38 on the sets (S_i) with $k = n/cd$ and $\varepsilon = 2^{-.1n/cd}$ to deduce that

$$\mathsf{corr}(f, g) \leq d2^{-.1n/cd} + (d-1)(2^{-.8n/cd} + 2^{-.1n/cd}) = 2^{-\Omega(n/cd)}. \qquad \blacktriangleleft$$

References

1 Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 11–19, 1985. doi:10.1109/SFCS.1985.19.

2 L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, pages 1–11, New York, NY, USA, 1989. Association for Computing Machinery. doi:10.1145/73007.73008.

3 Abhishek Bhrushundi, Prahladh Harsha, Pooya Hatami, Swastik Kopparty, and Mrinal Kumar. On Multilinear Forms: Bias, Correlation, and Tensor Rank. In Jarosław Byrka and Raghu Meka, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*, volume 176 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 29:1–29:23, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX/RANDOM.2020.29.

4 Jaroslaw Blasiok, Peter Ivanov, Yaonan Jin, Chin Ho Lee, Rocco A. Servedio, and Emanuele Viola. Fourier growth of structured $\mathcal{O}_2$-polynomials and applications. In Mary Wootters and Laura Sanità, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2021, August 16-18, 2021, University of Washington, Seattle, Washington, USA (Virtual Conference)*, volume 207 of *LIPIcs*, pages 53:1–53:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPICS.APPROX/RANDOM.2021.53.

5 Andrej Bogdanov, Zeev Dvir, Elad Verbin, and Amir Yehudayoff. Pseudorandomness for width-2 branching programs. *Theory of Computing*, 9(7):283–293, 2013. doi:10.4086/toc.2013.v009a007.

6 Eshan Chattopadhyay, Jesse Goodman, Vipul Goyal, Ashutosh Kumar, Xin Li, Raghu Meka, and David Zuckerman. Extractors and secret sharing against bounded collusion protocols. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1226–1242, 2020. doi:10.1109/FOCS46700.2020.00117.

7 Eshan Chattopadhyay and Jyun-Jie Liao. Hardness Against Linear Branching Programs and More. In Amnon Ta-Shma, editor, *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:27, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2023.9.

8 Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. Mining circuit lower bound proofs for meta-algorithms. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 262–273, 2014. doi:10.1109/CCC.2014.34.

9 Gil Cohen and Igor Shinkar. The complexity of dnf of parities. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 47–58, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2840728.2840734.

10 Jeff Ford and Anna Gál. Hadamard tensors and lower bounds on multiparty communication complexity. *Comput. Complex.*, 22(3):595–622, 2013. doi:10.1007/s00037-012-0052-6.

11 Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM Journal on Computing*, 42(3):1051–1076, 2013. doi:10.1137/110854990.

12 Svyatoslav Gryaznov, Pavel Pudlák, and Navid Talebanfard. Linear branching programs and directional affine extractors. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, Dagstuhl, DEU, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2022.4.

13 J. Hastad and M. Goldmann. On the power of small-depth threshold circuits. In *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, pages 610–618 vol.2, 1990. doi:10.1109/FSCS.1990.89582.

14 Pooya Hatami and William Hoza. Theory of unconditional pseudorandom generators. *Electron. Colloquium Comput. Complex.*, TR23-019, 2023. URL: https://eccc.weizmann.ac.il/report/2023/019, arXiv:TR23-019.

15 Pooya Hatami, William M. Hoza, Avishay Tal, and Roei Tell. Fooling constant-depth threshold circuits (extended abstract). In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 104–115, 2022. doi:10.1109/FOCS52979.2021.00019.

16 R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, pages 12–24, New York, NY, USA, 1989. Association for Computing Machinery. doi:10.1145/73007.73009.

17 Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36(5):1231–1247, 2007. doi:10.1137/S0097539705446846.

18 Ilan Komargodski, Ran Raz, and Avishay Tal. Improved average-case lower bounds for demorgan formula size. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 588–597, 2013. doi:10.1109/FOCS.2013.69.

19 Xin Li and Yan Zhong. Explicit Directional Affine Extractors and Improved Hardness for Linear Branching Programs. In Rahul Santhanam, editor, *39th Computational Complexity Conference (CCC 2024)*, volume 300 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:14, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2024.10.

20 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size ac0 circuits with n(1-o(1)) symmetric gates. In Leslie Ann Goldberg, Klaus Jansen, R. Ravi, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 640–651, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

21 M. Luby, B. Velickovic, and A. Wigderson. Deterministic approximate counting of depth-2 circuits. In *[1993] The 2nd Israel Symposium on Theory and Computing Systems*, pages 18–24, 1993. doi:10.1109/ISTCS.1993.253488.

22 Xin Lyu. Improved pseudorandom generators for ac0 circuits. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, Dagstuhl, DEU, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2022.34.

23 Noam Nisan. The communication complexity of threshold gates. Combinatorics, Paul Erdős is eighty, Vol. 1, 1993.

24 Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of computer and System Sciences*, 49(2):149–167, 1994. doi:10.1016/S0022-0000(05)80043-1.

25 Alexander Razborov and Avi Wigderson. w(log n) lower bounds on the size of depth-3 threshold cicuits with and gates at the bottom. *Information Processing Letters*, 45(6):303–307, 1993. doi:10.1016/0020-0190(93)90041-7.

26 Rocco A. Servedio and Li-Yang Tan. Luby-Velickovic-Wigderson Revisited: Improved Correlation Bounds and Pseudorandom Generators for Depth-Two Circuits. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 56:1–56:20, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2018.56.

27 Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *Theory Comput.*, 18:1–46, 2022. URL: https://theoryofcomputing.org/articles/v018a004/, doi:10.4086/TOC.2022.V018A004.

28 Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. doi:10.1137/050640941.

29 Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 124–127, 2008. doi:10.1109/CCC.2008.16.

30 Emanuele Viola. Correlation bounds against polynomials. *Electron. Colloquium Comput. Complex.*, TR22-142, 2022. URL: https://eccc.weizmann.ac.il/report/2022/142, arXiv:TR22-142.

31 Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for gf(2) polynomials and multiparty protocols. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 141–154, 2007. doi:10.1109/CCC.2007.15.

32 Thomas Watson. Pseudorandom generators for combinatorial checkerboards. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 232–242, 2011. doi:10.1109/CCC.2011.12.