Improved bounds on Fourier entropy and Min-entropy

Srinivasan Arunachalam* Sourav Chakraborty† Michal Koucký ‡
Nitin Saurabh§ Ronald de Wolf¶

Abstract

Given a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, define the *Fourier distribution* to be the distribution on subsets of $[n]$, where each $S \subseteq [n]$ is sampled with probability $\widehat{f}(S)^2$. The Fourier Entropy-Influence (FEI) conjecture of Friedgut and Kalai [FK96] seeks to relate two fundamental measures associated with the Fourier distribution: does there exist a universal constant $C > 0$ such that $\mathbb{H}(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$, where $\mathbb{H}(\hat{f}^2)$ is the Shannon entropy of the Fourier distribution of f and $\mathsf{Inf}(f)$ is the total influence of f.

In this paper we present three new contributions towards the FEI conjecture:

- We first consider the weaker *Fourier Min-entropy-Influence* (FMEI) conjecture posed by O'Donnell and others [OWZ11, O'D14] which asks if $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$, where $\mathbb{H}_\infty(\hat{f}^2)$ is the min-entropy of the Fourier distribution. We show $\mathbb{H}_\infty(\hat{f}^2) \leq 2\mathsf{C}^{\oplus}_{\min}(f)$, where $\mathsf{C}^{\oplus}_{\min}(f)$ is the minimum parity certificate complexity of f. We also show that for every $\varepsilon \geq 0$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq 2\log(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon))$, where $\|\widehat{f}\|_{1,\varepsilon}$ is the approximate spectral norm of f. As a corollary, we verify the FMEI conjecture for the class of read-k DNFs (for constant k). This improves upon a recent (independent) result of Shalev [Sha18].

- Our second contribution shows that $\mathbb{H}(\hat{f}^2) \leq 2\mathsf{aUC}^{\oplus}(f)$, where $\mathsf{aUC}^{\oplus}(f)$ is the average unambiguous parity certificate complexity of f. This improves upon several bounds shown by Chakraborty et al. [CKLS16].
  An important consequence of resolving the FEI conjecture is the long-standing conjecture of Mansour [Man95]. We show that a weaker question than the FEI conjecture would already imply Mansour's conjecture: is $\mathbb{H}(\hat{f}^2) \leq C \cdot \min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\}$?, where $\mathsf{C}^0(f)$ and $\mathsf{C}^1(f)$ are the zero and one certificate complexities of f respectively.

- Our third contribution is to understand better an implication of the FEI conjecture relating to the structure of polynomials that 1/3-approximate a Boolean function on the Boolean cube. We pose a conjecture: no *flat polynomial* (whose non-zero Fourier coefficients have the same magnitude) of degree d and sparsity $2^{\omega(d)}$ can 1/3-approximate a Boolean function. This conjecture is known to be true assuming FEI and we prove the conjecture unconditionally (i.e., without assuming the FEI conjecture) for a class of polynomials. We finally discuss an intriguing connection between our conjecture and the constant for the Bohnenblust-Hille inequality, which has been extensively studied in functional analysis.

*MIT. Work mostly done when at QuSoft, CWI, Amsterdam, the Netherlands, supported by ERC Consolidator Grant 615307 QPROGRESS. arunacha@mit.edu

†Indian Statistical Institute, Kolkata, India. sourav@isical.ac.in

‡Computer Science Institute of Charles University, Prague, Czech Republic. Partially supported by ERC Consolidator Grant 616787 LBCAD. koucky@iuuk.mff.cuni.cz

§Max Planck Institut für Informatik, Saarland Informatics Campus, Saarbrücken, Germany. Part of the work was done when the author was at IUUK, Prague and supported by the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 616787. Email: nsaurabh@mpi-inf.mpg.de.

¶QuSoft, CWI and University of Amsterdam, the Netherlands. Partially supported by ERC Consolidator Grant 615307 QPROGRESS. rdewolf@cwi.nl

1 Introduction

Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ naturally arise in many areas of theoretical computer science and mathematics such as learning theory, complexity theory, quantum computing, inapproximability, graph theory, extremal combinatorics, etc. Fourier analysis over the Boolean cube $\{-1,1\}^n$ is a powerful technique that has been used often to analyze problems in these areas. For a survey on the subject, see [O'D14, Wol08]. One of the most important and longstanding open problems in this field is the *Fourier Entropy-Influence* (FEI) conjecture, first formulated by Ehud Friedgut and Gil Kalai in 1996 [FK96]. The FEI conjecture seeks to relate the following two fundamental properties of a Boolean function f: the *Fourier entropy* of f and the *total influence* of f, which we define now.

For a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, Parseval's identity relates the *Fourier coefficients* $\{\widehat{f}(S)\}_S$ and the values $\{f(x)\}_x$ by

$$\sum_{S\subseteq[n]} \widehat{f}(S)^2 = \mathbb{E}_x[f(x)^2] = 1,$$

where the expectation is taken uniformly over the Boolean cube $\{-1,1\}^n$. An immediate implication of this equality is that the squared-Fourier coefficients $\{\widehat{f}(S)^2 : S \subseteq [n]\}$ can be viewed as a *probability distribution* over subsets $S \subseteq [n]$, which we often refer to as the *Fourier distribution*. The *Fourier entropy of* f (denoted $\mathbb{H}(\hat{f}^2)$) is then defined as the Shannon entropy of the Fourier distribution, i.e.,

$$\mathbb{H}(\hat{f}^2) := \sum_{S\subseteq[n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}.$$

The *total influence of* f (denoted $\mathsf{Inf}(f)$) measures the *expected size* of a subset $S \subseteq [n]$, where the expectation is taken according to the Fourier distribution, i.e.,

$$\mathsf{Inf}(f) = \sum_{S\subseteq[n]} |S|\ \widehat{f}(S)^2.$$

Intuitively, the Fourier entropy measures how "spread out" the Fourier distribution is over the 2^n subsets of $[n]$ and the total influence measures the concentration of the Fourier distribution on the "high" level coefficients. Informally, the FEI conjecture states that Boolean functions whose Fourier distribution is well "spread out" (i.e., functions with large Fourier entropy) must have significant Fourier weight on the high-degree monomials (i.e., their total influence is large). Formally, the FEI conjecture can be stated as follows:

Conjecture 1 (FEI Conjecture)**.** *There exists a universal constant $C > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$,*

$$\mathbb{H}(\hat{f}^2) \le C \cdot \mathsf{Inf}(f). \tag{1}$$

The original motivation of Friedgut and Kalai for the FEI conjecture came from studying threshold phenomena of monotone graph properties in random graphs [FK96]. For example, resolving the FEI conjecture would imply that every threshold interval of a monotone graph property on n vertices is of length at most $c(\log n)^{-2}$ (for some universal constant $c > 0$). The current best upper bound, proven by Bourgain and Kalai [BK97], is $c_\varepsilon(\log n)^{-2+\varepsilon}$ for every $\varepsilon > 0$.

Besides this application, the FEI conjecture is known to imply the famous Kahn-Kalai-Linial theorem [KKL88] (otherwise referred to as the KKL theorem). The KKL theorem was one of the first major applications of Fourier analysis to understand properties of Boolean functions and has since found its application in various areas of theoretical computer science.

Theorem 2 (KKL theorem). *For every $f : \{-1,1\}^n \to \{-1,1\}$, there exists an $i \in [n]$ such that* $\mathsf{Inf}_i(f) \geq \mathsf{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$.

See Section 2.1 for the definitions of these quantities. We discuss the implication of the FEI conjecture to the KKL theorem in more detail in Section 3. Another motivation to study the FEI conjecture is that a positive answer to this conjecture would resolve the notoriously hard conjecture of Mansour [Man95] from 1995.

Conjecture 3 (Mansour's conjecture). *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is computed by a t-term* DNF.[1] *Then for every $\varepsilon > 0$, there exists $\mathcal{T} \subseteq [n]$ such that $|\mathcal{T}| \leq t^{O(1/\varepsilon)}$ (i.e., $\mathcal{T}$ is polynomial in t) and $\sum_{T \in \mathcal{T}} \widehat{f}(T)^2 \geq 1 - \varepsilon$.*

A positive answer to Mansour's conjecture, along with the query algorithm of Gopalan et al. [GKK08b], would resolve a long-standing open question in computational learning theory of agnostically learning DNFs under the uniform distribution in polynomial time (up to any constant accuracy). We discuss this in more detail in Section 4.

More generally, the FEI conjecture implies that every Boolean function can be approximated (in ℓ_2-norm) by *sparse* polynomials over $\{-1,1\}$. In particular, for a Boolean function f and $\varepsilon > 0$, the FEI conjecture implies the existence of a polynomial p with $2^{O(\mathsf{Inf}(f)/\varepsilon)}$ monomials such that $\mathbb{E}_x[(f(x) - p(x))^2] \leq \varepsilon$. The current best known bound in this direction is $2^{O(\mathsf{Inf}(f)^2/\varepsilon^2)}$, proven by Friedgut [Fri98].[2]

Given the inherent difficulty in answering the FEI conjecture for arbitrary Boolean functions, there have been many recent works studying the conjecture for specific classes of Boolean functions. We give a brief overview of these results in the next section. Alongside the pursuit of resolving the FEI conjecture, O'Donnell and others [OWZ11, O'D14] have asked if a weaker question than the FEI conjecture, the *Fourier Min-entropy-Influence* (FMEI) conjecture can be resolved. The FMEI conjecture asks if the entropy-influence inequality in Eq. (1) holds when the entropy of the Fourier distribution is replaced by the *min-entropy* of the Fourier distribution (denoted $\mathbb{H}_\infty(\hat{f}^2)$). The min-entropy of $\{\widehat{f}(S)^2\}_S$ is defined as

$$\mathbb{H}_\infty(\hat{f}^2) := \min_{\substack{S \subseteq [n]: \\ \widehat{f}(S) \neq 0}} \left\{ \log \frac{1}{\widehat{f}(S)^2} \right\}$$

and thus it is easily seen that $\mathbb{H}_\infty(\hat{f}^2) \leq \mathbb{H}(\hat{f}^2)$. In fact, $\mathbb{H}_\infty(\hat{f}^2)$ could be much smaller compared to $\mathbb{H}(\hat{f}^2)$. For instance, consider the function $f(x) := x_1 \vee \mathsf{IP}(x_1, \ldots, x_n)$, then $\mathbb{H}_\infty(\hat{f}^2) = O(1)$ whereas $\mathbb{H}(\hat{f}^2) = \Omega(n)$. (IP is the inner-product-mod-2 function.) So the FMEI conjecture could be strictly weaker than the FEI conjecture, making it a natural candidate to resolve first.

Conjecture 4 (FMEI Conjecture). *There exists a universal constant $C > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$.*

Another way to formulate the FMEI conjecture is, suppose $f : \{-1,1\}^n \to \{-1,1\}$, then does there exist a Fourier coefficient $\widehat{f}(S)$ such that $|\widehat{f}(S)| \geq 2^{-O(\mathsf{Inf}(f))}$? By the *granularity* of Fourier coefficients it is well-known that *every* Fourier coefficient of a Boolean function f is an integral multiple of $2^{-\mathsf{deg}(f)}$, see [O'D14, Exercise 1.11] for a proof of this. (Here the $\mathsf{deg}(f)$ refers to the degree of the unique multilinear polynomial that represents f.) The FMEI conjecture asks if we

[1] A t-term DNF is a disjunction of at most t conjunctions of variables and their negations.

[2] Friedgut's Junta theorem says that f is ε-close to a junta on $2^{O(\mathsf{Inf}(f)/\varepsilon)}$ variables. We refer to [O'D14, Section 9.6, page 269, Friedgut's Junta Theorem] for details.

can prove a lower bound of $2^{-O(\mathsf{Inf}(f))}$ on *any one* Fourier coefficient and even this remains open. Proving the FMEI conjecture seems to require proving interesting structural properties of Boolean functions. In fact, as observed by [OWZ11], the FMEI conjecture suffices to imply the KKL theorem (see also Section 3).

Understanding the min-entropy of a Fourier distribution is important in its own right too. It was observed by Akavia et al. [ABG+14] that for a circuit class $\mathcal{C}$, tighter relations between min-entropy of $f \in \mathcal{C}$ and f_A defined as $f_A(x) := f(Ax)$, for an arbitrary linear transformation A, could enable us to translate lower bounds against the class $\mathcal{C}$ to the class $\mathcal{C} \circ \mathsf{MOD}_2$. In particular, they conjectured that min-entropy of f_A is only polynomially larger than f when $f \in \mathsf{AC}^0[\mathsf{poly}(n), O(1)]$. ($\mathsf{AC}^0[s, d]$ is the class of unbounded fan-in circuits of size at most s and depth at most d.) It is well-known that when $f \in \mathsf{AC}^0[s, d]$, $\mathbb{H}_\infty(\hat{f}^2)$ is at most $O((\log s)^{d-1} \cdot \log\log s)$ [LMN93, Bop97, Tal17]. Depending on the tightness of the relationship between $\mathbb{H}_\infty(\hat{f}^2)$ and $\mathbb{H}_\infty(\widehat{f_A}^2)$, one could obtain near-optimal lower bound on the size of $\mathsf{AC}^0[s, d] \circ \mathsf{MOD}_2$ circuits computing IP (inner-product-mod-2). This problem has garnered a lot of attention in recent times for a variety of reasons [SV10, SV12, ABG+14, CS16, CGJ+18]. The current best known lower bound for IP against $\mathsf{AC}^0[s, d] \circ \mathsf{MOD}_2$ is quadratic when $d = 4$, and only super-linear for all $d = O(1)$ [CGJ+18].

In the remaining part of this introduction, we first give a brief overview of prior work on the FEI conjecture in Section 1.1 and then describe our contributions in Section 1.2.

1.1 Prior Work

After Friedgut and Kalai [FK96] posed the FEI conjecture in 1996, there was not much work done towards resolving it, until the work of Klivans et al. [KLW10] in 2010. They showed that the FEI conjecture holds true for random DNF formulas. Since then, there have been many significant steps taken in the direction of resolving the FEI conjecture. We review some recent works here, referring the interested reader to the blog post of Kalai [Kal07] for additional discussions on the FEI conjecture.

The FEI conjecture is known to be true when we replace the universal constant C with $\log n$ in Eq. (1). In fact we know $\mathbb{H}(\hat{f}^2) \leq O(\mathsf{Inf}(f) \cdot \log n)$ for *real-valued* functions $f : \{-1, 1\}^n \to \mathbb{R}$ (see [OWZ11, KMS12] for a proof and [CKLS16] for an improvement of this statement).[3] If we strictly require C to be a universal constant, then the FEI conjecture is known to be false for real-valued functions. Instead, for real-valued functions an analogous statement called the *logarithmic Sobolev Inequality* [Gro75] is known to be true. The logarithmic Sobolev inequality states that for every $f : \{-1, 1\}^n \to \mathbb{R}$, we have $\mathsf{Ent}(f^2) \leq 2 \cdot \mathsf{Inf}(f)$, where $\mathsf{Ent}(f)$ is defined as $\mathsf{Ent}(f^2) = \mathbb{E}[f^2 \ln(f^2)] - \mathbb{E}[f^2] \ln(\mathbb{E}[f^2])$, where the expectation is taken over uniform $x \in \{-1, 1\}^n$.

Restricting to Boolean functions, the FEI conjecture is known to be true for the "standard" functions that arise often in analysis, such as AND, OR, Majority, Parity, Bent functions and Tribes. There have been many works on proving the FEI conjecture for specific classes of Boolean functions. O'Donnell et al. [OWZ11] showed that the FEI conjecture holds for symmetric Boolean functions and read-once decision trees. Keller et al. [KMS12] studied a generalization of the FEI conjecture when the Fourier coefficients are defined on biased product measures on the Boolean cube. Then, Chakraborty et al. [CKLS16] and O'Donnell and Tan [OT13], independently and simultaneously, proved the FEI conjecture for read-once formulas. In fact, O'Donnell and Tan proved an interesting composition theorem for the FEI conjecture (we omit the definition of composition theorem here, see [OT13] for more). For general Boolean functions, Chakraborty et al. [CKLS16] gave several

[3] For Boolean functions, the $\log n$-factor was improved by [GSTW16] to $\log(\mathsf{s}(f))$ (where $\mathsf{s}(f)$ is the sensitivity of the Boolean function f).

upper bounds on the Fourier entropy in terms of combinatorial quantities larger than the total influence, e.g., average parity-decision tree depth, etc.

Later Wan et al. [WWW14] used Shannon's source coding theorem [Sha48] (which characterizes entropy) to establish the FEI conjecture for read-k decision trees for constant k. Recently, Shalev [Sha18] improved the constant in the FEI inequality for read-k decision trees, and further verifies the conjecture when either the influence is *too low*, or the entropy is *too high*. The FEI conjecture is also verified for random Boolean functions by Das et al. [DPV11] and for random linear threshold functions (LTFs) by Chakraborty et al. [CKK+18].

There has also been some work in giving lower bounds on the constant C in the FEI conjecture. Hod [Hod17] gave a lower bound of $C > 6.45$ (the lower bound holds even when considering the class of monotone functions), improving upon the lower bound of O'Donnell and Tan [OT13].

However, there has not been much work on the FMEI conjecture. It was observed in [OWZ11, CKK+18] that the KKL theorem implies the FMEI conjecture for monotone functions and linear threshold functions. Finally, the FMEI conjecture for "regular" read-k DNFs was recently established by Shalev [Sha18].

1.2 Our Contributions

Our contributions in this paper are threefold, which we summarize below:

1.2.1 New upper bounds for the FMEI conjecture.

The FMEI conjecture is much less understood than the FEI conjecture. In fact, we are aware of only one very recent paper [Sha18] which studies the FMEI conjecture for a particular class of functions. Our first contribution is to give upper bounds on the min-entropy of general Boolean functions in terms of the minimum parity-certificate complexity (denoted $\mathsf{C}^{\oplus}_{\min}(f)$) and the approximate spectral norm of Boolean functions (denoted $\|\widehat{f}\|_{1,\varepsilon}$). The minimum parity-certificate complexity $\mathsf{C}^{\oplus}_{\min}(f)$ is also referred to as the parity kill number by O'Donnell et al. [OWZ+14] and is closely related to the communication complexity of XOR functions [ZS09, MO09, TWXZ13]. The approximate spectral norm $\|\widehat{f}\|_{1,\varepsilon}$ is related to the *quantum* communication complexity of XOR functions [LS09, Zha14]. In particular, it characterizes the bounded-error quantum communication complexity of XOR functions with constant $\mathbb{F}_2$-degree [Zha14]. (By $\mathbb{F}_2$-degree, we mean the degree of a function when viewed as a polynomial over $\mathbb{F}_2$.)

Theorem 5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then,*

1. *for every $\varepsilon \geq 0$, $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \log\left(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon)\right)$.*
2. *$\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \mathsf{C}^{\oplus}_{\min}(f)$.*
3. *$\mathbb{H}_\infty(\hat{f}^2) \leq 2(1+\log_2 3) \cdot R^{\oplus}_2(f)$.*

The proof of Theorem 5(1) expresses the quantity $\|\widehat{f}\|_{1,\varepsilon}$ as a (minimization) linear program. We consider the dual linear program and exhibit a feasible solution that achieves an optimum of $(1-\varepsilon)/\max_S |\widehat{f}(S)|$. This proves the desired inequality. In order to prove part (2) and (3) of the theorem, the idea is to consider a "simple" function g that has "good" correlation with f, and then upper bound the correlation between f and g using Plancherel's theorem (Fact 13) and the fact that g has a "simple" Fourier structure. For part (2), g is chosen to be the indicator function of an (affine) subspace where f is constant, whereas for part(3) the randomized parity decision tree computing f itself plays the role of g.

As a corollary (Corollary 20) of this theorem we also give upper bounds on the Rényi Fourier entropy of order $1+\delta$ (denoted $\mathbb{H}_{1+\delta}(\hat{f}^2)$) for all $\delta > 0$. Note that $\mathbb{H}_{1+\delta}(\hat{f}^2) \geq \mathbb{H}_\infty(\hat{f}^2)$ for every $\delta \geq 0$ and as $\delta \to \infty$, $\mathbb{H}_{1+\delta}(\hat{f}^2)$ converges to $\mathbb{H}_\infty(\hat{f}^2)$. Note that $\mathbb{H}_1(\hat{f}^2)$ is the standard Shannon entropy of the Fourier distribution.

We believe that these improved bounds on min-entropy of the Fourier distribution give a better understanding of Fourier coefficients of Boolean functions, and could be of independent interest. As a somewhat non-trivial application of Theorem 5 (in particular, part (2)) we verify the FMEI conjecture for read-k DNFs, for constant k. (A read-k DNF is a formula where each variable appears in at most k terms.)

Theorem 6. *For every Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ that can be expressed as a read-k DNF, we have*

$$\mathbb{H}_\infty(\hat{f}^2) \;\leq\; O(\log k) \cdot \mathsf{Inf}(f).$$

This theorem improves upon a recent (and independent) result of Shalev [Sha18] that establishes the FMEI conjecture for "regular" read-k DNFs (where regular means each term in the DNF has more or less the same number of variables, see [Sha18] for a precise definition). In order to prove Theorem 6, we essentially show that $\mathsf{Inf}(f)$ is at least $\mathsf{C}_{\min}(f)$, for read-k DNFs (Lemma 22), where $\mathsf{C}_{\min}(f)$ is the minimum certificate complexity of f. Now the proof of Theorem 6 follows in conjunction with Theorem 5(2).

1.2.2 Better upper bounds for the FEI conjecture.

Our second contribution is to give a better upper bound on the Fourier entropy $\mathbb{H}(\hat{f}^2)$ in terms of the *average unambiguous certificate complexity* of f (which we denote by $\mathsf{aUC}(f)$).

Informally, the unambiguous certificate complexity of f, denoted $\mathsf{UC}(f)$, is similar to the standard certificate complexity measure, except that the collection of certificates are now required to be *unambiguous*, i.e., every input should be consistent with a *unique* certificate. In other words, an unambiguous certificate is a monochromatic subcube partition of the Boolean cube. By the average unambiguous certificate complexity we mean the expected number of bits set by an unambiguous certificate on a uniformly random input. For formal definitions, we refer to Section 2. There have been many recent works on query complexity, giving upper and lower bounds on $\mathsf{UC}(f)$ in terms of other combinatorial measures such as decision-tree complexity, sensitivity, quantum query complexity, etc., see [Göö15, AKK16, BHT17] for more. Our main contribution here is an upper bound on $\mathbb{H}(\hat{f}^2)$ that improves upon all previously known upper bounds for the FEI conjecture in terms of combinatorial measures (see Section 1.1 for a discussion on these results).

Theorem 7. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then,*

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \mathsf{aUC}(f).$$

A new and crucial ingredient employed in the proof of the theorem is an analog of the law of large numbers in information theory, usually referred to as the *Asymptotic Equipartition Property (AEP)*. Employing information-theoretic techniques for the FEI conjecture seems very natural given that the conjecture seeks to bound the entropy of a distribution. Indeed, Keller et al. [KMS12, Section 3.1] envisioned a proof of the FEI conjecture itself using large deviation estimates and the tensor structure (explained below) in a stronger way, and Wan et al. [WWW14] used Shannon's source coding theorem [Sha48] to verify the conjecture for bounded-read decision trees.

In order to prove Theorem 7, we study the *tensorized* version of f, $f^M\colon \{-1,1\}^{Mn} \to \{-1,1\}$, which is defined as follows,

$$f^M(x^1,\dots,x^M) := f(x_1^1,\dots,x_n^1)\cdot f(x_1^2,\dots,x_n^2)\cdot f(x_1^M,\dots,x_n^M).$$

Similarly we define a *tensorized* version $\mathcal{C}^M$ of an unambiguous certificate $\mathcal{C}$ of f,[4] i.e., a direct product of M independent copies of $\mathcal{C}$. It is not hard to see that $\mathcal{C}^M$ is also an unambiguous certificate of f^M. To understand the properties of $\mathcal{C}^M$ we study $\mathcal{C}$ in a probabilistic manner. We observe that $\mathcal{C}$ naturally inherits a distribution $\mathbf{C}$ on its certificates when the underlying inputs $x \in \{-1,1\}^n$ are distributed uniformly. Using the asymptotic equipartition property with respect to $\mathbf{C}$, we infer that for every $\delta > 0$, there exists $M_0 > 0$ such that for all $M \geq M_0$, there are at most $2^{M(\mathsf{aUC}(f,\mathcal{C})+\delta)}$ certificates in $\mathcal{C}^M$ that together cover at least $1-\delta$ fraction of the inputs in $\{-1,1\}^{Mn}$. Furthermore, each of these certificates fixes at most $M(\mathsf{aUC}(f,\mathcal{C})+\delta)$ bits. Hence, a particular certificate can contribute to at most $2^{M(\mathsf{aUC}(f,\mathcal{C})+\delta)}$ Fourier coefficients of f^M. Combining both these bounds, all these certificates can overall contribute to at most $2^{2M(\mathsf{aUC}(f,\mathcal{C})+\delta)}$ Fourier coefficients of f^M. Let's denote this set of Fourier coefficients by $\mathcal{B}$. We then argue that the Fourier coefficients of f^M that are *not* in $\mathcal{B}$ have Fourier weight at most δ. This now allows us to bound the Fourier entropy of f^M as follows,

$$\mathbb{H}(\widehat{f^M}^2) \leq \log|\mathcal{B}| + \delta nM + \mathsf{H}(\delta),$$

where $\mathsf{H}(\delta)$ is the binary entropy function. Since $\mathbb{H}(\widehat{f^M}^2) = M\cdot\mathbb{H}(\hat{f}^2)$, we have

$$\mathbb{H}(\hat{f}^2) \leq 2(\mathsf{aUC}(f,\mathcal{C})+\delta) + \delta n + \frac{\mathsf{H}(\delta)}{M}.$$

By the AEP theorem, note that $\delta \to 0$ as $M \to \infty$. Thus, taking the limit as $M \to \infty$ we obtain our theorem.

Looking very finely into how certificates contribute to Fourier coefficients in the proof above, we further strengthen Theorem 7 by showing that we can replace $\mathsf{aUC}(f)$ by the *average unambiguous parity-certificate complexity* $\mathsf{aUC}^{\oplus}(f)$ of f. Here $\mathsf{aUC}^{\oplus}(f)$ is defined similar to $\mathsf{aUC}(f)$ except that instead of being defined in terms of monochromatic subcube partitions of f, we now partition the Boolean cube with monochromatic *affine subspaces*. (Observe that subcubes are also affine subspaces.)

Theorem 8. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any Boolean function. Then,*

$$\mathbb{H}(\hat{f}^2) \leq 2\cdot\mathsf{aUC}^{\oplus}(f).$$

The proof outline remains the same as in Theorem 7. However, a particular certificate in $\mathcal{C}^M$ *no longer* just fixes variables. Instead these parity certificates now fix parities over variables, and so potentially could involve all variables. Hence we cannot directly argue that all the certificates contribute to at most $2^{M(\mathsf{aUC}^{\oplus}(f,\mathcal{C})+\delta)}$ Fourier coefficients of f^M. Nevertheless, by the AEP theorem we still obtain that a typical parity-certificate fixes at most $M(\mathsf{aUC}^{\oplus}(f,\mathcal{C})+\delta)$ parities. Looking closely at the Fourier coefficients that a parity-certificate can contribute to, we now argue that such coefficients must lie in the linear span of the parities fixed by the parity-certificate. Therefore, a typical parity-certificate can overall contribute to at most $2^{M(\mathsf{aUC}^{\oplus}(f,\mathcal{C})+\delta)}$ Fourier coefficients of f^M. The rest of the proof now follows analogously.

[4] Recall an unambiguous certificate is a collection of certificates that partitions the Boolean cube $\{-1,1\}^n$.

Regarding Mansour's conjecture. We complement our improved bounds on $\mathbb{H}(\hat{f}^2)$ by further arguing that Theorem 8 seems to bring us closer to proving a version of Mansour's conjecture (Conjecture 3). Mansour's conjecture is an important motivation for proving the FEI conjecture, for more details on its significance, we refer to the discussion in Section 4.1. Here we observe that in fact a weaker statement than the FEI conjecture suffices to imply Mansour's conjecture. We note that it was implicit in previous works [Kal07, GKK08a, KLW10, OWZ11] that the full power of the FEI conjecture is not required to establish Mansour's conjecture. We believe our observation sharpens this relationship and shows that resolving Mansour's conjecture is a natural next step after our result towards resolving the FEI conjecture.

Lemma 9. *Suppose there exists a universal constant $\lambda > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\}.$$

Then, Mansour's conjecture (Conjecture 3) is true.

We note that $\min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\}$ could be exponentially smaller than $\mathsf{aUC}(f)$. This is witnessed by a read-once DNF over n variables where each term has *strictly less* than $\log n$ variables. For a proof, see [AHKÜ17, Theorem 6]. However, it is not clear whether $\min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\} \leq \mathsf{aUC}(f)$ holds for all Boolean functions.

1.2.3 Implications of the FEI conjecture and connections to the Bohnenblust-Hille inequality.

Our final contribution is to understand better the structure of polynomials that ε-approximate Boolean functions on the Boolean cube. To be more specific, for simplicity we fix ε to be $1/3$ and we consider polynomials p such that $|p(x) - f(x)| \leq 1/3$ for all $x \in \{-1,1\}^n$, where f is a Boolean function. Such polynomials have proved to be powerful and found diverse applications in theoretical computer science. The single most important measure associated with such polynomials is its *degree*. The *least* degree of a polynomial that $1/3$-approximates f is referred to as the *approximate degree* of f. Tight bounds on approximate degree have both algorithmic and complexity-theoretic implications, see for instance Sherstov's recent paper [She18] and references therein.

In this work we ask, suppose the FEI conjecture were true, what can be said about approximating polynomials? For instance, are these approximating polynomials p sparse in their Fourier domain, i.e., is the number of monomials in p, $|\{S : \widehat{p}(S) \neq 0\}|$, small? Do approximating polynomials have small spectral norm (i.e., small $\sum_S |\widehat{p}(S)|$)? In order to understand these questions better, we restrict ourselves to a class of polynomials called *flat* polynomials over $\{-1,1\}$, i.e., polynomials whose non-zero Fourier coefficients have the same magnitude.

We first observe that if a flat polynomial p $1/3$-approximates a Boolean function f, then the entropy of the Fourier distribution of f must be "large". In particular, we show that $\mathbb{H}(\hat{f}^2)$ must be at least as large as the logarithm of the Fourier sparsity of p (Claim 5.1). It then follows that assuming the FEI conjecture, a flat polynomial of degree d and sparsity $2^{\omega(d)}$ cannot $1/3$-approximate a Boolean function (Lemma 36). However, it is not clear to us how to obtain the same conclusion *unconditionally* (i.e., without assuming that the FEI conjecture is true) and, so we pose the following conjecture.

Conjecture 10. *No flat polynomial of degree d and sparsity $2^{\omega(d)}$ can $1/3$-approximate a Boolean function.*[5]

Since we could not solve the problem as posed above, we make progress in understanding this conjecture by further restricting ourselves to the class of *block-multilinear* polynomials. An n-variate polynomial is said to be *block-multilinear* if the input variables can be *partitioned* into disjoint blocks such that every monomial in the polynomial has *at most* one variable from each block. Such polynomials have been well-studied in functional analysis since the work of Bohnenblust and Hille [BH31], but more recently have found applications in quantum computing [AA18, Mon12], classical and quantum XOR games [BBLV13], and polynomial decoupling [OZ16]. In the functional analysis literature block-multilinear polynomials are known as *multilinear forms*. In an ingenious work [BH31], Bohnenblust and Hille showed that for every degree-d multilinear form $p : (\mathbb{R}^n)^d \to \mathbb{R}$, we have

$$\Big(\sum_{i_1,\ldots,i_d=1}^{n} |\widehat{p}_{i_1,\ldots,i_d}|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} \leq C_d \cdot \max_{x^1,\ldots,x^d \in [-1,1]^n} |p(x^1,\ldots,x^d)|, \tag{2}$$

where C_d is a constant that depends on d. In [BH31], they showed that it suffices to pick C_d to be exponential in d to satisfy the equation above. For $d = 2$, Eq. (2) generalizes Littlewood's famous $4/3$-inequality [Lit30]. Eq. (2) is commonly referred to as the Bohnenblust-Hille (BH) inequality and is known to have deep applications in various fields of analysis such as operator theory, complex analysis, etc. There has been a long line of work on improving the constant C_d in the BH inequality (to mention a few [DPS10, DFOC+11, DMFPSS12, ABPSS14, PE18]). The best known upper bound on C_d (we are aware of) is polynomial in d. It is also conjectured that it suffices to let C_d be a *universal* constant (independent of d) in order to satisfy Eq. (2).

In our context, using the best known bound on C_d in the BH-inequality implies that a flat block-multilinear polynomial of degree d and sparsity $2^{\omega(d \log d)}$ cannot $1/3$-approximate a Boolean function. However, from the discussion before Conjecture 10, we know that the FEI conjecture implies the following theorem.

Theorem 11. *If p is a flat block-multilinear polynomial of degree d and sparsity $2^{\omega(d)}$, then p cannot $1/8$-approximate a Boolean function.*

Moreover, the above theorem is also implied when the BH-constant C_d is assumed to be a universal constant. Our main contribution is to establish the above theorem *unconditionally*, i.e., neither assuming C_d is a universal constant nor assuming the FEI conjecture. In order to show the theorem, we show an inherent weakness of block-multilinear polynomials in approximating Boolean functions. More formally, we show the following.

Lemma 12. *Let p be a block-multilinear polynomial of degree-d that $1/8$-approximates a Boolean function f. Then, $\mathsf{deg}(f) \leq d$.*

Organization. In the remainder of the paper, we prove and elaborate on each of these results in more detail. In Section 3, we establish new upper bounds on Fourier min-entropy, and verify the FMEI conjecture for read-k DNFs. In Section 4, we give improved upper bounds on Fourier entropy, verify the FEI conjecture for functions with bounded average unambiguous certificate complexity,

[5] We remark that there exists a degree-d flat *Boolean* functions of sparsity 2^d. One simple example on 4 bits is the function $x_1(x_2+x_3)/2+x_4(x_2-x_3)/2$. By taking a $(d/2)$-fold product of this Boolean function on disjoint variables, we obtain our remark.

and elaborate on the connections between Fourier entropy vs. certificate complexity and Mansour's conjecture. In Section 5, we pose a conjecture which is a consequence of the FEI conjecture and make partial progress towards its resolution. We further discuss an intriguing connection between our conjecture and the constants in the Bohnenblust-Hille inequality. Finally in Section 6, we conclude with some open problems.

2 Preliminaries

Notation. We denote the set $\{1, 2, \ldots, n\}$ by $[n]$. A *partial assignment* of $[n]$ is a map $\tau : [n] \to \{-1, 1, *\}$. Define $|\tau| = |\tau^{-1}(1) \cup \tau^{-1}(-1)|$. A subcube of the Boolean cube $\{-1,1\}^n$ is a set of $x \in \{-1,1\}^n$ that agrees with some partial assignment τ, i.e., $\{x \in \{-1,1\}^n : \tau(i) = x_i \text{ for every } \tau(i) \neq *\}$.

2.1 Fourier Analysis.

We recall some definitions and basic facts from analysis of Boolean functions, referring to [O'D14, Wol08] for more. Consider the space of all functions from $\{-1,1\}^n$ to $\mathbb{R}$ equipped with the inner product defined as

$$\langle f, g \rangle \coloneqq \mathbb{E}_x[f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

For $S \subseteq [n]$, the character function $\chi_S : \{-1,1\}^n \to \{-1,1\}$ is defined as $\chi_S(x) \coloneqq \prod_{i \in S} x_i$. Then the set of character functions $\{\chi_S\}_{S \subseteq [n]}$ forms an orthonormal basis for the space of all real-valued functions on $\{-1,1\}^n$. Hence, every real-valued function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique Fourier expansion

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x).$$

The *degree* of f is defined as $\mathsf{deg}(f) = \max\{|S| : \widehat{f}(S) \neq 0\}$. The *spectral norm* of f is defined to be $\sum_S |\widehat{f}(S)|$. The *Fourier weight* of a function f on a set of coefficients $\mathcal{S} \subseteq 2^{[n]}$ is defined as $\sum_{S \in \mathcal{S}} \widehat{f}(S)^2$.

We note a well-known fact that follows from the orthonormality of the character functions χ_S.

Fact 13 (Plancherel's Theorem)**.** *For $f, g \colon \{-1,1\}^n \to \mathbb{R}$,*

$$\mathbb{E}_x[f(x)g(x)] = \sum_S \widehat{f}(S)\widehat{g}(S).$$

In particular, if $f \colon \{-1,1\}^n \to \{-1,1\}$ is Boolean-valued and $g = f$, we have Parseval's Identity $\sum_S \widehat{f}(S)^2 = \mathbb{E}[f(x)^2]$, which in turn equals 1. Hence $\sum_S \widehat{f}(S)^2 = 1$ and we can view $\{\widehat{f}(S)^2\}_S$ as a probability distribution, which allows us to discuss the Fourier entropy, min-entropy of the distribution $\{\widehat{f}(S)^2\}_S$, defined as

Definition 14. *For a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, its Fourier entropy (denoted $\mathbb{H}(\hat{f}^2)$) and min-entropy (denoted $\mathbb{H}_\infty(\hat{f}^2)$) are*

$$\mathbb{H}(\hat{f}^2) \coloneqq \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}, \quad \textit{and} \quad \mathbb{H}_\infty(\hat{f}^2) \coloneqq \min_{\substack{S \subseteq [n]: \\ \widehat{f}(S) \neq 0}} \left\{ \log \frac{1}{\widehat{f}(S)^2} \right\}.$$

Similarly, we can also define the Rényi Fourier entropy.

Definition 15 (Rényi Fourier entropy [Rén61]). *For $f : \{-1,1\}^n \to \{-1,1\}$, $\alpha \geq 0$ and $\alpha \neq 1$, the Rényi Fourier entropy of f of order α is defined as*

$$\mathbb{H}_\alpha(\hat{f}^2) := \frac{1}{1-\alpha} \log \left(\sum_{S \subseteq [n]} |\widehat{f}(S)|^{2\alpha} \right).$$

It is known that in the limit as $\alpha \to 1$, $\mathbb{H}_\alpha(\hat{f}^2)$ is the (Shannon) Fourier entropy $\mathbb{H}(\hat{f}^2)$ (see [Rén61] for a proof) and when $\alpha \to \infty$, observe that $\mathbb{H}_\alpha(\hat{f}^2)$ converges to $\mathbb{H}_\infty(\hat{f}^2)$. It is easily seen that $\mathbb{H}_\infty(\hat{f}^2) \leq \mathbb{H}(\hat{f}^2) \leq \mathbb{H}_{\frac{1}{2}}(\hat{f}^2) \leq \mathbb{H}_0(\hat{f}^2)$.

For $f : \{-1,1\}^n \to \{-1,1\}$, the *influence* of a coordinate $i \in [n]$, denoted $\mathsf{Inf}_i(f)$, is defined as

$$\mathsf{Inf}_i(f) = \Pr_x[f(x) \neq f(x^{(i)})] = \mathbb{E}_x\left[\left(\frac{f(x) - f(x^{(i)})}{2}\right)^2\right],$$

where the probability and expectation is taken according to the uniform distribution on $\{-1,1\}^n$ and $x^{(i)}$ is x with the i-th bit flipped. The *total influence* of f, denoted $\mathsf{Inf}(f)$ is

$$\mathsf{Inf}(f) = \sum_{i \in [n]} \mathsf{Inf}_i(f).$$

In terms of the Fourier coefficients of f, it can be shown, e.g., [KKL88], that $\mathsf{Inf}_i(f) = \sum_{S \ni i} \widehat{f}(S)^2$, and therefore

$$\mathsf{Inf}(f) = \sum_{S \subseteq [n]} |S| \widehat{f}(S)^2.$$

The *variance* of a real-valued function f is given by $\mathsf{Var}(f) = \sum_{S \neq \emptyset} \widehat{f}(S)^2 = 1 - \widehat{f}(\emptyset)^2$. It easily follows that $\mathsf{Var}(f) \leq \mathsf{Inf}(f)$.

2.2 Complexity measure of Boolean functions.

We introduce some basic complexity measures of Boolean functions which we use often, referring to [BW02] for more.

2.2.1 Sensitivity

For $x \in \{-1,1\}^n$, the *sensitivity* of f at x, denoted $\mathsf{s}_f(x)$, is defined to be the number of neighbors y of x in the Boolean hypercube (i.e., y is obtained by flipping *exactly* one bit of x) such that $f(y) \neq f(x)$. The sensitivity $\mathsf{s}(f)$ of f is $\max_x\{\mathsf{s}_f(x)\}$. The *average sensitivity* $\mathsf{as}(f)$ of f is defined to be $\mathbb{E}_x[\mathsf{s}_f(x)]$. By the linearity of expectation observe that

$$\mathbb{E}_x[\mathsf{s}_f(x)] = \sum_{i=1}^n \Pr_x[f(x) \neq f(x^{(i)})] = \sum_{i=1}^n \mathsf{Inf}_i(f) = \mathsf{Inf}(f),$$

so the average sensitivity of f equals the total influence of f. So, the FEI conjecture asks if $\mathbb{H}(\hat{f}^2) \leq C \cdot \mathsf{as}(f)$ for every Boolean function f.

2.2.2 Certificate complexity measures

Certificate complexity. For $x \in \{-1,1\}^n$, the *certificate complexity* of f at x, denoted $\mathsf{C}(f,x)$, is the minimum number of bits in x that needs to be fixed to ensure that the value of f is constant. The certificate complexity $\mathsf{C}(f)$ of f is $\max_x\{\mathsf{C}(f,x)\}$. The minimum certificate complexity of f is $\mathsf{C}_{\min}(f) = \min_x\{\mathsf{C}(f,x)\}$. The 0-certificate complexity $\mathsf{C}^0(f)$ of f is $\max_{x:f(x)=1}\{\mathsf{C}(f,x)\}$. Similarly, the 1-certificate complexity $\mathsf{C}^1(f)$ of f is $\max_{x:f(x)=-1}\{\mathsf{C}^1(f,x)\}$. Observe that for every $x \in \{-1,1\}^n$, $\mathsf{s}(f,x) \le \mathsf{C}(f,x)$. This gives $\mathsf{as}(f) \le \mathsf{aC}(f)$ and $s(f) \le \mathsf{C}(f)$.

Parity certificate complexity. Analogously, we define the *parity certificate complexity* $\mathsf{C}^{\oplus}(f,x)$ of f at x as the minimum number of parities on the input variables one has to fix in order to fix the value of f at x, i.e.,

$$\mathsf{C}^{\oplus}(f,x) := \min\{\mathsf{co\text{-}dim}(H) \mid x \in H, \text{ and } H \text{ is an affine subspace over } \mathbb{F}_2 \text{ on which } f \text{ is constant}\},$$

where $\mathsf{co\text{-}dim}(H)$ is the *co-dimension* of the affine subspace H. It is easily seen that $\mathsf{C}^{\oplus}(f,x) \le \mathsf{C}(f,x)$. We also define $\mathsf{C}^{\oplus}(f) := \max_x\{\mathsf{C}^{\oplus}(f,x)\}$, and $\mathsf{C}^{\oplus}_{\min}(f) := \min_x \mathsf{C}^{\oplus}(f,x)$.

Unambiguous certificate complexity. We now define the *unambiguous certificate complexity* of f. Let $\tau : [n] \to \{-1,1,*\}$ be a partial assignment. We refer to $S_\tau = \{x \in \{-1,1\}^n : x_i = \tau(i) \text{ for every } i \in [n]\backslash\tau^{-1}(*)\}$ as the subcube generated by τ. We call $C \subseteq \{-1,1\}^n$ a *subcube* of $\{-1,1\}^n$ if there exists a partial assignment τ such that $C = S_\tau$ and the co-dimension of C is the number of bits fixed by τ, i.e., $\mathsf{co\text{-}dim}(C) = |\{i \in [n] : \tau(i) \neq *\}|$. A set of subcubes $\mathcal{C} = \{C_1,\dots,C_m\}$ *partitions* $\{-1,1\}^n$ if the subcubes are disjoint and they cover $\{-1,1\}^n$, i.e., $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_i C_i = \{-1,1\}^n$.

An *unambiguous certificate* $\mathcal{U} = \{C_1,\dots,C_m\}$ (also referred to as a *subcube partition*) is a set of subcubes partitioning $\{-1,1\}^n$. We say $\mathcal{U}$ computes a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ if f is constant on each C_i (i.e., $f(x)$ is the same for all $x \in C_i$). For an unambiguous certificate $\mathcal{U}$, the *unambiguous certificate complexity* on input x (denoted $\mathsf{UC}(\mathcal{U},x)$), equals $\mathsf{co\text{-}dim}(C_i)$ for a C_i satisfying $x \in C_i$. Define the *average unambiguous certificate* of f with respect to $\mathcal{U}$ as $\mathsf{aUC}(f,\mathcal{U}) := \mathbb{E}_x[\mathsf{UC}(\mathcal{U},x)]$. Then, the *average unambiguous certificate complexity* of f is defined as

$$\mathsf{aUC}(f) := \min_{\mathcal{U}} \ \mathsf{aUC}(f,\mathcal{U}),$$

where the minimization is over all unambiguous certificates for f. Finally, the *unambiguous certificate complexity of* f is

$$\mathsf{UC}(f) := \min_{\mathcal{U}} \ \max_x \ \mathsf{UC}(\mathcal{U},x).$$

Note that since unambiguous certificates are more restricted than general certificates, we have $\mathsf{C}(f) \le \mathsf{UC}(f)$.

An unambiguous $\oplus$-certificate $\mathcal{U} = \{C_1,\dots,C_m\}$ for f is defined to be a collection of monochromatic *affine subspaces* that together partition the space $\{-1,1\}^n$. It is easily seen that a subcube is also an affine subspace. Analogously, for an unambiguous $\oplus$-certificate $\mathcal{U}$, on an input x, $\mathsf{UC}^{\oplus}(\mathcal{U},x) := \mathsf{co\text{-}dim}(C_i)$ for a C_i satisfying $x \in C_i$, and $\mathsf{aUC}^{\oplus}(f,\mathcal{U}) := \mathbb{E}_x[\mathsf{UC}^{\oplus}(\mathcal{U},x)]$. Similarly, we define $\mathsf{aUC}^{\oplus}(f)$ and $\mathsf{UC}^{\oplus}(f)$.

2.2.3 Degree and Decision tree complexity

Approximate degree. The *ε-approximate degree* of $f : \{-1,1\}^n \to \mathbb{R}$, denoted $\mathsf{deg}_\varepsilon(f)$, is defined to be the minimum degree among all multilinear real polynomial p such that $|f(x)-p(x)| \le \varepsilon$

for all $x \in \{-1,1\}^n$. Usually ε is chosen to be $1/3$, but it can be chosen to be any constant in $(0,1)$ (the choice of the constant affects the approximate degree only by a constant factor).

Deterministic decision tree. A deterministic decision tree for $f : \{-1,1\}^n \to \{-1,1\}$ is a rooted binary tree where each node is labelled by $i \in [n]$ and the leaves are labelled with an output bit $\{-1,1\}$. On input $x \in \{-1,1\}^n$, the tree proceeds at the ith node by evaluating the bit x_i and continuing with the subtree corresponding to the value of x_i. Once a leaf is reached, the tree outputs a bit. We say that a deterministic decision tree computes f if for all $x \in \{-1,1\}^n$ its output equals $f(x)$.

Randomized decision tree. A randomized decision tree for f is a probability distribution R_μ over deterministic decision trees for f. On input x, a decision tree is chosen according to R_μ, which is then evaluated on x. The complexity of the randomized tree is the largest depth among all deterministic trees with non-zero probability of being sampled according to R_μ.

We say that a randomized decision tree computes f with bounded-error, if for all $x \in \{-1,1\}^n$ its output equals $f(x)$ with probability at least $2/3$. $R_2(f)$ denotes the complexity of the optimal randomized decision tree that computes f with bounded-error, i.e., errs with probability at most $1/3$.

3 The Fourier Min-entropy-Influence conjecture

The *Fourier Min-entropy-Influence conjecture* (FMEI) is a natural weakening of the FEI conjecture that has received much less attention compared to the FEI conjecture. The FMEI conjecture was raised by O'Donnell and others in [OWZ11, O'D14] as a simpler question to tackle, given the hardness of resolving the FEI conjecture. Restating the FMEI conjecture below.

Conjecture 16 (FMEI conjecture)**.** *There exists a universal constant $C > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$.*

Although the FMEI conjecture is a natural first step towards proving the FEI conjecture, it is also interesting in its own right. The FMEI conjecture implies the famous KKL theorem [KKL88] as we show below. In fact we do not know of any proof of the KKL theorem that doesn't go through Hypercontractivity or logarithmic Sobolev inequalities, which makes proving the FMEI conjecture even more interesting. The KKL theorem states that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, there exists an index $i \in [n]$ such that $\mathsf{Inf}_i(f) \geq \mathsf{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$. We now show how a positive answer to the FMEI conjecture implies the KKL theorem:[6] for simplicity, assume that f is balanced (i.e., $\widehat{f}(\emptyset) = \mathbb{E}_x[f(x)] = 0$), then the FMEI conjecture implies the existence of $\emptyset \neq T \subseteq [n]$ such that $\widehat{f}(T)^2 \geq 2^{-C \cdot \mathsf{Inf}(f)}$. Furthermore, for every $i \in T$, we have

$$\mathsf{Inf}_i(f) = \sum_{S : S \ni i} \widehat{f}(S)^2 \geq \widehat{f}(T)^2 \geq 2^{-C \cdot \mathsf{Inf}(f)} \geq 2^{-Cn \cdot \max_j \{\mathsf{Inf}_j(f)\}}, \tag{3}$$

where the first inequality follows because T contains i, the second inequality follows from the FMEI conjecture and the last inequality because $\mathsf{Inf}(f) \leq n \cdot \max_j \mathsf{Inf}_j(f)$. However note that $\max_j\{\mathsf{Inf}_j(f)\}$ clearly upper bounds the left-hand-side of Eq. (3). Thus, we have $\max_j\{\mathsf{Inf}_j(f)\} \geq 2^{-Cn \cdot \max_j\{\mathsf{Inf}_j(f)\}}$. Rearranging this inequality, we obtain $\max_{j \in [n]} \mathsf{Inf}_j(f) \geq \Omega\left(\frac{\log n}{n}\right)$, which is the

[6] This argument has appeared before in [OWZ11].

KKL theorem for balanced functions. The proof can also be extended to non-balanced functions, which we omit here.

We now prove Theorem 17 which is our main contribution in this section (restated below for convenience). In the following theorem, we give upper bounds on $\mathbb{H}_\infty(\hat{f}^2)$ in terms of analytic and combinatorial measures of Boolean functions.

Theorem 17. *(Restatement of Theorem 5) Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then,*

1. *for every $\varepsilon \geq 0$, $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \log\left(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon)\right)$.*
2. *$\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \mathsf{C}^{\oplus}_{\min}(f)$.*
3. *$\mathbb{H}_\infty(\hat{f}^2) \leq 2(1+\log_2 3) \cdot R^{\oplus}_2(f)$.*

Before giving a proof, we first make a few remarks about the second statement in the theorem above. In Section 2.2, we saw that the total influence of a Boolean function f is equal to the average sensitivity of f. So the FMEI conjecture asks if $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{as}(f)$? Since we also know that for every $x \in \{-1,1\}^n$, we have $\mathsf{s}(f,x) \leq \mathsf{C}(f,x)$, a weaker question than the FMEI conjecture (with a larger right-hand-side) would be, is $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{aC}(f)$? In the theorem above, we give a positive answer to this question and in fact show that $\mathbb{H}_\infty(\hat{f}^2) \leq 2\mathsf{C}^{\oplus}_{\min}(f)$. Observe that $\mathsf{C}^{\oplus}_{\min}(f)$ is not only a lower bound on $\mathsf{aC}^{\oplus}(f)$ (and in turn $\mathsf{aC}(f)$), but it is the parity certificate complexity on the "easiest" input to f, unlike $\mathsf{C}^{\oplus}(f)$ where the complexity is measured according to the "hardest" input x to f. In our perspective, this brings us significantly closer to proving the FMEI conjecture. In fact, we identify a non-trivial class of Boolean functions for which $\mathsf{C}_{\min}(f)$ lower bounds $\mathsf{Inf}(f)$, and hence establish the FMEI conjecture for this class (Theorem 23).

Proof of Theorem 17. We prove the three parts separately as follows.

Part (1). Fix $\varepsilon \geq 0$ and $d \in [n]$. Given a Boolean function f, suppose p is a degree-d polynomial that minimizes

$$\|\widehat{f}\|_{1,\varepsilon,d} = \min\{\|\widehat{p}\|_1 : \mathsf{deg}(p) = d \text{ and } |p(x) - f(x)| \leq \varepsilon \text{ for every } x \in \{-1,1\}^n\},$$

where the minimization is over all polynomials. Alternatively, $\|\widehat{f}\|_{1,\varepsilon,d}$ can also be expressed as the following linear program and p minimizes this program.

$$\begin{array}{lll} \|\widehat{f}\|_{1,\varepsilon,d} = & \min & \sum_S |c_S| \\ & \text{subject to} & \left|f(x) - \sum_{S:|S|\leq d} c_S \chi_S(x)\right| \leq \varepsilon \quad \text{for every } x \in \{-1,1\}^n \\ & & c_S \in \mathbb{R} \quad \text{for every } S : |S| \leq d \end{array}$$

Note that for every $\varepsilon \geq 0$ and $d \geq \mathsf{deg}_\varepsilon(f)$ the above linear program is feasible. From standard manipulations, the dual of the linear program is as follows.

$$\begin{array}{lll} \max & \sum_{x\in\{-1,1\}^n} \phi(x) f(x) - \varepsilon \sum_{x\in\{-1,1\}^n} |\phi(x)| & \\ \text{subject to} & |\widehat{\phi}(S)| \leq \frac{1}{2^n} & \text{for every } S : |S| \leq d \\ & \phi(x) \in \mathbb{R} & \text{for every } x \in \{-1,1\}^n \end{array}$$

Observe that both linear programs are feasible for $d \geq \mathsf{deg}_\varepsilon(f)$. Therefore, from the duality theorem of linear programs, the objective value of any dual feasible solution lower bounds the primal optimum and, moreover, the two programs have the same optimum value. We thus obtain the following characterization of $\|\widehat{f}\|_{1,\varepsilon,d}$.[7]

Lemma 18. *Let $f : \{-1,1\}^n \to \{-1,1\}$, $\varepsilon \geq 0$, and $d \in [n]$ such that $d \geq \mathsf{deg}_\varepsilon(f)$. Then, $\|\widehat{f}\|_{1,\varepsilon,d} \geq T$ if and only if there exists a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ satisfying $|\widehat{\phi}(S)| \leq 2^{-n}$ for all $|S| \leq d$ and*

$$\sum_{x \in \{-1,1\}^n} \phi(x) f(x) - \varepsilon \sum_{x \in \{-1,1\}^n} |\phi(x)| \geq T.$$

Let us consider $\phi(x) = \frac{f(x)}{2^n \max_S |\widehat{f}(S)|}$. Clearly the dual constraints are satisfied, and the objective value for this choice of ϕ is given by

$$\sum_{x \in \{-1,1\}^n} \phi(x) f(x) - \varepsilon \sum_{x \in \{-1,1\}^n} |\phi(x)| = \frac{1-\varepsilon}{\max_S |\widehat{f}(S)|}.$$

The equality holds since $\phi(x)f(x) = |\phi(x)|$. Now, by Lemma 18, we have

$$\|\widehat{f}\|_{1,\varepsilon,d} \geq \frac{1-\varepsilon}{\max_S |\widehat{f}(S)|}.$$

Therefore, we obtain,

$$\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \log\left(\frac{\|\widehat{f}\|_{1,\varepsilon,d}}{1-\varepsilon}\right).$$

Since d is arbitrary, in order to ensure feasibility of the program we pick $d = \mathsf{deg}_\varepsilon(f)$. The first part of the theorem follows since $\|\widehat{f}\|_{1,\varepsilon,d} = \|\widehat{f}\|_{1,\varepsilon}$ for $d = \mathsf{deg}_\varepsilon(f)$.

Part (2). Suppose $\mathsf{C}^{\oplus}_{\min}(f) = k$. By definition of $\mathsf{C}^{\oplus}_{\min}(f)$, there exists an affine subspace $H \subseteq \{-1,1\}^n$ such that $\mathsf{co\text{-}dim}(H) = k$ and f is constant on H. Without loss of generality, assume that $f(x) = -1$ for every $x \in H$. Since $\mathsf{co\text{-}dim}(H)$ equals k, H is given by a set of k (linearly independent) parity constraints. That is, there exist k linearly independent vectors $S_1, \ldots, S_k \in \{0,1\}^n$, and $b_1, \ldots, b_k \in \{-1,1\}$, such that

$$H = \left\{ x \in \{-1,1\}^n : \text{ for every } j \in [k], \prod_{i \in \mathsf{supp}(S_j)} x_i = b_j \right\}.$$

Consider the indicator function $\mathbb{1}_H : \{-1,1\}^n \to \{-1,1\}$, which evaluates to -1 for every $x \in H$ and 1 otherwise. The Fourier expansion of $\mathbb{1}_H$ is easy to understand. Observe that H can be viewed as an AND over parities or negated-parities. For $j \in [k]$, let $y_j = \prod_{i \in \mathsf{supp}(S_j)} x_i$. It is now easily seen that

$$\mathbb{1}_H(x) = \mathsf{AND}(-b_1 y_1, -b_2 y_2, \ldots, -b_k y_k). \tag{4}$$

[7] We remark that similar linear program characterizations of approximate degree of Boolean functions have appeared before in the works of Sherstov [She11] and Bun and Thaler [BT13].

Recall that b_j is fixed, thus $-b_j y_j$ is either y_j or $-y_j$. Writing out the Fourier expansion for the AND function in Eq. (4), it follows that

$$\begin{aligned}\mathbb{1}_H(x) &= \left(1 - \frac{1}{2^{k-1}}\right) + \sum_{T \subseteq [k]\colon T \neq \emptyset} \frac{(-1)^{|T|+1}}{2^{k-1}} \prod_{j \in T} -b_j y_j \\ &= \left(1 - \frac{1}{2^{k-1}}\right) + \sum_{T \subseteq [k]\colon T \neq \emptyset} \left(\frac{-\prod_{j \in T} b_j}{2^{k-1}}\right) \prod_{j \in T} \left(\prod_{i \in \mathsf{supp}(S_j)} x_i\right).\end{aligned}$$

Now using the fact that $x_i^2 = 1$, observe that the monomial $\prod_{j \in T} \prod_{i \in \mathsf{supp}(S_j)} x_i$ simplifies to a multilinear monomial. We further observe that for each non-empty T, we can simplify $\prod_{j \in T} \prod_{i \in \mathsf{supp}(S_j)} x_i$ to a distinct multilinear monomial. This is a consequence of linear independence of $S_1, S_2, \ldots, S_k$. Let us denote the set of non-zero Fourier coefficients of $\mathbb{1}_H$ by $\mathcal{T}$. By what we argued just now, it follows that $|\mathcal{T}| \leq 2^k$. We are now ready to conclude the proof.

Lemma 19. *There exists a set $T \subseteq [n]$ such that $|\widehat{f}(T)| \geq \frac{1}{2^k}$.*

Proof. Let $\Pr_x[f(x) = -1] = p$. We consider the correlation between f and $\mathbb{1}_H$.

$$\langle f, \mathbb{1}_H \rangle = \mathbb{E}_x[f(x)\mathbb{1}_H(x)] = \frac{1}{2^k} + (-1)\left(p - \frac{1}{2^k}\right) + (1-p) = (1-2p) + \frac{1}{2^{k-1}} = \widehat{f}(\emptyset) + \frac{1}{2^{k-1}}. \quad (5)$$

On the other hand,

$$\langle f, \mathbb{1}_H \rangle = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{\mathbb{1}_H}(S) = \widehat{f}(\emptyset)\left(1 - \frac{1}{2^{k-1}}\right) + \sum_{T \in \mathcal{T}\colon T \neq \emptyset} \widehat{f}(T)\left(\frac{\prod_{j \in T} b_j}{2^{k-1}}\right). \quad (6)$$

Putting together Eq. (5) and Eq. (6) we have,

$$1 = -\widehat{f}(\emptyset) + \sum_{T \in \mathcal{T}\colon T \neq \emptyset} \widehat{f}(T)\left(\Pi_{j \in T} b_j\right) \leq \sum_{T \in \mathcal{T}} |\widehat{f}(T)|. \quad (7)$$

Since $|\mathcal{T}| = 2^k$, we obtain $\max_{T \in \mathcal{T}} |\widehat{f}(T)| \geq \frac{1}{2^k}$. □

This completes the proof of Part (2).

Part (3). Consider a randomized parity-decision tree R_μ computing f with probability at least $2/3$. Let $\mathcal{T}$ be the set of deterministic parity decision trees such that R_μ assigns a non-zero probability to every $T \in \mathcal{T}$. By Definition 2.2.3, it then follows that $\mathbb{E}_x \mathbb{E}_{T \sim \mu}[f(x)T(x)] \geq 1/3$. This also shows that

$$\frac{1}{3} \leq \mathbb{E}_x \mathbb{E}_{T \sim \mu}[f(x)T(x)] = \mathbb{E}_{T \sim \mu} \mathbb{E}_x[f(x)T(x)] = \mathbb{E}_{T \sim \mu}\left[\sum_{S \subseteq [n]} \widehat{f}(S)\widehat{T}(S)\right]. \quad (8)$$

On the other hand, one can upper bound the last expression in Eq. (8) as follows

$$\mathbb{E}_{T \sim \mu}\left[\sum_{S \subseteq [n]} \widehat{f}(S)\widehat{T}(S)\right] \leq \mathbb{E}_{T \sim \mu}\left[\sum_{S \subseteq [n]} |\widehat{f}(S)||\widehat{T}(S)|\right] \leq \left(\max_{S \subseteq [n]} |\widehat{f}(S)|\right) \mathbb{E}_{T \sim \mu}\left[\sum_{S \subseteq [n]} |\widehat{T}(S)|\right]. \quad (9)$$

Putting together Eq. (8) and Eq. (9) we have,

$$\frac{1}{3} \leq \left(\max_{S\subseteq[n]} |\widehat{f}(S)|\right) \mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} |\widehat{T}(S)|\right] \leq \left(\max_{S\subseteq[n]} |\widehat{f}(S)|\right) 2^{R_2^{\oplus}(f)}.$$

The second inequality follows from the fact that each T is a deterministic parity-decision tree of depth at most $R_2^{\oplus}(f)$, hence it easily follows that the spectral norm of the Fourier coefficients of T can be upper bounded by $2^{R_2^{\oplus}(f)}$ (for a proof of this, see [BOH90]). Rewriting the last inequality, we have $\max_{S\subseteq[n]} |\widehat{f}(S)| \geq \frac{1}{2^{R_2^{\oplus}(f)+\log 3}}$, which gives us the third part of the theorem. □

Using a well known fact that upper bounds Rényi entropy of order $1+\delta$ (for every $\delta > 0$) by a constant times the min-entropy of $\{\hat{f}(S)^2\}$, we deduce the following corollary. Since this fact works for all $\delta > 0$, it is tempting to say that we can improve the bounds in Theorem 17 from $\mathbb{H}_\infty(f)$ to $\mathbb{H}(f)$, but this relation between the Rényi entropies breaks down for $\delta = 0$.

Corollary 20. *Let $f : \{-1,1\}^n \to \{-1,1\}$, and $\delta > 0$. Then,*

1. *for every $\varepsilon \geq 0$, $\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\frac{1}{\delta}) \cdot \log\left(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon)\right)$.*
2. *$\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\frac{1}{\delta}) \cdot \mathsf{C}^{\oplus}_{\min}(f)$.*
3. *$\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\log 3)(1+\frac{1}{\delta}) \cdot R_2^{\oplus}(f)$.*

Proof. Use the fact that for any distribution P, $\mathbb{H}_{1+\delta}(P) \leq (1+\frac{1}{\delta})\,\mathbb{H}_\infty(P)$. Indeed, it is easily seen from the definition of Rényi entropy that $-\frac{1}{\delta}\log\left(\sum_j p_j^{1+\delta}\right) \leq -\frac{1+\delta}{\delta}\log\left(\max_j p_j\right)$, and thus the fact follows. We remark that a tighter analysis of the Rényi entropy can be used to improve the constants. □

As a corollary to Theorem 17(2), we now establish the FMEI conjecture for read-k DNFs, for constant k. A Boolean function is said to belong to the class of read-k DNF if it can be expressed as a DNF such that every variable (negated or un-negated) appears in at most k terms. We note that, independently, Shalev [Sha18] showed, among other things, that FMEI holds for "regular" read-k DNFs. However, we show it for the general class of read-k DNFs. We remark that this improvement crucially uses our *sharper* bound of $\mathsf{C}_{\min}$ on the min-entropy of $\{\hat{f}(S)^2\}_S$.

We will need the well known KKL theorem which we state below.

Theorem 21 ([KKL88])**.** *There exists a universal constant $c > 0$ such that for every $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathsf{Inf}(f) \geq c \cdot \mathsf{Var}(f) \cdot \log \frac{1}{\max_i \mathsf{Inf}_i(f)}.$$

The next lemma establishes a lower bound of *minimum* certificate size on the total influence of constant-read DNF. A similar argument appears in Shalev [Sha18] too.

Lemma 22. *There exists a universal constant $c > 0$ such that for all $f\colon \{-1,1\}^n \to \{-1,1\}$ that can be expressed as a read-k DNF, we have*

$$\mathsf{Inf}(f) \geq c \cdot \mathsf{Var}(f) \cdot \left(\mathsf{C}_{\min}(f) - 1 - \log k\right).$$

Proof. Suppose f is a t-term read-k DNF, then f can be written as $f(x) = \bigvee_{j=1}^{t} T_j(x)$, where $T_j(x)$ is a term. Recall, $\mathsf{Inf}_i(f) = \Pr_x[f(x) \neq f(x^{(i)})]$. Using the fact that f is a DNF we upper bound the $\mathsf{Inf}_i(f)$ as follows,

$$\mathsf{Inf}_i(f) = \Pr_x[f(x) \neq f(x^{(i)})] \leq \sum_{j=1}^{t} \Pr_x[T_j(x) \neq T_j(x^{(i)})].$$

Clearly when T_j is not defined over a variable x_i, $\Pr_x[T_j(x) \neq T_j(x^{(i)})] = 0$, and otherwise it equals $\frac{1}{2^{|T_j|-1}}$ because all other literals must be set to true in order to satisfy that term. Therefore, we have

$$\mathsf{Inf}_i(f) \leq \sum_{j=1}^{t} \Pr_x[T_j(x) \neq T_j(x^{(i)})] = \sum_{\substack{j=1: \\ x_i \text{ appears in } T_j}}^{t} \Pr_x[T_j(x) \neq T_j(x^{(i)})] \leq k2^{-(\mathsf{C}_{\min}(f)-1)}.$$

The second inequality follows because a variable appears in at most k terms and $|T_j| \geq \mathsf{C}_{\min}(f)$ for all j. Now using the KKL theorem (Theorem 21), we obtain

$$\mathsf{Inf}(f) \geq c \cdot \mathsf{Var}(f) \cdot \log \frac{1}{\max_i \mathsf{Inf}_i(f)} \geq c \cdot \mathsf{Var}(f) \cdot (\mathsf{C}_{\min}(f) - 1 - \log k).$$

This concludes the proof of the lemma. □

We now use Lemma 22 and Theorem 17(2) to show that the FMEI conjecture holds for read-k DNF.

Theorem 23. *(Restatement of Theorem 6) Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a read-k DNF. Then,*

$$\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f),$$

where $C = O(\log k)$.

Proof. We consider two cases based on whether $\mathsf{Var}(f)$ is "small"($< 1/2$) or "large"($\geq 1/2$).

Case 1 : $\mathsf{Var}(f) < 1/2$. Recall, $\mathsf{Var}(f) = 1 - \widehat{f}(\emptyset)^2$. Therefore, we have

$$\mathbb{H}_\infty(\hat{f}^2) \leq \log \frac{1}{\widehat{f}(\emptyset)^2} = \log \frac{1}{1 - \mathsf{Var}(f)} \leq (\log e)\frac{\mathsf{Var}(f)}{1 - \mathsf{Var}(f)} \leq (2 \log e)\mathsf{Var}(f) \leq (2 \log e)\mathsf{Inf}(f).$$

The second inequality uses the fact, for $x \in (0,1)$, $\log \frac{1}{1-x} \leq (\log e)\frac{x}{1-x}$.

Case 2 : $\mathsf{Var}(f) \geq 1/2$. Using Theorem 17(2) and Lemma 22 we bound the min-entropy as follows,

$$\mathbb{H}_\infty(\hat{f}^2) \leq 2\mathsf{C}_{\min}(f) \leq (2/c) \cdot \frac{\mathsf{Inf}(f)}{\mathsf{Var}(f)} + 2(1 + \log k),$$

where c is a universal constant. Since $\mathsf{Inf}(f) \geq \mathsf{Var}(f) \geq 1/2$, we further bound the last term in the above inequality to obtain

$$\mathbb{H}_\infty(\hat{f}^2) \leq ((4/c) + 4(1 + \log k)) \cdot \mathsf{Inf}(f).$$

This completes the proof of the theorem. □

4 Better bound on the FEI conjecture

In this section we give a new improved upper bound on the Fourier entropy of arbitrary Boolean functions. It is well-known that $\mathsf{Inf}(f)$ lower bounds many combinatorial measures associated with Boolean functions such as decision tree depth, certificate complexity, sensitivity, etc. Given the difficulty in resolving the FEI conjecture, it is natural to wonder if the Fourier entropy can be upper bounded by these larger measures. Indeed, Chakraborty et al. [CKLS16] established many bounds on the Fourier entropy, including *average parity-decision tree complexity*. We improve on their bounds by showing an upper bound of *average unambiguous parity-certificate complexity*. It is known that unambiguous certificate complexity can be quadratically smaller than decision tree complexity [AKK16].

A new and crucial ingredient of our proof is the following consequence of the law of large numbers, called the *Asymptotic Equipartition Property (AEP)* or *the Shannon-McMillan-Breiman theorem*. See Chapter 3 in the book [CT91] for more details.

Theorem 24 (Asymptotic Equipartition Property (AEP) Theorem)**.** *Let $\mathbf{X}$ be a random variable drawn from a distribution P and suppose $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_M}$ are independently and identically distributed copies of $\mathbf{X}$, then*

$$-\frac{1}{M}\log P(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M) \longrightarrow \mathbb{H}(\mathbf{X})$$

in probability as $M \to \infty$.

Definition 25. *Fix $\varepsilon \geq 0$. The* typical set $T_\varepsilon^{(M)}(\mathbf{X})$ *with respect to a distribution P is defined to be the set of sequences $(x_1, x_2, \ldots, x_M) \in \mathbf{X_1} \times \mathbf{X_2} \times \cdots \times \mathbf{X_M}$ such that*

$$2^{-M(\mathbb{H}(\mathbf{X})+\varepsilon)} \leq P(x_1, x_2, \ldots, x_M) \leq 2^{-M(\mathbb{H}(\mathbf{X})-\varepsilon)}.$$

The following properties of the typical set follows from the AEP.

Theorem 26. *[CT91, Theorem 3.1.2] Let $\varepsilon \geq 0$ and $T_\varepsilon^{(M)}(\mathbf{X})$ be a typical set with respect to P, then*

(i) $|T_\varepsilon^{(M)}(\mathbf{X})| \leq 2^{M(\mathbb{H}(\mathbf{X})+\varepsilon)}$.

(ii) Suppose $x_1, \ldots, x_M$ are drawn i.i.d. according to $\mathbf{X}$, then $\Pr[(x_1, \ldots, x_M) \in T_\varepsilon^{(M)}(\mathbf{X})] \geq 1-\varepsilon$ for M sufficiently large.

(iii) $|T_\varepsilon^{(M)}(\mathbf{X})| \geq (1-\varepsilon)2^{M(\mathbb{H}(\mathbf{X})-\varepsilon)}$ *for M sufficiently large.*

We recall from the preliminaries, an unambiguous certificate $\mathcal{C} = \{C_1, \ldots, C_t\}$ for f is a collection of monochromatic subcubes (with respect to f) that together partition the hypercube $\{-1,1\}^n$. The *average unambiguous certificate complexity* of f with respect to $\mathcal{C}$, denoted $\mathsf{aUC}(f, \mathcal{C})$, equals $\mathbb{E}_{x\in\{-1,1\}^n}[\mathsf{UC}(\mathcal{C}, x)]$. Further, $\mathsf{aUC}(f) = \min_{\mathcal{C}} \mathsf{aUC}(f, \mathcal{C})$.

We now proceed to prove our main theorem. In order to keep the presentation clear, we first prove a weaker upper bound of *average unambiguous certificate complexity* $\mathsf{aUC}(f)$, on the $\mathbb{H}(\hat{f}^2)$. We then sketch how to generalize the proof to establish the stronger upper bound of *average unambiguous parity-certificate complexity* $\mathsf{aUC}^{\oplus}(f)$.

Theorem 27. *(Restatement of Theorem 7) For every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \mathsf{aUC}(f).$$

Before we prove this inequality, we first give a sketch of the proof. Given $f\colon \{-1,1\}^n \to \{-1,1\}$ and an unambiguous certificate $\mathcal{C}$ for f, our first step is to consider the function $f^M : \{-1,1\}^{Mn} \to \{-1,1\}$ defined as the M-fold product of f. We then consider a random variable $\mathbf{C}$ (supported on $\mathcal{C}$) and let $T_\delta^{(M)}(\mathbf{C})$ be a typical set associated with M i.i.d copies of $\mathbf{C}$. Based on the typical set, we define a $2^{M\cdot(\mathsf{aUC}(f)+\delta)}$-sized set $\mathcal{B}$ of Fourier coefficients of f^M and show that $\mathcal{B}$ has fairly large Fourier weight. We then consider the re-normalized entropy when restricted to the Fourier coefficients in $\mathcal{B}$. By taking the limit ($M \to \infty$), we obtain $\mathbb{H}(\hat{f}^2) \le O(\mathsf{aUC}(f))$. We now fill in the details and prove the main theorem.

Proof. In fact, we will establish a stronger statement that for every unambiguous certificate $\mathcal{C}$ for f, we have

$$\mathbb{H}(\hat{f}^2) \le 2 \cdot \mathsf{aUC}(f, \mathcal{C}).$$

Let $\mathcal{C} := \{C_1, \dots, C_t\}$ be an unambiguous certificate of f. For every C_i, let $\tau(C_i)$ be the partial assignment $\tau(C_i) : [n] \to \{-1, 1, *\}$, corresponding to the bits fixed by C_i. Consider the Boolean function $f^M : \{-1,1\}^{Mn} \to \{-1,1\}$ given by the *M-fold iterated product* of f with itself over distinct variables, i.e.,

$$f^M(x^1, \dots, x^M) = f(x^1_1, \dots, x^1_n) \cdot f(x^2_1, \dots, x^2_n) \cdot f(x^M_1, \dots, x^M_n),$$

where $x^i \in \{-1,1\}^n$ for every $i \in [M]$. First, observe that $\mathbb{H}(\widehat{f^M}^2) = M \cdot \mathbb{H}(\hat{f}^2)$. Similarly, we also have $\mathsf{aUC}(f^M, \mathcal{C}^M) = M \cdot \mathsf{aUC}(f, \mathcal{C})$.

We now bound the Fourier entropy of f^M by showing that there is a "small" set of Fourier coefficients of f^M whose total Fourier weight is approximately 1.

Let $\mathbf{C}$ be a subcube-valued random variable that equals C_i with probability $2^{-|\tau(C_i)|}$.[8] Further, let $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_M$ be i.i.d. random copies of $\mathbf{C}$. For a choice of $\delta > 0$, let $T_\delta^{(M)}(\mathbf{C})$ be the typical set with respect to the distribution $\{2^{-\tau(C_i)}\}_{i\in[t]}$.

We now define a set $\mathcal{B}$ of Fourier coefficients of f^M, which we argue below to have large Fourier weight.

$$\mathcal{B} := \Big\{(S_1, \dots, S_M) \subseteq [n]^M \mid S_i \subseteq \mathsf{supp}(\tau(C_i)) \subseteq [n] \text{ for } i \in [M] \text{ and } (C_1, \dots, C_M) \in T_\delta^{(M)}(\mathbf{C})\Big\}.$$

Using Theorem 26 about typical sequences, we are now ready to bound the size of $\mathcal{B}$ as follows.

Claim 4.1. $|\mathcal{B}| \le 2^{2M(\mathsf{aUC}(f,\mathcal{C})+\delta)}$.

Proof. We first bound the size of $T_\delta^{(M)}(\mathbf{C})$ and then count contributions of a typical sequence $(C_1, \dots, C_M)$ to $\mathcal{B}^M$.

For the first bound, by the properties of the AEP Theorem 26 (i), the total number of typical sequences $|T_\delta^{(M)}(\mathbf{C})|$ is at most $2^{M(\mathbb{H}(\mathbf{C})+\delta)}$. For the second bound, observe that $(C_1, \dots, C_M)$ contributes a set $(S_1, \dots, S_M)$ to $\mathcal{B}^M$ if and only if $S_i \subseteq \mathsf{supp}(\tau(C_i))$ for all $i \in [M]$. Therefore, the maximum possible contribution of a typical sequence is bounded by

$$2^{|\tau(C_1)|+\cdots+|\tau(C_M)|} = (\Pr[\mathbf{C}_1 = C_1, \dots, \mathbf{C}_M = C_M])^{-1} \le 2^{M(\mathbb{H}(\mathbf{C})+\delta)},$$

where the equality is because the random variable $\mathbf{C}$ was sampled according to the distribution $\{2^{-|\tau(C_i)|}\}_{i\in[t]}$ and the inequality follows from Definition 25 of typical sets.

[8] Since $\{C_1, \dots, C_t\}$ are disjoint subcubes partitioning $\{-1,1\}^n$, we have that $\sum_{i=1}^t 2^{-|\tau(C_i)|} = 1$.

Combining both the upper bounds, we get $|\mathcal{B}| \leq 2^{2M(\mathbb{H}(\mathbf{C})+\delta)}$. Finally, by the definition of entropy, we have

$$\mathbb{H}(\mathbf{C}) = \sum_{i=1}^{t} 2^{-|\tau(C_i)|}|\tau(C_i)| = \frac{1}{2^n}\sum_{x\in\{-1,1\}^n} \mathsf{UC}(\mathcal{C}, x) = \mathsf{aUC}(f, \mathcal{C}),$$

where the second equality used that $\{C_1, \ldots, C_t\}$ formed an unambiguous certificate for f. Hence, the claim follows. □

We now claim that $\mathcal{B}$ is the "small" set of Fourier coefficients of f^M that has large Fourier weight. In order to quantitatively prove this, we show that the Fourier coefficients that are *not* in $\mathcal{B}$ have total Fourier weight at most δ.

Claim 4.2. $\sum_{(S_1,S_2,\ldots,S_M)\notin\mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \delta$.

Proof. We saw earlier that $\mathcal{C}^M$ is an unambiguous certificate of f^M. Let $\rho \in \mathcal{C}^M$ be a certificate of f^M, and $\mathbb{1}_\rho(z)$ be the $\{0,1\}$-valued function that is 1 if and only if z is consistent with the certificate ρ. Further we denote the value f^M takes on any input consistent with ρ by $f^M(\rho)$. We can then express f^M on input $z \in \{-1,1\}^{Mn}$ as follows

$$f^M(z) = \sum_{\rho\in\mathcal{C}^M} f^M(\rho)\cdot\mathbb{1}_\rho(z) = \sum_{\rho\in T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z) + \underbrace{\sum_{\rho\notin T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z)}_{:=g(z)}. \tag{10}$$

For $(S_1, \ldots, S_M) \subseteq [n]^M$, consider the expansion of the Fourier coefficient

$$\begin{aligned}
&\widehat{f^M}(S_1 \cup \cdots \cup S_M)\\
&= \mathbb{E}_z[f^M(z)\chi_{S_1\cup\cdots\cup S_M}(z)]\\
&= \mathbb{E}_z\left[\sum_{\rho\in T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z) + \sum_{\rho\notin T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z)\right]\\
&= \sum_{\rho\in T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{E}_z[\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z)] + \sum_{\rho\notin T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{E}_z[\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z)].
\end{aligned}$$

Now observe that for a fixed certificate ρ, we have $\mathbb{E}_z[\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z)] \neq 0$ if and only if ρ fixes the variables in $S_1 \cup \cdots \cup S_M$. By definition of $\mathcal{B}$ it now follows that, if $(S_1, \ldots, S_M) \notin \mathcal{B}$, then $\mathbb{E}_z[\mathbb{1}_\rho(z)\cdot\chi_{S_1\cup\cdots\cup S_M}(z)] = 0$ for $\rho \in T_\delta^{(M)}(\mathbf{C})$, and thus $\widehat{f^M}(S_1 \cup \cdots \cup S_M)$ gets contribution only from ρ that are *not* typical, i.e., $\rho \notin T_\delta^{(M)}(\mathbf{C})$.

In this direction, consider the function $g(z)$ defined in Eq. (10), which is $\{-1, 1, 0\}$-valued. Using the argument above, we have that if $(S_1, \ldots, S_M) \notin \mathcal{B}$, then $\widehat{f^M}(S_1 \cup \cdots \cup S_M) = \widehat{g}(S_1 \cup \cdots \cup S_M)$. Then, clearly,

$$\sum_{(S_1,S_2,\ldots,S_M)\notin\mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \sum_T \widehat{g}(T)^2. \tag{11}$$

Moreover by Parseval's Theorem (Fact 13), $\sum_T \widehat{g}(T)^2 = \mathbb{E}_z[g(z)^2]$. Therefore,

$$\sum_{(S_1,S_2,\ldots,S_M)\notin\mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \mathbb{E}_z[g(z)^2]$$
$$= \Pr_z[z \notin T_\delta^{(M)}(\mathbf{C})] \leq \delta,$$

where the first inequality uses Eq. (11) and Parseval, the second equality is because $g(z)^2 \in \{0,1\}$ and the last inequality follows from Theorem 26 (ii). □

We are now ready to finally bound the Fourier entropy of f and prove the theorem. We need the following well-known trick to bound entropy when the underlying distribution has large weight on a small support. Fix $\mathcal{S} \subseteq 2^{[n]}$ such that $\sum_{S\notin\mathcal{S}} \widehat{g}(S)^2 = \delta$. In order to upper bound the Fourier entropy, we first express $\mathbb{H}(\hat{g}^2)$ as follows

$$\mathbb{H}(\hat{g}^2) = \sum_{S\in\mathcal{S}} \widehat{g}(S)^2 \log\left(\frac{1}{\widehat{g}(S)^2}\right) + \sum_{S\notin\mathcal{S}} \widehat{g}(S)^2 \log\left(\frac{1}{\widehat{g}(S)^2}\right).$$

We renormalize the first expression in the sum by $(1-\delta)$ and the second expression by δ. By doing so, we get

$$\begin{aligned}\mathbb{H}(\hat{g}^2) &= (1-\delta)\,\mathbb{H}\left(\frac{\widehat{g}(S)^2}{1-\delta} : S \in \mathcal{S}\right) + \delta\,\mathbb{H}\left(\frac{\widehat{g}(S)^2}{\delta} : S \notin \mathcal{S}\right) \\ &\quad + \sum_{S\in\mathcal{S}} \widehat{g}(S)^2 \log(1-\delta) + \sum_{S\notin\mathcal{S}} \widehat{g}(S)^2 \log(\delta) \\ &= (1-\delta)\,\mathbb{H}\left(\frac{\widehat{g}(S)^2}{1-\delta} : S \in \mathcal{S}\right) + \delta\,\mathbb{H}\left(\frac{\widehat{g}(S)^2}{\delta} : S \notin \mathcal{S}\right) + \mathsf{H}(\delta),\end{aligned} \tag{12}$$

where the equality used $\sum_{S\in\mathcal{S}} \widehat{g}(S)^2 = 1-\delta$ and we denote $\mathsf{H}(p) := p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$.

We now use Eq. (12) when applied to the function f^M and set $\mathcal{S} = \mathcal{B}$. We then obtain

$$\begin{aligned}M \cdot \mathbb{H}(\hat{f}^2) = \mathbb{H}(\widehat{f^M}^2) &\leq (1-\delta)\cdot\log|\mathcal{B}| + \delta\cdot\log|\{S : S\notin\mathcal{B}\}| + \mathsf{H}(\delta) \\ &\leq 2M(\mathsf{aUC}(f,\mathcal{C}) + \delta) + \delta n M + \mathsf{H}(\delta),\end{aligned}$$

where the second inequality used Claim 4.1 and 4.2. Dividing by M on both sides, we get

$$\mathbb{H}(\hat{f}^2) \leq 2\cdot(\mathsf{aUC}(f,\mathcal{C}) + \delta) + \delta n + \frac{\mathsf{H}(\delta)}{M}.$$

By the AEP theorem we know that $\delta \to 0$ as $M \to \infty$. Therefore, allowing $M \to \infty$ and taking the limit gives us the theorem. □

We now discuss the generalization of this theorem by replacing the $\mathsf{aUC}(f)$ upper bound by *average unambiguous parity-certificate complexity*. Recall that an unambiguous $\oplus$-certificate $\mathcal{C} = \{C_1,\ldots,C_t\}$ for f is a collection of monochromatic *affine subspaces* that together partition the space $\{-1,1\}^n$. (Observe that a subcube is a special type of affine subspace.) Analogously, the *average unambiguous $\oplus$-certificate complexity* of f with respect to $\mathcal{C}$, denoted $\mathsf{aUC}^{\oplus}(f,\mathcal{C})$, equals $\mathbb{E}_x[\mathsf{UC}^{\oplus}(\mathcal{C},x)]$ and $\mathsf{aUC}^{\oplus}(f) := \min_{\mathcal{C}}\ \mathsf{aUC}^{\oplus}(f,\mathcal{C})$. Let A_i be the set of parities fixed by C_i for $i \in [t]$. A parity is defined over a subset of variables and thus, naturally, can be viewed as a vector in $\{0,1\}^n$.

Like in the proof of Theorem 27, we study the M-fold iterated product of $\mathcal{C}$. In order to find a "small" set of coefficients where the Fourier weight is concentrated, we define $\mathcal{B}$ differently. The Fourier expansion of f^M, given by Eq. (10), suggests the following definition.

For a set $S \subseteq [n]$, define $\mathbb{1}_S \in \{0,1\}^n$ to be the indicator vector representing S (i.e., $\mathbb{1}_S(j) = 1$ if and only if $j \in S$). Let $(S_1, \ldots, S_M)$ be an M-tuple where each $S_i \subseteq [n]$. Then, we define $\mathcal{B}$ by letting $(S_1, \ldots, S_M) \in \mathcal{B}$ if and only if there exists a typical sequence $(C_{i_1}, \ldots, C_{i_M}) \in T_\delta^{(M)}(\mathbf{C})$ such that for all $j \in [M]$, $\mathbb{1}_{S_j} \in \mathsf{span}\langle A_{i_j} \rangle$ (where by $\mathsf{span}\langle A_{i_j} \rangle$, we mean the linear $\mathbb{F}_2$-span of parities in A_{i_j}, when viewed as vectors). We recall that A_{i_j} is the set of parities fixed by C_{i_j}. Observe that the earlier definition of $\mathcal{B}$ is now a special case of this. With this definition of $\mathcal{B}$ the rest of the proof follows similarly to establish the following generalization.

Theorem 28. *(Restatement of Theorem 8) Let $f : \{-1,1\}^n \to \{-1,1\}$ be any Boolean function. Then,*

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \mathsf{aUC}^{\oplus}(f, \mathcal{C}),$$

where $\mathcal{C}$ is any unambiguous $\oplus$-certificate of f.

As a corollary to the above theorems we obtain that the FEI conjecture holds for the class of functions f with constant $\mathsf{aUC}^{\oplus}(f)$, and $\mathsf{Inf}(f) \geq 1$.

Corollary 29. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function such that $\mathsf{Inf}(f) \geq 1$. Then,*

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \mathsf{aUC}^{\oplus}(f) \cdot \mathsf{Inf}(f).$$

We note that the reduction (Proposition E.2) in [WWW14] shows that removing the requirement $\mathsf{Inf}(f) \geq 1$ from the above corollary will prove the FEI conjecture for all Boolean functions with $\mathsf{Inf}(f) \geq \log n$. Furthermore, if we could show the FEI conjecture for Boolean functions f where $\mathsf{aUC}^{\oplus}(f) = \omega(1)$ is a slow-growing function of n, again the padding argument in [WWW14] shows that we would be able to establish the FEI conjecture for all Boolean functions.

4.1 Discussions on certificate complexity and Mansour's conjecture

An important consequence of the FEI conjecture, among many, is a positive answer to the long-standing conjecture of Mansour.

Conjecture 30 (Mansour's Conjecture [Man94])**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function that is representable by a t-term DNF. For every constant $\varepsilon > 0$, there exists a polynomial p over $\{-1,1\}$ with sparsity $\mathsf{poly}(t)$ such that $\mathbb{E}_x[(f(x) - p(x))^2] \leq \varepsilon$. (The exponent in $\mathsf{poly}(t)$ can depend on $1/\varepsilon$.)*

In fact, Mansour's original conjecture states that sparsity of the polynomial p (in the conjecture above) can be taken to be $t^{O(\log \frac{1}{\varepsilon})}$. Mansour's conjecture has a number of important consequences. For instance, Gopalan et al. [GKK08b] showed that a positive answer to Mansour's conjecture (Conjecture 30) would imply that DNF formulas can be agnostically learned in polynomial time up to any constant error parameter. This has been a long-standing open question [GKK08a] in computational learning theory.

In the earlier section, we saw that $\mathbb{H}(\hat{f}^2) \leq O(\mathsf{aUC}^{\oplus}(f))$. An interesting follow-up question is if one could strengthen this upper bound to $O(\min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\})$. In this section we observe that this bound on the Fourier-entropy in terms of $\mathsf{C}^0(f), \mathsf{C}^1(f)$ (which is clearly weaker than the FEI conjecture) suffices to establish Mansour's conjecture.

We remark that it was implicit in previous works [Kal07, GKK08a, KLW10, OWZ11] that one doesn't need the full power of the FEI conjecture to establish Mansour's conjecture. However, the following question: *What is the weakest form of the FEI conjecture that still implies Mansour's conjecture?*, was left unexplored. Our observation sharpens this relationship and establishes Mansour's conjecture as a natural step towards resolving the FEI conjecture.

We now formally state the weaker conjecture than the FEI conjecture that suffices to imply Mansour's conjecture.

Conjecture 31. *There exists a universal constant $\lambda > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\}.$$

It is weaker than the FEI conjecture because $\mathsf{Inf}(f) \leq \min\{\mathsf{C}^0(f), \mathsf{C}^1(f)\}$ [Bop97, Tra09, Ama11]. To establish the implication we will use the following equivalent form of Conjecture 31.

Conjecture 32. *There exists a universal constant $\lambda > 0$ such that for every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \mathsf{C}^1(f).$$

Before establishing the implication, we quickly argue that Conjecture 31 and 32 are equivalent. It is easily seen that Conjecture 31 implies Conjecture 32. For the reverse direction, note that both f and $\neg f$ have the same Fourier-entropy, while $\mathsf{C}^0(f)$ and $\mathsf{C}^1(f)$ reverse roles.

We now establish that Conjecture 32 implies Mansour's conjecture.

Proposition 33. *Conjecture 32 implies Conjecture 30.*

Proof. Let f be a t-term DNF and suppose $\delta_1, \delta_2 > 0$ are constants which we pick later. Let g be a Boolean function obtained from f by dropping all terms of length more than $\log(4t/\delta_1)$ in the DNF for f. Over the uniform distribution, each term of length greater than $\log(4t/\delta_1)$ equals 1 with probability at most $\delta_1/4t$. Then $g(x)$ and $f(x)$ differ only if x is accepted by a term of length greater than $\log(4t/\delta_1)$. Since there are at most t terms, by a union bound, we get

$$\mathbb{E}_x[(f(x) - g(x))^2] \leq 4 \cdot t \cdot \frac{\delta_1}{4t} = \delta_1. \tag{13}$$

Using Conjecture 32 for the function g, we get $\mathbb{H}(\hat{g}^2) \leq \lambda \cdot \mathsf{C}^1(g) \leq \lambda \log(4t/\delta_1)$. We now construct a polynomial p by defining its Fourier coefficients as follows:

$$\widehat{p}(S) = \begin{cases} \widehat{g}(S) & \text{if } |\widehat{g}(S)| \geq 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}, \\ 0 & \text{otherwise.} \end{cases}$$

By Parseval's identity (Fact 13), it follows that the number of non-zero Fourier coefficients in p is at most $2^{\mathbb{H}(\hat{g}^2)/\delta_2}$. Additionally we have that

$$\mathbb{E}_x[(g(x) - p(x))^2] = \sum_S (\widehat{g}(S) - \widehat{p}(S))^2 = \sum_{S\colon |\widehat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \widehat{g}(S)^2, \tag{14}$$

and

$$\begin{aligned} \mathbb{H}(\hat{g}^2) = \sum_S \widehat{g}(S)^2 \log \frac{1}{\widehat{g}(S)^2} &\geq \sum_{S\colon |\widehat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \widehat{g}(S)^2 \log \frac{1}{\widehat{g}(S)^2} \\ &\geq \frac{\mathbb{H}(\hat{g}^2)}{\delta_2} \sum_{S\colon |\widehat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \widehat{g}(S)^2. \end{aligned} \tag{15}$$

Putting together Eq. (14) and (15), we get

$$\mathbb{E}_x[(g(x)-p(x))^2] = \sum_{S\colon |\widehat{g}(S)|<2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \widehat{g}(S)^2 \leq \delta_2. \tag{16}$$

Let us now compute $\mathbb{E}[(f-p)^2]$.

$$\begin{aligned}
\mathbb{E}[(f-p)^2] &= \mathbb{E}[(f-g)^2] + \mathbb{E}[2(f-g)(g-p)] + \mathbb{E}[(g-p)^2], \\
&\leq \delta_1 + 2\sqrt{\mathbb{E}[(f-g)^2]}\sqrt{\mathbb{E}[(g-p)^2]} + \delta_2, \\
&\leq \delta_1 + 2\sqrt{\delta_1\delta_2} + \delta_2, \\
&= (\sqrt{\delta_1} + \sqrt{\delta_2})^2,
\end{aligned}$$

where the first inequality used the Cauchy-Schwarz inequality, Eq. (13) and (16). The second inequality also used Eq. (13) and (16). By picking $\delta_1 = \delta_2 = \varepsilon/4$ we get $\mathbb{E}[(f-p)^2] \leq \varepsilon$, which ensures that p has the approximation needed for Mansour's conjecture. Additionally, the Fourier sparsity of p is at most

$$2^{\mathbb{H}(\hat{g}^2)/\delta_2} \leq 2^{\lambda \log(4t/\delta_1)/\delta_2} = \left(\frac{16t}{\varepsilon}\right)^{\frac{4\lambda}{\varepsilon}}.$$

□

We end this section with another open problem that could form an intermediate step towards resolving Mansour's conjecture. The following seemingly weaker conjecture than Conjecture 31 is not known to imply Mansour's conjecture.

Conjecture 34. *There exists a universal constant $\lambda > 0$ such that for any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$,*

$$\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \max\{\mathsf{C}^0(f), \mathsf{C}^1(f)\} = \lambda \cdot \mathsf{C}(f).$$

5 Implications of the FEI conjecture

The FEI and FMEI conjecture seem to impose a strong constraint on the Fourier spectrum of a Boolean function. For example, the FMEI conjecture (if true) would show the existence of a large Fourier coefficient in the spectrum of a Boolean function that has small average sensitivity. In the introduction we saw that the FEI conjecture implies the existence of a Fourier-sparse polynomial p that approximates a Boolean function f in ℓ_2-distance, i.e., $\mathbb{E}_x[(f(x)-p(x))^2]$ is small. (A particular case being Mansour's conjecture, which is also a consequence of the FEI conjecture.) In this section we discuss one implication of the FEI conjecture relating to the structure of polynomials that approximate Boolean functions in the ℓ_∞-distance, i.e, $|p(x)-f(x)|$ is small for every $x \in \{-1,1\}^n$. In particular, we consider the question "Do polynomials approximating a Boolean function in ℓ_∞-distance satisfy some property?". In this direction, we make progress by showing that if the FEI conjecture is true, then we can rule out a polynomial with "large" Fourier sparsity from representing or approximating Boolean functions. In this section, we also consider a class of polynomials and show that no polynomial in that class can 1/8-approximate a Boolean function (without assuming that the FEI conjecture is true).

Definition 35. *An n-variate multilinear polynomial p is said to be a* flat polynomial *if all its* non-zero *Fourier coefficients have the same magnitude.*

Lemma 36. *Let $\varepsilon \in (0, 1/2)$ be a constant, and suppose the FEI conjecture is true. Let p be a flat polynomial with degree d and sparsity $T = 2^{\omega(d)}$. Then p cannot ε-approximate any Boolean function.*

Proof. In order to prove the lemma, we crucially use the following claim.

Claim 5.1. *Let p be a flat polynomial with sparsity T and, further, suppose p ε-approximates a Boolean function f, then*

$$\mathbb{H}(\hat{f}^2) \geq \Omega(\log T). \tag{17}$$

We assume this claim and conclude the proof of the lemma. By contradiction, let us assume that p ε-approximates a Boolean function f, so $\mathsf{deg}_\varepsilon(f) \leq d$. Assuming the FEI conjecture is true, we have $\mathbb{H}(\hat{f}^2) = O(\mathsf{Inf}(f))$. Furthermore, using a result of Shi [Shi00], we have $\mathsf{Inf}(f) \leq O(\mathsf{deg}_\varepsilon(f))$ for every f and constant ε. So, we have $\mathbb{H}(\hat{f}^2) \leq O(\mathsf{deg}_\varepsilon(f)) \leq O(d)$. Using Claim 5.1, it follows that

$$\Omega(\log T) \leq \mathbb{H}(\hat{f}^2) \leq O(d).$$

But this upper bound of $T = 2^{O(d)}$ contradicts the assumption on T in the statement of the lemma. Hence, we conclude that p cannot ε-approximate any Boolean function.

Proof of Claim 5.1. Suppose p ε-approximates a Boolean function f. Without loss of generality suppose that all Fourier coefficients of p have magnitude $\alpha/\sqrt{T}$ for some $\alpha \in [(1-\varepsilon), (1+\varepsilon)]$. Such an α exists because, by Parseval's identity (Fact 13), we have $\sum_S \widehat{p}(S)^2 = \mathbb{E}_x[p(x)^2] \in [(1-\varepsilon)^2, (1+\varepsilon)^2]$ (where the inclusion assumes p ε-approximates f). Now consider the following set $\mathcal{A}$ of "large" Fourier coefficients,

$$\mathcal{A} := \left\{ S \ : \ |\widehat{f}(S)| \geq \frac{2\alpha}{\sqrt{T}} \right\}. \tag{18}$$

For every $S \in \mathcal{A}$, we have $|\widehat{f}(S) - \widehat{p}(S)| \geq |\widehat{f}(S)|/2$ and hence

$$\sum_{S \in \mathcal{A}} \widehat{f}(S)^2 \leq 4 \sum_{S \in \mathcal{A}} (\widehat{f}(S) - \widehat{p}(S))^2 \leq 4 \sum_{S \subseteq [n]} (\widehat{f}(S) - \widehat{p}(S))^2 = 4\, \mathbb{E}_x[(f(x) - p(x))^2] \leq 4\varepsilon^2, \tag{19}$$

where the equality uses Parseval's identity (Fact 13) and the last inequality uses that p ε-approximates f. From Eq. (19), we have $\sum_{S \notin \mathcal{A}} \widehat{f}(S)^2 \geq 1 - 4\varepsilon^2$. This gives us our desired lower bound on the Fourier entropy of f,

$$\begin{aligned}
\mathbb{H}(\hat{f}^2) = \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} \geq \sum_{S \notin \mathcal{A}} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} &\geq \sum_{S \notin \mathcal{A}} \widehat{f}(S)^2 \cdot \log \left(\frac{T}{4\alpha^2} \right) \\
&\geq (1 - 4\varepsilon^2) \log \left(\frac{T}{4\alpha^2} \right) \\
&\geq (1 - 4\varepsilon^2) \log \frac{T}{4(1+\varepsilon)^2} = \Omega(\log T).
\end{aligned}$$

The second inequality follows by the definition of $\mathcal{A}$ (in Eq. (18)) because for every $S \notin \mathcal{A}$ we have $|\widehat{f}(S)| < \frac{2\alpha}{\sqrt{T}}$, the third inequality is by negation of the inequalities in Eq. (19), the last inequality holds because $\alpha \leq (1+\varepsilon)$ by the definition of α, and the last equality assumes $\varepsilon < 1/2$. ☐

The proof of the claim concludes the proof of the lemma. ☐

Since we are still unable to resolve the FEI conjecture, an interesting intermediate question would be to *unconditionally* prove the following conjecture.

Conjecture 37. *No flat polynomial of degree d and sparsity $2^{\omega(d)}$ can ε-approximate a Boolean function.*

Although we have not been able to resolve this conjecture, we now discuss some partial progress towards resolving it. Additionally, we give an intriguing connection between this conjecture and the Bohnenblust-Hille inequality.

Partial progress towards resolving Conjecture 37. A first step towards disproving the conjecture would be to show that no flat polynomial of degree d and sparsity $\binom{n}{d}$ can approximate a Boolean function. We now show that this in fact already follows from results of Tal [Tal14, Tal17]. Among other results, Tal [Tal14, Claim 2.13] showed that, if a Boolean function f can be $1/3$-approximated in the ℓ_∞-distance by a degree-d polynomial, then it has exponentially decreasing Fourier tails above level $O(d)$. Now using the results[9] in [Tal17], we know

$$\sum_{S\colon |S|=d} |\widehat{f}(S)| \leq O(d)^d. \tag{20}$$

Suppose p is a degree-d flat polynomial where all degree-d multilinear monomials have non-zero coefficients. That is, the sparsity of p, denoted T, equals $\binom{n}{d}$. Furthermore, suppose that p $1/3$-approximates f in ℓ_∞-norm. This implies $|\widehat{p}(S)| = \alpha/\sqrt{T}$ for some $\alpha \in [2/3, 4/3]$. Then, we have

$$\sum_{S\colon |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \leq \left(\sum_{S\colon |S|=d} |\widehat{p}(S) - \widehat{f}(S)|^2\right)^{1/2} \cdot \sqrt{\binom{n}{d}} \leq \frac{1}{3} \cdot \sqrt{\binom{n}{d}},$$

where the first inequality is Cauchy-Schwarz and the second uses $|p(x) - f(x)| \leq 1/3$ for all x. On the other hand,

$$\begin{aligned}\sum_{S\colon |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \geq \sum_{S\colon |S|=d} (|\widehat{p}(S)| - |\widehat{f}(S)|) &= \sum_{S\colon |S|=d} |\widehat{p}(S)| - \sum_{S\colon |S|=d} |\widehat{f}(S)| \\ &\geq \frac{2}{3}\sqrt{\binom{n}{d}} - \sum_{S\colon |S|=d} |\widehat{f}(S)|,\end{aligned}$$

where the first inequality uses the reverse-triangle inequality and the last uses the lower bound on $|\widehat{p}(S)|$. From the above two inequalities we get

$$\sum_{S\colon |S|=d} |\widehat{f}(S)| \geq \frac{1}{3}\sqrt{\binom{n}{d}}.$$

This contradicts Eq. (20) when $d = o(n^{1/3})$. Thus, p cannot $1/3$-approximate a Boolean function. However, our conjecture asks if a similar result also holds when $T = 2^{\omega(d)}$.

Instead of considering arbitrary flat polynomials, we consider a restricted class of polynomials which are referred to as *block-multilinear polynomials*. An n-variate polynomial is said to be *block-multilinear* if the input variables can be *partitioned* into disjoint blocks $A_1, \ldots, A_d \subseteq \{x_1, \ldots, x_n\}$

[9] See [Tal17, Lemmas 29 and 34] for a precise statement of his results.

such that every monomial in the polynomial has *at most* one variable from each block. For the purposes of this paper we will assume that each block is of the same size. In other words, a block-multilinear polynomial $p : (\mathbb{R}^n)^d \to \mathbb{R}$ can be written as

$$p(x^1, \ldots, x^d) = \sum_{S \subseteq [nd] \,:\, \forall j \in [d],\, |S \cap x^j| \leq 1} \widehat{p}(S) \prod_{l \in S} x_l, \tag{21}$$

where $\widehat{p}(S) \in \mathbb{R}$ for every $S \subseteq [nd]$. Note that this is the standard Fourier decomposition of p if $x^i \in \{-1, 1\}^n$ for every $i \in [d]$. Clearly such a block-multilinear polynomial has degree at most d and sparsity at most $(n+1)^d$. Such polynomials have found applications in quantum computing [AA18, Mon12], classical and quantum XOR games [BBLV13], polynomial decoupling [OZ16] and in functional analysis which we discuss later. Our main contribution in this section is that we show a positive answer to Conjecture 37 for the class of flat block-multilinear polynomials.

Theorem 38. *(Restatement of Theorem 11) If p is an n-variate flat block-multilinear polynomial with degree d and sparsity $2^{\omega(d)}$, then p cannot $1/8$-approximate a Boolean function.*

We defer the proof of this theorem to the next section. We now continue with the relevance of block-multilinear polynomials and the theorem above to functional analysis literature.

Relation between Theorem 38 and the Bohnenblust-Hille (BH) inequality. Consider the block-multilinear polynomial $p : (\mathbb{R}^n)^d \to \mathbb{R}$ defined in Eq. (21), but we further assume that it has no monomial of degree $< d$. One way to show that p is not $1/8$-close to a Boolean function would be to show that there exists x' such that $|p(x')|$ is greater than $9/8$. Understanding if such an x' exists for such a polynomial p can be cast as the following maximization problem

$$\|p\| := \max_{x^1, \ldots, x^d \in [-1,1]^n} \Big| \sum_{i_1, \ldots, i_d = 1}^{n} \widehat{p}_{i_1, \ldots, i_d} x^1_{i_1} \cdots x^d_{i_d} \Big|. \tag{22}$$

In order to develop an intuition for the maximization problem, consider the case $d = 2$ and furthermore suppose $\widehat{p}_{i_1, i_2} \in \{-1, 1\}$ and $x^i \in \{-1, 1\}^n$, then giving a lower bound on $\|p\|$ is well-known in computer science as the so-called *unbalancing lights* problem (see [AS00, Section 2.5], where they show the existence of sign vectors such that $\|p\| \geq \sqrt{n}$). For larger d and arbitrary p, showing lower bounds on $\|p\|$ in Eq. (22) has been extensively studied in the functional analysis literature and is sometimes referred to as the "generalized unbalancing lights problem".

The first paper giving a lower bound to Eq. (21) was by H. F. Bohnenblust and E. Hille [BH31] in 1931. They gave a lower bound on the injective tensor norm of degree-d multilinear forms, which in our context translates to a lower bound on $\|p\|$. To be precise, their result states the following: for every n, d there exists a constant $C_d \geq 1$ such that, for every degree-d block-multilinear polynomial $p : (\mathbb{R}^n)^d \to \mathbb{R}$, we have

$$\Big(\sum_{i_1, \ldots, i_d = 1}^{n} |\widehat{p}_{i_1, \ldots, i_d}|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} \leq C_d \cdot \max_{x^1, \ldots, x^d \in [-1,1]^n} |p(x^1, \ldots, x^d)|. \tag{23}$$

The result of [BH31] showed that it suffices to pick

$$C_d = d^{\frac{d+1}{2d}} 2^{\frac{d-1}{2}}$$

in order to satisfy Eq. (23). In addition, their bound on C_d recovers the well-known Littlewood's $4/3$ inequality [Lit30] when $d = 2$. Since their seminal work, a lot of research in the functional analysis

literature (for the last 80 years!) has been in finding the optimal BH-*constants*, i.e., the smallest C_d for which Eq. (23) holds. We cite a few results [Mon12, ABPSS14, PE18, DFOC+11, DPS10], referring the reader to the references within these papers for more. It is a long-standing open question if C_d is a universal constant (there has been recent work [PE18] giving numerical evidence that this is the case).

After a series of works, Pellegrino and Seoane-Sepúlveda showed [PSS12] that it suffices to pick $C_d = \text{poly}(d)$. As far as we are aware, the best upper bound on C_d was shown by Diniz et al. [DMFPSS12] as $C_d = O(\sqrt{d})$. We also point the interested reader to [Mon12, Theorem 17] for a suboptimal, yet elegant proof that shows that it suffices to pick $C_d = O(d^{1.45})$ in order to satisfy Eq. (23).

We now discuss the relevance of the BH-inequality to Theorem 38. Consider an arbitrary flat block-multilinear polynomial $p : (\{-1,1\}^n)^d \to \mathbb{R}$ with degree d and sparsity T that $1/8$-approximates a Boolean function. That is, every non-zero Fourier coefficient $|\widehat{p}_{i_1,\ldots,i_d}|$ equals $\alpha/\sqrt{T}$ for some $\alpha \in [7/8, 9/8]$. Then using Eq. (23), we get

$$\left(\left(\frac{7/8}{\sqrt{T}}\right)^{\frac{2d}{d+1}} \cdot T\right)^{\frac{d+1}{2d}} \leq C_d \cdot \|p\| \leq (9/8) \cdot C_d.$$

With further simplification, we have

$$T \leq (9/7)^{2d} \cdot C_d^{2d}.$$

Using the result of Diniz et al. [DMFPSS12], $C_d = O(\sqrt{d})$, we get that the sparsity $T \leq 2^{O(d \log d)}$. However, from the FEI conjecture (cf. Claim 5.1) it follows that $T \leq 2^{O(d)}$. Thus, using the current best bound on C_d we cannot conclude Theorem 38. However, if the long-standing open question of C_d being a universal constant were true, then Theorem 38 follows. But proving C_d a universal constant seems to be a very hard problem. Nevertheless, in the next section, we will prove Theorem 38, while circumventing the barrier of improving the upper bound on C_d.

We remark that Theorem 38 does not say anything about the BH-inequality, since Theorem 38 states that a flat block-multilinear polynomial either takes a value in the range $(-1/8, 1/8)$ or takes a value of magnitude more that $9/8$, while the BH-inequality states that such a polynomial definitely takes a value of high magnitude ($> 9/8$) on at least one input.[10]

5.1 Proof of Theorem 38

We restate the theorem for convenience.

Theorem 38 (restated). *If p is an n-variate flat block-multilinear polynomial with degree d and sparsity $2^{\omega(d)}$, then p cannot $1/8$-approximate a Boolean function.*

Without loss of generality, let us assume d divides n. Let $\{A_1, \ldots, A_d\}$ be a partition of $\{x_1, \ldots, x_n\}$ and for simplicity suppose $A_1 = \{x_1, \ldots, x_k\}$ for $k = n/d$. Let $B = \{x_1, \ldots, x_n\} \backslash A_1$ denote the set of remaining variables. Then the monomials of p can be divided into those containing variables in A_1 and those independent of variables in A_1 and we can write p as follows

$$p(x) = q_0(x_B) + \sum_{i=1}^{k} x_i q_i(x_B), \tag{24}$$

[10] It is not clear to us if flat polynomials considered in Theorem 38 can take values close to 0, or if it is possible to show that flat polynomials *always* take a large value outside the Boolean cube.

where $q_0, \ldots, q_k$ are polynomials of degree at most $d-1$. From here on, for notational simplicity we simply rewrite x_B as z.

Lemma 39. *Let p be a degree-d block-multilinear polynomial that $1/8$-approximates a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$. Further, let $q_0, \ldots, q_k$ be as defined in Eq. (24). Then,*

(i) For every $z \in \{-1,1\}^{n-k}$, we have $q_i(z) \in [-9/8, -7/8] \cup [-1/8, 1/8] \cup [7/8, 9/8]$, i.e., q_i is a $\frac{1}{8}$-approximation to a $\{-1,0,1\}$-valued function.

(ii) For every $z \in \{-1,1\}^{n-k}$, there exists a unique $j \in \{0, \ldots, k\}$ which satisfies

$$|q_j(z)| \geq 7/8 \quad \text{and} \quad \sum_{i \neq j} |q_i(z)| \leq 1/8.$$

Proof. The proof of the first part is fairly straightforward, while the second part requires some calculations.

Proof of (i)**.** We first rewrite the $q_1, \ldots, q_k$ as follows:

$$q_i(z) = \frac{p(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k, z) - p(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_k, z)}{2},$$

for $i \in [k]$ and similarly we rewrite q_0 as

$$q_0(z) = \frac{p(x_1, \ldots, x_k, z) + p(-x_1, \ldots, -x_k, z)}{2}.$$

Now the first part follows from the fact that p is a $1/8$-approximation to the Boolean function f. So the q_is are a $1/8$-approximation to a $\{-1,0,1\}$-valued function.

Proof of (ii)**.** Fix $z \in \{-1,1\}^{n-k}$. First observe that for every $x_1, \ldots, x_k \in \{-1,1\}$, we have

$$p(x_1, \ldots, x_k, z) \in \underbrace{\left[q_0(z) - \sum_{i=1}^{k} |q_i(z)|,\ q_0(z) + \sum_{i=1}^{k} |q_i(z)| \right]}_{:= \mathcal{I}(z)}. \tag{25}$$

Furthermore, there exists a choice of $x_1, \ldots, x_k \in \{-1,1\}$ such that p evaluates to either one of the end points in the interval $\mathcal{I}(z)$. Moreover $\mathcal{I}(z)$ satisfies *exactly* one of the following containments:

(a) $\mathcal{I}(z) \subseteq [7/8, 9/8]$, or

(b) $\mathcal{I}(z) \subseteq [-9/8, -7/8]$, or

(c) $\mathcal{I}(z)$ intersects both the intervals $[-9/8, -7/8]$ and $[7/8, 9/8]$.

We now prove that in all the three cases, there exists $j \in \{0, \ldots, k\}$ such that $|q_j(z)| \geq 7/8$ and $\sum_{i \neq j} |q_i(z)| \leq 1/8$, hence proving the lemma statement.

Assuming Case (a) **and** (b)**.** The proofs for Case (a) and (b) are exactly the same, so we prove the lemma assuming $\mathcal{I}(z) \subseteq [7/8, 9/8]$. In this case, observe that $q_0(z) - \sum_{i=1}^{k} |q_i(z)| \geq 7/8$

and $q_0(z) + \sum_{i=1}^{k} |q_i(z)| \leq 9/8$. Clearly this implies $q_0(z) \geq 7/8$. Furthermore, from both the inequalities it follows that

$$\sum_{i=1}^{k} |q_i(z)| \leq \min\{q_0(z) - 7/8, 9/8 - q_0(z)\}.$$

Using the fact that $q_0(z) \in [7/8, 9/8]$, it follows that $\sum_{i=1}^{k} |q_i(z)| \leq 1/8$, which concludes the proof for Case (a).

Assuming Case (c)**.** Recall the definition of $p(x_1, \ldots, x_k, z)$ from Eq. (25). Let $a, b \in \{-1, 1\}^k$ be such that

$$p(a, z) = q_0(z) - \sum_{i=1}^{k} |q_i(z)| \quad \text{and } p(b, z) = q_0(z) + \sum_{i=1}^{k} |q_i(z)|. \tag{26}$$

It is not hard to see that $a = -b$. Consider a path from a to b on the Boolean hypercube. Since $p(a, z) \in [-9/8, -7/8]$ and $p(b, z) \in [7/8, 9/8]$, there exists an edge[11] on this path such that the value of p on this edge jumps from the interval $[-9/8, -7/8]$ to $[7/8, 9/8]$. For now the existence of such an edge is sufficient, below we explicitly construct such an edge. Let us denote the direction of this edge by $j \in [k]$. We now claim that $|q_j(z)| \geq 7/8$ and $\sum_{i \neq j} |q_i(z)| \leq 1/8$.

$|q_j(z)| \geq 7/8$ follows immediately, because the change in p on this edge is $2 \cdot |q_j(z)|$ (by Eq. (26)), and the change of value of p on this edge is $\geq 2 \cdot (7/8)$. Thus we have $|q_j(z)| \geq 7/8$.

We now prove $\sum_{i \neq j} |q_i(z)| \leq 1/8$ by considering two cases based on whether $\mathsf{sign}(q_0(z))$ is positive or negative. Since the proofs for both cases are similar, for simplicity we prove it assuming $\mathsf{sign}(q_0(z))$ is positive. We now define the edge between a, b where the value of p jumps from the interval $[-9/8, -7/8]$ to $[7/8, 9/8]$. Consider the edge in the jth direction given by setting all the variables except x_j as follows: $x_i = \mathsf{sign}(q_i(z))$ for $i \in [k] \setminus \{j\}$. Clearly the value of p only depends on x_j and equals $\sum_{i \neq j} |q_i(z)| + x_j q_j(z)$. When $x_j = \mathsf{sign}(q_j(z))$, p takes the value $\sum_{i \neq j} |q_i(z)| + |q_j(z)|$ which is in the interval $[7/8, 9/8]$ and

$$\sum_{i \neq j} |q_i(z)| + |q_j(z)| \in [7/8, 9/8], \text{ implies } \sum_{i \neq j} |q_i(z)| \leq 9/8 - |q_j(z)|. \tag{27}$$

Similarly, when $x_j = -\mathsf{sign}(q_j(z))$ then p takes the value $\sum_{i \neq j} |q_i(z)| - |q_j(z)|$ which is in the interval $[-9/8, -7/8]$, and

$$\sum_{i \neq j} |q_i(z)| - |q_j(z)| \in [-9/8, -7/8], \text{ implies } \sum_{i \neq j} |q_i(z)| \leq |q_j(z)| - 7/8. \tag{28}$$

From Eq. (27) and (28) we have $\sum_{i \neq j} |q_i(z)| \leq 1/8$.

The uniqueness of j in each case follows because $\sum_{i=0}^{k} |q_i(z)| \leq 9/8$. □

We now show that if $p(x)$ $\frac{1}{8}$-approximates a Boolean function f then $\mathsf{deg}(f) \leq d$.

Lemma 40. *Let p be a degree-d block-multilinear polynomial that $\frac{1}{8}$-approximates a Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$. Then,*

$$\mathsf{deg}(f) \leq d.$$

[11] An *edge* in the *direction* of $j \in [k]$ on the Boolean hypercube refers to a tuple $(w, w^{(j)})$ where $w, w^{(j)} \in \{-1, 1\}^k$ and $w^{(j)}$ is the bit string obtained by flipping the sign of the jth bit of w.

Proof. We will establish the proof by induction on d.

Base case $d = 1$. Let $p(x) = a_0 + \sum_{i=1}^{n} a_i x_i$, where $a_i \in \mathbb{R}$. Since p approximates f, it follows that $\mathsf{sign}(p(x)) = f(x)$ for every x. We now express $\mathsf{sign}(p(x))$ as a polynomial of degree at most 1. From Lemma 39 (ii), it follows that there exists a unique $i \in \{0\} \cup [n]$ such that

$$\mathsf{sign}(p(x)) = \begin{cases} \mathsf{sign}(a_i) \cdot x_i & \text{if } i \in [n], \\ \mathsf{sign}(a_0) & \text{otherwise.} \end{cases}$$

Thus the base case follows.

Inductive assumption. Assume the lemma statement for polynomials p of degree at most $d-1$.

Inductive step. Consider the decomposition of $p(x_1, \ldots, x_n)$

$$p(x) = q_0(z) + \sum_{i=1}^{k} x_i q_i(z).$$

Recall from the proof of Lemma 39 (i), that $q_i(z)$ can be expressed as a difference between two $(d-1)$-block-multilinear polynomials by fixing $x_1, \ldots, x_k$. That is,

$$\begin{aligned} q_0(z) &= \frac{p(1, \ldots, 1, z) + p(-1, \ldots, -1, z)}{2}, \\ q_i(z) &= \frac{p(1, \ldots, 1, 1, 1, \ldots, 1, z) - p(1, \ldots, 1, -1, 1, \ldots, 1, z)}{2} \qquad \text{for } i \in [k], \end{aligned} \tag{29}$$

where the -1 in the final expression $p(1, \ldots, 1, -1, 1, \ldots, 1, z)$ is at the ith coordinate. Clearly, it follows that $p(-1, \ldots, -1, z)$ and $p(1, \ldots, 1, z)$ are block-multilinear polynomials with $d-1$ parts that $\frac{1}{8}$-approximate Boolean functions $f(-1, \ldots, -1, z)$ and $f(1, \ldots, 1, z)$, respectively. Therefore, by the inductive assumption, both $\mathsf{deg}(f(-1, \ldots, -1, z))$ and $\mathsf{deg}(f(1, \ldots, 1, z))$ are $\leq d-1$. Additionally the function $\widetilde{q}_0$ defined as

$$\widetilde{q}_0(z) = \frac{f(1, \ldots, 1, z) + f(-1, \ldots, -1, z)}{2}.$$

is a $\{-1, 0, 1\}$-valued function satisfying $\mathsf{deg}(\widetilde{q}_0) \leq d-1$. Moreover, q_0 (as defined in Eq. (29)) is a $\frac{1}{8}$-approximation to $\widetilde{q}_0$. In a similar fashion, one can define $\widetilde{q}_i$ for all $i \in [k]$ which is a $\{-1, 0, 1\}$-valued function of degree at most $d-1$ that additionally satisfies that q_i (as defined in Eq. (29)) is a $\frac{1}{8}$-approximation to $\widetilde{q}_i$. Finally consider the polynomial $\widetilde{p}$ defined as follows,

$$\widetilde{p}(x) = \widetilde{q}_0(z) + \sum_{i=1}^{k} x_i \widetilde{q}_i(z).$$

Firstly note that $\widetilde{p}(x)$ is a polynomial that takes value in $\{-1, +1\}$ for all $x \in \{-1, +1\}^n$. This is because for all i, $q_i(z)$ $\frac{1}{8}$-approximates $\tilde{q}_i(z)$ and so from Lemma 39, for any given z there exists exactly one i such that $\tilde{q}_i(z)$ takes a non-zero value. So for a given x if j is the unique index from Lemma 39 such that $|q_j(z_x)| \geq 7/8$, then for that x we have $p(x) = x_j q_j(z)$.

From this observation we also get $|f(x) - \widetilde{p}(x)| \leq 1/4$ for any x. This is because for a given x if j is the unique index from Lemma 39 such that $|q_j(z_x)| \geq 7/8$, then $|f(x) - x_j q_j(z_x)| \leq 1/8$ (because of Lemma 39 (ii)) and $|x_j q_j(z_x) - x_j \tilde{q}_j(z_x)| \leq 1/8$ (by definition of $\tilde{q}_j$) and $p(x) = x_j q_j(z_x)$. Since both f and $\tilde{p}$ are $\{-1, +1\}$-valued functions and since $|f(x) - \tilde{p}(x)| \leq 1/4$, it follows that $\widetilde{p}(x) = f(x)$.

By construction, the degree of $\widetilde{p}$ is at most d. The lemma now follows as $\widetilde{p}(x) = f(x)$. ☐

Using Claim 5.1 and Lemma 40, it now follows that if p is a degree-d flat block-multilinear polynomial that 1/8-approximates a Boolean function, then the sparsity of p is at most $2^{O(d)}$. This completes the proof of Theorem 38.

6 Open probems

We list a few open problems which we believe are structurally interesting and could lead towards proving the FEI or FMEI conjecture. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function.

1. Does there exist a Fourier coefficient $S \subseteq [n]$ such that $|\widehat{f}(S)| \geq 2^{-O(\mathsf{deg}_{1/3}(f))}$? This would show $\mathbb{H}_\infty(\hat{f}^2) \leq O(\mathsf{deg}_{1/3}(f))$.

2. Does there exist a polynomial p with degree $d = O(\mathsf{deg}_{1/3}(f))$ such that $|p(x) - f(x)| \leq 1/3$ for every $x \in \{-1,1\}^n$ and $\sum_S |\widehat{p}(S)| \leq 2^{O(d)}$? In fact, can we even prove an upper bound of $2^{O(d \log d)}$ on the spectral norm of p?

3. Does there exist a universal constant $\lambda > 0$ such that $\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \min\{\mathsf{C}^1(f), \mathsf{C}^0(f)\}$? This would resolve Mansour's conjecture.

4. Can we show $\mathbb{H}(\hat{f}^2) \leq O(Q(f))$? (where $Q(f)$ is the 1/3-error quantum query complexity of f, which Beals et al. [BBC+01] showed to be at least $\mathsf{deg}_{1/3}(f)/2$).

Acknowledgements. Part of this work was carried out when NS and SC visited CWI, Amsterdam and SA visited University of Bristol (partially supported by EPSRC grant EP/L021005/1). SA thanks Ashley Montanaro for his hospitality. NS and SC would like to thank Satya Lokam for many helpful discussions on the Fourier entropy-Influence conjecture. SA and SC thank Jop Briët for pointing us to the literature on unbalancing lights and many useful discussions regarding Section 5. We also thank Penghui Yao and Avishay Tal for discussions during the course of this project.

References

[AA18] S. Aaronson and A. Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM Journal of Computing*, 47(3):982–1038, 2018. Earlier in STOC'15. arXiv:1411.5729. 9,28

[ABG+14] A. Akavia, A. Bogdanov, S. Guo, A.Kamath, and A.Rosen. Candidate weak pseudorandom functions in $AC^0 \circ MOD2$. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS'14, pages 251–260. ACM, 2014. 4

[ABPSS14] N. Albuquerque, F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. Sharp generalizations of the multilinear Bohnenblust-Hille inequality. *Journal of Functional Analysis*, 266(6):3276–3740, 2014. arXiv:1306.3362. 9,29

[AHKÜ17] S. R. Allen, L. Hellerstein, D. Kletenik, and T. Ünlüyurt. Evaluation of monotone DNF formulas. *Algorithmica*, 77(3):661–685, 2017. 8

[AKK16] A. Ambainis, M. Kokainis, and R. Kothari. Nearly optimal separations between communication (or query) complexity and partitions. In *31st Conference on Computational Complexity, CCC 2016*, pages 4:1–4:14, 2016. Combines arXiv:1512.01210 and arXiv:1512.00661. [6,19]

[Ama11] K. Amano. Tight bounds on the average sensitivity of k-CNF. *Theory of Computing*, 7(4):45–48, 2011. [24]

[AS00] N. Alon and J. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization, 2000. [28]

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98. quant-ph/9802049. [33]

[BBLV13] J. Briët, H. Buhrman, T. Lee, and T. Vidick. Multipartite entanglement in XOR games. *Quantum Information & Computation*, 13(3-4):334–360, 2013. arXiv:0911.4007. [9,28]

[BH31] H. F. Bohnenblust and E. Hille. On the absolute convergence of Dirichlet series. *Annals of Mathematics*, pages 600–622, 1931. [9,28]

[BHT17] S. Ben-David, P. Hatami, and A. Tal. Low-sensitivity functions from unambiguous certificates. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017*, pages 28:1–28:23, 2017. arXiv:1605.07084. [6]

[BK97] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis (GAFA)*, 7(3):438–461, 1997. [2]

[BOH90] Y. Brandman, A. Orlitsky, and J. Hennessy. A spectral lower bound technique for the size of decision trees and two-level and/or circuits. *IEEE Transactions of Computers*, 39(2):282–287, 1990. [17]

[Bop97] R. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997. [4,24]

[BT13] M. Bun and J. Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013*, pages 303–314, 2013. arXiv:1302.6191. [15]

[BW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002. [11]

[CGJ+18] M. Cheraghchi, E. Grigorescu, B. Juba, K. Wimmer, and N. Xie. $AC^0 \circ$ MOD2 lower bounds for the Boolean inner product. *Journal of Computer and System Sciences*, 97:45 – 59, 2018. [4]

[CKK+18] S. Chakraborty, S. Karmalkar, S. Kundu, S. V. Lokam, and N. Saurabh. Fourier entropy-influence conjecture for random linear threshold functions. In *LATIN 2018: Theoretical Informatics - 13th Latin American Symposium, 2018*, pages 275–289, 2018. [5]

[CKLS16] S. Chakraborty, R. Kulkarni, S.V. Lokam, and N. Saurabh. Upper bounds on Fourier entropy. *Theoretical Computer Science*, 654:92–112, 2016. Earlier version appeared in COCOON 2015. [1,4,19]

[CS16] G. Cohen and I. Shinkar. The complexity of DNF of parities. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 47–58. ACM, 2016. [4]

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991. [19]

[DFOC+11] A. Defant, L. Frerick, J. Ortega-Cerdá, M. Ounaïes, and K. Seip. The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive. *Annals of Mathematics*, 174(1):485–497, 2011. arXiv:0904.3540. [9,29]

[DMFPSS12] D. Diniz, G. A. Munoz-Fernndez, D. Pellegrino, and J. B. Seoane-Sepúlveda. The asymptotic growth of the constants in the Bohnenblust-Hille inequality is optimal. *Journal of Functional Analysis*, 263(2):415–428, 2012. arXiv:1108.1550. [9,29]

[DPS10] A. Defant, D. Popa, and U. Schwarting. Coordinatewise multiple summing operators in banach spaces. *Journal of Functional Analysis*, 259(1):220–242, 2010. [9,29]

[DPV11] B. Das, M. Pal, and V. Visavaliya. The entropy influence conjecture revisited, 2011. arxiv:1110.4301. [5]

[FK96] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996. [1,2,4]

[Fri98] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998. [3]

[GKK08a] P. Gopalan, A. Kalai, and A. R. Klivans. A query algorithm for agnostically learning DNF? In *21st Annual Conference on Learning Theory - COLT 2008*, pages 515–516, 2008. [8,23,24]

[GKK08b] P. Gopalan, A. T. Kalai, and A. Klivans. Agnostically learning decision trees. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 527–536, 2008. [3,23]

[Göö15] M. Göös. Lower bounds for clique vs. independent set. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015*, pages 1066–1076, 2015. [6]

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. [4]

[GSTW16] P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson. Degree and sensitivity: Tails of two distributions. In *31st Conference on Computational Complexity, CCC 2016*, pages 13:1–13:23, 2016. arxiv: 1604.07432. [4]

[Hod17] R. Hod. Improved lower bounds for the Fourier entropy/influence conjecture via lexicographic functions, 2017. arxiv:1711.00762. [5]

[Kal07] G. Kalai. The entropy/influence conjecture. Terence Tao's blog: https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjectur 2007. 4,8,24

[KKL88] J. Kahn, G. Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988. 2,11,13,17

[KLW10] A. Klivans, H. Lee, and A. Wan. Mansour's conjecture is true for random DNF formulas. In *Proceedings of the 23rd Conference on Learning Theory*, pages 368–380, 2010. 4,8,24

[KMS12] N. Keller, E. Mossel, and T. Schlank. A note on the entropy/influence conjecture. *Discrete Mathematics*, 312(22):3364 – 3372, 2012. arXiv:1105.2651. 4,6

[Lit30] J. E. Littlewood. On bounded bilinear forms in an infinite number of variables. *The Quarterly Journal of Mathematics*, 1:164–174, 1930. 9,28

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993. 4

[LS09] T. Lee and A. Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009. 5

[Man94] Y. Mansour. Learning Boolean functions via the Fourier transform. In V. Roychowdhury, K-Y Siu, and A. Orlitsky, editors, *Theoretical Advances in Neural Computation and Learning*, pages 391–424. Springer US, 1994. 23

[Man95] Y. Mansour. An $n^{O(\log\log n)}$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995. 1,3

[MO09] A. Montanaro and T. Osborne. On the communication complexity of XOR functions, 2009. arXiv:0909.3392. 5

[Mon12] A. Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12):122206, 2012. arXiv:1208.0161. 9,28,29

[O'D14] R. O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 1,2,3,10,13

[OT13] R. O'Donnell and L-Y. Tan. A composition theorem for the Fourier entropy-influence conjecture. In *Proceedings of Automata, Languages and Programming - 40th International Colloquium*, pages 780–791, 2013. arXiv:1304.1347. 4,5

[OWZ11] R. O'Donnell, J. Wright, and Y. Zhou. The Fourier entropy-influence conjecture for certain classes of Boolean functions. In *Proceedings of Automata, Languages and Programming - 38th International Colloquium*, pages 330–341, 2011. 1,3,4,5,8,13,24

[OWZ+14] R. O'Donnell, J. Wright, Y. Zhao, X. Sun, and L-Y. Tan. A composition theorem for parity kill number. In *IEEE 29th Conference on Computational Complexity, CCC 2014*, pages 144–154, 2014. arXiv:1312.2143. 5

[OZ16] R. O'Donnell and Y. Zhao. Polynomial bounds for decoupling, with applications. In *31st Conference on Computational Complexity, CCC 2016*, pages 24:1–24:18, 2016. arXiv:1512.01603. 9,28

[PE18] D. Pellegrino and V. T. Eduardo. Towards sharp Bohnenblust-Hille constants. *Communications in Contemporary Mathematics*, 20(3):1750029, 2018. arXiv:1604.07595. 9,29

[PSS12] D. Pellegrino and J. Seoane-Sepúlveda. New upper bounds for the constants in the Bohnenblust-Hille inequality. *Journal of Mathematical Analysis and Applications*, 386(1):300–307, 2012. arXiv:1010.0461v3. 29

[Rén61] A. Rényi. On measures of entropy and information. *Hungarian Academy of Sciences*, 1961. 11

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 5,6

[Sha18] G. Shalev. On the Fourier Entropy Influence conjecture for extremal classes. arxiv:1806.03646, 2018. 1,5,6,17

[She11] A. A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. arXiv:0906.4291. Earlier version in STOC'08. 15

[She18] Alexander A. Sherstov. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 311–324, 2018. 8

[Shi00] Y. Shi. Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Information Processing Letters*, 75(1–2):79–83, 2000. arXiv:quant-ph/9904107. 26

[SV10] R. Shaltiel and E. Viola. Hardness amplification proofs require majority. *SIAM Journal on Computing*, 39(7):3122–3154, 2010. 4

[SV12] R. A. Servedio and E. Viola. On a special case of rigidity. Manuscript: http://eccc.hpi-web.de/report/2012/144, 2012. 4

[Tal14] A. Tal. Shrinkage of de morgan formulae by spectral techniques. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 551–560, 2014. 27

[Tal17] A. Tal. Tight bounds on the Fourier spectrum of AC^0. In *32nd Computational Complexity Conference, CCC 2017*, pages 15:1–15:31, 2017. 4,27

[Tra09] P. Traxler. Variable influences in conjunctive normal forms. In *Theory and Applications of Satisfiability Testing - SAT 2009*, pages 101–113. Springer Berlin Heidelberg, 2009. 24

[TWXZ13] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 658–667, 2013. arXiv:1304.1245. 5

[Wol08] R. de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory of Computing*, 2008. ToC Library, Graduate Surveys 1. [2,10]

[WWW14] A. Wan, J. Wright, and C. Wu. Decision trees, protocols and the entropy-influence conjecture. In *Innovations in Theoretical Computer Science, ITCS'14*, pages 67–80, 2014. arXiv:1312.3003. [5,6,23]

[Zha14] S. Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*, pages 1878–1885, 2014. arXiv:1307.6738. [5]

[ZS09] Z. Zhang and Y. Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009. [5]