Electronic Colloquium on Computational Complexity, Report No. 99 (2014)

The Complexity of DNF of Parities

Gil Cohen* Igor Shinkar**

August 7, 2014

Abstract

We study depth 3 circuits of the form $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$, or equivalently – DNF of parities. This model was first explicitly studied by Jukna (CPC'06) who obtained a $2^{\Omega(n)}$ lower bound for explicit functions in this model. Several related models have gained attention in the last few years, such as parity decision trees, the parity kill number and $\mathsf{AC}^0 \circ \mathsf{XOR}$ circuits.

For a function $f\colon \{0,1\}^n \to \{0,1\}$, we denote by $\mathsf{DNF}_\oplus(f)$ the least integer s for which there exists an $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$ circuit, with OR gate of fan-in s, that computes f. We summarize some of our results:

- For any affine disperser $f\colon \{0,1\}^n \to \{0,1\}$ for dimension k, it holds that $\mathsf{DNF}_\oplus(f) \geq 2^{n-2k}$. By plugging Shaltiel's affine disperser (FOCS'11) we obtain an explicit $2^{n-n^{o(1)}}$ lower bound.
- We give a non-trivial general upper bound by showing that $\mathsf{DNF}_\oplus(f) \leq O(2^n/n)$ for any function f on n bits. This bound is shown to be tight up to an $O(\log n)$ factor.
- We show that for any symmetric function f it holds that $\mathsf{DNF}_\oplus(f) \leq 1.5^n \cdot \mathrm{poly}(n)$. Furthermore, there exists a symmetric function f for which this bound is tight up to a polynomial factor.
- For threshold functions we show tighter bounds. For example, we show that the majority function has $\mathsf{DNF}_\oplus$ complexity of $2^{n/2} \cdot \mathrm{poly}(n)$. This is also tight up to a polynomial factor.
- For the inner product function IP on n inputs we show that $\mathsf{DNF}_\oplus(\mathrm{IP}) = 2^{n/2} - 1$. Previously, Jukna gave a lower bound of $\Omega(2^{n/4})$ for the $\mathsf{DNF}_\oplus$ complexity of this function. We further give bounds for any low degree polynomial.
- Finally, we obtain a $2^{n-o(n)}$ average case lower bound for the parity decision tree model using affine extractors.

*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. gil.cohen@weizmann.ac.il. Supported by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

**Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. igor.shinkar@weizmann.ac.il. Research supported by ERC grant number 239985.

ISSN 1433-8092

Contents

1 Introduction

In this paper we study depth 3 circuits of the form $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$, where all gates have unbounded fan-in. Note that such a circuit computes a DNF applied to linear combinations of the input variables. Thus, for a function $f\colon \{0,1\}^n \to \{0,1\}$ we denote by $\mathsf{DNF}_\oplus(f)$ the minimum top gate fan-in over all circuits of the above form that compute f. Why not define $\mathsf{DNF}_\oplus(f)$ as the minimum number of gates required by an $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$ circuit for computing f? There are three answers to this question, which also shed more light on this model of computation.

1. There is an equivalent, yet more combinatorial meaning, to the $\mathsf{DNF}_\oplus$ complexity of a function f the way it is defined above; $\mathsf{DNF}_\oplus(f)$ is the least number of affine subspaces required to cover exactly $f^{-1}(1)$. This is because every input to the top gate is an $\mathsf{AND} \circ \mathsf{XOR}$ circuit, and such a circuit computes the indicator of an affine subspace (we allow the use of constants).

2. Although potentially the fan-in of the AND and XOR gates can be arbitrary large, one can in fact assume it is bounded by n. Indeed, since $\mathsf{AND} \circ \mathsf{XOR}$ circuit computes the indicator function of some affine subspace, one can always replace the circuit with an equivalent circuit where the fan-in of the AND and XOR gates is at most n. Thus, the minimum number of gates required by an $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$ circuit for computing a given function f is bounded above by $\mathsf{DNF}_\oplus(f) \cdot n^2$. In this paper we are mainly interested in functions with exponentially large $\mathsf{DNF}_\oplus(f)$ complexity and we do not mind such polynomial factors.

3. The size of a DNF is defined as the number of terms it contains, which is the top gate fan-in when represented as an $\mathsf{OR} \circ \mathsf{AND}$ circuit. So the current definition is analogous to the definition of the DNF complexity of a function.

To the best of our knowledge, the $\mathsf{DNF}_\oplus$ complexity of a function was first explicitly considered by Jukna [Juk06] (see also Chapter 11 of [Juk12]). Jukna applies graph theoretic arguments and gives $2^{\Omega(n)}$ lower bounds on the $\mathsf{DNF}_\oplus$ complexity for several explicit and natural functions. For example, for all even n, a lower bound of $\Omega(2^{n/4})$ is given for the $\mathsf{DNF}_\oplus$ complexity of the inner product function $\mathrm{IP}(x) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n$, where addition is over $\mathbb{F}_2$. [1] A similar result was obtained by Grolmusz [Gro94] based on communication complexity arguments. In [Juk06] it is also shown that the disjointness function $\mathrm{disj}\colon \{0,1\}^n \to \{0,1\}$ defined by $\mathrm{disj}(x) = 1$ if and only if $x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n = 0$, where addition is over $\mathbb{N}$, has $\mathsf{DNF}_\oplus$ complexity of $\Omega(2^{0.016n})$.

More generally, Jukna characterizes the $\mathsf{DNF}_\oplus$ complexity of functions that represent bipartite graphs in a certain way. The downside of this technique is that it cannot yield lower bounds stronger than $2^{n/2}$, whereas one can show that most functions on n inputs have $\mathsf{DNF}_\oplus$ complexity $\Omega(2^n/(n \cdot \log n))$, as we discuss later in the introduction.

[1] Jukna's lower bound holds even if one replace the top OR gate by any threshold gate.

Several related models have been considered in the literature. For example, the parity decision tree model, defined by Kushilevitz and Mansour [KM93] in the context of learning theory, has received a significant attention in the last few years [MO09, ZS10, TWXZ13, OST+13, STV14], with motivation coming mainly from communication complexity. We elaborate on the relation between the $\mathsf{DNF}_\oplus$ model and parity decision tree model in Section 7, and give new results for it. Another example would be a recent work of Akavia *et al.* [ABG+14], who considered $\mathsf{AC}^0 \circ \mathsf{XOR}$ circuits, which strictly generalizes the $\mathsf{DNF}_\oplus$ model. Their motivation comes from cryptography. A work of O'Donnell *et al.* [OST+13] is related to the *width* of $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{XOR}$ circuits, whereas the $\mathsf{DNF}_\oplus$ complexity is about understanding the *size* of such circuits. We also mention the work of Grolmusz [Gro94] who studied depth 3 circuits where the top gate is a threshold gate, the middle layer contains AND gates, and the bottom layer is composed of MOD_m gates, for some integer m.

1.1 Our Results

In this paper we further study the $\mathsf{DNF}_\oplus$ model. We also obtain results for the parity decision tree model. In the remaining of the introduction we elaborate on our contributions.

Almost Optimal Lower Bounds via Affine Dispersers

The first result of this paper states that good affine dispersers have a very high $\mathsf{DNF}_\oplus$ complexity. An *affine disperser* for dimension k is a function $f\colon \{0,1\}^n \to \{0,1\}$ with the following property: For every affine subspace $U \subseteq \{0,1\}^n$ of dimension k, f restricted to U is not constant. Using a standard probabilistic argument, one can show the existence of affine dispersers for dimension $k = \log_2(n) + \log_2\log_2(n) + O(1)$. In terms of explicit constructions, the state of the art affine disperser is due to Shaltiel [Sha11]. Shaltiel's disperser works for dimension as low as $k = 2^{\log^{0.9} n}$, which although not logarithmic is still sub-polynomial.

Affine dispersers can be thought of as a "linear analogue" to Ramsey graphs, and are very natural pseudorandom objects which gained some attention by researchers in the past few years (e.g., [BKS+05, BSK12, Sha11, Li11, CT14]). Nevertheless, the only application of them we are aware of is a lower bound of $3n - O(k)$ by Demenkov and Kulikov [DK11] for circuits over the full basis. By plugging an affine disperser for sub-linear dimension this gives $(3 - o(1)) \cdot n$ lower bound, matching and simplifying a result of Blum [Blu83], and is still the state of the art lower bound for this model. Here we give another application of affine dispersers, as captured by the following lemma.

Theorem 1.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be an affine disperser for dimension k. Then,*

1. $\mathsf{DNF}_\oplus(f) \geq 2^{n-2k}$.
2. $\max\left(\mathsf{DNF}_\oplus(f), \mathsf{DNF}_\oplus(1-f)\right) \geq 2^{n-k}$, *where $1-f$ is the negation of f.*

A clarification regarding Theorem 1.1 is in order. The first item of the theorem tells us that if f is an affine disperser for dimension k then $\mathsf{DNF}_\oplus(f) \geq 2^{n-2k}$. This by itself is

already enough to yield almost optimal explicit lower bounds for the $\mathsf{DNF}_\oplus$ model. Indeed, since Shaltiel's disperser is an affine disperser for dimension $k = n^{o(1)}$ one obtains a $2^{n-n^{o(1)}}$ lower bound (we remark that although Shaltiel's disperser is explicit in the computational sense, its description is not at all simple! We believe it is interesting to find an explicit function in the computational sense, that also has a simple description. We discuss this in Section 2).

Nevertheless, it is not clear whether the factor of 2 in the exponent of 2^{n-2k} is necessary. Moreover, as we exemplify next, in some cases this factor of 2 is highly undesired. So, although Theorem 1.1 does not guarantee a lower bound of 2^{n-k} for any affine disperser, the second item of the theorem (which has a "one line proof") does guarantee such a lower bound for either f or its negation. We note that if f is explicit then so is its negation, and as a result, we have this amusing scenario where any explicit affine disperser for dimension k yields an explicit lower bound of 2^{n-k}, though we do not necessarily know whether this lower bound comes from f or its negation. [2]

One application of item 2 (in which item 1 is meaningless) is in proving that $\mathsf{DNF}_\oplus(\mathrm{IP}) \geq \Omega(2^{n/2})$, which improves the lower bound of $\Omega(2^{n/4})$ obtained by Jukna. Indeed, it is a well-known fact that IP is an affine disperser for dimension $n/2+1$ (see Appendix A), and so the second item of Theorem 1.1 implies that either IP or its negation have $\mathsf{DNF}_\oplus$ complexity of $\Omega(2^{n/2})$. We further discuss the inner product later in the introduction (see also Section 6), where we show that the bound holds in fact for both functions. We also give a second proof for this fact.

Item 2 of Theorem 1.1 also implies the existence of a function f on n inputs such that $\mathsf{DNF}_\oplus(f) \geq \Omega(2^n/(n \cdot \log n))$. This can be seen by taking an affine disperser for dimension $k = \log_2(n) + \log_2 \log_2(n) + O(1)$, which is promised to exists by the probabilistic method. It is worth mentioning that by using a counting argument (which is the most common way of proving such lower bounds) one would get a weaker lower bound of $\Omega(2^n/n^2)$.

We remark that while a random function on n inputs have $\mathsf{DNF}_\oplus$ complexity $2^n/(n \cdot \log n)$, with high probability, Theorem 1.1 gives an explicit *property* of a random function that causes its $\mathsf{DNF}_\oplus$ complexity to be large.

An Upper Bound for All Functions

Clearly the $\mathsf{DNF}_\oplus$ complexity of any function f on n bits is bounded above by 2^n. Indeed, one can take the union of points in $f^{-1}(1)$ as these are affine subspaces of dimension 0. The next theorem gives an upper bound of $O(2^n/n)$ for the $\mathsf{DNF}_\oplus$ complexity of all functions on n bits.

Theorem 1.2. *For any function $f\colon \{0,1\}^n \to \{0,1\}$ it holds that $\mathsf{DNF}_\oplus(f) \leq O(2^n/n)$.*

The combinatorial meaning of Theorem 1.2 is that any set (namely, $f^{-1}(1)$) can always be covered by not $O(2^n/n)$ affine subspaces, regardless of its structure. By the lower bound

[2] In fact, if one knows toward which value f is biased, then one can tell which of f or $1-f$ obtains the 2^{n-k} lower bound. Still, it is not always clear if a given construction of an affine disperser is biased towards 1, and we prefer to give a statement that could be used in a "black-box" fashion.

mentioned above, it follows that this upper bound is tight up to an $O(\log n)$ factor. Our proof for the upper bound makes use of the Gowers norm, and does not seem to be related to (or to follow from) the classical $O(2^n/n)$ upper bound of Lupanov [Lup58] for fan-in 2 circuits over Boolean circuits.

The $\mathsf{DNF}_\oplus$ Complexity of Symmetric and Threshold Functions

We continue to study the $\mathsf{DNF}_\oplus$ complexity of natural classes of functions. Our next result gives a non-trivial upper bound for any symmetric function.

Theorem 1.3. *For any symmetric function $f\colon \{0,1\}^n \to \{0,1\}$ it holds that*

$$\mathsf{DNF}_\oplus(f) \leq 1.5^n \cdot \mathrm{poly}(n).$$

Moreover, there exists a symmetric function $g\colon \{0,1\}^n \to \{0,1\}$ such that $\mathsf{DNF}_\oplus(g) \geq \Omega(1.5^n/\sqrt{n})$.

Theorem 1.3 states that any symmetric function has $\mathsf{DNF}_\oplus$ complexity at most $1.5^n \cdot \mathrm{poly}(n)$, and this is tight for the class of symmetric functions. One may still ask whether a better bound can be obtained for the natural subclass of threshold functions. Consider for example the majority function $\mathrm{Maj}\colon \{0,1\}^n \to \{0,1\}$, where n is an odd positive integer. Is the upper bound of $1.5^n \cdot \mathrm{poly}(n)$ tight for Maj? It is not hard to show that $\mathsf{DNF}_\oplus(\mathrm{Maj}) \geq \Omega(2^{n/2}) \geq \Omega(1.414^n)$. To see this we use the fact an affine subspaces of dimension d cannot be contained in Hamming balls of radius $d-1$ (see Fact 4.2). Since $\mathrm{Maj}^{-1}(1)$ is the Hamming ball of radius $(n+1)/2$ centered at the all ones vector, any affine subspace that participates in the covering of $\mathrm{Maj}^{-1}(1)$ must have dimension at most $(n+1)/2$. Thus, $\mathsf{DNF}_\oplus(\mathrm{Maj}) \geq 2^{n-1}/2^{(n+1)/2} = \Omega(2^{n/2})$.

Which of these bounds, if any, is the right one? Our next theorem states that it is the lower bound that is tight. In fact, it gives a tight bound (up to polynomial factors) for the $\mathsf{DNF}_\oplus$ complexity of any threshold function, where the threshold is at least $1/2$. In the theorem below, $\mathsf{Th}_\tau\colon \{0,1\}^n \to \{0,1\}$ is the function such that $\mathsf{Th}_\tau(x) = 1$ if and only if $|x| \geq \tau n$, where $\tau \in [0,1]$ is an integer multiple of $1/n$, and H is the Shannon binary entropy function defined by $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$.

Theorem 1.4. *For any integer n and any $1/2 \leq \tau \leq 1$ that is an integer multiple of $1/n$*

$$2^{(H(\tau)-(1-\tau))\cdot n} \cdot \mathrm{poly}(n^{-1}) \leq \mathsf{DNF}_\oplus(\mathsf{Th}_\tau) \leq 2^{(H(\tau)-(1-\tau))\cdot n} \cdot \mathrm{poly}(n).$$

By plugging $\tau = 1/2$ in Theorem 1.4 we obtain that, up to $\mathrm{poly}(n)$ factors, the $\mathsf{DNF}_\oplus$ complexity of Maj on n inputs is $2^{n/2}$. It is interesting to compare this result with Quine [Qui53] classical result stating that the DNF complexity of Maj is $\binom{n}{n/2} = \Theta(2^n/\sqrt{n})$.

Theorem 1.4 only applies for $\tau \geq 1/2$, and there is a reason for that. Indeed, using arguments similar to the lower bound for $\mathsf{DNF}_\oplus(\mathrm{Maj})$ sketched above, one can show that for any $\tau < 1/2$ it holds that $\mathsf{DNF}_\oplus(\mathsf{Th}_\tau) \geq 2^{\tau n}$. Roughly speaking, this asymmetry of the $\mathsf{DNF}_\oplus$ complexity between threshold functions with $\tau \geq 1/2$ and $\tau < 1/2$ follows from the

fact that the set of functions with low $\mathsf{DNF}_\oplus$ complexity is not closed under negation. For that one should also consider the $\mathsf{CNF}_\oplus$ complexity, defined in the natural way. A similar phenomenon occurs in the standard DNF and CNF complexity measures. We elaborate on that in Section 7.

While the lower bound in Theorem 1.3 holds for all symmetric functions, the construction we give that matches this lower bound only uses parity gates with fan-in 2 (and the upper bound in Theorem 1.4 requires only fan-in 3). Note that one can replace each fan-in 2 parity gate with a width 2 constant size CNF. By collapsing levels one then obtains a depth 3 Boolean circuit of size $1.5^n \cdot \mathrm{poly}(n)$, with fan-in 2 bottom layer gates, that computes the given symmetric function. The latter is a result that is attributed to Paturi *et al.* [PSZ97], and Theorem 1.3 reproduce it.

The Inner Product Function and Low Degree Polynomials

As mentioned above, the second item in Theorem 1.1 implies a lower bound of $\Omega(2^{n/2})$ for $\mathsf{DNF}_\oplus(\mathrm{IP})$ (or for $\mathsf{DNF}_\oplus(1-\mathrm{IP})$). We would like to get a more precise bound. It is easy to see that $\mathsf{DNF}_\oplus(\mathrm{IP}) \leq 2^{n/2}-1$. Indeed, when fixing the variables $\{x_i : i \text{ odd}\}$ to zeros we are getting the constant zero function, while for any other fix of the variables $\{x_i : i \text{ odd}\}$ we get a linear function in the $\{x_i : i \text{ even}\}$. Therefore, we can write IP as

$$\mathrm{IP}(x_1,\ldots,x_n) = \bigvee_{(\alpha_1,\alpha_3,\ldots,\alpha_{n-1})\neq\vec{0}} \left(\left(\bigwedge_{i \text{ odd}} x_i = \alpha_i\right) \wedge \left(\bigoplus_{i \text{ odd}} \alpha_i x_{i+1} = 1\right)\right).$$

In the following theorem we show that this is best possible.

Theorem 1.5. *For any even integer n, it holds that $\mathsf{DNF}_\oplus(\mathrm{IP}) = 2^{n/2}-1$.*

We prove Theorem 1.5 in Section 6. So one completely understands the $\mathsf{DNF}_\oplus$ complexity of IP. In fact, one can prove something more general. Namely, for any degree 2 polynomial over $\mathbb{F}_2$ we have $\mathsf{DNF}_\oplus(f) = \Theta(1/\mathrm{bias}(f))$ (see Corollary 6.3). So the $\mathsf{DNF}_\oplus$ complexity of degree 2 polynomials is also well understood. What about higher degrees? In the following theorem we give a non-trivial upper bound for the $\mathsf{DNF}_\oplus$ complexity of degree $d > 2$ polynomials. The bound is meaningful for functions with degree up to roughly $\log\log n$.

Theorem 1.6. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function that has degree $d \geq 3$ as a polynomial over $\mathbb{F}_2$. Then,*

$$\mathsf{DNF}_\oplus(f) \leq 2^{n-\Omega(n^{1/(d-1)!})}.$$

On the other hand, with high probability, a random degree d polynomial over $\mathbb{F}_2$ has $\mathsf{DNF}_\oplus$ complexity $2^{n-O(n^{1/(d-1)})}$.

We also prove stronger upper bounds for biased low degree polynomials (see Section 6 for more details).

The Parity Decision Tree Model

As mentioned, Kushilevitz and Mansour [KM93] introduced the notion of parity decision trees, which received a significant attention in the last few years [MO09, ZS10, TWXZ13, OST+13, STV14]. Roughly speaking, these are decision trees where each node contains not a variable but rather a linear function of some subset of the variables (see, e.g., [STV14] for the formal definition). The parity decision tree complexity of a function $f\colon \{0,1\}^n \to \{0,1\}$, which we denote here by $\mathsf{DT}_\oplus(f)$, is the least size (that is, number of leaves) required by a parity decision tree for computing f.

In this paper we give an average case hardness result for the $\mathsf{DT}_\oplus$ model. Namely, we give an explicit construction for a function f such that any function with not too large $\mathsf{DT}_\oplus$ complexity has a small correlation with f. For this we need to recall the definition of an affine extractor, which is a strengthening of affine dispersers.

An *affine extractor* for dimension k with bias ε is a function $f\colon \{0,1\}^n \to \{0,1\}$ with the following property. For every affine subspace $U \subseteq \mathbb{F}_2^n$ of dimension k, the bias of f restricted to U, defined as $\mathrm{bias}(f|_U) = |\mathbb{E}_{u\sim U}[(-1)^{f(u)}]|$, is at most ε. A standard probabilistic argument shows the existence of an affine extractor with bias ε for dimension $k = \log_2(n/\varepsilon^2) + \log_2\log_2(n/\varepsilon^2) + O(1)$. The state of the art explicit constructions for affine extractors [Bou07, Yeh11, Li11] works for dimension $k = O(n/\sqrt{\log\log n})$, with bias that is exponentially small in n.

In the theorem below, the distance between two functions, denoted by $\mathrm{dist}(f,g)$, is defined as the fraction of points in the hypercube on which the functions disagree.

Theorem 1.7. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be an affine extractor for dimension k, with bias $\varepsilon \leq 1/2$. Then, for any $g\colon \{0,1\}^n \to \{0,1\}$ such that $\mathsf{DT}_\oplus(g) \leq \varepsilon \cdot 2^{n-k}$ it holds that $\mathrm{dist}(f,g) \geq 1/2 - 4\varepsilon$.*

By plugging the efficiently constructible affine extractors mentioned above, one obtains an explicit average case lower bound of $2^{n-o(n)}$ for the $\mathsf{DT}_\oplus$ model. As in the case of lower bounds for the $\mathsf{DNF}_\oplus$ model, it is also interesting to obtain explicit lower bounds that have a succinct and simple description. One example comes from the inner product function. Indeed, the fact that IP is an affine extractor for dimension $n/2 + c$, with error exponentially small in c, is considered a folklore (see Appendix A). Thus, Theorem 1.7 implies that IP has no more than ε correlation with functions having $\mathsf{DT}_\oplus$ complexity $O(\varepsilon^2 \cdot 2^{n/2})$. To break the "$n/2$ barrier", one can consider the function $\mathsf{Tr}(x^7)$ [3], which clearly has a simple description. Ben-Sasson and Kopparty [BSK12] proved that this function is an affine extractor for dimension $2n/5 + O(\log^2(1/\varepsilon))$, when n is odd. [4] Thus for, say, a constant ε, one obtains a simple and explicit average case lower bound of $\Omega(2^{0.6n})$ for this model.

Like in the case of affine dispersers, although affine extractors are natural objects, to the best of our knowledge so far they only found two applications in the literature. One was

[3] Tr is the trace function from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$. We also assume some underlying isomorphism between the vector space $\mathbb{F}_2^n$ and the field $\mathbb{F}_{2^n}$.

[4] To get this dependency in ε, one needs to use a result by Haramaty and Shpilka [HS10]. See also Theorem 6.2 in [CT14].

given by Ben-Sasson and Zewi [BSZ11], who gave a construction of a two-source extractor based on affine extractors. Conditioned on the Polynomial Freiman-Ruzsa conjecture from additive combinatorics, this two-source extractor is shown to support min-entropy lower than the state of the art extractor by Bourgain [Bou05]. The second application was shown by Viola [Vio14] for the purpose of constructing extractors for circuit sources.

A proof for Theorem 1.7 is given in Section 7, where we also relate the $\mathsf{DNF}_\oplus$ and $\mathsf{DT}_\oplus$ complexity measures. We summary here by saying that interesting results regarding the DNF and the decision tree complexity "goes through" in the analogous models with parity gates with hardly any change in the proofs. Still, we feel it is worthwhile to point out these relations between $\mathsf{DNF}_\oplus$ and $\mathsf{DT}_\oplus$.

1.2 Preliminaries

Throughout the paper, for readability sake, we suppress floor and ceiling. For integers n, k such that $0 \leq k \leq n$, we denote by L_k the k'th level of the n dimensional hypercube, namely, $L_k = \{x \in \{0,1\}^n \colon |x| = k\}$, where $|x|$ is the Hamming weight of x. Note that n is suppressed in this notation. This will not cause a confusion because n will always be clear from the context. The bias of a Boolean function $f \colon D \to \{0,1\}$, defined on some domain D, is given by $\mathrm{bias}(f) \triangleq |D|^{-1} \cdot |\sum_{x \in D} (-1)^{f(x)}|$.

We make some use of basic results from Fourier analysis of Boolean functions. We follow the standard notations set in O'Donnell's book [O'D14].

2 Almost Optimal Lower Bounds via Affine Dispersers

We start this section by proving Theorem 1.1. To this end we prove the following lemma.

Lemma 2.1. *Let $A \subset \{0,1\}^n$ be a set of size $t < 2^n$. Then, for any integer $\ell \geq 0$ such that*

$$(t+1) \cdot 2^{\ell-1} < 2^n, \tag{2.1}$$

there exists an affine subspace $V_\ell \subset \{0,1\}^n$ such that $A \cap V_\ell = \emptyset$.

Proof. We construct the affine subspaces $(V_\ell)_\ell$ by induction on ℓ. We start with the base case $\ell = 0$. As we assume that $|A| = t < 2^n$, there exists a point in $\{0,1\}^n \setminus A$, which is an affine subspace of dimension $\ell = 0$, as desired.

We now construct V_ℓ given $V_{\ell-1}$, assuming ℓ satisfies Equation 2.1. Let $\Delta_1, \ldots, \Delta_{\ell-1}$ be linearly independent vectors such that $V_{\ell-1} = \Delta_0 + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})$, for some shift vector Δ_0. We wish to find a vector Δ_ℓ, independent of $\Delta_1, \ldots, \Delta_{\ell-1}$, such that $\Delta_0 + \mathrm{Span}(\Delta_1, \ldots, \Delta_\ell)$ does not intersect A. This will be our V_ℓ. To this end, consider the set of "good shifts"

$$X = \{x \in \{0,1\}^n \mid (x + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})) \cap A = \emptyset\}.$$

We note that $\Delta_0 \in X$. Moreover, if x is another point in X then by setting $\Delta_\ell = x + \Delta_0$ we get that

$$\begin{aligned}\Delta_0 + \mathrm{Span}(\Delta_1, \ldots, \Delta_\ell) &= (\Delta_0 + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})) \cup (\Delta_0 + \Delta_\ell + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})) \\ &= (\Delta_0 + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})) \cup (x + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1}))\,,\end{aligned}$$

and since both $x, \Delta_0 \in X$, it follows that the set above does not intersect A. So, all that is left to prove is the existence of an $x \in X$ such that $\Delta_\ell = x + \Delta_0$ is linearly independent of $\Delta_1, \ldots, \Delta_{\ell-1}$, or equivalently, that $X \setminus V_{\ell-1} \neq \emptyset$. To this end, note that $x \in X$ if and only if $x \notin A + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})$. So,

$$\begin{aligned}|X| &= 2^n - |A + \mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})| \\ &\geq 2^n - |A| \cdot |\mathrm{Span}(\Delta_1, \ldots, \Delta_{\ell-1})| \\ &= 2^n - t \cdot 2^{\ell-1}.\end{aligned}$$

Thus,

$$|X \setminus V_{\ell-1}| \geq |X| - |V_{\ell-1}| \geq 2^n - (t+1) \cdot 2^{\ell-1} > 0,$$

where the last inequality follows by Equation (2.1). Thus, $X \setminus V_{\ell-1} \neq \emptyset$, which concludes the proof. □

We now turn to the proof of Theorem 1.1.

Proof of Theorem 1.1. We start with the proof of the first item. Let f be an affine disperser for dimension k and set $s = \mathsf{DNF}_\oplus(f)$. Thus, f is the union of s affine subspaces. Since f is an affine disperser for dimension k, each of these affine subspaces has dimension at most $k-1$. Thus, $|f^{-1}(1)| \leq s \cdot 2^{k-1}$. By Lemma 2.1, for every integer ℓ such that $(|f^{-1}(1)|+1) \cdot 2^{\ell-1} < 2^n$, there exists an affine subspace of dimension ℓ, restricted to which f is the constant 0. Since f is an affine disperser for dimension k, we get that

$$(s \cdot 2^{k-1} + 1) \cdot 2^{k-1} \geq (|f^{-1}(1)| + 1) \cdot 2^{k-1} \geq 2^n.$$

Thus, $s \geq 2^{n-2k}$ as desired.

We now turn to prove the second item, which is actually even simpler to prove. As before, if f is an affine disperser for dimension k with $\mathsf{DNF}_\oplus(f) = s$ then $|f^{-1}(1)| \leq s \cdot 2^{k-1}$. So $\mathsf{DNF}_\oplus(f) \geq |f^{-1}(1)| \cdot 2^{1-k}$. Now note that if f is an affine disperser for dimension k then so does $1 - f$, and so applying the argument above for $1 - f$ gives $\mathsf{DNF}_\oplus(1 - f) \geq |(1-f)^{-1}(1)| \cdot 2^{1-k}$. Clearly, $\max(|f^{-1}(1)|, |(1-f)^{-1}(1)|) \geq 2^{n-1}$ and so we get that $\max(\mathsf{DNF}_\oplus(f), \mathsf{DNF}_\oplus(1-f)) \geq 2^{n-k}$, as stated.

□

A discussion regarding explicitness.

As mentioned, by plugging Shaltiel's disperser to Theorem 1.1 it follows that there exists an efficiently computable function $f\colon \{0,1\}^n \to \{0,1\}$ such that $\mathsf{DNF}_\oplus(f) \geq 2^{n-n^{o(1)}}$. The

affine disperser of Shaltiel mentioned above, which we denote by Sha, is explicit in the computational sense. Namely, given $x \in \{0,1\}^n$, one can compute $\mathsf{Sha}(x)$ in time $\mathrm{poly}(|x|)$. However, the description of the function Sha is not at all simple. Thus, it is natural to ask for an explicit lower bound for the $\mathsf{DNF}_\oplus$ model, which also has a simple description. Ben-Sasson and Kopparty [BSK12] gave such a construction that works for dimension $\Omega(n^{4/5})$ (though we omit its description here as it would require setting some notations). Thus, there is a "simple" and explicit lower bound of $2^{n-\Omega(n^{4/5})}$ for the $\mathsf{DNF}_\oplus$ model.

We next show a very simple and explicit function that has $\mathsf{DNF}_\oplus$ complexity $\Omega(2^{2n/3})$ (just to break the "$n/2$ barrier"). In [BSK12] the authors showed that for an odd n, the function $\mathsf{Tr}(x^{15})$, which evidently has a very simple description, is an affine disperser for dimension $n/3 + 10$. By examining the proof of the second item in Theorem 1.1, one can see that for any affine disperser f for dimension k, $\mathsf{DNF}_\oplus(f) \geq |f^{-1}(1)| \cdot 2^{1-k}$. Now, $\mathsf{Tr}(x^{15})$ is a non-constant degree 4 polynomial over $\mathbb{F}_2$, and so it obtains the value 1 on at least 2^{-4} fraction of the hypercube. Thus, it follows that $\mathsf{DNF}_\oplus(\mathsf{Tr}(x^{15})) \geq \Omega(2^{2n/3})$.

3 An Upper Bound for All Functions

In this section we prove Theorem 1.2. To this end we need the following lemma.

Lemma 3.1. *Let $A \subseteq \{0,1\}^n$ be a set of size $\varepsilon \cdot 2^n$, for any $\varepsilon > 2^{-n/4}$. Then there exists an affine subspace $V \subseteq A$ of dimension $\dim(V) \geq \log(n) - \log\log(1/\varepsilon) - 2$.*

We defer the proof of Lemma 3.1 and first prove Theorem 1.2.

Proof of Theorem 1.2. In order to prove Theorem 1.2, fix a function $f : \{0,1\}^n \to \{0,1\}$, and let $A = f^{-1}(1)$. We want to show that there are $s = O(2^n/n)$ affine subspaces $V_1, \ldots, V_s$ contained in A that cover A, i.e. $A = \cup_{i=1}^{s} V_i$. We choose the affine subspaces greedily as follows:

1. Set $A_1 = A$.
2. Set $i = 1$.
3. Repeat
   (a) Pick an affine subspace $V_i \subseteq A_i$ of maximal dimension.
   (b) Set $A_{i+1} = A_i \setminus V_i$.
   (c) Increment i by 1.

   until $|A_i| \leq 2^n/n$.
4. Output $V_1, \ldots, V_{i-1}$ together with the singleton affine subspaces for each point in A_i.

We claim that the above procedure terminates in at most $O(2^n/n)$ iterations. For every integer $t \geq 0$ define $i_t = \min\{i : A_i \leq 2^{n-t}\}$. Note that $i_0 = 1$. By Lemma 3.1, for every $i < i_t$ we have $|V_i| \geq n/4t$. Therefore,

$$i_t - i_{t-1} \leq \frac{|A_{i_{t-1}}|}{\min_{i \leq i_t} |V_i|} \leq \frac{2^{n-t}}{n/4t}\,.$$

The algorithm terminates in the iteration $i = i_{\log n}$, which is at most

$$i_{\log n} = 1 + \sum_{t=1}^{\log n}(i_t - i_{t-1}) \leq 1 + \sum_{t=1}^{\log n} \frac{2^{n-t}}{n/4t} \leq 1 + \frac{4 \cdot 2^n}{n} \cdot \sum_{t=1}^{\infty} \frac{t}{2^t} = 1 + 8 \cdot \frac{2^n}{n}\,.$$

Therefore, the procedure above outputs at most $1 + 8 \cdot 2^n/n + |A_{i_{\log n}}| = 1 + 9 \cdot 2^n/n$ affine subspaces. This completes the proof of the theorem. □

We now return to the proof of Lemma 3.1. We need the following definition of *degree d norm* of a function, also known as the uniformity norm or Gowers norm of a function.

Definition 3.2 (Degree d norm)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function. The d'th norm of f is defined as*

$$U_d(f) = \left(\mathop{\mathbf{E}}_{x,y_1,\ldots,y_d \in \{0,1\}^n} \left[\prod_{S \subseteq [d]} f(x + \sum_{i \in S} y_i) \right] \right)^{1/2^d}\,.$$

The following proposition is standard and can be found, e.g., in [VW07].

Proposition 3.3 (Gowers-Cauchy-Schwartz inequality)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function. Then, $U_1(f) = |\mathbf{E}[f(x)]|$, and for any positive integer i it holds that $U_i(f) \leq U_{i+1}(f)$.*

We are now ready to prove Lemma 3.1.

Proof of Lemma 3.1. Let $f : \{0,1\}^n \to \{0,1\}$ be the indicator function of the set A. Then, for any $d \in \mathbb{N}$, by Proposition 3.3 we have $U_d(f) \geq \mathbf{E}[f(x)] = \varepsilon$. Therefore, since f is Boolean, by the definition of $U_d(f)$ we have

$$\mathop{\mathbf{Pr}}_{x,y_1,\ldots,y_d \in \{0,1\}^n}[x + \mathrm{Span}(y_1,\ldots,y_d) \subseteq A] = \mathop{\mathbf{E}}_{x,y_1,\ldots,y_d \in \{0,1\}^n} \left[\prod_{S \subseteq [d]} f(x + \sum_{i \in S} y_i) \right] \geq \varepsilon^{2^d}.$$

Note that uniformly random vectors $y_1,\ldots,y_d \in \{0,1\}^n$ are linearly independent with probability $(1-2^{-n})(1-2^{1-n})(1-2^{2-n})\cdots(1-2^{d-1-n}) > 1 - 2^{d-n}$. Therefore, for uniformly random $x, y_1,\ldots,y_d \in \{0,1\}^n$, with probability at least $\varepsilon^{2^d} - 2^{d-n}$, the affine subspace $x + \mathrm{Span}(y_1,\ldots,y_d)$ is contained in A, and its dimension is d. Thus, as long as $\varepsilon^{2^d} > 2^{d-n}$ the set A contains an affine subspace of dimension d. It is easy to check that this indeed holds for $d = \lceil \log(n) - \log\log(1/\varepsilon) - 2 \rceil$. □

4 The $\mathsf{DNF}_\oplus$ Complexity of Symmetric Functions

In this section we prove Theorem 1.3. To this end we will need the following two facts.

Fact 4.1. *Let $n > 1$ be an integer. Let $p \in (0,1)$ be such that pn is an integer. Let $q = 1-p$. Then,*

$$\frac{1}{\sqrt{8npq}} \leq \binom{n}{pn} \cdot 2^{-H(p)n} \leq \frac{1}{\sqrt{3npq}}\,.$$

In particular,

$$\frac{1}{\sqrt{8n}} \leq \binom{n}{pn} \cdot 2^{-H(p)n} \leq 1\,.$$

A proof for Fact 4.1 can be found in, e.g., [CT12].

Fact 4.2. *Let $u+U$ be an affine subspace of dimension d. Then $u+U$ is not contained in any ball of radius $d-1$. In particular, there exist vectors $u_1, u_2 \in u+U$ such that $|u_1| \geq d$ and $|u_2| \leq n-d$.*

Proof of Fact 4.2. Consider a $d \times n$ generating matrix A for the subspace U. By performing a Gaussian elimination on A and permuting the columns of the resulted matrix, we can assume that the first d columns of A form the identity matrix. This can be done because such operations have no affect on Hamming weights nor on the dimension. Now let B be a ball of radius $d-1$ around the point u_0, that is, $B = \{x \in \{0,1\}^n : |x - u_0| \leq d-1\}$. Let v be the vector in U that disagrees with u_0 on the first d entries. Such a vector exists by the structure of A deduced above. Clearly $v \notin B$ since $|v - u_0| \geq d$. The proof follows. □

We start by showing that the "moreover direction" of Theorem 1.3 holds. Consider the function $g_k \colon \{0,1\}^n \to \{0,1\}$ defined by $g_k(x) = 1$ if and only if $|x| = k$, where $0 \leq k \leq n$ is an integer.

Claim 4.3. *For any integers n and $0 \leq k \leq n$ it holds that $\mathsf{DNF}_\oplus(g_k) \geq \binom{n}{k} \cdot 2^{-\min(k,n-k)}$.*

Proof of Claim 4.3. First note that for any k, $\mathsf{DNF}_\oplus(g_k) = \mathsf{DNF}_\oplus(g_{n-k})$ since any covering for g_k can be translated to a covering for g_{n-k} by adding the all ones vector to the shifts of all affine subspaces in the covering. It is therefore enough to show that for all $k \leq n/2$, $\mathsf{DNF}_\oplus(g_k) \geq \binom{n}{k} \cdot 2^{-k}$. Note that $\mathsf{DNF}_\oplus(g_k)$ is the minimum number of affine subspaces such that their union equals L_k. Since $|L_k| = \binom{n}{k}$ the proof follows by Fact 4.2 which guarantees that any affine subspace in this covering has dimension at most k. □

Consider now the function g_k for $k = 2n/3$. By Claim 4.3 and Fact 4.1 it follows that

$$\mathsf{DNF}_\oplus(g_k) \geq \frac{\binom{n}{2n/3}}{2^{n/3}} \geq \Omega\left(\frac{1}{\sqrt{n}} \cdot 2^{n\cdot(H(2/3)-1/3)}\right) = \Omega\left(\frac{1}{\sqrt{n}} \cdot 1.5^n\right),$$

which concludes the "moreover direction" of Theorem 1.3. To prove the more interesting direction, we show that Claim 4.3 is tight up to $\mathrm{poly}(n)$ factors.

Lemma 4.4. *For any integers n and $0 \leq k \leq n$ it holds that*

$$\mathsf{DNF}_{\oplus}(g_k) \leq \binom{n}{k} \cdot 2^{-\min(k,n-k)} \cdot n.$$

Proof of Lemma 4.4. We start by presenting the proof for $k = n/2$ (assuming that n is even). The proof for this special case is slightly simpler than the proof for general k, and already demonstrates the proof idea.

For every permutation $\sigma \in S_n$ we define the affine subspace $V_\sigma \subseteq \{0,1\}^n$ that contains all points subject to the following $n/2$ affine constraints:

$$\begin{cases} 1 &= x_{\sigma(1)} + x_{\sigma(2)} \\ 1 &= x_{\sigma(3)} + x_{\sigma(4)} \\ &\vdots \\ 1 &= x_{\sigma(n-1)} + x_{\sigma(n)} \,. \end{cases}$$

Note that for every $\sigma \in S_n$, it holds that $V_\sigma \subseteq L_{n/2}$. We show next that by choosing $m = n \cdot \binom{n}{n/2} \cdot 2^{-n/2}$ random permutations $\sigma_1, \ldots, \sigma_m$ uniformly and independently from S_n, with high probability $L_{n/2} = \cup_{i=1}^{m} V_{\sigma_i}$. To see this fix $x \in L_{n/2}$. By symmetry, for a uniformly random $\sigma \in S_n$ we have that

$$\mathop{\mathbf{Pr}}_{\sigma}[x \in V_\sigma] = \frac{|V_\sigma|}{|L_{n/2}|} = \frac{2^{n/2}}{\binom{n}{n/2}}.$$

Therefore,

$$\mathop{\mathbf{Pr}}_{\sigma_1,\ldots,\sigma_m}\left[x \notin \bigcup_{i=1}^{m} V_{\sigma_i}\right] \leq \left(1 - \frac{2^{n/2}}{\binom{n}{n/2}}\right)^m \leq e^{-n},$$

where the last inequality holds by the choice of m. Hence, by the union bound

$$\mathop{\mathbf{Pr}}_{\sigma_1,\ldots,\sigma_m}\left[\exists x \in L_{n/2} \text{ such that } x \notin \bigcup_{i=1}^{m} V_{\sigma_i}\right] \leq \binom{n}{n/2} \cdot e^{-n} < 1.$$

Therefore, there exist m affine subspaces, of the above form, such that their union equals $L_{n/2}$, and so $\mathsf{DNF}_{\oplus}(g_{n/2}) \leq n \cdot \binom{n}{n/2} \cdot 2^{-n/2}$.

We proceed now for general $k \leq n/2$. As mentioned, this would also conclude the proof for all $k \geq n/2$. Let $t = n - 2k$. For every permutation $\sigma \in S_n$ define the affine subspace V_σ as the set of points that obey the following affine constraints:

$$\begin{cases} \begin{cases} 0 &= x_{\sigma(1)} \\ &\vdots \\ 0 &= x_{\sigma(t)} \end{cases} \\ \begin{cases} 1 &= x_{\sigma(t+1)} + x_{\sigma(t+2)} \\ &\vdots \\ 1 &= x_{\sigma(n-1)} + x_{\sigma(n)} \,. \end{cases} \end{cases}$$

Clearly, $V_\sigma \subseteq L_k$ for every permutation $\sigma \in S_n$. Moreover, for every fixed $x \in L_k$ we have that $\mathbf{Pr}_\sigma[x \in V_\sigma] = |V_\sigma|/|L_k| = 2^k/\binom{n}{k}$. Therefore, if we choose $m = n \cdot \binom{n}{k} \cdot 2^{-k}$ random permutations $\sigma_1, \ldots, \sigma_m$ uniformly and independently at random then

$$\mathop{\mathbf{Pr}}_{\sigma_1,\ldots,\sigma_m}\left[x \notin \bigcup_{i=1}^{m} V_{\sigma_i}\right] \leq \left(1 - \frac{2^k}{\binom{n}{k}}\right)^m \leq e^{-n},$$

where the last inequality follows by our choice of m. By taking the union bound over all $x \in L_k$ we get

$$\mathop{\mathbf{Pr}}_{\sigma_1,\ldots,\sigma_m}\left[\exists x \in L_k \text{ such that } x \notin \bigcup_{i=1}^{m} V_{\sigma_i}\right] \leq \binom{n}{k} \cdot e^{-n} < 1.$$

Therefore, $\mathsf{DNF}_\oplus(g_k) \leq \binom{n}{k} \cdot 2^{-k} \cdot n$. □

We now deduce Theorem 1.3 from Lemma 4.4. Any symmetric function $f\colon \{0,1\}^n \to \{0,1\}$ can be written as the union of some subset of $\{g_0, \ldots, g_n\}$. Thus, $\mathsf{DNF}_\oplus(f) \leq \sum_{k=0}^{n} \mathsf{DNF}_\oplus(g_k)$. By Lemma 4.4 and Fact 4.1,

$$\mathsf{DNF}_\oplus(g_k) \leq \binom{n}{k} \cdot 2^{-\min(k,n-k)} \cdot n \leq 2^{H(k/n)\cdot n - \min(k,n-k)} \cdot n.$$

One can show that the maximum over $0 \leq k \leq n$ of the expression $H(k/n)\cdot n - \min(k, n-k)$ that appears in the exponent above is obtained at $k = n/3$ and $k = 2n/3$. This maximum value is $(H(1/3) - 1/3) \cdot n = \log_2(1.5) \cdot n$. Thus, for all $0 \leq k \leq n$, $\mathsf{DNF}_\oplus(g_k) \leq 1.5^n \cdot n$ and so $\mathsf{DNF}_\oplus(f) \leq O(1.5^n \cdot n^2)$. We remark that by a more careful argument one can show that $\mathsf{DNF}_\oplus(f) \leq O(1.5^n \cdot n)$.

5 The $\mathsf{DNF}_\oplus$ Complexity of Threshold Functions

In this section we prove Theorem 1.4.

Proof. As in the proof of Theorem 1.3, it is enough to show how to cover any level ωn for $\tau \leq \omega \leq 1$ in the hypercube by $2^{(H(\tau)-(1-\tau))\cdot n} \cdot \mathrm{poly}(n)$ affine subspaces. However, as apposed to the way this was done in the proof of Theorem 1.3, we may now consider affine subspaces that are not restricted to level ωn, and points may "leak" to higher levels. This is the leverage we exploit so to obtain stronger results for threshold functions.

The proof is more delicate than the proof of Theorem 1.3, and for the purpose of covering $L_{\omega n}$, for different $\tau \leq \omega \leq 1$, we need to consider affine subspaces with different structure. Moreover, we first handle levels ωn such that $\tau \leq \omega \leq (3\tau + 1)/4$. We then show how to handle the higher levels.

Given $1/2 \leq \tau \leq 1$ and $\tau \leq \omega \leq (3\tau+1)/4$, define [5]

$$\begin{aligned}\alpha &= 2\omega - 1\\ \beta &= 3\tau - 4\omega + 1\\ \gamma &= 2\omega - 2\tau.\end{aligned}$$

Note that by our choice of τ, ω it holds that $0 \leq \alpha, \beta, \gamma \leq 1$. For a permutation $\sigma \in S_n$ let V_σ be the affine subspace of $\{0,1\}^n$ that is defined by the following set of affine constraints:

$$\left\{\begin{array}{l}\left\{\begin{array}{rcl}1 &=& x_{\sigma(1)}\\ &\vdots&\\ 1 &=& x_{\sigma(\alpha n)}\end{array}\right.\\ \left\{\begin{array}{rcl}1 &=& x_{\sigma(\alpha n+1)} + x_{\sigma(\alpha n+2)}\\ &\vdots&\\ 1 &=& x_{\sigma((\alpha+2\beta)n-1)} + x_{\sigma((\alpha+2\beta)n)}\end{array}\right.\\ \left\{\begin{array}{rcl}1 &=& x_{\sigma((\alpha+2\beta)n+1)} + x_{\sigma((\alpha+2\beta)n+2)} + x_{\sigma((\alpha+2\beta)n+3)}\\ &\vdots&\\ 1 &=& x_{\sigma(n-2)} + x_{\sigma(n-1)} + x_{\sigma(n)}.\end{array}\right.\end{array}\right.$$

Namely, we have αn constraints on 1 variable, βn constraints on two variables and γn constraints on three variables, where each variable appears in exactly one constraint as σ is a permutation. Note that $\alpha + 2\beta + 3\gamma = 1$, so this set of constraints acts on all entries of x, and is well-defined in the sense that it does not operate on invalid entries of x.

We note that if x satisfies all the constraints above then $|x| \geq \tau n$. Indeed, each entry of x appears in exactly one constraint, the number of constraints is exactly $\alpha + \beta + \gamma = \tau$ and each satisfied constraint implies that at least one of the entries of x that participate in the constraint is 1. Thus, for any permutation σ, the affine subspace V_σ is contained in $\mathsf{Th}_\tau^{-1}(1)$, and so taking union of V_σ for several permutations σ would never cover undesired points.

We now compute the number of points in level ωn covered by V_σ. Let $x \in V_\sigma \cap L_{\omega n}$. Since $x \in V_\sigma$, the first αn constraints contribute αn to the Hamming weight of x. The following βn constraints contribute βn to the Hamming weight of x (as exactly one of the variables in such a constraint on two variables is 1). Since $|x| = \omega n$ and since each of the remaining γn constraints must have either 1 or 3 variables set to 1, it holds that exactly δn of these constraints have one variable with value 1 and the rest have all 3 variables set to 1, where

$$\delta = \frac{\alpha + \beta + 3\gamma - \omega}{2} = \frac{3\omega - 3\tau}{2}\,.$$

[5] For ease of notation, we treat variables such as $\tau, \omega, \alpha, \beta$ and γ as if they were some real numbers in $[0,1]$ and ignore the issue of rounding to integer multiplication of $1/n$. This does not affect any of the results.

Moreover, there is a one to one mapping between the points in $V_\sigma \cap L_{\omega n}$ and the number of ways to choose which of the two variables in each of the βn constraints will have value 1, which δn out of the γn constraints will have exactly one variable set to 1 and which variable out of the three would that be. Hence,

$$|V_\sigma \cap L_{\omega n}| = 2^{\beta n} \cdot \binom{\gamma n}{\delta n} \cdot 3^{\delta n} \geq \frac{1}{\sqrt{8n}} \cdot 2^{(\beta+\gamma\cdot H(\delta/\gamma)+\delta\cdot\log_2 3)\cdot n} = \frac{1}{8\sqrt{n}} \cdot 2^{(1-\tau)\cdot n},$$

where the inequality follows by Fact 4.1, and the last equality follows by our choice of α, β and γ. Let

$$m = \sqrt{8} \cdot n^{1.5} \cdot 2^{(H(\omega)-(1-\tau))\cdot n} \leq \sqrt{8} \cdot n^{1.5} \cdot 2^{(H(\tau)-(1-\tau))\cdot n}.$$

As in the proof of Theorem 1.3, it follows that if $\sigma_1, \ldots, \sigma_m$ are permutations sampled uniformly and independently at random, then for any $x \in L_{\omega n}$

$$\mathop{\mathbf{Pr}}_{\sigma_1,\ldots,\sigma_m}\left[x \notin \bigcup_{i=1}^{m} V_{\sigma_i}\right] \leq \left(1 - \frac{|V_{\sigma_1} \cap L_{\omega n}|}{|L_{\omega n}|}\right)^m \leq e^{-n}.$$

Hence, by the union bound over all $x \in L_{\omega n}$, there exist m permutations $\sigma_1, \ldots, \sigma_m$ such that $V_{\sigma_1}, \ldots V_{\sigma_m}$ are all contained in $\mathsf{Th}_\tau^{-1}(1)$ and which their union covers $L_{\omega n}$. We conclude that the union of all levels $L_{\omega n}$ for $\tau \leq \omega \leq (3\tau+1)/4$ can be covered by $O(n^{2.5} \cdot 2^{(H(\tau)-(1-\tau))\cdot n})$ affine subspaces, all contained in $\mathsf{Th}_\tau^{-1}(1)$.

We note that the constraint $\omega \leq (3\tau+1)/4$ is necessary for our set of affine constraints to be well-defined. Indeed, for $\omega > (3\tau+1)/4$, we have $\beta < 0$ which makes no sense since βn represents the number of constraints on two variables. The values of α, β and γ were obtained by some optimization which was spared from the reader. The fact that this optimization process resulted in taking a negative β for high values of ω suggests that we should do as close to it as we can, and indeed to handle the higher levels $\omega > (3\tau+1)/4$ we take $\beta = 0$, namely sets of constraints on only one or three variables.

Let $\sigma \in S_n$ be a permutation. Consider a system of affine constraints were αn constraints are on one variable, and the rest, say γn of them, are on three variables. Namely,

$$\begin{cases} \begin{cases} 1 &= x_{\sigma(1)} \\ &\vdots \\ 1 &= x_{\sigma(\alpha n)} \end{cases} \\ \begin{cases} 1 &= x_{\sigma(\alpha n+1)} + x_{\sigma(\alpha n+2)} + x_{\sigma(\alpha n+3)} \\ &\vdots \\ 1 &= x_{\sigma(n-2)} + x_{\sigma(n-1)} + x_{\sigma(n)}\,. \end{cases} \end{cases}$$

One must take $\alpha + 3\gamma = 1$ since each entry of x should appear in exactly one constraint. Moreover, since we want that any solution to these set of constraints will be contained in $\mathsf{Th}_\tau^{-1}(1)$, one must take $\alpha + \gamma = \tau$. Solving gives

$$\alpha = \frac{3\tau - 1}{2}\,, \qquad \gamma = \frac{1-\tau}{2}.$$

Note that the affine subspaces defined by these constraints is the same for all levels, as apposed to the case $\omega \leq (3\tau+1)/4$, where the number of constraints of each type depended on ω. By our choice of α, γ, for every permutation σ it holds that $V_\sigma \subseteq \mathsf{Th}_\tau^{-1}(1)$.

We now turn to compute $|V_\sigma \cap L_{\omega n}|$. Consider x in this intersection. Suppose δn of the constraints on three variables contain exactly one entry of x with value 1. Then, $\omega = \alpha + \delta + 3(\gamma - \delta)$ and so

$$\delta = \frac{\alpha + 3\gamma - \omega}{2} = \frac{1-\omega}{2}.$$

Moreover, as in the previous case, there is a one to one mapping between the points in $V_\sigma \cap L_{\omega n}$ and the number of ways to choose δn constraints from the γn constraints on three variables and choosing which unique variable in a triplet is set to 1. Thus,

$$|V_\sigma \cap L_{\omega n}| = \binom{\gamma n}{\delta n} \cdot 3^{\delta n} \geq \frac{1}{\sqrt{8n}} \cdot 2^{\mu \cdot n},$$

where

$$\mu = \frac{1-\tau}{2} \cdot H\left(\frac{1-\omega}{1-\tau}\right) + \log_2(3) \cdot \frac{1-\omega}{2}.$$

It is left to show that $H(\omega) - \mu \leq H(\tau) - (1-\tau)$ for $(3\tau+1)/4 < \omega \leq 1$, and the proof will follow. Thus, one needs to prove that for this range of ω,

$$\phi_\tau(\omega) \triangleq H(\tau) + \frac{1-\omega}{2} \cdot \log_2(3) - H(\omega) - \frac{1-\tau}{2} \cdot \left(2 - H\left(\frac{1-\omega}{1-\tau}\right)\right) \geq 0.$$

One can verify that

$$\phi_\tau\left(\frac{3\tau+1}{4}\right) = H(\tau) - H\left(\frac{3\tau+1}{4}\right) \geq 0,$$

where the last inequality holds since $1/2 \leq \tau \leq (3\tau+1)/4$. Thus, to complete the proof it is enough to show that $\phi_\tau(\omega)$ is monotone increasing in the interval $(1/2, 1)$. To this end we consider the derivative of $\phi_\tau(\omega)$ (note that $\phi_\tau(\omega)$ is infinitely differentiable in its domain),

$$\begin{aligned}
\frac{d}{d\omega}\phi_\tau(\omega) &= \log_2\left(\frac{\omega}{1-\omega}\right) + \frac{1}{2}\log_2\left(\frac{1-\omega}{\omega-\tau}\right) - \frac{1}{2}\log_2(3) \\
&= \frac{1}{2}\log_2\left(\frac{\omega^2}{3(1-\omega)(\omega-\tau)}\right) \\
&\geq \frac{1}{2}\log_2\left(\frac{\omega^2}{3(1-\omega)(\omega-1/2)}\right).
\end{aligned}$$

The proof then follows as one can easily verify that the expression inside the $\log(\cdot)$ is at least 1 for any $1/2 < \omega < 1$. □

6 The Inner Product Function and Low Degree Polynomials

We start this section by proving Theorem 1.5. As mentioned in the introduction, $\mathsf{DNF}_{\oplus}(\mathrm{IP}) \leq 2^{n/2}-1$, so it is enough to prove a matching lower bound. We first show two proofs that almost achieve this lower bound. We will obtain the tight bound afterwards.

The first proof uses a lemma of Akavia *et al.* [ABG+14]. To state it we move to the $\{\pm 1\}$ representation of functions. Namely, we consider functions of the form $f\colon \{0,1\}^n \to \{\pm 1\}$. Based on a result by Jackson [Jac97], Akavia *et al.* [ABG+14] proved the following lemma.

Lemma 6.1 ([ABG+14])**.** *For any function $f\colon \{0,1\}^n \to \{\pm 1\}$ it holds that*

$$\max_{S\subseteq[n]} |\hat{f}(S)| \geq \frac{1}{2\mathsf{DNF}_{\oplus}(f)+1}\,.$$

Since all the Fourier coefficients of IP in the $\{\pm 1\}$ representation have absolute value $2^{-n/2}$, Lemma 6.1 immediately implies that $\mathsf{DNF}_{\oplus}(\mathrm{IP}) \geq (2^{n/2}-1)/2$, which is almost tight – only factor 2 away from the upper bound.

The second proof, for which we gave a rough sketch in the introduction, is based on the well-known fact that $\mathrm{IP}\colon \{0,1\}^n \to \{0,1\}$ is an affine disperser for dimension $n/2+1$ (see Appendix A). By the proof of the second item in Theorem 1.1, we get that

$$\mathsf{DNF}_{\oplus}(\mathrm{IP}) \geq \frac{|\mathrm{IP}^{-1}(1)|}{2^{(n/2+1)-1}} = \frac{1}{2}\cdot(2^{n/2}-1),$$

which is also a factor of 2 away from the upper bound. Next, by being slightly more careful, we give the proof for the exact bound of $2^{n/2}-1$. To this end we consider again the $\{\pm 1\}$ representation of functions. We use the following result regarding small bias sets.

Lemma 6.2 ([PR04, AS10, AC13] (see Lemma 4.5 in [AC13]))**.** *Let S be an ε-biased set. Then, for any affine subspace U it holds that*

$$\left|\frac{|S\cap U|}{|S|} - \frac{|U|}{2^n}\right| \leq \varepsilon. \tag{6.1}$$

Proof of Theorem 1.5. Recall that a set S is ε-biased if and only if $|S|^{-1}\cdot|\sum_{s\in S}[(-1)^{\langle s,\alpha\rangle}]| \leq \varepsilon$ for all non-zero $\alpha \in \{0,1\}^n$. One can easily show that this is equivalent of saying that $|\widehat{1_S}(T)| \leq \varepsilon\cdot\frac{|S|}{2^{n-1}}$ for all $\emptyset \neq T \subseteq [n]$, where 1_S is the indicator function for the set S in the $\{\pm 1\}$ representation. Recall also that all non-zero Fourier coefficients of IP in the $\{\pm 1\}$ representation have absolute value of $2^{-n/2}$. By plugging this to Equation (6.1) and rearranging we get that

$$\left|\left|\mathrm{IP}^{-1}(-1)\cap U\right| - \frac{|\mathrm{IP}^{-1}(-1)|}{2^n}\cdot|U|\right| \leq 2^{n/2-1}. \tag{6.2}$$

Assume now that U is an affine subspace that is contained in $\mathrm{IP}^{-1}(-1)$. Using the fact that $|\mathrm{IP}^{-1}(-1)| = 2^{n-1} - 2^{n/2-1}$, Equation (6.2) implies that $|U|/2 < (1/2 + 2^{-n/2-1}) \cdot |U| \leq 2^{n/2-1}$. It follows that $|U| < 2^{n/2}$, and hence since the size of U is a power of 2, we conclude that $|U| \leq 2^{n/2-1}$. Therefore $\mathsf{DNF}_{\oplus}(\mathrm{IP}) \geq \frac{|\mathrm{IP}^{-1}(-1)|}{2^{n/2-1}} = 2^{n/2} - 1$, as stated. □

We now deduce a result regarding general quadratic polynomials.

Corollary 6.3. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function that has degree 2 as a polynomial over $\mathbb{F}_2$. Let $\delta = \mathrm{bias}(f) = \left|\mathbb{E}_{x \sim \{0,1\}^n}[(-1)^{f(x)}]\right|$. Then, $\mathsf{DNF}_{\oplus}(f) = \Theta(1/\delta)$.*

Proof. Dickson's theorem ([Dic01], Theorem 199) states that, up to linear transformations, all degree 2 polynomials over $\mathbb{F}_2$ are essentially the inner product function. More precisely, any degree 2 polynomial $f(x)$ over $\mathbb{F}_2$ can be written us $f(x) = \ell_0(x) + \mathrm{IP}(\ell_1(x), \ell_2(x), \ldots, \ell_r(x))$, where the ℓ_i's are independent linear functions, and $\delta = \Theta(2^{-r/2})$. Theorem 1.5 implies that the inner product function on r inputs has $\mathsf{DNF}_{\oplus}$ complexity of $\Theta(2^{r/2}) = \Theta(1/\delta)$. The proof follows since applying the inner product function on r independent linear functions (rather than on the input bits) does not change its $\mathsf{DNF}_{\oplus}$ complexity. □

The $\mathsf{DNF}_{\oplus}$ complexity of degree $d \geq 3$ polynomials

Next we prove Theorem 1.6. The first part of the theorem readily follows from the following result by Cohen and Tal [CT14].

Theorem 6.4 ([CT14], Theorem 2)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function that has degree d as a polynomial over $\mathbb{F}_2$. Then, there exists a partition of $\{0,1\}^n$ to affine subspaces, each of dimension $\Omega(n^{1/(d-1)!})$, such that f is constant on each part.*

Proof of Theorem 1.6. To deduce the first part of Theorem 1.6 from Theorem 6.4 note that the latter implies that $f^{-1}(1)$ can be written as a union of at most $|f^{-1}(1)|/2^{\Omega(n^{1/(d-1)!})} \leq 2^{n-\Omega(n^{1/(d-1)!})}$ affine subspaces. As for the moreover part, a result by Ben-Eliezer *et al.* [BEHL09] implies that a random degree d polynomial is, with high probability, an affine disperser for dimension $k = O(n^{1/(d-1)})$. This together with Theorem 1.1 imply the moreover part. □

As in the case of quadratic polynomials, one can obtain a stronger result for the case of biased polynomials of constant degree. We start with the case of degree 3 polynomials.

Theorem 6.5. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function that has degree 3 as a polynomial over $\mathbb{F}_2$. Assume that $\mathrm{bias}(f) = \delta$. Then,*

$$\mathsf{DNF}_{\oplus}(f) \leq 2^{n/2 + O(\log^4(1/\delta))}.$$

For the proof of Theorem 6.5 we need the following structural result by Haramaty and Shpilka [HS10].

Theorem 6.6 ([HS10])**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a cubic polynomial with bias δ. Then, there exists $c = O(\log^4(1/\delta))$ such that f can be written as*

$$f(x) = \sum_{i=1}^{c} \ell_i(x) q_i(x) + g(\ell_1(x), \ldots, \ell_c(x)),$$

where the ℓ_i's are linear functions, the q_i's are quadratic polynomials and g is a cubic polynomial.

Theorem 6.5 readily follows from Theorem 6.6. Indeed, for any fixing of $\ell_1(x), \ldots, \ell_c(x)$ the function f is reduced to a sum of quadratic polynomials, which is by itself a quadratic polynomial, and as mentioned above, has $\mathsf{DNF}_\oplus$ complexity at most $O(2^{n/2})$. One can then take the union over all (at most 2^c) appropriate fixings of $\ell_1(x), \ldots, \ell_c(x)$.

For biased polynomials of higher degrees one can prove the following theorem.

Theorem 6.7. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a degree d polynomial with bias δ. Then, there is some $c = c(d,\delta)$ (independent of n) such that*

$$\mathsf{DNF}_\oplus(f) \leq 2^{n - \frac{n^{1/(d-2)!}}{c^c} + c}.$$

The proof of Theorem 6.7 readily follows by a structural result for biased low degree polynomials of Kaufman and Lovett [KL08] combined with Theorem 4 in [CT14]. We omit the details.

7 The Parity Decision Tree Model

As mentioned, the parity decision tree model, defined by Kushilevitz and Mansour [KM93], has received a significant attention in the last few years [MO09, ZS10, TWXZ13, OST+13, STV14]. We start this section by proving Theorem 1.7 and then discuss the relation between the $\mathsf{DT}_\oplus$ complexity of a function and its $\mathsf{DNF}_\oplus$ complexity.

Proof of Theorem 1.7. Let g be a function with $\mathsf{DT}_\oplus(g) = s \leq \varepsilon \cdot 2^{n-k}$ such that $\mathrm{dist}(f,g) = \delta$. We want to show that $\delta \geq 1/2 - 4\varepsilon$. Since $\mathsf{DT}_\oplus(g) = s$, the function g is an indicator function of the *distinct* union of s affine subspaces $U_1, \ldots, U_s$. Let $I \subseteq [s]$ be the set of indices for which $\dim(U_i) \geq k$.

Let $c = |f^{-1}(1) \setminus g^{-1}(1)|$ be the number of ones obtained by f outside $\cup_i U_i$. For $i \in [s]$, let a_i be the number of ones obtained by f on U_i, and let b_i be the number of zeros obtained by f on U_i. Since f is an affine extractor for dimension k, with bias ε, it holds that for any $i \in I$,

$$\mathrm{bias}(f|_{U_i}) = \left|\frac{a_i - b_i}{a_i + b_i}\right| \leq \varepsilon.$$

Thus, for all $i \in I$,

$$a_i \leq \left(\frac{1+\varepsilon}{1-\varepsilon}\right) \cdot b_i \leq (1+4\varepsilon) \cdot b_i,$$

and so

$$\sum_{i\in I} a_i \leq (1+4\varepsilon)\cdot\sum_{i\in I} b_i \leq (1+4\varepsilon)\cdot(\delta\cdot 2^n - c), \tag{7.1}$$

where the last inequality follows since each $i \in [s]$ contributes b_i to the (non-relative) distance between f and g. Note that for this inequality we also use the disjointness of the U_i's. Now, by Equation (7.1) and by our assumption that $s \leq \varepsilon \cdot 2^{n-k}$, we have that

$$\begin{aligned}
|f^{-1}(1)| &\leq c + s\cdot 2^k + \sum_{i\in I} a_i \\
&\leq c + s\cdot 2^k + (1+4\varepsilon)\cdot(\delta\cdot 2^n - c) \\
&\leq s\cdot 2^k + (1+4\varepsilon)\cdot\delta\cdot 2^n \\
&\leq (\delta + \varepsilon + 4\varepsilon\delta)\cdot 2^n.
\end{aligned}$$

On the other hand, f must have bias at most ε on $\{0,1\}^n$, and so $|f^{-1}(1)| \geq (1/2 - \varepsilon/2)\cdot 2^n$. Thus,

$$\delta \geq \frac{1-3\varepsilon}{2(1+4\varepsilon)} \geq \frac{1}{2} - 4\varepsilon,$$

as desired. □

The relation between $\mathsf{DT}_\oplus$ and $\mathsf{DNF}_\oplus$

One can easily see that $\mathsf{DNF}_\oplus(f) \leq \mathsf{DT}_\oplus(f)$ for any function f, and a natural question is what can be said in the other direction. Let $\mathsf{DNF}(f), \mathsf{DT}(f)$ denote the size of the smallest DNF and smallest decision tree for computing f, respectively. Jukna *et al.* [JRSW99] gave an exponential separation between the DNF and DT complexity. Their proof is based on the observation that $\mathsf{DT}(f) \geq \|\hat{f}\|_1$, while on the other hand, there is a function with very large spectral norm that is computable by a small DNF. The Tribes function is one such example, where the DNF complexity is $O(n/\log n)$ while $\|\widehat{\mathrm{Tribes}}\|_1 \geq 2^{\Omega(n/\log n)}$. We observe that the arguments of Jukna *et al.* also gives an exponential separation between $\mathsf{DNF}_\oplus$ and $\mathsf{DT}_\oplus$ (and in fact even a separation between DNF and $\mathsf{DT}_\oplus$). This is because one can show that $\mathsf{DT}_\oplus(f) \geq \Omega(\|\hat{f}\|_1)$ (see the exercises of Chapter 4 in O'Donnell book [O'D14]).

It is worth mentioning that proving lower bounds on the $\mathsf{DT}_\oplus$ complexity of a function via the spectral norm cannot give bounds better than $2^{n/2}$, whereas Theorem 1.7 (and even Theorem 1.1) yield lower bounds of the form $2^{n-o(n)}$.

Let $\mathsf{CNF}(f)$ denote the size of the smallest CNF for computing f. A result of Ehrenfeucht and Haussler [EH89] states that an exponential separation such as above cannot occur when the CNF complexity of f is also small. More precisely, it is shown that $\mathsf{DT}(f) \leq n^{O(\log^2(\mathsf{DNF}(f)+\mathsf{CNF}(f)))}$ (see also a proof due to Savický in [Juk12], Theorem 14.32). By inspecting the proof, one can verify that the same relation also holds in the analog $\mathsf{DNF}_\oplus, \mathsf{CNF}_\oplus$ and $\mathsf{DT}_\oplus$ models. Namely, $\mathsf{DT}_\oplus(f) \leq n^{O(\log^2(\mathsf{DNF}_\oplus(f)+\mathsf{CNF}_\oplus(f)))}$. Since the verification is straightforward, we omit the proof.

We mention one more result in the standard DNF, CNF and DT models that "goes through" in the analog models with parities. So far we only discussed the size of DNFs,

CNFs and decision trees. However, one can also consider the width of DNFs and CNFs and the depth of a decision tree. For a function f, we denote by $\mathsf{C}_1(f)$ the least integer w for which there exists a width w DNF that computes f. One similarly defines $\mathsf{C}_0(f)$ as the least integer w for which there exists a width w CNF that computes f. The least integer d for which there exists a depth d decision tree that computes f is denoted by $\mathsf{D}(f)$.

Several recent papers [TWXZ13, OST+13, STV14] have studied the relation between the analog models with parities and properties of the Fourier spectrum of a function. Here we only want to point out the following relation. A classical result, that was rediscovered by several researchers [BI87, Tar89, HH91], states that $\mathsf{D}(f) \leq \mathsf{C}_0(f) \cdot \mathsf{C}_1(f)$ (see also Theorem 14.3 in [Juk12]). Again, by inspection one can verify that this result also holds in the analog model with parities.

8 Open Problems

Average case lower bounds for $\mathsf{DNF}_\oplus$. In Theorem 1.1 we proved a worst case lower bound of $2^{n-n^{o(1)}}$ for the $\mathsf{DNF}_\oplus$ model using affine dispersers. We used affine extractors to prove average case lower bounds of $2^{n-o(n)}$ for the weaker $\mathsf{DT}_\oplus$ model in Theorem 1.7. It would be interesting to prove average case lower bounds also for the $\mathsf{DNF}_\oplus$ model. We believe that proving a $2^{\Omega(n)}$ average case lower bound for this model using affine extractors is attainable. We note that an average case lower bound of the form $2^{\Omega(\sqrt{n})}$ for the majority function follows by standard arguments, and this is tight even for DNFs, as was shown by O'Donnell and Wimmer [OW07].

Affine dispersers and depth 3 Boolean circuits. In this paper we gave explicit lower bounds for the $\mathsf{DNF}_\oplus$ model using affine dispersers. We ask what is the least size of a depth 3 Boolean circuit (namely, using only $\mathsf{AND}, \mathsf{OR}$ and NOT gates) for computing an affine disperser (or affine extractors) for dimension $O(\log n)$? The best known explicit lower bound for the latter model is $2^{\Omega(\sqrt{n})}$. One function that gives this bound is the XOR function [Hås86]. Improving this lower bound is a major open problem in circuit complexity with interesting consequences [Val83] (see also Chapter 11 in [Juk12]).

It is implicit in [Raz88, Sav95] that a polynomial-size depth 3 circuit of the form $\mathsf{XOR} \circ \mathsf{AND} \circ \mathsf{XOR}$, with top fan-in $O(n \log n)$, can compute an affine extractor for dimension $O(\log n)$ (see also an appendix of [CT14] for an explicit statement). Since the XOR function on m inputs can be computed by depth 3 circuits with size $2^{O(\sqrt{m})}$, one can replace the XOR gates in the circuit above, and by collapsing layers, obtain a depth 5 Boolean circuit with size $2^{O(\sqrt{n \log n})}$ that computes an affine extractor for dimension $O(\log n)$. We ask whether depth 3 Boolean circuits with similar size can do as well?

Acknowledgement

We wish to thank Elazar Goldenberg for discussions at early stages of this work. We thank Johan Håstad for some helpful conversations regarding this work, in particular for a key idea which enabled us to prove Theorem 1.4.

References

[ABG+14] A. Akavia, A. Bogdanov, S. Guo, A. Kamath, and A. Rosen. Candidate weak pseudorandom functions in AC0 ∘ MOD2. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 251–260. ACM, 2014.

[AC13] N. Alon and G. Cohen. On rigid matrices and U-polynomials. In *Conference on Computational Complexity (CCC), 2013 IEEE*, pages 197–206. IEEE, 2013.

[AS10] V. Arvind and S. Srinivasan. The remote point problem, small bias space, and expanding generator sets. In *27th International Symposium on Theoretical Aspects of Computer Science-STACS 2010*, pages 59–70, 2010.

[BEHL09] I. Ben-Eliezer, R. Hod, and S. Lovett. Random low degree polynomials are hard to approximate. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 366–377. Springer, 2009.

[BI87] M. Blum and R. Impagliazzo. Generic oracles and oracle classes. In *28th Annual Symposium on Foundations of Computer Science, 1987.*, pages 118–126. IEEE, 1987.

[BKS+05] 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 thirty-seventh annual ACM symposium on Theory of computing*, pages 1–10. ACM, 2005.

[Blu83] N. Blum. A boolean function requiring 3n network size. *Theoretical Computer Science*, 28(3):337–345, 1983.

[Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[Bou07] J. Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

[BSK12] E. Ben-Sasson and S. Kopparty. Affine dispersers from subspace polynomials. *SIAM Journal on Computing*, 41(4):880–914, 2012.

[BSZ11] E. Ben-Sasson and N. Zewi. From affine to two-source extractors via approximate duality. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 177–186. ACM, 2011.

[CT12] T. M. Cover and A. J. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[CT14] G. Cohen and A. Tal. Two structural results for low degree polynomials and applications. *arXiv preprint arXiv:1404.0654*, 2014.

[Dic01] L. E. Dickson. *Linear groups with an exposition of the Galois field theory*. B.G Teubner's Sammlung von Lehrbuchern auf dem Gebiete der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. B.G. Teubner, 1901.

[DK11] E. Demenkov and A. S. Kulikov. An elementary proof of a 3n-o(n) lower bound on the circuit complexity of affine dispersers. In *Mathematical Foundations of Computer Science 2011*, pages 256–265. Springer, 2011.

[EH89] A. Ehrenfeucht and D. Haussler. Learning decision trees from random examples. *Information and Computation*, 82(3):231–246, 1989.

[Gro94] V. Grolmusz. A weight-size trade-off for circuits with MOD m gates. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 68–74. ACM, 1994.

[Hås86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

[HH91] J. Hartmanis and L. A. Hemachandra. One-way functions and the nonisomorphism of NP-complete sets. *Theoretical Computer Science*, 81(1):155–163, 1991.

[HS10] E. Haramaty and A. Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

[Jac97] J. C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997.

[JRSW99] S. Jukna, A. Razborov, P. Savickỳ, and I. Wegener. On P versus NP ∩ co-NP for decision trees and read-once branching programs. *Computational Complexity*, 8(4):357–370, 1999.

[Juk06] S. Jukna. On graph complexity. *Combinatorics, Probability and Computing*, 15(06):855–876, 2006.

[Juk12] S. Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springerverlag Berlin Heidelberg, 2012.

[KL08] T. Kaufman and S. Lovett. Worst case to average case reductions for polynomials. In *Foundations of Computer Science (FOCS), 2008 49th Annual IEEE Symposium on*, pages 166–175. IEEE, 2008.

[KM93] E. Kushilevitz and Y. Mansour. Learning decision trees using the fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993.

[Li11] X. Li. A new approach to affine extractors and dispersers. In *Computational Complexity (CCC), 2011 IEEE 26th Annual Conference on*, pages 137–147. IEEE, 2011.

[Lup58] O. Lupanov. A method of circuit synthesis. *Izvestia vuz Radio zike*, 1:120–140, 1958.

[MO09] A. Montanaro and T. Osborne. On the communication complexity of XOR functions. *arXiv preprint arXiv:0909.3392*, 2009.

[O'D14] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OST+13] R. O'Donnell, X. Sun, L. Y. Tan, J. Wright, and Y. Zhao. A composition theorem for parity kill number. *arXiv preprint arXiv:1312.2143*, 2013.

[OW07] R. ODonnell and K. Wimmer. Approximation by DNF: examples and counterexamples. In *Automata, Languages and Programming*, pages 195–206. Springer, 2007.

[PR04] P. Pudlák and V. Rödl. Pseudorandom sets and explicit constructions of Ramsey graphs. *Quad. Mat*, 13:327–346, 2004.

[PSZ97] R. Paturi, M. E. Saks, and F. Zane. Exponential lower bounds for depth 3 boolean circuits. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 86–91. ACM, 1997.

[Qui53] W. Quine. *Two theorems about truth functions*. Sociedade Matematica Mexicana, 1953.

[Raz88] A. Razborov. Bounded-depth formulas over $\{\wedge, \oplus\}$ and some combinatorial problems. *Complexity of Algorithms and Applied Mathematical Logic (in Russian). Ser. Voprosy Kibernetiky (Problems in Cybernetics), S. I. Adian, Ed., Moscow*, pages 149–166, 1988.

[Sav95] P. Savickỳ. Improved Boolean formulas for the Ramsey graphs. *Random Structures & Algorithms*, 6(4):407–415, 1995.

[Sha11] R. Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 247–256. IEEE, 2011.

[STV14] A. Shpilka, A. Tal, and B. Volk. On the structure of boolean functions with small spectral norm. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 37–48. ACM, 2014.

[Tar89] G. Tardos. Query complexity, or why is it difficult to separate $NP^A \cap coNP^A$ from P^A by random a oracle A? *Combinatorica*, 9(4):385–392, 1989.

[TWXZ13] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. *arXiv preprint arXiv:1304.1245*, 2013.

[Val83] L. G. Valiant. Exponential lower bounds for restricted monotone circuits. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 110–117. ACM, 1983.

[Vio14] E. Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

[VW07] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Computational Complexity, 2007. CCC'07. Twenty-Second Annual IEEE Conference on*, pages 141–154. IEEE, 2007.

[Yeh11] A. Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

[ZS10] Z. Zhang and Y. Shi. On the parity complexity measures of boolean functions. *Theoretical Computer Science*, 411(26):2612–2618, 2010.

A The Inner Product Function is an Affine Extractor

In this section, for completeness, we give two proofs for the following folklore result.

Theorem A.1. *Let $n \geq 2$ be an even integer and let $c \geq 1$ be an integer. Then, the inner product function* IP *on n inputs is an affine extractor for dimension $k = n/2 + c$ with bias $\varepsilon \leq 2^{-c}$. In particular,* IP *is an affine disperser for dimension $n/2 + 1$.*

Proof 1. Let U be an affine subspace with dimension $k = n/2 + c$. In order to prove that IP is balanced on U, we make the following observation.

Observation A.2. *Let $f : \{0,1\}^n \to \{\pm 1\}$ be a function, and let $g : \{0,1\}^{n-1} \to \{\pm 1\}$ be obtained by restricting f to some $n-1$ dimensional affine subspace. Suppose for concreteness that the affine subspace is $\{x \in \{0,1\}^n : x_n = \sum_{j \in J} x_j + b\}$ for some $J \subseteq [n-1]$ and*

$b \in \{0,1\}$. *Then, we may write* g *as* $g(x_1,\ldots,x_{n-1}) = f(x_1,\ldots,x_{n-1},\sum_{j\in J} x_j + b)$, *and hence for any* $S \subseteq [n-1]$ *we have* $\hat{g}(S) = \hat{f}(S) + (-1)^b \cdot \hat{f}((S \cup \{n\})\Delta J)$.

By applying this observation repeatedly, we conclude that if g is the restriction of f to the affine subspace U, then each Fourier coefficient of g is a sum/difference of 2^{n-k} Fourier coefficient of f. In particular, for $f = \mathrm{IP}$, since $|\widehat{\mathrm{IP}}(S)| = -2^{n/2}$ for all $S \subseteq [n]$, each Fourier coefficient of its restriction $\mathrm{IP}|_U$ is at most $2^{n-k} \cdot 2^{-n/2} = 2^{-c}$ in absolute value. In particular, the bias of $\mathrm{IP}|_U$, which is equal to the empty coefficient of the restricted function, is at most $\mathrm{bias}(\mathrm{IP}|_U) = \widehat{\mathrm{IP}}(\emptyset) \leq 2^{-c}$, as required. □

Next we give an alternative proof for Theorem A.1. The proof gives a bound of $2^{-c}+2^{-n/2}$ on the bias ε, which is slightly weaker than the bound 2^{-c} stated in Theorem A.1. Still, we find the proof interesting.

Proof 2. Let U be an affine subspace with dimension $k = n/2 + c$. We will use the fact that in the $\{\pm 1\}$ representation, all Fourier coefficients of IP have absolute value of $2^{-n/2}$. Thus, similarly to the way it was done in the proof of Theorem 1.5, one can show that $\mathrm{IP}^{-1}(1)$ is an ε-biased set with $\varepsilon = \frac{2^{n/2-1}}{|\mathrm{IP}^{-1}(1)|}$. Lemma 6.2 states that for any ε-biased set S and for any affine subspace U it holds that

$$\left| \frac{|S \cap U|}{|S|} - \frac{|U|}{2^n} \right| \leq \varepsilon. \tag{A.1}$$

Thus, by plugging $|U| = 2^{n/2+c}$ and $\varepsilon = \frac{2^{n/2-1}}{|\mathrm{IP}^{-1}(1)|}$ to Equation (A.1) and rearranging we get

$$\left| \frac{|\mathrm{IP}^{-1}(-1) \cap U|}{|U|} - \frac{|\mathrm{IP}^{-1}(-1)|}{2^n} \right| \leq \varepsilon \cdot \frac{|\mathrm{IP}^{-1}(-1)|}{|U|} = 2^{-c-1}. \tag{A.2}$$

Since $|\mathrm{IP}^{-1}(-1)|/2^n = 1/2 - 2^{-n/2-1}$ it follows that the bias of IP on U is

$$\mathrm{bias}(\mathrm{IP}|_U) = 2 \cdot \left| \frac{|\mathrm{IP}^{-1}(-1) \cap U|}{|U|} - \frac{1}{2} \right| \leq 2^{-c} + 2^{-n/2}.$$

□

ECCC ISSN 1433-8092
http://eccc.hpi-web.de