Multilinear Formulas, Maximal-Partition Discrepancy and Mixed-Sources Extractors

Ran Raz[*] Amir Yehudayoff[†]

Abstract

We study a new method for proving lower bounds for subclasses of arithmetic circuits. Roughly speaking, the lower bound is proved by bounding the correlation between the coefficients' vector of a polynomial and the coefficients' vector of any product of two polynomials with disjoint sets of variables. We prove lower bounds for several old and new subclasses of circuits.

Monotone Circuits: We prove a tight $2^{\Omega(n)}$ lower bound for the size of monotone arithmetic circuits. The highest previous lower bound was $2^{\Omega(\sqrt{n})}$.

Orthogonal Formulas: We prove a tight $2^{\Omega(n)}$ lower bound for the size of orthogonal multilinear formulas (defined, motivated, and studied by Aaronson). Previously, nontrivial lower bounds were only known for subclasses of orthogonal multilinear formulas.

Non-Cancelling Formulas: We define and study the new model of *non-cancelling multilinear formulas*. Roughly speaking, in this model one is not allowed to sum two polynomials that almost cancel each other. The non-cancelling multilinear model is a generalization of both the monotone model and the orthogonal model. We prove lower bounds of $n^{\Omega(1)}$ for the *depth* of non-cancelling multilinear formulas.

[*] Faculty of Mathematics and Computer Science, Weizmann Institute, Rehovot, Israel. Email: ran.raz@weizmann.ac.il. Research supported by grants from the Binational Science Foundation (BSF), the Israel Science Foundation (ISF), and the Minerva Foundation.

[†] Faculty of Mathematics and Computer Science, Weizmann Institute, Rehovot, Israel. Email: amir.yehudayoff@weizmann.ac.il. Research supported by grants from the Binational Science Foundation (BSF), the Israel Science Foundation (ISF), the Minerva Foundation, and the Israel Ministry of Science (IMOS) - Eshkol Fellowship.

Noise-Resistant Formulas: We define and study the new model of *noise-resistant multilinear formulas*. Roughly speaking, noise-resistant formulas are formulas that compute the required polynomial even in the presence of noise. We prove lower bounds of $n^{\Omega(1)}$ for the *depth* of noise-resistant multilinear formulas.

One ingredient of our proof is an explicit map $f : \{0,1\}^n \to \{0,1\}$ that has exponentially small discrepancy for every partition of $\{1,\dots,n\}$ into two sets of roughly the same size. We give two additional applications of this property to extractors construction and to communication complexity.

Mixed-Source Extractors: We define a new type of extractors which we call *mixed-2-source extractors*. A mixed-2-source is a source of randomness whose bits arrive from two independent sources, but they arrive in a fixed but *unknown* order. We are able to extract a linear number of almost perfect random bits from such sources.

Communication Complexity: We prove a tight $\Omega(n)$ lower bound for the randomized best-partition communication complexity of f, with error exponentially close to 1/2.

1 Introduction

Arithmetic circuits are the standard model for computing polynomials. Proving super-polynomial lower bounds for the size of arithmetic circuits and formulas is an outstanding open problem. Here we study a new method for proving lower bounds for several subclasses of circuits. Roughly speaking, this method is based on bounding the correlation between the coefficients' vector of a polynomial and the coefficients' vector of any product of two polynomials with disjoint sets of variables.

We prove tight exponential size lower bounds for two previously studied models of arithmetic circuits: monotone circuits (that are circuits that use only positive real numbers), and orthogonal multilinear formulas (that are formulas that are only allowed to sum two polynomials whose coefficients' vectors are orthogonal). We also prove $n^{\Omega(1)}$ lower bounds for the depth of two new models of multilinear formulas: non-cancelling formulas, and noise-resistant formulas. Here are rough definitions of these two models: Non-cancelling formulas are formulas that are not allowed to sum two polynomials that almost cancel each other – the non-cancelling model is a generalization of both the monotone model and the orthogonal model. Noise-resistant formulas are formulas that compute well even when some small noise occurred during the computation.

One important ingredient of our proof is an explicit map $f : \{0,1\}^n \to \{0,1\}$ that has exponentially small maximal-partition discrepancy (see Section 1.1.2 for a formal definition). This notion is also related to extractors construction and to communication complexity, as we will describe below.

1.1 Definitions and Motivations

We start by giving the formal definitions of the notions needed to state our results. We also give some motivation for considering these notions.

1.1.1 Arithmetic Circuits

We start by some general definitions regarding arithmetic circuits. An *arithmetic circuit* Φ over the field of complex numbers $\mathbb{C}$ and over the set of variables $X = \{x_1, \ldots, x_n\}$ is a directed acyclic graph as follows: Every vertex of in-degree 0 is labelled by either a field element or a variable. Every other vertex is of in-degree 2, and is labelled by either + or ×. There is a unique vertex in Φ of out-degree 0. An *arithmetic formula* is an arithmetic circuit whose underlying graph is a binary tree (whose edges are directed from the leaves to the root).

The *size* of Φ is the number of vertices in Φ. We denote the size of Φ by $|\Phi|$. The *depth* of a vertex v in Φ is the length of the longest directed path reaching v. We denote the depth of v by $\mathrm{depth}(v)$. The *depth* of Φ is the maximal depth of a gate in Φ. The vertices of Φ are also called *gates*. Gates of in-degree 0 are also called *input* gates. Gates labelled by + are called *sum* gates, and gates labelled by × are called *product* gates. The gate of out-degree 0 is called the *output* gate. If there is a directed edge from a gate v to a gate u, then v is called a *child* of u.

An arithmetic circuit computes a polynomial in a natural way. An input gate computes the polynomial it is labelled by (i.e., the variable or the field element). A sum gate computes the sum of the two polynomials computed by its two children. A product gate computes the product of the two polynomials computed by its two children. For a gate v in Φ, denote by Φ_v the sub-circuit of Φ rooted at v. Denote by X_v the set of variables that occur in Φ_v. Denote by $\widehat{\Phi}_v$ the polynomial in $\mathbb{C}[X_v]$ computed by v in Φ. Denote by $\widehat{\Phi}$ the polynomial computed by the output gate of Φ.

We now turn to the the different models of arithmetic circuits we consider.

Monotone Circuits. A polynomial $f \in \mathbb{R}[X]$ is called *monotone*, if the coefficients of all the monomials in f are non-negative. A well known example of a monotone polynomial is the permanent. An arithmetic circuit is called *monotone*, if all the field elements labeling its input gates are positive real numbers.

The model of monotone circuits has been studied in many papers, for example [Sc, Sn, ShSn, JSn, SV, TT, V]. In particular, Shamir and Snir proved a $2^{\Omega(\sqrt{n})}$ lower bound for the size of monotone circuits

[ShSn], and this is the best lower bound previously known. Moreover, Valiant showed that one 'negation' gate is exponentially powerful [V].

Multilinear Formulas. A polynomial $f \in \mathbb{C}[X]$ is called *multilinear*, if the degree of every variable in f is at most 1. We say that an arithmetic circuit is *multilinear*, if the polynomial computed by each of its gates is multilinear. We say that an arithmetic circuit is *syntactically multilinear*, if for every product gate v in it with children v_1 and v_2, the two sets X_{v_1} and X_{v_2} are disjoint (for a discussion of the difference between these two notions of multilinear computation see [RSY]).

Multilinear polynomials are common (e.g., determinant, and permanent). The natural way to compute a multilinear polynomial is via a multilinear computation, as the use of high powers during the computation requires non-intuitive cancellations. We note, however, that this intuition is false for monotone circuits for example (where "a single minus gate adds a lot of power").

The multilinear model was first studied by Nisan and Wigderson [NW]. Later [R04a] proved a super-polynomial lower bound for the size of multilinear arithmetic formulas for the determinant and the permanent. Furthermore, [R04b] proved a super-polynomial separation between the size of multilinear arithmetic circuits and formulas. The proof of this separation was later simplified in [RY], that also showed that syntactically multilinear arithmetic circuits of size poly(n) are (without loss of generality) of depth $O(\log^2(n))$ ([RY] following [VSBR]).

Proving super-polynomial lower bounds for the size of multilinear arithmetic circuits is an open problem (the best lower bound known for syntactically multilinear arithmetic circuits is $\Omega(n^{4/3}/\log^2(n))$ [RSY]). Since syntactically multilinear arithmetic circuits can be balanced [RY], proving $\omega(\log^2(n))$ lower bounds for the depth of syntactically multilinear arithmetic formulas will give a super-polynomial lower bound for the size of syntactically multilinear arithmetic circuits. This motivates proving lower bounds for the depth of subclasses of multilinear formulas, as we do here.

Before defining the rest of the models of circuits that we consider, we need to explain how to view a polynomial as a vector.

Polynomials as Vectors. Let $n \in \mathbb{N}$ be an integer. We denote $[n] = \{1, \ldots, n\}$. For the rest of this paper, we will sometimes interchange between subsets of $[n]$, subsets of $X = \{x_1, \ldots, x_n\}$ and monic multilinear monomials in the variables X (a monic monomial is a monomial whose coefficient is 1). For example, a set $T \subseteq [n]$ is also the set $\{x_i \ : \ i \in T\}$ as well as the monomial $\prod_{i \in T} x_i$.

We will focus on the following two vector spaces over the field $\mathbb{C}$.

1. The vector space of multilinear polynomials in $\mathbb{C}[X']$, where $X' \subseteq X$ (thinking of a polynomial as the vector of its coefficients). For example, for a gate v in a multilinear formula Φ over the field $\mathbb{C}$ and over the set of variables X, we think of the polynomial $\widehat{\Phi}_v$ also as a vector.

2. The vector space of maps from $\{1,-1\}^T$ to $\mathbb{C}$, where $T \subseteq [n]$.

For two vectors w, w' (as above), the *inner product* of w and w' is

$$\langle w, w' \rangle = \sum_t w(t)\overline{w'(t)},$$

where the sum is over all the coordinates t of the vectors (and for $\alpha \in \mathbb{C}$, we denote by $\overline{\alpha}$ the complex conjugate of α). Define the *correlation* of w and w' as

$$\operatorname{cor}(w, w') = \left|\langle w, w' \rangle\right|.$$

The vectors w and w' are called *orthogonal*, if $\operatorname{cor}(w, w') = 0$. The *norm* of the vector w is

$$\|w\| = \sqrt{\langle w, w \rangle}.$$

Orthogonal and Non-Cancelling Formulas. Orthogonal syntactically multilinear formulas were first defined and motivated by Aaronson [A]. He suggested a connection between such formulas and a certain type of quantum computations. More specifically, he defined a family of quantum states which he calls *orthogonal tree* states. He advocates that such states represent 'natural' quantum states. Orthogonal tree states can also be thought of as polynomials that are computed by orthogonal syntactically multilinear formulas. Aaronson studied the orthogonal model and proved lower bounds for what he calls *manifestly orthogonal* formulas, that are a subclass of orthogonal formulas.

Part of the motivation for considering monotone circuits is to understand what can circuits do without *any* cancellations of monomials. It seems natural to generalize this notion. One way to generalize this notion is given by the non-cancelling model. In fact, the non-cancelling model is more general than both the monotone model (in which there are no cancellations at all) and the orthogonal model (discussed above).

Every sum gate v in an arithmetic formula Φ sums two polynomials, say f_1 and f_2. Roughly, the non-cancelling condition says that the norm of $f_1 + f_2$ is not negligible compared to the norm of both f_1 and f_2. What does this mean? Well, in the case where the norm of $f_1 + f_2$ is negligible compared to the norms of both f_1 and f_2, the two polynomials are 'almost' the same (with opposite signs), except for a negligible part in which they may differ. Loosely speaking, this condition could be interpreted as a 'deep' understanding Φ (or the designer of Φ) has about the computation of $\widehat{\Phi}$.

The fact that we succeed in proving polynomial lower bounds for the depth of non-cancelling syntactically multilinear formulas, and not for (general) multilinear formulas, gives more 'evidence' to the 'fact' that we need to understand the cancellations of monomials better.

Here are the formal definitions of these models. For $\tau > 0$, we say that a sum gate v in an arithmetic formula Φ is *τ-non-cancelling*, if

$$\|\widehat{\Phi}_v\| \geq \tau \cdot \max(\|\widehat{\Phi}_{v_1}\|, \|\widehat{\Phi}_{v_2}\|), \tag{1.1}$$

where v_1 and v_2 are the two children of v. Stated differently, v is non-cancelling, if it does not subtract two polynomials that are 'almost' the same. We say that Φ is *τ-non-cancelling*, if every sum gate in Φ is τ-non-cancelling.

We say that an arithmetic formula Φ is *orthogonal*, if for every sum gate v in Φ with children v_1 and v_2,

$$\mathrm{cor}(\widehat{\Phi}_{v_1}, \widehat{\Phi}_{v_2}) = 0;$$

that is, the polynomials $\widehat{\Phi}_{v_1}$ and $\widehat{\Phi}_{v_2}$ are orthogonal (as vectors of coefficients). So, an orthogonal arithmetic formula is, in particular, 1-non-cancelling.

Remark 1.1. *We note that for every two vectors f and g, since*

$$\|f + g\| \geq \big|\|f\| - \|g\|\big|,$$

it holds that for $\tau \leq 1$

$$\|f + g\| \geq \tau \cdot \min(\|f\|, \|g\|) \quad \Rightarrow \quad \|f + g\| \geq \frac{\tau}{2} \cdot \max(\|f\|, \|g\|).$$

So, using minimum instead of maximum in (1.1) is the same, up to a factor of 2.

The Noise-Resistant Model. Our main motivation for this model is that it seems natural to assume that in any 'real' implementation of an arithmetic formula over $\mathbb{C}$ noise will occur. In fact, it seems that there are two ways to implement an arithmetic computation over the field of complex numbers: either by an analog circuit, which are bound to have some noise in it, or by a digital circuit, which yields the finite representation of complex numbers (floating point, for instance). Both of these ways seem to have an intrinsic noise in them. So, in order to compute (or even approximate) a map $g : \{1, -1\}^n \to \mathbb{C}$ in a way that will be resilient to the noise introduced by practical implementations, we want to find an arithmetic formula that is noise-resistant to computing g.

It seems natural to think that if the noise is much smaller than the size of the formula, then the formula computes almost the same polynomial even when noise occurs. Thus, one could expect that a polynomial size formula is always noise-resistant for exponentially small noise. This, however, is not necessarily true. In this paper, we prove lower bounds for the size of formulas that are noise-resistant for exponentially small noise.

We note that in other computation models defined over $\mathbb{C}$ (such as quantum circuits) a noise model was studied, and various interesting results were obtained.

Given an input t, say in $\{1,-1\}^n$, an arithmetic formula Φ gives a natural way for computing the value of the polynomial $\widehat{\Phi}$ in t. Upon realizing this computation of $\widehat{\Phi}(t)$ in the 'real world', it seems reasonable to assume that some noise will occur. A natural model for this noise is that each edge in the formula introduces a small noise into the computation. Given Φ we will think of a noisy version of Φ as the same as Φ, except that each edge of the noisy version is multiplied by a value that is close to 1 (that we think of as noise).

We note that, since we are proving lower bounds, if we assume a weaker noise model, our results become stronger. Hence, we want the noise model to be as weak as possible. We hence assume that the noise have the following two restrictions: only sum gates introduce noise, and the noise is a positive real number that is independent of the input.

We now turn to the formal definition of the noise model. For a gate v in an arithmetic formula Φ, and for $0 \leq \varepsilon \leq 1$, we will define below $\mathrm{N}_\varepsilon(\Phi_v)$ to be the set of maps from $\{1,-1\}^{X_v}$ to $\mathbb{C}$ that are the outputs of all the noisy versions of Φ_v on inputs in $\{1,-1\}^{X_v}$. Elements of $\mathrm{N}_\varepsilon(\Phi_v)$ will be called *ε-noisy values* of Φ_v. Before the definition, we make the following remark.

Remark 1.2. *The polynomial $\widehat{\Phi}_v$ naturally defines a map ϕ_v from $\{1,-1\}^{X_v}$ to $\mathbb{C}$. For $t \in \{1,-1\}^n$, the value of $\phi_v(t)$ is the value of the polynomial $\widehat{\Phi}_v$ after substituting $x_i = t_i$. Since only variables in X_v occur in $\widehat{\Phi}_v$, the map ϕ_v is indeed from $\{1,-1\}^{X_v}$ to $\mathbb{C}$.*

The definition of $\mathrm{N}_\varepsilon(\Phi_v)$ is inductively as follows.

- If v is an input gate,
$$\mathrm{N}_\varepsilon(\Phi_v) = \{\phi_v\},$$
where ϕ_v is the map from $\{1,-1\}^{X_v}$ to $\mathbb{C}$ defined by $\widehat{\Phi}_v$ – see Remark 1.2 (and so there is no noise in input gates). For example, if $\widehat{\Phi}_v = x_i$, then $\phi_v(1) = 1$ and $\phi_v(-1) = -1$.

Otherwise, v has two children v_1 and v_2. We note that although ϕ_{v_i} is a map from $\{1,-1\}^{X_{v_i}}$ to $\mathbb{C}$ we can naturally think of it as a map from $\{1,-1\}^{X_v}$ to $\mathbb{C}$ (for every $t \in \{1,-1\}^{X_v}$, set $\phi_{v_i}(t)$ to be $\phi_{v_i}(t')$, where t' is the restriction of t to the entries in X_{v_i}), and so the following is well defined.

- If v is a product gate,

$$\mathrm{N}_\varepsilon(\Phi_v) = \{\phi_{v_1} \cdot \phi_{v_2} \ : \ \phi_{v_1} \in \mathrm{N}_\varepsilon(\Phi_{v_1}) \ , \ \phi_{v_2} \in \mathrm{N}_\varepsilon(\Phi_{v_2})\}$$

  (and so there is no noise in edges going into product gates).

- If v is a sum gate,

$$\mathrm{N}_\varepsilon(\Phi_v) = \{(1+\alpha_1)\cdot\phi_{v_1} + (1+\alpha_2)\cdot\phi_{v_2} \ : \ \phi_{v_1} \in \mathrm{N}_\varepsilon(\Phi_{v_1}) \ , \ \phi_{v_2} \in \mathrm{N}_\varepsilon(\Phi_{v_2})\} \, ,$$

  where α_1, α_2 are arbitrary *real* values such that

$$0 \le \alpha_1 \le \varepsilon \ \text{ and } \ 0 \le \alpha_2 \le \varepsilon$$

  (and so the edges going into sum gates introduce a noise of 'magnitude' at most ε).

For a map $g : \{1,-1\}^n \to \mathbb{C}$, we say that Φ is *ε-noise-resistant to computing g*, if every ε-noisy value of Φ is 'correlated' with g; that is, for every $\phi \in \mathrm{N}_\varepsilon(\Phi)$,

$$\mathrm{cor}(\phi, g) \ge \varepsilon \cdot \|\phi\| \cdot \|g\| \tag{1.2}$$

(where we think of ϕ and g as maps from $\{1,-1\}^n$ to $\mathbb{C}$). So, for Φ to be noise-resistant to computing g, we only require all noisy values of Φ to be *weakly* correlated with g. We note that we could have introduced a new parameter (other than ε, that could, perhaps, be closer to 1) to bound the correlation in (1.2). We do not do so for simplicity of notation (and, once again, this only makes the lower bounds stronger).

Reading the definition above the reader may ask herself whether noise-resistant formulas exist. One example of a formula that is noise-resistant is a formula that is a sum of monomials – Every two *different* multilinear monomials m and m' in the variables X admit $\mathrm{cor}(\phi_m, \phi_{m'}) = 0$, where ϕ_m and $\phi_{m'}$ are the maps from $\{1,-1\}^n$ to $\mathbb{C}$ defined by m and m' respectively – see Remark 1.2. Thus, a polynomial of the form $\sum_i c_i m_i$ is not very sensitive to small changes in the c_i's (where the m_i's are distinct monic monomials and the c_i's are their coefficients).

1.1.2 Maximal-Partition Discrepancy

The discrepancy of a matrix is a well known and useful property, since it measures (in some sense) the amount of pseudo-randomness in a matrix. In computer science, it is connected to randomized communication complexity, extractors construction, and more. In combinatorics, it is connected to Ramsey theory.

The notion of maximal-partition discrepancy is a stricter measure of pseudo-randomness. We use known ideas to show that maximal-partition discrepancy is connected to communication complexity and extractors construction. Furthermore, we show a new connection between maximal-partition discrepancy and proving lower bounds for subclasses of arithmetic formulas.

We first recall the definition of the discrepancy of a matrix. Let M be an $N \times N'$ matrix with entries in $\{0,1\}$. A *rectangle* R in M is a set of the form $R = Y \times Z \subseteq [N] \times [N']$. The *discrepancy of a rectangle* R in M is the difference between the number of 1's and the number of 0's in R divided by the size of M; that is,

$$\textsc{Disc}_R(M) = \frac{1}{N \cdot N'} \cdot \left| \sum_{(y,z) \in R} (-1)^{M(y,z)} \right| .$$

The *discrepancy* of M is

$$\textsc{Disc}(M) = \max_R \textsc{Disc}_R(M),$$

where the maximum is over all rectangles R in M.

We now define maximal-partition discrepancy. Let f be a map from $\{0,1\}^n$ to $\{1,-1\}$, and let A be a subset of $\{1,\ldots,n\}$ of size k (we think of A as a partition of $\{1,\ldots,n\}$ into A and $\{1,\ldots,n\} \setminus A$). For $y \in \{0,1\}^k$ and $z \in \{0,1\}^{n-k}$, define f_A to be the $2^k \times 2^{n-k}$ matrix whose (y,z) entry is $f((y,z)_A)$, where $(y,z)_A$ is the unique vector in $\{0,1\}^n$ whose restriction to the entries in A is y and restriction to the entries not in A is z. The maximal-partition discrepancy of f is the maximal discrepancy of f_A among all sets A of size $n/3 \leq |A| \leq 2n/3$

1.1.3 Mixed-2-Source Extractors

Chor and Goldreich were among the first to consider weak sources of randomness, which are sources with min-entropy k [CG]. Extracting randomness from one weak source is impossible (as long as $k \leq n-1$). So, other sources of randomness were considered, such as two independent weak sources, and a few independent sources. We note that the study of extracting randomness from a few independent sources has advanced significantly lately [BIW, BKSSW, BRSW, R05, R] due to the well known sum-product theorem [BKT].

We introduce and analyze a new class of sources, that is a generalization of two independent sources, which we call mixed-2-sources. Given two independent sources of $n/2$ bits each and total min-entropy k, [CG] showed that the Hadamard matrix gives efficient extraction of one random bit for $k > n/2$ (we omit the dependency on the error term). The state of the art, due to Bourgain [Bo$^+$], is a 2-source extractor that gives a linear number of almost perfect bits for $k > n(1-\delta)/2$ (for some constant $\delta > 0$).

One way of thinking of a mixed-2-source extractor is as an extractor that works also when the bits of the two random sources arrive in a fixed but unknown order. This seems to be a natural relaxation of the well known notion of 2-source extractors, although, as far as we know, it has not been considered before. We also note that the Hadamard matrix does not give a mixed-2-source extractor even for $k = n - 4$ (in fact, the Hadamard extractor can be made constant for such a k).

We start with a few preliminary definitions and notation. Let μ be a distribution on $\{0,1\}^n$, and denote by $t \sim \mu$ an element distributed by μ. The *min-entropy* of μ is

$$\mathrm{H}_\infty(\mu) = \min_{t \in \{0,1\}^n} \log\left(\frac{1}{\mu(t)}\right);$$

that is, the min-entropy of μ is $k > 0$, if the most probable element in μ has probability 2^{-k}. We denote by U_n the uniform distribution on $\{0,1\}^n$. The *statistical distance* between μ and the uniform distribution U_n is

$$\|\mu - \mathrm{U}_n\|_1 = \sum_{t \in \{0,1\}^n} |\mu(x) - \mathrm{U}_n(x)|.$$

For two vectors t and t' in $\{0,1\}^n$, denote by $t \circ t' \in \{0,1\}^{2n}$ the concatenation of t and t'. For a one-to-one map π from $[2n]$ to $[2n]$, denote by $(t \circ t')_\pi \in \{0,1\}^{2n}$ the reordering of $t \circ t'$ according to π; that is, for every $i \in [2n]$, the i'th entry in $(t \circ t')_\pi$ is $(t \circ t')_{\pi(i)}$.

We now give the definition of a mixed-2-source extractor. For $n, m \in \mathbb{N}$ and $k, \varepsilon > 0$, a map $\textsc{Ext} : \{0,1\}^{2n} \to \{0,1\}^m$ is called a *mixed-2-source extractor* with k min-entropy requirement and error ε, if for every μ and μ', two independent distributions on $\{0,1\}^n$ such that

$$\mathrm{H}_\infty(\mu) + \mathrm{H}_\infty(\mu') \geq k,$$

and for every one-to-one map π from $[2n]$ to $[2n]$,

$$\|\textsc{Ext}((t \circ t')_\pi) - \mathrm{U}_m\|_1 \leq \varepsilon,$$

where $t \sim \mu$ and $t' \sim \mu'$.

A mixed-2-source extractor is stronger than a 2-source extractor. More specifically, a 2-source extractor is promised to extract random bits only when π is the identity map. We note that we think of π as being a fixed (but unknown) order in which the bits from the two random sources arrive.

1.1.4 Best-Partition Communication Complexity

Communication complexity was defined by Yao [Y], and has been studied extensively since. Different models of communications complexity are related to various areas in computer science. In particular, best-partition communication complexity is related to time/space tradeoffs for Very Large Scale Integration Circuits and to the width of branching programs (see [J]). Lam and Ruzzo proved an $\Omega(n)$ lower bound for randomized best-partition communication complexity, with error polynomially close to 1/2 [LR].

We now define the framework of randomized best-partition communication complexity. There are two players, Alice and Bob, that share a public random string of bits. There is a fixed boolean function $g : \{0,1\}^n \to \{0,1\}$ that they both know (and assume that n is even). Let A and B be a partition of $[n]$ into two sets of equal size. Given an input $x \in \{0,1\}^n$, Alice gets $x_A \in \{0,1\}^{n/2}$ and Bob gets $x_B \in \{0,1\}^{n/2}$ (where x_A is x restricted to the entries in A and x_B is x restricted to the entries in B). Alice does not know x_B and Bob does not know x_A. Their common goal is to compute $g(x)$.

The *randomized communication complexity of g with error $0 \leq \varepsilon \leq 1/2$, with respect to A and B*, is the number of bits Alice and Bob need to exchange in order to compute g (as above) with two-sided error ε (the error means that the probability of outputting the wrong answer is at most ε). The *randomized best-partition communication complexity of g with error ε* is the *minimal* randomized communication complexity of g with error ε, with respect to A and B, among all partitions of $[n]$ to two sets A and B of equal size.

1.2 Results

We start with a construction of a 'pseudo-random' polynomial, which will be used in the proofs of all our other results. For the rest of this section, $n = 12sp$ is an integer, where $p \in \mathbb{N}$ is prime and $s \in \mathbb{N}$ is a large enough constant (given in Theorem 2.1), and f is the multilinear polynomial over the set of variables $X = \{x_1, \ldots, x_n\}$ with coefficients in $\{1, -1\}$ defined below in Section 2.1. Roughly speaking, one can compute the coefficient of a monomial m in f as follows. Think of m as a zero-one vector in the natural way. Partition m to a constant number of blocks of equal size. Think of each of these blocks as a field element in the appropriate field. The coefficient of m in f is the first bit[1] of the field element that is obtained by multiplying all the blocks. In particular, f is in VNP, Valiant's algebraic analog of NP. We will also use the map g from $\{1,-1\}^n$ to $\{1,-1\}$ defined by

$$\forall\ t \in \{1,-1\}^n \qquad g(t) \text{ is the coefficient of the monomial } \prod_{\substack{i\in[n]:\\ t_i=-1}} x_i \text{ in } f. \tag{1.3}$$

The property of f that we use is given by the following theorem. A multilinear polynomial $f' \in \mathbb{C}[X]$ is called a *product* polynomial, if there exist two disjoint sets $X_1, X_2 \subseteq X$ of size at least $n/3$ each, and two polynomials $f_1 \in \mathbb{C}[X_1]$ and $f_2 \in \mathbb{C}[X_2]$ such that

$$f' = f_1 \cdot f_2 \tag{1.4}$$

(see Section 3.1 for more details).

Theorem 1.3. *Every product multilinear polynomial $f' \in \mathbb{C}[X]$ admits*

$$cor(f,f') \leq 2^{-\Omega(n)}\|f\|\|f'\|,$$

where f is the polynomial defined in Section 2.1, and we think of f and f' as vectors of coefficients.

The proof of Theorem 1.3 is in Section 2.2. A key ingredient in the proof is an exponential sum estimate of Bourgain, Glibichuk and Konyagin [BoGK]. A corollary of Theorem 1.3 is that g has small maximal-partition discrepancy, the corollary is proved in Section 2.1. To prove the corollary, we observe that correlation with product polynomials is a generalization of maximal-partition discrepancy: roughly, every rectangle R can be represented by a product polynomial f', in the sense that the discrepancy with respect to R is the correlation with f' (for a formal statement see Section 2.1). We now explain the main difference between correlation with product polynomials and maximal-partition discrepancy. Every rectangle can be thought of as its characteristic vector, which has zero-one entries. On the other hand, a product polynomial can be thought of as a vector with arbitrary complex entries. The difference between these two notions is thus in that for maximal-partition discrepancy we consider correlation with zero-one vectors, and in correlation with product polynomials we consider arbitrary complex vectors.

Corollary 1.4. *The maximal-partition discrepancy of g is $2^{-\Omega(n)}$, where g is the map defined in (1.3).*

[1] More precisely, $(-1)^b$, where b is the first bit of of the field element that is obtained by multiplying all the blocks.

We have learned that an alternative construction of a function with exponentially small maximal-partition discrepancy – that is based on expanders graphs and the Hadamard matrix – is implicit in the work of Hayes [H] (unpublished manuscript). The same construction was independently discovered and suggested to us by Wigderson [W].

We now turn to our results regarding arithmetic circuits. The polynomial f has negative coefficients, and so it can not be computed by a monotone circuit. However, we can use f to define a new polynomial $F \in \mathbb{C}[X]$ with coefficients in $\{0, 1\}$, for which we will also be able to prove lower bounds. The polynomial F is defined as follows: for a monic monomial m in the variables X, the coefficient of m in F is

$$\frac{f_m + 1}{2} \in \{0, 1\},$$

where f_m is the coefficient of m in f.

The following theorem gives a tight lower bound for the size of monotone arithmetic circuits for F (a monotone multilinear polynomial always has a monotone circuit of size $2^{O(n)}$).

Theorem 1.5. *Let Φ be a monotone arithmetic circuit over the field $\mathbb{R}$ and over the set of variables X computing the polynomial F defined above. Then,*

$$|\Phi| = 2^{\Omega(n)}.$$

We note that since F is multilinear, any monotone circuit for F is, in particular, syntactically multilinear. However, monotone circuits are much more restricted than syntactically multilinear circuits, and indeed our results for monotone circuits are much stronger. We also note that the lower bound proof for monotone circuits already gives the spirit of the proofs for multilinear formulas (we explain it in more detail after Theorem 1.8 below). The proof of Theorem 1.5 is in Section 3.2.

The following theorem gives a tradeoff between the depth and τ (the "amount of non-cancelling") for a syntactically multilinear arithmetic formula computing f. For example, a $2^{-\sqrt{n}}$-non-cancelling syntactically multilinear arithmetic formula that is at least $2^{-\sqrt{n}}$ correlated with f is of depth $\Omega(\sqrt{n})$. Note that the smaller τ is, the less restricted the formula is. So, for proving a lower bound, the smaller τ is, the stronger the lower bound is.

Theorem 1.6. *Let $\tau, c > 0$, and let Φ be a τ-non-cancelling syntactically multilinear arithmetic formula of depth $d \in \mathbb{N}$ over the field $\mathbb{C}$ and over the set of variables X such that*

$$cor(\widehat{\Phi}, f) \geq c \cdot \|\widehat{\Phi}\| \cdot \|f\|,$$

where f is the polynomial defined in Section 2.1, and we think of $\widehat{\Phi}$ and f as vectors of coefficients. Then,

$$|\Phi| \cdot \tau^{-d} \geq c \cdot 2^{\Omega(n)}.$$

In particular, if $\tau < 2$ and $c \geq 1/2$,

$$d = \Omega\left(\frac{n}{\log(2/\tau)}\right),$$

and if $\tau \geq 1$ and $c \geq 1/2$,

$$|\Phi| = 2^{\Omega(n)}.$$

Since we do not know how to balance arithmetic formulas in the non-cancelling model, Theorem 1.6 does not imply an exponential lower bound for the size (for small τ). However, since every orthogonal arithmetic formula is 1-non-cancelling, we have the following exponential lower bound for the size of orthogonal syntactically multilinear arithmetic formulas computing f.

Corollary 1.7. *Let Φ be an orthogonal syntactically multilinear arithmetic formula over the field $\mathbb{C}$ and over the set of variables X computing f, where f is the polynomial defined in Section 2.1. Then,*

$$|\Phi| = 2^{\Omega(n)}.$$

A similar trade-off holds for a noise-resistant computation of f. For example, a syntactically multilinear arithmetic formula that is $2^{-\sqrt{n}}$-noise-resistant to computing g is of depth $\Omega(\sqrt{n})$.

Theorem 1.8. *Let $0 < \varepsilon < 1$, and let Φ be a syntactically multilinear arithmetic formula of depth $d \in \mathbb{N}$ over the field $\mathbb{C}$ and over the set of variables X that is ε-noise-resistant to computing g, where g is defined in (1.3). Then,*

$$d = \Omega\left(\frac{n}{\log(2/\varepsilon)}\right).$$

The proofs of the lower bounds are in Section 3. We now describe the main ideas behind the proofs. To do so, we focus on the lower bound proof for monotone circuits, as the proofs for multilinear formulas are more technical. The high-level idea is the so called *discrepancy method*: roughly speaking, the proof follows by defining a notion of discrepancy for monotone circuits so that a small circuit computes a polynomial with large discrepancy, and so every polynomial with small discrepancy requires large monotone circuits. Here is a more detailed description of the proof.

The first step is showing that polynomials that are computed by monotone circuits have a ‘special’ structure: they can be represented as a sum of what we call *product* polynomials (see Section 1.1.2 for the definition of product polynomials). More specifically, a multilinear polynomial h that is computed by a monotone circuit of size roughly s can be written as $h = \sum_{i=1}^{s} h_i$ for product polynomials $h_1, \ldots, h_s$. This representation of monotone circuits follows by an inductive argument. Intuitively, it shows that small monotone circuits are well correlated with product polynomials.

The second step of the proof employs the fact that f (defined below) has 'small' correlation with any product polynomial to conclude the lower bound: Let F be the monotone polynomial defined above using f, that is, F_m is $(f_m + 1)/2$ for every monomial m. Assume that F has more 1's than 0's; that is, $\mathrm{cor}(f, F)$ is large. Also assume that F can be computed by a monotone circuit of size s. Using the first step, $F = \sum_{i=1}^{s} h_i$ for product polynomials $h_1, \ldots, h_s$. Thus, $\mathrm{cor}(f, h_i) \geq \mathrm{cor}(f, F)/s$ for some i. Since f has small correlation with any product polynomial, and in particular with h_i, it follows that s must be large.

We note that there are some subtleties that we did not address in the discussion above. We also note that correlation between functions is usually defined with respect to the output of the functions. Here we define the correlation between polynomial as the correlation between their coefficients.

To prove the lower bounds for multilinear formulas, we need a structural understanding, that is similar in nature to (but more technical than) the one given in the first step above. To make this understanding formal, we define a new type of formulas, which we call *sum trees*. The role of these sum trees is similar to the role of the sum given in the first step above.

Finally, we state the results regarding extractors and communication complexity.

The following theorem gives an efficient map that extracts a linear number of almost perfect random bits from a mixed-2-source of randomness of high min-entropy.

Theorem 1.9. *There exists a constant $\beta > 0$ such that the following holds. Let $n = 12sp$ be an even integer, where $p \in \mathbb{N}$ is prime and $s \in \mathbb{N}$ is the constant given in Theorem 2.1. Then, there exists an explicit mixed-2-source extractor* $\textsc{Ext} : \{0,1\}^n \to \{0,1\}^m$ *with $m = \lfloor \beta n \rfloor$, that is computable in deterministic polynomial time with $(n - 3m)$ min-entropy requirement and error 2^{-2m}.*

The proof of Theorem 1.9 is in Section 4.

The following theorem lower bounds the randomized best-partition communication complexity of g.

Theorem 1.10. *There exists a constant $\beta > 0$ such that for every $\varepsilon \leq 1/2 - 2^{-\beta n}$, the randomized best-partition communication complexity of g with error ε is $\Omega(n)$, where g is the map defined in (1.3).*

The proof of Theorem 1.10 follows using standard methods in communication complexity (see, e.g., [KN]), and using the exponentially small maximal-partition discrepancy of g.

2 The Explicit Construction

In this section we construct a multilinear polynomial f that is 'uncorrelated' with any multilinear product polynomial, that is, every product multilinear polynomial $f' \in \mathbb{C}[X]$ admits

$$\mathrm{cor}(f, f') \leq 2^{-\Omega(n)} \|f\| \|f'\|$$

(see Theorem 1.3).

2.1 Definition of f

We start with a few preliminaries.

Let $p \in \mathbb{N}$ be a prime integer, and let $\mathbb{F} = \mathrm{GF}(2^p)$ be the field of size 2^p. Every $y \in \mathbb{F}$ can be thought of as a vector $(y_1, \ldots, y_p) \in \{0,1\}^p$. The *inner product* of two field elements $y = (y_1, \ldots, y_p)$ and $z = (z_1, \ldots, z_p)$ is defined as

$$\langle y, z \rangle = \sum_{i \in [p]} y_i z_i \in \{0,1\}$$

(where the sum is modulo 2). For $z \in \mathbb{F}$, define the map $\psi_z : \mathbb{F} \to \mathbb{C}$ as

$$\forall\ y \in \mathbb{F} \quad \psi_z(y) = (-1)^{\langle z, y \rangle}.$$

So, every y and y' in $\mathbb{F}$ admit

$$\psi_z(y + y') = \psi_z(y) \cdot \psi_z(y'). \tag{2.1}$$

The map ψ_z is called *an additive character* of $\mathbb{F}$. If z is non-zero, then ψ_z is called a *non-trivial* additive character of $\mathbb{F}$. So, the image of a non-trivial character is $\{1, -1\}$.

Let $n = 12sp$ be an integer, where $s \in \mathbb{N}$ is the constant given in Theorem 2.1. Let $X = \{x_1, \ldots, x_n\}$ be a set of variables, and let $\mathbb{F}$ be the field of size 2^p. Recall that we think of field elements in $\mathbb{F}$ also as vectors in $\{0,1\}^p$. For a multilinear monomial m over the set of variables X and for $i \in [12s]$, we denote by $y_i = y_i(m) \in \mathbb{F}$ the field element defined as

$$\forall\ j \in [p] \quad (y_i)_j = \text{the degree of } x_{p(i-1)+j} \text{ in } m.$$

Roughly speaking, given a monomial m, we interpret m as a $\{0,1\}^n$ vector, and then we cut m into $12s$ blocks of equal size; these blocks are $y_1, \ldots, y_{12s}$.

Let ψ be an arbitrary non-trivial additive character of $\mathbb{F}$ (we note that the fact that ψ is arbitrary will be used in Section 4 in the proof that the extractor works).

We define the multilinear polynomial $f \in \mathbb{C}[X]$ by defining the coefficients of the monomials in f. Roughly speaking, to define the coefficient of a monomial in f, we partition the monomial to $12s$ equal parts, each of size p, we view each of these parts as a field element, and we multiply them. Formally, let m be a monic multilinear monomial over the set of variables X. For every $i \in [12s]$, let $y_i = y_i(m) \in \mathbb{F}$ be the field element defined above. Define the coefficient of m in f to be

$$\psi(y_1 \cdot y_2 \cdots y_{12s}) \in \{1, -1\} .$$

Before proving Theorem 1.3, we show how small correlation with product polynomials implies low maximal-partition discrepancy.

Proof of Corollary 1.4. The idea is to simulate a rectangle with a product polynomial.

Let A be a subset of $[n]$ of size k, where $n/3 \leq k \leq 2n/3$, and let $\bar{A} = [n] \setminus A$. Recall that g_A is the $2^k \times 2^{n-k}$ matrix whose (y, z) entry, $y \in \{0,1\}^k$ and $z \in \{0,1\}^{n-k}$, is $g((y,z)_A)$, where $(y,z)_A$ is the unique vector w in $\{0,1\}^n$ such that $w_A = y$ and $w_{\bar{A}} = z$ (the vector w_A, e.g., is the restriction of w to the entries in A). Let $R = Y \times Z \subset \{0,1\}^k \times \{0,1\}^{n-k}$ be a rectangle so that $\textsc{Disc}(g_A) = \textsc{Disc}_R(g_A)$.

Let X_1 be the set of variables x_i so that $i \in A$, and let X_2 be the complement of X_1. Define the polynomial $f_1 \in \mathbb{R}[X_1]$ so that it simulates Y, that is, the coefficient of the monomial $\prod_{i : y_i = 1} x_i$ in f_1 is 1 if and only if $y \in Y$ (otherwise, it is 0). Similarly, define the polynomial $f_2 \in \mathbb{R}[X_2]$ so that it simulates Z. The polynomial $f_1 f_2$ is a product polynomial that simulates R. Theorem 1.3 implies

$$\begin{aligned} \textsc{Disc}(g_A) = 2^{-n} \Big| \sum_{(y,z) \in R} (-1)^{g_A(y,z)} \Big| &= 2^{-n} \Big| \sum_{w} f_1(w_A) f_2(w_{\bar{A}}) f(w) \Big| \\ &= 2^{-n} \mathrm{cor}(f, f_1 f_2) \leq 2^{-n} 2^{-\Omega(n)} 2^{n/2} 2^{n/2} = 2^{-\Omega(n)}, \end{aligned}$$

where we used $\|f\|, \|f_1 f_2\| \leq 2^{n/2}$. The corollary follows, as A is arbitrary. □

2.2 Proof of Theorem 1.3

Before proving the theorem, we give a high level view of it. Since discrepancy is a well-known notion, to explain the structure of the proof, we consider discrepancy rather than correlation with product polynomials. For every set $A \subset [n]$ of size roughly $n/2$, we have a matrix M_A. Our goal is to prove that the discrepancy of every M_A is small. How do we show that the discrepancy of a matrix is small? Well, it is known that Hadamard matrices (i.e., matrices whose rows are orthogonal) have low discrepancy (see, e.g., [LS]). The matrix M_A turns out to be close enough to a Hadamard matrix, so that it also has small discrepancy: most of the rows of M_A are almost orthogonal. To prove this we partition the rows of M_A to two sets S_1, S_2, according to equations (2.5) and (2.6) below. The set S_1 turns out to be small enough so that we can ignore it; see Proposition 2.3 below. To prove that S_2 consists of almost orthogonal vectors we use the following exponential sum estimate[2] from [BoGK, Bo]; see Proposition 2.4 below. Roughly, the exponential sum estimate tells us that vectors of certain form are orthogonal.

Theorem 2.1. *There exist two constants, an integer $s \in \mathbb{N}$ and $\beta > 0$, such that for every prime $p \in \mathbb{N}$, for every family of sets $A_1, \ldots, A_s \subseteq GF(2^p)$ of size at least $2^{p/4}$ each, for every non-zero field element $z \in GF(2^p)$, and for every non-trivial additive character ψ of $GF(2^p)$,*

$$\left| \sum_{y_1 \in A_1, \ldots, y_s \in A_s} \psi(z \cdot y_1 \cdot y_2 \cdots y_s) \right| \leq 2^{-\beta \cdot p} \cdot |A_1| \cdot |A_2| \cdots |A_s|.$$

Recall the Cauchy-Schwarz inequality: for every $N \in \mathbb{N}$ and for every two vectors $(w_1, \ldots, w_N)$ and $(t_1, \ldots, t_N)$ in $\mathbb{C}^N$,

$$\Big| \sum_{\ell \in [N]} w_\ell \overline{t_\ell} \Big|^2 \leq \Big(\sum_{\ell \in [N]} |w_\ell|^2 \Big) \Big(\sum_{\ell \in [N]} |t_\ell|^2 \Big).$$

In this proof we use the following notation. For a multilinear polynomial F in $\mathbb{C}[X]$ and for a multilinear monomial m in the variables X, we denote by $F(m) \in \mathbb{C}$ the coefficient of m in F. This may be misleading, as F is also a function, but we do so for simplicity of notation. We note that in this section we will think of a polynomial always as a vector of coefficients, and not as a function.

Let $f' \in \mathbb{C}[X]$ be a product multilinear polynomial. Thus, there exists a partition of X into two sets A and B (i.e., $A \cup B = X$ and $A \cap B = \emptyset$) of size at least $n/3$ each, and two multilinear polynomials $g \in \mathbb{C}[A]$ and $h \in \mathbb{C}[B]$ such that

$$f' = gh.$$

The proof continues as follows. We will identify two sets $A_1 \subseteq A$ and $B_1 \subseteq B$ that will enable us to use the exponential sum estimate of [BoGK] to bound the correlation between f and f'. We will then give some notation, and finally we will bound the correlation between f and f'.

[2] We state a weaker result than the result of [BoGK].

2.2.1 Identifying A_1 and B_1

For $i \in [12s]$, set

$$X(i) = \left\{ x_{(i-1)p+j} \ : \ j \in [p] \right\},$$

and set

$$A(i) = A \cap X(i) \text{ and } B(i) = B \cap X(i).$$

The following proposition will give A_1 and B_1 (see (2.2) and (2.3) below).

Proposition 2.2. *There exists a set $I \subseteq [12s]$ of size s such that for every $i \in I$,*

$$|A(i)| \geq p/4.$$

Proof. Let I' be the set of $i \in [12s]$ such that $|A(i)| \geq p/4$. Since $|A| \geq n/3$, we have

$$4sp \leq |A| \leq |I'| \cdot p + (12s - |I'|) \cdot p/4,$$

which implies $|I'| > s$. Set I to be a subset of I' of size s. □

Let $I \subseteq [12s]$ be the set given by Proposition 2.2, and let $J = [12s] \setminus I$. Set

$$A_1 = \bigcup_{i \in I} A(i) \quad \text{and} \quad A_2 = A \setminus A_1, \tag{2.2}$$

and set

$$B_1 = \bigcup_{i \in J} B(i) \quad \text{and} \quad B_2 = B \setminus B_1. \tag{2.3}$$

So, since $|B| \geq n/3$, every $i \in I$ admits $|B(i)| \leq p$ and $|I| = s$, we have

$$|B_1| \geq |B| - sp \geq 3sp. \tag{2.4}$$

Notation. For a set of variables $T \subseteq X$, we write t (or t') when t (or t') is a monic multilinear monomial in the variables T. For example, b_1 (or b'_1) is a monic multilinear monomial in the variables B_1. Recall that $f(m) \in \mathbb{C}$ is the coefficient of the monomial m in f, and recall that for $i \in [12s]$, the field element $y_i = y_i(m) \in \mathbb{F}$ is defined as

$$\forall\ j \in [p] \quad (y_i)_j = \text{the degree of } x_{p(i-1)+j} \text{ in } m.$$

For a monomial a_2 over the set of variables A_2, and for two monomials b_1 and b'_1 over the set of variables B_1, we denote

$$\mathrm{Z}(a_2, b_1, b'_1) = \prod_{i \in J} y_i(a_2 b_1) - \prod_{i \in J} y_i(a_2 b'_1) \in \mathbb{F}.$$

Denote by $S(a_2)$ the set of pairs (b_1, b'_1) such that $\mathrm{Z}(a_2, b_1, b'_1) = 0$. Denote

$$S_1 = \left\{a_2 \ : \ |S(a_2)| > 2^{2|B_1| - p/12}\right\}, \tag{2.5}$$

and denote

$$S_2 = \left\{a_2 \ : \ |S(a_2)| \leq 2^{2|B_1| - p/12}\right\} \tag{2.6}$$

(the complement set of S_1).

Here is some intuition for the definitions above. Recall the high-level description of the proof given in the beginning of the section. To prove the theorem, we use the Cauchy-Schwarz inequality to "isolate" certain parts of the sum we wish to bound. This is a common use of the Cauchy-Schwarz inequality, and the field element Z comes up naturally after applying the Cauchy-Schwarz inequality. The reason we are interested in considering S_1 and S_2 separately is that the only case in which Theorem 2.1 does not hold is when $z = 0$. For S_1, when many Z's are zero, we can not use Theorem 2.1, and we need to use a different consideration, which turns out to be a counting argument. For S_2, when not many Z's are zero, we can use Theorem 2.1 to complete the proof.

2.2.2 Bounding the Correlation Between f and f'

Recall that

$$\mathrm{cor}(f, f') = \mathrm{cor}(f, gh) = \left| \sum_{a_1, a_2, b_1, b_2} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right|,$$

where the sum is over all monomials a_1 in the variables A_1, all monomials a_2 in the variables A_2, all monomials b_1 in the variables B_1 and all monomials b_2 in the variables B_2.

Denote

$$C_1 = \left| \sum_{a_2 \in S_1} \sum_{a_1, b_1, b_2} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right| ,$$

and

$$C_2 = \left| \sum_{a_2 \in S_2} \sum_{a_1, b_1, b_2} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right| .$$

Therefore,

$$\mathrm{cor}(f, f') \leq C_1 + C_2.$$

We bound the correlation between f and f' by bounding C_1 and C_2.

Proposition 2.3. *There exists a constant $\beta_1 > 0$ such that*

$$C_1 \leq 2^{-\beta_1 p} \|f\| \|f'\|.$$

Proposition 2.4. *There exists a constant $\beta_2 > 0$ such that*

$$C_2 \leq 2^{-\beta_2 p} \|f\| \|f'\|.$$

The proof of Proposition 2.3 is in Section 2.2.3, and the proof of Proposition 2.4 is in Section 2.2.4. Using Propositions 2.3 and 2.4, since $p = \Omega(n)$, we have

$$\mathrm{cor}(f, f') \leq 2^{-\Omega(n)} \|f\| \|f'\|,$$

which completes the proof of Theorem 1.3. □

2.2.3 Proof of Proposition 2.3

Recall that we want to bound from above

$$C_1 = \left| \sum_{a_2 \in S_1} \sum_{a_1, b_1, b_2} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right| ,$$

where

$$S_1 = \left\{ a_2 \ : \ |S(a_2)| > 2^{2|B_1| - p/12} \right\} .$$

First, we will bound the size of S_1 from above. We denote by S the set of triplets (a_2, b_1, b_1') such that $\mathrm{Z}(a_2, b_1, b_1') = 0$. To bound the size of S_1 we bound the size of S.

Claim 2.5. *For every large enough p,*

$$|S| \leq 2^{2|B_1|+|A_2|-p/6}.$$

Proof. We will first bound the number of triplets (a_2, b_1, b'_1) such that

$$\prod_{i \in J} y_i(a_2 b'_1) = 0. \tag{2.7}$$

Since

$$A_2 \cup B_1 = \bigcup_{i \in J} X(i),$$

and since $|J| = 11s$, all the monomials of the form $a_2 b'_1$ are all the 2^{11sp} monomials in the variables $\bigcup_{i \in J} X(i)$. Note that for every monomial $a_2 b'_1$,

$$\forall\ i \in J \ \ y_i(a_2 b'_1) \neq 0 \quad \Leftrightarrow \quad \prod_{i \in J} y_i(a_2 b'_1) \neq 0,$$

and that for every $i \in J$, the number of pairs (a_2, b'_1) for which $y_i(a_2 b'_1) = 0$ is $2^{|B_1|+|A_2|-p}$. So, by the union bound, the number of pairs (a_2, b'_1) for which (2.7) holds is at most

$$|J| 2^{|B_1|+|A_2|-p} = 11s 2^{|B_1|+|A_2|-p}.$$

Hence, the number of triplets (a_2, b_1, b'_1) for which (2.7) holds is at most

$$2^{|B_1|} \cdot 11s 2^{|B_1|+|A_2|-p} = 11s 2^{2|B_1|+|A_2|-p}.$$

We will now bound the number of triplets in S for which (2.7) does not hold. Since $|B_1| \geq 3sp$ (see (2.4)), there exists $j \in J$ such that

$$|B(j)| \geq p/4.$$

The number of triplets in S for which (2.7) does not hold is at most the number of triplets (a_2, b_1, b'_1) in S such that

$$y_j(a_2 b'_1) = \frac{\prod_{i \in J} y_i(a_2 b_1)}{\prod_{i \in J \setminus \{j\}} y_i(a_2 b'_1)}$$

(note that $\prod_{i \in J \setminus \{j\}} y_i(a_2 b'_1)$ is non-zero). So, the number of triplets in S for which (2.7) does not hold is at most

$$2^{2|B_1|+|A_2|-p/4}.$$

We conclude that, for large enough p,

$$|S| \leq 11s2^{2|B_1|+|A_2|-p} + 2^{2|B_1|+|A_2|-p/4} \leq 2^{2|B_1|+|A_2|-p/6}.$$

□

The following corollary bounds the size of S_1.

Corollary 2.6. *For every large enough p,*

$$\left|S_1\right| \leq 2^{|A_2|-p/12}.$$

Proof. Using Claim 2.5, for every large enough p,

$$2^{2|B_1|+|A_2|-p/6} \geq |S| = \sum_{a_2} |S(a_2)| > |S_1| \cdot 2^{2|B_1|-p/12}.$$

So, for large enough p,

$$|S_1| \leq 2^{|A_2|-p/12}.$$

□

Back to the proof of Proposition 2.3. Recall that

$$C_1 = \left| \sum_{a_2 \in S_1} \sum_{a_1,b_1,b_2} f(a_1a_2b_1b_2)\overline{f'(a_1a_2b_1b_2)} \right|.$$

By the Cauchy-Schwarz inequality,

$$C_1 \leq \sqrt{\sum_{a_2 \in S_1} \sum_{a_1,b_1,b_2} |f(a_1a_2b_1b_2)|^2} \sqrt{\sum_{a_2 \in S_1} \sum_{a_1,b_1,b_2} |f'(a_1a_2b_1b_2)|^2}.$$

Since the coefficients of f are in $\{1,-1\}$ and since the sum is only over $a_2 \in S_1$,

$$C_1 \leq \sqrt{|S_1| 2^{|A_1|+|B_1|+|B_2|}} \|f'\|.$$

By Corollary 2.6, for every large enough p,

$$C_1 \leq 2^{(|A_1|+|A_2|+|B_1|+|B_2|)/2-p/24} \|f'\|.$$

Thus, since $\|f\| = 2^{n/2}$ and since $|A_1|+|A_2|+|B_1|+|B_2| = n$, there exists a constant $\beta_1 > 0$ such that

$$C_1 \leq 2^{-\beta_1 p} \|f\| \|f'\|,$$

which completes the proof of the proposition. □

2.2.4 Proof of Proposition 2.4

Recall that we want to bound from above

$$C_2 = \left| \sum_{a_2 \in S_2} \sum_{a_1,b_1,b_2} f(a_1a_2b_1b_2)\overline{g(a_1a_2)h(b_1b_2)} \right|,$$

where

$$S_2 = \left\{ a_2 \;:\; |S(a_2)| \leq 2^{2|B_1|-p/12} \right\}.$$

We first prove the following claim.

Claim 2.7. *There exists a constant $\beta_3 > 0$ such that for every multilinear monomial a_2 over the set of variables A_2, and for every multilinear monomial b_2 over the set of variables B_2,*

$$\sum_{b_1,b_1'} \left| \sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)} \right|^2 \leq 2^{2|A_1|}\Big(|S(a_2)| + 2^{2|B_1|-\beta_3 p} \Big).$$

Proof. Let a_2 be a multilinear monomial over the set of variables A_2, and let b_2 be a multilinear monomial over the set of variables B_2. For every $i \in J$, we have that y_i does not depend on the variables in either A_1 or B_2. Similarly, for every $i \in I$, we have that y_i does not depend on the variables in either A_2 or B_1. Let a_1 be a multilinear monomial over the set of variables A_1, and let b_1 and b_1' be two multilinear monomials over the set of variables B_1. Thus,

$$\begin{aligned} \prod_{i\in[12s]} y_i(a_1a_2b_1b_2) - \prod_{i\in[12s]} y_i(a_1a_2b_1'b_2) &= \prod_{i\in I} y_i(a_1b_2) \prod_{i\in J} y_i(a_2b_1) - \prod_{i\in I} y_i(a_1b_2) \prod_{i\in J} y_i(a_2b_1') \\ &= \prod_{i\in I} y_i(a_1b_2) \Big(\prod_{i\in J} y_i(a_2b_1) - \prod_{i\in J} y_i(a_2b_1') \Big) \\ &= \mathrm{Z}(a_2,b_1,b_1') \prod_{i\in I} y_i(a_1b_2) \end{aligned}$$

(by the definition of $\mathrm{Z}(a_2,b_1,b_1')$). Recall that

$$\mathrm{Z}(a_2,b_1,b_1') = 0 \Leftrightarrow (b_1,b_1') \in S(a_2). \tag{2.8}$$

Thus, by the definition of f, since ψ is an additive character of $\mathbb{F}$ (using (2.1)),

$$\left| \sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)} \right| = \left| \sum_{a_1} \psi\left(\mathrm{Z}(a_2,b_1,b_1') \prod_{i\in I} y_i(a_1b_2) \right) \right|.$$

Denote by $i_1, \ldots, i_s$ the elements of I. For all $j \in [s]$, denote

$$A_1(j) = A_1 \cap X(i_j).$$

So, $A_1(1), \ldots, A_1(s)$ is a partition of A_1. In the following sums $a_1(j)$ is a monomial in the variables $A_1(j)$. By Proposition 2.2, for all $j \in [s]$,

$$|A_1(j)| \geq p/4.$$

Therefore, if $(b_1, b_1') \notin S(a_2)$, then, by (2.8) and by Theorem 2.1, there exists a constant $\alpha > 0$ such that

$$\begin{aligned}\left|\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\right| &= \left|\sum_{a_1(1),\cdots,a_1(s)} \psi\left(\mathrm{Z}(a_2,b_1,b_1')\prod_{i\in I} y_i(a_1b_2)\right)\right| \\ &< 2^{-\alpha p+|A_1(1)|+|A_1(2)|\cdots+|A_1(s)|} \\ &= 2^{-\alpha p+|A_1|}.\end{aligned}$$

Also if $(b_1, b_1') \in S(a_2)$, then

$$\left|\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\right| \leq 2^{|A_1|}.$$

Therefore,

$$\sum_{b_1,b_1'}\left|\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\right|^2 \leq |S(a_2)|2^{2|A_1|} + 2^{2|B_1|}2^{-2\alpha p+2|A_1|}.$$

So, there exists a constant $\beta_3 > 0$ such that

$$\sum_{b_1,b_1'}\left|\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\right|^2 \leq 2^{2|A_1|}\left(|S(a_2)| + 2^{2|B_1|-\beta_3 p}\right).$$

□

We will use the following corollary.

Corollary 2.8. *There exists a constant $\beta_4 > 0$ such that*

$$\sum_{a_2\in S_2}\sum_{b_2}\left|\sum_{a_1,b_1} f(a_1a_2b_1b_2)\overline{g(a_1a_2)h(b_1b_2)}\right|^2 \leq 2^{|A_1|+|B_1|-\beta_4 p}\|g\|^2\|h\|^2.$$

Proof. Denote

$$R = \sum_{a_2\in S_2}\sum_{b_2}\left|\sum_{a_1,b_1} f(a_1a_2b_1b_2)\overline{g(a_1a_2)h(b_1b_2)}\right|^2 .$$

So,

$$R \;=\; \sum_{a_2\in S_2}\sum_{b_2}\left|\sum_{a_1}\overline{g(a_1a_2)}\sum_{b_1} f(a_1a_2b_1b_2)\overline{h(b_1b_2)}\right|^2 .$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
R \;&\leq\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\Big(\sum_{a_1}\left|\sum_{b_1} f(a_1a_2b_1b_2)\overline{h(b_1b_2)}\right|^2\Big)\\
&=\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\Big(\sum_{b_1,b_1'}\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)h(b_1b_2)}h(b_1'b_2)\Big)\\
&=\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\Big(\sum_{b_1,b_1'}\overline{h(b_1b_2)}h(b_1'b_2)\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\Big).
\end{aligned}$$

Again, using the Cauchy-Schwarz inequality,

$$R \;\leq\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\sqrt{\sum_{b_1,b_1'}|\overline{h(b_1b_2)}h(b_1'b_2)|^2}\sqrt{\sum_{b_1,b_1'}\left|\sum_{a_1} f(a_1a_2b_1b_2)\overline{f(a_1a_2b_1'b_2)}\right|^2}.$$

So, using Claim 2.7,

$$R \;\leq\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\sqrt{\left|\sum_{b_1}|h(b_1b_2)|^2\right|^2}\sqrt{2^{2|A_1|}\Big(|S(a_2)|+2^{2|B_1|-\beta_3 p}\Big)}.$$

So, by the definition of S_2, for large enough p, there exists a constant $\beta_4 > 0$ such that

$$\begin{aligned}
R \;&\leq\; \sum_{a_2\in S_2}\sum_{b_2}\Big(\sum_{a_1}|g(a_1a_2)|^2\Big)\Big(\sum_{b_1}|h(b_1b_2)|^2\Big)\sqrt{2^{2|A_1|}\Big(2^{2|B_1|-p/12}+2^{2|B_1|-\beta_3 p}\Big)}\\
&\leq\; 2^{|A_1|+|B_1|-\beta_4 p}\|g\|^2\|h\|^2.
\end{aligned}$$

□

Back to the proof of Proposition 2.4. Recall that

$$C_2 = \left| \sum_{a_2 \in S_2} \sum_{b_2} 1 \sum_{a_1, b_1} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right|.$$

So, using Corollary 2.8 and the Cauchy-Schwarz inequality,

$$\begin{aligned} C_2 &\leq \sqrt{\sum_{a_2 \in S_2} \sum_{b_2} 1^2} \sqrt{\sum_{a_2 \in S_2} \sum_{b_2} \left| \sum_{a_1, b_1} f(a_1 a_2 b_1 b_2) \overline{g(a_1 a_2) h(b_1 b_2)} \right|^2} \\ &\leq 2^{|A_2|/2 + |B_2|/2} 2^{|A_1|/2 + |B_1|/2 - \beta_4 p/2} \|g\| \|h\|. \end{aligned}$$

By Claim 3.2, we have $\|f'\| = \|g\| \|h\|$. Thus, since $\|f\| = 2^{n/2}$ and since $|A_1| + |A_2| + |B_1| + |B_2| = n$, there exists a constant $\beta_2 > 0$ such that

$$C_2 \leq 2^{-\beta_2 p} \|f\| \|f'\|,$$

which completes the proof of the proposition. □

3 Lower Bounds for Arithmetic Circuits

3.1 Product Polynomials and Norms

Let $n \geq 3$ be an integer, and let $X = \{x_1, \ldots, x_n\}$. We say that a multilinear polynomial $f \in \mathbb{C}[X]$ is a *product* polynomial, if there exist two disjoint sets $X_1, X_2 \subseteq X$ of size at least $n/3$ each, and two polynomials $f_1 \in \mathbb{C}[X_1]$ and $f_2 \in \mathbb{C}[X_2]$ such that

$$f = f_1 \cdot f_2 \tag{3.1}$$

We say that a variable $x \in X$ *occurs* in a polynomial $f \in \mathbb{C}[X]$, if the degree of x in f is at least 1.

We will use the following claim.

Claim 3.1. *Let $n \geq 3$ be an integer, and let $X = \{x_1, \ldots, x_n\}$. Let $f \in \mathbb{C}[X]$ be a product polynomial. Let $T \subseteq X$ be such that $|T| \leq n/3$. Let $g \in \mathbb{C}[T]$ be a polynomial such that $f \cdot g$ is multilinear. Then, the polynomial $f \cdot g$ is a product polynomial as well.*

Proof. Let $f = f' \cdot f''$, where $f' \in \mathbb{C}[X']$ and $f'' \in \mathbb{C}[X'']$ are the two polynomials given by the fact that f is a product polynomial. Let $T' \subseteq X'$ be the set of variables in X' that occur in f', and let $T'' \subseteq X''$ be the set of variables in X'' that occur in f''. So, f' is in $\mathbb{C}[T']$ and f'' is in $\mathbb{C}[T'']$. Assume without loss of generality that $|T'| \geq |T''|$. Since $f' \cdot f'' \cdot g$ is multilinear, the sets T', T'' and T are pairwise disjoint. Consider two cases:

1. $|T''| \geq n/3$ (and hence $|T'| \geq n/3$). Thus, $f \cdot g = f' \cdot (f'' \cdot g)$ is a product polynomial (with the sets T' and $T'' \cup T$).

2. $|T''| < n/3$. Thus, $|T'' \cup T| < 2n/3$. Since f is a product polynomial, $|T'| \leq 2n/3$. So, let S'' be a subset of $X \setminus T'$ of size at least $n/3$ and at most $2n/3$, such that $T'' \cup T \subseteq S''$, and let $S' = X \setminus S''$. Thus, $f \cdot g = f' \cdot (f'' \cdot g)$ is a product polynomial (with the sets S' and S''). □

The following claim shows that the norm is multiplicative (in a certain case).

Claim 3.2. *Let f and g be two polynomials in $\mathbb{C}[X]$ such that $f \cdot g$ is multilinear. Then,*

$$\|f \cdot g\| = \|f\| \cdot \|g\|.$$

Proof. For a polynomial F and a monomial m, we denote by F_m the coefficient of m in F. Denote by A the set of variables that occur in f, and denote by B the set of variables that occur in g. Since $f \cdot g$ is multilinear, the sets A and B are disjoint. Furthermore,

$$\|f \cdot g\|^2 = \sum_{a,b} |[f \cdot g]_{a \cdot b}|^2 = \sum_{a,b} |f_a \cdot g_b|^2 = \left(\sum_a |f_a|^2\right)\left(\sum_b |g_b|^2\right) = \|f\|^2 \cdot \|g\|^2,$$

where the sums are over all multilinear monomials a in the variables A, and all multilinear monomials b in the variables B. □

3.2 Monotone Arithmetic Circuits

In this section we prove Theorem 1.5 that gives a tight lower bound for the size of monotone arithmetic circuits. The proof of this theorem already gives a lot of the details needed to prove the lower bounds for the various other models we consider.

3.2.1 The Structure of Monotone Circuits

In this section we prove the following lemma about the structure of monotone syntactically multilinear circuits.

Lemma 3.3. *Let $n \geq 3$ be an integer. Let Φ be a monotone syntactically multilinear arithmetic circuit with $s \in \mathbb{N}$ edges over the field $\mathbb{R}$ and over the set of variables $X = \{x_1, \dots, x_n\}$. Then, there exist $s+1$ monotone product polynomials $g_1, \dots, g_{s+1} \in \mathbb{R}[X]$ such that*

$$\widehat{\Phi} = \sum_{i \in [s+1]} g_i$$

(the definition of a product polynomial is in Section 3.1).

Proof. The proof follows by induction on the number of edges in Φ.

Assume without loss of generality that Φ has a unique output gate v computing $\widehat{\Phi}$.

Induction Base: The gate v is an input gate.

Since $n \geq 3$, the polynomial $\widehat{\Phi}$ is a product polynomial. Thus, the lemma follows with $g_1 = \widehat{\Phi}$.

Induction Step: The gate v is not an input gate.

If $|X_v| \leq 2n/3$, then $\widehat{\Phi}$ is a product polynomial, and the lemma follows with $g_1 = \widehat{\Phi}$ (since $s \geq 0$).

Assume that $|X_v| > 2n/3$. Every gate u in Φ with children u_1 and u_2 admits $|X_u| \leq |X_{u_1}| + |X_{u_2}|$. Thus, there exists a gate u in Φ such that

$$n/3 \leq |X_u| \leq 2n/3$$

(u is the first gate that satisfies the above, going down in Φ from v, when each step is to the child with the maximal number of variables).

Let Ψ be the circuit Φ after substituting a new variable y instead of u. Since Φ is monotone and syntactically multilinear, there exists a monotone multilinear polynomial h_1 in the set of variables $X \setminus X_u$ such that

$$\widehat{\Psi} = h_1 \cdot y + h_2,$$

where h_2 is the polynomial computed by Ψ after substituting $y = 0$. By the definition of Ψ,

$$\widehat{\Phi} = h_1 \cdot \widehat{\Phi}_u + h_2.$$

Since $\widehat{\Phi}_u$ is monotone and since $n/3 \leq |X_u| \leq 2n/3$, the polynomial $h_1 \cdot \widehat{\Phi}_u$ is both monotone and a product polynomial.

Denote by Ψ_0 the circuit Ψ after substituting $y = 0$. The circuit Ψ_0 is a monotone syntactically multilinear circuit for h_2 and it has at most $s - 1$ edges. By induction, there are s monotone product polynomials $g_1, \ldots, g_s \in \mathbb{R}[X]$ such that

$$h_2 = \sum_{i \in [s]} g_i.$$

Thus, setting $g_{s+1} = h_1 \cdot \widehat{\Phi}_u$, the lemma follows. □

3.2.2 Proof of Theorem 1.5

For a monomial m in the variables X and a polynomial $h \in \mathbb{R}[X]$, we denote (in this section) by $h(m)$ the coefficient of m in h (this may be misleading, as h is also a function, but we do so for simplicity of notation.) Let f be the polynomial defined in Section 2.1, and let F be the polynomial defined as

$$F(m) = \frac{f(m) + 1}{2} \in \{0, 1\},$$

for every monomial m in the variables X. Let Φ be a monotone arithmetic circuit over the field $\mathbb{R}$ and over the set of variables X computing F. Since Φ is monotone and F multilinear, we can assume without loss of generality that Φ is also syntactically multilinear. By Lemma 3.3, since the in-degree of Φ is at most 2, there exist at most $s = 2|\Phi| + 1$ monotone product polynomials $g_1, \ldots, g_s \in \mathbb{R}[X]$ such that

$$F = \sum_{i \in [s]} g_i.$$

By the definition of F, since

$$\sum_m f(m) \geq 0,$$

where the sum is over all multilinear monomials in the variables X, we have (recall that $|f(m)| = 1$),

$$\langle F, f \rangle = \sum_m \frac{f(m) + 1}{2} f(m) = \sum_m \frac{1}{2} + \sum_m \frac{f(m)}{2} \geq 2^{n-1}.$$

Since the polynomials $g_1, \ldots, g_s$ are monotone, for every monomial m the following holds.

- If $f(m) = -1$ (which implies $F(m) = 0$), then $g_i(m) = 0$, for every $i \in [s]$.

- If $f(m) = 1$ (which implies $F(m) = 1$), then $0 \leq g_i(m) \leq 1$, for every $i \in [s]$.

Thus, for every $i \in [s]$, we have $\langle g_i, f\rangle \geq 0$ and $\|g_i\| \leq \|f\|$. Hence, since

$$\sum_{i\in[s]} \langle g_i, f\rangle = \langle F, f\rangle \geq 2^{n-1},$$

there exists $j \in [s]$ such that

$$\langle g_j, f\rangle \geq 2^{n-1}/s.$$

Since g_j is a product polynomial and since $\|g_j\| \leq \|f\|$, using Theorem 1.3,

$$\langle g_j, f\rangle \leq 2^{-\Omega(n)}\|g_j\|\|f\| \leq 2^{-\Omega(n)}\|f\|^2 = 2^{-\Omega(n)}2^n.$$

So, since $s \leq 2|\Phi| + 1$,

$$|\Phi| = 2^{\Omega(n)},$$

and the theorem follows. □

3.3 Sum Trees

In this section we define and study sum trees. We first show that every syntactically multilinear arithmetic formula can be thought of as a sum tree with certain properties. We then show that sum trees do not increase the correlation with a given polynomial during their computation. This will enable us to bound the correlation between the polynomials computed by non-cancelling or noise-resistant syntactically multilinear arithmetic formulas and a certain family of polynomials. In the next section we will use this bound on the correlation to prove lower bounds for non-cancelling and noise-resistant arithmetic formulas.

A *sum tree* Ψ over the field $\mathbb{C}$ and over the set of variables $X = \{x_1, \ldots, x_n\}$ is a directed binary tree (whose edges are directed from the leaves to the root) as follows: Every leaf in Ψ is labelled by a polynomial in $\mathbb{C}[X]$. All vertices of in-degree 2 in Ψ are labelled by $+$.

The notation and definitions of sum trees are the same as of arithmetic formulas. We will now give a few examples. Every gate v in a sum tree computes a polynomial $\widehat{\Psi}_v$ in $\mathbb{C}[X_v]$ (where leaves compute the polynomials they are labelled by). A sum tree Ψ is τ-non-cancelling if every sum gate v with two children v_1 and v_2 in it (these are all the inner gates of Ψ) admits

$$\|\Psi_v\| \geq \tau \cdot \max(\|\Psi_{v_1}\|, \|\Psi_{v_2}\|).$$

The set of noisy values of a sum tree $\mathrm{N}_\varepsilon(\Psi_v)$ is defined the same as for formulas. We note that in the case of sum tree an input gate u computes an arbitrary polynomials $\widehat{\Psi}_u$, and so the set of noisy values of u is composed of a single element which is the map from $\{1,-1\}^{X_u}$ to $\mathbb{C}$ defined by $\widehat{\Psi}_u$ (see Remark 1.2).

3.3.1 Multilinear Arithmetic Formulas as Sum Trees

We now show that every syntactically multilinear arithmetic formula can be transformed to a sum tree in which the input gates are labelled by product polynomials (for the definition of a product polynomial see Section 3.1). We note that for every polynomial, there is a sum tree Ψ of size 1 computing it. However, the input gate of Ψ is not (necessarily) labelled by a product polynomial.

Theorem 3.4. *Let $n \geq 3$ be an integer, and let $\tau, \varepsilon > 0$. Let Φ be a τ-non-cancelling syntactically multilinear arithmetic formula over the field $\mathbb{C}$ and over the set of variables $X = \{x_1, \ldots, x_n\}$. Then, there exists a τ-non-cancelling sum tree Ψ of size at most $|\Phi|$ and of depth at most the depth of Φ over the field $\mathbb{C}$ and over the set of variables X computing $\widehat{\Phi}$ such that every input gate in Ψ is labelled by a product polynomial. Furthermore,*

$$\mathrm{N}_\varepsilon(\Psi) \subseteq \mathrm{N}_\varepsilon(\Phi).$$

Proof. We will in fact prove the following claim. Let v be a gate in Φ. Then, there exists a τ-non-cancelling sum tree Ψ_v of size at most $|\Phi_v|$ and of depth at most $\mathrm{depth}(v)$ over the field $\mathbb{C}$ and over the set of variables X_v computing $\widehat{\Phi}_v$ such that every input gate in Ψ_v is labelled by a product polynomial. Furthermore,

$$\mathrm{N}_\varepsilon(\Psi_v) \subseteq \mathrm{N}_\varepsilon(\Phi_v).$$

The proof will follow by induction on the size of Φ_v. Consider the following four cases:

Case one: v is an input gate. Set Ψ_v to be an input gate labelled $\widehat{\Phi}_v$. So, Ψ_v is a sum tree of size 1 and of depth 0 over the set of variables X_v computing $\widehat{\Phi}_v$ such that (since $n \geq 3$) the input gate of Ψ_v is labelled by a product polynomial. Since Ψ_v has no sum gates, it is τ-non-cancelling. Furthermore, since there is no noise in input gates, $\mathrm{N}_\varepsilon(\Psi_v) \subseteq \mathrm{N}_\varepsilon(\Phi_v)$.

Case two: v is a sum gate with children v_1 and v_2. By induction, there exist two sum trees Ψ_{v_1} and Ψ_{v_2} with the above properties. Set $\Psi_v = \Psi_{v_1} + \Psi_{v_2}$. By induction,

$$\widehat{\Psi}_v = \widehat{\Psi}_{v_1} + \widehat{\Psi}_{v_2} = \widehat{\Phi}_{v_1} + \widehat{\Phi}_{v_2} = \widehat{\Phi}_v.$$

Furthermore, since Φ is τ-non-cancelling,

$$\|\widehat{\Psi}_v\| \geq \tau \cdot \max(\|\widehat{\Psi}_{v_1}\|, \|\widehat{\Psi}_{v_2}\|).$$

So, by induction, Ψ_v is a τ-non-cancelling sum tree of size at most $|\Phi_v|$ and of depth at most depth(v) over the set of variables X_v computing $\widehat{\Phi}_v$ such that the input gates of Ψ_v are labelled by product polynomials. Furthermore, let $\psi_v \in \mathrm{N}_\varepsilon(\Psi_v)$. Thus, there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ that admit $0 \leq \alpha_1 \leq \varepsilon$ and $0 \leq \alpha_2 \leq \varepsilon$ such that

$$\psi_v = (1+\alpha_1)\cdot\psi_{v_1} + (1+\alpha_2)\cdot\psi_{v_2},$$

where $\psi_{v_1} \in \mathrm{N}_\varepsilon(\Psi_{v_1})$ and $\psi_{v_2} \in \mathrm{N}_\varepsilon(\Psi_{v_2})$. By induction, $\psi_{v_1} \in \mathrm{N}_\varepsilon(\Phi_{v_1})$ and $\psi_{v_2} \in \mathrm{N}_\varepsilon(\Phi_{v_2})$, and so $\psi_v \in \mathrm{N}_\varepsilon(\Phi_v)$. Thus, $\mathrm{N}_\varepsilon(\Psi_v) \subseteq \mathrm{N}_\varepsilon(\Phi_v)$.

Case three: v is a product gate with children v_1 and v_2 such that the sets X_{v_1} and X_{v_2} are of size at least $n/3$ each. Since Φ is syntactically multilinear, $X_{v_1} \cap X_{v_2} = \emptyset$. So, the polynomial $\widehat{\Phi}_v = \widehat{\Phi}_{v_1} \cdot \widehat{\Phi}_{v_2}$ is a product polynomial. Set Ψ_v to be an input gate labelled by $\widehat{\Phi}_v$. So, Ψ_v is a sum tree of size 1 and of depth 0 over the set of variables X_v computing $\widehat{\Phi}_v$ such that the input gate of Ψ_v is labelled by a product polynomial. Since Ψ_v has no sum gates, it is τ-non-cancelling. Furthermore, since there is no noise in input gates, $\mathrm{N}_\varepsilon(\Psi_v) \subseteq \mathrm{N}_\varepsilon(\Phi_v)$.

Case four: v is a product gate with two children v_1 and v_2 such that (without loss of generality) $|X_{v_2}| < n/3$. By induction, there exists a sum tree $\Psi' = \Psi_{v_1}$ satisfying the above properties with respect to v_1. Recall that for a gate u in Ψ', we defined $\widehat{\Psi}'_u$ to be the polynomial in $\mathbb{C}[X_{v_1}]$ that u computes in Ψ'. Set $\Psi = \Psi_v$ (we denote Ψ_v by Ψ, for simplicity of notation) to be the same as Ψ', except that each input gate u in Ψ' is labelled in Ψ by

$$\widehat{\Psi}'_u \cdot \widehat{\Phi}_{v_2}.$$

There is a one-to-one correspondence between gates in Ψ' and gates in Ψ. We think of a gate u both as a gate in Ψ' and as a gate in Ψ. It follows by induction (on the structure of Ψ) that each gate u admits

$$\widehat{\Psi}_u = \widehat{\Psi}'_u \cdot \widehat{\Phi}_{v_2}.$$

So, if u_1 and u_2 are the children of u, using Claim 3.2, since $X_{v_1} \cap X_{v_2} = \emptyset$, and since Ψ' is τ-non-cancelling,

$$\|\widehat{\Psi}_u\| = \|\widehat{\Psi}'_u\| \cdot \|\widehat{\Phi}_{v_2}\| \geq \tau \cdot \max(\|\widehat{\Psi}_{u_1}\|, \|\widehat{\Psi}_{u_2}\|).$$

So, Ψ is τ-non-cancelling. By induction, $\widehat{\Psi}' = \widehat{\Phi}_{v_1}$, which implies $\widehat{\Psi} = \widehat{\Phi}_v$. For every input gate u in Ψ, since $\widehat{\Psi}'_u$ is a product polynomial in $\mathbb{C}[X_{v_1}]$, since $X_{v_1} \cap X_{v_2} = \emptyset$, and since $|X_{v_2}| < n/3$, using Claim 3.1, it follows that $\widehat{\Psi}_u = \widehat{\Psi}'_u \cdot \widehat{\Phi}_{v_2}$ is a product polynomial. So, Ψ is a sum tree of size at most $|\Phi_v|$ and of depth at most depth(v) over the set of variables X_v computing $\widehat{\Phi}_v$ such that the input gates of Ψ are labelled by product polynomials.

Furthermore, let $\psi \in \mathrm{N}_\varepsilon(\Psi)$, and let $\phi_{v_2} \in \mathrm{N}_\varepsilon(\Phi_{v_2})$ be the map defined by $\widehat{\Phi}_{v_2}$. It follows by induction (on the structure of Ψ) that there exists $\psi' \in \mathrm{N}_\varepsilon(\Psi')$ such that

$$\psi = \psi' \cdot \phi_{v_2}.$$

By induction, $\psi' \in \mathrm{N}_\varepsilon(\Phi_{v_1})$, and so $\psi \in \mathrm{N}_\varepsilon(\Phi_v)$. Thus, $\mathrm{N}_\varepsilon(\Psi) \subseteq \mathrm{N}_\varepsilon(\Phi_v)$. □

3.3.2 Sum Trees Do Not Increase Correlation

In the previous section we showed that without loss of generality every syntactically multilinear arithmetic formula is a sum tree, whose input gates are labelled by product polynomials. We now bound the correlation between a polynomial computed by a sum tree and a given polynomial, using the correlations in the input gates. The intuition behind the theorem is that sum trees, as their name suggests, are just a sum of *input* polynomials. Now, if a given polynomial has small correlation with all the input polynomials, then a small sum tree should not increase the correlation by much. However, when the input polynomials have huge coefficients, the correlation upper bound is meaningless. Nevertheless, when the sum tree is either non-cancelling or noise-resistant, this does not happen (as the output polynomial has small coefficients).

Theorem 3.5. *Let $n \in \mathbb{N}$ be an integer, let $\tau > 0$ and let $0 < \varepsilon \leq 1$. Let Ψ be a τ-non-cancelling sum tree of depth d over the field $\mathbb{C}$ and over the set of variables $X = \{x_1, \ldots, x_n\}$. Let $\delta > 0$, and let f be a polynomial in $\mathbb{C}[X]$ such that for every input gate u in Ψ,*

$$cor(\widehat{\Psi}_u, f) \leq \delta \cdot \|\widehat{\Psi}_u\| \cdot \|f\|.$$

Then,

$$cor(\widehat{\Psi}, f) \leq \delta \cdot \|\widehat{\Psi}\| \cdot \|f\| \cdot |\Psi| \cdot \tau^{-d}.$$

Furthermore, let g be a map from $\{1, -1\}^n$ to $\mathbb{C}$ such that for every input gate u in Ψ,

$$cor(\psi_u, g) \leq \delta \cdot \|\psi_u\| \cdot \|g\|,$$

where $\psi_u : \{1, -1\}^n \to \mathbb{C}$ is the unique element of $\mathrm{N}_\varepsilon(\Psi_u)$ (recall that ψ_u is the map defined by the polynomial $\widehat{\Psi}_u$ – see Remark 1.2). Then, there exists $\psi \in \mathrm{N}_\varepsilon(\Psi)$ such that

$$cor(\psi, g) \leq \delta \cdot \|\psi\| \cdot \|g\| \cdot (\varepsilon/6)^{-d}.$$

Proof. The proof follows by induction on the size of Ψ. Let v be the root of Ψ, and consider the following two cases:

Case one: v is an input gate.
Since $|\Psi| = 1$ and since $d = 0$,

$$\mathrm{cor}(\widehat{\Psi}, f) \leq \delta \cdot \|\widehat{\Psi}\| \cdot \|f\| = \delta \cdot \|\widehat{\Psi}\| \cdot \|f\| \cdot |\Psi| \cdot \tau^{-d},$$

and

$$\mathrm{cor}(\psi, g) \leq \delta \cdot \|\psi\| \cdot \|g\| \cdot (\varepsilon/6)^{-d},$$

where $\psi \in \mathrm{N}_\varepsilon(\Psi)$.

Case two: v is a sum gate with children v_1 and v_2.
By induction,

$$\mathrm{cor}(\widehat{\Psi}_{v_1}, f) \leq \delta \cdot \|\widehat{\Psi}_{v_1}\| \cdot \|f\| \cdot |\Psi_{v_1}| \cdot \tau^{-d+1}$$

and

$$\mathrm{cor}(\widehat{\Psi}_{v_2}, f) \leq \delta \cdot \|\widehat{\Psi}_{v_2}\| \cdot \|f\| \cdot |\Psi_{v_2}| \cdot \tau^{-d+1}.$$

So,

$$\begin{aligned}\mathrm{cor}(\widehat{\Psi}, f) &= \mathrm{cor}(\widehat{\Psi}_{v_1} + \widehat{\Psi}_{v_2}, f) \leq \mathrm{cor}(\widehat{\Psi}_{v_1}, f) + \mathrm{cor}(\widehat{\Psi}_{v_2}, f) \\ &\leq \delta \cdot \max(\|\widehat{\Psi}_{v_1}\|, \|\widehat{\Psi}_{v_2}\|) \cdot \|f\| \cdot (|\Psi_{v_1}| + |\Psi_{v_2}|) \cdot \tau^{-d+1}.\end{aligned}$$

Since Ψ is τ-non-cancelling,

$$\max(\|\widehat{\Psi}_{v_1}\|, \|\widehat{\Psi}_{v_2}\|) \leq \tau^{-1} \|\widehat{\Psi}\|.$$

So, since $|\Psi_{v_1}| + |\Psi_{v_2}| \leq |\Psi|$,

$$\mathrm{cor}(\widehat{\Psi}, f) \leq \delta \cdot \|f\| \cdot \|\widehat{\Psi}\| \cdot |\Psi| \cdot \tau^{-d}.$$

Similarly, there exist $\psi_{v_1} \in \mathrm{N}_\varepsilon(\Psi_{v_1})$ and $\psi_{v_2} \in \mathrm{N}_\varepsilon(\Psi_{v_2})$ such that

$$\mathrm{cor}(\psi_{v_1}, g) \leq \delta \cdot \|\psi_{v_1}\| \cdot \|g\| \cdot (\varepsilon/6)^{-d+1}$$

and

$$\mathrm{cor}(\psi_{v_2}, g) \leq \delta \cdot \|\psi_{v_2}\| \cdot \|g\| \cdot (\varepsilon/6)^{-d+1}.$$

Assume without loss of generality that $\|\psi_{v_1}\| \geq \|\psi_{v_2}\|$. There are two possibilities:

1.
$$\|\psi_{v_1} + \psi_{v_2}\| \geq \varepsilon/2 \cdot \|\psi_{v_1}\|.$$

Then, we set $\psi = \psi_{v_1} + \psi_{v_2}$, and so $\psi \in \mathrm{N}_\varepsilon(\Psi)$. Thus,

$$\|\psi\| \geq \varepsilon/2 \cdot \|\psi_{v_1}\|.$$

2.
$$\|\psi_{v_1} + \psi_{v_2}\| < \varepsilon/2 \cdot \|\psi_{v_1}\|.$$

Then, we set $\psi = (1+\varepsilon)\psi_{v_1} + \psi_{v_2}$, and so $\psi \in \mathrm{N}_\varepsilon(\Psi)$. Thus,

$$\|\psi\| \geq \varepsilon \cdot \|\psi_{v_1}\| - \|\psi_{v_1} + \psi_{v_2}\| > \varepsilon/2 \cdot \|\psi_{v_1}\|.$$

So, since $\varepsilon \leq 1$,

$$\operatorname{cor}(\psi, g) \leq (1+\varepsilon) \cdot \operatorname{cor}(\psi_{v_1}, g) + \operatorname{cor}(\psi_{v_2}, g) \leq 3\delta \cdot \|\psi_{v_1}\| \cdot \|g\| \cdot (\varepsilon/6)^{-d+1} \leq \delta \cdot \|\psi\| \cdot \|g\| \cdot (\varepsilon/6)^{-d}.$$

□

3.4 Lower Bounds for Multilinear Formulas

In this section we prove the two lower bounds for non-cancelling and for noise-resistant syntactically multilinear arithmetic formulas.

Proof of Theorem 1.6. By Theorem 3.4, there exists a τ-non-cancelling sum tree Ψ of size at most $|\Phi|$ and of depth at most d over the field $\mathbb{C}$ and over the set of variables X computing $\widehat{\Phi}$ such that every input gate in Ψ is labelled by a product multilinear polynomial. So, by Theorem 1.3, every input gate u in Ψ admits

$$\operatorname{cor}(\widehat{\Psi}_u, f) \leq 2^{-\Omega(n)} \cdot \|\widehat{\Psi}_u\| \cdot \|f\|.$$

So, by Theorem 3.5, since $\widehat{\Psi} = \widehat{\Phi}$,

$$c \cdot \|\widehat{\Psi}\| \cdot \|f\| \leq \operatorname{cor}(\widehat{\Psi}, f) \leq 2^{-\Omega(n)} \cdot \|\widehat{\Psi}\| \cdot \|f\| \cdot |\Psi| \cdot \tau^{-d}.$$

Thus, since $|\Psi| \leq |\Phi|$,

$$|\Phi| \cdot \tau^{-d} \geq c \cdot 2^{\Omega(n)}.$$

Furthermore, since $|\Phi| \leq 2^d$, setting $c = 1/2$ and assuming $\tau < 2$,

$$d = \Omega\left(\frac{n}{\log(2/\tau)}\right).$$

□

Proof of Theorem 1.8. By Theorem 3.4, there exists a sum tree Ψ of size at most $|\Phi|$ and of depth at most d over the field $\mathbb{C}$ and over the set of variables X such that every input gate in Ψ is labelled by a product polynomial, and such that $\mathrm{N}_\varepsilon(\Psi) \subseteq \mathrm{N}_\varepsilon(\Phi)$.

Let u be an input gate in Ψ, and let ψ_u be the unique element of $\mathrm{N}_\varepsilon(\Psi_u)$ (recall that ψ_u is the map defined by the polynomial $\widehat{\Psi}_u$ – see Remark 1.2). Since $\widehat{\Psi}_u$ is a product polynomial, ψ_u is the vector of coefficients of a product polynomial (different than $\widehat{\Psi}_u$). So, by Theorem 1.3, and by the definition of g,

$$\mathrm{cor}(\psi_u, g) \leq 2^{-\Omega(n)} \cdot \|\psi_u\| \cdot \|g\|.$$

So, since Φ is ε-noise-resistant to computing g, and by Theorem 3.5, there exists $\psi \in \mathrm{N}_\varepsilon(\Psi)$ such that

$$\varepsilon \cdot \|\psi\| \cdot \|g\| \leq \mathrm{cor}(\psi, g) \leq 2^{-\Omega(n)} \cdot \|\psi\| \cdot \|g\| \cdot (\varepsilon/6)^{-d}.$$

So,

$$d = \Omega\left(\frac{n}{\log(2/\varepsilon)}\right).$$

□

4 Mixed-2-Source Extractors

In this section we construct a mixed-2-source extractor.

4.1 The Extractor

Let $n = 12sp$ be an integer, where $p \in \mathbb{N}$ is prime and $s \in \mathbb{N}$ is the constant given in Theorem 2.1. Let β_0 be the constant in the $\Omega(\cdot)$ in Corollary 1.4 and set

$$\beta = \beta_0/8$$

(also assume that $\beta \leq 1/8$). Let

$$m = \lfloor \beta \cdot n \rfloor \quad \text{and} \quad k = n - 3m.$$

Recall that m is the length of the output of the extractor and that k is the min-entropy requirement.

We think of $\{0,1\}^p$ as the field $\mathbb{F}$ of size 2^p (see Section 2.1). For $t \in \{0,1\}^n$ and $i \in [12s]$, define $y_i = y_i(t) \in \mathbb{F}$ as

$$\forall\ j \in [p] \quad (y_i)_j = t_{p(i-1)+j}.$$

Define the map F from $\{0,1\}^n$ to $\mathbb{F}$ by

$$F(t) = F(y_1, \ldots, y_{12s}) = y_1 \cdot y_2 \cdots y_{12s}.$$

Roughly, F cuts the input into $12s$ blocks of equal size, and multiplies these blocks as field elements. The extractor $\textsc{Ext} : \{0,1\}^n \to \{0,1\}^m$ is defined as the m most significant bits of $F(\cdot)$. That is,

$$\textsc{Ext}(t) = (F_1(t), \ldots, F_m(t)),$$

where $F_i(\cdot)$ is the i'th coordinate of $F(\cdot)$, for every $i \in [m]$. Note that $\textsc{Ext}(\cdot)$ can be computed in deterministic polynomial time. Also note that m and k are as required by Theorem 1.9.

4.2 Proof of Theorem 1.9

The proof of the theorem follows by an argument known as Vazirani's XOR lemma.

Let μ_1 and μ_2 be two independent distributions on $\{0,1\}^{n/2}$ (recall that n is even) such that

$$\mathrm{H}_\infty(\mu_1) = k_1 \quad \mathrm{H}_\infty(\mu_2) = k_2 \quad \text{and} \quad k_1 + k_2 \geq k.$$

Assume without loss of generality that μ_1 is a uniform distribution on a set $A_1 \subseteq \{0,1\}^{n/2}$, that μ_2 is a uniform distribution on a set $A_2 \subseteq \{0,1\}^{n/2}$, and that

$$|A_1| \cdot |A_2| \geq (2^{k_1} - 1)(2^{k_2} - 1) \geq 2^{k-1},$$

where the last inequality follows since both k_1 and k_2 are at most $n/2$ and since $6m + 4 \leq n$ (μ_1 and μ_2 can be written as a convex combination of such distributions - see Remark 4.1 below).

Remark 4.1. *The set of distributions with min-entropy k' form a convex body. Thus, every distribution with min-entropy k' can be written as a convex combination of the extreme points of this body. In addition, if $2^{k'}$ is an integer, then the extreme points of this body are exactly the distributions that are uniform on a set of size $2^{k'}$.*

Let $t_1 \sim \mu_1$ and let $t_2 \sim \mu_2$. Thus, t_1 is a uniform element of A_1 and t_2 is a uniform element of A_2. Let π be a one-to-one map from $[n]$ to $[n]$, and denote $t = (t_1 \circ t_2)_\pi$ (recall that the i'th entry in $(t_1 \circ t_2)_\pi$ is $(t_1 \circ t_2)_{\pi(i)}$). Thus, t is the input for the extractor.

Denote by W the random variable $\textsc{Ext}(t)$. To prove Theorem 1.9 we need to show that W is close to uniform; i.e.,

$$\|W - \mathrm{U}_m\|_1 \leq 2^{-2m}$$

(W means the distribution on $\{0,1\}^m$ defined by W). The proof has three main steps. The first step is to show that every XOR of the bits of W is almost uniform. The second step is to use Parseval's equality and conclude that the distance in 2-norm of W from uniform is small. The third step is to use Cauchy-Schwarz inequality to conclude that the statistical distance of W from uniform is small.

4.2.1 Every XOR of the Bits of W Is Almost Uniform

We will denote by W_S the XOR of all the entries of W that are in S. Formally, for $S \subseteq [m]$, denote

$$F_S = \bigoplus_{i \in S} F_i,$$

and denote

$$W_S = F_S(t),$$

where $t = (t_1 \circ t_2)_\pi$, $t_1 \sim \mu_1$ and $t_2 \sim \mu_2$.

In this section we will prove that for every nonempty $S \subseteq [m]$,

$$\|W_S - \mathrm{U}_1\|_1 \leq 2^{-3m} \tag{4.1}$$

(W_S means the distribution on $\{0,1\}$ defined by W_S). The proof will follow using the small maximal-partition discrepancy of f (see Section 1.1.2 for definitions).

The map π defines a partition of $[n]$ to two sets $\pi^{-1}(\{1, \ldots, n/2\})$ and $\pi^{-1}(\{n/2+1, \ldots, n\})$. This partition defines a $2^{n/2} \times 2^{n/2}$ matrix M whose (r_1, r_2) entry is $F_S((r_1 \circ r_2)_\pi)$, where $r_1, r_2 \in \{0,1\}^{n/2}$.

Recall that $f(\cdot)$ is defined as $\psi(F(\cdot))$, for an arbitrary non-trivial character ψ, and note that $(-1)^{F_S(\cdot)} = \psi(F(\cdot))$, where $\psi(\cdot)$ is a non-trivial character of $\mathbb{F}$. Corollary 1.4 tells us that the maximal-partition discrepancy of F_S is at most $2^{-\beta_0 n}$, which implies that

$$\textsc{Disc}(M) \leq 2^{-\beta_0 n}.$$

The sets A_1 and A_2 define a rectangle R in M. The random variable W_S is a uniform element of R. Thus,

$$\|W_S - \mathrm{U}_1\|_1 = \frac{2^n}{|A_1||A_2|}\textsc{Disc}_R(M) \leq 2^{n-(k-1)-\beta_0 n} \leq 2^{-3m},$$

as claimed (where the last inequality follows since $6m + 1 \leq \beta_0 n$).

4.2.2 Distance of Ext from U_m in 2-Norm is Small

By Parseval's equality and by (4.1),

$$\sum_{g\in\{0,1\}^m} (\Pr[W = g] - \mathrm{U}_m(g))^2 = 2^{-m} \sum_{S\subseteq[m]:S\neq\emptyset} (\|W_S - \mathrm{U}_1\|_1)^2 \leq 2^{-6m} \tag{4.2}$$

(the following remark gives additional details, for completeness).

Remark 4.2. *We recall some definitions regarding Fourier transform. We think of* $\mathrm{G} \overset{\text{def}}{=} \{0,1\}^m$ *as an abelian group (with addition of vectors over* GF(2)*). For every* $S \subseteq [m]$*, the map* ψ_S *from* G *to* $\mathbb{C}$ *defined as*

$$\forall\ g = (g_1, \ldots, g_m) \in \mathrm{G} \quad \psi_S(g) = (-1)^{\sum_{i\in S} g_i}$$

is a character of G*. The set of characters of* G*,* $\{\psi_S\}_{S\subseteq[m]}$*, form an orthonormal basis for the vector space of maps from* G *to* $\mathbb{C}$ *with respect to the inner product*

$$\langle \chi, \chi' \rangle = 2^{-m} \sum_{g\in\mathrm{G}} \chi(g) \cdot \overline{\chi'(g)},$$

where χ *and* χ' *are maps from* G *to* $\mathbb{C}$*. Thus, every map* $\chi : \mathrm{G} \to \mathbb{C}$ *can be written as*

$$\chi = \sum_{S\subseteq[m]} \widehat{\chi}(S) \cdot \psi_S,$$

where

$$\widehat{\chi}(S) = \langle \chi, \psi_S \rangle$$

(the map $\widehat{\chi}(\cdot)$ is called the Fourier transform of χ), and we have Parseval's equality:

$$\sum_{g \in \mathrm{G}} |\chi(g)|^2 = 2^m \sum_{S \subseteq [m]} |\widehat{\chi}(S)|^2.$$

Denote by $\mathrm{U} \overset{\mathrm{def}}{=} \mathrm{U}_m$ the uniform distribution on G. Since $\mathrm{U} = 2^{-m} \cdot \psi_\emptyset$, for every $S \subseteq [m]$,

$$\widehat{\mathrm{U}}(S) = \begin{cases} 0 & S \neq \emptyset \\ 2^{-m} & S = \emptyset. \end{cases}$$

By Parseval's equality,

$$\sum_{g \in \mathrm{G}} (Pr[W = g] - \mathrm{U}(g))^2 = \sum_{g \in \mathrm{G}} ([\mathrm{P} - \mathrm{U}](g))^2 = 2^m \sum_{S \subseteq [m]} (\widehat{[\mathrm{P} - \mathrm{U}]}(S))^2,$$

where $\mathrm{P}(g) = Pr[W = g]$. Note that

$$\widehat{\mathrm{P}}(\emptyset) = 2^{-m}$$

and that

$$\widehat{\mathrm{P}}(S) = 2^{-m}\,\mathbb{E}\left[(-1)^{W_S}\right] = 2^{-m}\|W_S - \mathrm{U}_1\|_1,$$

for every non-empty $S \subseteq [m]$. Thus, by linearity of Fourier transform,

$$\sum_{g \in \mathrm{G}} (Pr[W = g] - \mathrm{U}(g))^2 = 2^m \sum_{S \subseteq [m] : S \neq \emptyset} (\widehat{\mathrm{P}}(S))^2 = 2^{-m} \sum_{S \subseteq [m] : S \neq \emptyset} (\|W_S - \mathrm{U}_1\|_1)^2.$$

4.2.3 Completing the Proof

By Cauchy-Schwarz inequality, using (4.2),

$$\Big(\sum_{g \in \{0,1\}^m} \big| \Pr[W = g] - \mathrm{U}_m(g) \big| \Big)^2 \leq 2^m \sum_{g \in \{0,1\}^m} (\Pr[W = g] - \mathrm{U}_m(g))^2 \leq 2^{-5m}.$$

Thus,

$$\|W - \mathrm{U}_m\|_1 \leq 2^{-2m},$$

which completes the proof. □

Acknowledgement

We wish to thank Scott Aaronson and Avi Wigderson for helpful conversations.

References

[A] S. Aaronson. Multilinear Formulas and Skepticism of Quantum Computing. *STOC 2004*: 118-127.

[BIW] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting Randomness Using Few Independent Sources. In Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, pages 384393, 2004.

[BKSSW] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers, and Extractors. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pages 110, 2005.

[BRSW] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson. 2 Source Dispersers for no(1) Entropy and Ramsey Graphs beating the Frankl-Wilson Construction. In Proceedings of the 38th Annual ACM Symposium on Theory of Computing, 2006.

[Bo] J. Bourgain. On the Construction of Affine Extractors. *Manuscript*, 2005.

[Bo^+] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. International Journal of Number Theory, 1:132, 2005.

[BoGK] J. Bourgain, A. A. Glibichuk and S. V. Konyagin. Estimates for the Number of Sums and Products and for Exponential Sums in Fields of Prime Order. *Journal of London Mathematical Society*, 2: 380-398, 2006.

[BKT] J. Bourgain, N. Katz, and T. Tao. A Sum-Product Estimate in Finite Fields, and Appli- cations. Geometric and Functional Analysis, 14:2757, 2004.

[CG] B. Chor and O. Goldreich. Unbiased Bits from Sources of Weak Randomness and Proba- bilistic Communication Complexity. SIAM Journal on Computing, 17(2):230-261, 1988.

[H] T. P. Hayes. Separating the k-party communication hierarchy: an application of the Zarankiewicz problem. Manuscript, 2001.

[JSn] M. Jerrum and M. Snir. Some Exact Complexity Results for Straight-Line Computations over Semirings. *Journal of the ACM*, Volume 29, Issue 3: 874 - 897, 1982.

[J] S. Jukna. On the P versus NP intersected with co-NP question in communication complexity. Information Processing Letters 96(6):202-206, 2005.

[KN] E. Kushilevitz and N. Nisan. Communication complexity. *Cambridge University Press*, 1997.

[LR] T. W. Lam and W. L. Ruzzo. Results on communication complexity classes. J. Computer Systems Sciences, 44: 324342, 1992.

[LS] T. Lee and A. Shraibman. Disjointness Is Hard in the Multi-party Number-on-the-Forehead Model. Proceedings of the 23rd CCC, pp. 81–91, 2008.

[NW] N. Nisan and A. Wigderson. Lower Bounds on Arithmetic Circuits via Partial Derivatives. *Computational Complexity*, 6: 217-234, 1996 (preliminary version in Proceeding of the 36th FOCS 1995).

[R] A. Rao. Extractors for a Constant Number of Polynomially SmallMin-entropy Independent Sources. In Proceedings of the 38th Annual ACM Symposium on Theory of Computing, 2006.

[R02] R.Raz. On the Complexity of Matrix Product. *SIAM Journal of Computing* 32(5) (2003), pp. 1356-1369 (also in the Proceeding of the 34th STOC, 2002, pp. 144-151).

[R04a] R. Raz. Multi-Linear Formulas for Permanent and Determinant are of Super-Polynomial Size. Proceeding of the 36th STOC: 633-641, 2004.

[R04b] R. Raz. Separation of Multilinear Circuit and Formula Size. Theory Of Computing Vol. 2, article 6, 2006, and *Proceeding of the 45th FOCS*: 344-351, 2004 (title: "Multilinear-$NC_1 \neq$ Multilinear-NC_2").

[R05] R. Raz. Extractors with Weak Random Seeds. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pages 1120, 2005.

[RSY] R. Raz, A. Shpilka and A. Yehudayoff. A Lower Bound for the Size of Syntactically Multilinear Arithmetic Circuits. *ECCC Report* TR06-060.

[RY] R. Raz and A. Yehudayoff. Balancing Syntactically Multilinear Arithmetic Circuits. *Manuscript*, 2007.

[Sc] C. P. Schnorr. A Lower Bound on the Number of Additions in Monotone Computations. Theor. Comput. Sci., 2, pages 305–315, 1976.

[SV] R. Sengupta and H. Venkateswaran. A Lower Bound for Monotone Arithmetic Circuits Computing 0-1 Permanent. Theor. Comput. Sci., 209 (1-2), pages 389–398, 1998.

[ShSn] E. Shamir and M. Snir. On the Depth Complexity of Formulas. *Journal Theory of Computing Systems*, Volume 13, Number 1: 301-322, 1979.

[Sn] M. Snir. On the Size Complexity of Monotone Formulas. Tech Rep CSR-46-79, Univ. of Edinburgh, Edinburgh, Scotland, 1979.

[TT] P. Tiwari and M. Tompa. A Direct Version of Shamir and Snir's Lower Bounds on Monotone Circuit Depth. Inf. Process. Lett., 49, 5, pages 243–248, 1994.

[V] L. G. Valinat. Negation can be exponentially powerful. *Proceedings of the eleventh annual ACM symposium on theory of computing.* : 189 - 196, 1979.

[VSBR] L. G. Valiant, S. Skyum, S. Berkowitz, C. Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. SIAM J. Comput. 12(4): 641-644, 1983.

[W] A. Wigderson. *Private communication.*

[Y] A. C. Yao. Some complexity questions related to distributive computing. In Proceedings of the eleventh annual ACM symposium on Theory of computing, pages: 209 - 213, 1979.