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 $f\colon \{-1,1\}^n \to \{-1,1\}$, define the *Fourier distribution* to be the distribution on subsets of $[n]$, where each $S \subseteq [n]$ is sampled with probability $\widehat{f}(S)^2$. The Fourier Entropy-influence (FEI) conjecture of Friedgut and Kalai [28] seeks to relate two fundamental measures associated with the Fourier distribution: does there exist a universal constant $C > 0$ such that $\mathbb{H}(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$, where $\mathbb{H}(\hat{f}^2)$ is the Shannon entropy of the Fourier distribution of f and $\mathsf{Inf}(f)$ is the total influence of f?

In this article, we present three new contributions toward the FEI conjecture:

(1) Our first contribution shows that $\mathbb{H}(\hat{f}^2) \leq 2 \cdot \mathsf{aUC}^{\oplus}(f)$, where $\mathsf{aUC}^{\oplus}(f)$ is the average unambiguous parity-certificate complexity of f. This improves upon several bounds shown by Chakraborty et al. [20]. We further improve this bound for unambiguous DNFs. We also discuss how our work makes Mansour's conjecture for DNFs a natural next step toward resolution of the FEI conjecture.
(2) We next consider the weaker *Fourier Min-entropy-influence* (FMEI) conjecture posed by O'Donnell and others [50, 53], which asks if $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$, where $\mathbb{H}_\infty(\hat{f}^2)$ is the min-entropy of the Fourier distribution. We show $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \mathsf{C}^{\oplus}_{\min}(f)$, where $\mathsf{C}^{\oplus}_{\min}(f)$ is the minimum parity-certificate complexity of f. We also show that for all $\varepsilon \geq 0$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq 2\log(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon))$, where $\|\widehat{f}\|_{1,\varepsilon}$ is the approximate spectral norm of f. As a corollary, we verify the FMEI conjecture for the class of read-k DNFs (for constant k).
(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

Srinivasan Arunachalam: Work done when at QuSoft, CWI, Amsterdam, supported by ERC Consolidator Grant No. 615307 QPROGRESS and MIT-IBM Watson AI Lab under the project *Machine Learning in Hilbert space*. Also partially supported by EPSRC Grant No. EP/L021005/1 while a visitor at University of Bristol. Michal Koucký: Partially supported by ERC Consolidator Grant No. 616787 LBCAD, and GAČR Grant No. 19-27871X. Nitin Saurabh: Part of the work was done when the author was at IUUK Prague supported by the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant No. 616787, and at Max Planck Institut für Informatik, Saarbrücken, Germany. Ronald de Wolf: Partially supported by ERC Consolidator Grant No. 615307 QPROGRESS (ended Feb. 2019) and by NWO under QuantERA project QuantAlgo 680-91-034 and the Quantum Software Consortium.

Authors' addresses: S. Arunachalam, IBM Research, New York, USA; email: Srinivasan.Arunachalam@ibm.com; S. Chakraborty, Indian Statistical Institute, Kolkata, India; email: sourav@isical.ac.in; M. Koucký, Computer Science Institute of Charles University, Prague, Czech Republic; email: koucky@iuuk.mff.cuni.cz; N. Saurabh, Technion-IIT, Haifa, Israel; email: nitinsau@cs.technion.ac.il; R. de Wolf, QuSoft, CWI and University of Amsterdam, Amsterdam, the Netherlands; email: 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.
© 2021 Association for Computing Machinery.

1942-3454/2021/08-ART22 $15.00
https://doi.org/10.1145/3470860

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**; • **Theory of computation** → *Oracles and decision trees;*

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

ACM Reference format:
Srinivasan Arunachalam, Sourav Chakraborty, Michal Koucký, Nitin Saurabh, and Ronald de Wolf. 2021. Improved Bounds on Fourier Entropy and Min-entropy. *ACM Trans. Comput. Theory* 13, 4, Article 22 (August 2021), 40 pages.
https://doi.org/10.1145/3470860

1 INTRODUCTION

Boolean functions $f\colon \{-1,1\}^n \to \{-1,1\}$ naturally arise in many areas of theoretical computer science and mathematics such as learning theory, complexity theory, quantum computing, inapproximability, graph theory, extremal combinatorics, and so on. 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 References [50, 69]. 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 [28]. The FEI conjecture seeks to relate the following two fundamental properties of a Boolean function $f\colon$ the *Fourier entropy* of f and the *total influence* of f, which we define now.

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

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

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

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

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

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

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

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 $f : \{-1,1\}^n \to \{-1,1\}$,*

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

The original motivation of Friedgut and Kalai for the FEI conjecture came from studying threshold phenomena of monotone graph properties in random graphs [28]. 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 [13] proved a bound of $c_\varepsilon(\log n)^{-2+\varepsilon}$ for every $\varepsilon > 0$. A very recent work of Kelman et al. [39] 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 [36] (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 $f : \{-1,1\}^n \to \{-1,1\}$, there exists an $i \in [n]$ such that $\mathsf{Inf}_i(f) \geq \mathsf{Var}(f) \cdot \Omega(\frac{\log n}{n})$, where $\mathsf{Var}(f) := 1 - \mathbb{E}_x[f(x)]^2 = 1 - \widehat{f}(\emptyset)^2$.*

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 [47] from 1995.

Conjecture 1.3 (Mansour's Conjecture). *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is computed by a t-term* DNF.[1] *Then for every $\varepsilon > 0$, there exists 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. [31], 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 ℓ_2-norm) by *sparse* polynomials over $\{-1,1\}$. In particular, for a Boolean function f and $\varepsilon > 0$, the FEI conjecture implies the existence of a polynomial p with $2^{O(\mathsf{Inf}(f)/\varepsilon)}$ monomials such that $\mathbb{E}_x[(f(x) - p(x))^2] \leq \varepsilon$. The current best known bound in this direction is $2^{O(\mathsf{Inf}(f)\log(\mathsf{Inf}(f))/\varepsilon)}$ [39].

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 [50, 53] 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 Equation (1) holds when the entropy of the Fourier distribution is replaced by the *min-entropy* of the Fourier distribution (denoted $\mathbb{H}_\infty(\hat{f}^2)$).

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

The min-entropy of $\{\widehat{f}(S)^2\}_S$ is defined as

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

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

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

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

Another way to formulate the FMEI conjecture is, suppose $f : \{-1,1\}^n \to \{-1,1\}$, then does there exist a Fourier coefficient $\widehat{f}(S)$ such that $|\widehat{f}(S)| \geq 2^{-O(\mathsf{Inf}(f))}$? By the *granularity* of Fourier coefficients it is well-known that *every* Fourier coefficient of a Boolean function f is an integral multiple of $2^{-\deg(f)}$, see Reference [50, Exercise 1.11] for a proof of this. (Here the $\deg(f)$ refers to the degree of the unique multilinear polynomial that represents f.) The FMEI conjecture asks if we can prove a lower bound of $2^{-O(\mathsf{Inf}(f))}$ on *any one* Fourier coefficient, and even this remains open. Proving the FMEI conjecture seems to require proving interesting structural properties of Boolean functions. It was observed by Reference [53] that the FMEI conjecture suffices to imply the KKL theorem for balanced functions (see also Section 4). In fact, it suffices[2] for most of the implications that follows from the FEI conjecture (see Reference [39]).

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 $C \circ \mathsf{MOD}_2$. In particular, they conjectured that min-entropy of f_A is only polynomially larger than f when $f \in \mathsf{AC}^0[\mathsf{poly}(n), O(1)]$. ($\mathsf{AC}^0[s,d]$ is the class of unbounded fan-in circuits of size at most s and depth at most d.) It is well-known that when $f \in \mathsf{AC}^0[s,d]$, $\mathbb{H}_\infty(\hat{f}^2)$ is at most $O((\log s)^{d-1} \cdot \log\log s)$ [12, 42, 65]. Depending on the tightness of the relationship between $\mathbb{H}_\infty(\hat{f}^2)$ and $\mathbb{H}_\infty(\widehat{f_A}^2)$, one could obtain near-optimal lower bound on the size of $\mathsf{AC}^0[s,d]\circ\mathsf{MOD}_2$ circuits computing IP (inner-product-mod-2). This problem has garnered a lot of attention in recent times for a variety of reasons [2, 21, 22, 57, 59]. The current best known lower bound for IP against $\mathsf{AC}^0[s,d]\circ\mathsf{MOD}_2$ is quadratic when $d = 4$, and only super-linear for all $d = O(1)$ [21].

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 [28] posed the FEI conjecture in 1996, there was not much work done toward resolving it, until the work of Klivans et al. [40] 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 [37] for additional discussions on the FEI conjecture.

[2]With a minor strengthening: the large Fourier coefficient is required to be associated with a Fourier character of size $O(\mathsf{Inf}(f))$.

The FEI conjecture is known to be true when we replace the universal constant C with $\log n$ in Equation (1). In fact, we know $\mathbb{H}(\hat{f}^2) \leq O(\mathsf{Inf}(f) \cdot \log n)$ for *real-valued* functions $f : \{-1,1\}^n \to \mathbb{R}$ (see References [38, 53] for a proof and Reference [20] for an improvement of this statement).[3] If we strictly require C to be a universal constant, then the FEI conjecture is known to be false for real-valued functions. Instead, for real-valued functions an analogous statement called the *logarithmic Sobolev Inequality* [34] is known to be true. The logarithmic Sobolev inequality states that for every $f : \{-1,1\}^n \to \mathbb{R}$, we have $\mathsf{Ent}(f^2) \leq 2 \cdot \mathsf{Inf}(f)$, where $\mathsf{Ent}(f)$ is defined as $\mathsf{Ent}(f) = \mathbb{E}[f \ln(f)] - \mathbb{E}[f] \ln(\mathbb{E}[f])$, 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. [53] showed that the FEI conjecture holds for symmetric Boolean functions and read-once decision trees. Keller et al. [38] 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. [20] and O'Donnell and Tan [51], 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 Reference [51] for more). For general Boolean functions, Chakraborty et al. [20] gave several upper bounds on the Fourier entropy in terms of combinatorial quantities larger than the total influence, e.g., average decision tree depth, and so on, and sometimes even quantities that could be much smaller than influence, namely, average parity-decision tree depth.

Later Wan et al. [68] used Shannon's source coding theorem [60] (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 [58] 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. [24] and for random linear threshold functions (LTFs) by Chakraborty et al. [19].

There has also been some work in giving lower bounds on the constant C in the FEI conjecture. Hod [35] 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 [51].

However, there has not been much work on the FMEI conjecture. It was observed in References [19, 53] 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 [58].

1.2 Our Contributions

Our contributions in this article 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 article is to give a better upper bound on the Fourier entropy $\mathbb{H}(\hat{f}^2)$ in terms of $\mathsf{aUC}(f)$, the *average unambiguous certificate complexity* of f. Informally, the unambiguous certificate complexity $\mathsf{UC}(f)$ 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

[3] For Boolean functions, the $\log n$-factor was improved by Reference [33] to $\log(\mathsf{s}(f))$ (where $\mathsf{s}(f)$ is the sensitivity of the Boolean function f).

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, $\mathrm{aUC}(f)$, 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 $\mathrm{UC}(f)$ in terms of other combinatorial measures such as decision-tree complexity, sensitivity, quantum query complexity, and so on, see References [6, 10, 29] for more. It follows from definitions that $\mathrm{UC}(f)$ lower bounds decision tree complexity. However, it is known that $\mathrm{UC}(f)$ can be quadratically smaller than decision tree complexity [6]. Our main contribution here is an improved upper bound of *average unambiguous certificate complexity* $\mathrm{aUC}(f)$ on $\mathbb{H}(\hat{f}^2)$. This improves upon the previously known bound of average decision tree depth on $\mathbb{H}(\hat{f}^2)$ [20].

Theorem 1.5. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then,*

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

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. [38, 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. [68] used Shannon's source coding theorem [60] to verify the conjecture for bounded-read decision trees.

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

$$f^M(x^1,\ldots,x^M) := f(x^1_1,\ldots,x^1_n)\cdot f(x^2_1,\ldots,x^2_n)\cdots f(x^M_1,\ldots,x^M_n).$$

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

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

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

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

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

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

Looking finely into how certificates contribute to Fourier coefficients in the proof above, we further strengthen Theorem 1.5 by showing that we can replace $\mathrm{aUC}(f)$ by the *average unambiguous parity-certificate complexity* $\mathrm{aUC}^{\oplus}(f)$ of f. Here $\mathrm{aUC}^{\oplus}(f)$ is defined similar to $\mathrm{aUC}(f)$ 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 $\mathbb{H}(\hat{f}^2)$ [20]. It is easily seen that $\mathrm{aUC}^{\oplus}(f)$ lower bounds the average parity-decision tree depth.

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

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

The proof outline remains the same as in Theorem 1.5. However, a particular certificate in $\mathscr{C}^M$ 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 $2^{M(\mathrm{aUC}^{\oplus}(f,\mathscr{C})+\delta)}$ Fourier coefficients of f^M. Nevertheless, by the AEP theorem, we still obtain that a typical parity-certificate fixes at most $M(\mathrm{aUC}^{\oplus}(f,\mathscr{C}) + \delta)$ 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 $2^{M(\mathrm{aUC}^{\oplus}(f,\mathscr{C})+\delta)}$ Fourier coefficients of f^M. The rest of the proof now follows analogously.

1.3.1 Further Extension to Unambiguous DNF*s.* Consider an unambiguous certificate $\mathscr{C} = \{C_1, \ldots, C_t\}$ of f. It covers both 1 and -1 inputs of f. Suppose $\{C_1, \ldots, C_{t_1}\}$ for some $t_1 < t$ is a partition of $f^{-1}(-1)$ and $\{C_{t_1+1}, \ldots, C_t\}$ 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 $\{C_1, \ldots, C_{t_1}\}$ 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

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

where 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 $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function and $\mathscr{C} = \{C_1, \ldots, C_t\}$ be a monochromatic affine subspace partition of $\{-1,1\}^n$ with respect to f such that $\{C_1, \ldots, C_{t_1}\}$ for some $t_1 < t$ is an affine subspace partition of $f^{-1}(-1)$ and $\{C_{t_1+1}, \ldots, C_t\}$ is an affine subspace partition of*

$f^{-1}(1)$. *Further,* $p := \Pr_x[f(x) = 1]$. *Then,*

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

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 $\{C_1, \ldots, C_{t_1}, f^{-1}(1)\}$, which naturally inherits a distribution $\mathbf{C}$ 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, unlike before, the probability of observing a strongly typical sequence does not capture the number of coefficients it could contribute to $\mathcal{B}$. Here, we use stronger properties guaranteed by the strong AEP. In particular, it guarantees that the *empirical* distribution of a typical sequence is close to the distribution of $\mathbf{C}$. In contrast, the (weak) AEP only guarantees that the *empirical* entropy of a typical sequence is close to the entropy of $\mathbf{C}$. Using the stronger property we can now lower bound the magnitude of any non-zero Fourier coefficient in the Fourier expansion of the indicator function of a strongly typical sequence. We then use Parseval's Identity (Fact 2.1) to deduce an upper bound on its 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 negligible Fourier weight, which can be done as before. Using the two properties, we can now complete the proof.

We also note that a similar bound for the general DNF representation, i.e., when $\{C_1, \ldots, C_{t_1}\}$ is an arbitrary DNF representation of f where the C_is need not be disjoint, suffices to establish Mansour's conjecture (Conjecture 1.3). In fact, following the analogy, Theorem 1.7 implies a bound of "average 1-certificate complexity" in the general case. In this direction, in Section 3.3, we observe that a weaker bound of 1-certificate complexity, i.e., showing $\mathbb{H}(\widehat{f^2}) \leq O(\mathbf{C}_1(f))$, would already suffice to answer Mansour's conjecture positively. For details, see Section 3.3.

1.4 New Upper Bounds for the FMEI Conjecture

Given the hardness of obtaining better upper bounds on the Fourier entropy of a Boolean function, we make progress on a weaker conjecture, the FMEI conjecture. The FMEI conjecture is much less studied than the FEI conjecture. In fact, we are aware of only one very recent paper [58] that studies the FMEI conjecture for a particular class of functions. Our second contribution is to give upper bounds on the min-entropy of general Boolean functions in terms of the minimum parity-certificate complexity (denoted $\mathbf{C}^{\oplus}_{\min}(f)$) and the approximate spectral norm of Boolean functions (denoted $\|\widehat{f}\|_{1,\varepsilon}$). The minimum parity-certificate complexity $\mathbf{C}^{\oplus}_{\min}(f)$ is also referred to as the parity kill number by O'Donnell et al. [52] and is closely related to the communication complexity of XOR functions [49, 67, 71]. The approximate spectral norm $\|\widehat{f}\|_{1,\varepsilon}$ is related to the *quantum* communication complexity of XOR functions [41, 70]. In particular, it characterizes the bounded-error quantum communication complexity of XOR functions with constant $\mathbb{F}_2$-degree [70]. (By $\mathbb{F}_2$-degree, we mean the degree of a function when viewed as a polynomial over $\mathbb{F}_2$.)

Theorem 1.8. *Let* $f\colon \{-1,1\}^n \to \{-1,1\}$ *be a Boolean function. Then,*

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

(ii) $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \mathbf{C}^{\oplus}_{\min}(f)$.

(iii) $\mathbb{H}_\infty(\hat{f}^2) \le 2(1+\log_2 3)\cdot R_2^{\oplus}(f)$.[5]

The proof of Theorem 1.8 (i) expresses the quantity $\|\widehat{f}\|_{1,\varepsilon}$ as a (minimization) linear program. We consider the dual linear program and exhibit a feasible solution that achieves an optimum of $(1-\varepsilon)/\max_S |\widehat{f}(S)|$. This proves the desired inequality. To prove part (ii) and (iii) of the theorem, the idea is to consider a "simple" function g that has "good" correlation with f, and then upper bound the correlation between f and g using Plancherel's theorem (Fact 2.1) and the fact that g has a "simple" Fourier structure. For part (ii), g is chosen to be the indicator function of an (affine) subspace where f is constant, whereas for part (iii) the randomized parity decision tree computing f itself plays the role of g.[6]

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 $\mathbb{H}_{1+\delta}(\hat{f}^2) \ge \mathbb{H}_\infty(\hat{f}^2)$ for every $\delta \ge 0$ and as $\delta \to \infty$, $\mathbb{H}_{1+\delta}(\hat{f}^2)$ converges to $\mathbb{H}_\infty(\hat{f}^2)$. Also, $\mathbb{H}_1(\hat{f}^2)$, obtained by taking the limit as $\delta \to 0$, 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 $f\colon \{-1,1\}^n \to \{-1,1\}$ that can be expressed as a read-k DNF, we have*

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

This theorem improves upon a recent (and independent) result of Shalev [58] 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 Reference [58] for a precise definition). To prove Theorem 1.9, we essentially show that $\mathsf{Inf}(f)$ is at least as large as $\mathsf{C}_{\min}(f)$, for read-k DNFs (Lemma 4.8), where $\mathsf{C}_{\min}(f)$ 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. [39] shows that the FMEI conjecture holds up to a logarithmic factor in $\mathsf{Inf}(f)$. In general, they show that almost all the Fourier weight lies on Fourier coefficients of size at most $O(\widetilde{\mathsf{Inf}(f)})$ and weight at least $2^{-O\left(\widetilde{\mathsf{Inf}(f)}\log\left(1+\widetilde{\mathsf{Inf}(f)}\right)\right)}$, where $\widetilde{\mathsf{Inf}(f)} := \frac{\mathsf{Inf}(f)}{\mathsf{Var}(f)}$. 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 $\mathsf{C}^{\oplus}_{\min}(f)$ is much smaller than $\widetilde{\mathsf{Inf}(f)}\log(1+\widetilde{\mathsf{Inf}(f)})$ (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. 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. [39] essentially choose g to be the truncation of the Fourier spectrum of f to the Fourier coefficients of size at most $O(\widetilde{\mathsf{Inf}(f)})$ and weight *at most* $2^{-\Omega\left(\widetilde{\mathsf{Inf}(f)}\log\left(1+\widetilde{\mathsf{Inf}(f)}\right)\right)}$. 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

[5] $R_2^{\oplus}(f)$ is the randomized parity-decision tree complexity of f (we define this formally in Section 2).

[6] 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 formulation of $\mathbb{H}_\infty(\hat{f}^2)$ might help obtain better bounds, such as fractional block sensitivity.

weight on characters of size at most $O(\widetilde{\mathbf{Inf}(f)})$, by Markov's inequality, is a large constant (at least $1 - \mathrm{Var}(f)/100$).

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 ε-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 [61] 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 $\sum_S |\widehat{p}(S)|$)? To understand these questions better, we restrict ourselves to a class of polynomials called *flat* polynomials over $\{-1,1\}^n$, 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 [11], but more recently have found applications in quantum computing [1, 48], classical and quantum XOR games [15], and polynomial decoupling [54]. In the functional analysis literature homogeneous block-multilinear polynomials are known as *multilinear forms*. In an ingenious work [11], Bohnenblust and Hille showed that for every degree-d multilinear form $p\colon (\mathbb{R}^n)^d \to \mathbb{R}$, we have

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

where C_d is a constant that depends on d. In Reference [11], they showed that it suffices to pick C_d to be exponential in d to satisfy the equation above. For $d = 2$, Equation (2) generalizes Littlewood's

famous 4/3-inequality [43]. Equation (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, and so on. There has been a long line of work on improving the constant C_d in the BH inequality (to mention a few [3, 8, 25, 26, 56]). 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) to satisfy Equation (2).

In our context, using the best known bound on C_d in the BH-inequality implies that a flat homogeneous 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 1/3-approximate a Boolean function.*

Moreover, the above theorem, with an added restriction of homogeneity on p, 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. 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 1/3-approximates a Boolean function f. Then,* $\deg(f) \leq d$.

Organization. In the remainder of the article, 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) that is a consequence of the FEI conjecture and make partial progress (Theorem 1.12) toward 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 $\{1, 2, \ldots, n\}$ by $[n]$. A *partial assignment* of $[n]$ is a map $\tau : [n] \to \{-1, 1, *\}$. Define $|\tau| = |\tau^{-1}(1) \cup \tau^{-1}(-1)|$. A subcube of the Boolean cube $\{-1, 1\}^n$ is a set of $x \in \{-1, 1\}^n$ that agrees with some partial assignment τ, i.e., $\{x \in \{-1, 1\}^n : x_i = \tau(i)$ for every i with $\tau(i) \neq *\}$.

Fourier Analysis. We recall some definitions and basic facts from analysis of Boolean functions, referring to References [50, 69] 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 $\chi_S : \{-1, 1\}^n \to \{-1, 1\}$ 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 $f : \{-1, 1\}^n \to \mathbb{R}$ 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 $\mathcal{S}$ 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* $f, g\colon \{-1,1\}^n \to \mathbb{R}$,

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

In particular, if $f\colon \{-1,1\}^n \to \{-1,1\}$ is Boolean-valued and $g = f$, then 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 follows.

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

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

Similarly, we can also define the Rényi Fourier entropy.

Definition 2.3 (Rényi Fourier Entropy). For $f\colon \{-1,1\}^n \to \{-1,1\}$, $\alpha \geq 0$ and $\alpha \neq 1$, the Rényi Fourier entropy of f of order α is defined as

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

It is known that in the limit as $\alpha \to 1$, $\mathbb{H}_\alpha(\hat{f}^2)$ is the (Shannon) Fourier entropy $\mathbb{H}(\hat{f}^2)$ (see Reference [23, Chapter 17, Section 8]) and when $\alpha \to \infty$, observe that $\mathbb{H}_\alpha(\hat{f}^2)$ converges to $\mathbb{H}_\infty(\hat{f}^2)$. It is easily seen that

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

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

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

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

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

In terms of the Fourier coefficients of f, it can be shown, e.g., Reference [36], that $\mathsf{Inf}_i(f) = \sum_{S \ni i} \widehat{f}(S)^2$, and therefore

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

The *variance* of a real-valued function f is given by $\mathrm{Var}(f) := \sum_{S \neq \emptyset} \widehat{f}(S)^2 = 1 - \widehat{f}(\emptyset)^2$. It easily follows that $\mathrm{Var}(f) \leq \mathsf{Inf}(f)$. We now introduce some basic complexity measures of Boolean functions that we use often, referring to Reference [16] for more.

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

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

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

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

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

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

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

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

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

as the subcube generated by τ. We call $C \subseteq \{-1,1\}^n$ a *subcube* of $\{-1,1\}^n$ if there exists a partial assignment τ such that $C = S_\tau$ and the co-dimension of C is the number of bits fixed by τ, i.e., co-dim$(C) = |\{i \in [n] : \tau(i) \neq *\}|$. A set of subcubes $\mathscr{C} = \{C_1, \ldots, C_m\}$ *partitions* $\{-1,1\}^n$ if the subcubes are disjoint and they cover $\{-1,1\}^n$, i.e., $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_i C_i = \{-1,1\}^n$.

An *unambiguous certificate* $\mathcal{U} = \{C_1, \ldots, C_m\}$, also referred to as a *subcube partition*, is a set of subcubes partitioning $\{-1,1\}^n$. We say $\mathcal{U}$ computes a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ if f is constant on each C_i (i.e., $f(x)$ is the same for all $x \in C_i$). For an unambiguous certificate $\mathcal{U}$, the

unambiguous certificate complexity $\mathsf{UC}(\mathcal{U}, x)$ on input x equals co-dim(C_i) for the C_i satisfying $x \in C_i$. Define the *average unambiguous certificate complexity* of f with respect to $\mathcal{U}$ as

$$\mathsf{aUC}(f, \mathcal{U}) := \mathbb{E}_x[\mathsf{UC}(\mathcal{U}, x)].$$

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

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

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

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

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

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

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

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

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

A parity-decision tree for f is similar to a deterministic decision tree, except that each node in the tree is labelled by 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 for $i \in S$ and continues with the subtree corresponding to the value of $\oplus_{i \in S} x_i$. Note that if the subsets at each node have size $|S| = 1$, then we get the standard deterministic decision tree model.

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

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

Information Theory. We need some preliminaries from Information theory. We start with the following consequence of the law of large numbers, called the *Asymptotic Equipartition Property (AEP)* or *the Shannon-McMillan-Breiman theorem*. See Chapter 3 in the book [23] for more details.

Theorem 2.4 (Asymptotic Equipartition Property (AEP) Theorem). *Let* $\mathbf{X}$ *be a random variable drawn from a distribution* P *and suppose* $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_M}$ *are independently and identically distributed copies of* $\mathbf{X}$*, then*

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

in probability as $M \to \infty$.

We now recall the notion of typical sequences.

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 $(x_1, x_2, \ldots, x_M) \in \mathbf{X_1} \times \mathbf{X_2} \times \cdots \times \mathbf{X_M}$ such that

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

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

Theorem 2.6 ([23, Theorem 3.1.2]). *Let* $\varepsilon \geq 0$ *and* $T_\varepsilon^{(M)}(\mathbf{X})$ *be a typical set with respect to* P*, then*

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

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

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

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 ([23, 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 $\rho = (x_1, x_2, \ldots, x_M) \in \mathbf{X_1} \times \mathbf{X_2} \times \cdots \times \mathbf{X_M}$ such that for every x, $N(x;\rho) = 0$ 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 ρ.

The strongly typical set shares similar properties with its (weak) typical counterpart, which we state now. See Reference [23, 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* $\delta \to 0$ *as* $\varepsilon \to 0$*, and the following hold*:

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

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

for M *sufficiently large.*

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

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

(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 article: a new improved upper bound on the Fourier entropy of Boolean functions. It is well-known that $\mathsf{Inf}(f)$ lower bounds many combinatorial measures associated with Boolean functions such as decision tree depth, certificate complexity, sensitivity, and so on. 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. [20] 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*. To keep the presentation clear, we first prove a weaker upper bound of *average unambiguous certificate complexity* $\mathsf{aUC}(f)$ 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* $\mathsf{aUC}^{\oplus}(f)$.

3.1 Average Unambiguous Certificate Complexity

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

Theorem 3.1 (Restatement of Theorem 1.5). *For every Boolean function $f:\{-1,1\}^n \to \{-1,1\}$, we have*

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

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

Proof. It suffices to show that for every unambiguous certificate $\mathscr{C}$ for f, we have

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

Let $\mathscr{C} := \{C_1,\ldots,C_t\}$ be an unambiguous certificate of f. For every C_i, let $\tau(C_i)$ be the partial assignment $\tau(C_i):[n]\to\{-1,1,*\}$, corresponding to the bits fixed by C_i. Let $\mathsf{supp}(\tau(C_i)) \subseteq [n]$ be the set of indices where $\tau(C_i)\neq *$. As a shorthand, we denote $|\mathsf{supp}(\tau(C_i))|$ by $|\tau(C_i)|$. Consider the Boolean function $f^M:\{-1,1\}^{Mn}\to\{-1,1\}$ given by the *M-fold iterated product* of f with itself over distinct variables, i.e.,

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

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

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

Let $\mathbf{C}$ be a subcube-valued random variable that equals C_i with probability $2^{-|\tau(C_i)|}$.[7] Further, let $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_M$ be i.i.d. random copies of $\mathbf{C}$. For a choice of $\delta > 0$, let $T_\delta^{(M)}(\mathbf{C})$ be the typical set with respect to the distribution $\{2^{-|\tau(C_i)|}\}_{i\in[t]}$. We now define a set $\mathcal{B}$ of Fourier coefficients of f^M, which we argue below to have large Fourier weight:

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

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

Claim 3.1.

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

Proof of Claim. We first bound the size of $T_\delta^{(M)}(\mathbf{C})$ and then count contributions of a typical sequence $(C_1, \ldots, C_M)$ 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 is at most $2^{M(\mathbb{H}(\mathbf{C})+\delta)}$. For the second bound, observe that $(C_1, \ldots, C_M)$ contributes a set $(S_1, \ldots, S_M)$ to $\mathcal{B}$ if and only if $S_i \subseteq \mathrm{supp}(\tau(C_i))$ for all $i \in [M]$. Therefore, the maximum possible contribution of a typical sequence is bounded by

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

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 inequality follows from Definition 2.5 of typical sets.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\sum_{S\notin\mathcal{S}} \widehat{g}(S)^2 = \delta$. To upper bound the Fourier entropy, we first express $\mathbb{H}(\hat{g}^2)$ as follows:

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

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

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

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

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

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

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

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

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

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* $\mathsf{aUC}^{\oplus}(f)$. Recall that an unambiguous $\oplus$-certificate $\mathscr{C} = \{C_1, \ldots, C_t\}$ for f is a collection of monochromatic *affine subspaces* that together partition the space $\{-1,1\}^n$. (Observe that a subcube is a special type of affine subspace.) Analogously, the *average unambiguous $\oplus$-certificate complexity* of f with respect to $\mathscr{C}$, denoted by $\mathsf{aUC}^{\oplus}(f,\mathscr{C})$, equals $\mathbb{E}_x[\mathsf{UC}^{\oplus}(\mathscr{C},x)]$ and $\mathsf{aUC}^{\oplus}(f) := \min_{\mathscr{C}} \mathsf{aUC}^{\oplus}(f,\mathscr{C})$. 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 $\mathscr{C}$. 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 Equation (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 $(S_1, \ldots, S_M)$ be an M-tuple where each $S_i \subseteq [n]$. Then, we define $\mathcal{B}$ by letting $(S_1, \ldots, S_M) \in \mathcal{B}$ if and only if there exists a typical sequence $(C_{i_1}, \ldots, C_{i_M}) \in T_\delta^{(M)}(\mathbf{C})$ such that for all $j \in [M]$, $\mathbb{1}_{S_j} \in \mathrm{span}\langle A_{i_j}\rangle$ (where by $\mathrm{span}\langle A_{i_j}\rangle$, 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 $f\colon \{-1,1\}^n \to \{-1,1\}$ be any Boolean function. Then,*

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

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

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

We note that the reduction in Reference [68, Proposition E.2] shows that removing the requirement $\mathsf{Inf}(f) \ge 1$ from the above inequality will prove the FEI conjecture for all Boolean functions with $\mathsf{Inf}(f) \ge \log n$. Furthermore, if we could show the FEI conjecture for Boolean functions f where $\mathsf{aUC}^{\oplus}(f) = \omega(1)$ is a slow-growing function of n, again the padding argument in Reference [68] 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 $\mathscr{C} := \{C_1, \ldots, C_t\}$ be an unambiguous certificate of f. It covers both 1 and -1 inputs of f. Suppose $\{C_1, \ldots, C_{t_1}\}$ for some $t_1 < t$ is a subcube partition of $f^{-1}(-1)$ and $\{C_{t_1+1}, \ldots, C_t\}$ 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 $\{C_1, \ldots, C_{t_1}\}$ 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 $\tau(C_i)\colon [n] \to \{-1, 1, *\}$ corresponding to the bits fixed by C_i, and $|\tau(C_i)|$ denotes the size of the set of indices where $\tau(C_i) \ne *$. We know

that

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

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

Proposition 3.3. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function and $\mathscr{C} = \{C_1,\dots,C_t\}$ be an unambiguous certificate of f. Then,*

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

Therefore, it is 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 $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function and $\mathscr{C} = \{C_1,\dots,C_t\}$ 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,*

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

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_is 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

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

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 $\{C_1, \ldots, C_{t_1}\}$ 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 $\mathbf{C}$ that equals one of the C_i, $1 \le i \le t_1$, with probability $2^{-|\tau(C_i)|}$ and equals the set $f^{-1}(1) := \{x \mid f(x) = 1\}$ with probability p.[8] We consider the M-fold iterated product of this partition $\{C_1, \ldots, C_{t_1}, f^{-1}(1)\}$ 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 ρ. Furthermore, for $\rho \in \mathcal{P}$, let $f^M(\rho) \in \{-1,1\}$ be the value of f^M on the part ρ. Then, on an input $z \in \{-1,1\}^{Mn}$,

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

where $T_\varepsilon^{*(M)}(\mathbf{C})$ is the strongly typical set with respect to the distribution of $\mathbf{C}$ 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:

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

Claim 3.3.

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

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

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

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

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

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

Therefore,

$$\text{sparsity}(\widehat{\mathbb{1}_\rho}) \cdot \min_{(S_1,\ldots,S_M)} \widehat{\mathbb{1}_\rho}(S_1 \cup \cdots \cup S_M)^2 \le \Pr_z\left[\mathbb{1}_\rho(z) = 1\right] \le 2^{-M(\mathbb{H}(\mathbf{C})-\delta)}, \tag{10}$$

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 $\rho \in \{C_1, \ldots, C_{t_1}, f^{-1}(1)\}^M$, the Fourier expansion of the indicator function $\mathbb{1}_\rho$ is a product of Fourier expansion of indicator functions of $C_1, \ldots, C_{t_1}$ 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

[8] Observe that $\sum_{i=1}^{t_1} 2^{-|\tau(C_i)|} = 1 - p$, hence $\mathbf{C}$ is a valid random variable.

$1 - \sum_{i=1}^{t_1} \mathbb{1}_{C_i}$. Hence, a lower bound on the magnitude of a non-zero Fourier coefficient of $\mathbb{1}_\rho$ is given by

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

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

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

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 $T_\varepsilon^{*(M)}(\mathbf{C})$, we obtain the claimed bound on the size of $\mathcal{B}$, thereby completing the proof of the claim. □

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

Claim 3.4.

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

Proof of Claim. We omit the proof as it is similar to the proof of Claim 3.2. □

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

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

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 $\{C_1,\ldots,C_t\}$ forms a monochromatic affine subspace partition. We state the generalization below without the proof.

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

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

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 [46]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function that is representable by a t-term* DNF. *For every constant $\varepsilon > 0$, there exists a polynomial p over $\{-1,1\}$ with sparsity* $\mathsf{poly}(t)$ *such that $\mathbb{E}_x[(f(x)-p(x))^2] \le \varepsilon$. (The exponent in* $\mathsf{poly}(t)$ *can depend on $1/\varepsilon$.)*

In fact, Mansour's original conjecture states that sparsity of the polynomial p (in the conjecture above) can be taken to be $t^{O(\log \frac{1}{\varepsilon})}$. Mansour's conjecture has a number of important consequences. For instance, Gopalan et al. [31] 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 [30] in computational learning theory.

Earlier, we saw that $\mathbb{H}(\hat{f}^2) \le O(\mathrm{aUC}^{\oplus}(f))$. 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 $O(\min\{\mathrm{C}^0(f), \mathrm{C}^1(f)\})$. In this section, we observe that this bound on the Fourier-entropy in terms of $\mathrm{C}^0(f), \mathrm{C}^1(f)$ (which is clearly weaker than the FEI conjecture) suffices to establish Mansour's conjecture.

We remark that it was implicit in previous works [30, 37, 40, 53] that one does not 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 toward 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 $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbb{H}(\hat{f}^2) \le \lambda \cdot \min\left\{\mathrm{C}^0(f), \mathrm{C}^1(f)\right\}.$$

It is weaker than the FEI conjecture, because $\mathsf{Inf}(f) \le \min\{\mathrm{C}^0(f), \mathrm{C}^1(f)\}$ [5, 12, 66]. 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 $f : \{-1,1\}^n \to \{-1,1\}$, we have*

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

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 $\mathrm{C}^0(f)$ and $\mathrm{C}^1(f)$ 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 Reference [32], 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. It is also worthwhile to mention the recent works on DNF sparsification [44, 45] that can be used to establish the reverse implication too (see Reference [32, Lemma 6.4]).

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

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

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

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

By Parseval's identity (Fact 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

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

where the inequality follows from the Markov inequality.

Using traingle inequality on ℓ_2-norm, we can easily bound $\mathbb{E}[(f-p)^2]$:

$$\mathbb{E}\left[(f-p)^2\right] = \mathbb{E}\left[(f-g+g-p)^2\right] \leq (\sqrt{\delta_1}+\sqrt{\delta_2})^2,$$

where the inequality used substitutions from Equations (12) and (13). 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

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

We end this section with another open problem that could form an intermediate step toward 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 $f\colon \{-1,1\}^n \to \{-1,1\}$,*

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

4 BETTER BOUNDS ON FOURIER MIN-ENTROPY

The 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 References [50, 53] 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 $f\colon \{-1,1\}^n \to \{-1,1\}$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \mathsf{Inf}(f)$.*

Although the FMEI conjecture is a natural first step toward proving the FEI conjecture, it is also interesting in its own right. For example, it is known to imply the famous KKL theorem [36]. In

fact, until very recently,[9] we did not know of any proof of the KKL theorem that did not go through hypercontractivity or logarithmic Sobolev inequalities, which makes proving the FMEI conjecture even more interesting. The KKL theorem states that for every Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, there exists an index $i \in [n]$ such that $\mathsf{Inf}_i(f) \geq \mathsf{Var}(f) \cdot \Omega(\frac{\log n}{n})$. We now present an argument of Reference [53] that shows how a positive answer to the FMEI conjecture implies the KKL theorem for balanced functions: assume that f is balanced (i.e., $\widehat{f}(\emptyset) = \mathbb{E}_x[f(x)] = 0$). Then the FMEI conjecture implies the existence of $\emptyset \neq T \subseteq [n]$ such that $\widehat{f}(T)^2 \geq 2^{-C \cdot \mathsf{Inf}(f)}$. Furthermore, for every $i \in T$, we have

$$\mathsf{Inf}_i(f) = \sum_{S : S \ni i} \widehat{f}(S)^2 \geq \widehat{f}(T)^2 \geq 2^{-C \cdot \mathsf{Inf}(f)} \geq 2^{-Cn \cdot \max_j \{\mathsf{Inf}_j(f)\}}, \tag{14}$$

where the first inequality follows, because T contains i, the second inequality follows from the FMEI conjecture and the last inequality, because $\mathsf{Inf}(f) \leq n \cdot \max_j\{\mathsf{Inf}_j(f)\}$. However, note that $\max_j\{\mathsf{Inf}_j(f)\}$ clearly upper bounds the left-hand side of Equation (14). Thus, we have

$$\max_{j \in [n]} \left\{\mathsf{Inf}_j(f)\right\} \geq 2^{-Cn \cdot \max_j \{\mathsf{Inf}_j(f)\}}.$$

Rearranging this inequality, we obtain $\max_{j\in[n]}\{\mathsf{Inf}_j(f)\} \geq \Omega(\frac{\log n}{n})$, which is the KKL theorem for balanced functions. The proof can also be extended to non-balanced functions though using the FEI conjecture (see Reference [53]).

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

Theorem 4.2 (Restatement of Theorem 1.8). *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then,*

(i) For every $\varepsilon \geq 0$, $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \log(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon))$,
(ii) $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \mathrm{C}^{\oplus}_{\min}(f)$,
(iii) $\mathbb{H}_\infty(\hat{f}^2) \leq 2(1+\log_2 3) \cdot R^{\oplus}_2(f)$.

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

Proof of Theorem 4.2. We prove the three parts separately as follows.

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

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

[9] Eldan and Gross [27] developed new techniques using stochastic analysis that give a new way to prove the KKL theorem among other things.

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

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

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

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

Observe that both linear programs are feasible for $d \ge \deg_\varepsilon(f)$. 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 $\|\widehat{f}\|_{1,\varepsilon,d}$.[10]

Lemma 4.3. *Let $f : \{-1,1\}^n \to \{-1,1\}$, $\varepsilon \ge 0$, and $d \in [n]$ such that $d \ge \deg_\varepsilon(f)$. Then, $\|\widehat{f}\|_{1,\varepsilon,d} \ge T$ if and only if there exists a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ satisfying $|\widehat{\phi}(S)| \le 2^{-n}$ for all $|S| \le d$ and*

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

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

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

The equality holds, since $\phi(x) f(x) = |\phi(x)|$. Now, by Lemma 4.3, we have

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

Therefore, we obtain

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

Since d is arbitrary, 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 $\mathrm{C}^{\oplus}_{\min}(f) = k$. By definition of $\mathrm{C}^{\oplus}_{\min}(f)$, there exists an affine subspace $H \subseteq \{-1,1\}^n$ such that 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 co-dim(H) equals k, H is given by a set of k (linearly independent)

[10] We remark that similar linear program characterizations of approximate degree of Boolean functions have appeared before in the works of Sherstov [62] and Bun and Thaler [17].

parity constraints. That is, there exist k linearly independent vectors $S_1, \ldots, S_k \in \{0,1\}^n$, and $b_1, \ldots, b_k \in \{-1,1\}$, such that

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

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

$$\mathbb{1}_H(x) = \mathrm{AND}(-b_1 y_1, -b_2 y_2, \ldots, -b_k y_k). \tag{15}$$

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 Equation (15), it follows that

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

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

$$\langle f, \mathbb{1}_H \rangle = \mathbb{E}_x\left[f(x) \cdot \mathbb{1}_H(x)\right] = \frac{1}{2^k} + (-1)\left(p - \frac{1}{2^k}\right) + (1-p) = (1-2p) + \frac{1}{2^{k-1}} = \widehat{f}(\emptyset) + \frac{1}{2^{k-1}}. \tag{16}$$

However,

$$\langle f, \mathbb{1}_H \rangle = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \widehat{\mathbb{1}_H}(S) = \widehat{f}(\emptyset) \cdot \left(1 - \frac{1}{2^{k-1}}\right) + \sum_{T \in \mathcal{T} :\, T \neq \emptyset} \widehat{f}(T) \cdot \left(\frac{-\prod_{j \in T} b_j}{2^{k-1}}\right). \tag{17}$$

Putting together Equations (16) and (17), we have

$$1 = -\widehat{f}(\emptyset) + \sum_{T \in \mathcal{T} :\, T \neq \emptyset} \widehat{f}(T) \cdot \left(-\Pi_{j \in T} b_j\right) \leq \sum_{T \in \mathcal{T}} |\widehat{f}(T)|. \tag{18}$$

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

This completes the proof of part (ii).

Remark 4.5. We note that in fact the proof of part (ii), in particular Equation (18), shows that the following bounds hold when H is a subcube.

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

$$\left|\widehat{f}(S)\right| \geq \frac{1 - |\widehat{f}(\emptyset)|}{2^{\mathrm{C}_{\min}(f)} - 1} \geq \frac{\mathrm{Var}(f)}{2^{\mathrm{C}_{\min}(f)+1}}.$$

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

$$\left|\widehat{f}(S)\right| \geq \frac{1 + |\widehat{f}(\emptyset)|}{2^{\max\{\mathrm{C}^0_{\min}(f), \mathrm{C}^1_{\min}(f)\}} - 1} \geq \frac{\sqrt{\mathrm{Var}(f)}}{2^{\max\{\mathrm{C}^0_{\min}(f), \mathrm{C}^1_{\min}(f)\}}}.$$

This inequality follows from using either f or $-f$ in Equation (18), depending on whether $\widehat{f}(\emptyset)$ is positive or not.

The size bound on S follows, because when H is a subcube the Fourier expansion of $\mathbb{1}_H$ is supported on monomials of degree at most co-dim(H). We end the remark here and continue with the proof of the theorem.

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

$$\frac{1}{3} \leq \mathbb{E}_x \mathbb{E}_{T\sim\mu}[f(x)T(x)] = \mathbb{E}_{T\sim\mu} \mathbb{E}_x[f(x)T(x)] = \mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} \widehat{f}(S)\widehat{T}(S)\right]. \tag{19}$$

However, one can upper bound the last expression in Equation (19) as follows:

$$\mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} \widehat{f}(S)\widehat{T}(S)\right] \leq \mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} |\widehat{f}(S)||\widehat{T}(S)|\right] \leq \left(\max_{S\subseteq[n]} |\widehat{f}(S)|\right) \mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} |\widehat{T}(S)|\right]. \tag{20}$$

Putting together Equations (19) and (20), we have

$$\frac{1}{3} \leq \left(\max_{S\subseteq[n]} |\widehat{f}(S)|\right) \mathbb{E}_{T\sim\mu}\left[\sum_{S\subseteq[n]} |\widehat{T}(S)|\right] \leq \left(\max_{S\subseteq[n]} |\widehat{f}(S)|\right) 2^{R^{\oplus}_2(f)}.$$

The second inequality follows from the fact that each T is a deterministic parity-decision tree of depth at most $R^{\oplus}_2(f)$, hence it easily follows that the spectral norm of the Fourier coefficients of T can be upper bounded by $2^{R^{\oplus}_2(f)}$ (for a proof of this, see Reference [14]). Rewriting the last inequality, we have $\max_{S\subseteq[n]} |\widehat{f}(S)| \geq \frac{1}{2^{R^{\oplus}_2(f)+\log 3}}$, which gives us the third part of the theorem. □

Using a well-known fact that upper bounds Rényi entropy of order $1+\delta$ (for every $\delta > 0$) by a constant times the min-entropy of $\{\hat{f}(S)^2\}$, we deduce the following corollary. Since this fact works for all $\delta > 0$, it is tempting to say that we can improve the bounds in Theorem 4.2 from $\mathbb{H}_\infty(\hat{f}^2)$ to $\mathbb{H}(\hat{f}^2)$, but this relation between the Rényi entropies breaks down for $\delta = 0$.

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

(i) For every $\varepsilon \geq 0$, $\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\frac{1}{\delta}) \cdot \log(\|\widehat{f}\|_{1,\varepsilon}/(1-\varepsilon))$,
(ii) $\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\frac{1}{\delta}) \cdot \mathrm{C}^{\oplus}_{\min}(f)$,
(iii) $\mathbb{H}_{1+\delta}(\hat{f}^2) \leq 2(1+\log 3)(1+\frac{1}{\delta}) \cdot R^{\oplus}_2(f)$.

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

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 [58] 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 $\mathrm{C}_{\min}$ on the min-entropy of $\{\hat{f}(S)^2\}_S$.

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

THEOREM 4.7 ([28, 36]). *There exists a universal constant $c > 0$ such that for every $f : \{-1,1\}^n \to \{-1,1\}$, we have*

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

The next lemma establishes a lower bound of *minimum* certificate size on the total influence of constant-read DNF. A similar argument appears in Shalev [58] too.

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

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

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

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

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 to satisfy that term. Therefore, we have

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

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

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

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 $f : \{-1,1\}^n \to \{-1,1\}$ be a read-k DNF. Then,*

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

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

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

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

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

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

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

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

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

This completes the proof of the theorem. □

5 IMPLICATIONS OF THE FEI CONJECTURE

The FEI and FMEI conjecture seem to impose a strong constraint on the Fourier spectrum of a Boolean function. For example, the FMEI conjecture (if true) would show the existence of a large Fourier coefficient in the spectrum of every Boolean function that has small average sensitivity. In the Introduction, we saw that the FEI conjecture implies the existence of a Fourier-sparse polynomial p that approximates a Boolean function f in ℓ_2-distance, i.e., $\mathbb{E}_x[(f(x)-p(x))^2]$ is small. (A particular case being Mansour's conjecture, which is also a consequence of the FEI conjecture.) In this section, we discuss one implication of the FEI conjecture relating to the structure of polynomials that approximate Boolean functions in the ℓ_∞-distance, i.e., $|p(x)-f(x)|$ is small for every $x \in \{-1,1\}^n$. In particular, we consider the question "Do polynomials approximating a Boolean function in ℓ_∞-distance satisfy some property?" In this direction, we make progress by showing that if the FEI conjecture is true, then we can rule out polynomials with "large" Fourier sparsity from representing or approximating Boolean functions. In this section, we also consider a class of polynomials and show that no polynomial in that class can 1/3-approximate a Boolean function (without assuming that the FEI conjecture is true).

Definition 5.1. An n-variate multilinear polynomial p is said to be a *flat polynomial* if all its *non-zero* coefficients have the same magnitude.

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 d and sparsity $T = 2^{\omega(d)}$. Then p cannot ε-approximate any Boolean function.*

Proof. To prove the lemma, we crucially use the following claim.

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

$$\mathbb{H}(\hat{f}^2) \geq \Omega(\log T), \tag{21}$$

where the constant in $\Omega(\cdot)$ depends on ε.

We assume this claim and conclude the proof of the lemma. By contradiction, let us assume that p ε-approximates a Boolean function f, so $\deg_\varepsilon(f) \leq d$. Assuming the FEI conjecture is true, we

have $\mathbb{H}(\hat{f}^2) = O(\mathsf{Inf}(f))$. Furthermore, using a result of Shi [63], we have $\mathsf{Inf}(f) \leq O(\deg_\varepsilon(f))$ for every f and constant ε. So, we have $\mathbb{H}(\hat{f}^2) \leq O(\deg_\varepsilon(f)) \leq O(d)$. Using Claim 5.1, it follows that

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

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

We now prove the claim.

Proof of Claim 5.1. Suppose p ε-approximates a Boolean function f and that all non-zero coefficients of p have magnitude $\frac{\alpha}{\sqrt{T}}$ for some $\alpha \in [(1-\varepsilon), (1+\varepsilon)]$. Such an α exists, because, by Parseval's identity (Fact 2.1), we have

$$\sum_S \widehat{p}(S)^2 = \mathbb{E}_x\left[p(x)^2\right] \in \left[(1-\varepsilon)^2, (1+\varepsilon)^2\right],$$

where the inclusion assumes p ε-approximates f. Now consider the following set $\mathcal{A}$ of "large" Fourier coefficients,

$$\mathcal{A} := \left\{S \;:\; |\widehat{f}(S)| \geq \frac{2\alpha}{\sqrt{T}}\right\}. \tag{22}$$

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

$$\sum_{S\in\mathcal{A}} \widehat{f}(S)^2 \leq 4\sum_{S\in\mathcal{A}}\left(\widehat{f}(S) - \widehat{p}(S)\right)^2 \leq 4\sum_{S\subseteq[n]}\left(\widehat{f}(S) - \widehat{p}(S)\right)^2 = 4\,\mathbb{E}_x\left[(f(x)-p(x))^2\right] \leq 4\varepsilon^2, \tag{23}$$

where the equality uses Parseval's identity (Fact 2.1) and the last inequality uses that p ε-approximates f. From Equation (23), we have $\sum_{S\notin\mathcal{A}} \widehat{f}(S)^2 \geq 1 - 4\varepsilon^2$. This gives us our desired lower bound on the Fourier entropy of f,

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

The second inequality follows by the definition of $\mathcal{A}$ (Equation (22)), the third inequality by negation of the inequalities in Equation (23), the last inequality holds, because $\alpha \leq (1+\varepsilon)$ by the definition of α, and the last equality assumes $\varepsilon \in (0, 1/2)$. □

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 ε-approximate a Boolean function.*

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

Partial progress toward resolving Conjecture 5.3. As a first step toward proving the conjecture, we observe that a *weaker* version of the conjecture holds when further restricting the approximating polynomials to be *homogeneous*. A polynomial is said to be *homogeneous* if every monomial in the polynomial has the same degree. More formally, we observe the following.

Proposition 5.4. *No flat homogeneous polynomial of degree d and sparsity $2^{\omega(d\log d)}$ can $1/3$-approximate a Boolean function.*

Proof. Let p be a flat homogeneous polynomial of degree d that $1/3$-approximates a Boolean function f. Further, suppose the sparsity of p is T. This implies $|\widehat{p}(S)| = \frac{\alpha}{\sqrt{T}}$ for some $\alpha \in [2/3, 4/3]$. Then, we have

$$\sum_{S:\ |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \leq \left(\sum_{S:\ |S|=d} |\widehat{p}(S) - \widehat{f}(S)|^2 \right)^{1/2} \cdot \sqrt{T} \leq \frac{1}{3}\sqrt{T}, \tag{24}$$

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

$$\begin{aligned} \sum_{S:\ |S|=d} |\widehat{p}(S) - \widehat{f}(S)| \geq \sum_{S:\ |S|=d} \left(|\widehat{p}(S)| - |\widehat{f}(S)| \right) &= \sum_{S:\ |S|=d} |\widehat{p}(S)| - \sum_{S:\ |S|=d} |\widehat{f}(S)| \\ &\geq \frac{2}{3}\sqrt{T} - \sum_{S:\ |S|=d} |\widehat{f}(S)|, \end{aligned} \tag{25}$$

where the first inequality uses the reverse-triangle inequality and the last uses the lower bound on $|\widehat{p}(S)|$. Combining Equations (24) and (25), we obtain

$$\sum_{S:\ |S|=d} |\widehat{f}(S)| \geq \frac{1}{3}\sqrt{T}. \tag{26}$$

However, from results of Tal (see Reference [64, Claim 2.13] and Reference [65, Lemmas 29 and 34]), we know

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

Thus, from Equation (26), we obtain

$$T \leq 2^{O(d\log d)}. \qquad \square$$

Conjecture 5.3 asks if a similar result also holds when $T = 2^{\omega(d)}$. We now consider a restricted class of polynomials called *block-multilinear polynomials*. An n-variate polynomial is said to be *block-multilinear* if the input variables can be *partitioned* into disjoint blocks $A_1, \ldots, A_d \subseteq \{x_1, \ldots, x_n\}$ such that every monomial in the polynomial has *at most* one variable from each block. Without loss of generality, we will assume that each block is of the same size. In other words, a block-multilinear polynomial $p\colon (\mathbb{R}^n)^d \to \mathbb{R}$ can be written as

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

where $\widehat{p}(S) \in \mathbb{R}$ for every $S \subseteq [nd]$. Note that this is the standard Fourier decomposition of p if $x^i \in \{-1, 1\}^n$ for every $i \in [d]$. Clearly, such a block-multilinear polynomial has degree at most d and sparsity at most $(n+1)^d$. Such polynomials have found applications in quantum computing [1, 48], classical and quantum XOR games [15], polynomial decoupling [54], and in functional analysis, which we discuss later. Our main contribution in this section is that we show a positive answer to Conjecture 5.3 for the class of flat block-multilinear polynomials.

Theorem 5.5 (Restatement of Theorem 1.12). *If p is an n-variate flat block-multilinear polynomial with degree d and sparsity $2^{\omega(d)}$, then p cannot $1/3$-approximate a Boolean function.*

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

Relation between Theorem 5.5 and the Bohnenblust-Hille (BH) inequality. Consider a homogeneous block-multilinear polynomial $p\colon (\mathbb{R}^n)^d \to \mathbb{R}$. One way to show that p is not $1/3$-close to a Boolean function would be to show that there exists x' such that $|p(x')|$ is greater than $4/3$. Understanding if such an x' exists for such a polynomial p can be cast as the following maximization problem

$$\|p\| := \max_{x^1,\ldots,x^d \in [-1,1]^n} \left| \sum_{i_1,\ldots,i_d=1}^{n} \widehat{p}_{i_1,\ldots,i_d} x^1_{i_1} \cdots x^d_{i_d} \right|. \tag{28}$$

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 Reference [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 Equation (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 Equation (28) was by Bohnenblust and Hille [11] 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 homogeneous block-multilinear polynomial $p\colon (\mathbb{R}^n)^d \to \mathbb{R}$, we have

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

The result of Reference [11] showed that it suffices to pick $C_d = d^{\frac{d+1}{2d}} 2^{\frac{d-1}{2}}$ to satisfy Equation (29). In addition, their bound on C_d recovers the well-known Littlewood's $4/3$ inequality [43] 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 Equation (29) holds. We cite a few results [3, 8, 25, 26, 48, 56], 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 [56] giving numerical evidence that this is the case).

After a series of works, Pellegrino and Seoane-Sepúlveda showed [55] 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. [8] as $C_d = O(d^{0.365})$. We also point the interested reader to Reference [48, Theorem 17] for a suboptimal, yet elegant proof that shows that it suffices to pick $C_d = O(d^{1.45})$ to satisfy Equation (29).

We now discuss the relevance of the BH-inequality to a special case of Theorem 5.5 where the approximating polynomial is also assumed to be homogeneous. Consider an arbitrary flat homogeneous block-multilinear polynomial $p\colon (\{-1, 1\}^n)^d \to \mathbb{R}$ with degree d and sparsity T that $1/3$-approximates a Boolean function. That is, every non-zero Fourier coefficient $|\widehat{p}_{i_1,\ldots,i_d}|$ equals $\alpha/\sqrt{T}$ for some $\alpha \in [2/3, 4/3]$. Then using Equation (29), we get

$$\left(\left(\frac{2/3}{\sqrt{T}} \right)^{\frac{2d}{d+1}} \cdot T \right)^{\frac{d+1}{2d}} \leq C_d \cdot \|p\| \leq (4/3) \cdot C_d.$$

With further simplification, we have

$$T \leq 2^{2d} \cdot C_d^{2d}.$$

Using the result of Bayart et al. [8], $C_d = O(d^{0.365})$, 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 even conclude the special case of Theorem 5.5. However, if the long-standing open question of C_d being a universal constant were true, then this special case follows. But proving C_d a universal constant seems to be a very hard problem. Nevertheless, in the next section, we will prove Theorem 5.5, while circumventing the barrier of improving the upper bound on C_d.

We remark that Theorem 5.5 does not say anything about the BH-inequality, since Theorem 5.5 states that a flat block-multilinear polynomial either takes a value in the range $(-2/3, 2/3)$ or takes a value of magnitude more that $4/3$, while the BH-inequality states that such a polynomial definitely takes a value of high magnitude ($> 4/3$) on at least one input.

5.1 Proof of Theorem 5.5

We restate the theorem for convenience.

Theorem 5.5 (Restated). *If p is an n-variate flat block-multilinear polynomial with degree d and sparsity $2^{\omega(d)}$, then p cannot $1/3$-approximate a Boolean function.*

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

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

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

Lemma 5.6. *Let p be a degree-d block-multilinear polynomial that $1/3$-approximates a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$. Further, let $q_0, \ldots, q_k$ be as defined in Equation (30). Then,*

(I) For every $z \in \{-1,1\}^{n-k}$, we have $q_i(z) \in [-4/3, -2/3] \cup [-1/3, 1/3] \cup [2/3, 4/3]$, i.e., q_i is a $\frac{1}{3}$-approximation to a $\{-1, 0, 1\}$-valued function.

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

$$|q_j(z)| \geq 2/3 \quad \text{and} \quad \sum_{i \neq j} |q_i(z)| \leq 1/3.$$

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

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

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

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

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

Now the first part follows from the fact that p is a 1/3-approximation to the Boolean function f. So the q_is are 1/3-approximation to a $\{-1, 0, 1\}$-valued function. □

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

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

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

(a) $\mathcal{I}(z) \subseteq [2/3, 4/3]$, or
(b) $\mathcal{I}(z) \subseteq [-4/3, -2/3]$, or
(c) $\mathcal{I}(z)$ intersects both the intervals $[-4/3, -2/3]$ and $[2/3, 4/3]$.

We now prove that in all the three cases, there exists $j \in \{0, \ldots, k\}$ such that $|q_j(z)| \geq 2/3$ and $\sum_{i \neq j} |q_i(z)| \leq 1/3$, 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 $\mathcal{I}(z) \subseteq [2/3, 4/3]$. In this case, observe that $q_0(z) - \sum_{i=1}^{k} |q_i(z)| \geq 2/3$ and $q_0(z) + \sum_{i=1}^{k} |q_i(z)| \leq 4/3$. Clearly, this implies $q_0(z) \geq 2/3$. Furthermore, from both the inequalities it follows that

$$\sum_{i=1}^{k} |q_i(z)| \leq \min\left\{q_0(z) - 2/3, 4/3 - q_0(z)\right\}.$$

Using the fact that $q_0(z) \in [2/3, 4/3]$, it follows that $\sum_{i=1}^{k} |q_i(z)| \leq 1/3$, which concludes the proof for Case (a).

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

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

It is not hard to see that $a = -b$. Consider a path from a to b on the Boolean hypercube. Since $p(a, z) \in [-4/3, -2/3]$ and $p(b, z) \in [2/3, 4/3]$, there exists an edge[11] on this path such that the value of p on this edge jumps from the interval $[-4/3, -2/3]$ to $[2/3, 4/3]$. 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 $|q_j(z)| \geq 2/3$ and $\sum_{i \neq j} |q_i(z)| \leq 1/3$.

The first claim $|q_j(z)| \geq 2/3$ follows immediately, because the change in p on this edge is $2 \cdot |q_j(z)|$ (by Equation (32)), and the change of value of p on this edge is $\geq 2 \cdot (2/3)$. Thus, we have $|q_j(z)| \geq 2/3$.

We now prove $\sum_{i \neq j} |q_i(z)| \leq 1/3$ by considering two cases based on whether $\mathrm{sign}(q_0(z))$ is positive or negative. Since the proofs for both cases are similar, for simplicity, we prove it assuming $\mathrm{sign}(q_0(z))$ is positive. We now define the edge between a, b where the value of p jumps from the interval $[-4/3, -2/3]$ to $[2/3, 4/3]$. Consider the edge in the jth direction given by setting all the variables except x_j as follows: $x_i = \mathrm{sign}(q_i(z))$ for $i \in [k] \setminus \{j\}$. Clearly, the value of p only depends

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

on x_j and equals $\sum_{i\neq j}|q_i(z)|+x_jq_j(z)$. 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 $[2/3,4/3]$ and

$$\sum_{i\neq j}|q_i(z)|+|q_j(z)|\in[2/3,4/3],\ \text{implies}\ \sum_{i\neq j}|q_i(z)|\leq 4/3-|q_j(z)|. \tag{33}$$

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 $[-4/3,-2/3]$, and

$$\sum_{i\neq j}|q_i(z)|-|q_j(z)|\in[-4/3,-2/3],\ \text{implies}\ \sum_{i\neq j}|q_i(z)|\leq |q_j(z)|-2/3. \tag{34}$$

From Equations (33) and (34), we have $\sum_{i\neq j}|q_i(z)|\leq 1/3$. The uniqueness of j in each case follows, because $\sum_{i=0}^{k}|q_i(z)|\leq 4/3$. □

We now show that if $p(x)$ $\frac{1}{3}$-approximates a Boolean function f then $\deg(f)\leq d$.

Lemma 5.7. *Let p be a degree-d block-multilinear polynomial that $\frac{1}{3}$-approximates a Boolean function $f\colon\{-1,1\}^n\to\{-1,1\}$. Then,* $\deg(f)\leq d$.

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

Base case. We have $d=1$. Let $p(x)=a_0+\sum_{i=1}^{n}a_ix_i$, 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 5.6 (II), it follows that there exists a unique $i\in\{0\}\cup[n]$ such that

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

Thus, the base case follows.

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

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

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

Recall from the proof of Lemma 5.6 (I), that $q_i(z)$ can be expressed as a difference between two $(d-1)$-block-multilinear polynomials by fixing $x_1,\ldots,x_k$. That is,

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

where the -1 in the expression $p(1,\ldots,1,-1,1,\ldots,1,z)$ is at the ith coordinate. Clearly, it follows that $p(-1,\ldots,-1,z)$ and $p(1,\ldots,1,z)$ are block-multilinear polynomials with $d-1$ parts that $\frac{1}{3}$-approximate Boolean functions $f(-1,\ldots,-1,z)$ and $f(1,\ldots,1,z)$, respectively. Therefore, by the inductive assumption, both $\deg(f(-1,\ldots,-1,z))$ and $\deg(f(1,\ldots,1,z))$ are $\leq d-1$. Additionally, the function $\widetilde{q}_0$ defined as

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

is a $\{-1,0,1\}$-valued function satisfying $\deg(\widetilde{q}_0) \leq d-1$. Moreover, q_0 (as defined in Equation (35)) is a $\frac{1}{3}$-approximation to $\widetilde{q}_0$. In a similar fashion, one can define $\widetilde{q}_i$ for all $i \in [k]$, which is a $\{-1,0,1\}$-valued function of degree at most $d-1$ that additionally satisfies that q_i (as defined in Equation (35)) is a $\frac{1}{3}$-approximation to $\widetilde{q}_i$. Finally consider the polynomial $\widetilde{p}$ defined as follows:

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

First, note that $\widetilde{p}(x)$ is a polynomial that takes value in $\{-1,+1\}$ for all $x \in \{-1,+1\}^n$. This is because for all i, $q_i(z)$ $\frac{1}{3}$-approximates $\widetilde{q}_i(z)$ and so from Lemma 5.6, for any given z there exists exactly one i such that $\widetilde{q}_i(z)$ takes a non-zero value. So, for a given x if j is the unique index from Lemma 5.6 such that $|q_j(z)| \geq 2/3$, then for that x, we have $\widetilde{p}(x) = x_j \widetilde{q}_j(z)$.

From this observation, we also get $|f(x) - \widetilde{p}(x)| \leq 2/3$ for any $x \in \{-1,1\}^n$. This is because for a given x if j is the unique index from Lemma 5.6 such that $4/3 \geq |q_j(z)| \geq 2/3$, then $f(x) = \mathrm{sign}(x_j q_j(z))$ and $|f(x) - x_j q_j(z)| \leq 1/3$ (because of Lemma 5.6 (II)), $|x_j q_j(z) - x_j \widetilde{q}_j(z)| \leq 1/3$ (by definition of $\widetilde{q}_j$) and $\widetilde{p}(x) = x_j \widetilde{q}_j(z)$. Since both f and $\widetilde{p}$ are $\{-1,+1\}$-valued functions and $|f(x) - \widetilde{p}(x)| \leq 2/3$, it follows that $\widetilde{p}(x) = f(x)$.

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

Using Claim 5.1 and Lemma 5.7, it now follows that if p is a degree-d flat block-multilinear polynomial that 1/3-approximates a Boolean function, then the sparsity of p is at most $2^{O(d)}$. This completes the proof of Theorem 5.5.

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 that we believe are structurally interesting and could lead toward proving the FEI or FMEI conjecture. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function.

(1) Does there exist a Fourier coefficient $S \subseteq [n]$ such that $|\widehat{f}(S)| \geq 2^{-O(\deg_{1/3}(f))}$? 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. [9] showed to be at least $\deg_{1/3}(f)/2$)?
(3) Does there exist a universal constant $\lambda > 0$ such that $\mathbb{H}(\hat{f}^2) \leq \lambda \cdot \min\{\mathrm{C}^1(f), \mathrm{C}^0(f)\}$? 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 Question (2) above. However, in a very recent work [18], it is shown that $\log(\|\widehat{f}\|_{1,\varepsilon})$ 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 N.S. and S.C. visited CWI, Amsterdam. S.A. 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. S.A. thanks Ashley Montanaro for his hospitality. N.S. and S.C. thank Satya Lokam for many helpful discussions on the Fourier entropy-Influence conjecture. S.A. and S.C. 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 Reference [8]. Finally, we thank the anonymous reviewers for many helpful comments that greatly improved the presentation of the article. An extended abstract of this manuscript appeared in the conference proceedings of STACS 2020 [7].

REFERENCES

[1] S. Aaronson and A. Ambainis. 2018. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM J. Comput.* 47, 3 (2018), 982–1038.

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

[3] N. Albuquerque, F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. 2014. Sharp generalizations of the multilinear Bohnenblust-Hille inequality. *J. Funct. Anal.* 266, 6 (2014), 3276–3740.

[4] N. Alon and J. Spencer. 2000. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization.

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

[6] A. Ambainis, M. Kokainis, and R. Kothari. 2016. Nearly optimal separations between communication (or query) complexity and partitions. In *Proceedings of the 31st Conference on Computational Complexity (CCC'16)*. 4:1–4:14.

[7] S. Arunachalam, S. Chakraborty, M. Koucký, N. Saurabh, and R. de Wolf. 2020. Improved bounds on fourier entropy and min-entropy. In *Proceedings of the 37th International Symposium on Theoretical Aspects of Computer Science (STACS'20) (LIPIcs, Vol. 154)*. 45:1–45:19.

[8] F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. 2014. The bohr radius of the n-dimensional polydisk is equivalent to $\sqrt{(\log n)/n}$. *Adv. Math.* 264 (2014), 726–746.

[9] 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.

[10] S. Ben-David, P. Hatami, and A. Tal. 2017. Low-sensitivity functions from unambiguous certificates. In *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS'17)*. 28:1–28:23.

[11] H. F. Bohnenblust and E. Hille. 1931. On the absolute convergence of dirichlet series. *Ann. Math.* 32, 3 (July 1931), 600–622.

[12] R. Boppana. 1997. The average sensitivity of bounded-depth circuits. *Inform. Process. Lett.* 63, 5 (1997), 257–261.

[13] J. Bourgain and G. Kalai. 1997. Influences of variables and threshold intervals under group symmetries. *Geom. Funct. Anal.* 7, 3 (1997), 438–461.

[14] 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. *IEEE Trans. Comput.* 39, 2 (1990), 282–287.

[15] J. Briët, H. Buhrman, T. Lee, and T. Vidick. 2013. Multipartite entanglement in XOR games. *Quant. Info. Comput.* 13, 3-4 (2013), 334–360. Retrieved from https://arXiv:0911.4007.

[16] H. Buhrman and R. de Wolf. 2002. Complexity measures and decision tree complexity: A survey. *Theor. Comput. Sci.* 288, 1 (2002), 21–43.

[17] M. Bun and J. Thaler. 2013. Dual lower bounds for approximate degree and markov-bernstein inequalities. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP'13)*. 303–314.

[18] S. Chakraborty, A. Chattopadhyay, N. S. Mande, and M. Paraashar. 2020. Quantum query-to-communication simulation needs a logarithmic overhead. In *Proceedings of the 35th Computational Complexity Conference (CCC'20) (LIPIcs, Vol. 169)*. 32:1–32:15.

[19] S. Chakraborty, S. Karmalkar, S. Kundu, S. V. Lokam, and N. Saurabh. 2018. Fourier entropy-influence conjecture for random linear threshold functions. In *Proceedings of the 13th Latin American Theoretical Informatics Symposium (LATIN'18)*. 275–289.

[20] S. Chakraborty, R. Kulkarni, S.V. Lokam, and N. Saurabh. 2016. Upper bounds on fourier entropy. *Theor. Comput. Sci.* 654 (2016), 92–112.
[21] M. Cheraghchi, E. Grigorescu, B. Juba, K. Wimmer, and N. Xie. 2018. $AC^0 \circ MOD2$ lower bounds for the Boolean inner product. *J. Comput. System Sci.* 97 (2018), 45–59.
[22] G. Cohen and I. Shinkar. 2016. The complexity of DNF of parities. In *Proceedings of the 7th Conference on Innovations in Theoretical Computer Science (ITCS'16)*. ACM, 47–58.
[23] T. M. Cover and J. A. Thomas. 1991. *Elements of Information Theory*. John Wiley & Sons.
[24] B. Das, M. Pal, and V. Visavaliya. 2011. The Entropy Influence Conjecture Revisited. Retrieved from https://arxiv:1110.4301.
[25] A. Defant, L. Frerick, J. Ortega-Cerdá, M. Ounaïes, and K. Seip. 2011. The bohnenblust-hille inequality for homogeneous polynomials is hypercontractive. *Ann. Math.* 174, 1 (2011), 485–497.
[26] A. Defant, D. Popa, and U. Schwarting. 2010. Coordinatewise multiple summing operators in banach spaces. *J. Funct. Anal.* 259, 1 (2010), 220–242.
[27] R. Eldan and R. Gross. 2020. Concentration on the Boolean hypercube via pathwise stochastic analysis. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC'20)*. 208–221.
[28] E. Friedgut and G. Kalai. 1996. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* 124, 10 (1996), 2993–3002.
[29] M. Göös. 2015. Lower bounds for clique vs. independent set. In *Proceedings of the IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS'15)*. 1066–1076.
[30] P. Gopalan, A. Kalai, and A. R. Klivans. 2008. A query algorithm for agnostically learning DNF? In *Proceedings of the 21st Annual Conference on Learning Theory (COLT'08)*. 515–516.
[31] P. Gopalan, A. T. Kalai, and A. Klivans. 2008. Agnostically learning decision trees. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC'08)*. 527–536.
[32] P. Gopalan, R. Meka, and O. Reingold. 2013. DNF sparsification and a faster deterministic counting algorithm. *Comput. Complex.* 22, 2 (June 2013), 275–310.
[33] P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson. 2016. Degree and sensitivity: Tails of two distributions. In *Proceedings of the 31st Conference on Computational Complexity (CCC'16)*. 13:1–13:23.
[34] L. Gross. 1975. Logarithmic sobolev inequalities. *Amer. J. Math.* 97, 4 (1975), 1061–1083.
[35] R. Hod. 2017. Improved Lower Bounds for the Fourier Entropy/Influence Conjecture via Lexicographic Functions. Retrieved from https://arxiv:1711.00762.
[36] J. Kahn, G. Kalai, and Nathan Linial. 1988. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (FOCS'88)*. 68–80.
[37] G. Kalai. 2007. The Entropy/Influence Conjecture. Retrieved from https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/.
[38] N. Keller, E. Mossel, and T. Schlank. 2012. A note on the entropy/influence conjecture. *Discrete Math.* 312, 22 (2012), 3364 – 3372.
[39] E. Kelman, G. Kindler, N. Lifshitz, D. Minzer, and M. Safra. 2020. Towards a proof of the fourier-entropy conjecture?. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS'20)*. IEEE, 247–258.
[40] 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 Theory (COLT'10)*. 368–380.
[41] T. Lee and A. Shraibman. 2009. Lower bounds in communication complexity. *Found. Trends Theor. Comput. Sci.* 3, 4 (2009), 263–398.
[42] N. Linial, Y. Mansour, and N. Nisan. 1993. Constant depth circuits, fourier transform, and learnability. *J. ACM* 40, 3 (July 1993), 607–620.
[43] J. E. Littlewood. 1930. On bounded bilinear forms in an infinite number of variables. *Quart. J. Math.* 1 (1930), 164–174.
[44] S. Lovett, K. Wu, and J. Zhang. 2020. Decision list compression by mild random restrictions. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC'20)*. ACM, 247–254.
[45] S. Lovett and J. Zhang. 2019. DNF sparsification beyond sunflowers. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC'19)*. ACM, 454–460.
[46] Y. Mansour. 1994. Learning Boolean functions via the fourier transform. In *Theoretical Advances in Neural Computation and Learning*, V. Roychowdhury, K.-Y. Siu, and A. Orlitsky (Eds.). Springer U.S., 391–424.
[47] 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.
[48] A. Montanaro. 2012. Some applications of hypercontractive inequalities in quantum information theory. *J. Math. Phys.* 53, 12 (2012), 122206.
[49] A. Montanaro and T. Osborne. 2009. On the communication complexity of XOR functions. Retrieved from https://arXiv:0909.3392.

[50] R. O'Donnell. 2014. *Analysis of Boolean Functions*. Cambridge University Press.
[51] R. O'Donnell and L.-Y. Tan. 2013. A composition theorem for the fourier entropy-influence conjecture. In *Proceedings of the 40th International Colloquium on Automata, Languages and Programming (ICALP'13)*. 780–791.
[52] R. O'Donnell, J. Wright, Y. Zhao, X. Sun, and L.-Y. Tan. 2014. A composition theorem for parity kill number. In *Proceedings of the 29th Conference on Computational Complexity (CCC'14)*. 144–154.
[53] R. O'Donnell, J. Wright, and Y. Zhou. 2011. The fourier entropy-influence conjecture for certain classes of Boolean functions. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP'11)*. 330–341.
[54] R. O'Donnell and Y. Zhao. 2016. Polynomial bounds for decoupling, with applications. In *Proceedings of the 31st Conference on Computational Complexity (CCC'16)*. 24:1–24:18.
[55] D. Pellegrino and J. Seoane-Sepúlveda. 2012. New upper bounds for the constants in the bohnenblust-hille inequality. *J. Math. Anal. Appl.* 386, 1 (2012), 300–307.
[56] D. Pellegrino and E. V. Teixeira. 2018. Towards sharp Bohnenblust-Hille constants. *Commun. Contemp. Math.* 20, 3 (2018), 1750029.
[57] R. A. Servedio and E. Viola. 2012. On a special case of rigidity. Retrieved from http://eccc.hpi-web.de/report/2012/144.
[58] G. Shalev. 2018. On the Fourier Entropy Influence Conjecture for Extremal Classes. Retrieved from https://arxiv:1806.03646.
[59] R. Shaltiel and E. Viola. 2010. Hardness amplification proofs require majority. *SIAM J. Comput.* 39, 7 (2010), 3122–3154.
[60] C. E. Shannon. 1948. A mathematical theory of communication. *Bell Syst. Techn. J.* 27, 3 (1948), 379–423.
[61] A. Sherstov. 2018. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC'18)*. 311–324.
[62] A. A. Sherstov. 2011. The pattern matrix method. *SIAM J. Comput.* 40, 6 (2011), 1969–2000.
[63] Y. Shi. 2000. Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Inform. Process. Lett.* 75, 1–2 (2000), 79–83.
[64] A. Tal. 2014. Shrinkage of de morgan formulae by spectral techniques. In *Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science (FOCS'14)*. 551–560.
[65] A. Tal. 2017. Tight bounds on the fourier spectrum of AC^0. In *Proceedings of the 32nd Computational Complexity Conference (CCC'17)*. 15:1–15:31.
[66] P. Traxler. 2009. Variable influences in conjunctive normal forms. In *Proceedings of 12th International Conference on Theory and Applications of Satisfiability Testing (SAT'09)*. Springer, Berlin, 101–113.
[67] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. 2013. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS'13)*. 658–667.
[68] A. Wan, J. Wright, and C. Wu. 2014. Decision trees, protocols and the entropy-influence conjecture. In *Proceedings of the 5th Innovations in Theoretical Computer Science (ITCS'14)*. 67–80.
[69] R. de Wolf. 2008. A brief introduction to fourier analysis on the Boolean cube. *Theory Comput. Library Grad. Surveys* 1 (2008), 1–20.
[70] S. Zhang. 2014. Efficient quantum protocols for XOR functions. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*. 1878–1885.
[71] Z. Zhang and Y. Shi. 2009. Communication complexities of symmetric XOR functions. *Quant. Info. Comput.* 9, 3 (2009), 255–263.

Received August 2020; revised February 2021; accepted April 2021