A composition theorem for parity kill number

Ryan O'Donnell
Carnegie Mellon University

Xiaorui Sun
Columbia University

Li-Yang Tan
Columbia University

John Wright
Carnegie Mellon University

Yu Zhao
Carnegie Mellon University

April 11, 2014

Abstract

In this work, we study the parity complexity measures $\mathsf{C}^{\oplus}_{\min}[f]$ and $\mathsf{DT}^{\oplus}[f]$. $\mathsf{C}^{\oplus}_{\min}[f]$ is the *parity kill number* of f, the fewest number of parities on the input variables one has to fix in order to "kill" f, i.e. to make it constant. $\mathsf{DT}^{\oplus}[f]$ is the depth of the shortest *parity decision tree* which computes f. These complexity measures have in recent years become increasingly important in the fields of communication complexity [ZS09, MO09, ZS10, TWXZ13] and pseudorandomness [BSK12, Sha11, CT13].

Our main result is a composition theorem for $\mathsf{C}^{\oplus}_{\min}$. The k-th power of f, denoted $f^{\circ k}$, is the function which results from composing f with itself k times. We prove that if f is not a parity function, then

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \Omega(\mathsf{C}_{\min}[f]^k).$$

In other words, the parity kill number of f is essentially supermultiplicative in the *normal* kill number of f (also known as the minimum certificate complexity).

As an application of our composition theorem, we show lower bounds on the parity complexity measures of $\mathsf{Sort}^{\circ k}$ and $\mathsf{HI}^{\circ k}$. Here Sort is the sort function due to Ambainis [Amb06], and HI is Kushilevitz's hemi-icosahedron function [NW95]. In doing so, we disprove a conjecture of Montanaro and Osborne [MO09] which had applications to communication complexity and computational learning theory. In addition, we give new lower bounds for conjectures of [MO09, ZS10] and [TWXZ13].

1 Introduction

Recent work on the Log-Rank Conjecture has shown the importance of two related Boolean function complexity measures: sparsity and parity decision tree (PDT) depth. The sparsity of a Boolean function, denoted $\mathsf{sparsity}[\widehat{f}]$, is the number of nonzero coefficients in its Fourier transform. A parity decision tree is a decision tree in which the nodes are allowed to query arbitrary parities of the input variables. The PDT depth of a Boolean function, denoted $\mathsf{DT}^{\oplus}[f]$, is the depth of the shortest PDT which computes f. These two quantities were linked in the papers of [MO09] and [ZS10], both of which posed the following question:

Given a sparse Boolean function, must it have a short parity decision tree?

As a lower bound, any PDT computing f must have depth at least $\frac{1}{2}\log(\mathsf{sparsity}[\widehat{f}])$, and [MO09, ZS10] conjectured that there exists a PDT which is only polynomially worse—depth $\log(\mathsf{sparsity}[\widehat{f}])^k$ for some absolute constant k. Settling this question in the affirmative would prove the Log-Rank Conjecture for an important class of functions known as XOR functions (introduced in [ZS09]). Unfortunately, at present we are very far from deciding this question. The best known upper-bound is $\mathsf{DT}^{\oplus}[f] \leq O\left(\sqrt{\mathsf{sparsity}[\widehat{f}]} \cdot \log(\mathsf{sparsity}[\widehat{f}])\right)$ by [TWXZ13] (see also [STV14, Lov13]), only a square root better than the trivial $\mathsf{DT}^{\oplus}[f] \leq \mathsf{sparsity}[\widehat{f}]$ bound.

A quantity intimately related to $\mathsf{DT}^{\oplus}[f]$ is the *parity kill number* of a Boolean function f, denoted $\mathsf{C}^{\oplus}_{\min}[f]$ (for reasons we will soon explain). This is the fewest number of parities on the input variables one has to fix in order to "kill" f, i.e. to make it constant. There are several equivalent ways to reformulate this definition. Perhaps the most familiar is in terms of *parity certificate complexity*, a generalization of the "normal" certificate complexity measure. Given an input $x \in \mathbb{F}_2^n$, the certificate complexity of f on x is the minimum number of bits x_i one has to read to be certain of the value of $f(x)$. Formally,

$$\mathsf{C}[f,x] := \min\{\operatorname{codim}(C) : C \ni x,\ C \text{ is a subcube on which } f \text{ is constant}\}.$$

We define the minimum certificate complexity of f to be $\mathsf{C}_{\min}[f] := \min_x\{\mathsf{C}[f,x]\}$. This is the minimum number of input bits one has to fix to force f to be a constant. The parity certificate complexity of f on x is defined analogously, as follows:

$$\mathsf{C}^{\oplus}[f,x] := \min\{\operatorname{codim}(H) : H \ni x,\ H \text{ is an affine subspace on which } f \text{ is constant}\},$$

and therefore $\mathsf{C}^{\oplus}_{\min}[f] = \min_x\{\mathsf{C}^{\oplus}[f,x]\}$. We note here that $\mathsf{C}_{\min}[f] \geq \mathsf{C}^{\oplus}_{\min}[f]$ always.

Given a parity decision tree T for f, the parities that T reads on input $x \in \mathbb{F}_2^n$ form a parity certificate for x. As a result, $\mathsf{C}^{\oplus}_{\min}[f]$ lower-bounds the length of any root-to-leaf path in any parity decision tree for f. In particular, $\mathsf{DT}^{\oplus}[f] \geq \mathsf{C}^{\oplus}_{\min}[f]$. Thus, to lower-bound $\mathsf{DT}^{\oplus}[f]$, it suffices to lower-bound $\mathsf{C}^{\oplus}_{\min}[f]$. Remarkably, the reverse is true as well: a recent result by Tsang et al. [TWXZ13] has shown that to *upper*-bound $\mathsf{DT}^{\oplus}[f]$, it suffices to *upper*-bound $\mathsf{C}^{\oplus}_{\min}[f]$[1]. More formally, they showed:

Theorem 1. *Suppose that $\mathsf{C}^{\oplus}_{\min}[f] \leq M[f]$ for all Boolean functions f, where $M[f]$ is some downward non-increasing complexity measure. Then $\mathsf{DT}^{\oplus}[f] \leq M[f] \cdot \log(\mathsf{sparsity}[\widehat{f}])$ for all f.*

[1] A similar argument of translating a best-case bound into a worst-case bound was recently used by Lovett in [Lov13] to show a new upper-bound for the Log-Rank Conjecture. He showed that any total Boolean function with rank r has a communication protocol of complexity $O(\sqrt{r} \cdot \log(r))$

Here by downward non-increasing we mean that $M[f'] \le M[f]$ whenever f' can be derived from f by fixing some parities on the input variables. Theorem 1 implies that to prove the conjecture of [MO09, ZS10], it suffices to show a bound of the form $\mathsf{C}^{\oplus}_{\min}[f] \le \log(\mathsf{sparsity}[\widehat{f}])^k$, for some absolute constant k. This motivates studying the properties of $\mathsf{C}^{\oplus}_{\min}[f]$.

Another area in which parity kill number features prominently is pseudorandomness. A common scenario in this area deals with randomness extraction, in which one has access to a source that outputs mildly random bits, and the goal is to extract from these bits a set of truly random bits. A variety of tools have been developed to accomplish this goal in different settings, one of which is the *affine disperser*. An affine disperser of dimension d is simply a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ with $\mathsf{C}^{\oplus}_{\min}[f] \ge n - d - 1$. Generally, one hopes to design dispersers with low dimension or, equivalently, a high parity kill number. An affine disperser f is “pseudorandom” in the sense that given inputs from a source which is supported on some large enough affine subspace H, f will always be non-constant. Affine dispersers have been constructed with sublinear dimension [BSK12], and the state of the art is a disperser with dimension $n^{o(1)}$ [Sha11]. The study of affine dispersers has gone hand-in-hand with studying the parity kill number of $\mathbb{F}_2$-polynomials; see [CT13] for an example.

Let $\mathsf{DT}[f]$ denote the depth of the shortest decision tree computing f. As $\mathsf{DT}[f]$ is such a simple and well-understood complexity measure, one might hope to carry over intuition, and, when possible, even results, about $\mathsf{DT}[f]$ to the case of $\mathsf{DT}^{\oplus}[f]$. In some cases, this hope has borne fruit: an example is the following theorem from [BTW13], which until recently was only known to hold for decision trees.

Theorem 2. *Let f be a Boolean function. Then $\sum_{i=1}^{n} \widehat{f}(i) \le O(\mathsf{DT}^{\oplus}[f]^{1/2})$.*

Another example is the OSSS inequality for decision trees [OSSS05], which can also be shown to hold for parity decision trees by a straightforward adaptation of the proof of [JZ11]. However, these few instances of similarity appear to be the deceptive minority rather than the majority. On the whole, parity decision trees seem to have a much richer and more counterintuitive structure than normal decision trees, and many questions which are trivial for decision trees become interesting for parity decision trees.

1.1 Boolean function composition and powering

One of the most basic operations one can perform on two Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $g : \mathbb{F}_2^m \to \mathbb{F}_2$, is to *compose* them, producing the new function $f \circ g : \mathbb{F}_2^{m \cdot n} \to \mathbb{F}_2$. On input $y = (y^{(1)}, \cdots, y^{(n)}) \in (\mathbb{F}_2^m)^n$,

$$(f \circ g)(y) := f(g(y^{(1)}), \cdots, g(y^{(n)})).$$

Using this, we can construct the k-th *power* $f^{\circ k}$ of a Boolean function recursively: $f^{\circ 1} := f$, and $f^{\circ k} := f \circ f^{\circ k-1}$. Boolean function powering is a simple tool for generating families of Boolean functions, and it is especially useful in proving lower bounds. It has found application in a variety of areas, from communication complexity [NW95] and Boolean function analysis [OT13] to computational learning theory [Tal13b] and quantum query complexity [HLS07]. For a comprehensive introduction to the subject of Boolean function composition and powering, see [Tal13b].

Decision tree depth is multiplicative with respect to composition and powering: $\mathsf{DT}[f \circ g] = \mathsf{DT}[f] \cdot \mathsf{DT}[g]$, and $\mathsf{DT}[f^{\circ k}] = \mathsf{DT}[f]^k$. On the other hand, $\mathsf{C}_{\min}$ is supermultiplicative: $\mathsf{C}_{\min}[f \circ g] \ge \mathsf{C}_{\min}[f] \cdot \mathsf{C}_{\min}[g]$, and $\mathsf{C}_{\min}[f^{\circ k}] \ge \mathsf{C}_{\min}[f]^k$ (for simple proofs of these facts, see [Tal13b]). How might $\mathsf{DT}^{\oplus}$ and $\mathsf{C}^{\oplus}_{\min}$ behave under composition and powering?

Given arbitrary Boolean functions f and g, consider their composition $f \circ g$. Let us try to construct a small parity certificate for $(f \circ g)(y)$, i.e. a way to fix a small number of parities on the variables in y to make $f \circ g$ constant. To begin, consider a minimum (non-parity) certificate for $f(x_1, \ldots, x_n)$. This certificate consists of a set of coordinates $\mathcal{J} \subseteq [n]$, where $|\mathcal{J}| = \mathsf{C}_{\min}[f]$, and for each $i \in \mathcal{J}$ a fixing $x_i = b_i$, for $b_i \in \mathbb{F}_2$. The guarantee is that if each x_i in $\mathcal{J}$ is set according to this certificate then f is forced to be a constant. Now we will write down a parity certificate for $f \circ g$ which, for each $i \in \mathcal{J}$, fixes $g(y^{(i)})$ to have value b_i. The obvious way to do this is to separately write down the minimum parity certificate for $g(y^{(i)})$ which sets $g(y^{(i)}) = b_i$, for each $i \in \mathcal{J}$. This gives a parity certificate for $f \circ g$ of size at least $\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$; we will call this the *trivial certificate*. Note that if we used this process to construct a parity certificate for $f^{\circ k}$, it would have size at least $\mathsf{C}_{\min}[f]^{(k-1)} \cdot \mathsf{C}^{\oplus}_{\min}[f]$. In particular, the size of the trivial certificate is essentially supermultiplicative in $\mathsf{C}_{\min}[f]$.

Let us consider trying to improve on the trivial certificate for the powered function $f^{\circ k}$. The trivial certificate seems to only weakly use the power of parities. Potentially, significantly shorter certificates could exist which combine the parity certificates for the various $f(y^{(i)})$'s in clever ways. Indeed, depending on the identiy of f, it is sometimes possible to take small "shortcuts" when making the trivial certificate and save on a small number of parities. However, using these shortcuts on $f^{\circ k}$ yields a parity certificate whose size is still essentially supermultiplicative in $\mathsf{C}_{\min}[f]$. Thus, on the whole there isn't an obvious way to improve on the trivial certificate in any substantive way. It is tempting then to conjecture that $\mathsf{C}^{\oplus}_{\min}$ is in fact supermultiplicative in $\mathsf{C}_{\min}$, and if this were true we could prove it by showing optimality of the trivial certificate.

Unfortunately, this intuition does not hold in general. When f is a parity function, $f^{\circ k}$ is also a parity function, for all k. In this case, $\mathsf{C}^{\oplus}_{\min}[f] = 1$ even though $\mathsf{C}_{\min}[f]^{(k-1)} \cdot \mathsf{C}^{\oplus}_{\min}[f]$, the size of the trivial certificate, may be quite large. Our main result is that if we rule out this one pathological case, then $\mathsf{C}^{\oplus}_{\min}[f]$ is indeed supermultiplicative in $\mathsf{C}_{\min}[f]$:

Theorem 3. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function which is not a parity. Then*

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \Omega(\mathsf{C}_{\min}[f]^k).$$

Here, the constant in the $\Omega(\cdot)$ depends on the function f.

Note that as $\mathsf{C}_{\min}[f] \geq \mathsf{C}^{\oplus}_{\min}[f]$, this is a stronger statement than both $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] = \Omega(\mathsf{C}^{\oplus}_{\min}[f]^k)$ and $\mathsf{C}_{\min}[f^{\circ k}] = \Omega(\mathsf{C}_{\min}[f]^k)$. In addition, because $\mathsf{DT}^{\oplus}[f] \geq \mathsf{C}^{\oplus}_{\min}[f]$, this shows that $\mathsf{DT}^{\oplus}[f] \geq \Omega(\mathsf{C}_{\min}[f]^{(k-1)})$. That the constant in the $\Omega(\cdot)$ depends on f follows from the fact that the bound we show is of the form $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq C \cdot \mathsf{C}_{\min}[f]^{(k-1)}$, where $C > 0$ is an absolute constant *independent* of f. The example of the trivial certificate shows that we cannot improve this lower bound to $C \cdot \mathsf{C}_{\min}[f]^k$, where $C > 0$ is independent of f. However, as is typically the case for Boolean function powering, Theorem 3 is sufficient for our applications.

Most of the work in proving Theorem 3 comes from proving the following two-function composition theorem for $\mathsf{C}^{\oplus}_{\min}[f \circ g]$:

Theorem 4. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $g : \mathbb{F}_2^m \to \mathbb{F}_2$ be Boolean functions. If $\mathsf{C}^{\oplus}_{\min}[g] \geq 2$, then*

$$\mathsf{C}^{\oplus}_{\min}[f \circ g] \geq \mathsf{C}^{\oplus}_{\min}[f] + \mathsf{C}_{\min}[f].$$

One oddity of this theorem is that the right-hand side of the inequality does not depend on g (the only dependence on g is in the hypothesis $\mathsf{C}^{\oplus}_{\min}[g] \geq 2$). Though this theorem is sufficient for our applications, it is interesting to consider how tight it might be. The example of the trivial certificate suggests a composition theorem of the form "$\mathsf{C}^{\oplus}_{\min}[f \circ g] \geq \mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$". However, it is possible to construct functions f and g for which $\mathsf{C}^{\oplus}_{\min}[f \circ g] \approx \frac{1}{2}\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$. We discuss this further in Section 6.

After proving Theorem 4, we first prove a stronger version of Theorem 3 in the special case when $\mathsf{C}^{\oplus}_{\min}[f] \geq 2$:

Theorem 5. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function with $\mathsf{C}^{\oplus}_{\min}[f] \geq 2$. Then*

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \frac{\mathsf{C}_{\min}[f]^k - \mathsf{C}_{\min}[f]}{\mathsf{C}_{\min}[f] - 1} + \mathsf{C}^{\oplus}_{\min}[f] \geq \mathsf{C}_{\min}[f]^{(k-1)}.$$

The general theorem then follows from a simple reduction to this case. As we will see, Theorem 5 obtains quantitatively tight bounds for certain functions f.

While Theorems 3 and 5 give a lower bound on $\mathsf{DT}^{\oplus}[f^{\circ k}]$ via the inequality $\mathsf{DT}^{\oplus}[f^{\circ k}] \geq \mathsf{C}^{\oplus}_{\min}[f^{\circ k}]$, sometimes we can get a better lower bound if we know some additional information about f. In this case, we use the following theorem:

Theorem 6. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function satisfying $f(\mathbf{0}) = 0$. If f is not a parity function, then*

$$\mathsf{C}^{\oplus}[f^{\circ k}, \mathbf{0}] \geq \Omega(\mathsf{C}[f, \mathbf{0}]^k).$$

In particular, we note that the LHS of the inequality is a lower bound on $\mathsf{DT}^{\oplus}[f]$. Here, the constant in the $\Omega(\cdot)$ depends on the function f.

1.2 Applications

For our main application of Theorem 3, we disprove one conjecture in communication complexity and show lower bounds for two related conjectures. Let us begin by stating the conjectures. The first we introduced above:

Conjecture 1 ([MO09, ZS10])**.** *For every Boolean function f, $\mathsf{DT}^{\oplus}[f] \leq O(\log(\mathsf{sparsity}[\widehat{f}])^k)$, for some absolute constant k.*

The next conjecture was introduced in [MO09] as a possible means of proving Conjecture 1. It states, roughly, that for any Boolean function f, there is always a parity one can query to "collapse" a large part of f's Fourier transform onto itself.

Conjecture 2 (Montanaro–Osborne)**.** *There exists universal constants $C > 0, K \in [0, 1]$ such that the following holds: for every Boolean function with $\mathsf{sparsity}[\widehat{f}] \geq C$ there exists $\beta \in \mathbb{F}_2^n$ such that*

$$\left|\mathrm{supp}(\widehat{f}) \cap (\mathrm{supp}(\widehat{f}) + \beta)\right| \geq K \cdot \mathsf{sparsity}[\widehat{f}],$$

where $\mathrm{supp}(\widehat{f}) = \{\alpha : \widehat{f}(\alpha) \neq 0\}$, and $\mathrm{supp}(\widehat{f}) + \beta = \{\alpha + \beta : \alpha \in \mathrm{supp}(\widehat{f})\}$.

If this conjecture were true, then one could construct a good parity decision tree for f by always querying the parity associated with the β guaranteed by the conjecture. After $\log(\mathsf{sparsity}[\widehat{f}])$ queries, the restricted function would have constant sparsity. As a result, this conjecture is strong enough to imply Conjecture 1 with $k = 1$, i.e. $\mathsf{DT}^{\oplus}[f] \leq O(\log(\mathsf{sparsity}[\widehat{f}]))$. We remark that Conjecture 1 with $k = 1$ also has implications outside of communication complexity: together with the inequality of Theorem 2 and the Fourier-analytic learning algorithm of [OS07], they imply an efficient algorithm for learning poly(n)-sparse monotone functions from uniform random examples. This would represent a significant advance on a major open problem in learning theory, that of efficiently learning poly(n)-term monotone DNF formulas.

The final conjecture upper bounds $\mathsf{C}_{\min}[f]$ in terms of $\|\widehat{f}\|_1 := \sum_\alpha |\widehat{f}(\alpha)|$ (this is Conjecture 27 in [TWXZ13]):

Conjecture 3 ([TWXZ13])**.** *For every Boolean function f, $\mathsf{C}^{\oplus}_{\min}[f] \leq O(\log(\|\widehat{f}\|_1)^k)$, for some absolute constant k.*

Combined with Theorem 1, this implies Conjecture 1 with exponent $(k+1)$:

$$\mathsf{DT}^{\oplus}[f] \le O(\log(\|\widehat{f}\|_1)^k \cdot \log(\mathsf{sparsity}[\widehat{f}])) \le O(\log(\mathsf{sparsity}[\widehat{f}])^{k+1}),$$

where we have used here the inequality $\|\widehat{f}\|_1 \le \mathsf{sparsity}[\widehat{f}]$. The authors of [TWXZ13] point out that they don't know of a counterexample to Conjecture 3 even in the case of $k = 1$ (which was true also for Conjecture 1).

To prove lower bounds for these conjectures, we consider a pair of functions and the function families generated by powering them. The first of these functions is the Sort function. This function was introduced by Ambainis in [Amb06], in which the family of functions $\mathsf{Sort}^{\circ k}$ was used to provide a separation between polynomial degree and quantum query complexity (see also [LLS06, HLS07]). Applying Theorem 3 to $\mathsf{Sort}^{\circ k}$ yields the following corollary:

Corollary 1.1. *For infinitely many n, there exists a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfying*

$$\mathsf{C}^{\oplus}_{\min}[f] = \Omega((\log(\mathsf{sparsity}[\widehat{f}]))^{\log_2 3}) = \Omega(\log(\|\widehat{f}\|_1)^{\log_2 3}).$$

This example shows that a lower bound of $k \ge \log_2 3 \approx 1.58$ is necessary for Conjecture 3. In fact, by using Theorem 5, we can exactly calculate both $\mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}^{\circ k}]$ and $\mathsf{DT}^{\oplus}[\mathsf{Sort}^{\circ k}]$ (see Section 5 for full details).

The second function we consider is Kushilevitz's hemi-icosahedron function HI. The family of functions $\mathsf{HI}^{\circ k}$ has provided the best known lower bounds for a variety of problems (e.g. [NW95, HKP11]). Applying Theorem 6 to $\mathsf{HI}^{\circ k}$ yields:

Corollary 1.2. *For infinitely many n, there exists a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfying*

$$\mathsf{DT}^{\oplus}[f] = \Omega((\log(\mathsf{sparsity}[\widehat{f}]))^{\log_3 6}).$$

This example shows that a lower bound of $k \ge \log_3 6 \approx 1.63$ is necessary for Conjecture 1. In addition, both Corollaries 1.1 and 1.2 provide examples of functions for which $\mathsf{DT}^{\oplus}[f] = \omega(\log(\mathsf{sparsity}[\widehat{f}]))$, disproving Conjecture 2.

For full details of these functions and the lower bounds, see Section 5. Independent of this work, Noga Ron-Zewi, Amir Shpilka, and Ben Lee Volk have also proven Corollary 1.2 using a family of functions related to $\mathsf{HI}^{\circ k}$ [RZSV13]. With their kind permission, we have reproduced their argument in Appendix A.

1.3 Organization

Section 2 contains definitions and notations. The most technical part of the paper is Section 3, which contains the proof of Theorem 4. Section 4 contains some consequences of Theorem 4, most importantly Theorems 3, 5, and 6. In Section 5, we lower bound the parity complexity measures of $\mathsf{Sort}^{\circ k}$ and $\mathsf{HI}^{\circ k}$, proving Corollaries 1.1 and 1.2. The alternate proof of Corollary 1.2 by Ron-Zewi, Shpilka, and Volk can be found in Appendix A.

2 Preliminaries

2.1 Fourier analysis over the Boolean hypercube

We will be concerned with the Fourier representation of Boolean functions and its relevant complexity measures. In this context it will be convenient to view the output of f as real numbers

$-1, 1 \in \mathbb{R}$ instead of elements of $\mathbb{F}_2$, where we associate $0 \in \mathbb{F}_2$ with $1 \in \mathbb{R}$, and $1 \in \mathbb{F}_2$ with $-1 \in \mathbb{R}$. Throughout this paper we will often switch freely between the two representations.

Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ has a unique representation as a multilinear polynomial

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\chi_\alpha(x) \quad \text{where } \chi_\alpha(x) = (-1)^{\langle x, \alpha \rangle},$$

known as the Fourier transform of f. The numbers $\widehat{f}(\alpha)$ are the Fourier coefficients of f, and we refer to the 2^n functions $\chi_\alpha : \mathbb{F}_2^n \to \{-1, 1\}$ as the Fourier characters. We write $\text{supp}(\widehat{f}) = \{\alpha \in \mathbb{F}_2^n \colon \widehat{f}(\alpha) \neq 0\}$ to denote the support of the Fourier spectrum of f. The *Fourier sparsity* of f, which we denote as $\mathsf{sparsity}[\widehat{f}]$, is the cardinality of its Fourier spectrum $\text{supp}(\widehat{f})$.

The *spectral* 1*-norm of* f is defined to be

$$\|f\|_1 := \sum_{\alpha \in \mathbb{F}_2^n} |\widehat{f}(\alpha)|.$$

For Boolean functions, we have $\mathsf{sparsity}[\widehat{f}] \geq \|f\|_1$.

2.2 Parity complexity measures

In this section, we define some relevant complexity measures. We begin with parity decision tree complexity.

Definition 7 (Parity decision trees)**.** A parity decision tree (PDT) is a binary tree where each internal node is labelled by a subset $\alpha \subseteq [n]$, and each leaf is labelled by a bit $b \in \mathbb{F}_2$. A PDT computes a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ the natural way: on input $x \in \mathbb{F}_2^n$, it computes $\langle x, \alpha \rangle$ where α is the subset at the root. If $\langle x, \alpha \rangle = 1$ the right subtree is recursively evaluated, and if $\langle x, \alpha \rangle = 0$ the left subtree is recursively evaluated. When a leaf is reached the corresponding bit $b \in \mathbb{F}_2$ is the output of the function.

Definition 8 (Parity decision tree complexity)**.** Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. The *parity decision tree complexity of* f, denoted $\mathsf{DT}^{\oplus}[f]$, is the depth of the shallowest parity decision tree computing f.

Definition 9 (Certificate complexity)**.** Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. For every $x \in \mathbb{F}_2^n$, the *certificate complexity* and *parity certificate complexity of* f *at* x are defined to be

$$\begin{aligned} \mathsf{C}[f, x] &:= \min\{\text{codim}(C) \colon C \ni x, \text{ where } C \text{ is a subcube on which } f \text{ is constant}\} \\ \mathsf{C}^{\oplus}[f, x] &:= \min\{\text{codim}(H) \colon H \ni x, \text{ where } H \text{ an affine subspace within which } f \text{ is constant}\}. \end{aligned}$$

The *certificate complexity* and *parity certificate complexity* of f are

$$\mathsf{C}[f] := \max\{\mathsf{C}[f, x] \colon x \in \mathbb{F}_2^n\} \quad \text{and} \quad \mathsf{C}^{\oplus}[f] := \max\{\mathsf{C}^{\oplus}[f, x] \colon x \in \mathbb{F}_2^n\}$$

The *minimum certificate complexity* and *minimum parity certificate complexity* of f are

$$\mathsf{C}_{\min}[f] := \min\{\mathsf{C}[f, x] \colon x \in \mathbb{F}_2^n\} \quad \text{and} \quad \mathsf{C}^{\oplus}_{\min}[f] := \min\{\mathsf{C}^{\oplus}[f, x] \colon x \in \mathbb{F}_2^n\}$$

The complexity measures are related as follows:

Fact 2.1. *The parity complexity measures satisfy* $\mathsf{C}^{\oplus}_{\min}[f] \leq \mathsf{C}^{\oplus}[f] \leq \mathsf{DT}^{\oplus}[f]$ *for every Boolean function* f*.*

Fact 2.2. *For every Boolean function* f *and integer* $k \geq 1$*, we have* $\mathsf{C}_{\min}[f^{\circ k}] \geq \mathsf{C}_{\min}[f]^k$*.*

Fact 2.3. *For every Boolean function* f *and integer* $k \geq 1$*, we have* $\mathsf{C}[f^{\circ k}, \mathbf{0}] \geq \mathsf{C}[f, \mathbf{0}]^k$*.*

Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_d\} \subseteq \mathbb{F}_2^n$ be a linearly independent set of vectors, and $\sigma : \mathcal{B} \to \mathbb{F}_2$. We write $A[\mathcal{B}, \sigma]$ to denote the affine subspace

$$A[\mathcal{B}, \sigma] := \{x \in \mathbb{F}_2^n \colon \langle x, \alpha_i \rangle = \sigma(\alpha_i) \text{ for all } 1 \leq i \leq d\}$$

of co-dimension d. Note that $A[\mathcal{B}, \sigma]$ is a linear subspace if σ is the constant 0 function.

We say that coordinate $i \in [n]$ is *relevant* in an affine subspace H if there is an $x \in \mathbb{F}_2^n$ such that $x \in H$ but $x + \boldsymbol{e}_i \notin H$, and if not we say that i is *irrelevant*.

Proposition 2.4. *Let* $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *be a Boolean function and* $H \subseteq \mathbb{F}_2^n$ *be an affine subspace on which* f *is constant. Then* $\mathsf{C}_{\min}[f]$ *is at most the number of relevant coordinates in* H*.*

Proof. Without loss of generality, suppose coordinates $i \in [k]$ are relevant in H and the others are irrelevant. Fix an arbitrary $x \in H$ and consider

$$C = \{y \in \mathbb{F}_2^n \colon y_i = x_i \text{ for all } i \in [k]\},$$

Note that $C \subseteq H$, since any $y \in C$ differs from x only on the irrelevant coordinates of H. Therefore C is a subcube of co-dimension k on which f is constant, and so $\mathsf{C}_{\min}[f] \leq \mathsf{C}[f, x] \leq k$. □

3 Supermultiplicativity of parity certificate complexity

Theorem 4. *Let* $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *and* $g : \mathbb{F}_2^m \to \mathbb{F}_2$ *be Boolean functions. If* $\mathsf{C}^{\oplus}_{\min}[g] \geq 2$*, then*

$$\mathsf{C}^{\oplus}_{\min}[f \circ g] \geq \mathsf{C}^{\oplus}_{\min}[f] + \mathsf{C}_{\min}[f].$$

Our proof uses the following strategy: given an affine subspace H on which $f \circ g$ is constant, we generate an affine subspace H^* on which f is constant. We do this by removing each g from $f \circ g$ one-by-one. Our key step is in showing that every time we remove a g on the outer layer, if that g was relevant to H, then removing it reduces the codimension of H by at least one. This step we formalize as Proposition 3.1 below.

Proposition 3.1. *Let* $f^* : \mathbb{F}_2^n \times \mathbb{F}_2 \to \mathbb{F}_2$ *and* $g : \mathbb{F}_2^k \to \mathbb{F}_2$ *be Boolean functions where* $\mathsf{C}^{\oplus}_{\min}[g] \geq 2$*. Define* $f : \mathbb{F}_2^n \times \mathbb{F}_2^k \to \mathbb{F}_2$ *to be:*

$$f(x, y) = f^*(x, g(y)).$$

For any affine subspace $H \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^k$ *on which* f *is constant, there exists an affine subspace* $H^* \subseteq \mathbb{F}_2^n \times \mathbb{F}_2$ *on which* f^* *is constant such either:*

1. $\operatorname{codim}(H^*) \leq \operatorname{codim}(H) - 1$*, or*
2. *the* $(n+1)$*-st coordinate is irrelevant in* H^* *and* $\operatorname{codim}(H^*) \leq \operatorname{codim}(H)$*.*

Furthermore, among the first n x*-coordinates, any coordinate that was irrelevant in* H *remains irrelevant in* H^* *as well.*

Proof of Theorem 4 assuming Proposition 3.1. Consider $f \circ g$. Let $H \subseteq \mathbb{F}_2^{n \cdot m}$ be an affine subspace of minimum co-dimension on which $f \circ g$ is constant, and so $\mathrm{codim}(H) = \mathsf{C}^{\oplus}_{\min}[f \circ g]$. Applying Proposition 3.1 to each of the n base functions g that f is composed with, we get an affine subspace $H^* \subseteq \mathbb{F}_2^n$ on which f is constant. Note that the first condition of Proposition 3.1 must hold at least $\mathsf{C}_{\min}[f]$ times in this process of deriving H^* from H, since there are at least $\mathsf{C}_{\min}[f]$ relevant variables in H^* by Proposition 2.4. Therefore

$$\begin{aligned}\mathsf{C}^{\oplus}_{\min}[f] &\leq \mathrm{codim}(H^*)\\ &\leq \mathrm{codim}(H) - \mathsf{C}_{\min}[f]\\ &= \mathsf{C}^{\oplus}_{\min}[f \circ g] - \mathsf{C}_{\min}[f].\end{aligned}$$

Rearranging this inequality completes the proof. □

3.1 Proof of Proposition 3.1

We begin with a pair of technical lemmas.

Lemma 3.2. *Let $g : \mathbb{F}_2^3 \to \mathbb{F}_2$. There exists an affine subspace $H \subseteq \mathbb{F}_2^3$ of codimension at most one such that $g(x) = a_0 \oplus a_1 x_1 \oplus a_2 x_2 \oplus a_3 x_3$ for all $x \in H$, where $a_0, a_1, a_2, a_3 \in \mathbb{F}_2$.*

Proof. Since the only arity-two Boolean functions with $\mathbb{F}_2$-degree two are AND (two-bit conjunction) and OR_2 (two-bit conjunction), we may assume that the restriction of f to any subcube of co-dimension one yields either AND_2 or OR_2. It follows that f must be isomorphic to either

$$\begin{aligned}\mathsf{MAJ}(x_1, x_2, x_3) &= 1 \text{ iff at least two input bits are } 1\\ \mathsf{NAE}(x_1, x_2, x_3) &= 1 \text{ iff } x_1 \neq x_2 \text{ or } x_2 \neq x_3,\end{aligned}$$

both of which satisfy the lemma since they are computed by parity decision trees of depth 2. □

Lemma 3.3. *Let H be an affine subspace of $\mathbb{F}_2^n \times \mathbb{F}_2^k$. There exists an invertible linear transformation $L = L_\ell \otimes L_r$ on $\mathbb{F}_2^n \times \mathbb{F}_2^k$, $\mathcal{B}^* \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^k$, and $\sigma^* : \mathcal{B}^* \to \mathbb{F}_2$ such that $A[\mathcal{B}^*, \sigma^*] = \{Lx \colon x \in H\}$, and $\mathcal{B}^*$ can be partitioned into $\mathcal{B}^* = \mathcal{B}^*_x \sqcup \mathcal{B}^*_y \sqcup \mathcal{B}^*_{x,y}$, where*

- $\mathcal{B}^*_{x,y} = \{(\boldsymbol{e}_i, \boldsymbol{e}_i) \colon 1 \leq i \leq t\}$
- $\mathcal{B}^*_x = \{(\boldsymbol{e}_j, \mathbf{0}) \colon t+1 \leq j \leq t'\}$
- $\mathcal{B}^*_y = \{(\mathbf{0}, \boldsymbol{e}_k) \colon t+1 \leq k \leq t''\}$,

and $t + (t' - t) + (t'' - t) = \mathrm{codim}(H)$.

Proof. Let $H = A[\mathcal{B}, \sigma]$, where $\mathcal{B} = \{(\alpha_1, \beta_1), \ldots, (\alpha_d, \beta_d)\} \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^k$. First, we claim that we may assume without loss of generality that the multisets of vectors

$$\begin{aligned}\mathcal{B}_\ell &= \{\alpha \in \mathbb{F}_2^n - \{\mathbf{0}\} \colon (\alpha, \beta) \in \mathcal{B} \text{ for some } \beta \in \mathbb{F}_2^k\}\\ \mathcal{B}_r &= \{\beta \in \mathbb{F}_2^k - \{\mathbf{0}\} \colon (\alpha, \beta) \in \mathcal{B} \text{ for some } \alpha \in \mathbb{F}_2^n\}\end{aligned}$$

are each linearly independent. Indeed, suppose there exists $\alpha_{i_1}, \ldots, \alpha_{i_k} \in \mathcal{B}_\ell$ such that $\alpha_{i_1} + \ldots + \alpha_{i_k} = \mathbf{0}$ (an identical argument applies for $\mathcal{B}_r$). Since $\mathcal{B}$ is linearly independent, there must exist some $j \in [k]$ such that $\beta_{i_j} \neq \mathbf{0}$. We note that H remains the same if we replace $(\alpha_{i_j}, \beta_{i_j})$ with $(\mathbf{0}, \beta_{i_1} + \ldots + \beta_{i_k})$, and if we set $\sigma^*(\mathbf{0}, \beta_{i_1} + \ldots + \beta_{i_k}) = \sigma(\alpha_{i_1}, \beta_{i_1}) + \ldots + \sigma(\alpha_{i_k}, \beta_{i_k})$. In addition, $\beta_{i_1} + \ldots + \beta_{i_k}$ can be written as a linear combination of the other elements in $\mathcal{B}_r$ if and only if β_{i_j}

can. Therefore, the number of elements in $\mathcal{B}_\ell \cup \mathcal{B}_r$ that can be written as a linear combination of the others decreases by one. Performing this replacement iteratively, the process must eventually terminate with $\mathcal{B}_\ell$ and $\mathcal{B}_r$ both being linearly independent.

When $\mathcal{B}_\ell$ and $\mathcal{B}_r$ are linearly independent, it is straightforward to define invertible linear transformations L_ℓ on $\mathbb{F}_2^n$ mapping $\mathcal{B}_\ell$ to $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{|\mathcal{B}_\ell|}\}$ and L_r on $\mathbb{F}_2^k$ mapping $\mathcal{B}_r$ to $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{|\mathcal{B}_r|}\}$ accordingly, so that the invertible linear transformation L on $\mathbb{F}_2^n \times \mathbb{F}_2^k$ given by

$$L(x, y) = \big((L_\ell^{-1})^T x, (L_r^{-1})^T y\big)$$

satisfies the conditions of the lemma. □

Now we prove Proposition 3.1.

Proof of Proposition 3.1. Let the input variables of $f^* : \mathbb{F}_2^n \times \mathbb{F}_2 \to \mathbb{F}_2$ be $x_1, \ldots, x_n \in \mathbb{F}_2^n$ and $z \in \mathbb{F}_2$, and the input variables of $g : \mathbb{F}_2^k \to \mathbb{F}_2$ be $y_1, \ldots, y_k \in \mathbb{F}_2^k$. By Lemma 3.3, we may assume that $H = A[\mathcal{B}, \sigma]$ where $\mathcal{B} = \mathcal{B}_x \sqcup \mathcal{B}_y \sqcup \mathcal{B}_{x,y}$ and

- $\mathcal{B}_{x,y} = \{(\boldsymbol{e}_i, \boldsymbol{e}_i) \colon 1 \le i \le t\}$
- $\mathcal{B}_x = \{(\boldsymbol{e}_j, \boldsymbol{0}) \colon t + 1 \le j \le t'\}$
- $\mathcal{B}_y = \{(\boldsymbol{0}, \boldsymbol{e}_k) \colon t + 1 \le k \le t''\}$,

and $t + (t' - t) + (t'' - t) = \mathrm{codim}(H)$. Let

$$\begin{aligned} C_x &= \{x \in \mathbb{F}_2^n \colon x_j = \sigma(\boldsymbol{e}_j, \boldsymbol{0}) \text{ for all } t + 1 \le j \le t'\} \\ C_y &= \{y \in \mathbb{F}_2^k \colon y_k = \sigma(\boldsymbol{0}, \boldsymbol{e}_k) \text{ for all } t + 1 \le k \le t''\} \end{aligned}$$

be subcubes of $\mathbb{F}_2^n$ and $\mathbb{F}_2^k$ of co-dimension $|\mathcal{B}_x|$ and $|\mathcal{B}_y|$ respectively. Note that H comprises exactly the pairs $(x, y) \in C_x \times C_y$ satisfying $x_i \oplus y_i = \sigma(\boldsymbol{e}_i, \boldsymbol{e}_i)$ for all $1 \le i \le t$.

3.1.1 Case 1: $|\mathcal{B}_y| \ge 1$ and $|\mathcal{B}_{x,y}| = 0$.

First suppose there exists $b \in \mathbb{F}_2$ such that $g(y) = b$ for all $y \in C_y$; by our assumption on g we have $|\mathcal{B}_y| \ge \mathsf{C}^{\oplus}_{\min}[g] \ge 2$. We claim that f^* is constant on

$$H^* = \{(x, z) \colon x \in C_x \text{ and } z = b\}$$

of co-dimension $|\mathcal{B}_x| + 1 = (|\mathcal{B}| - |\mathcal{B}_y|) + 1 \le |\mathcal{B}| - 1$. Indeed, suppose there exists $(x, b), (x', b) \in H^*$ such that $f^*(x, b) \ne f^*(x', b)$. Then for any $y \in C_y$ we have $(x, y), (x', y) \in H$ and $f(x, y) \ne f(x', y)$.

On the other hand, suppose g is not constant on C_y. In this case we claim that f^* is constant on $H^* = \{(x, z) \colon x \in C_x\}$ of co-dimension $|\mathcal{B}_x| = |\mathcal{B}| - |\mathcal{B}_y| \le |\mathcal{B}| - 1$. Again, suppose there exists $(x, z), (x', z') \in H^*$ such that $f^*(x, z) \ne f^*(x', z')$. Selecting $y, y' \in C_y$ such that $g(y) = z$ and $g(y') = z'$, we get $(x, y), (x', y') \in H$ such that $f(x, y) \ne f(x, y')$.

3.1.2 Case 2: $|\mathcal{B}_y| \ge 1$ and $|\mathcal{B}_{x,y}| \ge 1$.

We define subcubes $C'_x \subseteq C_x$ and $C'_y \subseteq C_y$:

$$\begin{aligned} C'_x &= \{x \in C_x \colon x_i = 0 \text{ for all } 1 \le i \le t - 1\} \\ C'_y &= \{y \in C_y \colon y_i = \sigma(\boldsymbol{e}_i, \boldsymbol{e}_i) \text{ for all } 1 \le i \le t - 1\}. \end{aligned}$$

Note that C'_x has co-dimension $|\mathcal{B}_x| + |\mathcal{B}_{x,y}| - 1 \leq |\mathcal{B}| - 2$. Furthermore, to show that a pair $(x, y) \in C'_x \times C'_y$ falls in H it suffices to ensure $x_t \oplus y_t = \sigma(\boldsymbol{e}_t, \boldsymbol{e}_t)$. We consider two possibilities: (i) there exists $a_0, a_t \in \mathbb{F}_2$ such that $g(y) = a_0 \oplus a_t y_t$ for all $y \in C'_y$, and otherwise (ii) there exists $b \in \mathbb{F}_2$ such that g is non-constant on $C'_y \cap \{y \in \mathbb{F}_2^k : y_t = b\}$.

(i) We claim that f^* is constant on

$$H^* = \{(x, z) : x \in C'_x \text{ and } z = a_0 \oplus a_t(x_t \oplus \sigma(\boldsymbol{e}_t, \boldsymbol{e}_t))\}.$$

of co-dimension $(|\mathcal{B}_x| + |\mathcal{B}_{x,y}| - 1) + 1 \leq |\mathcal{B}| - 1$. Indeed, suppose $f(x, z) \neq f(x', z')$ for some $(x, z), (x', z') \in H^*$. Selecting $y, y' \in C'_y$ such that $y_t = (x_t \oplus \sigma(\boldsymbol{e}_t, \boldsymbol{e}_t)) \oplus a_0$ and $y'_t = (x'_t \oplus \sigma(\boldsymbol{e}_t, \boldsymbol{e}_t)) \oplus a_0$, we get $(x, y), (x', y') \in H$ such that $f(x, y) \neq f(x', y')$.

(ii) In this case we claim that f^* is constant on

$$H^* = \{(x, z) : x \in C'_x \text{ and } x_t = \sigma(\boldsymbol{e}_t, \boldsymbol{e}_t) \oplus b\}.$$

Suppose $f(x, z) \neq f(x', z')$ for some $(x, z), (x', z') \in H^*$. Selecting $y, y' \in C'_y \cap \{y \in \mathbb{F}_2^k : y_t = b\}$ satisfying $g(y) = z$ and $g(y') = z'$, we get $(x, y), (x', y') \in H$ such that $f(x, y) \neq f(x', y')$.

3.1.3 Case 3: $|\mathcal{B}_y| = 0$ and $|\mathcal{B}_{x,y}| \geq 1$.

First suppose there exists $b_1, \ldots, b_t \in \mathbb{F}_2$ such that g is non-constant on the subcube $C'_y = \{y \in \mathbb{F}_2^k : y_i = b_i \text{ for all } 1 \leq i \leq t\}$. In this case we claim that f^* is constant on

$$H^* = \{(x, z) : x \in C_x \text{ and } x_i = \sigma(\boldsymbol{e}_i, \boldsymbol{e}_i) \oplus b_i \text{ for all } 1 \leq i \leq t\}.$$

Indeed, suppose there exists $(x, z), (x', z') \in H^*$ such that $f^*(x, z) \neq f^*(x', z')$. Select $y, y' \in C'_y$ satisfying $g(y) = z$ and $g(y') = z'$, we get $(x, y), (x', y') \in H$ such that $f(x, y) \neq f(x', y')$. Note that although $\operatorname{codim}(H^*)$ may be as large as $|\mathcal{B}|$, we have that H^* is a subcube in $\mathbb{F}_2^n \times \mathbb{F}_2$ where the $(n+1)$-st coordinate is irrelevant, satisfying the second condition of the theorem statement.

Finally, if no such subcube C'_y exists then g is a junta over its first t coordinates. It is straightforward to verify that $t \geq 3$, since every 2-junta has $\mathsf{C}^{\oplus}_{\min}$ at most 1. Consider the sub-function $g' : \mathbb{F}_2^3 \to \mathbb{F}_2$ where $g'(y_1, y_2, y_3) := g(y_1, y_2, y_3, 0, \ldots, 0)$. Applying Lemma 3.2 to g', we get that there exists $\alpha \in \mathbb{F}_2^3 \times \mathbf{0}^{k-3}$ and $a_0, a_1, a_2, a_3, b \in \mathbb{F}_2$ such that

$$g'(y) = a_0 \oplus a_1 y_1 \oplus a_2 y_2 \oplus a_3 y_3 \text{ for all } y \text{ satisfying } \langle y, \alpha \rangle = b. \tag{1}$$

Exactly two elements of $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ form a linearly independent set with α. We suppose without loss of generality that they are $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, and so $\boldsymbol{e}_3 = \alpha + c_1 \boldsymbol{e}_1 + c_2 \boldsymbol{e}_2$ for some $c_1, c_2, \in \mathbb{F}_2$.

We claim that f^* is constant on the affine subspace H^* comprising $(x, z) \in \mathbb{F}_2^n \times \mathbb{F}_2$ satisfying all of the following conditions:

I. $x \in C_x$.

II. $x_i = \sigma(\boldsymbol{e}_i, \boldsymbol{e}_i)$ for all $4 \leq i \leq t$.

III. $x_3 = \sigma(\boldsymbol{e}_3, \boldsymbol{e}_3) \oplus b \oplus c_1(x_1 \oplus \sigma(\boldsymbol{e}_1, \boldsymbol{e}_1)) \oplus c_2(x_2 \oplus \sigma(\boldsymbol{e}_2, \boldsymbol{e}_2))$.

IV. $z = a_0 \oplus a_1(x_1 \oplus \sigma(\boldsymbol{e}_1, \boldsymbol{e}_1)) \oplus a_2(x_2 \oplus \sigma(\boldsymbol{e}_2, \boldsymbol{e}_2)) \oplus a_3(x_3 \oplus \sigma(\boldsymbol{e}_3, \boldsymbol{e}_3))$.

Note that H^* has co-dimension $|\mathcal{B}_x| + (t-3) + 1 + 1 = |\mathcal{B}| - 1$. Once again, suppose $f^*(x,z) \neq f^*(x',z')$ where $(x,z),(x',z') \in H^*$. Selecting $y \in \mathbb{F}_2^3 \times \mathbf{0}^{k-3}$ satisfying

$$y_1 = x_1 \oplus \sigma(\boldsymbol{e}_1,\boldsymbol{e}_1), \quad y_2 = x_2 \oplus \sigma(\boldsymbol{e}_2,\boldsymbol{e}_2), \quad \langle y,\alpha\rangle = b, \tag{2}$$

and likewise y' for x', we claim that $(x,y),(x',y') \in H$ and $f(x,y) \neq f(x',y')$.

We show that $(x,y) \in H$ by checking that $x_i \oplus y_i = \sigma(\boldsymbol{e}_i,\boldsymbol{e}_i)$ for all $1 \leq i \leq t$; the argument for (x',y') is identical. Since $y_i = 0$ for all $i \geq 4$, condition (II) of H^* ensures that $x_i \oplus y_i = \sigma(\boldsymbol{e}_i,\boldsymbol{e}_i)$ for these i's. The conditions (2) on y_1 and y_2 above ensure that $x_i \oplus y_i = \sigma(\boldsymbol{e}_i,\boldsymbol{e}_i)$ for $i \in \{1,2\}$. For $i=3$, we use the fact that

$$\begin{aligned} y_3 &= \langle y,\boldsymbol{e}_3\rangle \\ &= b \oplus c_1 y_1 \oplus c_2 y_2 \\ &= b \oplus c_1(x_1 \oplus \sigma(\boldsymbol{e}_1,\boldsymbol{e}_1)) \oplus c_2(x_2 \oplus \sigma(\boldsymbol{e}_2,\boldsymbol{e}_2)), \end{aligned}$$

and see that condition (III) on H^* in fact ensures $x_3 \oplus y_3 = \sigma(\boldsymbol{e}_3,\boldsymbol{e}_3)$.

To complete the proof it remains to argue that $g(y) = z$; again an identical argument establishes $g(y') = z'$. This follows by combining (1) and (2) with condition (IV) on H^*:

$$\begin{aligned} g(y) = g'(y) &= a_0 \oplus a_1 y_1 \oplus a_2 y_2 \oplus a_3 y_3 \\ &= a_0 \oplus a_1(x_1 \oplus \sigma(\boldsymbol{e}_1,\boldsymbol{e}_1)) \oplus a_2(x_2 \oplus \sigma(\boldsymbol{e}_2,\boldsymbol{e}_2)) \oplus a_3(x_3 \oplus \sigma(\boldsymbol{e}_3,\boldsymbol{e}_3)) \\ &= z. \end{aligned}$$

Here the second equality is by (1), the third by (2), and the final by condition (IV) on H^*. □

Remark 10. It can be checked that in all cases, if H is a linear subspace on which f is constantly 0, then H^* is a linear subspace on which f^* is constantly 0 as well. Therefore, a straightforward modification of the Proof of Theorem 4 using Proposition 3.1 (and Fact 2.3) yields the following incomparable statement:

Theorem 11. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $g : \mathbb{F}_2^m \to \mathbb{F}_2$ be Boolean functions satisfying $f(\mathbf{0}) = g(\mathbf{0}) = 0$. If $\mathsf{C}^{\oplus}_{\min}[g] \geq 2$, then*

$$\mathsf{C}^{\oplus}[f \circ g,\mathbf{0}] \geq \mathsf{C}^{\oplus}[f,\mathbf{0}] + \mathsf{C}[f,\mathbf{0}].$$

In particular, we note that the LHS of the inequality is a lower bound on $\mathsf{DT}^{\oplus}[f \circ g]$.

4 Some consequences of Theorem 4

We will now derive some easy consequences of Theorem 4.

Theorem 5. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function with $\mathsf{C}^{\oplus}_{\min}[f] \geq 2$. Then*

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \frac{\mathsf{C}_{\min}[f]^k - \mathsf{C}_{\min}[f]}{\mathsf{C}_{\min}[f] - 1} + \mathsf{C}^{\oplus}_{\min}[f] \geq \mathsf{C}_{\min}[f]^{(k-1)}.$$

Proof. We apply Theorem 4 and split $f^{\circ k}$ as $f^{\circ k} = f^{\circ k-1} \circ f$.

$$\begin{aligned} \mathsf{C}^{\oplus}_{\min}[f^{\circ k} \circ f] &\geq \mathsf{C}^{\oplus}_{\min}[f^{\circ k-1}] + \mathsf{C}_{\min}[f^{\circ k-1}] \\ &\geq \mathsf{C}^{\oplus}_{\min}[f^{\circ k-1}] + \mathsf{C}_{\min}[f]^{(k-1)}. \end{aligned}$$

Here we have used Theorem 4 for the first inequality and the supermultiplicativity of $\mathsf{C}_{\min}$ (Fact 2.2) for the second. Solving this recurrence completes the proof. □

Next, we have our main theorem.

Theorem 3. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function which is not a parity. Then*

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \Omega(\mathsf{C}_{\min}[f]^k).$$

Here, the constant in the $\Omega(\cdot)$ depends on the function f.

To prove this, we will need the following fact, which is easy to prove:

Fact 4.1. *Suppose $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is not a parity and $\mathsf{C}_{\min}[f] \geq 2$. Then $\mathsf{C}^{\oplus}_{\min}[f \circ f] \geq 2$.*

Using this, we can prove Theorem 3.

Proof of Theorem 3. If $\mathsf{C}_{\min}[f] = 1$, then $\mathsf{C}_{\min}[f]^k = 1$ as well, and so the theorem trivially holds. From now on, we will assume that $\mathsf{C}_{\min}[f] \geq 2$. We may write $f^{\circ k} = f^{\circ(k-2)} \circ (f \circ f)$. By Fact 4.1, $\mathsf{C}^{\oplus}_{\min}[f \circ f] \geq 2$. As a result, we can apply Theorem 4 to show that

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] = \mathsf{C}^{\oplus}_{\min}[f^{\circ(k-2)} \circ (f \circ f)] \geq \mathsf{C}_{\min}[f^{\circ(k-2)}] \geq \mathsf{C}_{\min}[f]^{(k-2)}.$$

This proves Theorem 3 using $\mathsf{C}_{\min}[f]^2$ as the constant in the $\Omega(\cdot)$. □

This proof gives the bound $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \mathsf{C}_{\min}[f]^{(k-2)}$. Though this is sufficient for most (if not all) applications, it is possible to slightly improve on the bound it gives using a more sophisticated argument. At a high level, if we try using the proof of Theorem 5 on a function f for which $\mathsf{C}^{\oplus}_{\min}[f] = 1$, then it is possible when applying Proposition 3.1 to fall into case 1 without actually reducing the codimension of H by one. Whenever this happens, the argument essentially makes no progress, and if this always happens then there's nothing we can say about $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}]$. Fortunately, in the case when f is not a parity function, it is possible to use an amortized-analysis-style argument to show that a constant fraction of the case 1s *do* result in reducing the codimension of H. This allows us to slightly improve on the bound $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \mathsf{C}_{\min}[f]^{(k-2)}$:

Lemma 4.2. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function which is not a parity. Then*

$$\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq C \cdot \mathsf{C}_{\min}[f]^{(k-1)},$$

where $C > 0$ is an absolute constant independent of f.

As the proof of this is more complicated than the proof of Theorem 3, we choose to omit it. We note that by the example of the trivial certificate in Section 1.1, this gives the correct dependence on $\mathsf{C}_{\min}[f]$.

Now we have the issue of performing a similar "bootstrapping" on Theorem 11 to produce Theorem 6. Theorem 11 follows from Theorem 4 by Remark 10. As we are just reusing the proof of Theorem 4 to prove Theorem 3, the same remark holds here. As a result, we have the following theorem.

Theorem 6. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function satisfying $f(\mathbf{0}) = 0$. If f is not a parity function, then*

$$\mathsf{C}^{\oplus}[f^{\circ k}, \mathbf{0}] \geq \Omega(\mathsf{C}[f, \mathbf{0}]^k).$$

In particular, we note that the LHS of the inequality is a lower bound on $\mathsf{DT}^{\oplus}[f]$. Here, the constant in the $\Omega(\cdot)$ depends on the function f.

We end with a remark.

Remark 12. Theorem 3 shows that $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}]$ has nontrivial exponential growth, except in the following cases:

1. f is a parity function.
2. $\mathsf{C}_{\min}[f] = 1$, which has the following two subcases:
   (a) There exists a bit b and an input x_i such that $x_i = b \Rightarrow f(x) = b$.
   (b) There does *not* exist a bit b and an input x_i such that $x_i = b \Rightarrow f(x) = b$.

It is easy to see that in cases 1 and 2a, $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] = 1$ for all k. This is not so clear for case 2b, however. In fact, we can show that in case 2b, $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}]$ has nontrivial exponential growth. To see this, let us assume first that k is even (a similar argument can be made when k is odd), in which case we can write $f^{\circ k} = (f \circ f)^{k/2}$. Now, because we're in case 2b, $\mathsf{C}_{\min}[f \circ f] \geq 2$. Thus, we can apply Theorem 3 to see that $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] \geq \Omega(\mathsf{C}_{\min}[f \circ f]^{k/2}) \geq \Omega(2^{k/2})$. In summary, our results show that for any function f, either $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}] = 1$ for trivial reasons (i.e., f falls in case 1 or 2a), or $\mathsf{C}^{\oplus}_{\min}[f^{\circ k}]$ has nontrivial exponential growth.

5 Lower bounds for specific functions

In this section, we show lower bounds on the parity complexity measures of $\mathsf{Sort}^{\circ k}$ and $\mathsf{HI}^{\circ k}$. Together, these prove Corollaries 1.1 and 1.2.

5.1 The Sort function

The Sort function of Ambainis [Amb06] is defined as follows.

Definition 13. $\mathsf{Sort} : \mathbb{F}_2^4 \to \mathbb{F}_2$ outputs 1 if $x_1 \geq x_2 \geq x_3 \geq x_4$ or $x_1 \leq x_2 \leq x_3 \leq x_4$. Otherwise, $\mathsf{Sort}(x_1, x_2, x_3, x_4) = 0$.

Viewing Sort as a function mapping $\{-1,1\}^4 \to \{-1,1\}$, its Fourier expansion is the degree-2 homogeneous polynomial

$$\mathsf{Sort}(x_1, x_2, x_3, x_4) = \frac{x_1x_2 + x_2x_3 + x_3x_4 - x_4x_1}{2}. \tag{3}$$

It is easy to check that $\mathsf{C}_{\min}[\mathsf{Sort}] = 3$, and so our Theorem 3 implies that $\mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}^{\circ k}] \geq \Omega(3^k)$. To compute the sparsity of $\mathsf{Sort}^{\circ k}$, we first note that Equation 3 gives the recurrence

$$\mathsf{sparsity}[\widehat{\mathsf{Sort}^{\circ k}}] = 4 \cdot \mathsf{sparsity}[\widehat{\mathsf{Sort}^{\circ (k-1)}}]^2.$$

Solving this gives $\mathsf{sparsity}[\widehat{\mathsf{Sort}^{\circ k}}] = 4^{2^k - 1}$. In particular, $\log(\mathsf{sparsity}(\widehat{\mathsf{Sort}^{\circ k}})) = O(2^k)$. Together, these facts imply the first equality in Corollary 1.1.

Corollary 1.1. $\mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}^{\circ k}] = \Omega((\log(\mathsf{sparsity}[\widehat{\mathsf{Sort}^{\circ k}}])^{\log_2 3}) = \Omega(\log(\|\widehat{\mathsf{Sort}^{\circ k}}\|_1)^{\log_2 3})$.

For the second equality, it is easy to check that every nonzero Fourier coefficient of $\mathsf{Sort}^{\circ k}$ has equal weight (up to differences in sign). Thus, $\|\widehat{\mathsf{Sort}^{\circ k}}\|_1 = \sqrt{\mathsf{sparsity}[\widehat{\mathsf{Sort}^{\circ k}}]}$, which gives the second equality.

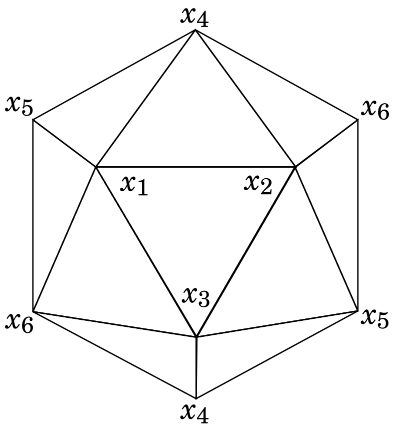
Remark 14. It is also possible to verify that $\mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}] = 2$. Thus, the more refined bound of Theorem 5 shows that

$$\mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}^{\circ k}] \geq \frac{3^k+1}{2},$$

which is matched *exactly* by a parity decision tree for $\mathsf{Sort}^{\circ k}$ of depth $\frac{1}{2}(3^k+1)$. In other words, our analysis shows that $\mathsf{DT}^{\oplus}[\mathsf{Sort}^{\circ k}] = \mathsf{C}^{\oplus}_{\min}[\mathsf{Sort}^{\circ k}] = \frac{1}{2}(3^k+1)$, and in particular, every leaf in the optimal parity decision tree computing $\mathsf{Sort}^{\circ k}$ has maximal depth.

5.2 The HI function

Definition 15. The hemi-icosahedron function $\mathsf{HI} : \mathbb{F}_2^6 \to \mathbb{F}_2$ of Kushilevitz [NW95] is defined as follows: $\mathsf{HI}(x) = 1$ if the Hamming weight $\|x\|$ of x is 1, 2 or 6, and $\mathsf{HI}(x) = 0$ if $\|x\|$ is 0, 4 or 5. Otherwise (i.e. $\|x\| = 3$), $\mathsf{HI}(x) = 1$ if and only if one of the ten facets in the following diagram has all three of its vertices 1:



Viewing HI as a function mapping $\{-1,1\}^6 \to \{-1,1\}$, its Fourier expansion is the degree-3 polynomial

$$\begin{aligned}\mathsf{HI}(x_1,\ldots,x_6) =& \frac{1}{4}\Big(-\sum_i x_i + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_6 + x_1x_4x_5 \\ &+ x_1x_5x_6 + x_2x_3x_5 + x_2x_4x_6 + x_2x_5x_6 + x_3x_4x_5 + x_3x_4x_6\Big).\end{aligned}$$

Because $\mathsf{HI}(\mathbf{0}) = 0$ and $\mathsf{HI}(x) = 1$ for every string x of Hamming weight one, $\mathsf{C}[\mathsf{HI},\mathbf{0}] = 6$. As a result, our Theorem 6 implies that $\mathsf{DT}^{\oplus}[\mathsf{HI}^{\circ k}] \geq \Omega(6^k)$. As for its sparsity, we refer to the following fact.

Fact 5.1. $\mathsf{sparsity}\big(\widehat{\mathsf{HI}^{\circ k}}\big) \leq 4^{3^k}$.

Proof. We will first show that any Boolean function f computed by a degree-d polynomial has sparsity at most 4^d. This is true because any such polynomial is 2^{-d}-granular, meaning that every coefficient is an integer multiple of 2^{-d} (this fact is exercise 12 in chapter 1 of [O'D13]). Finally, by Parseval's equation,

$$1 = \sum_{\alpha\in\mathbb{F}_2^n} \widehat{f}(\alpha)^2 = \sum_{\alpha:\widehat{f}(\alpha)\neq 0} \widehat{f}(\alpha)^2 \geq \mathsf{sparsity}[\widehat{f}] \cdot \left(\frac{1}{2^d}\right)^2.$$

Rearranging, $\mathsf{sparsity}[f] \leq 4^d$.

We saw above that HI is a degree-3 polynomial, so $\mathsf{HI}^{\circ k}$ is a degree-3^k polynomial. This means that $\mathsf{sparsity}\big(\widehat{\mathsf{HI}^{\circ k}}\big) \leq 4^{3^k}$. □

In particular, $\log(\mathsf{sparsity}(\widehat{\mathsf{HI}^{\circ k}})) = O(3^k)$. Putting these facts together, we get Corollary 1.2:

Corollary 1.2. $\mathsf{DT}^{\oplus}[\mathsf{HI}^{\circ k}] = \Omega((\log(\mathsf{sparsity}[\widehat{\mathsf{HI}^{\circ k}}]))^{\log_3 6})$.

6 Future directions

One obvious future direction is to improve the two-function version of our composition theorem given in Theorem 4. As mentioned in Section 1.1, it is possible to construct functions f and g for which $\mathsf{C}^{\oplus}_{\min}[f \circ g] \approx \frac{1}{2}\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$. The construction is as follows: let $f : \mathbb{F}_2^2 \to \mathbb{F}_2$ is the two-bit parity function and $g : \mathbb{F}_2^m \to \mathbb{F}_2$ be a Boolean function (to be chosen later). Then fixing $y_i^{(1)} = y_2^{(2)}$ for each $i \in [m]$ will cause $f(g(y^{(1)}), g(y^{(2)}))$ to be a constant. This shows that $\mathsf{C}^{\oplus}_{\min}[f \circ g] \leq m$. Now, we can choose g to be an affine disperser of sublinear dimension (e.g., [BSK12] or [Sha11]), meaning that $\mathsf{C}^{\oplus}_{\min}[g] = m - o(m)$, in which case $\mathsf{C}^{\oplus}_{\min}[f \circ g] \lessapprox \frac{1}{2}\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$. (This example can be extended to the case when f is the parity over any even number of variables.) To our knowledge, this is the largest known gap between $\mathsf{C}^{\oplus}_{\min}[f \circ g]$ and $\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$, and so it is entirely possible that a composition theorem of the form $\mathsf{C}^{\oplus}_{\min}[f \circ g] \geq \frac{1}{2}\mathsf{C}_{\min}[f] \cdot \mathsf{C}^{\oplus}_{\min}[g]$ is true. Subsequent to our publishing of this paper, Avishay Tal [Tal13a] proved the result, which improves on our Theorem 4 in the case when $\mathsf{C}^{\oplus}_{\min}[g] \geq 9$:

$$\mathsf{C}^{\oplus}_{\min}[f \circ g] \geq \mathsf{C}_{\min}[f] \cdot \left(\log_2\left(\mathsf{C}^{\oplus}_{\min}[g]\right) - 2\right) + \mathsf{C}^{\oplus}_{\min}[f].$$

This is done by an improved analysis of our Case 3 (Section 3.1.3) when g is a junta over its first t coordinates.

Another future direction is to prove a composition theorem for $\mathsf{DT}^{\oplus}$. With respect to function composition, $\mathsf{DT}[f]$ is a more nicely behaved complexity measure than $\mathsf{C}_{\min}[f]$. This is because $\mathsf{DT}[f^{\circ k}] = \mathsf{DT}[f]^k$ exactly, whereas $\mathsf{C}_{\min}[f^{\circ k}]$ is only $\geq \mathsf{C}_{\min}[f]^k$. On the other hand, our paper shows a composition theorem for $\mathsf{C}^{\oplus}_{\min}[f]$ but leaves as an open problem proving a similar composition theorem for $\mathsf{DT}^{\oplus}[f]$. Though our results imply that $\mathsf{DT}^{\oplus}[f]$ is supermultiplicative in $\mathsf{C}_{\min}[f]$, it is trivial to construct functions for which $\mathsf{C}_{\min}[f]$ is small but $\mathsf{DT}^{\oplus}[f]$ is quite large. Thus, a composition theorem for $\mathsf{DT}^{\oplus}[f]$ might prove to be useful.

A final direction, pointed out to us by an anonymous reviewer, is to investigate a weaker form of Conjecture 2 in which the condition on the intersections is replaced with

$$\left|\mathrm{supp}(\widehat{f}) \cap (\mathrm{supp}(\widehat{f}) + \beta)\right| \geq \frac{\mathsf{sparsity}[\widehat{f}]}{\log\left(\mathsf{sparsity}[\widehat{f}]\right)^d},$$

where d is some absolute constant. Though this is weaker than Conjecture 2, it is still strong enough to imply Conjecture 1. The $\mathsf{Sort}^{\circ k}$ function can be used to show that d must be ≥ 1 for this to be true, and it is an interesting open problem whether this lower bound on d can be improved.

Acknowledgments. We would like to thank Rocco Servedio for helpful discussions. We also would like to thank Noga Ron-Zewi, Amir Shpilka, and Ben Lee Volk for allowing us to reproduce their argument in Appendix A.

A Communication complexity proof of Corollary 1.2

In this section, we give the alternate proof of Corollary 1.2 due to Ron-Zewi, Shpilka, and Volk [RZSV13]. Let $\wedge : \mathbb{F}_2^2 \to \mathbb{F}_2$ be the two-bit AND function. The function family they consider is $h_k := \mathsf{HI}^{\circ k} \circ \wedge$. Their lower bound is:

Lemma A.1. $\mathsf{DT}^{\oplus}[h_k] = \Omega((\log(\mathsf{sparsity}[h_k]))^{\log_3 6})$.

Proof. Let us first calculate the sparsity of h_k. As we saw in Section 5.2, $\mathsf{HI}^{\circ k}$ is a degree-3^k polynomial. Because $\wedge$ is a degree-2 polynomial, the degree of h_k is $2 \cdot 3^k$. By a similar argument as in Fact 5.1, this means that $\mathsf{sparsity}[\widehat{h_k}] \leq 4^{2 \cdot 3^k}$. In particular, $\log(\mathsf{sparsity}[\widehat{h_k}]) \leq O(3^k)$.

Now we will show a lower bound on $\mathsf{DT}^{\oplus}[h_k]$. The main facts that we will use about HI are that $\mathsf{HI}(\mathbf{0}) = 0$ and $\mathsf{HI}(x) = 1$ for every string x of Hamming weight one. These imply that $\mathsf{HI}^{\circ k}(\mathbf{0}) = 0$ and $\mathsf{HI}^{\circ k}(x) = 1$ for every string x of Hamming weight one.

Set $n := 6^k$, the number of variables of $\mathsf{HI}^{\circ k}$. Let us group the input variables of h_k into two strings $x, y \in \mathbb{F}_2^n$ and write

$$h_k(x, y) = \mathsf{HI}^{\circ k}(x_1 \wedge y_1, x_2 \wedge y_2, \ldots, x_n \wedge y_n).$$

Consider the communication complexity scenario in which Alice is given x and Bob is given y, and they are asked to compute $h_k(x, y)$. If they had a parity decision tree for h_k of depth d, then they could compute $h_k(x, y)$ using $O(d)$ bits of communication. Define the *intersection size* of x and y to be the number of indices i for which $x_i \wedge y_i = 1$. It is easy to see that computing h_k is equivalent to solving the Set Disjointness problem, at least when x and y are guaranteed to have intersection size 0 or 1 (this follows because $\mathsf{HI}^{\circ k}(\mathbf{0}) = 0$ and $\mathsf{HI}^{\circ k}(x) = 1$ for every string x of Hamming weight one). It is known that even in this special case, Set Disjointness requires $\Omega(n)$ bits of communication [KS92] (see also [Raz92]). As a result, $d = \Omega(n)$, meaning that $\mathsf{DT}^{\oplus}[h_k] = \Omega(6^k)$. Combining this with the above bound of $\log(\mathsf{sparsity}[\widehat{h_k}]) \leq O(3^k)$ yields the lemma. □

References

[Amb06] Andris Ambainis. Polynomial degree vs. quantum query complexity. *Journal of Computer and System Sciences*, 72(2):220–238, 2006.

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

[BTW13] Eric Blais, Li-Yang Tan, and Andrew Wan. Analysis of Boolean functions via information theory. Manuscript, 2013.

[CT13] Gil Cohen and Avishay Tal. Two structural results for low degree polynomials and applications. In *Electronic Colloquium on Computational Complexity TR13-155*, 2013.

[HKP11] Pooya Hatami, Raghav Kulkarni, and Denis Pankratov. Variations on the sensitivity conjecture. *Theory of Computing Library Graduate Surveys*, 3, 2011.

[HLS07] Peter Hoyer, Troy Lee, and Robert Spalek. Negative weights make adversaries stronger. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 526–535, 2007.

[JZ11] Rahul Jain and Shengyu Zhang. The influence lower bound via query elimination. *Theory of Computing*, 7:147–153, 2011.

[KS92] Bala Kalyanasundaram and Georg Schintger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

[LLS06] Sophie Laplante, Troy Lee, and Mario Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15(2):163–196, 2006.

[Lov13] Shachar Lovett. Communication is bounded by root of rank. Technical report, arXiv:1306.1877, 2013.

[MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. Technical report, arXiv:0909.3392, 2009.

[NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.

[O'D13] Ryan O'Donnell. *Analysis of Boolean functions.* 2013.

[OS07] Ryan O'Donnell and Rocco Servedio. Learning monotone decision trees in polynomial time. *SIAM Journal on Computing*, 37(3):827–844, 2007.

[OSSS05] Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco A Servedio. Every decision tree has an influential variable. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 31–39, 2005.

[OT13] Ryan O'Donnell and Li-Yang Tan. A composition theorem for the Fourier Entropy-Influence conjecture. In *Proceedings of the 40th International Colloquium on Automata, Languages and Programming*, pages 780–791, 2013.

[Raz92] Alexander Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[RZSV13] Noga Ron-Zewi, Amir Shpilka, and Ben Lee Volk. Personal communication, 2013.

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

[STV14] Amir Shpilka, Avishay Tal, and Ben Lee Volk. On the structure of Boolean functions with small spectral norm. In *Proceedings of the 5th Innovations in Theoretical Computer Science*, 2014.

[Tal13a] Avishay Tal. Personal communication, 2013.

[Tal13b] Avishay Tal. Properties and applications of Boolean function composition. In *Proceedings of the 4th Innovations in Theoretical Computer Science*, pages 441–454, 2013.

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

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information and Computation*, 9(3&4):255–263, 2009.

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