

# Improved bounds on Fourier entropy and Min-entropy

SRINIVASAN ARUNACHALAM, IBM Research, USA

SOURAV CHAKRABORTY, Indian Statistical Institute, India

MICHAL KOUCKÝ, Computer Science Institute of Charles University, Czech Republic

NITIN SAURABH, Technion - IIT, Israel

RONALD DE WOLF, QuSoft, CWI and University of Amsterdam, the Netherlands

Given a Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , 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 [27] seeks to relate two fundamental measures associated with the Fourier distribution: does there exist a universal constant  $C > 0$  such that  $\mathbb{H}(\widehat{f}^2) \leq C \cdot \text{Inf}(f)$ , where  $\mathbb{H}(\widehat{f}^2)$  is the Shannon entropy of the Fourier distribution of  $f$  and  $\text{Inf}(f)$  is the total influence of  $f$ ?

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

- (1) Our first contribution shows that  $\mathbb{H}(\widehat{f}^2) \leq 2 \cdot \text{aUC}^\oplus(f)$ , where  $\text{aUC}^\oplus(f)$  is the average unambiguous parity-certificate complexity of  $f$ . This improves upon several bounds shown by Chakraborty et al. [19]. We further improve this bound for unambiguous DNFs. We also discuss how our work makes Mansour's conjecture for DNFs a natural next step towards resolution of the FEI conjecture.
- (2) We next consider the weaker *Fourier Min-entropy-Influence* (FMEI) conjecture posed by O'Donnell and others [47, 50] which asks if  $\mathbb{H}_\infty(\widehat{f}^2) \leq C \cdot \text{Inf}(f)$ , where  $\mathbb{H}_\infty(\widehat{f}^2)$  is the min-entropy of the Fourier distribution. We show  $\mathbb{H}_\infty(\widehat{f}^2) \leq 2 \cdot C_{\min}^\oplus(f)$ , where  $C_{\min}^\oplus(f)$  is the minimum parity-certificate complexity of  $f$ . We also show that for all  $\varepsilon \geq 0$ , we have  $\mathbb{H}_\infty(\widehat{f}^2) \leq 2 \log(\|\widehat{f}\|_{1, \varepsilon} / (1 - \varepsilon))$ , 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  $k$ ).
- (3) Our third contribution is to better understand implications of the FEI conjecture for 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 discuss an intriguing connection between our conjecture and the constant for the Bohnenblust-Hille inequality, which has been extensively studied in functional analysis.

CCS Concepts: • **Computing methodologies** → **Representation of Boolean functions**; • **Mathematics of computing** → *Information theory*.

Additional Key Words and Phrases: Fourier analysis of Boolean functions, FEI conjecture, query complexity, polynomial approximation, approximate degree, certificate complexity

---

Authors' addresses: Srinivasan Arunachalam, IBM Research, USA, Srinivasan.Arunachalam@ibm.com; Sourav Chakraborty, Indian Statistical Institute, Kolkata, India, sourav@isical.ac.in; Michal Koucký, Computer Science Institute of Charles University, Prague, Czech Republic, koucky@iuuk.mff.cuni.cz; Nitin Saurabh, Technion - IIT, Haifa, Israel, nitinsau@cs.technion.ac.il; Ronald de Wolf, QuSoft, CWI and University of Amsterdam, the Netherlands, rdewolf@cwi.nl.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2020 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

**ACM Reference Format:**

Srinivasan Arunachalam, Sourav Chakraborty, Michal Koucký, Nitin Saurabh, and Ronald de Wolf. 2020. Improved bounds on Fourier entropy and Min-entropy. 1, 1 (June 2020), 40 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

**1 INTRODUCTION**

Boolean functions  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  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 [47, 66]. 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 [27]. 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  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , 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  $\mathbb{H}(\widehat{f}^2)$ ) is then defined as the Shannon entropy of the Fourier distribution, i.e.,

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

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

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

Combinatorially  $\text{Inf}(f)$  is the same as the average sensitivity  $\text{as}(f)$  of  $f$ . In particular, for  $i \in [n]$ , define  $\text{Inf}_i(f)$  to be the probability that on a uniformly random input flipping the  $i$ -th bit changes the function value. Then,  $\text{Inf}(f)$  is defined to be  $\sum_{i=1}^n \text{Inf}_i(f)$ .

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

$$\mathbb{H}(\widehat{f}^2) \leq C \cdot \text{Inf}(f). \tag{1}$$

The original motivation of Friedgut and Kalai for the FEI conjecture came from studying threshold phenomena of monotone graph properties in random graphs [27]. 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  $c > 0$ ). Bourgain and Kalai [12] proved a bound of  $c_\varepsilon(\log n)^{-2+\varepsilon}$  for every  $\varepsilon > 0$ . A very recent work of Kelman et al. [38] improved this bound to  $O((\log \log n)^2/(\log n)^2)$  by establishing nearly optimal bounds on Fourier Min-entropy. In general, it implies sharper bounds on the threshold interval of transitively symmetric functions.

Besides this application, the FEI conjecture is known to imply the famous Kahn-Kalai-Linial theorem [35] (otherwise referred to as the KKL theorem). The KKL theorem was one of the first major applications of Fourier analysis to understanding properties of Boolean functions and has since found many application in various areas of theoretical computer science.

**THEOREM 1.2 (KKL THEOREM).** *For every  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , there exists an  $i \in [n]$  such that  $\text{Inf}_i(f) \geq \text{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$ , where  $\text{Var}(f) := 1 - \mathbb{E}_x[f(x)]^2 = 1 - \widehat{f}(0)^2$ .*

We discuss the implication of the FEI conjecture to the KKL theorem in more detail in Section 4. Another motivation to study the FEI conjecture is that a positive answer to this conjecture would resolve the notoriously hard conjecture of Mansour [44] from 1995.

**CONJECTURE 1.3 (MANSOUR'S CONJECTURE).** *Suppose  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  is computed by a  $t$ -term DNF.<sup>1</sup> Then for every  $\varepsilon > 0$ , there exists a family  $\mathcal{T}$  of subsets of  $[n]$  such that  $|\mathcal{T}| \leq t^{O(1/\varepsilon)}$  (i.e., size of  $\mathcal{T}$  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. [30], 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 3.3.

More generally, the FEI conjecture implies that every Boolean function can be approximated (in  $\ell_2$ -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  $2^{O(\text{Inf}(f)/\varepsilon)}$  monomials such that  $\mathbb{E}_x[(f(x) - p(x))^2] \leq \varepsilon$ . The current best known bound in this direction is  $2^{O(\text{Inf}(f)^2/\varepsilon^2)}$ , proven by Friedgut [26].<sup>2</sup>

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 [47, 50] 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  $\mathbb{H}_\infty(\widehat{f}^2)$ ). The min-entropy of  $\{\widehat{f}(S)^2\}_S$  is defined as

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

and thus it is easily seen that  $\mathbb{H}_\infty(\widehat{f}^2) \leq \mathbb{H}(\widehat{f}^2)$ . In fact,  $\mathbb{H}_\infty(\widehat{f}^2)$  could be much smaller compared to  $\mathbb{H}(\widehat{f}^2)$ . For instance, consider the function  $f(x) := x_1 \vee \text{IP}(x_1, \dots, x_n)$ ; then  $\mathbb{H}_\infty(\widehat{f}^2) = O(1)$  whereas  $\mathbb{H}(\widehat{f}^2) = \Omega(n)$ . (IP is the

<sup>1</sup>A  $t$ -term DNF is a disjunction of at most  $t$  conjunctions of variables and their negations.

<sup>2</sup>Friedgut's Junta theorem says that  $f$  is  $\varepsilon$ -close to a junta on  $2^{O(\text{Inf}(f)/\varepsilon)}$  variables. We refer to [47, Section 9.6, page 269, Friedgut's Junta Theorem] for details.

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 1.4 (FMEI CONJECTURE). *There exists a universal constant  $C > 0$  such that for every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have*

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

Another way to formulate the FMEI conjecture is, suppose  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , then does there exist a Fourier coefficient  $\hat{f}(S)$  such that  $|\hat{f}(S)| \geq 2^{-O(\text{Inf}(f))}$ ? By the *granularity* of Fourier coefficients it is well-known that every Fourier coefficient of a Boolean function  $f$  is an integral multiple of  $2^{-\text{deg}(f)}$ , see [47, Exercise 1.11] for a proof of this. (Here the  $\text{deg}(f)$  refers to the degree of the unique multilinear polynomial that represents  $f$ .) The FMEI conjecture asks if we can prove a lower bound of  $2^{-O(\text{Inf}(f))}$  on any one Fourier coefficient, and even this remains open. Proving the FMEI conjecture seems to require proving interesting structural properties of Boolean functions. It was observed by [50] that the FMEI conjecture suffices to imply the KKL theorem (see also Section 4). In fact, it suffices<sup>3</sup> for most of the implications that follows from the FEI conjecture (see [38]).

Understanding the min-entropy of a Fourier distribution is important in its own right too. It was observed by Akavia et al. [2] that for a circuit class  $C$ , tighter relations between min-entropy of  $f \in C$  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  $C$  to the class  $C \circ \text{MOD}_2$ . In particular, they conjectured that min-entropy of  $f_A$  is only polynomially larger than  $f$  when  $f \in \text{AC}^0[\text{poly}(n), O(1)]$ . ( $\text{AC}^0[s, d]$  is the class of unbounded fan-in circuits of size at most  $s$  and depth at most  $d$ .) It is well-known that when  $f \in \text{AC}^0[s, d]$ ,  $\mathbb{H}_\infty(\hat{f}^2)$  is at most  $O((\log s)^{d-1} \cdot \log \log s)$  [11, 41, 62]. 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  $\text{AC}^0[s, d] \circ \text{MOD}_2$  circuits computing IP (inner-product-mod-2). This problem has garnered a lot of attention in recent times for a variety of reasons [2, 20, 21, 54, 56]. The current best known lower bound for IP against  $\text{AC}^0[s, d] \circ \text{MOD}_2$  is quadratic when  $d = 4$ , and only super-linear for all  $d = O(1)$  [20].

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

## 1.1 Prior work

After Friedgut and Kalai [27] posed the FEI conjecture in 1996, there was not much work done towards resolving it, until the work of Klivans et al. [39] 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 [36] 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  $\mathbb{H}(\hat{f}^2) \leq O(\text{Inf}(f) \cdot \log n)$  for real-valued functions  $f: \{-1, 1\}^n \rightarrow \mathbb{R}$  (see [37, 50] for a proof and [19] for an improvement of this statement).<sup>4</sup> 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* [33] is known to be true. The logarithmic Sobolev inequality states that for every  $f: \{-1, 1\}^n \rightarrow \mathbb{R}$ ,

<sup>3</sup>With a minor strengthening: the large Fourier coefficient is required to be associated with a Fourier character of size  $O(\text{Inf}(f))$ .

<sup>4</sup>For Boolean functions, the  $\log n$ -factor was improved by [32] to  $\log(s(f))$  (where  $s(f)$  is the sensitivity of the Boolean function  $f$ ).

we have  $\text{Ent}(f^2) \leq 2 \cdot \text{Inf}(f)$ , where  $\text{Ent}(f)$  is defined as  $\text{Ent}(f) = \mathbb{E}[f \ln(f)] - \mathbb{E}[f] \ln(\mathbb{E}[f])$ , 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. [50] showed that the FEI conjecture holds for symmetric Boolean functions and read-once decision trees. Keller et al. [37] 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. [19] and O’Donnell and Tan [48], 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 [48] for more). For general Boolean functions, Chakraborty et al. [19] gave several upper bounds on the Fourier entropy in terms of combinatorial quantities larger than the total influence, e.g., average decision tree depth, etc., and sometimes even quantities that could be much smaller than influence, namely, average parity-decision tree depth.

Later Wan et al. [65] used Shannon’s source coding theorem [57] (which characterizes entropy) to establish the FEI conjecture for read- $k$  decision trees for constant  $k$ . Using their novel interpretation of the FEI conjecture they also reproved O’Donnell-Tan’s composition theorem in an elegant way. Recently, Shalev [55] improved the constant in the FEI inequality for read- $k$  decision trees, and further verified 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. [23] and for random linear threshold functions (LTFs) by Chakraborty et al. [18].

There has also been some work in giving lower bounds on the constant  $C$  in the FEI conjecture. Hod [34] 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 [48].

However, there has not been much work on the FMEI conjecture. It was observed in [18, 50] 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 [55].

## 1.2 Our contributions

Our contributions in this paper are threefold, which we summarize in the following sections.

### 1.3 Better upper bounds for the FEI conjecture

Our first and main contribution of this paper is to give a better upper bound on the Fourier entropy  $\mathbb{H}(\hat{f}^2)$  in terms of  $\text{aUC}(f)$ , the *average unambiguous certificate complexity* of  $f$ . Informally, the unambiguous certificate complexity  $\text{UC}(f)$  of  $f$  is similar to the standard certificate complexity measure, except that the collection of certificates is 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,  $\text{aUC}(f)$ , we mean the expected number of bits set by an unambiguous certificate on a uniformly random input.

There have been many recent works on query complexity, giving upper and lower bounds on  $\text{UC}(f)$  in terms of other combinatorial measures such as decision-tree complexity, sensitivity, quantum query complexity, etc., see [6, 9, 28] for more. It follows from definitions that  $\text{UC}(f)$  lower bounds decision tree complexity. However, it is known that  $\text{UC}(f)$  can be quadratically smaller than decision tree complexity [6]. Our main contribution here is an improved upper bound

of average unambiguous certificate complexity  $\text{aUC}(f)$  on  $\mathbb{H}(\hat{f}^2)$ . This improves upon the previously known bound of average decision tree depth on  $\mathbb{H}(\hat{f}^2)$  [19].

THEOREM 1.5. *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function. Then,*

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

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)* theorem (Theorem 2.4). 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. [37, 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. [65] used Shannon's source coding theorem [57] to verify the conjecture for bounded-read decision trees.

In order to prove Theorem 1.5, we study the *tensorized* version of  $f$ ,  $f^M: \{-1, 1\}^{Mn} \rightarrow \{-1, 1\}$ , which is defined as follows,

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

Similarly we define a *tensorized* version  $\mathcal{C}^M$  of an unambiguous certificate  $\mathcal{C}$  of  $f$ ,<sup>5</sup> 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 induces 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  $2^{M(\text{aUC}(f, \mathcal{C}) + \delta)}$  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  $M(\text{aUC}(f, \mathcal{C}) + \delta)$  bits. Hence, a particular certificate can contribute to at most  $2^{M(\text{aUC}(f, \mathcal{C}) + \delta)}$  Fourier coefficients of  $f^M$ . Combining both these bounds, all these certificates can overall contribute to at most  $2^{2M(\text{aUC}(f, \mathcal{C}) + \delta)}$  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  $\delta$ . This now allows us to bound the Fourier entropy of  $f^M$  as follows,

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

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

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

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

Looking finely into how certificates contribute to Fourier coefficients in the proof above, we further strengthen Theorem 1.5 by showing that we can replace  $\text{aUC}(f)$  by the *average unambiguous parity-certificate complexity*  $\text{aUC}^\oplus(f)$  of  $f$ . Here  $\text{aUC}^\oplus(f)$  is defined similar to  $\text{aUC}(f)$  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.) This strengthening also improves upon the previously known bound of average parity-decision tree depth on  $\mathbb{H}(\hat{f}^2)$  [19]. It is easily seen that  $\text{aUC}^\oplus(f)$  lower bounds the average parity-decision tree depth.

<sup>5</sup>Recall an unambiguous certificate is a collection of certificates that partitions the Boolean cube  $\{-1, 1\}^n$ .

THEOREM 1.6. *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be any Boolean function. Then,*

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

The proof outline remains the same as in Theorem 1.5. However, a particular certificate in  $\mathcal{C}^M$  no longer fixes just 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  $2^{M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)}$  Fourier coefficients of  $f^M$ . Nevertheless, by the AEP theorem we still obtain that a typical parity-certificate fixes at most  $M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)$  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  $2^{M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)}$  Fourier coefficients of  $f^M$ . The rest of the proof now follows analogously.

*1.3.1 Further extension to unambiguous DNFs.* Consider an unambiguous certificate  $\mathcal{C} = \{C_1, \dots, C_t\}$  of  $f$ . It covers both 1 and  $-1$  inputs of  $f$ . Suppose  $\{C_1, \dots, C_{t_1}\}$  for some  $t_1 < t$  is a partition of  $f^{-1}(-1)$  and  $\{C_{t_1+1}, \dots, C_t\}$  is a partition of  $f^{-1}(1)$ . To represent  $f$ , it suffices to consider  $\bigvee_{i=1}^{t_1} C_i$ . This is a DNF representation of  $f$  with the additional property that  $\{C_1, \dots, C_{t_1}\}$  forms a partition of  $f^{-1}(-1)$ . We call such a representation an *unambiguous* DNF. In general, a DNF representation need not satisfy this additional property.

Using the equivalence of total influence and average sensitivity, one can easily observe that

$$\text{Inf}(f) \leq 2 \cdot \min \left\{ \sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}, \sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} \right\} \leq \text{aUC}(f, \mathcal{C}),$$

where  $\text{co-dim}(\cdot)$  denotes the co-dimension of an affine space. Note that the quantity  $\sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}$ , in a certain sense, is “average unambiguous 1-certificate complexity” and, similarly,  $\sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}$  captures “average unambiguous 0-certificate complexity”.

Building on our ideas from the main theorem in the previous section and using a *stronger* version (Theorem 2.8) of the AEP theorem we nearly (see Remark 3.5) establish the aforementioned improved bound of the smaller quantity between “average unambiguous 1-certificate complexity” and “average unambiguous 0-certificate complexity” on the Fourier entropy. Formally, we prove the following.

THEOREM 1.7. *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function and  $\mathcal{C} = \{C_1, \dots, C_t\}$  be a monochromatic affine subspace partition of  $\{-1, 1\}^n$  with respect to  $f$  such that  $\{C_1, \dots, C_{t_1}\}$  for some  $t_1 < t$  is an affine subspace partition of  $f^{-1}(-1)$  and  $\{C_{t_1+1}, \dots, C_t\}$  is an affine subspace partition of  $f^{-1}(1)$ . Further,  $p := \Pr_x[f(x) = 1]$ . Then,*

$$\mathbb{H}(\hat{f}^2) \leq \begin{cases} 2 \left( \sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + p \cdot \max_{i \in \{1, \dots, t_1\}} \text{co-dim}(C_i) \right), \\ 2 \left( \sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + (1-p) \cdot \max_{i \in \{t_1+1, \dots, t\}} \text{co-dim}(C_i) \right). \end{cases}$$

The outline for the proof of Theorem 1.7 remains the same as before, but it differs in implementation details. We sketch them now. Analogous to the proof of the main theorem we consider a partition of inputs with respect to  $f$  and its tensorized version. Motivated by the DNF representation, we study the following partition  $\{C_1, \dots, C_{t_1}, f^{-1}(1)\}$  which naturally inherits a distribution  $\mathbb{C}$  given by the uniform distribution on the underlying inputs. Again we build a “small” set  $\mathcal{B}$  of Fourier coefficients of  $f^M$  based on the Fourier expansions of strongly typical sequences. However,

365 unlike before, the probability of observing a strongly typical sequence doesn't capture the number of coefficients it  
 366 could contribute to  $\mathcal{B}$ . Here, we use stronger properties guaranteed by the strong AEP. In particular, it guarantees  
 367 that the *empirical* distribution of a typical sequence is close to the distribution of  $C$ . In contrast, the (weak) AEP only  
 368 guarantees that the *empirical* entropy of a typical sequence is close to the entropy of  $C$ . Using the stronger property  
 369 we can now lower bound the magnitude of any non-zero Fourier coefficient in the Fourier expansion of the indicator  
 370 function of a strongly typical sequence. We then use Parseval's Identity (Fact 2.1) to deduce an upper bound on its  
 371 Fourier sparsity, which in turn is used to bound the size of  $\mathcal{B}$ . We also need to argue that coefficients *not* in  $\mathcal{B}$  have  
 372 negligible Fourier weight, which can be done as before. Using the two properties, we can now complete the proof.  
 373

374 We also note that a similar bound for the general DNF representation, i.e., when  $\{C_1, \dots, C_{t_1}\}$  is an arbitrary DNF  
 375 representation of  $f$  where the  $C_i$ s need not be disjoint, suffices to establish Mansour's conjecture (Conjecture 1.3). In  
 376 fact, following the analogy, Theorem 1.7 implies a bound of "average 1-certificate complexity" in the general case. In  
 377 this direction, in Section 3.3 we observe that a weaker bound of 1-certificate complexity, i.e., showing  $\mathbb{H}(\hat{f}^2) \leq O(C_1(f))$ ,  
 378 would already suffice to answer Mansour's conjecture positively. For details, see Section 3.3.  
 379

#### 382 1.4 New upper bounds for the FMEI conjecture

383 Given the hardness of obtaining better upper bounds on the Fourier entropy of a Boolean function, we make progress  
 384 on a weaker conjecture, the FMEI conjecture. The FMEI conjecture is much less studied than the FEI conjecture. In fact,  
 385 we are aware of only one very recent paper [55] which studies the FMEI conjecture for a particular class of functions.  
 386 Our second contribution is to give upper bounds on the min-entropy of general Boolean functions in terms of the  
 387 minimum parity-certificate complexity (denoted  $C_{\min}^{\oplus}(f)$ ) and the approximate spectral norm of Boolean functions  
 388 (denoted  $\|\hat{f}\|_{1,\varepsilon}$ ). The minimum parity-certificate complexity  $C_{\min}^{\oplus}(f)$  is also referred to as the parity kill number  
 389 by O'Donnell et al. [49] and is closely related to the communication complexity of XOR functions [46, 64, 68]. The  
 390 approximate spectral norm  $\|\hat{f}\|_{1,\varepsilon}$  is related to the *quantum* communication complexity of XOR functions [40, 67]. In  
 391 particular, it characterizes the bounded-error quantum communication complexity of XOR functions with constant  
 392  $\mathbb{F}_2$ -degree [67]. (By  $\mathbb{F}_2$ -degree, we mean the degree of a function when viewed as a polynomial over  $\mathbb{F}_2$ .)  
 393

394  
 395  
 396  
 397 **THEOREM 1.8.** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function. Then,*

- 398  
 399 (i) *For every  $\varepsilon \geq 0$ ,  $\mathbb{H}_{\infty}(\hat{f}^2) \leq 2 \cdot \log(\|\hat{f}\|_{1,\varepsilon}/(1-\varepsilon))$ .*  
 400 (ii)  $\mathbb{H}_{\infty}(\hat{f}^2) \leq 2 \cdot C_{\min}^{\oplus}(f)$ .  
 401 (iii)  $\mathbb{H}_{\infty}(\hat{f}^2) \leq 2(1 + \log_2 3) \cdot R_2^{\oplus}(f)$ .<sup>6</sup>  
 402  
 403

404 The proof of Theorem 1.8 (i) expresses the quantity  $\|\hat{f}\|_{1,\varepsilon}$  as a (minimization) linear program. We consider the dual  
 405 linear program and exhibit a feasible solution that achieves an optimum of  $(1-\varepsilon)/\max_S |\hat{f}(S)|$ . This proves the desired  
 406 inequality. In order to prove part (ii) and (iii) of the theorem, the idea is to consider a "simple" function  $g$  that has "good"  
 407 correlation with  $f$ , and then upper bound the correlation between  $f$  and  $g$  using Plancherel's theorem (Fact 2.1) and the  
 408 fact that  $g$  has a "simple" Fourier structure. For part (ii),  $g$  is chosen to be the indicator function of an (affine) subspace  
 409 where  $f$  is constant, whereas for part (iii) the randomized parity decision tree computing  $f$  itself plays the role of  $g$ .<sup>7</sup>  
 410  
 411

412  
 413 <sup>6</sup> $R_2^{\oplus}(f)$  is the randomized parity-decision tree complexity of  $f$  (we define this formally in Section 2).

414 <sup>7</sup>We remark here that there exists simpler proof of part (i), along the lines of parts (ii) and (iii). However, we believe that the linear-programming  
 415 formulation of  $\mathbb{H}_{\infty}(\hat{f}^2)$  might help obtain better bounds, such as fractional block sensitivity.  
 416



As a corollary (Corollary 4.6) of this theorem we also obtain upper bounds on the Rényi Fourier entropy  $\mathbb{H}_{1+\delta}(\hat{f}^2)$  of order  $1 + \delta$  for all  $\delta > 0$ . We recall that  $\mathbb{H}_{1+\delta}(\hat{f}^2) \geq \mathbb{H}_{\infty}(\hat{f}^2)$  for every  $\delta \geq 0$  and as  $\delta \rightarrow \infty$ ,  $\mathbb{H}_{1+\delta}(\hat{f}^2)$  converges to  $\mathbb{H}_{\infty}(\hat{f}^2)$ . Also  $\mathbb{H}_1(\hat{f}^2)$ , obtained by taking the limit as  $\delta \rightarrow 0$ , is the standard Shannon entropy.

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 1.8 (in particular, part (ii)) 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 1.9.** *For every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  that can be expressed as a read- $k$  DNF, we have*

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

This theorem improves upon a recent (and independent) result of Shalev [55] 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 [55] for a precise definition). In order to prove Theorem 1.9, we essentially show that  $\text{Inf}(f)$  is at least as large as  $C_{\min}(f)$ , for read- $k$  DNFs (Lemma 4.8), where  $C_{\min}(f)$  is the minimum certificate complexity of  $f$ . Now the proof of Theorem 1.9 follows in conjunction with Theorem 1.8 (ii).

**New research work since the completion of this work:** A very recent work of Kelman et al. [38] shows that the FMEI conjecture holds up to a logarithmic factor in  $\text{Inf}(f)$ . In general they show that almost all the Fourier weight lies on Fourier coefficients of size at most  $O(\overline{\text{Inf}(f)})$  and weight at least  $2^{-O(\overline{\text{Inf}(f)} \log(1 + \overline{\text{Inf}(f)}))}$ , where  $\overline{\text{Inf}(f)} := \frac{\text{Inf}(f)}{\text{Var}(f)}$ . We remark that though these results are very close to the FMEI conjecture, they are incomparable to our bounds. In particular one could construct examples where that  $C_{\min}^{\oplus}(f)$  is much smaller than  $\overline{\text{Inf}(f)} \log(1 + \overline{\text{Inf}(f)})$  (e.g., the inner-product-mod-2 function, AND, etc.) and vice versa (e.g., Majority).

However we remark that there is a similarity in the high-level proof idea of their work and ours. In order to capture a Fourier coefficient of large weight consider the inner product between the function  $f$  and another function  $g$ . In our case we choose  $g$  to be a subfunction (of  $f$ ) that has large inner product with  $f$ , whereas Kelman et al. [38] essentially choose  $g$  to be the truncation of the Fourier spectrum of  $f$  to the Fourier coefficients of size at most  $O(\overline{\text{Inf}(f)})$  and weight at most  $2^{-\Omega(\overline{\text{Inf}(f)} \log(1 + \overline{\text{Inf}(f)}))}$ . They then present a non-trivial analysis to show that the inner product between  $f$  and  $g$  is very small. This is sufficient to reach the conclusion since the total Fourier weight on characters of size at most  $O(\overline{\text{Inf}(f)})$ , by Markov’s inequality, is a large constant (at least  $1 - \text{Var}(f)/100$ ).

## 1.5 Implications of the FEI conjecture and connections to the Bohnenblust-Hille inequality

Our final contribution is to better understand the structure of polynomials that  $\varepsilon$ -approximate Boolean functions on the Boolean cube. To be more specific, for simplicity we fix  $\varepsilon = 1/3$  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 [59] 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 5.2). 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 1.10. *No flat polynomial of degree  $d$  and sparsity  $2^{\omega(d)}$  can 1/3-approximate a Boolean function.*

*Remark 1.11.* We remark that there exist 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.

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 [10], but more recently have found applications in quantum computing [1, 45], classical and quantum XOR games [14], and polynomial decoupling [51]. In the functional analysis literature block-multilinear polynomials are known as *multilinear forms*. In an ingenious work [10], Bohnenblust and Hille showed that for every degree- $d$  multilinear form  $p: (\mathbb{R}^n)^d \rightarrow \mathbb{R}$ , we have

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

where  $C_d$  is a constant that depends on  $d$ . In [10], 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 [42]. 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 [3, 7, 24, 25, 53]). 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 1.10, we know that the FEI conjecture implies the following theorem.

THEOREM 1.12. *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 here 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 1.13. Let  $p$  be a block-multilinear polynomial of degree- $d$  that  $1/8$ -approximates a Boolean function  $f$ . Then,  $\deg(f) \leq d$ .

*Organization.* In the remainder of the paper, we prove and elaborate on each of these results in more detail. We start with preliminaries in Section 2. In Section 3, we give improved upper bounds on Fourier entropy (Theorems 1.5, 1.6 and 1.7), verify the FEI conjecture for functions with bounded average unambiguous certificate complexity, and elaborate on the connections between Fourier entropy vs. certificate complexity and Mansour's conjecture. In Section 4, we establish new upper bounds on Fourier min-entropy (Theorem 1.8), and verify the FMEI conjecture for read- $k$  DNFs (Theorem 1.9). In Section 5, we pose a conjecture (Conjecture 1.10) which is a consequence of the FEI conjecture and make partial progress (Theorem 1.12) towards its resolution. We further discuss an intriguing connection between our conjecture and the constants in the Bohnenblust-Hille inequality. Finally we conclude with some open problems in Section 6.

## 2 PRELIMINARIES

*Notation.* We denote the set  $\{1, 2, \dots, n\}$  by  $[n]$ . A *partial assignment* of  $[n]$  is a map  $\tau: [n] \rightarrow \{-1, 1, *\}$ . 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  $\tau$ , i.e.,  $\{x \in \{-1, 1\}^n : x_i = \tau(i) \text{ for every } i \text{ with } \tau(i) \neq *\}$ .

*Fourier Analysis.* We recall some definitions and basic facts from analysis of Boolean functions, referring to [47, 66] for more. Consider the space of all functions from  $\{-1, 1\}^n$  to  $\mathbb{R}$  equipped with the inner product defined as

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

For  $S \subseteq [n]$ , the character function  $\chi_S: \{-1, 1\}^n \rightarrow \{-1, 1\}$  is defined as  $\chi_S(x) := \prod_{i \in S} x_i$ . 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  $f: \{-1, 1\}^n \rightarrow \mathbb{R}$  has a unique Fourier expansion

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

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

$$\|\widehat{f}\|_{1, \varepsilon} = \min \left\{ \sum_S |\widehat{p}(S)| : |p(x) - f(x)| \leq \varepsilon \text{ for every } x \in \{-1, 1\}^n \right\}.$$

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

FACT 2.1 (PLANCHEREL'S THEOREM). For any  $f, g: \{-1, 1\}^n \rightarrow \mathbb{R}$ ,

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

In particular, if  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  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 and min-entropy of the distribution  $\{\widehat{f}(S)^2\}_S$ , defined as

573 *Definition 2.2.* For a Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , its Fourier entropy (denoted  $\mathbb{H}(\hat{f}^2)$ ) and min-entropy  
574 (denoted  $\mathbb{H}_\infty(\hat{f}^2)$ ) are  
575

$$576 \quad \mathbb{H}(\hat{f}^2) := \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}, \quad \text{and} \quad \mathbb{H}_\infty(\hat{f}^2) := \min_{\substack{S \subseteq [n]: \\ \widehat{f}(S) \neq 0}} \left\{ \log \frac{1}{\widehat{f}(S)^2} \right\}.$$

579 Similarly, we can also define the Rényi Fourier entropy.

580 *Definition 2.3 (Rényi Fourier entropy).* For  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ ,  $\alpha \geq 0$  and  $\alpha \neq 1$ , the Rényi Fourier entropy of  $f$  of  
581 order  $\alpha$  is defined as  
582

$$583 \quad \mathbb{H}_\alpha(\hat{f}^2) := \frac{1}{1-\alpha} \log \left( \sum_{S \subseteq [n]} |\widehat{f}(S)|^{2\alpha} \right).$$

584 It is known that in the limit as  $\alpha \rightarrow 1$ ,  $\mathbb{H}_\alpha(\hat{f}^2)$  is the (Shannon) Fourier entropy  $\mathbb{H}(\hat{f}^2)$  (see [22, Chapter 17, Section 8])  
585 and when  $\alpha \rightarrow \infty$ , observe that  $\mathbb{H}_\alpha(\hat{f}^2)$  converges to  $\mathbb{H}_\infty(\hat{f}^2)$ . It is easily seen that  
586

$$587 \quad \mathbb{H}_\infty(\hat{f}^2) \leq \mathbb{H}_1(\hat{f}^2) \leq \mathbb{H}_{\frac{1}{2}}(\hat{f}^2) \leq \mathbb{H}_0(\hat{f}^2).$$

589 For  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , the *influence* of a coordinate  $i \in [n]$ , denoted  $\text{Inf}_i(f)$ , is defined as  
590

$$591 \quad \text{Inf}_i(f) = \Pr_{x \in \{-1, 1\}^n} [f(x) \neq f(x^{(i)})] = \mathbb{E}_x \left[ \left( \frac{f(x) - f(x^{(i)})}{2} \right)^2 \right],$$

592 where the probability and expectation is taken according to the uniform distribution on  $\{-1, 1\}^n$  and  $x^{(i)}$  is  $x$  with the  
593  $i$ -th bit flipped. The *total influence* of  $f$ , denoted  $\text{Inf}(f)$ , is  
594

$$595 \quad \text{Inf}(f) = \sum_{i \in [n]} \text{Inf}_i(f).$$

596 In terms of the Fourier coefficients of  $f$ , it can be shown, e.g., [35], that  $\text{Inf}_i(f) = \sum_{S \ni i} \widehat{f}(S)^2$ , and therefore  
597

$$598 \quad \text{Inf}(f) = \sum_{S \subseteq [n]} |S| \widehat{f}(S)^2.$$

599 The *variance* of a real-valued function  $f$  is given by  $\text{Var}(f) := \sum_{S \neq \emptyset} \widehat{f}(S)^2 = 1 - \widehat{f}(\emptyset)^2$ . It easily follows that  $\text{Var}(f) \leq$   
600  $\text{Inf}(f)$ . We now introduce some basic complexity measures of Boolean functions which we use often, referring to [15]  
601 for more.  
602

603 *Sensitivity.* For  $x \in \{-1, 1\}^n$ , the *sensitivity* of  $f$  at  $x$ , denoted  $s_f(x)$ , is defined to be the number of neighbors  $y$  of  $x$   
604 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  
605  $s(f)$  of  $f$  is  $\max_x \{s_f(x)\}$ . The *average sensitivity*  $\text{as}(f)$  of  $f$  is defined to be  $\mathbb{E}_x [s_f(x)]$ . By the linearity of expectation  
606 observe that  
607

$$608 \quad \mathbb{E}_x [s_f(x)] = \sum_{i=1}^n \Pr_x [f(x) \neq f(x^{(i)})] = \sum_{i=1}^n \text{Inf}_i(f) = \text{Inf}(f),$$

609 so the average sensitivity of  $f$  equals the total influence of  $f$ . As a result, the FEI conjecture asks if  $\mathbb{H}(\hat{f}^2) \leq C \cdot \text{as}(f)$   
610 for every Boolean function  $f$ .  
611

612 *Certificate complexity.* For  $x \in \{-1, 1\}^n$ , the *certificate complexity* of  $f$  at  $x$ , denoted  $C(f, x)$ , is the minimum number  
613 of bits in  $x$  that needs to be fixed to ensure that the value of  $f$  is constant. The certificate complexity  $C(f)$  of  $f$  is  
614

625  $\max_x \{C(f, x)\}$ . The minimum certificate complexity of  $f$  is  $C_{\min}(f) = \min_x \{C(f, x)\}$ . The 0-certificate complexity  
 626  $C^0(f)$  of  $f$  is  $\max_{x:f(x)=1} \{C(f, x)\}$ . Similarly, the 1-certificate complexity  $C^1(f)$  of  $f$  is  $\max_{x:f(x)=-1} \{C^1(f, x)\}$ . Observe  
 627 that for every  $x \in \{-1, 1\}^n$ ,  $s(f, x) \leq C(f, x)$ . This gives  $s(f) \leq C(f)$  and  $as(f) \leq aC(f)$  where  $aC(f)$  denotes the  
 628 average certificate complexity of  $f$ . As before, the average is taken with respect to the uniform distribution on  $\{-1, 1\}^n$ .  
 629

630  
 631 *Parity-certificate complexity.* Analogously, we define the *parity-certificate complexity*  $C^\oplus(f, x)$  of  $f$  at  $x$  as the  
 632 minimum number of parities on the input variables one has to fix in order to fix the value of  $f$  at  $x$ , i.e.,  
 633

$$634 \quad C^\oplus(f, x) := \min\{\text{co-dim}(H) \mid H \text{ is an affine subspace on which } f \text{ is constant and } x \in H\},$$

635 where  $\text{co-dim}(H)$  is the *co-dimension* of the affine subspace  $H$ . It is easily seen that  $C^\oplus(f, x) \leq C(f, x)$ . We also define  
 636  $C^\oplus(f) := \max_x \{C^\oplus(f, x)\}$ , and  $C_{\min}^\oplus(f) := \min_x \{C^\oplus(f, x)\}$ .  
 637  
 638

639 *Unambiguous certificate complexity.* We now define the *unambiguous certificate complexity* of  $f$ . Let  $\tau: [n] \rightarrow \{-1, 1, *\}$   
 640 be a partial assignment. We refer to  
 641

$$642 \quad S_\tau = \{x \in \{-1, 1\}^n : x_i = \tau(i) \text{ for every } i \in [n] \setminus \tau^{-1}(*)\}$$

643 as the subcube generated by  $\tau$ . We call  $C \subseteq \{-1, 1\}^n$  a *subcube* of  $\{-1, 1\}^n$  if there exists a partial assignment  $\tau$  such  
 644 that  $C = S_\tau$  and the co-dimension of  $C$  is the number of bits fixed by  $\tau$ , i.e.,  $\text{co-dim}(C) = |\{i \in [n] : \tau(i) \neq *\}|$ . A set of  
 645 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  
 646  $i \neq j$  and  $\cup_i C_i = \{-1, 1\}^n$ .  
 647  
 648

649 An *unambiguous certificate*  $\mathcal{U} = \{C_1, \dots, C_m\}$ , also referred to as a *subcube partition*, is a set of subcubes partitioning  
 650  $\{-1, 1\}^n$ . We say  $\mathcal{U}$  computes a Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  if  $f$  is constant on each  $C_i$  (i.e.,  $f(x)$  is the  
 651 same for all  $x \in C_i$ ). For an unambiguous certificate  $\mathcal{U}$ , the *unambiguous certificate complexity*  $\text{UC}(\mathcal{U}, x)$  on input  $x$   
 652 equals  $\text{co-dim}(C_i)$  for the  $C_i$  satisfying  $x \in C_i$ . Define the *average unambiguous certificate complexity* of  $f$  with respect  
 653 to  $\mathcal{U}$  as  
 654  
 655

$$656 \quad a\text{UC}(f, \mathcal{U}) := \mathbb{E}_x[\text{UC}(\mathcal{U}, x)].$$

657 Then, the *average unambiguous certificate complexity* of  $f$  is defined as  
 658

$$659 \quad a\text{UC}(f) := \min_{\mathcal{U}} a\text{UC}(f, \mathcal{U}),$$

660 where the minimization is over all unambiguous certificates for  $f$ . Finally, the *unambiguous certificate complexity* of  $f$  is  
 661  
 662

$$663 \quad \text{UC}(f) := \min_{\mathcal{U}} \max_x \text{UC}(\mathcal{U}, x).$$

664 Note that since unambiguous certificates are more restricted than general certificates, we have  $C(f) \leq \text{UC}(f)$ .  
 665

666 An unambiguous  $\oplus$ -certificate  $\mathcal{U} = \{C_1, \dots, C_m\}$  for  $f$  is defined to be a collection of monochromatic *affine subspaces*  
 667 that together partition the space  $\{-1, 1\}^n$ . It is easily seen that a subcube is also an affine subspace. Analogously,  
 668 for an unambiguous  $\oplus$ -certificate  $\mathcal{U}$ , on an input  $x$ ,  $\text{UC}^\oplus(\mathcal{U}, x) := \text{co-dim}(C_i)$  for the  $C_i$  satisfying  $x \in C_i$ , and  
 669  $a\text{UC}^\oplus(f, \mathcal{U}) := \mathbb{E}_x[\text{UC}^\oplus(\mathcal{U}, x)]$ . Similarly, we define  $a\text{UC}^\oplus(f)$  and  $\text{UC}^\oplus(f)$ .  
 670  
 671

672 DNFs. A DNF (disjunctive normal form) is a disjunction (OR) of conjunctions (ANDs) of variables and their negations.  
 673 An *unambiguous DNF* is a DNF that satisfies the additional property that: on every  $(-1)$ -input, exactly one of the  
 674 conjunctions outputs  $-1$ .  
 675  
 676

677 *Approximate degree.* The  $\varepsilon$ -approximate degree of  $f: \{-1, 1\}^n \rightarrow \mathbb{R}$ , denoted  $\deg_\varepsilon(f)$ , is defined to be the minimum  
 678 degree among all multilinear real polynomials  $p$  such that  $|f(x) - p(x)| \leq \varepsilon$  for all  $x \in \{-1, 1\}^n$ . Usually  $\varepsilon$  is chosen to  
 679 be  $1/3$ , but it can be chosen to be any constant in  $(0, 1)$ , without significantly changing the model.  
 680

681 *Deterministic decision trees.* A deterministic decision tree for  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  is a rooted binary tree where  
 682 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  
 683 proceeds at a node with label  $i$  by evaluating the bit  $x_i$  and continuing with the subtree corresponding to the value of  $x_i$ .  
 684 Once a leaf is reached, the tree outputs a bit. We say that a deterministic decision tree computes  $f$  if for all  $x \in \{-1, 1\}^n$   
 685 its output equals  $f(x)$ .  
 686

687 A parity-decision tree for  $f$  is similar to a deterministic decision tree, except that each node in the tree is labelled by  
 688 a subset  $S \subseteq [n]$ . On input  $x \in \{-1, 1\}^n$ , the tree proceeds at a node with label  $S$  by evaluating the parity of the bits  $x_i$   
 689 for  $i \in S$  and continues with the subtree corresponding to the value of  $\bigoplus_{i \in S} x_i$ . Note that if the subsets at each node  
 690 have size  $|S| = 1$ , then we get the standard deterministic decision tree model.  
 691

692 *Randomized decision trees* A randomized decision tree for  $f$  is a probability distribution  $R_\mu$  over deterministic decision  
 693 trees for  $f$ . On input  $x$ , a decision tree is chosen according to  $R_\mu$ , which is then evaluated on  $x$ . The complexity of the  
 694 randomized tree is the largest depth among all deterministic trees with non-zero probability of being sampled according  
 695 to  $R_\mu$ . One can similarly define a randomized parity-decision tree as a probability distribution  $R_\mu^\oplus$  over deterministic  
 696 parity-decision trees for  $f$ .  
 697

698 We say that a randomized decision tree computes  $f$  with bounded-error if for all  $x \in \{-1, 1\}^n$  its output equals  $f(x)$   
 699 with probability at least  $2/3$ .  $R_2(f)$  (respectively,  $R_2^\oplus(f)$ ) denotes the complexity of the optimal randomized (respectively,  
 700 parity) decision tree that computes  $f$  with bounded-error, i.e., errs with probability at most  $1/3$ .  
 701  
 702

703 *Information Theory.* We need some preliminaries from Information theory. We start with the following consequence  
 704 of the law of large numbers, called the *Asymptotic Equipartition Property (AEP)* or the *Shannon-McMillan-Breiman*  
 705 *theorem*. See Chapter 3 in the book [22] for more details.  
 706

707 **THEOREM 2.4 (ASYMPTOTIC EQUIPARTITION PROPERTY (AEP) THEOREM).** *Let  $\mathbf{X}$  be a random variable drawn from a*  
 708 *distribution  $P$  and suppose  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M$  are independently and identically distributed copies of  $\mathbf{X}$ , then*  
 709

$$710 \quad -\frac{1}{M} \log P(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M) \longrightarrow \mathbb{H}(\mathbf{X})$$

711  
 712 *in probability as  $M \rightarrow \infty$ .*  
 713

714 We now recall the notion of typical sequences.

715 *Definition 2.5.* 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  
 716  $(x_1, x_2, \dots, x_M) \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_M$  such that  
 717

$$718 \quad 2^{-M(\mathbb{H}(\mathbf{X})+\varepsilon)} \leq P(x_1, x_2, \dots, x_M) \leq 2^{-M(\mathbb{H}(\mathbf{X})-\varepsilon)}.$$

719  
 720 The following properties of the typical set follows from the AEP.

721 **THEOREM 2.6 ([22, THEOREM 3.1.2]).** *Let  $\varepsilon \geq 0$  and  $T_\varepsilon^{(M)}(\mathbf{X})$  be a typical set with respect to  $P$ , then*

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

723 (ii) *Suppose  $x_1, \dots, x_M$  are drawn i.i.d. according to  $\mathbf{X}$ , then*

$$724 \quad \Pr \left[ (x_1, \dots, x_M) \in T_\varepsilon^{(M)}(\mathbf{X}) \right] \geq 1 - \varepsilon,$$

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 also require the following stronger version of typical sequences and asymptotic equipartition property.

*Definition 2.7 ([22, Chapter 11, Section 2]).* Let  $\mathbf{X}$  be a random variable drawn according to a distribution  $P$ . Fix  $\varepsilon > 0$ . The *strongly typical set*  $T_\varepsilon^{*(M)}(\mathbf{X})$  is defined to be the set of sequences  $\rho = (x_1, x_2, \dots, x_M) \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_M$  such that  $N(x; \rho) = 0$  if  $P(x) = 0$ , and otherwise

$$\left| \frac{N(x; \rho)}{M} - P(x) \right| \leq \frac{\varepsilon}{|\mathbf{X}|},$$

where  $N(x; \rho)$  is defined as the number of occurrences of  $x$  in  $\rho$ .

The strongly typical set shares similar properties with its (weak) typical counterpart which we state now. See [22, Chapter 11, Section 2] for a proof of this theorem.

**THEOREM 2.8 (STRONG AEP THEOREM).** *Following the notation in Definition 2.7, let  $T_\varepsilon^{*(M)}(\mathbf{X})$  be a strongly typical set. Then, there exists  $\delta > 0$  such that  $\delta \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , and the following hold:*

(i) *Suppose  $x_1, \dots, x_M$  are drawn i.i.d. according to  $\mathbf{X}$ , then*

$$\Pr \left[ (x_1, \dots, x_M) \in T_\varepsilon^{*(M)}(\mathbf{X}) \right] \geq 1 - \varepsilon,$$

for  $M$  sufficiently large.

(ii) *If  $(x_1, \dots, x_M) \in T_\varepsilon^{*(M)}(\mathbf{X})$ , then*

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

(iii) *For  $M$  sufficiently large,*

$$(1 - \varepsilon)2^{M(\mathbb{H}(\mathbf{X}) - \delta)} \leq |T_\varepsilon^{*(M)}(\mathbf{X})| \leq 2^{M(\mathbb{H}(\mathbf{X}) + \delta)}.$$

### 3 BETTER BOUNDS ON FOURIER ENTROPY

In this section we prove the main result of this paper: a new improved upper bound on the Fourier entropy of Boolean functions. It is well-known that  $\text{Inf}(f)$  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. [19] 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*. In order to keep the presentation clear, we first prove a weaker upper bound of *average unambiguous certificate complexity*  $\text{aUC}(f)$  on  $\mathbb{H}(\hat{f}^2)$ . We then explain how to generalize the proof to establish the stronger upper bound of *average unambiguous parity-certificate complexity*  $\text{aUC}^\oplus(f)$ .

#### 3.1 Average unambiguous certificate complexity

We recall an unambiguous certificate  $\mathcal{C} = \{C_1, \dots, C_t\}$  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 by  $\text{aUC}(f, \mathcal{C})$ , equals  $\mathbb{E}_{x \in \{-1, 1\}^n} [\text{UC}(\mathcal{C}, x)]$ . Further,  $\text{aUC}(f) = \min_{\mathcal{C}} \{\text{aUC}(f, \mathcal{C})\}$ . We now prove the main theorem.

781 **THEOREM 3.1 (RESTATEMENT OF THEOREM 1.5).** *For every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have*

$$782 \quad \mathbb{H}(\hat{f}^2) \leq 2 \cdot \text{aUC}(f).$$

783  
784 Before we prove this inequality, we first give a sketch of the proof. Given  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  and an unambiguous  
785 certificate  $\mathcal{C}$  for  $f$ , our first step is to consider the function  $f^M: \{-1, 1\}^{Mn} \rightarrow \{-1, 1\}$  defined as the  $M$ -fold product  
786 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  
787 copies of  $\mathbf{C}$ . Based on the typical set, we define a  $2^{2M \cdot (\text{aUC}(f) + \delta)}$ -sized set  $\mathcal{B}$  of Fourier coefficients of  $f^M$  and show  
788 that  $\mathcal{B}$  has fairly large Fourier weight. We then consider the re-normalized entropy when restricted to the Fourier  
789 coefficients in  $\mathcal{B}$ . By taking the limit ( $M \rightarrow \infty$ ), we obtain  $\mathbb{H}(\hat{f}^2) \leq O(\text{aUC}(f))$ . We now fill in the details.

790  
791  
792 **PROOF.** It suffices to show that for every unambiguous certificate  $\mathcal{C}$  for  $f$ , we have

$$793 \quad \mathbb{H}(\hat{f}^2) \leq 2 \cdot \text{aUC}(f, \mathcal{C}).$$

794  
795 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  $\tau(C_i): [n] \rightarrow$   
796  $\{-1, 1, *\}$ , corresponding to the bits fixed by  $C_i$ . Let  $\text{supp}(\tau(C_i)) \subseteq [n]$  be the set of indices where  $\tau(C_i) \neq *$ . As a  
797 shorthand, we denote  $|\text{supp}(\tau(C_i))|$  by  $|\tau(C_i)|$ . Consider the Boolean function  $f^M: \{-1, 1\}^{Mn} \rightarrow \{-1, 1\}$  given by the  
798  $M$ -fold iterated product of  $f$  with itself over distinct variables, i.e.,

$$800 \quad f^M(x^1, \dots, x^M) = f(x_1^1, \dots, x_n^1) \cdot f(x_1^2, \dots, x_n^2) \cdots f(x_1^M, \dots, x_n^M),$$

801  
802 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  $\text{aUC}(f^M, \mathcal{C}^M) =$   
803  $M \cdot \text{aUC}(f, \mathcal{C})$ .

804  
805 We now bound the Fourier entropy of  $f^M$  by showing that there is a “small” set of Fourier coefficients of  $f^M$  whose  
806 total Fourier weight is approximately 1.

807  
808 Let  $\mathbf{C}$  be a subcube-valued random variable that equals  $C_i$  with probability  $2^{-|\tau(C_i)|}$ .<sup>8</sup> Further, let  $C_1, C_2, \dots, C_M$   
809 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  
810  $\{2^{-\tau(C_i)}\}_{i \in [t]}$ . We now define a set  $\mathcal{B}$  of Fourier coefficients of  $f^M$ , which we argue below to have large Fourier weight.

$$811 \quad \mathcal{B} := \left\{ (S_1, \dots, S_M) \subseteq [n]^M \mid S_i \subseteq \text{supp}(\tau(C_i)) \subseteq [n] \text{ for } i \in [M] \text{ and } (C_1, \dots, C_M) \in T_\delta^{(M)}(\mathbf{C}) \right\}. \quad (3)$$

812  
813 Using Theorem 2.6 about typical sequences, we are now ready to bound the size of  $\mathcal{B}$  as follows.

814  
815 **CLAIM 3.1.**

$$816 \quad |\mathcal{B}| \leq 2^{2M(\text{aUC}(f, \mathcal{C}) + \delta)}.$$

817  
818 **PROOF OF CLAIM.** We first bound the size of  $T_\delta^{(M)}(\mathbf{C})$  and then count contributions of a typical sequence  $(C_1, \dots, C_M)$   
819 to  $\mathcal{B}$ . For the first bound, by the properties of the AEP Theorem 2.6 (i), the total number  $|T_\delta^{(M)}(\mathbf{C})|$  of typical sequences  
820 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  $\mathcal{B}$  if and only  
821 if  $S_i \subseteq \text{supp}(\tau(C_i))$  for all  $i \in [M]$ . Therefore, the maximum possible contribution of a typical sequence is bounded by

$$822 \quad 2^{|\tau(C_1)| + \dots + |\tau(C_M)|} = (\Pr[\mathbf{C}_1 = C_1, \dots, \mathbf{C}_M = C_M])^{-1} \leq 2^{M(\mathbb{H}(\mathbf{C}) + \delta)},$$

823  
824 where the equality is because the random variable  $\mathbf{C}$  is sampled according to the distribution  $\{2^{-|\tau(C_i)|}\}_{i \in [t]}$  and the  
825 inequality follows from Definition 2.5 of typical sets.

826  
827 <sup>8</sup>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  $|\mathcal{B}| \leq 2^{2M(\mathbb{H}(\mathcal{C})+\delta)}$ . Finally, by the definition of entropy, we have

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

where the second equality used that  $\{C_1, \dots, C_t\}$  formed an unambiguous certificate for  $f$ .  $\square$

We now claim that  $\mathcal{B}$  is a “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  $\delta$ . In what follows, with some abuse of notation, we will denote a set  $S \subseteq [n]$  and the corresponding subset  $\{x_i \mid i \in S\}$  of variables by the same set  $S$ .

CLAIM 3.2.

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

PROOF OF CLAIM. 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  $\rho$ . Further we denote the value  $f^M$  takes on any input consistent with  $\rho$  by  $f^M(\rho)$ . We can then express  $f^M$  on an input  $z \in \{-1, 1\}^{Mn}$  as follows

$$f^M(z) = \sum_{\rho \in \mathcal{C}^M} f^M(\rho) \cdot \mathbb{1}_\rho(z) = \sum_{\rho \in T_\delta^{(M)}(\mathcal{C})} f^M(\rho) \cdot \mathbb{1}_\rho(z) + \underbrace{\sum_{\rho \notin T_\delta^{(M)}(\mathcal{C})} f^M(\rho) \cdot \mathbb{1}_\rho(z)}_{:=g(z)}. \quad (4)$$

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

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

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

In this direction, consider the function  $g(z)$  defined in Eq. (4), which is  $\{-1, 1, 0\}$ -valued. Using the argument above, we have that if  $(S_1, \dots, S_M) \notin \mathcal{B}$ , then  $\widehat{f^M}(S_1 \cup \dots \cup S_M) = \widehat{g}(S_1 \cup \dots \cup S_M)$ . Then, clearly,

$$\sum_{(S_1, S_2, \dots, S_M) \notin \mathcal{B}} \widehat{f^M}(S_1 \cup \dots \cup S_M)^2 \leq \sum_T \widehat{g}(T)^2. \quad (5)$$

Moreover by Parseval’s Theorem (Fact 2.1),  $\sum_T \widehat{g}(T)^2 = \mathbb{E}_z \left[ g(z)^2 \right]$ . Therefore,

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

where the first inequality uses Eq. (5) and Parseval, the second equality is because  $g(z)^2 \in \{0, 1\}$  and the last inequality follows from Theorem 2.6 (ii).  $\square$

We are now ready to bound the Fourier entropy of  $f$  and complete the proof. We need the following well-known trick to bound entropy when the underlying distribution has large weight on a small support. Let  $\mathcal{S}$  be a set of Fourier characters such that  $\sum_{S \notin \mathcal{S}} \widehat{g}(S)^2 = \delta$ . In order to upper bound the Fourier entropy, we first express  $\mathbb{H}(\widehat{g}^2)$  as follows

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

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

$$\begin{aligned} \mathbb{H}(\widehat{g}^2) &= (1 - \delta) \mathbb{H} \left( \frac{\widehat{g}(S)^2}{1 - \delta} : S \in \mathcal{S} \right) + \delta \mathbb{H} \left( \frac{\widehat{g}(S)^2}{\delta} : S \notin \mathcal{S} \right) - \sum_{S \in \mathcal{S}} \widehat{g}(S)^2 \log(1 - \delta) - \sum_{S \notin \mathcal{S}} \widehat{g}(S)^2 \log(\delta), \\ &= (1 - \delta) \mathbb{H} \left( \frac{\widehat{g}(S)^2}{1 - \delta} : S \in \mathcal{S} \right) + \delta \mathbb{H} \left( \frac{\widehat{g}(S)^2}{\delta} : S \notin \mathcal{S} \right) + \mathbb{H}(\delta), \end{aligned} \quad (6)$$

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

We now use Eq. (6) when applied to the function  $f^M$  and set  $\mathcal{S} = \mathcal{B}$ . We then obtain

$$M \cdot \mathbb{H}(\widehat{f}^2) = \mathbb{H}(\widehat{f^M}^2) \leq \log |\mathcal{B}| + \delta \cdot \log |\{S : S \notin \mathcal{B}\}| + \mathbb{H}(\delta) \leq 2M(\text{aUC}(f, \mathcal{C}) + \delta) + \delta nM + \mathbb{H}(\delta),$$

where the second inequality used Claims 3.1 and 3.2. Dividing by  $M$  on both sides, we get

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

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

**3.1.1 Extension to  $\oplus$ -certificate complexity.** We now discuss a strengthening of Theorem 3.1 where we improve the upper bound to *average unambiguous parity-certificate complexity*  $\text{aUC}^\oplus(f)$ . Recall that an unambiguous  $\oplus$ -certificate  $\mathcal{C} = \{C_1, \dots, C_t\}$  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 by  $\text{aUC}^\oplus(f, \mathcal{C})$ , equals  $\mathbb{E}_x[\text{UC}^\oplus(\mathcal{C}, x)]$  and  $\text{aUC}^\oplus(f) := \min_{\mathcal{C}} \text{aUC}^\oplus(f, \mathcal{C})$ . 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 3.1, 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. (4), 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  $(S_1, \dots, S_M)$  be an  $M$ -tuple where each  $S_i \subseteq [n]$ . Then, we define  $\mathcal{B}$  by letting  $(S_1, \dots, S_M) \in \mathcal{B}$  if and only if there exists a typical sequence  $(C_{i_1}, \dots, C_{i_M}) \in T_\delta^{(M)}(\mathcal{C})$  such that for all  $j \in [M]$ ,  $\mathbb{1}_{S_j} \in \text{span}\langle A_{i_j} \rangle$  (where by  $\text{span}\langle A_{i_j} \rangle$ , 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 3.2 (RESTATEMENT OF THEOREM 1.6). *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be any Boolean function. Then,*

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

We remark that as a corollary to the theorem it follows that the FEI conjecture holds for the class of functions  $f$  with bounded  $\text{aUC}^\oplus(f)$ , and  $\text{Inf}(f) \geq 1$ . That is, for a Boolean function  $f$  with  $\text{Inf}(f) \geq 1$ , we have

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

We note that the reduction in [65, Proposition E.2] shows that removing the requirement  $\text{Inf}(f) \geq 1$  from the above inequality will prove the FEI conjecture for all Boolean functions with  $\text{Inf}(f) \geq \log n$ . Furthermore, if we could show the FEI conjecture for Boolean functions  $f$  where  $\text{aUC}^\oplus(f) = \omega(1)$  is a slow-growing function of  $n$ , again the padding argument in [65] shows that we would be able to establish the FEI conjecture for all Boolean functions.

### 3.2 Further improving the bound: unambiguous DNFs

In this section, we further improve the bounds obtained in the previous section by considering the DNF representation of  $f$ . Let  $\mathcal{C} := \{C_1, \dots, C_t\}$  be an unambiguous certificate of  $f$ . It covers both 1 and  $-1$  inputs of  $f$ . Suppose  $\{C_1, \dots, C_{t_1}\}$  for some  $t_1 < t$  is a subcube partition of  $f^{-1}(-1)$  and  $\{C_{t_1+1}, \dots, C_t\}$  is a subcube partition of  $f^{-1}(1)$ . To represent  $f$ , it suffices to consider  $\bigvee_{i=1}^{t_1} C_i$ . This is a DNF representation of  $f$  with an additional property that  $\{C_1, \dots, C_{t_1}\}$  forms a partition of  $f^{-1}(-1)$ . We call such a representation an *unambiguous* DNF. In general, a DNF representation need not satisfy this additional property.

As before, for every  $C_i$ , let  $\tau(C_i)$  be the partial assignment  $\tau(C_i): [n] \rightarrow \{-1, 1, *\}$  corresponding to the bits fixed by  $C_i$ , and  $|\tau(C_i)|$  denotes the size of the set of indices where  $\tau(C_i) \neq *$ . We know that

$$\text{Inf}(f) \leq \text{aUC}(f, \mathcal{C}) = \sum_{i=1}^t |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} = \sum_{i=1}^{t_1} |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} + \sum_{i=t_1+1}^t |\tau(C_i)| \cdot 2^{-|\tau(C_i)|}.$$

However, the following better bound on the influence is easily seen from the equivalence of influence and average sensitivity.

PROPOSITION 3.3. *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function and  $\mathcal{C} = \{C_1, \dots, C_t\}$  be an unambiguous certificate of  $f$ . Then,*

$$\text{Inf}(f) \leq 2 \cdot \min \left\{ \sum_{i=1}^{t_1} |\tau(C_i)| 2^{-|\tau(C_i)|}, \sum_{i=t_1+1}^t |\tau(C_i)| 2^{-|\tau(C_i)|} \right\} \leq \text{aUC}(f, \mathcal{C}).$$

Therefore, it's natural to ask whether we can improve the upper bound in Theorem 3.1 to

$$\min \left\{ \sum_{i=1}^{t_1} |\tau(C_i)| 2^{-|\tau(C_i)|}, \sum_{i=t_1+1}^t |\tau(C_i)| 2^{-|\tau(C_i)|} \right\} \text{ (up to a constant factor).}$$

We remark that the quantity  $\sum_{i=1}^{t_1} |\tau(C_i)| 2^{-|\tau(C_i)|}$ , in a certain sense, is “average unambiguous 1-certificate complexity” and, similarly,  $\sum_{i=t_1+1}^t |\tau(C_i)| 2^{-|\tau(C_i)|}$  captures “average unambiguous 0-certificate complexity”. Moreover, answering this question is of significance to tackling Mansour’s conjecture positively. For more details, see Section 3.3.

Building on our ideas from the main theorem in the previous section and using a *stronger* version of the AEP theorem (Theorem 2.8) we nearly establish the aforementioned improved bound of the smaller quantity between “average

unambiguous 1-certificate complexity” and “average unambiguous 0-certificate complexity” on the Fourier entropy. Formally, we prove the following.

**THEOREM 3.4.** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function and  $\mathcal{C} = \{C_1, \dots, C_t\}$  be an unambiguous certificate of  $f$  such that  $\{C_1, \dots, C_{t_1}\}$  for some  $t_1 < t$  is a subcube partition of  $f^{-1}(-1)$  and  $\{C_{t_1+1}, \dots, C_t\}$  is a subcube partition of  $f^{-1}(1)$ . Further,  $p := \Pr_x[f(x) = 1]$ . Then,*

$$\mathbb{H}(\hat{f}^2) \leq \begin{cases} 2 \left( \sum_{i=1}^{t_1} |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} + p \cdot \max_{i \in \{1, \dots, t_1\}} |\tau(C_i)| \right), \\ 2 \left( \sum_{i=t_1+1}^t |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} + (1-p) \cdot \max_{i \in \{t_1+1, \dots, t\}} |\tau(C_i)| \right). \end{cases}$$

*Remark 3.5.* An unsatisfactory part of the bound above is the presence of the term  $p \cdot \max |\tau(C_i)|$ . This is because when  $\max |\tau(C_i)|$  term is *not* weighted by  $p$ , it becomes a trivial bound on entropy. Ideally, one would like to get rid of this term altogether, possibly at the expense of increasing the constant factor in the first summand.

Moreover, a similar bound when  $\{C_1, \dots, C_{t_1}\}$  is an arbitrary DNF representation of  $f$  (i.e., when the  $C_i$ s need not be disjoint) suffices to prove Mansour’s conjecture (Conjecture 3.7). For details, see Section 3.3.

**PROOF OF THEOREM 3.4.** We only prove the inequality

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \left( \sum_{i=1}^{t_1} |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} + p \cdot \max_{i \in [t_1]} |\tau(C_i)| \right). \quad (7)$$

In order to obtain the other inequality in the theorem statement, we can replace  $f$  by  $\neg f$  and observe that Fourier entropy of  $f$  and  $\neg f$  is the same. We recall  $\{C_1, \dots, C_{t_1}\}$  is a subcube partition of  $f^{-1}(-1)$  and  $p = \Pr_x[f(x) = 1]$ .

Like in the proof of Theorem 3.1, we define a random variable over subsets of  $\{-1, 1\}^n$  such that the support of the random variable partitions the space. In particular, define a random variable  $C$  that equals one of the  $C_i$ ,  $1 \leq i \leq t_1$ , with probability  $2^{-|\tau(C_i)|}$  and equals the set  $f^{-1}(1) := \{x \mid f(x) = 1\}$  with probability  $p$ .<sup>9</sup> We consider the  $M$ -fold iterated product of this partition  $\{C_1, \dots, C_{t_1}, f^{-1}(1)\}$  of  $\{-1, 1\}^n$ , which gives us a partition  $\mathcal{P}$  of the space  $\{-1, 1\}^{Mn}$  such that  $f^M$  is constant on each part. Overall there are  $(t_1 + 1)^M$  parts in  $\mathcal{P}$ . For a part  $\rho \in \mathcal{P}$ , let  $\mathbb{1}_\rho(z)$  be the  $\{0, 1\}$ -valued function that is 1 if and only if  $z \in \{-1, 1\}^{Mn}$  is in  $\rho$ . Furthermore, for  $\rho \in \mathcal{P}$ , let  $f^M(\rho) \in \{-1, 1\}$  be the value of  $f^M$  on the part  $\rho$ . Then, on an input  $z \in \{-1, 1\}^{Mn}$ ,

$$f^M(z) = \sum_{\rho \in \mathcal{P}} f^M(\rho) \cdot \mathbb{1}_\rho(z) = \sum_{\rho \in T_\varepsilon^{*(M)}(C)} f^M(\rho) \cdot \mathbb{1}_\rho(z) + \sum_{\rho \notin T_\varepsilon^{*(M)}(C)} f^M(\rho) \cdot \mathbb{1}_\rho(z). \quad (8)$$

where  $T_\varepsilon^{*(M)}(C)$  is the strongly typical set with respect to the distribution of  $C$  and  $\varepsilon > 0$ . Similar to the proof of Theorem 3.1, we now define a set  $\mathcal{B}$  of Fourier coefficients of  $f^M$ , which we argue below to be of “small” size and “large” Fourier weight.

$$\mathcal{B} := \left\{ (S_1, \dots, S_M) \subseteq [n]^M \mid \widehat{\mathbb{1}_\rho}(S_1 \cup \dots \cup S_M) \neq 0 \text{ for some } \rho \in T_\varepsilon^{*(M)}(C) \right\}. \quad (9)$$

**CLAIM 3.3.**

$$|\mathcal{B}| \leq 2^{2M\delta} \cdot 2^{2M \left( \sum_{i=1}^{t_1} |\tau(C_i)| (2^{-|\tau(C_i)| + \varepsilon'} + (p + \varepsilon') \max_{i=1}^{t_1} |\tau(C_i)| \right)},$$

<sup>9</sup>Observe that  $\sum_{i=1}^{t_1} 2^{-|\tau(C_i)|} = 1 - p$ , hence  $C$  is a valid random variable.

where  $\varepsilon' = \varepsilon/(t_1 + 1)$  and  $\delta > 0$  is such that  $\delta \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

PROOF OF CLAIM. Similar to Claim 3.1 we first bound the size of the strongly typical set  $T_\varepsilon^{*(M)}(\mathcal{C})$  and then count contributions of a strongly typical sequence  $\rho$  to  $\mathcal{B}$ . The contribution of  $\rho$  can be upper bounded by the sparsity of the Fourier expansion of the indicator function  $\mathbb{1}_\rho$ . Therefore,

$$|\mathcal{B}| \leq |T_\varepsilon^{*(M)}(\mathcal{C})| \cdot \max_{\rho \in \mathcal{P}} \{\text{Fourier sparsity of } \mathbb{1}_\rho\}.$$

By the strong AEP theorem, Theorem 2.8 (iii), we have  $|T_\varepsilon^{*(M)}(\mathcal{C})| \leq 2^{M(\mathbb{H}(\mathcal{C})+\delta)}$ . We now bound the Fourier sparsity of  $\mathbb{1}_\rho$  for a strongly typical  $\rho$ . By Parseval's Theorem (Fact 2.1), we have

$$\sum_{(S_1, \dots, S_M)} \widehat{\mathbb{1}}_\rho(S_1 \cup \dots \cup S_M)^2 = \Pr_z [\mathbb{1}_\rho(z) = 1].$$

Therefore,

$$\text{sparsity}(\widehat{\mathbb{1}}_\rho) \cdot \min_{(S_1, \dots, S_M)} \widehat{\mathbb{1}}_\rho(S_1 \cup \dots \cup S_M)^2 \leq \Pr_z [\mathbb{1}_\rho(z) = 1] \leq 2^{-M(\mathbb{H}(\mathcal{C})-\delta)}, \quad (10)$$

where the second inequality follows from the strong AEP theorem, Theorem 2.8 (ii), and  $\text{sparsity}(\widehat{\mathbb{1}}_\rho)$  denotes the Fourier sparsity of  $\mathbb{1}_\rho$ . To obtain an upper bound on the sparsity we now establish a lower bound on the magnitude of the non-zero Fourier coefficients of  $\mathbb{1}_\rho$ .

Since  $\rho \in \{C_1, \dots, C_{t_1}, f^{-1}(1)\}^M$ , the Fourier expansion of the indicator function  $\mathbb{1}_\rho$  is a product of Fourier expansion of indicator functions of  $C_1, \dots, C_{t_1}$  and  $f^{-1}(1)$ . The indicator function of  $C_i$  is just an AND over the appropriate subset of literals, and the indicator function of  $f^{-1}(1)$  is  $1 - \sum_{i=1}^{t_1} \mathbb{1}_{C_i}$ . Hence, a lower bound on the magnitude of a non-zero Fourier coefficient of  $\mathbb{1}_\rho$  is given by

$$\prod_{i=1}^{t_1} \left( \frac{1}{2^{|\tau(C_i)|}} \right)^{N(C_i; \rho)} \cdot \left( \frac{1}{2^{\max_{i \in \{1, \dots, t_1\}} \{|\tau(C_i)|\}}} \right)^{N(f^{-1}(1); \rho)},$$

where  $N(C_i; \rho)$  (respectively,  $N(f^{-1}(1); \rho)$ ) is the number of occurrences of  $C_i$  (respectively,  $f^{-1}(1)$ ) in  $\rho$ . Therefore, squaring the above lower bound and using it in Eq. (10) we obtain

$$\begin{aligned} \text{sparsity}(\widehat{\mathbb{1}}_\rho) &\leq \frac{1}{2^{M(\mathbb{H}(\mathcal{C})-\delta)}} \cdot \prod_{i=1}^{t_1} 2^{2|\tau(C_i)| \cdot N(C_i; \rho)} \cdot 2^{2 \max_{i \in [t_1]} \{|\tau(C_i)|\} \cdot N(f^{-1}(1); \rho)}, \\ &= \frac{1}{2^{M(\mathbb{H}(\mathcal{C})-\delta)}} \cdot 2^{2 \cdot (\sum_{i=1}^{t_1} |\tau(C_i)| \cdot N(C_i; \rho) + \max_{i \in [t_1]} \{|\tau(C_i)|\} \cdot N(f^{-1}(1); \rho))}, \\ &\leq \frac{1}{2^{M(\mathbb{H}(\mathcal{C})-\delta)}} \cdot 2^{2M \cdot (\sum_{i=1}^{t_1} |\tau(C_i)| \cdot (2^{-|\tau(C_i)|+\varepsilon'} + \max_{i \in [t_1]} \{|\tau(C_i)|\} \cdot (p+\varepsilon'))}, \end{aligned} \quad (11)$$

where the last inequality follows from the strong AEP theorem and  $\varepsilon' = \varepsilon/(t_1 + 1)$ . Now using inequality (11) along with the bound on the size of the strongly typical set  $T_\varepsilon^{*(M)}(\mathcal{C})$  we obtain the claimed bound on the size of  $\mathcal{B}$ , thereby completing the proof of the claim.  $\square$

The next claim shows that the Fourier coefficients *not* in  $\mathcal{B}$  have low total Fourier weight.

CLAIM 3.4.

$$\sum_{(S_1, S_2, \dots, S_M) \notin \mathcal{B}} \widehat{f^M}(S_1 \cup \dots \cup S_M)^2 \leq \varepsilon.$$

PROOF OF CLAIM. We omit the proof as it is similar to the proof of Claim 3.2.  $\square$

Now bounding the Fourier entropy of  $f^M$  as in the proof of Theorem 3.1 and taking the limit as  $M \rightarrow \infty$ , we obtain the following bound on the Fourier entropy of  $f$ ,

$$\mathbb{H}(\hat{f}^2) \leq 2 \cdot \left( \sum_{i=1}^{t_1} |\tau(C_i)| \cdot 2^{-|\tau(C_i)|} + p \cdot \max_{i \in [t_1]} |\tau(C_i)| \right).$$

$\square$

**3.2.1 Extension to affine subspace partition.** Since our techniques are oblivious to the change of basis, again analogous to Theorem 3.2 we obtain the following generalization of Theorem 3.4 to the setting when  $\{C_1, \dots, C_t\}$  forms a monochromatic affine subspace partition. We state the generalization below without the proof.

**THEOREM 3.6 (RESTATEMENT OF THEOREM 1.7).** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function and  $\mathcal{C} = \{C_1, \dots, C_t\}$  be a monochromatic affine subspace partition of  $\{-1, 1\}^n$  with respect to  $f$  such that  $\{C_1, \dots, C_{t_1}\}$  for some  $t_1 < t$  is an affine subspace partition of  $f^{-1}(-1)$  and  $\{C_{t_1+1}, \dots, C_t\}$  is an affine subspace partition of  $f^{-1}(1)$ . Further,  $p := \Pr_x[f(x) = 1]$ . Then,*

$$\mathbb{H}(\hat{f}^2) \leq \begin{cases} 2 \left( \sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + p \cdot \max_{i \in \{1, \dots, t_1\}} \text{co-dim}(C_i) \right), \\ 2 \left( \sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + (1-p) \cdot \max_{i \in \{t_1+1, \dots, t\}} \text{co-dim}(C_i) \right). \end{cases}$$

### 3.3 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 3.7 (MANSOUR'S CONJECTURE [43]).** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  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  $\text{poly}(t)$  such that  $\mathbb{E}_x \left[ (f(x) - p(x))^2 \right] \leq \varepsilon$ . (The exponent in  $\text{poly}(t)$  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. [30] showed that a positive answer to Mansour's conjecture (Conjecture 3.7) 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 [29] in computational learning theory.

Earlier we saw that  $\mathbb{H}(\hat{f}^2) \leq O(\text{aUC}^\oplus(f))$ . In fact we saw a somewhat stronger bound (Theorem 3.6); informally, capturing "average unambiguous 1-parity-certificate complexity". An interesting follow-up question in light of such results, in particular Theorems 3.1 and 3.4, is if one could strengthen these upper bound to  $O(\min\{C^0(f), C^1(f)\})$ . In this section we observe that this bound on the Fourier-entropy in terms of  $C^0(f), C^1(f)$  (which is clearly weaker than the FEI conjecture) suffices to establish Mansour's conjecture.

We remark that it was implicit in previous works [29, 36, 39, 50] 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 next 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 3.8. *There exists a universal constant  $\lambda > 0$  such that for every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have*

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

It is weaker than the FEI conjecture because  $\text{Inf}(f) \leq \min \{C^0(f), C^1(f)\}$  [5, 11, 63]. To establish the implication we will use the following equivalent form of Conjecture 3.8.

CONJECTURE 3.9. *There exists a universal constant  $\lambda > 0$  such that for every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have*

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

Before establishing the implication, we quickly argue that Conjectures 3.8 and 3.9 are equivalent. It is easily seen that Conjecture 3.8 implies Conjecture 3.9. For the reverse direction, note that both  $f$  and  $\neg f$  have the same Fourier-entropy, while  $C^0(f)$  and  $C^1(f)$  reverse roles.

We also note that Conjecture 3.9 is readily seen to imply the weaker version of "Mansour's conjecture for width" posed by [31] which in turn is known to imply Conjecture 3.7. However, for the sake of completeness, we present here a direct argument to verify this implication.

PROPOSITION 3.10. *Conjecture 3.9 implies Conjecture 3.7.*

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

$$\mathbb{E}_x \left[ (f(x) - g(x))^2 \right] \leq 4 \cdot t \cdot \frac{\delta_1}{4t} = \delta_1. \quad (12)$$

Using Conjecture 3.9 for the function  $g$ , we get  $\mathbb{H}(\hat{g}^2) \leq \lambda \cdot C^1(g) \leq \lambda \log(4t/\delta_1)$ . We now construct a polynomial  $p$  by defining its Fourier coefficients as follows:

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

By Parseval's identity (Fact 2.1), 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

$$\mathbb{E}_x \left[ (g(x) - p(x))^2 \right] = \sum_S (\hat{g}(S) - \hat{p}(S))^2 = \sum_{S: |\hat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \hat{g}(S)^2, \quad (13)$$

and

$$\begin{aligned} \mathbb{H}(\hat{g}^2) &= \sum_S \hat{g}(S)^2 \log \frac{1}{\hat{g}(S)^2} \geq \sum_{S: |\hat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \hat{g}(S)^2 \log \frac{1}{\hat{g}(S)^2} \\ &\geq \frac{\mathbb{H}(\hat{g}^2)}{\delta_2} \sum_{S: |\hat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \hat{g}(S)^2. \end{aligned} \quad (14)$$

Putting together Eq. (13) and (14), we get

$$\mathbb{E}_x [(g(x) - p(x))^2] = \sum_{S: |\hat{g}(S)| < 2^{-\mathbb{H}(\hat{g}^2)/(2\delta_2)}} \hat{g}(S)^2 \leq \delta_2. \quad (15)$$

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. (12) and (15). The second inequality also used Eq. (12) and (15). 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 3.8 is not known to imply Mansour's conjecture.

CONJECTURE 3.11. *There exists a universal constant  $\lambda > 0$  such that for any Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ ,*

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

#### 4 BETTER BOUNDS ON FOURIER MIN-ENTROPY

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 [47, 50] as a simpler question to tackle, given the hardness of resolving the FEI conjecture. We restate the FMEI conjecture below.

CONJECTURE 4.1 (FMEI CONJECTURE). *There exists a universal constant  $C > 0$  such that for every Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have  $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \text{Inf}(f)$ .*

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 [35] 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



1249  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , there exists an index  $i \in [n]$  such that  $\text{Inf}_i(f) \geq \text{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$ . We now show how a  
 1250 positive answer to the FMEI conjecture implies the KKL theorem:<sup>10</sup> for simplicity, assume that  $f$  is balanced (i.e.,  
 1251  $\widehat{f}(\emptyset) = \mathbb{E}_x[f(x)] = 0$ ). Then the FMEI conjecture implies the existence of  $\emptyset \neq T \subseteq [n]$  such that  $\widehat{f}(T)^2 \geq 2^{-C \cdot \text{Inf}(f)}$ .  
 1252 Furthermore, for every  $i \in T$ , we have  
 1253

$$1254 \quad \text{Inf}_i(f) = \sum_{S: S \ni i} \widehat{f}(S)^2 \geq \widehat{f}(T)^2 \geq 2^{-C \cdot \text{Inf}(f)} \geq 2^{-Cn \cdot \max_j \{\text{Inf}_j(f)\}}, \quad (16)$$

1257 where the first inequality follows because  $T$  contains  $i$ , the second inequality follows from the FMEI conjecture and  
 1258 the last inequality because  $\text{Inf}(f) \leq n \cdot \max_j \{\text{Inf}_j(f)\}$ . However note that  $\max_j \{\text{Inf}_j(f)\}$  clearly upper bounds the  
 1259 left-hand-side of Eq. (16). Thus, we have  
 1260

$$1261 \quad \max_{j \in [n]} \{\text{Inf}_j(f)\} \geq 2^{-Cn \cdot \max_j \{\text{Inf}_j(f)\}}.$$

1263 Rearranging this inequality, we obtain  $\max_{j \in [n]} \{\text{Inf}_j(f)\} \geq \Omega\left(\frac{\log n}{n}\right)$ , which is the KKL theorem for balanced functions.  
 1264 The proof can also be extended to non-balanced functions (see [50]).  
 1265

1266 We now prove Theorem 4.2, which is our main contribution in this section. In the following theorem, we give upper  
 1267 bounds on  $\mathbb{H}_\infty(\widehat{f}^2)$  in terms of analytic and combinatorial measures of Boolean functions.  
 1268

1269  
 1270 **THEOREM 4.2 (RESTATEMENT OF THEOREM 1.8).** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a Boolean function. Then,*

$$1272 \quad (i) \text{ For every } \varepsilon \geq 0, \mathbb{H}_\infty(\widehat{f}^2) \leq 2 \cdot \log\left(\|\widehat{f}\|_{1, \varepsilon} / (1 - \varepsilon)\right).$$

$$1273 \quad (ii) \mathbb{H}_\infty(\widehat{f}^2) \leq 2 \cdot C_{\min}^\oplus(f).$$

$$1274 \quad (iii) \mathbb{H}_\infty(\widehat{f}^2) \leq 2(1 + \log_2 3) \cdot R_2^\oplus(f).$$

1276  
 1277 Before giving a proof, we first make a few remarks about the second statement in the theorem above. The FMEI  
 1278 conjecture asks if  $\mathbb{H}_\infty(\widehat{f}^2) \leq C \cdot \text{as}(f)$ ? Since we also know that for every  $x \in \{-1, 1\}^n$ , we have  $s(f, x) \leq C(f, x)$ , a  
 1279 weaker question than the FMEI conjecture (with a larger right-hand side) would be, is  $\mathbb{H}_\infty(\widehat{f}^2) \leq C \cdot \text{aC}(f)$ ? In the  
 1280 theorem above, we give a positive answer to this question and in fact show that  $\mathbb{H}_\infty(\widehat{f}^2) \leq 2 \cdot C_{\min}^\oplus(f)$ . Observe that  
 1281  $C_{\min}^\oplus(f)$  is not only a lower bound on  $\text{aC}(f)$  (and in turn  $\text{aC}(f)$ ), but it is the parity certificate complexity on the  
 1282 “easiest” input to  $f$ , unlike  $C^\oplus(f)$  where the complexity is measured according to the “hardest” input  $x$  to  $f$ . In our  
 1283 perspective, this brings us closer to proving the FMEI conjecture. In fact, we identify a non-trivial class of Boolean  
 1284 functions for which  $C_{\min}(f)$  lower bounds  $\text{Inf}(f)$ , and hence establish the FMEI conjecture for this class (Theorem 4.9).  
 1285  
 1286  
 1287  
 1288  
 1289

1290 **PROOF OF THEOREM 4.2.** We prove the three parts separately as follows.

1291  
 1292 *Part (i).* Fix  $\varepsilon \geq 0$  and  $d \in [n]$ . Given a Boolean function  $f$ , suppose  $p$  is a degree- $d$  polynomial that minimizes

$$1293 \quad \|\widehat{f}\|_{1, \varepsilon, d} = \min \left\{ \|\widehat{p}\|_1 : \deg(p) \leq d \text{ and } |p(x) - f(x)| \leq \varepsilon \text{ for every } x \in \{-1, 1\}^n \right\},$$

1298  
 1299 <sup>10</sup>This argument has appeared before in [50].

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.

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

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

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

Observe that both linear programs are feasible for  $d \geq \deg_\varepsilon(f)$ . 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  $\|\widehat{f}\|_{1,\varepsilon,d}$ .<sup>11</sup>

LEMMA 4.3. *Let  $f: \{-1,1\}^n \rightarrow \{-1,1\}$ ,  $\varepsilon \geq 0$ , and  $d \in [n]$  such that  $d \geq \deg_\varepsilon(f)$ . Then,  $\|\widehat{f}\|_{1,\varepsilon,d} \geq T$  if and only if there exists a polynomial  $\phi: \{-1,1\}^n \rightarrow \mathbb{R}$  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  $\phi$  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 4.3, we have

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

Therefore, we obtain,

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

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

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

<sup>11</sup>We remark that similar linear program characterizations of approximate degree of Boolean functions have appeared before in the works of Sherstov [58] and Bun and Thaler [16].

independent vectors  $S_1, \dots, S_k \in \{0, 1\}^n$ , and  $b_1, \dots, b_k \in \{-1, 1\}$ , such that

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

Consider the indicator function  $\mathbb{1}_H: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , 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  $y_j = \prod_{i \in \text{supp}(S_j)} x_i$ . It is now easily seen that

$$\mathbb{1}_H(x) = \text{AND}(-b_1 y_1, -b_2 y_2, \dots, -b_k y_k). \quad (17)$$

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. (17), it follows that

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

Now using the fact that  $x_i^2 = 1$ , observe that the monomial  $\prod_{j \in T} \prod_{i \in \text{supp}(S_j)} x_i$  simplifies to a multilinear monomial. We further observe that for each non-empty  $T$ , we can simplify  $\prod_{j \in T} \prod_{i \in \text{supp}(S_j)} x_i$  to a distinct multilinear monomial. This is a consequence of linear independence of  $S_1, S_2, \dots, S_k$ . 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 4.4. *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  $\mathbb{1}_H$ .

$$\langle f, \mathbb{1}_H \rangle = \mathbb{E}_x[f(x) \cdot \mathbb{1}_H(x)] = \frac{1}{2^k} + (-1) \left(p - \frac{1}{2^k}\right) + (1-p) = (1-2p) + \frac{1}{2^{k-1}} = \widehat{f}(\emptyset) + \frac{1}{2^{k-1}}. \quad (18)$$

On the other hand,

$$\langle f, \mathbb{1}_H \rangle = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \widehat{\mathbb{1}_H}(S) = \widehat{f}(\emptyset) \cdot \left(1 - \frac{1}{2^{k-1}}\right) + \sum_{T \in \mathcal{T}: T \neq \emptyset} \widehat{f}(T) \cdot \left(\frac{-\prod_{j \in T} b_j}{2^{k-1}}\right). \quad (19)$$

Putting together Eq. (18) and Eq. (19) we have,

$$1 = -\widehat{f}(\emptyset) + \sum_{T \in \mathcal{T}: T \neq \emptyset} \widehat{f}(T) \cdot (-\prod_{j \in T} b_j) \leq \sum_{T \in \mathcal{T}} |\widehat{f}(T)|. \quad (20)$$

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

This completes the proof of part (ii).

*Remark 4.5.* We note that in fact the proof of part (ii), in particular Eq. (20), shows that the following bounds hold.

(a) There exists a non-empty set  $S \subseteq [n]$  of size at most  $C_{\min}(f)$  such that

$$|\widehat{f}(S)| \geq \frac{1 - |\widehat{f}(\emptyset)|}{2^{C_{\min}(f)} - 1} \geq \frac{\text{Var}(f)}{2^{C_{\min}(f)+1}}.$$

(b) There exists a non-empty set  $S \subseteq [n]$  of size at most  $\max\{C_{\min}^0(f), C_{\min}^1(f)\}$  such that

$$|\widehat{f}(S)| \geq \frac{1 + |\widehat{f}(\emptyset)|}{2^{\max\{C_{\min}^0(f), C_{\min}^1(f)\}} - 1} \geq \frac{\sqrt{\text{Var}(f)}}{2^{\max\{C_{\min}^0(f), C_{\min}^1(f)\}}}.$$

We end the remark here and continue with the proof of the theorem.

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

$$\frac{1}{3} \leq \mathbb{E}_x \mathbb{E}_{T \sim \mu}[f(x)T(x)] = \mathbb{E}_{T \sim \mu} \mathbb{E}_x[f(x)T(x)] = \mathbb{E}_{T \sim \mu} \left[ \sum_{S \subseteq [n]} \widehat{f}(S) \widehat{T}(S) \right]. \quad (21)$$

On the other hand, one can upper bound the last expression in Eq. (21) as follows

$$\mathbb{E}_{T \sim \mu} \left[ \sum_{S \subseteq [n]} \widehat{f}(S) \widehat{T}(S) \right] \leq \mathbb{E}_{T \sim \mu} \left[ \sum_{S \subseteq [n]} |\widehat{f}(S)| |\widehat{T}(S)| \right] \leq \left( \max_{S \subseteq [n]} |\widehat{f}(S)| \right) \mathbb{E}_{T \sim \mu} \left[ \sum_{S \subseteq [n]} |\widehat{T}(S)| \right]. \quad (22)$$

Putting together Eq. (21) and Eq. (22) 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 [13]). 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.  $\square$

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  $\{\widehat{f}(S)^2\}$ , 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 4.2 from  $\mathbb{H}_\infty(\widehat{f}^2)$  to  $\mathbb{H}(\widehat{f}^2)$ , but this relation between the Rényi entropies breaks down for  $\delta = 0$ .

**COROLLARY 4.6.** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , and  $\delta > 0$ . Then,*

(i) *For every  $\varepsilon \geq 0$ ,  $\mathbb{H}_{1+\delta}(\widehat{f}^2) \leq 2 \left(1 + \frac{1}{\delta}\right) \cdot \log(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon))$ .*

(ii)  $\mathbb{H}_{1+\delta}(\widehat{f}^2) \leq 2 \left(1 + \frac{1}{\delta}\right) \cdot C_{\min}^\oplus(f)$ .

(iii)  $\mathbb{H}_{1+\delta}(\widehat{f}^2) \leq 2(1 + \log 3) \left(1 + \frac{1}{\delta}\right) \cdot R_2^\oplus(f)$ .

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

As a corollary to Theorem 4.2 (ii), 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 [55] 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  $C_{\min}$  on the min-entropy of  $\{\widehat{f}(S)^2\}_S$ .

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

**THEOREM 4.7** ([27, 35]). *There exists a universal constant  $c > 0$  such that for every  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ , we have*

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

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 [55] too.

**LEMMA 4.8.** *There exists a universal constant  $c > 0$  such that for all  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  that can be expressed as a read- $k$  DNF, we have*

$$\text{Inf}(f) \geq c \cdot \text{Var}(f) \cdot (C_{\min}(f) - 1 - \log k).$$

**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,  $\text{Inf}_i(f) = \Pr_x[f(x) \neq f(x^{(i)})]$ . Using the fact that  $f$  is a DNF we upper bound the  $\text{Inf}_i(f)$  as follows,

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

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

$$\text{Inf}_i(f) \leq \sum_{j=1}^t \Pr_x[T_j(x) \neq T_j(x^{(i)})] = \sum_{\substack{j=1 \\ x_i \text{ appears in } T_j}}^t \Pr_x[T_j(x) \neq T_j(x^{(i)})] \leq k 2^{-(C_{\min}(f)-1)}.$$

The second inequality follows because a variable appears in at most  $k$  terms, and  $|T_j| \geq C_{\min}(f)$  for all  $j$ . Now using the KKL theorem (Theorem 4.7), we obtain

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

This concludes the proof of the lemma. □

We now use Lemma 4.8 and Theorem 4.2 (ii) to show that FMEI conjecture holds for read- $k$  DNF.

**THEOREM 4.9 (RESTATEMENT OF THEOREM 1.9).** *Let  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$  be a read- $k$  DNF. Then,*

$$\mathbb{H}_{\infty}(\hat{f}^2) \leq C \cdot \text{Inf}(f), \quad \text{where } C = O(\log k).$$

**PROOF.** We consider two cases based on whether  $\text{Var}(f)$  is “small” ( $< 1/2$ ) or “large” ( $\geq 1/2$ ).

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

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

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.*  $\text{Var}(f) \geq 1/2$ . Using Theorem 4.2 (ii) and Lemma 4.8 we bound the min-entropy as follows,

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

1509 where  $c$  is a universal constant. Since  $\text{Inf}(f) \geq \text{Var}(f) \geq 1/2$ , we further bound the last term in the above inequality to  
 1510 obtain

$$1511 \quad \mathbb{H}_\infty(\hat{f}^2) \leq ((4/c) + 4(1 + \log k)) \cdot \text{Inf}(f).$$

1512  
 1513  
 1514 This completes the proof of the theorem. □

## 1515 5 IMPLICATIONS OF THE FEI CONJECTURE

1516 The FEI and FMEI conjecture seem to impose a strong constraint on the Fourier spectrum of a Boolean function. For  
 1517 example, the FMEI conjecture (if true) would show the existence of a large Fourier coefficient in the spectrum of every  
 1518 Boolean function that has small average sensitivity. In the introduction we saw that the FEI conjecture implies the  
 1519 existence of a Fourier-sparse polynomial  $p$  that approximates a Boolean function  $f$  in  $\ell_2$ -distance, i.e.,  $\mathbb{E}_x[(f(x) - p(x))^2]$   
 1520 is small. (A particular case being Mansour’s conjecture, which is also a consequence of the FEI conjecture.) In this  
 1521 section we discuss one implication of the FEI conjecture relating to the structure of polynomials that approximate  
 1522 Boolean functions in the  $\ell_\infty$ -distance, i.e.  $|p(x) - f(x)|$  is small for every  $x \in \{-1, 1\}^n$ . In particular, we consider the  
 1523 question “Do polynomials approximating a Boolean function in  $\ell_\infty$ -distance satisfy some property?”. In this direction,  
 1524 we make progress by showing that if the FEI conjecture is true, then we can rule out polynomials with “large” Fourier  
 1525 sparsity from representing or approximating Boolean functions. In this section, we also consider a class of polynomials  
 1526 and show that no polynomial in that class can  $1/8$ -approximate a Boolean function (without assuming that the FEI  
 1527 conjecture is true).  
 1528

1529  
 1530  
 1531  
 1532  
 1533 *Definition 5.1.* An  $n$ -variate multilinear polynomial  $p$  is said to be a *flat polynomial* if all its *non-zero* coefficients  
 1534 have the same magnitude.

1535  
 1536  
 1537 **LEMMA 5.2.** *Let  $\varepsilon \in (0, 1/2)$  be a constant, and suppose the FEI conjecture is true. Let  $p$  be a flat polynomial with degree*  
 1538 *and sparsity  $T = 2^{\omega(d)}$ . Then  $p$  cannot  $\varepsilon$ -approximate any Boolean function.*

1539  
 1540 **PROOF.** In order to prove the lemma, we crucially use the following claim.

1541  
 1542 **CLAIM 5.1.** *Let  $p$  be a flat polynomial with sparsity  $T$  and, further, suppose  $p$   $\varepsilon$ -approximates a Boolean function  $f$ . Then,*

$$1543 \quad \mathbb{H}(\hat{f}^2) \geq \Omega(\log T), \tag{23}$$

1544  
 1545 where the constant in  $\Omega(\cdot)$  depends on  $\varepsilon$ .

1546  
 1547 We assume this claim and conclude the proof of the lemma. By contradiction, let us assume that  $p$   $\varepsilon$ -approximates a  
 1548 Boolean function  $f$ , so  $\text{deg}_\varepsilon(f) \leq d$ . Assuming the FEI conjecture is true, we have  $\mathbb{H}(\hat{f}^2) = O(\text{Inf}(f))$ . Furthermore,  
 1549 using a result of Shi [60], we have  $\text{Inf}(f) \leq O(\text{deg}_\varepsilon(f))$  for every  $f$  and constant  $\varepsilon$ . So, we have  $\mathbb{H}(\hat{f}^2) \leq O(\text{deg}_\varepsilon(f)) \leq$   
 1550  $O(d)$ . Using Claim 5.1, it follows that

$$1551 \quad \Omega(\log T) \leq \mathbb{H}(\hat{f}^2) \leq O(d).$$

1552  
 1553 But this upper bound of  $T = 2^{O(d)}$  contradicts the assumption on  $T$  in the statement of the lemma. Hence, we conclude  
 1554 that  $p$  cannot  $\varepsilon$ -approximate any Boolean function. □

1555  
 1556 We now prove the claim.

1557  
 1558  
 1559  
 1560 Manuscript submitted to ACM

PROOF OF CLAIM 5.1. Suppose  $p$   $\varepsilon$ -approximates a Boolean function  $f$ . Without loss of generality suppose that all non-zero coefficients of  $p$  have magnitude  $\frac{\alpha}{\sqrt{T}}$  for some  $\alpha \in [(1 - \varepsilon), (1 + \varepsilon)]$ . Such an  $\alpha$  exists because, by Parseval's identity (Fact 2.1), 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  $p$   $\varepsilon$ -approximates  $f$ . Now consider the following set  $\mathcal{A}$  of "large" Fourier coefficients,

$$\mathcal{A} := \left\{ S : |\widehat{f}(S)| \geq \frac{2\alpha}{\sqrt{T}} \right\}. \quad (24)$$

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

$$\sum_{S \in \mathcal{A}} \widehat{f}(S)^2 \leq 4 \sum_{S \in \mathcal{A}} (\widehat{f}(S) - \widehat{p}(S))^2 \leq 4 \sum_{S \subseteq [n]} (\widehat{f}(S) - \widehat{p}(S))^2 = 4 \mathbb{E}_x [(f(x) - p(x))^2] \leq 4\varepsilon^2, \quad (25)$$

where the equality uses Parseval's identity (Fact 2.1) and the last inequality uses that  $p$   $\varepsilon$ -approximates  $f$ . From Eq. (25), 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  $f$ ,

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

The second inequality follows by the definition of  $\mathcal{A}$  (Eq. (24)), the third inequality by negation of the inequalities in Eq. (25), the last inequality holds because  $\alpha \leq (1 + \varepsilon)$  by the definition of  $\alpha$ , and the last equality assumes  $\varepsilon \in (0, 1/2)$ .  $\square$

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

CONJECTURE 5.3. *No flat polynomial of degree  $d$  and sparsity  $2^{\omega(d)}$  can  $\varepsilon$ -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 5.3.* 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 [61, 62]. Among other results, Tal [61, Claim 2.13] showed that, if a Boolean function  $f$  can be  $1/3$ -approximated in the  $\ell_\infty$ -distance by a degree- $d$  polynomial, then it has exponentially decreasing Fourier tails above level  $O(d)$ . Now using the results<sup>12</sup> in [62], we know

$$\sum_{S: |S|=d} |\widehat{f}(S)| \leq O(d)^d. \quad (26)$$

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  $\ell_\infty$ -norm. This implies

<sup>12</sup>See [62, Lemmas 29 and 34] for a precise statement of his results.

1613  $|\widehat{p}(S)| = \frac{\alpha}{\sqrt{T}}$  for some  $\alpha \in [2/3, 4/3]$ . Then, we have

1614  
1615  
1616 
$$\sum_{S: |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \leq \left( \sum_{S: |S|=d} |\widehat{p}(S) - \widehat{f}(S)|^2 \right)^{1/2} \cdot \sqrt{\binom{n}{d}} \leq \frac{1}{3} \cdot \sqrt{\binom{n}{d}},$$

1617  
1618 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,

1619  
1620 
$$\sum_{S: |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \geq \sum_{S: |S|=d} (|\widehat{p}(S)| - |\widehat{f}(S)|) = \sum_{S: |S|=d} |\widehat{p}(S)| - \sum_{S: |S|=d} |\widehat{f}(S)|$$

1621  
1622 
$$\geq \frac{2}{3} \sqrt{\binom{n}{d}} - \sum_{S: |S|=d} |\widehat{f}(S)|,$$

1623  
1624  
1625

1626 where the first inequality uses the reverse-triangle inequality and the last uses the lower bound on  $|\widehat{p}(S)|$ . From the

1627 above two inequalities we get

1628  
1629 
$$\sum_{S: |S|=d} |\widehat{f}(S)| \geq \frac{1}{3} \sqrt{\binom{n}{d}}.$$

1630  
1631

1632 This contradicts Eq. (26) when  $d = o(n^{1/3})$ . Thus,  $p$  cannot  $1/3$ -approximate a Boolean function. However, our conjecture

1633 asks if a similar result also holds when  $T = 2^{\omega(d)}$ .

1634  
1635 Instead of considering arbitrary flat polynomials, we consider a restricted class of polynomials which are referred to

1636 as *block-multilinear polynomials*. An  $n$ -variate polynomial is said to be *block-multilinear* if the input variables can be

1637 *partitioned* into disjoint blocks  $A_1, \dots, A_d \subseteq \{x_1, \dots, x_n\}$  such that every monomial in the polynomial has *at most one*

1638 variable from each block. For the purposes of this paper we will assume that each block is of the same size. In other

1639 words, a block-multilinear polynomial  $p: (\mathbb{R}^n)^d \rightarrow \mathbb{R}$  can be written as

1640  
1641 
$$p(x^1, \dots, x^d) = \sum_{S \subseteq [nd]: \forall j \in [d], |S \cap x^j| \leq 1} \widehat{p}(S) \prod_{l \in S} x_l, \quad (27)$$

1642  
1643

1644 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

1645 every  $i \in [d]$ . Clearly such a block-multilinear polynomial has degree at most  $d$  and sparsity at most  $(n+1)^d$ . Such

1646 polynomials have found applications in quantum computing [1, 45], classical and quantum XOR games [14], polynomial

1647 decoupling [51] and in functional analysis which we discuss later. Our main contribution in this section is that we show

1648 a positive answer to Conjecture 5.3 for the class of flat block-multilinear polynomials.

1649  
1650  
1651

1652 **THEOREM 5.4 (RESTATEMENT OF THEOREM 1.12).** *If  $p$  is an  $n$ -variate flat block-multilinear polynomial with degree  $d$*

1653 *and sparsity  $2^{\omega(d)}$ , then  $p$  cannot  $1/8$ -approximate a Boolean function.*

1654  
1655

1656 We defer the proof of this theorem to the next section. We now continue with the relevance of block-multilinear

1657 polynomials and the theorem above to functional analysis literature.

1658  
1659

1660 *Relation between Theorem 5.4 and the Bohnenblust-Hille (BH) inequality.* Consider the block-multilinear polynomial

1661  $p: (\mathbb{R}^n)^d \rightarrow \mathbb{R}$  defined in Eq. (27), but we further assume that it has no monomial of degree  $< d$ . One way to show

1662 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  $9/8$ .

1663  
1664



Understanding if such an  $x'$  exists for such a polynomial  $p$  can be cast as the following maximization problem

$$\|p\| := \max_{x^1, \dots, x^d \in [-1, 1]^n} \left| \sum_{i_1, \dots, i_d=1}^n \widehat{p}_{i_1, \dots, i_d} x_{i_1}^1 \cdots x_{i_d}^d \right|. \quad (28)$$

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 [4, 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. (28) 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. (27) was by H. F. Bohnenblust and E. Hille [10] 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  $p: (\mathbb{R}^n)^d \rightarrow \mathbb{R}$ , we have

$$\left( \sum_{i_1, \dots, i_d=1}^n |\widehat{p}_{i_1, \dots, i_d}|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \leq C_d \cdot \max_{x^1, \dots, x^d \in [-1, 1]^n} |p(x^1, \dots, x^d)|. \quad (29)$$

The result of [10] showed that it suffices to pick  $C_d = d^{\frac{d+1}{2d}} 2^{\frac{d-1}{2}}$  in order to satisfy Eq. (29). In addition, their bound on  $C_d$  recovers the well-known Littlewood’s 4/3 inequality [42] 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. (29) holds. We cite a few results [3, 7, 24, 25, 45, 53], referring the reader to the references within these papers for more. It is a long-standing open question whether  $C_d$  is a universal constant (there has been recent work [53] giving numerical evidence that this is the case).

After a series of works, Pellegrino and Seoane-Sepúlveda showed [52] 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 Bayart et al. [7] as  $C_d = O(d^{0.365})$ . We also point the interested reader to [45, 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. (29).

We now discuss the relevance of the BH-inequality to Theorem 5.4. Consider an arbitrary flat block-multilinear polynomial  $p: (\{-1, 1\}^n)^d \rightarrow \mathbb{R}$  with degree  $d$  and sparsity  $T$  that 1/8-approximates a Boolean function. That is, every non-zero Fourier coefficient  $|\widehat{p}_{i_1, \dots, i_d}|$  equals  $\alpha/\sqrt{T}$  for some  $\alpha \in [7/8, 9/8]$ . Then using Eq. (29), 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 Bayart et al. [7],  $C_d = O(d^{0.365})$ , 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 5.4. However, if the long-standing open question of  $C_d$  being a universal constant were true, then Theorem 5.4 follows. But proving  $C_d$  a universal constant seems to be a very hard problem. Nevertheless, in the next section, we will prove Theorem 5.4, while circumventing the barrier of improving the upper bound on  $C_d$ .

We remark that Theorem 5.4 does not say anything about the BH-inequality, since Theorem 5.4 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 than  $9/8$ ,

1717 while the BH-inequality states that such a polynomial definitely takes a value of high magnitude ( $> 9/8$ ) on at least one  
 1718 input.<sup>13</sup>  
 1719

## 1720 5.1 Proof of Theorem 5.4

1721 We restate the theorem for convenience.

1722 *Theorem 5.4 (restated).* If  $p$  is an  $n$ -variate flat block-multilinear polynomial with degree  $d$  and sparsity  $2^{\omega(d)}$ , then  $p$   
 1723 cannot  $1/8$ -approximate a Boolean function.

1724 Without loss of generality, let us assume  $d$  divides  $n$ . Let  $\{A_1, \dots, A_d\}$  be a partition of  $\{x_1, \dots, x_n\}$  and for simplicity  
 1725 suppose  $A_1 = \{x_1, \dots, x_k\}$  for  $k = n/d$ . Let  $B = \{x_1, \dots, x_n\} \setminus A_1$  denote the set of remaining variables. Then the  
 1726 monomials of  $p$  can be divided into those containing variables in  $A_1$  and those independent of variables in  $A_1$ . So we  
 1727 can write  $p$  as follows  
 1728

$$1729 \quad p(x) = q_0(x_B) + \sum_{i=1}^k x_i q_i(x_B), \quad (30)$$

1730 where  $q_0, \dots, q_k$  are polynomials of degree at most  $d - 1$ . From here on, for notational simplicity we simply rewrite  $x_B$   
 1731 as  $z$ .

1732 **LEMMA 5.5.** Let  $p$  be a degree- $d$  block-multilinear polynomial that  $1/8$ -approximates a Boolean function  $f: \{-1, 1\}^n \rightarrow$   
 1733  $\{-1, 1\}$ . Further, let  $q_0, \dots, q_k$  be as defined in Eq. (30). Then,

1734 (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  $1/8$ -approximation to a  
 1735  $\{-1, 0, 1\}$ -valued function.

1736 (II) For every  $z \in \{-1, 1\}^{n-k}$ , there exists a unique  $j \in \{0, \dots, k\}$  which satisfies

$$1737 \quad |q_j(z)| \geq 7/8 \quad \text{and} \quad \sum_{i \neq j} |q_i(z)| \leq 1/8.$$

1738 **PROOF.** The proof of the first part is fairly straightforward, while the second part requires some calculations.

1739 **PROOF OF (I).** We first rewrite  $q_1, \dots, q_k$  as follows:

$$1740 \quad q_i(z) = \frac{p(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_k, z) - p(x_1, \dots, x_{i-1}, -1, x_{i+1}, \dots, x_k, z)}{2},$$

1741 for  $i \in [k]$  and similarly we rewrite  $q_0$  as

$$1742 \quad q_0(z) = \frac{p(x_1, \dots, x_k, z) + p(-x_1, \dots, -x_k, z)}{2}.$$

1743 Now the first part follows from the fact that  $p$  is a  $1/8$ -approximation to the Boolean function  $f$ . So the  $q_i$ s are a  
 1744  $1/8$ -approximation to a  $\{-1, 0, 1\}$ -valued function.  $\square$

1745 **PROOF OF (II).** Fix  $z \in \{-1, 1\}^{n-k}$ . First observe that for every  $x_1, \dots, x_k \in \{-1, 1\}$ , we have

$$1746 \quad p(x_1, \dots, x_k, z) \in \underbrace{\left[ q_0(z) - \sum_{i=1}^k |q_i(z)|, q_0(z) + \sum_{i=1}^k |q_i(z)| \right]}_{:= I(z)}. \quad (31)$$

1766 <sup>13</sup>It is not clear to us if flat polynomials considered in Theorem 5.4 can take values close to 0, or if it is possible to show that flat polynomials *always* take  
 1767 a large value outside the Boolean cube.

Furthermore, there exists a choice of  $x_1, \dots, x_k \in \{-1, 1\}$  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  $j \in \{0, \dots, k\}$  such that  $|q_j(z)| \geq 7/8$  and  $\sum_{i \neq j} |q_i(z)| \leq 1/8$ , hence proving the lemma statement.

*Cases (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).

*Case (c).* Recall the definition of  $p(x_1, \dots, x_k, z)$  from Eq. (31). Let  $a, b \in \{-1, 1\}^k$  be such that

$$p(a, z) = q_0(z) - \sum_{i=1}^k |q_i(z)| \quad \text{and} \quad p(b, z) = q_0(z) + \sum_{i=1}^k |q_i(z)|. \quad (32)$$

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<sup>14</sup> 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  $\sum_{i \neq j} |q_i(z)| \leq 1/8$ .

The first claim  $|q_j(z)| \geq 7/8$  follows immediately, because the change in  $p$  on this edge is  $2 \cdot |q_j(z)|$  (by Eq. (32)), 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  $\text{sign}(q_0(z))$  is positive or negative. Since the proofs for both cases are similar, for simplicity we prove it assuming  $\text{sign}(q_0(z))$  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  $j$ th direction given by setting all the variables except  $x_j$  as follows:  $x_i = \text{sign}(q_i(z))$  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  $x_j = \text{sign}(q_j(z))$ ,  $p$  takes the value  $\sum_{i \neq j} |q_i(z)| + |q_j(z)|$  which is in the interval  $[7/8, 9/8]$  and

$$\sum_{i \neq j} |q_i(z)| + |q_j(z)| \in [7/8, 9/8], \text{ implies } \sum_{i \neq j} |q_i(z)| \leq 9/8 - |q_j(z)|. \quad (33)$$

Similarly, when  $x_j = -\text{sign}(q_j(z))$  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

$$\sum_{i \neq j} |q_i(z)| - |q_j(z)| \in [-9/8, -7/8], \text{ implies } \sum_{i \neq j} |q_i(z)| \leq |q_j(z)| - 7/8. \quad (34)$$

From Eq. (33) and (34) we have  $\sum_{i \neq j} |q_i(z)| \leq 1/8$ . The uniqueness of  $j$  in each case follows because  $\sum_{i=0}^k |q_i(z)| \leq 9/8$ .  $\square$

<sup>14</sup>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  $j$ th bit of  $w$ .

1821

□

1822

1823

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

1824

1825

LEMMA 5.6. *Let  $p$  be a degree- $d$  block-multilinear polynomial that  $\frac{1}{8}$ -approximates a Boolean function  $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ . Then,  $\deg(f) \leq d$ .*

1826

1827

1828

PROOF. We will establish the proof by induction on  $d$ .

1829

1830

1831

*Base case.* We have  $d = 1$ . 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  $\text{sign}(p(x)) = f(x)$  for every  $x$ . We now express  $\text{sign}(p(x))$  as a polynomial of degree at most 1. From Lemma 5.5 (II), it follows that there exists a unique  $i \in \{0\} \cup [n]$  such that

1832

1833

1834

1835

1836

1837

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

Thus the base case follows.

1838

1839

1840

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

1841

1842

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

1843

1844

1845

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

1846

1847

Recall from the proof of Lemma 5.5 (I), that  $q_i(z)$  can be expressed as a difference between two  $(d - 1)$ -block-multilinear polynomials by fixing  $x_1, \dots, x_k$ . That is,

1848

1849

1850

1851

1852

$$\begin{aligned} q_0(z) &= \frac{p(1, \dots, 1, z) + p(-1, \dots, -1, z)}{2}, \\ q_i(z) &= \frac{p(1, \dots, 1, 1, 1, \dots, 1, z) - p(1, \dots, 1, -1, 1, \dots, 1, z)}{2} \quad \text{for } i \in [k], \end{aligned} \tag{35}$$

1853

1854

1855

1856

1857

1858

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

1859

1860

$$\tilde{q}_0(z) = \frac{f(1, \dots, 1, z) + f(-1, \dots, -1, z)}{2}$$

1861

1862

1863

1864

1865

is a  $\{-1, 0, 1\}$ -valued function satisfying  $\deg(\tilde{q}_0) \leq d - 1$ . Moreover,  $q_0$  (as defined in Eq. (35)) is a  $\frac{1}{8}$ -approximation to  $\tilde{q}_0$ . In a similar fashion, one can define  $\tilde{q}_i$  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. (35)) is a  $\frac{1}{8}$ -approximation to  $\tilde{q}_i$ . Finally consider the polynomial  $\tilde{p}$  defined as follows,

1866

1867

1868

1869

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

1870

1871

1872

Firstly note that  $\tilde{p}(x)$  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 5.5, 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 5.5 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  $|f(x) - \tilde{p}(x)| \leq 1/4$  for any  $x$ . This is because for a given  $x$  if  $j$  is the unique index from Lemma 5.5 such that  $|q_j(z_x)| \geq 7/8$ , then  $|f(x) - x_j q_j(z_x)| \leq 1/8$  (because of Lemma 5.5 (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  $\tilde{p}(x) = f(x)$ .

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

Using Claim 5.1 and Lemma 5.6, 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 5.4.

## 6 CONCLUSION

We gave improved upper bounds on Fourier entropy of Boolean functions in terms of average unambiguous (parity)-certificate complexity, and as a corollary verified the FEI conjecture for functions with bounded average unambiguous (parity)-certificate complexity. We established many bounds on Fourier min-entropy in terms of analytic and combinatorial measures, namely minimum certificate complexity, logarithm of the approximate spectral norm and randomized (parity)-decision tree complexity. As a corollary to this, we verified the FMEI conjecture for read- $k$  DNFs. We also studied structural implications of the FEI conjecture on approximating polynomials. In particular, we proved that flat block-multilinear polynomials of degree  $d$  and sparsity  $2^{\omega(d)}$  can not approximate Boolean functions.

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

- (1) Does there exist a Fourier coefficient  $S \subseteq [n]$  such that  $|\hat{f}(S)| \geq 2^{-O(\deg_{1/3}(f))}$ ? This would show  $\mathbb{H}_\infty(\hat{f}^2) \leq O(\deg_{1/3}(f))$ .
- (2) Can we show  $\mathbb{H}(\hat{f}^2) \leq O(Q(f))$ ? Or,  $\mathbb{H}_\infty(\hat{f}^2) \leq O(Q(f))$ ? (where  $Q(f)$  is the  $1/3$ -error quantum query complexity of  $f$ , which Beals et al. [8] showed to be at least  $\deg_{1/3}(f)/2$ ).
- (3) Does there exist a universal constant  $\lambda > 0$  such that  $\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \min\{C^1(f), C^0(f)\}$ ? This would resolve Mansour's conjecture.

In an earlier version of this manuscript we suggested that bounding the logarithm of the approximate spectral norm by  $O(\deg_{1/3}(f))$  or  $O(Q(f))$  might be an approach to answer Question (1) or (2) above. However, in a very recent work [17] it is shown that  $\log(\|\hat{f}\|_{1,\epsilon})$  could be as large as  $\Omega(Q(f) \cdot \log n)$ , thus nullifying the suggested approach.

## ACKNOWLEDGMENTS

Part of this work was carried out when NS and SC visited CWI, Amsterdam. SA did most of this work while a Postdoc at Center for Theoretical Physics at MIT, PhD student at QuSoft, CWI, Amsterdam, and a visitor at 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, and Fernando Vieira Costa Júnior for pointing us to the reference [7]. Finally, we thank the anonymous reviewers for many helpful comments. An extended abstract of this manuscript appeared in the conference proceedings of STACS 2020.

1925 *Funding.*

1926 Srinivasan Arunachalam: Work done when at QuSoft, CWI, Amsterdam, supported by ERC Consolidator Grant 615307  
 1927 QPROGRESS and MIT-IBM Watson AI Lab under the project *Machine Learning in Hilbert space*.  
 1928

1929 Michal Koucký: Partially supported by ERC Consolidator Grant 616787 LBCAD, and GAČR grant 19-27871X.

1930 Nitin Saurabh: Part of the work was done when the author was at IUUK Prague supported by the European Union's  
 1931 Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 616787, and at Max Planck Institut für  
 1932 Informatik, Saarbrücken, Germany.  
 1933

1934 Ronald de Wolf: Partially supported by ERC Consolidator Grant 615307 QPROGRESS (ended Feb 2019) and by NWO  
 1935 under QuantERA project QuantAlgo 680-91-034 and the Quantum Software Consortium.  
 1936

1937 **REFERENCES**

- 1938 [1] S. Aaronson and A. Ambainis. 2018. Forrelation: A Problem That Optimally Separates Quantum from Classical Computing. *SIAM Journal of*  
 1939 *Computing* 47, 3 (2018), 982–1038. Earlier in STOC'15. arXiv:1411.5729.
- 1940 [2] A. Akavia, A. Bogdanov, S. Guo, A. Kamath, and A. Rosen. 2014. Candidate Weak Pseudorandom Functions in  $AC^0 \circ MOD_2$ . In *Proceedings of the 5th*  
 1941 *Conference on Innovations in Theoretical Computer Science (ITCS'14)*. ACM, 251–260.
- 1942 [3] N. Albuquerque, F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. 2014. Sharp generalizations of the multilinear Bohnenblust-Hille inequality.  
 1943 *Journal of Functional Analysis* 266, 6 (2014), 3276–3740. arXiv:1306.3362.
- 1944 [4] N. Alon and J. Spencer. 2000. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization.
- 1945 [5] K. Amano. 2011. Tight Bounds on the Average Sensitivity of  $k$ -CNF. *Theory of Computing* 7, 4 (2011), 45–48.
- 1946 [6] A. Ambainis, M. Kokainis, and R. Kothari. 2016. Nearly Optimal Separations Between Communication (or Query) Complexity and Partitions. In *31st*  
 1947 *Conference on Computational Complexity, CCC 2016*. 4:1–4:14. Combines arXiv:1512.01210 and arXiv:1512.00661.
- 1948 [7] F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. 2014. The Bohr radius of the  $n$ -dimensional polydisk is equivalent to  $\sqrt{(\log n)/n}$ . *Advances in*  
 1949 *Mathematics* 264 (2014), 726–746.
- 1950 [8] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. 2001. Quantum Lower Bounds by Polynomials. *J. ACM* 48, 4 (2001), 778–797. Earlier  
 1951 version in FOCS'98. quant-ph/9802049.
- 1952 [9] S. Ben-David, P. Hatami, and A. Tal. 2017. Low-Sensitivity Functions from Unambiguous Certificates. In *8th Innovations in Theoretical Computer*  
 1953 *Science Conference, ITCS 2017*. 28:1–28:23. arXiv:1605.07084.
- 1954 [10] H. F. Bohnenblust and E. Hille. 1931. On the absolute convergence of Dirichlet series. *Annals of Mathematics* (1931), 600–622.
- 1955 [11] R. Boppana. 1997. The Average Sensitivity of Bounded-Depth Circuits. *Inform. Process. Lett.* 63, 5 (1997), 257–261.
- 1956 [12] J. Bourgain and G. Kalai. 1997. Influences of Variables and Threshold Intervals under Group Symmetries. *Geometric and Functional Analysis (GAFA)*  
 7, 3 (1997), 438–461.
- 1957 [13] Y. Brandman, A. Orlitsky, and J. Hennessy. 1990. A Spectral Lower Bound Technique for the Size of Decision Trees and Two-Level AND/OR Circuits.  
 1958 *IEEE Transactions of Computers* 39, 2 (1990), 282–287.
- 1959 [14] J. Briët, H. Buhrman, T. Lee, and T. Vidick. 2013. Multipartite entanglement in XOR games. *Quantum Information & Computation* 13, 3-4 (2013),  
 1960 334–360. arXiv:0911.4007.
- 1961 [15] H. Buhrman and R. de Wolf. 2002. Complexity Measures and Decision Tree Complexity: A Survey. *Theoretical Computer Science* 288, 1 (2002), 21–43.
- 1962 [16] M. Bun and J. Thaler. 2013. Dual Lower Bounds for Approximate Degree and Markov-Bernstein Inequalities. In *Automata, Languages, and*  
 1963 *Programming - 40th International Colloquium, ICALP 2013*. 303–314. arXiv:1302.6191.
- 1964 [17] S. Chakraborty, A. Chattopadhyay, N. Mandé, and M. Paraashar. 2019. Quantum Query-to-Communication simulation needs a logarithmic overhead.  
 1965 arXiv:1909.10428.
- 1966 [18] S. Chakraborty, S. Karmalkar, S. Kundu, S. V. Lokam, and N. Saurabh. 2018. Fourier Entropy-Influence Conjecture for Random Linear Threshold  
 1967 Functions. In *LATIN 2018: Theoretical Informatics - 13th Latin American Symposium, 2018*. 275–289.
- 1968 [19] S. Chakraborty, R. Kulkarni, S. V. Lokam, and N. Saurabh. 2016. Upper bounds on Fourier entropy. *Theoretical Computer Science* 654 (2016), 92–112.  
 1969 The first version appeared as a technical report TR13-052 on ECCS in 2013.
- 1970 [20] M. Cheraghchi, E. Grigorescu, B. Juba, K. Wimmer, and N. Xie. 2018.  $AC^0 \circ MOD_2$  lower bounds for the Boolean Inner Product. *J. Comput. System*  
 1971 *Sci.* 97 (2018), 45 – 59.
- 1972 [21] G. Cohen and I. Shinkar. 2016. The Complexity of DNF of Parities. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer*  
 1973 *Science* (Cambridge, Massachusetts, USA) (ITCS '16). ACM, 47–58.
- 1974 [22] T. M. Cover and J. A. Thomas. 1991. *Elements of Information Theory*. John Wiley & Sons.
- 1975 [23] B. Das, M. Pal, and V. Visavaliya. 2011. The Entropy Influence Conjecture Revisited. arxiv:1110.4301.
- 1976 [24] A. Defant, L. Frerick, J. Ortega-Cerdá, M. Ounaies, and K. Seip. 2011. The Bohnenblust-Hille inequality for homogeneous polynomials is hypercon-  
 1977 tractive. *Annals of Mathematics* 174, 1 (2011), 485–497. arXiv:0904.3540.

1976 Manuscript submitted to ACM

- 1977 [25] A. Defant, D. Popa, and U. Schwaning. 2010. Coordinatewise multiple summing operators in Banach spaces. *Journal of Functional Analysis* 259, 1  
1978 (2010), 220–242.
- 1979 [26] E. Friedgut. 1998. Boolean Functions With Low Average Sensitivity Depend On Few Coordinates. *Combinatorica* 18, 1 (1998), 27–35.
- 1980 [27] E. Friedgut and G. Kalai. 1996. Every Monotone Graph Property has a Sharp Threshold. *Proc. Amer. Math. Soc.* 124, 10 (1996), 2993–3002.
- 1981 [28] M. Göös. 2015. Lower Bounds for Clique vs. Independent Set. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015*.  
1982 1066–1076.
- 1983 [29] P. Gopalan, A. Kalai, and A. R. Klivans. 2008. A Query Algorithm for Agnostically Learning DNF?. In *21st Annual Conference on Learning Theory -  
1984 COLT 2008*. 515–516.
- 1985 [30] P. Gopalan, A. T. Kalai, and A. Klivans. 2008. Agnostically Learning Decision Trees. In *Proceedings of the 40th annual ACM symposium on Theory of  
1986 computing (STOC '08)*. 527–536.
- 1987 [31] Parikshit Gopalan, Raghu Meka, and Omer Reingold. 2013. DNF sparsification and a faster deterministic counting algorithm. *computational  
1988 complexity* 22, 2 (01 Jun 2013), 275–310.
- 1989 [32] P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson. 2016. Degree and Sensitivity: Tails of Two Distributions. In *31st Conference on Computational  
1990 Complexity, CCC 2016*. 13:1–13:23. arxiv: 1604.07432.
- 1991 [33] L. Gross. 1975. Logarithmic Sobolev inequalities. *American Journal of Mathematics* 97, 4 (1975), 1061–1083.
- 1992 [34] R. Hod. 2017. Improved Lower Bounds for the Fourier Entropy/Influence Conjecture via Lexicographic Functions. arxiv:1711.00762.
- 1993 [35] J. Kahn, G. Kalai, and Nathan Linial. 1988. The Influence of Variables on Boolean Functions. In *Proceedings of the 29th Annual IEEE Symposium on  
1994 Foundations of Computer Science*. 68–80.
- 1995 [36] G. Kalai. 2007. The Entropy/Influence Conjecture. Terence Tao’s blog: <https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/>.
- 1996 [37] N. Keller, E. Mossel, and T. Schrank. 2012. A note on the Entropy/Influence conjecture. *Discrete Mathematics* 312, 22 (2012), 3364 – 3372.  
arXiv:1105.2651.
- 1997 [38] Esty Kelman, Guy Kindler, Noam Lifshitz, Dor Minzer, and Muli Safra. 2019. Towards a Proof of the Fourier–Entropy Conjecture? *CoRR* abs/1911.10579  
1998 (2019). arXiv:1911.10579 <http://arxiv.org/abs/1911.10579>
- 1999 [39] A. Klivans, H. Lee, and A. Wan. 2010. Mansour’s Conjecture is True for Random DNF Formulas. In *Proceedings of the 23rd Conference on Learning  
2000 Theory*. 368–380.
- 2001 [40] T. Lee and A. Shraibman. 2009. Lower Bounds in Communication Complexity. *Foundations and Trends in Theoretical Computer Science* 3, 4 (2009),  
2002 263–398.
- 2003 [41] N. Linial, Y. Mansour, and N. Nisan. 1993. Constant Depth Circuits, Fourier Transform, and Learnability. *J. ACM* 40, 3 (July 1993), 607–620.
- 2004 [42] J. E. Littlewood. 1930. On bounded bilinear forms in an infinite number of variables. *The Quarterly Journal of Mathematics* 1 (1930), 164–174.
- 2005 [43] Y. Mansour. 1994. Learning Boolean Functions via the Fourier Transform. In *Theoretical Advances in Neural Computation and Learning*, V. Roy-  
2006 chowdhury, K-Y Siu, and A. Orlitsky (Eds.). Springer US, 391–424.
- 2007 [44] Y. Mansour. 1995. An  $n^{O(\log \log n)}$  Learning Algorithm for DNF under the Uniform Distribution. *J. Comput. System Sci.* 50, 3 (1995), 543–550.
- 2008 [45] A. Montanaro. 2012. Some applications of hypercontractive inequalities in quantum information theory. *J. Math. Phys.* 53, 12 (2012), 122206.  
arXiv:1208.0161.
- 2009 [46] A. Montanaro and T. Osborne. 2009. On the communication complexity of XOR functions. arXiv:0909.3392.
- 2010 [47] R. O’Donnell. 2014. *Analysis of Boolean Functions*. Cambridge University Press.
- 2011 [48] R. O’Donnell and L-Y. Tan. 2013. A Composition Theorem for the Fourier Entropy-Influence Conjecture. In *Proceedings of Automata, Languages and  
2012 Programming - 40th International Colloquium*. 780–791. arXiv:1304.1347.
- 2013 [49] R. O’Donnell, J. Wright, Y. Zhao, X. Sun, and L-Y. Tan. 2014. A Composition Theorem for Parity Kill Number. In *IEEE 29th Conference on Computational  
2014 Complexity, CCC 2014*. 144–154. arXiv:1312.2143.
- 2015 [50] R. O’Donnell, J. Wright, and Y. Zhou. 2011. The Fourier Entropy-Influence Conjecture for certain classes of Boolean Functions. In *Proceedings of  
2016 Automata, Languages and Programming - 38th International Colloquium*. 330–341.
- 2017 [51] R. O’Donnell and Y. Zhao. 2016. Polynomial Bounds for Decoupling, with Applications. In *31st Conference on Computational Complexity, CCC 2016*.  
24:1–24:18. arXiv:1512.01603.
- 2018 [52] D. Pellegrino and J. Seoane-Sepúlveda. 2012. New upper bounds for the constants in the Bohnenblust-Hille inequality. *J. Math. Anal. Appl.* 386, 1  
2019 (2012), 300–307. arXiv:1010.0461v3.
- 2020 [53] D. Pellegrino and E. V. Teixeira. 2018. Towards sharp Bohnenblust-Hille constants. *Communications in Contemporary Mathematics* 20, 3 (2018),  
2021 1750029. arXiv:1604.07595.
- 2022 [54] R. A. Servedio and E. Viola. 2012. On a special case of rigidity. Manuscript: <http://eccc.hpi-web.de/report/2012/144>.
- 2023 [55] G. Shalev. 2018. On the Fourier Entropy Influence Conjecture for Extremal Classes. (2018). arxiv:1806.03646.
- 2024 [56] R. Shaltiel and E. Viola. 2010. Hardness Amplification Proofs Require Majority. *SIAM J. Comput.* 39, 7 (2010), 3122–3154.
- 2025 [57] C. E. Shannon. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 3 (1948), 379–423.
- 2026 [58] A. A. Sherstov. 2011. The Pattern Matrix Method. *SIAM J. Comput.* 40, 6 (2011), 1969–2000. arXiv:0906.4291. Earlier version in STOC’08.
- 2027 [59] Alexander A. Sherstov. 2018. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC  
2028 2018, Los Angeles, CA, USA, June 25-29, 2018*. 311–324.

- 2029 [60] Y. Shi. 2000. Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Inform.*  
2030 *Process. Lett.* 75, 1–2 (2000), 79–83. arXiv:quant-ph/9904107.
- 2031 [61] A. Tal. 2014. Shrinkage of De Morgan Formulae by Spectral Techniques. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS*  
2032 *2014, Philadelphia, PA, USA, October 18–21, 2014*. 551–560.
- 2033 [62] A. Tal. 2017. Tight Bounds on the Fourier Spectrum of  $AC^0$ . In *32nd Computational Complexity Conference, CCC 2017*. 15:1–15:31.
- 2034 [63] P. Traxler. 2009. Variable Influences in Conjunctive Normal Forms. In *Theory and Applications of Satisfiability Testing - SAT 2009*. Springer Berlin  
2035 Heidelberg, 101–113.
- 2036 [64] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. 2013. Fourier Sparsity, Spectral Norm, and the Log-Rank Conjecture. In *54th Annual IEEE Symposium*  
2037 *on Foundations of Computer Science, FOCS 2013*. 658–667. arXiv:1304.1245.
- 2038 [65] A. Wan, J. Wright, and C. Wu. 2014. Decision Trees, Protocols and the Entropy-Influence Conjecture. In *Innovations in Theoretical Computer Science,*  
2039 *ITCS'14*. 67–80. arXiv:1312.3003.
- 2040 [66] R. de Wolf. 2008. A Brief Introduction to Fourier Analysis on the Boolean Cube. *Theory of Computing* (2008). ToC Library, Graduate Surveys 1.
- 2041 [67] S. Zhang. 2014. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete*  
2042 *Algorithms, SODA 2014*. 1878–1885. arXiv:1307.6738.
- 2043 [68] Z. Zhang and Y. Shi. 2009. Communication complexities of symmetric XOR functions. *Quantum Information & Computation* 9, 3 (2009), 255–263.

2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080