Electronic Colloquium on Computational Complexity, Report No. 44 (2016)

Structure of protocols for XOR functions

Kaave Hosseini *
Computer Science and Engineering
University of California, San Diego
skhossei@ucsd.edu

Shachar Lovett †
Computer Science and Engineering
University of California, San Diego
slovett@ucsd.edu

March 18, 2016

Abstract

Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Its associated XOR function is the two-party function $f_\oplus(x,y) = f(x \oplus y)$. We show that, up to polynomial factors, the deterministic communication complexity of $f_\oplus$ is equal to the parity decision tree complexity of f. This relies on a novel technique of entropy reduction for protocols, combined with existing techniques in Fourier analysis and additive combinatorics.

1 Introduction

Let $F : X \times Y \to \{0,1\}$ be a boolean function and suppose Alice and Bob are holding $x \in X$ and $y \in Y$, respectively. A natural question capturing the essence of communication complexity is the following: How much communication between Alice and Bob is required to compute $F(x,y)$ in the worst case? One of the fundamental open problems in communication complexity, the log-rank conjecture, links this question to the rank of F as a real matrix.

Conjecture 1.1. *(Log-rank conjecture [LS93]) Is it true that for every boolean function $F : X \times Y \to \{0,1\}$,*

$$D(F) \leq \text{polylog}(rank(F))$$

where $D(\cdot)$ is the deterministic communication complexity.

Yet, after over 30 years of active research, we are far from settling this conjecture, directing attention towards solving log-rank for special classes of boolean functions. A natural and important such class is the so called XOR functions.

Let $\mathbb{F}_2^n$ be the n-dimensional vector space over the field of two elements. For a given function $f : \mathbb{F}_2^n \to \{0,1\}$ define its XOR function as $f_\oplus(x,y) = f(x+y)$. This class of functions is large enough to capture many interesting examples (e.g., equality and Hamming distance functions), but it is also especially attractive for it allows use of tools from discrete Fourier analysis. This is because the eigenvalues of $f_\oplus$ as a matrix are the same as the Fourier coefficients of f; therefore, $\text{rank}(f_\oplus) = |\text{supp}(\widehat{f})|$. Moreover, if $A \times B \subset \mathbb{F}_2^n \times \mathbb{F}_2^n$ is a monochromatic rectangle in $f_\oplus$, then f

*Supported by NSF CAREER award 1350481.
†Supported by NSF CAREER award 1350481 and a Sloan fellowship.

ISSN 1433-8092

is constant on all of $A + B$, where the sum set $A + B$ is defined as $\{a + b : a \in A, b \in B\}$. This directly links communication complexity of XOR functions to the structure of sum sets in additive combinatorics. We will discuss this relation in more details later.

Going back to the log-rank conjecture for XOR functions, an interesting approach to settle the conjecture is via another complexity measure, called the parity decision tree complexity (PDT in short), denoted $\mathrm{pdt}(\cdot)$. A PDT for a boolean function f is an extension of the usual notion of decision trees. While in a regular decision tree, intermediate nodes query variables, in a parity decision tree they are allowed to query an arbitrary linear function of the inputs. If a function f has a PDT of depth k, then it admits a natural protocol for its associated XOR function, which computes $f_\oplus(x, y)$ by simulating the PDT for f evaluated on $x \oplus y$. This shows that $D(f_\oplus) \leq 2 \cdot \mathrm{pdt}(f)$.

Our main interest in this work is whether this relation can be reversed. Namely, is is true that an efficient deterministic protocol for a XOR functions implies a low depth parity decision tree for the boolean function. Our main result is a polynomial relation between the two.

Theorem 1.2 (Main theorem)**.** *For any $f : \mathbb{F}_2^n \to \{0, 1\}$ we have* $\mathrm{pdt}(f) \leq O(D(f_\oplus)^6)$.

1.1 Proof overview

Fix $f : \mathbb{F}_2^n \to \{0, 1\}$, where we assume that $f_\oplus$ has an efficient deterministic protocol. Our goal is to design a low depth PDT for f.

Reduction to monochromatic subspaces. Note that if f has a PDT of depth k, then in particular, the leaves of the PDT determine affine subspaces of co-dimension $\leq k$ on which f is contant. We call such subspaces *monochromatic subspaces* for f. From here onwards, we use "subspace" as a shorthand for "affine subspace".

It turns out that in order to design a PDT for f, it suffices to show that there exists a large monochromatic subspace for f. This follows from [TWXZ13] who showed (among others) that if f is constant on a subspace V, then the Fourier sparsity of f restricted to any coset of V reduces by at least a factor of two. This is sufficient for our application, as the existence of an efficient deterministic protocol for $f_\oplus$ implies in particular that f has low Fourier sparsity. This reduces Theorem 1.2 to the following question, which is the main problem we investigate in this paper.

Question 1.3. *Let $f : \mathbb{F}_2^n \to \{0, 1\}$ with with $D(f_\oplus) \leq k$. Find a subspace V of co-dimesion* $\mathrm{poly}(k)$ *on which f is constant.*

In the next few paragraphs we give a brief discussion of how to find such a subspace. We first describe a natural approach, which only tries to exploit the existence of a large monochromatic rectangle for $f_\oplus$ (many techniques in communication complexity follow this approach; in the randomized settings, one needs to replace "monochromatic rectangle" with "biased rectangle"). However, as we discuss below, a direct application of this technique fails, and a more careful application requires unproven conjectures in additive combinatorics. As such, we follow a different route, which exploits the entire structure of the protocol. This is somewhat uncommon in communication complexity, and we view this is as a conceptual contribution of this work.

Using a large monochromatic rectangle, and why it fails. The existence of an efficient deterministic protocol for $f_\oplus$ implies that it is constant on a large rectangle $A \times B$. This implies that f is constant on $A + B$. As a first attempt, one may hope that if $A, B \subset \mathbb{F}_2^n$ are large sets,

then $A + B$ must contain a large subspace. This would directly imply that f is constant on this subspace. Unfortunately this is false, as the following example of Green [Gre04] shows.

Example 1.4. *Let $A = B = \mathcal{B}(n/2 - \sqrt{n})$ where $\mathcal{B}(r) \subset \{0,1\}^n$ is the hamming ball of radius r. Then $|A| = |B| = \Omega(2^n)$, $A + B = \mathcal{B}(n - 2\sqrt{n})$ but the largest subspace contained in $A + B$ has co-dimension $2\sqrt{n}$ (for example, such a subspace can be obtained by fixing the first $2\sqrt{n}$ bits to zero).*

The situation becomes better when looking on sum sets of more than two sets. Sanders [San12] showed that for a set $A \subset \mathbb{F}_2^n$ with $|A| \geq \varepsilon 2^n$, $4A = A + A + A + A$ contains a subspace of co-dimension $O(\log^4(1/\varepsilon))$. As Yao showed [Yao15], this result directly implies that a deterministic protocol for the 4-party function $F(x, y, z, w) = f(x \oplus y \oplus z \oplus w)$ with complexity k implies the existence of a parity decision tree of depth $O(k^5)$ for f.

Going back to two-fold sum sets, we can use the assumption that f has few nonzero Fourier coefficients, and require to find only a near-monochromatic subspace in $A + B$. Such a subspace would automatically imply the existence of a large monochromatic subspace for f. The reason is that if f is a boolean function with at most 2^k non-zero Fourier coefficients, then all the Fourier coefficients have the form $a/2^k$ for an integer a (for a proof see [GOS+11]). In particular, if V is a subspace on which $\mathbb{E}[f|_V] < 2^{-k}$ then in fact $f|_V = 0$, and if $\mathbb{E}[f|_V] > 1 - 2^{-k}$ then in fact $f|_V = 1$.

Therefore, it is enough to show that if $A, B \subset \mathbb{F}_2^n$ are large sets, then $A + B$ contains most of a large subspace. Working out the details, it turns out that we would need the following conjecture.

Conjecture 1.5. *Let $A \subset \mathbb{F}_2^n$ of size $|A| \geq \varepsilon 2^n$. Then for any $\delta > 0$ there exists a subspace V such that $|2A \cap V| \geq (1 - \delta)|V|$, where the co-dimension of V is at most* $\mathrm{polylog}(1/\varepsilon\delta)$.

For this and related conjectures see [SS14] (in particular section 9, the paragraph on correlations of $2A, 3A, 4A$). We note that two partial results towards Conjecture 1.5 are known, both due to Sanders:

- [San10] proves the existence of a subspace with co-dimension $O((1/\varepsilon)\log(1/\delta))$.
- [San12] proves the existence of a subspace with co-dimension $O((1/\delta^2)\log^4(1/\varepsilon))$.

Unfortunately, neither version is strong enough for our application. If $f_\oplus$ has a deterministic protocol which sends k bits, then the largest monochromatic rectangle has size $|A|, |B| \geq 2^{n-k}$. We thus have $\varepsilon = 2^{-k}$. Furthermore, $f_\oplus$ has at most 2^k nonzero Fourier coefficients, which means that we need a subspace which is 2^{-k} close to monochromatic. This means that we need to set $\delta < 2^{-k}$. As our goal is to get a subspace of co-dimension $\mathrm{poly}(k)$, we need poly-logarithmic dependency on both ε and δ.

Our approach: utilizing the entire protocol. We circumvent the need to use unproven conjectures by devising an alternative route, which exploits the entire structure of the protocol. Fix a deterministic protocol for $f_\oplus$ which sends k bits, and let $K = 2^k$. Let $A_i \times B_i$ for $i \in [K]$ be the partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ induced by the protocol. By our assumption $f_\oplus$ is constant on each $A_i \times B_i$, which means that f is constant on each $A_i + B_i$.

Let $\mu = \mathbb{E}[f]$ be the average of f on the entire space, and assume without loss of generality that $\mu \geq 1/2$. We may use the existence of a large monochromatic rectangle to find a large subspace V on which the average of f is far from the global average. Concretely, let $A \times B$ be the largest

rectangle on which f equals to zero. It can be shown that $|A|, |B| \geq 2^{n-2k}$. The results of [San12] imply the existence of a subspace V such that $|V \cap (A+B)| \geq (3/4)|V|$, where the co-dimension of V is $O(k^4)$. This implies that $\mathbb{E}[f|_V] \leq 1/4$.

Thus, f is not pseudo-random with respect to the subspace V. More concretely, it cannot be the case that

$$\frac{|A_i \cap V|}{|V|} \approx \frac{|A_i|}{2^n}, \frac{|B_i \cap V|}{|V|} \approx \frac{|B_i|}{2^n} \qquad \forall i \in [K],$$

as this would imply that $\mathbb{E}[f|_V] \approx \mathbb{E}[f]$. In fact, the same holds if we replace V with any coset of V, as the restriction of $f_\oplus$ to $(V+w) \times (V+w)$ still computes $f|_V$. We exploit this lack of pseudo-randomness, and show that f simplifies when restricted to a typical coset of V.

Technically, we show this by analyzing the entropy of the protocol. Let $p_i = \frac{|A_i \times B_i|}{2^{2n}}$ denote the density of each rectangle. We define the entropy of the partition $P = \{A_i \times B_i : i \in [K]\}$ to be the Shannon entropy of the induced distribution, namely

$$h(P) = h(p_1, \ldots, p_K) = \sum_{i=1}^{k} p_i \log(1/p_i).$$

Our main technical lemma (Lemma 3.1) shows whenever the average of a function f is far from its average on a subspace V, then if we restrict $\mathcal{P}$ to a typical coset $(V+w_1) \times (V+w_2)$ of $V \times V$ then the entropy of the restricted partition reduces by a constant. Concretely, if we assume that $\mathbb{E}[f] \geq 1/2, \mathbb{E}[f|_V] \leq 1/4$ then

$$\mathbb{E}_{w_1, w_2 \in \mathbb{F}_2^n} \left[h(P|_{(V+w_1)\times(V+w_2)}) \right] \leq h(P) - 2^{-25}.$$

In particular, there exists a coset $(V+w_1) \times (V+w_2)$ on which the entropy decreases by at least 2^{-25}. We may now iterate this process. As originally we have $h(P) \leq k$ (since the partition $\mathcal{P}$ is to $K = 2^k$ rectangles), after $O(k)$ iterations we would reach a constant function on a subspace of co-dimension $O(k^5)$.

Paper organization. We give some preliminary definitions in Section 2. We state and prove our main technical lemma, Lemma 3.1, in Section 3. We apply it to prove Theorem 1.2 in Section 4. We discuss some open problems in Section 5.

2 Preliminaries

Partitions. A labeled partition (or simply a partition) P of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ is a family $P = \{(R_i, z_i) : i \in [m]\}$ where $R_i = A_i \times B_i$, $A_i, B_i \subset \mathbb{F}_2^n$, such that $\{R_i : i \in [m]\}$ forms a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$, and $z_i \in \{0,1\}$. We also view P as a function $P : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \{0,1\}$, where $P(x,y) = z_i$, where i is the unique index for which $(x,y) \in R_i$. For a subspace V over $\mathbb{F}_2$ we extend these definitions to $P : V \times V \to \{0,1\}$, by identifying $V \cong \mathbb{F}_2^n$ for $n = \dim(V)$.

Restriction to subspaces. Let $V \subset \mathbb{F}_2^n$ be a linear subspace, and let W be the quotient subspace $\mathbb{F}_2^n/V$, so that $\{V+w : w \in W\}$ are the cosets of V in $\mathbb{F}_2^n$. Let $P = \{(R_i, z_i) : i \in [m]\}$ be a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$. For any $w', w'' \in W$ we define $P|_{V,w',w''}$ to be a partition of $V \times V$, induced by the restriction of P to $(V+w') \times (V+w'')$, shifted to $V \times V$. That is,

$$P|_{V,w',w''} = \{((R_i + (w', w'')) \cap (V \times V), z_i) : i \in [m]\}$$

where $R + (w', w'') = \{(x + w', y + w'') : (x, y) \in R\}$.

Entropy of partitions. Let $p_1, \ldots, p_m$ be a probability distribution, that is, $p_i \geq 0$ and $\sum p_i = 1$. Its entropy is

$$h(p_1, \ldots, p_m) = \sum_{i=1}^{m} p_i \log(1/p_i)$$

where here and throughout the paper, logarithms are in base two. The entropy of a partition is the entropy of the distribution it induces on rectangles. Let $P = \{(R_i, z_i) : i \in [m]\}$ be a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$. Let $p_i = |R_i|/2^{2n}$. We define

$$h(P) = h(p_1, \ldots, p_m).$$

Partition averaged function. Given a partition P of $\mathbb{F}_2^n \times \mathbb{F}_2^n$, we define its averaged function $f_P : \mathbb{F}_2^n \to [0, 1]$ by

$$f_P(x) = \mathbb{E}_{y \in \mathbb{F}_2^n} P(y, x + y).$$

Note that if P corresponds to a deterministic protocol for a XOR function $f_\oplus$ then $f_P = f$.

2.1 Technical claims

Claim 2.1. *Let X be a random variable, finitely supported on $[0, \infty)$. Assume that $\mathbb{E}X = 1$ and $\mathbb{E}|X - 1| = \eta$. Then*

$$\mathbb{E}[X \log X] \geq \eta^2/2.$$

Proof. We apply Pinsker's inequality: for any two distributions $p_1, \ldots, p_m$ and $q_1, \ldots, q_m$,

$$\sum p_i \log(p_i/q_i) \geq (1/2) \left(\sum |p_i - q_i| \right)^2.$$

Assume that X is supported on $x_1, \ldots, x_m \in [0, \infty)$. Let $p_i = x_i \Pr[X = x_i]$ and $q_i = \Pr[X = x_i]$. Note that $p_1, \ldots, p_m$ and $q_1, \ldots, q_m$ are indeed distributions,

$$\sum_i p_i \log(p_i/q_i) = \sum_i \Pr[X = x_i] x_i \log x_i = \mathbb{E}[X \log X]$$

and

$$\sum_i |p_i - q_i| = \sum_i \Pr[X = x_i] |x_i - 1| = \mathbb{E}|X - 1|.$$

Thus $\mathbb{E}[X \log X] \geq \mathbb{E}[|X - 1|]^2/2 = \eta^2/2$. □

3 Entropy decrease lemma

Let $V \subset \mathbb{F}_2^n$ be a linear subspace and let W be a dual subspace, so that $\mathbb{F}_2^n = V + W$. Let P be a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$. Define

$$h_V(P) := \mathbb{E}_{w', w'' \in W}[h(P|_{V, w', w''})].$$

We will show that always $h_V(P) \leq h(P)$, and furthermore, the gap is noticeable if the average of f_P on V differs from its global average. Recall that $f_P : \mathbb{F}_2^n \to [0, 1]$ is defined by $f_P(x) = \mathbb{E}_{y \in \mathbb{F}_2^n} P(y, x + y)$. Let $\mathbb{E}[f_P] = \mathbb{E}[P]$ denote the global average value of f_P, and let $\mathbb{E}[f_P|V] = \mathbb{E}_{x \in V} f_P(x)$ denote its average restricted to V.

Lemma 3.1. *Let $V \subset \mathbb{F}_2^n$ be a linear subspace, P a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$. Assume that*

$$|\mathbb{E}[f_P] - \mathbb{E}[f_P|V]| \geq \varepsilon.$$

Then

$$h(P) - h_V(P) \geq \varepsilon^5/20000.$$

We assume without loss of generality that $\mathbb{E}[f_P] \geq \mathbb{E}[f_P|V] + \varepsilon$, the other case being analogous (flip the labels of all rectangles in P). We first set some notations. Let $P = \{(R_i, z_i) : i \in [m]\}$ where $R_i = A_i \times B_i$. Define

$$\alpha_i := \frac{|A_i|}{2^n}, \quad \beta_i := \frac{|B_i|}{2^n}, \quad p_i := \alpha_i \beta_i.$$

Note that $h(P) = \sum_i p_i \log(1/p_i)$. For each $w', w'' \in W$ define

$$\alpha_{i|w'} := \frac{|A_i \cap (V + w')|}{|V|}, \quad \beta_{i|w''} := \frac{|B_i \cap (V + w'')|}{|V|}, \quad p_{i|w',w''} := \alpha_{i|w'}\beta_{i|w''}.$$

Note that $\mathbb{E}_{w'}[\alpha_{i|w'}] = \alpha_i$, $\mathbb{E}_{w''}[\beta_{i|w''}] = \beta_i$ and $\mathbb{E}_{w',w''}[p_{i|w',w''}] = p_i$, where here and throughout the section, $w', w'' \in W$ are uniformly chosen. Furthermore, define

$$q_{i|w',w''} = \frac{p_{i|w',w''}}{p_i}.$$

Note that $\mathbb{E}_{w',w''}[q_{i|w',w''}] = 1$ for all $i \in [m]$.

Claim 3.2. $h(P) - h_V(P) = \sum_i p_i \cdot \mathbb{E}_{w',w'' \in W}\left[q_{i|w',w''} \log(q_{i|w',w''})\right]$.

Proof. Using the fact that $\mathbb{E}_{w',w''}[p_{i|w',w''}] = p_i$ for all $i \in [m]$, we have

$$\begin{aligned} h(P) - h_V(P) &= \sum_i \left\{p_i \log(1/p_i) - E_{w',w'' \in W}\left[p_{i|w',w''} \log(1/p_{i|w',w''})\right]\right\} \\ &= \sum_i \mathbb{E}_{w',w'' \in W}\left[p_{i|w',w''} \log(p_{i|w',w''}/p_i)\right] \\ &= \sum_i p_i \cdot \mathbb{E}_{w',w'' \in W}\left[q_{i|w',w''} \log(q_{i|w',w''})\right]. \end{aligned}$$

□

Let us shorthand $\Delta_i := \mathbb{E}_{w',w'' \in W}\left[q_{i|w',w''} \log(q_{i|w',w''})\right]$. Note that as the function $x \log x$ is convex, and $\mathbb{E}_{w',w''}[q_{i|w',w''}] = 1$, we obtain that

$$\Delta_i \geq 0 \qquad \forall i \in [m]. \tag{1}$$

In particular, $h(P) - h_V(P) \geq 0$ holds for any subspace V. However, our goal is to show under the conditions of Lemma 3.1, that there is a significant gap.

To this end, define $I := \{i \in [m] : z_i = 1\}$. We may express $\mathbb{E}[f_P]$ and $\mathbb{E}[f_P|V]$ as

$$\mathbb{E}[f_P] = \mathbb{E}_{x,y \in \mathbb{F}_2^n} P(x, y) = \sum_{i \in I} p_i \tag{2}$$

and

$$\mathbb{E}[f_P|V] = \mathbb{E}_{v\in V, x\in\mathbb{F}_2^n} P(x, v+x) = \mathbb{E}_{w\in W}\mathbb{E}_{v',v''\in V} P(v'+w, v''+w) = \mathbb{E}_{w\in W}\sum_{i\in I} p_{i|w,w}. \quad (3)$$

The main idea is that we cannot have $\alpha_{i|w} \approx \alpha_i, \beta_{i|w} \approx \beta_i$ for all $w \in W, i \in [m]$, as otherwise $p_i \approx p_{i|w,w}$ and $\mathbb{E}[f_P] \approx \mathbb{E}[f_P|V]$, contradicting our assumption. Thus, for a typical $i \in I$, either $\{\alpha_{i|w} : w \in W\}$ or $\{\beta_{i|w} : w \in W\}$ must have a noticable variance. To formalize this let $\eta > 0$ to be determined later and define

$$\begin{aligned} I_1 &:= \{i \in I : \mathbb{E}_{w\in W}|\alpha_{i|w} - \alpha_i| \geq \eta^2\alpha_i\} \\ I_2 &:= \{i \in I : \mathbb{E}_{w\in W}|\beta_{i|w} - \beta_i| \geq \eta^2\beta_i\} \\ I_3 &:= I \setminus (I_1 \cup I_2). \end{aligned}$$

Claim 3.3. *If $i \in I_1 \cup I_2$ then $\Delta_i \geq \eta^4/2$.*

Proof. We show the proof only for $i \in I_1$, as the proof for $i \in I_2$ is analogous. Fix $i \in I_1$. Let $q_{i|w'} = \mathbb{E}_{w''\in W}[q_{i|w',w''}] = \frac{\alpha_{i|w'}}{\alpha_i}$. By convexity of the function $x \log x$ we have

$$\Delta_i = \mathbb{E}_{w',w''\in W}\left[q_{i|w',w''}\log(q_{i|w',w''})\right] \geq \mathbb{E}_{w'\in W}\left[q_{i|w'}\log(q_{i|w'})\right].$$

Let X be a random variable, given by $X = q_{i|w'}$ for a uniformly chosen $w' \in W$. Then X is finitely supported on $[0,\infty)$ and satisfies $\mathbb{E}X = 1$ and $\mathbb{E}|X-1| \geq \eta^2$. By Claim 2.1 we have

$$\Delta_i = \mathbb{E}[X\log X] \geq \eta^4/2.$$

□

So, we need to show that between I_1 and I_2 we have a noticeable fraction of the probability mass. This is given by the following claim.

Claim 3.4. $\sum_{i\in I_1\cup I_2} p_i \geq \varepsilon - 5\eta$.

Proof. Fix $i \in I_3$. As $i \notin I_1$ we have that $\mathbb{E}_{w\in W}|\alpha_{i|w} - \alpha_i| \leq \eta^2\alpha_i$. Thus, by Markov inequality we have that $\Pr_{w\in W}[|\alpha_{i|w} - \alpha_i| \leq \eta\alpha_i] \geq 1-\eta$. Similarly, as $i \notin I_2$ we have $\Pr_{w\in W}[|\beta_{i|w} - \beta_i| \leq \eta\beta_i] \geq 1-\eta$. Let

$$S_i = \{w \in W : |\alpha_{i|w} - \alpha_i| \leq \eta\alpha_i \text{ and } |\beta_{i|w} - \beta_i| \leq \eta\beta_i\}.$$

By the union bound, $|S_i| \geq (1-2\eta)|W|$.

Next, fix $w \in S_i$ and let $a_{i,w} = \alpha_{i|w}/\alpha_i$ and $b_{i,w} = \beta_{i|w}/\beta_i$. Note that $a_{i,w}, b_{i,w} \in [1-\eta, 1+\eta]$. We have

$$\left|\frac{p_{i|w,w}}{p_i} - 1\right| = |a_{i,w}b_{i,w} - 1| \leq a_{i,w}|b_{i,w} - 1| + |a_{i,w} - 1| \leq (1+\eta)\eta + \eta \leq 3\eta.$$

Thus, for any $i \in I_3$ we have

$$\mathbb{E}_{w\in W}[p_{i|w,w}] = \frac{1}{|W|}\sum_{w\in W} p_{i|w,w} \geq \frac{1}{|W|}\sum_{w\in S_i} p_{i|w,w} \geq \frac{|S_i|}{|W|}(1-3\eta)p_i \geq (1-5\eta)p_i.$$

Hence

$$\mathbb{E}[f_P|V] = \mathbb{E}_{w\in W}\sum_{i\in I} p_{i|w,w} \geq \mathbb{E}_{w\in W}\sum_{i\in I_3} p_{i|w,w} \geq (1-5\eta)\sum_{i\in I_3} p_i \geq \sum_{i\in I_3} p_i - 5\eta.$$

On the other hand, we know by our assumptions that

$$\mathbb{E}[f_P|V] \leq \mathbb{E}[f_P] - \varepsilon = \sum_{i\in I} p_i - \varepsilon.$$

We conclude that

$$\sum_{i\in I_3} p_i \leq \sum_{i\in I} p_i - \varepsilon + 5\eta.$$

The lemma follows as $\sum_{i\in I_1\cup I_2} p_i = \sum_{i\in I} p_i - \sum_{i\in I_3} p_i$. □

We now conclude the proof of Lemma 3.1. Set $\eta = \varepsilon/10$ so that $\sum_{i\in I_1\cup I_2} p_i \geq \varepsilon/2$. Thus

$$h(P) - h_V(P) = \sum_i p_i\Delta_i \geq \sum_{i\in I_1\cup I_2} p_i\Delta_i \geq \sum_{i\in I_1\cup I_2} p_i\eta^4/2 \geq \varepsilon^5/20000.$$

4 Deterministic protocols for XOR functions

Let $f : \mathbb{F}_2^n \to \{0,1\}$ be a boolean function. The associated XOR function is $f_\oplus(x,y) = f(x+y)$. A deterministic protocol for $f_\oplus$ corresponds to a partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ to rectangles $A_i \times B_i$, such that $f_\oplus$ is constant on each $A_i \times B_i$. Equivalently, f is constant on each $A_i + B_i$. Let $\mathrm{D}^\oplus(f)$ denote the minimum complexity of a deterministic protocol which computes $f_\oplus$.

We restate Theorem 1.2, which we prove in this section, for the convenience of the reader.

Theorem 1.2 (Main theorem). *For any $f : \mathbb{F}_2^n \to \{0,1\}$ we have $\mathrm{pdt}(f) \leq O(\mathrm{D}^\oplus(f)^6)$.*

Tsang et al. [TWXZ13] showed that in order to design a parity decision tree, it suffices to find a large subspace on which the function is constant; and then recurse. For completeness, we reproduce their argument. Let $\mathrm{rank}(f)$ denote the rank of the real matrix $M_{x,y} = f(x+y)$. It equals the number of nonzero Fourier coefficients of f. Note that $\log \mathrm{rank}(f) \leq \mathrm{D}^\oplus(f)$.

Lemma 4.1. *Let $T : \mathbb{N} \to \mathbb{N}$ be a function for which the following holds. For any function $f : \mathbb{F}_2^n \to \{0,1\}$, if $\mathrm{D}^\oplus(f) = k$ then there exists an affine subspace V of co-dimension $T(k)$ on which f is constant. Then for any function $f : \mathbb{F}_2^n \to \{0,1\}$, $\mathrm{pdt}(f) \leq T(\mathrm{D}^\oplus(f)) \cdot \mathrm{D}^\oplus(f)$.*

Proof. The main idea is that if f is constant on V, then its rank on any coset of V reduces by at least a factor of two, which then allows for induction. To see that, assume that $\mathrm{rank}(f) = r$. Then

$$f(x) = \sum_{i=1}^{r} \hat{f}(\alpha_i)(-1)^{\langle x,\alpha_i\rangle},$$

for some $\alpha_1, \ldots, \alpha_r \in \mathbb{F}_2^n$. We know by assumption that f is constant on an affine subspace V of co-dimension $t = T(\mathrm{D}^\oplus(f))$. We may assume that V is linear subspace, by replacing $f(x)$ with $f(x+v)$ for some $v \in V$ (note that this does not change $\mathrm{D}^\oplus$ or $\mathrm{rank}(f)$). Let W be the quotient subspace $\mathbb{F}_2^n/V$ so that $\dim(W) = t$ and $\mathbb{F}_2^n = V + W$. Note that any $x \in \mathbb{F}_2^n$ can be uniquely

decomposed as $x = v + w$ with $v \in V, w \in W$. Let $\pi_V : \mathbb{F}_2^n \to V$ and $\pi_W : \mathbb{F}_2^n \to W$ be the projection maps to V and W, respectively, mapping $x = v + w$ to $\pi_V(x) = v$ and $\pi_W(x) = w$. Then

$$f|_V(v) = \sum_{i=1}^{r} \hat{f}(\alpha_i)(-1)^{\langle v, \pi_V(\alpha_i)\rangle},$$

In particular, as f is constant on V, it must be the case that for every non-zero α_i there exists some α_j such that $\pi_V(\alpha_i) = \pi_V(\alpha_j)$, or equivalently $\alpha_i + \alpha_j \in W$. Thus

$$|\{\pi_V(\alpha_i) : i \in [r]\}| \leq \frac{r+1}{2}.$$

Let $V + w$ be any coset of V. Then

$$f|_{V+w}(v + w) = \sum_{i=1}^{r} \hat{f}(\alpha_i)(-1)^{\langle w, \pi_W(\alpha_i)\rangle}(-1)^{\langle v, \pi_V(\alpha_i)\rangle}.$$

In particular, $\mathrm{rank}(f|_{V+w}) \leq |\{\pi_V(\alpha_i) : i \in [r]\}| \leq \frac{\mathrm{rank}(f)+1}{2}$.

We now construct the parity decision tree for f. We first query $w = \pi_W(x)$, which requires depth $\dim(W) = T(\mathrm{D}^{\oplus}(f))$. Each restricted function $f|_{V+w}$ has $\mathrm{D}^{\oplus}(f|_{V+w}) \leq \mathrm{D}^{\oplus}(f)$ and $\mathrm{rank}(f|_{V+w}) \leq \frac{\mathrm{rank}(f)+1}{2}$, and hence by induction can be computed by a parity decision tree of depth at most $T(\mathrm{D}^{\oplus}(f)) \cdot (\log(\mathrm{rank}(f)) + 1) \leq T(\mathrm{D}^{\oplus}(f)) \cdot (\mathrm{D}^{\oplus}(f) + 1)$. The lemma follows. □

From now on, we focus on the task of finding a large subspace on which f is constant. We use the following result of Sanders [San12] (see also [CS10] and [CŁS13]).

Theorem 4.2. *Let $A, B \subset \mathbb{F}_2^n$ be sets of size $|A|, |B| \geq 2^n/K$. For any $\eta > 0$, there exists an affine subspace V of co-dimension $d \leq O(\log(K)^4/\eta)$ such that*

$$|(A + B) \cap V| \geq (1 - \eta)|V|.$$

We also need the following lemma from [GOS+11], which shows that low rank boolean functions cannot be too close to constant without actually being constant.

Lemma 4.3 (Theorem 12 in [GOS+11])**.** *Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a function which has at most 2^s nonzero Fourier coefficients. Then all the Fourier coefficients of f are of the form $\frac{a}{2^s}$ where $a \in \mathbb{Z}$. In particular, if $\mathbb{E}[f] < 2^{-s}$ then $f \equiv 0$, and if $\mathbb{E}[f] > 1 - 2^{-s}$ then $f \equiv 1$.*

Definition 4.4 (XOR entropy)**.** *The XOR entropy of a function $f : \mathbb{F}_2^n \to \{0, 1\}$ is the minimum entropy of a partition P of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ corresponding to deterministic XOR protocols for which $f_P = f$. We denote it by $\mathrm{h}^{\oplus}(f)$.*

The following lemma allows for decrease in the XOR entropy.

Lemma 4.5. *Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a non-constant function. Then there exists an affine subspace V of co-dimension $\dim(V) = O(\mathrm{D}^{\oplus}(f)^4)$ such that*

$$\mathrm{h}^{\oplus}(f|_V) \leq \mathrm{h}^{\oplus}(f) - 2^{-25}.$$

Theorem 1.2 follows immediately by applying Lemma 4.5 iteratively, as $\mathrm{D}^{\oplus}(f|V) \leq \mathrm{D}^{\oplus}(f)$ and entropy is never negative, we deduce the existence of a subspace V of co-dimension $O(\mathrm{D}^{\oplus}(f)^5)$ such that $f|V$ is constant, and by Lemma 4.1 we conclude that $\mathrm{pdt}(f) \leq O(\mathrm{D}^{\oplus}(f)^6)$.

Proof. Let $k = \mathrm{D}^{\oplus}(f)$. Assume without loss of generality that $\mathbb{E}[f] \geq 1/2$ (otherwise replace f with $1 - f$). By Lemma 4.3, $\mathbb{E}[f] \leq 1 - 2^{-k}$. Fixing a XOR protocol for f with at most 2^k many rectangles, and considering all the 0-rectangles in it, there must exist a rectangle $A \times B$ such that $f(A + B) = 0$ and $|A \times B| \geq 2^{2n-2k}$. In particular, $|A|, |B| \geq 2^{n-2k}$. Applying Theorem 4.2 to A, B with $K = 2^{2k}, \eta = 1/4$, we deduce the existence of an affine subspace V of co-dimension $O(k^4)$ such that $|(A + B) \cap V| \geq (3/4)|V|$. In particular, $\mathbb{E}[f|V] \leq 1/4$. We may assume that V is a linear subspace, by replacing $f(x)$ with $f(x + v)$ for some $v \in V$. Let W be the dual of V so that $\mathbb{F}_2^n = V + W$. Let P be the partition of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ corresponding to the deterministic XOR protocol for f, so that $h(P) = \mathrm{h}^{\oplus}(f)$. Applying Lemma 3.1 we obtain that

$$\mathbb{E}_{w',w'' \in W}[h(P|_{V,w',w''})] = h_V(P) \leq h(P) - 2^{-25}.$$

In particular, there exists a choice of $w', w'' \in W$ such that

$$h(P|_{(V+w')\times(V+w'')}) \leq h(P) - 2^{-25}.$$

Note that $P|_{(V+w')\times(V+w'')}$ is a deterministic protocol for $f|_{V+w'+w''}$, and hence

$$\mathrm{h}^{\oplus}(f|_{V+w'+w''}) \leq \mathrm{h}^{\oplus}(f) - 2^{-25}.$$

□

5 Open problems

There are two natural open problems which stem directly from our work. The first is whether our result can be extended to randomized protocols vs randomized parity decision trees. Some partial results follow directly from our technique (concretely, a parity decision tree which approximates the function under a product distribution) but the general result still seems to be elusive.

Problem 5.1. *Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a function. Assume that $f_{\oplus}$ has a randomized protocol with complexity k. Does there exist a randomized parity decision tree of depth* $\mathrm{poly}(k)$ *which computes f?*

The second question asks about what happens if we replace XOR with other gadgets. Sherstov [She11] showed that for many gadgets, including some natural 2-bit gadgets, efficient protocols imply low-degree approximating polynomials, which by the work of Nisan and Szegedy [NS94] imply efficient (standard) decision trees. This however does not hold for 1-bit gadgets. Except for XOR functions, the other class of gadgets that can be considered are AND gadgets (any other 1-bit gadget is either trivial or equivalent to either XOR or AND).

That is, for a boolean function $f : \{0, 1\}^n \to \{0, 1\}$ define its corresponding AND function as $f_{\wedge}(x, y) = f(x \wedge y)$, where $\wedge$ is bitwise AND function. An example of an AND function is disjointness. The analog class of decision trees are AND decision trees, where each internal node may query the AND of a subset of the inputs or their negations.

Problem 5.2. *Let $f : \mathbb{F}_2^n \to \{0, 1\}$ be a function. Assume that $f_{\wedge}$ has a deterministic / randomized protocol with complexity k. Does there exist a deterministic / randomized AND decision tree of depth* $\mathrm{poly}(k)$ *which computes f?*

Acknowledgements. We thank Hamed Hatami and Pooya Hatami for early stage discussions about this project.

References

[CŁS13] Ernie Croot, Izabella Łaba, and Olof Sisask. Arithmetic progressions in sumsets and L_p-almost-periodicity. *Combinatorics, Probability and Computing*, 22(03):351–365, 2013.

[CS10] Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geometric and functional analysis*, 20(6):1367–1396, 2010.

[GOS+11] Parikshit Gopalan, Ryan O'Donnell, Rocco A Servedio, Amir Shpilka, and Karl Wimmer. Testing fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011.

[Gre04] Ben Green. Spectral structure of sets of integers. In *Fourier analysis and convexity*, pages 83–96. Springer, 2004.

[LS93] László Lovăsz and Michael Saks. Communication complexity and combinatorial lattice theory. *Journal of Computer and System Sciences*, 47(2):322–349, 1993.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational complexity*, 4(4):301–313, 1994.

[San10] Tom Sanders. Green's sumset problem at density one half. *arXiv preprint arXiv:1003.5649*, 2010.

[San12] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[SS14] Tomasz Schoen and Olof Sisask. Roth's theorem for four variables and additive structures in sums of sparse sets. *arXiv preprint arXiv:1408.2568*, 2014.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 658–667. IEEE, 2013.

[Yao15] Penghui Yao. Parity decision tree complexity and 4-party communication complexity of xor-functions are polynomially equivalent. *arXiv preprint arXiv:1506.02936*, 2015.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de