

Improved bounds on Fourier entropy and Min-entropy

Srinivasan Arunachalam

MIT, USA

arunacha@mit.edu

Sourav Chakraborty

Indian Statistical Institute, Kolkata, India

sourav@isical.ac.in

Michal Koucký

Computer Science Institute of Charles University, Prague, Czech Republic

koucky@iuuk.mff.cuni.cz

Nitin Saurabh

Max Planck Institut für Informatik, Saarland Informatics Campus, Saarbrücken, Germany

nsaurabh@mpi-inf.mpg.de

Ronald de Wolf

QuSoft, CWI and University of Amsterdam, the Netherlands

rdewolf@cwi.nl

Abstract

Given a Boolean function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, define the *Fourier distribution* to be the distribution on subsets of $[n]$, where each $S \subseteq [n]$ is sampled with probability $\hat{f}(S)^2$. The Fourier Entropy-Influence (FEI) conjecture of Friedgut and Kalai [24] seeks to relate two fundamental measures associated with the Fourier distribution: does there exist a universal constant $C > 0$ such that $\mathbb{H}(\hat{f}^2) \leq C \cdot \text{Inf}(f)$, where $\mathbb{H}(\hat{f}^2)$ is the Shannon entropy of the Fourier distribution of f and $\text{Inf}(f)$ is the total influence of f ?

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

- (i) Our first contribution shows that $\mathbb{H}(\hat{f}^2) \leq 2 \cdot \text{aUC}^\oplus(f)$, where $\text{aUC}^\oplus(f)$ is the average unambiguous parity-certificate complexity of f . This improves upon several bounds shown by Chakraborty et al. [16]. We further improve this bound for unambiguous DNFs.
- (ii) We next consider the weaker *Fourier Min-entropy-Influence* (FMEI) conjecture posed by O'Donnell and others [43, 40] which asks if $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \text{Inf}(f)$, where $\mathbb{H}_\infty(\hat{f}^2)$ is the min-entropy of the Fourier distribution. We show $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \cdot \text{C}_{\min}^\oplus(f)$, where $\text{C}_{\min}^\oplus(f)$ is the minimum parity certificate complexity of f . We also show that for all $\varepsilon \geq 0$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq 2 \log(\|\hat{f}\|_{1,\varepsilon}/(1-\varepsilon))$, where $\|\hat{f}\|_{1,\varepsilon}$ 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).
- (iii) Our third contribution is to better understand implications of the FEI conjecture for the structure of polynomials that $1/3$ -approximate a Boolean function on the Boolean cube. We pose a conjecture: no *flat polynomial* (whose non-zero Fourier coefficients have the same magnitude) of degree d and sparsity $2^{\omega(d)}$ can $1/3$ -approximate a Boolean function. This conjecture is known to be true assuming FEI and we prove the conjecture unconditionally (i.e., without assuming the FEI conjecture) for a class of polynomials. We discuss an intriguing connection between our conjecture and the constant for the Bohnenblust-Hille inequality, which has been extensively studied in functional analysis.

2012 ACM Subject Classification Computing methodologies \rightarrow Representation of Boolean functions; Theory of computation \rightarrow Models of computation; Mathematics of computing \rightarrow Information theory

Keywords and phrases Fourier analysis of Boolean functions, FEI conjecture, query complexity, polynomial approximation, approximate degree, certificate complexity.

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23



© Srinivasan Arunachalam, Sourav Chakraborty, Michal Koucký, Nitin Saurabh, and Ronald de Wolf; licensed under Creative Commons License CC-BY

42nd Conference on Very Important Topics (CVIT 2016).

Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23–23:18



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

23: Improved bounds on Fourier entropy and Min-entropy

48 **Related Version** A full version [5] of the paper is available at <https://arxiv.org/abs/1809.09819>.

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

52 *Michal Koucký*: Partially supported by ERC Consolidator Grant 616787 LBCAD, and GAČR grant
53 19-27871X

54 *Nitin Saurabh*: Part of the work was done when the author was at IUUK, Prague and supported by
55 the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no.
56 616787

57 *Ronald de Wolf*: Partially supported by ERC Consolidator Grant 615307 QPROGRESS (ended Feb
58 2019) and by NWO under QuantERA project QuantAlgo 680-91-034 and the Quantum Software
59 Consortium.

60 **Acknowledgements** Part of this work was carried out when NS and SC visited CWI, Amsterdam and
61 SA was part of MIT (supported by MIT-IBM Watson AI Lab under the project *Machine Learning*
62 *in Hilbert Space*) and University of Bristol (partially supported by EPSRC grant EP/L021005/1).
63 SA thanks Ashley Montanaro for his hospitality. NS and SC would like to thank Satya Lokam
64 for many helpful discussions on the Fourier entropy-Influence conjecture. SA and SC thank Jop
65 Briët for pointing us to the literature on unbalancing lights and many useful discussions regarding
66 Section 3.3. We also thank Penghui Yao and Avishay Tal for discussions during the course of this
67 project, and Fernando Vieira Costa Júnior for pointing us to the reference [6]. Finally, we thank the
68 anonymous reviewers for many helpful comments.

1 Introduction

Boolean functions $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ naturally arise in many areas of theoretical computer science and mathematics such as learning theory, complexity theory, quantum computing, inapproximability, graph theory, extremal combinatorics, etc. Fourier analysis over the Boolean cube $\{-1, 1\}^n$ is a powerful technique that has been used often to analyze problems in these areas. For a survey on the subject, see [40, 54]. 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 [24]. The FEI conjecture seeks to relate the following two fundamental properties of a Boolean function f : the *Fourier entropy* of f and the *total influence* of f , which we define now.

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

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

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

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

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

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

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

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

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

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

The original motivation of Friedgut and Kalai for the FEI conjecture came from studying threshold phenomena of monotone graph properties in random graphs [24]. 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 $c > 0$).

23:2 Improved bounds on Fourier entropy and Min-entropy

108 The current best upper bound, proven by Bourgain and Kalai [11], is $c_\varepsilon(\log n)^{-2+\varepsilon}$ for
 109 every $\varepsilon > 0$.

110 Besides this application, the FEI conjecture is known to imply the famous Kahn-Kalai-
 111 Linial theorem [30] (otherwise referred to as the KKL theorem). The KKL theorem was one
 112 of the first major applications of Fourier analysis to understanding properties of Boolean
 113 functions and has since found many application in various areas of theoretical computer
 114 science.

115 ► **Theorem 1.2 (KKL theorem).** *For every $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, there exists an $i \in [n]$
 116 such that $\text{Inf}_i(f) \geq \text{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right)$.*

117 See Section 2 for the definitions of these quantities. Another motivation to study the FEI
 118 conjecture is that a positive answer to this conjecture would resolve the notoriously hard
 119 conjecture of Mansour [37] from 1995.

120 ▷ **Conjecture 1.3 (Mansour’s conjecture).** Suppose $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is computed by
 121 a t -term DNF.¹ Then for every $\varepsilon > 0$, there exists a family \mathcal{T} of subsets of $[n]$ such that
 122 $|\mathcal{T}| \leq t^{O(1/\varepsilon)}$ (i.e., size of \mathcal{T} is polynomial in t) and $\sum_{T \in \mathcal{T}} \hat{f}(T)^2 \geq 1 - \varepsilon$.

123 A positive answer to Mansour’s conjecture, along with the query algorithm of Gopalan et
 124 al. [26], would resolve a long-standing open question in computational learning theory of
 125 agnostically learning DNFs under the uniform distribution in polynomial time (up to any
 126 constant accuracy).

127 More generally, the FEI conjecture implies that every Boolean function can be approxi-
 128 mated (in ℓ_2 -norm) by *sparse* polynomials over $\{-1, 1\}$. In particular, for a Boolean function
 129 f and $\varepsilon > 0$, the FEI conjecture implies the existence of a polynomial p with $2^{O(\text{Inf}(f)/\varepsilon)}$
 130 monomials such that $\mathbb{E}_x[(f(x) - p(x))^2] \leq \varepsilon$. The current best known bound in this direction
 131 is $2^{O(\text{Inf}(f)^2/\varepsilon^2)}$, proven by Friedgut [23].²

132 Given the inherent difficulty in answering the FEI conjecture for arbitrary Boolean
 133 functions, there have been many recent works studying the conjecture for specific classes of
 134 Boolean functions. We give a brief overview of these results in the next section. Alongside the
 135 pursuit of resolving the FEI conjecture, O’Donnell and others [43, 40] have asked if a weaker
 136 question than the FEI conjecture, the *Fourier Min-entropy-Influence* (FMEI) conjecture can
 137 be resolved. The FMEI conjecture asks if the entropy-influence inequality in Eq. (1) holds
 138 when the entropy of the Fourier distribution is replaced by the *min-entropy* of the Fourier
 139 distribution (denoted $\mathbb{H}_\infty(\hat{f}^2)$). The min-entropy of $\{\hat{f}(S)^2\}_S$ is defined as

$$140 \quad \mathbb{H}_\infty(\hat{f}^2) := \min_{\substack{S \subseteq [n]: \\ \hat{f}(S) \neq 0}} \left\{ \log \frac{1}{\hat{f}(S)^2} \right\}$$

141 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
 142 compared to $\mathbb{H}(\hat{f}^2)$. For instance, consider the function $f(x) := x_1 \vee \text{IP}(x_1, \dots, x_n)$; then
 143 $\mathbb{H}_\infty(\hat{f}^2) = O(1)$ whereas $\mathbb{H}(\hat{f}^2) = \Omega(n)$. (IP is the inner-product-mod-2 function.) So the
 144 FMEI conjecture could be strictly weaker than the FEI conjecture, making it a natural
 145 candidate to resolve first.

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

² Friedgut’s Junta theorem says that f is ε -close to a junta on $2^{O(\text{Inf}(f)/\varepsilon)}$ variables. We refer to [40, Section 9.6, page 269, Friedgut’s Junta Theorem] for details.

146 ▷ Conjecture 1.4 (FMEI Conjecture). There exists a universal constant $C > 0$ such that for
 147 every Boolean function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, we have $\mathbb{H}_\infty(\hat{f}^2) \leq C \cdot \text{Inf}(f)$.

148 Another way to formulate the FMEI conjecture is, suppose $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, then
 149 does there exist a Fourier coefficient $\hat{f}(S)$ such that $|\hat{f}(S)| \geq 2^{-O(\text{Inf}(f))}$? By the *granularity*
 150 of Fourier coefficients it is well-known that *every* Fourier coefficient of a Boolean function f
 151 is an integral multiple of $2^{-\text{deg}(f)}$, see [40, Exercise 1.11] for a proof of this. (Here the $\text{deg}(f)$
 152 refers to the degree of the unique multilinear polynomial that represents f .) The FMEI
 153 conjecture asks if we can prove a lower bound of $2^{-O(\text{Inf}(f))}$ on *any one* Fourier coefficient, and
 154 even this remains open. Proving the FMEI conjecture seems to require proving interesting
 155 structural properties of Boolean functions. In fact, as observed by [43], the FMEI conjecture
 156 suffices to imply the KKL theorem.

157 Understanding the min-entropy of a Fourier distribution is important in its own right too.
 158 It was observed by Akavia et al. [2] that for a circuit class \mathcal{C} , tighter relations between min-
 159 entropy of $f \in \mathcal{C}$ and f_A defined as $f_A(x) := f(Ax)$, for an arbitrary linear transformation
 160 A , could enable us to translate lower bounds against the class \mathcal{C} to the class $\mathcal{C} \circ \text{MOD}_2$.
 161 In particular, they conjectured that min-entropy of f_A is only polynomially larger than
 162 f when $f \in \text{AC}^0[\text{poly}(n), O(1)]$. ($\text{AC}^0[s, d]$ is the class of unbounded fan-in circuits of size
 163 at most s and depth at most d .) It is well-known that when $f \in \text{AC}^0[s, d]$, $\mathbb{H}_\infty(\hat{f}^2)$ is at
 164 most $O((\log s)^{d-1} \cdot \log \log s)$ [35, 10, 51]. Depending on the tightness of the relationship
 165 between $\mathbb{H}_\infty(\hat{f}^2)$ and $\mathbb{H}_\infty(\hat{f}_A^2)$, one could obtain near-optimal lower bound on the size of
 166 $\text{AC}^0[s, d] \circ \text{MOD}_2$ circuits computing IP (inner-product-mod-2). This problem has garnered
 167 a lot of attention in recent times for a variety of reasons [48, 46, 2, 18, 17]. The current
 168 best known lower bound for IP against $\text{AC}^0[s, d] \circ \text{MOD}_2$ is quadratic when $d = 4$, and only
 169 super-linear for all $d = O(1)$ [17].

170 **Organization** We end this introduction with an overview of prior work on the FEI and
 171 FMEI conjecture in Section 1.1. We then describe our contributions and sketch the proofs in
 172 Section 3. We conclude in Section 4. Due to lack of space the proofs have been omitted. We
 173 refer to the full version [5] for any omission from this version.

174 1.1 Prior work

175 After Friedgut and Kalai [24] posed the FEI conjecture in 1996, there was not much work
 176 done towards resolving it, until the work of Klivans et al. [33] in 2010. They showed that
 177 the FEI conjecture holds true for random DNF formulas. Since then, there have been many
 178 significant steps taken in the direction of resolving the FEI conjecture. We review some
 179 recent works here, referring the interested reader to the blog post of Kalai [31] for additional
 180 discussions on the FEI conjecture.

181 The FEI conjecture is known to be true when we replace the universal constant C
 182 with $\log n$ in Eq. (1). In fact we know $\mathbb{H}(\hat{f}^2) \leq O(\text{Inf}(f) \cdot \log n)$ for *real-valued* functions
 183 $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ (see [43, 32] for a proof and [16] for an improvement of this statement).³
 184 If we strictly require C to be a universal constant, then the FEI conjecture is known to be
 185 false for real-valued functions. Instead, for real-valued functions an analogous statement
 186 called the *logarithmic Sobolev Inequality* [28] is known to be true. The logarithmic Sobolev
 187 inequality states that for every $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, we have $\text{Ent}(f^2) \leq 2 \cdot \text{Inf}(f)$, where $\text{Ent}(f)$

³ For Boolean functions, the $\log n$ -factor was improved by [27] to $\log(\mathfrak{s}(f))$ (where $\mathfrak{s}(f)$ is the sensitivity of the Boolean function f).

23:4 Improved bounds on Fourier entropy and Min-entropy

188 is defined as $\text{Ent}(f) = \mathbb{E}[f \ln(f)] - \mathbb{E}[f] \ln(\mathbb{E}[f])$, where the expectation is taken over uniform
189 $x \in \{-1, 1\}^n$.

190 Restricting to Boolean functions, the FEI conjecture is known to be true for the “standard”
191 functions that arise often in analysis, such as AND, OR, Majority, Parity, Bent functions and
192 Tribes. There have been many works on proving the FEI conjecture for specific classes of
193 Boolean functions. O’Donnell et al. [43] showed that the FEI conjecture holds for symmetric
194 Boolean functions and read-once decision trees. Keller et al. [32] studied a generalization of
195 the FEI conjecture when the Fourier coefficients are defined on biased product measures on the
196 Boolean cube. Then, Chakraborty et al. [16] and O’Donnell and Tan [41], independently and
197 simultaneously, proved the FEI conjecture for read-once formulas. In fact, O’Donnell and Tan
198 proved an interesting composition theorem for the FEI conjecture (we omit the definition of
199 composition theorem here, see [41] for more). For general Boolean functions, Chakraborty et
200 al. [16] gave several upper bounds on the Fourier entropy in terms of combinatorial quantities
201 larger than the total influence, e.g., average decision tree depth, etc., and sometimes even
202 quantities that could be much smaller than influence, namely, average parity-decision tree
203 depth.

204 Later Wan et al. [53] used Shannon’s source coding theorem [49] (which characterizes
205 entropy) to establish the FEI conjecture for read- k decision trees for constant k . Using their
206 novel interpretation of the FEI conjecture they also reproved O’Donnell-Tan’s composition
207 theorem in an elegant way. Recently, Shalev [47] improved the constant in the FEI inequality
208 for read- k decision trees, and further verified the conjecture when either the influence is
209 *too low*, or the entropy is *too high*. The FEI conjecture is also verified for random Boolean
210 functions by Das et al. [20] and for random linear threshold functions (LTFs) by Chakraborty
211 et al. [15].

212 There has also been some work in giving lower bounds on the constant C in the FEI
213 conjecture. Hod [29] gave a lower bound of $C > 6.45$ (the lower bound holds even when
214 considering the class of monotone functions), improving upon the lower bound of O’Donnell
215 and Tan [41].

216 However, there has not been much work on the FMEI conjecture. It was observed
217 in [43, 15] that the KKL theorem implies the FMEI conjecture for monotone functions and
218 linear threshold functions. Finally, the FMEI conjecture for “regular” read- k DNFs was
219 recently established by Shalev [47].

220 **2 Preliminaries**

221 **Notation.** We denote the set $\{1, 2, \dots, n\}$ by $[n]$. A *partial assignment* of $[n]$ is a map
222 $\tau : [n] \rightarrow \{-1, 1, *\}$. Define $|\tau| = |\tau^{-1}(1) \cup \tau^{-1}(-1)|$. A subcube of the Boolean cube
223 $\{-1, 1\}^n$ is a set of $x \in \{-1, 1\}^n$ that agrees with some partial assignment τ , i.e., $\{x \in$
224 $\{-1, 1\}^n : x_i = \tau(i) \text{ for every } i \text{ with } \tau(i) \neq *\}$.

225 **Fourier Analysis.** We recall some definitions and basic facts from analysis of Boolean
226 functions, referring to [40, 54] for more. Consider the space of all functions from $\{-1, 1\}^n$ to
227 \mathbb{R} equipped with the inner product defined as

$$228 \quad \langle f, g \rangle := \mathbb{E}_x[f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \{-1, 1\}^n} f(x)g(x).$$

229 For $S \subseteq [n]$, the character function $\chi_S : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is defined as $\chi_S(x) := \prod_{i \in S} x_i$.
230 Then the set of character functions $\{\chi_S\}_{S \subseteq [n]}$ forms an orthonormal basis for the space of

231 all real-valued functions on $\{-1, 1\}^n$. Hence, every real-valued function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$
 232 has a unique Fourier expansion

$$233 \quad f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x).$$

234 The *degree* of f , denoted $\deg(f)$, is defined as $\max\{|S| : \widehat{f}(S) \neq 0\}$. The *spectral norm* of
 235 f is defined to be $\sum_S |\widehat{f}(S)|$. The *Fourier weight* of a function f on a set of coefficients
 236 $\mathcal{S} \subseteq 2^{[n]}$ is defined as $\sum_{S \in \mathcal{S}} \widehat{f}(S)^2$. The *approximate spectral norm* of a Boolean function f
 237 is defined as

$$238 \quad \|\widehat{f}\|_{1,\varepsilon} = \min \left\{ \sum_S |\widehat{p}(S)| : |p(x) - f(x)| \leq \varepsilon \text{ for every } x \in \{-1, 1\}^n \right\}.$$

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

240 \triangleright **Fact 2.1** (Plancherel's Theorem). For any $f, g : \{-1, 1\}^n \rightarrow \mathbb{R}$,

$$241 \quad \mathbb{E}_x[f(x)g(x)] = \sum_S \widehat{f}(S)\widehat{g}(S).$$

242 In particular, if $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is Boolean-valued and $g = f$, we have Parseval's
 243 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
 244 view $\{\widehat{f}(S)^2\}_S$ as a probability distribution, which allows us to discuss the Fourier entropy
 245 and min-entropy of the distribution $\{\widehat{f}(S)^2\}_S$, defined as

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

$$248 \quad \mathbb{H}(\widehat{f}^2) := \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}, \quad \text{and} \quad \mathbb{H}_\infty(\widehat{f}^2) := \min_{\substack{S \subseteq [n]: \\ \widehat{f}(S) \neq 0}} \left\{ \log \frac{1}{\widehat{f}(S)^2} \right\}.$$

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

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

$$252 \quad \mathbb{H}_\alpha(\widehat{f}^2) := \frac{1}{1-\alpha} \log \left(\sum_{S \subseteq [n]} |\widehat{f}(S)|^{2\alpha} \right).$$

253 It is known that in the limit as $\alpha \rightarrow 1$, $\mathbb{H}_\alpha(\widehat{f}^2)$ is the (Shannon) Fourier entropy $\mathbb{H}(\widehat{f}^2)$
 254 (see [19, Chapter 17, Section 8]) and when $\alpha \rightarrow \infty$, observe that $\mathbb{H}_\alpha(\widehat{f}^2)$ converges to $\mathbb{H}_\infty(\widehat{f}^2)$.
 255 It is easily seen that $\mathbb{H}_\infty(\widehat{f}^2) \leq \mathbb{H}(\widehat{f}^2) \leq \mathbb{H}_{\frac{1}{2}}(\widehat{f}^2) \leq \mathbb{H}_0(\widehat{f}^2)$.

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

$$258 \quad \text{Inf}_i(f) = \Pr_{x \in \{-1, 1\}^n} [f(x) \neq f(x^{(i)})] = \mathbb{E}_x \left[\left(\frac{f(x) - f(x^{(i)})}{2} \right)^2 \right],$$

259 where the probability and expectation is taken according to the uniform distribution on
 260 $\{-1, 1\}^n$ and $x^{(i)}$ is x with the i -th bit flipped. The *total influence* of f , denoted $\text{Inf}(f)$, is

$$261 \quad \text{Inf}(f) = \sum_{i \in [n]} \text{Inf}_i(f).$$

23:6 Improved bounds on Fourier entropy and Min-entropy

262 In terms of the Fourier coefficients of f , it can be shown, e.g., [30], that $\text{Inf}_i(f) = \sum_{S \ni i} \widehat{f}(S)^2$,
 263 and therefore

$$264 \quad \text{Inf}(f) = \sum_{S \subseteq [n]} |S| \widehat{f}(S)^2.$$

265 The *variance* of a real-valued function f is given by $\text{Var}(f) = \sum_{S \neq \emptyset} \widehat{f}(S)^2$. It easily
 266 follows that $\text{Var}(f) \leq \text{Inf}(f)$. We will also need the following version of the well-known KKL
 267 theorem.

268 ► **Theorem 2.4** (KKL Theorem, [30]). *There exists a universal constant $c > 0$ such that for*
 269 *every $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$, we have*

$$270 \quad \text{Inf}(f) \geq c \cdot \text{Var}(f) \cdot \log \frac{1}{\max_i \text{Inf}_i(f)}.$$

271 We now introduce some basic complexity measures of Boolean functions which we use
 272 often, referring to [13] for more.

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

$$277 \quad \mathbb{E}_x[s_f(x)] = \sum_{i=1}^n \Pr_x[f(x) \neq f(x^{(i)})] = \sum_{i=1}^n \text{Inf}_i(f) = \text{Inf}(f),$$

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

280 **Certificate complexity.** For $x \in \{-1, 1\}^n$, the *certificate complexity* of f at x , denoted
 281 $C(f, x)$, is the minimum number of bits in x that needs to be fixed to ensure that the
 282 value of f is constant. The certificate complexity $C(f)$ of f is $\max_x \{C(f, x)\}$. The mini-
 283 mum certificate complexity of f is $C_{\min}(f) = \min_x \{C(f, x)\}$. The 0-certificate complexity
 284 $C^0(f)$ of f is $\max_{x: f(x)=1} \{C(f, x)\}$. Similarly, the 1-certificate complexity $C^1(f)$ of f is
 285 $\max_{x: f(x)=-1} \{C^1(f, x)\}$. Observe that for every $x \in \{-1, 1\}^n$, $s(f, x) \leq C(f, x)$. This gives
 286 $s(f) \leq C(f)$ and $\text{as}(f) \leq \text{aC}(f)$ where $\text{aC}(f)$ denotes the *average* certificate complexity of f .
 287 As before, the average is taken with respect to the uniform distribution on $\{-1, 1\}^n$.

288 **Parity certificate complexity.** Analogously, we define the *parity certificate complexity*
 289 $C^\oplus(f, x)$ of f at x as the minimum number of parities on the input variables one has
 290 to fix in order to fix the value of f at x , i.e.,

$$291 \quad C^\oplus(f, x) := \min\{\text{co-dim}(H) \mid H \text{ is an affine subspace on which } f \text{ is constant and } x \in H\},$$

292 where $\text{co-dim}(H)$ is the *co-dimension* of the affine subspace H . It is easily seen that
 293 $C^\oplus(f, x) \leq C(f, x)$. We also define $C^\oplus(f) := \max_x \{C^\oplus(f, x)\}$, and $C_{\min}^\oplus(f) := \min_x C^\oplus(f, x)$.

294 **Unambiguous certificate complexity.** We now define the *unambiguous certificate complexity*
 295 of f . Let $\tau : [n] \rightarrow \{-1, 1, *\}$ be a partial assignment. We refer to $S_\tau = \{x \in \{-1, 1\}^n :$
 296 $x_i = \tau(i)$ for every $i \in [n] \setminus \tau^{-1}(*)\}$ as the subcube generated by τ . We call $C \subseteq \{-1, 1\}^n$
 297 a *subcube* of $\{-1, 1\}^n$ if there exists a partial assignment τ such that $C = S_\tau$ and the
 298 co-dimension of C is the number of bits fixed by τ , i.e., $\text{co-dim}(C) = |\{i \in [n] : \tau(i) \neq *\}|$. A
 299 set of subcubes $\mathcal{C} = \{C_1, \dots, C_m\}$ *partitions* $\{-1, 1\}^n$ if the subcubes are disjoint and they
 300 cover $\{-1, 1\}^n$, i.e., $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_i C_i = \{-1, 1\}^n$.

301 An *unambiguous certificate* $\mathcal{U} = \{C_1, \dots, C_m\}$ (also referred to as a *subcube partition*)
 302 is a set of subcubes partitioning $\{-1, 1\}^n$. We say \mathcal{U} computes a Boolean function $f :$
 303 $\{-1, 1\}^n \rightarrow \{-1, 1\}$ if f is constant on each C_i (i.e., $f(x)$ is the same for all $x \in C_i$). For
 304 an unambiguous certificate \mathcal{U} , the *unambiguous certificate complexity* on input x (denoted
 305 $\text{UC}(\mathcal{U}, x)$), equals $\text{co-dim}(C_i)$ for the C_i satisfying $x \in C_i$. Define the *average unambiguous*
 306 *certificate complexity* of f with respect to \mathcal{U} as $\text{aUC}(f, \mathcal{U}) := \mathbb{E}_x[\text{UC}(\mathcal{U}, x)]$. Then, the *average*
 307 *unambiguous certificate complexity* of f is defined as

$$308 \quad \text{aUC}(f) := \min_{\mathcal{U}} \text{aUC}(f, \mathcal{U}),$$

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

$$311 \quad \text{UC}(f) := \min_{\mathcal{U}} \max_x \text{UC}(\mathcal{U}, x).$$

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

314 An unambiguous \oplus -certificate $\mathcal{U} = \{C_1, \dots, C_m\}$ for f is defined to be a collection of
 315 monochromatic *affine subspaces* that together partition the space $\{-1, 1\}^n$. It is easily seen
 316 that a subcube is also an affine subspace. Analogously, for an unambiguous \oplus -certificate
 317 \mathcal{U} , on an input x , $\text{UC}^\oplus(\mathcal{U}, x) := \text{co-dim}(C_i)$ for the C_i satisfying $x \in C_i$, and $\text{aUC}^\oplus(f, \mathcal{U}) :=$
 318 $\mathbb{E}_x[\text{UC}^\oplus(\mathcal{U}, x)]$. Similarly, we define $\text{aUC}^\oplus(f)$ and $\text{UC}^\oplus(f)$.

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

322 **Approximate degree.** The ε -*approximate degree* of $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, denoted $\text{deg}_\varepsilon(f)$,
 323 is defined to be the minimum degree among all multilinear real polynomials p such that
 324 $|f(x) - p(x)| \leq \varepsilon$ for all $x \in \{-1, 1\}^n$. Usually ε is chosen to be $1/3$, but it can be chosen to
 325 be any constant in $(0, 1)$, without significantly changing the model.

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

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

23:8 Improved bounds on Fourier entropy and Min-entropy

337 **Randomized decision tree.** A randomized decision tree for f is a probability distribution
 338 R_μ over deterministic decision trees for f . On input x , a decision tree is chosen according
 339 to R_μ , which is then evaluated on x . The complexity of the randomized tree is the largest
 340 depth among all deterministic trees with non-zero probability of being sampled according to
 341 R_μ . One can similarly define a randomized parity-decision tree as a probability distribution
 342 R_μ^\oplus over deterministic parity-decision trees for f .

343 We say that a randomized decision tree computes f with bounded-error if for all $x \in$
 344 $\{-1, 1\}^n$ its output equals $f(x)$ with probability at least $2/3$. $R_2(f)$ (resp. $R_2^\oplus(f)$) denotes
 345 the complexity of the optimal randomized (resp. parity) decision tree that computes f with
 346 bounded-error, i.e., errs with probability at most $1/3$.

347 **Information Theory.** We now recall the following consequence of the law of large numbers,
 348 called the *Asymptotic Equipartition Property (AEP)* or the *Shannon-McMillan-Breiman*
 349 *theorem*. See Chapter 3 in the book [19] for more details.

350 ► **Theorem 2.5** (Asymptotic Equipartition Property (AEP) Theorem). *Let \mathbf{X} be a random*
 351 *variable drawn from a distribution P and suppose $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M$ are independently and*
 352 *identically distributed copies of \mathbf{X} , then*

$$353 \quad -\frac{1}{M} \log P(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M) \longrightarrow \mathbb{H}(\mathbf{X})$$

354 *in probability as $M \rightarrow \infty$.*

355 ► **Definition 2.6.** *Fix $\varepsilon \geq 0$. The typical set $T_\varepsilon^{(M)}(\mathbf{X})$ with respect to a distribution P is*
 356 *defined to be the set of sequences $(x_1, x_2, \dots, x_M) \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_M$ such that*

$$357 \quad 2^{-M(\mathbb{H}(\mathbf{X})+\varepsilon)} \leq P(x_1, x_2, \dots, x_M) \leq 2^{-M(\mathbb{H}(\mathbf{X})-\varepsilon)}.$$

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

359 ► **Theorem 2.7** ([19, Theorem 3.1.2]). *Let $\varepsilon \geq 0$ and $T_\varepsilon^{(M)}(\mathbf{X})$ be a typical set with respect*
 360 *to P , then*

- 361 (i) $|T_\varepsilon^{(M)}(\mathbf{X})| \leq 2^{M(\mathbb{H}(\mathbf{X})+\varepsilon)}$.
- 362 (ii) *Suppose x_1, \dots, x_M are drawn i.i.d. according to \mathbf{X} , then*
 363 $\Pr[(x_1, \dots, x_M) \in T_\varepsilon^{(M)}(\mathbf{X})] \geq 1 - \varepsilon$ *for M sufficiently large.*
- 364 (iii) $|T_\varepsilon^{(M)}(\mathbf{X})| \geq (1 - \varepsilon)2^{M(\mathbb{H}(\mathbf{X})-\varepsilon)}$ *for M sufficiently large.*

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

367 ► **Definition 2.8** ([19, Chapter 11, Section 2]). *Let \mathbf{X} be a random variable drawn according*
 368 *to a distribution P . Fix $\varepsilon > 0$. The strongly typical set $T_\varepsilon^{*(M)}(\mathbf{X})$ is defined to be the set*
 369 *of sequences $\rho = (x_1, x_2, \dots, x_M) \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_M$ such that $N(x; \rho) = 0$ if $P(x) = 0$,*
 370 *and otherwise*

$$371 \quad \left| \frac{N(x; \rho)}{M} - P(x) \right| \leq \frac{\varepsilon}{|\mathbf{X}|},$$

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

373 The strongly typical set shares similar properties with its (weak) typical counterpart which
 374 we state now. See [19, Chapter 11, Section 2] for a proof of this theorem.

375 ▶ **Theorem 2.9** (Strong AEP Theorem). *Following the notation in Definition 2.8, let $T_\varepsilon^{*(M)}(\mathbf{X})$*
 376 *be a strongly typical set. Then, there exists $\delta > 0$ such that $\delta \rightarrow 0$ as $\varepsilon \rightarrow 0$, and the following*
 377 *hold:*

- 378 (i) *Suppose x_1, \dots, x_M are drawn i.i.d. according to \mathbf{X} , then*
 379 $\Pr[(x_1, \dots, x_M) \in T_\varepsilon^{*(M)}(\mathbf{X})] \geq 1 - \varepsilon$ *for M sufficiently large.*
 380 (ii) *If $(x_1, \dots, x_M) \in T_\varepsilon^{*(M)}(\mathbf{X})$, then*

$$381 \quad 2^{-M(\mathbb{H}(\mathbf{X})+\delta)} \leq P(x_1, \dots, x_M) \leq 2^{-M(\mathbb{H}(\mathbf{X})-\delta)}.$$

- 382 (iii) *For M sufficiently large,*

$$383 \quad (1 - \varepsilon)2^{M(\mathbb{H}(\mathbf{X})-\delta)} \leq |T_\varepsilon^{*(M)}(\mathbf{X})| \leq 2^{M(\mathbb{H}(\mathbf{X})+\delta)}.$$

384 **3 Our Contributions**

385 Our contributions in this paper are threefold, which we elaborate on below:

386 **3.1 Better upper bounds for the FEI conjecture**

387 Our first and main contribution of this paper is to give a better upper bound on the Fourier
 388 entropy $\mathbb{H}(\hat{f}^2)$ in terms of $\text{aUC}(f)$, the *average unambiguous certificate complexity* of f .
 389 Informally, the unambiguous certificate complexity $\text{UC}(f)$ of f is similar to the standard
 390 certificate complexity measure, except that the collection of certificates is now required to be
 391 *unambiguous*, i.e., every input should be consistent with a *unique* certificate. In other words,
 392 an unambiguous certificate is a monochromatic subcube partition of the Boolean cube. By
 393 the average unambiguous certificate complexity, $\text{aUC}(f)$, we mean the expected number of
 394 bits set by an unambiguous certificate on a uniformly random input.

395 There have been many recent works on query complexity, giving upper and lower bounds
 396 on $\text{UC}(f)$ in terms of other combinatorial measures such as decision-tree complexity, sensitivity,
 397 quantum query complexity, etc., see [25, 4, 8] for more. It follows from definitions that $\text{UC}(f)$
 398 lower bounds decision tree complexity. However, it is known that $\text{UC}(f)$ can be quadratically
 399 smaller than decision tree complexity [4]. Our main contribution here is an improved upper
 400 bound of *average unambiguous certificate complexity* $\text{aUC}(f)$ on $\mathbb{H}(\hat{f}^2)$. This improves upon
 401 the previously known bound of average decision tree depth on $\mathbb{H}(\hat{f}^2)$ [16].

402 ▶ **Theorem 3.1.** *Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be a Boolean function. Then,*

$$403 \quad \mathbb{H}(\hat{f}^2) \leq 2 \cdot \text{aUC}(f).$$

404 A new and crucial ingredient employed in the proof of the theorem is an analog of the law
 405 of large numbers in information theory, usually referred to as the *Asymptotic Equipartition*
 406 *Property (AEP)* theorem (Theorem 2.5). Employing information-theoretic techniques for the
 407 FEI conjecture seems very natural given that the conjecture seeks to bound the entropy of a
 408 distribution. Indeed, Keller et al. [32, Section 3.1] envisioned a proof of the FEI conjecture
 409 itself using large deviation estimates and the tensor structure (explained below) in a stronger
 410 way, and Wan et al. [53] used Shannon's source coding theorem [49] to verify the conjecture
 411 for bounded-read decision trees.

412 In order to prove Theorem 3.1, we study the *tensorized* version of f , $f^M : \{-1, 1\}^{Mn} \rightarrow$
 413 $\{-1, 1\}$, which is defined as follows,

$$414 \quad f^M(x^1, \dots, x^M) := f(x_1^1, \dots, x_n^1) \cdot f(x_1^2, \dots, x_n^2) \cdots f(x_1^M, \dots, x_n^M).$$

23:10 Improved bounds on Fourier entropy and Min-entropy

415 Similarly we define a *tensorized* version \mathcal{C}^M of an unambiguous certificate \mathcal{C} of f ,⁴ i.e., a direct
 416 product of M independent copies of \mathcal{C} . It is not hard to see that \mathcal{C}^M is also an unambiguous
 417 certificate of f^M . To understand the properties of \mathcal{C}^M we study \mathcal{C} in a probabilistic manner.
 418 We observe that \mathcal{C} naturally inherits a distribution \mathbf{C} on its certificates when the underlying
 419 inputs $x \in \{-1, 1\}^n$ are distributed uniformly. Using the asymptotic equipartition property
 420 with respect to \mathbf{C} , we infer that for every $\delta > 0$, there exists $M_0 > 0$ such that for all
 421 $M \geq M_0$, there are at most $2^{M(\text{aUC}(f, \mathcal{C}) + \delta)}$ certificates in \mathcal{C}^M that together cover at least
 422 $1 - \delta$ fraction of the inputs in $\{-1, 1\}^{Mn}$. Furthermore, each of these certificates fixes at most
 423 $M(\text{aUC}(f, \mathcal{C}) + \delta)$ bits. Hence, a particular certificate can contribute to at most $2^{M(\text{aUC}(f, \mathcal{C}) + \delta)}$
 424 Fourier coefficients of f^M . Combining both these bounds, all these certificates can overall
 425 contribute to at most $2^{2M(\text{aUC}(f, \mathcal{C}) + \delta)}$ Fourier coefficients of f^M . Let's denote this set of
 426 Fourier coefficients by \mathcal{B} . We then argue that the Fourier coefficients of f^M that are *not* in
 427 \mathcal{B} have Fourier weight at most δ . This now allows us to bound the Fourier entropy of f^M as
 428 follows,

$$429 \quad \mathbb{H}(\widehat{f^M}^2) \leq \log |\mathcal{B}| + \delta n M + \mathbb{H}(\delta),$$

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

$$431 \quad \mathbb{H}(\widehat{f^2}) \leq 2(\text{aUC}(f, \mathcal{C}) + \delta) + \delta n + \frac{\mathbb{H}(\delta)}{M}.$$

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

434 Looking finely into how certificates contribute to Fourier coefficients in the proof above,
 435 we further strengthen Theorem 3.1 by showing that we can replace $\text{aUC}(f)$ by the *average*
 436 *unambiguous parity-certificate complexity* $\text{aUC}^\oplus(f)$ of f . Here $\text{aUC}^\oplus(f)$ is defined similar to
 437 $\text{aUC}(f)$ except that instead of being defined in terms of monochromatic subcube partitions
 438 of f , we now partition the Boolean cube with monochromatic *affine subspaces*. (Observe that
 439 subcubes are also affine subspaces.) This strengthening also improves upon the previously
 440 known bound of average parity-decision tree depth on $\mathbb{H}(\widehat{f^2})$ [16]. It is easily seen that
 441 $\text{aUC}^\oplus(f)$ lower bounds the average parity-decision tree depth.

442 ► **Theorem 3.2.** *Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be any Boolean function. Then,*

$$443 \quad \mathbb{H}(\widehat{f^2}) \leq 2 \cdot \text{aUC}^\oplus(f).$$

444 The proof outline remains the same as in Theorem 3.1. However, a particular certificate
 445 in \mathcal{C}^M no longer fixes just variables. Instead these parity certificates now fix parities over
 446 variables, and so potentially could involve all variables. Hence we cannot directly argue
 447 that all the certificates contribute to at most $2^{M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)}$ Fourier coefficients of f^M .
 448 Nevertheless, by the AEP theorem we still obtain that a typical parity-certificate fixes at
 449 most $M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)$ parities. Looking closely at the Fourier coefficients that a parity-
 450 certificate can contribute to, we now argue that such coefficients must lie in the linear span of
 451 the parities fixed by the parity-certificate. Therefore, a typical parity-certificate can overall
 452 contribute to at most $2^{M(\text{aUC}^\oplus(f, \mathcal{C}) + \delta)}$ Fourier coefficients of f^M . The rest of the proof now
 453 follows analogously.

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

454 ▶ Remark 3.3. As a corollary to the theorem we obtain that the FEI conjecture holds for the
 455 class of functions f with constant $\text{aUC}^\oplus(f)$, and $\text{Inf}(f) \geq 1$. That is, for a Boolean function
 456 f with $\text{Inf}(f) \geq 1$, we have

$$457 \quad \mathbb{H}(\hat{f}^2) \leq 2 \cdot \text{aUC}^\oplus(f) \cdot \text{Inf}(f).$$

458 We note that the reduction in [53, Proposition E.2] shows that removing the requirement
 459 $\text{Inf}(f) \geq 1$ from the above inequality will prove the FEI conjecture for all Boolean functions
 460 with $\text{Inf}(f) \geq \log n$. Furthermore, if we could show the FEI conjecture for Boolean functions
 461 f where $\text{aUC}^\oplus(f) = \omega(1)$ is a slow-growing function of n , again the padding argument in
 462 [53] shows that we would be able to establish the FEI conjecture for all Boolean functions.

463 Further extension to unambiguous DNFs

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

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

$$472 \quad \text{Inf}(f) \leq 2 \cdot \min \left\{ \sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}, \sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} \right\} \leq \text{aUC}(f, \mathcal{C}),$$

473 where $\text{co-dim}(\cdot)$ denotes the co-dimension of an affine space. Note that the quantity
 474 $\sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}$, in a certain sense, is “average unambiguous 1-certificate
 475 complexity” and, similarly, $\sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)}$ captures “average unambiguous
 476 0-certificate complexity”.

477 Building on our ideas from the main theorem in the previous section and using a *stronger*
 478 version of the AEP theorem (Theorem 2.9) we essentially establish the aforementioned im-
 479 proved bound of the smaller quantity between “average unambiguous 1-certificate complexity”
 480 and “average unambiguous 0-certificate complexity” on the Fourier entropy. Formally, we
 481 prove the following.

482 ▶ **Theorem 3.4.** *Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be a Boolean function and $\mathcal{C} = \{C_1, \dots, C_t\}$ be a*
 483 *monochromatic affine subspace partition of $\{-1, 1\}^n$ with respect to f such that $\{C_1, \dots, C_{t_1}\}$*
 484 *for some $t_1 < t$ is an affine subspace partition of $f^{-1}(-1)$ and $\{C_{t_1+1}, \dots, C_t\}$ is an affine*
 485 *subspace partition of $f^{-1}(1)$. Further, $p := \Pr_x[f(x) = 1]$. Then,*

$$486 \quad \mathbb{H}(\hat{f}^2) \leq \begin{cases} 2 \left(\sum_{i=1}^{t_1} \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + p \cdot \max_{i \in \{1, \dots, t_1\}} \text{co-dim}(C_i) \right), \\ 2 \left(\sum_{i=t_1+1}^t \text{co-dim}(C_i) \cdot 2^{-\text{co-dim}(C_i)} + (1-p) \cdot \max_{i \in \{t_1+1, \dots, t\}} \text{co-dim}(C_i) \right). \end{cases}$$

487 We remark that to truly claim the bound of “average unambiguous 1-certificate complexity”
 488 one would like to remove the additive term $p \cdot \max_{i \in \{1, \dots, t_1\}} \text{co-dim}(C_i)$ from the stated bound
 489 in the above theorem. This is because when the $\max_i \text{co-dim}(C_i)$ term is *not* weighted by

490 p , it becomes a trivial bound on entropy. Ideally, one would like to get rid of this term
 491 altogether, possibly at the expense of increasing the constant factor in the first summand.

492 We also note that a similar bound for the general DNF representation, i.e., when
 493 $\{C_1, \dots, C_{t_1}\}$ is an arbitrary DNF representation of f where the C_i s need not be disjoint,
 494 suffices to establish Mansour’s conjecture (Conjecture 1.3). In fact, following the analogy,
 495 Theorem 3.4 implies a bound of “average 1-certificate complexity” in the general case. In
 496 this direction, we observe that a weaker bound of 1-certificate complexity, i.e., showing
 497 $\mathbb{H}(\hat{f}^2) \leq O(C_1(f))$, would already suffice to answer Mansour’s conjecture positively. We refer
 498 to the full version [5] for a detailed discussion on this.

499 The outline for the proof of Theorem 3.4 remains the same as before, but it differs in
 500 implementation details. We sketch them now. Analogous to the proof of the main theorem
 501 we consider a partition of inputs with respect to f and its tensorized version. Motivated
 502 by the DNF representation, we study the following partition $\{C_1, \dots, C_{t_1}, f^{-1}(1)\}$ which
 503 naturally inherits a distribution \mathbf{C} given by the uniform distribution on the underlying inputs.
 504 Again we build a “small” set \mathcal{B} of Fourier coefficients of f^M based on the Fourier expansions
 505 of strongly typical sequences. However, unlike before, the probability of observing a strongly
 506 typical sequence doesn’t capture the number of coefficients it could contribute to \mathcal{B} . Here,
 507 we use stronger properties guaranteed by the strong AEP. In particular, it guarantees that
 508 the *empirical* distribution of a typical sequence is close to the distribution of \mathbf{C} . In contrast,
 509 the (weak) AEP only guarantees that the *empirical* entropy of a typical sequence is close
 510 to the entropy of \mathbf{C} . Using the stronger property we can now lower bound the magnitude
 511 of any non-zero Fourier coefficient in the Fourier expansion of the indicator function of a
 512 strongly typical sequence. We then use Parseval’s Identity (Fact 2.1) to deduce an upper
 513 bound on its Fourier sparsity, which in turn is used to bound the size of \mathcal{B} . We also need to
 514 argue that coefficients *not* in \mathcal{B} have negligible Fourier weight, which can be done as before.
 515 Using the two properties, we can now complete the proof.

516 3.2 New upper bounds for the FMEI conjecture

517 Given the hardness of obtaining better upper bounds on the Fourier entropy of a Boolean
 518 function, we make progress on a weaker conjecture, the FMEI conjecture. The FMEI
 519 conjecture is much less studied than the FEI conjecture. In fact, we are aware of only
 520 one recent paper [47] which studies the FMEI conjecture for a particular class of functions.
 521 Our second contribution is to give upper bounds on the min-entropy of general Boolean
 522 functions in terms of the minimum parity-certificate complexity (denoted $C_{\min}^{\oplus}(f)$) and the
 523 approximate spectral norm of Boolean functions (denoted $\|\hat{f}\|_{1,\varepsilon}$). The minimum parity-
 524 certificate complexity $C_{\min}^{\oplus}(f)$ is also referred to as the parity kill number by O’Donnell et
 525 al. [42] and is closely related to the communication complexity of XOR functions [56, 39, 52].
 526 The approximate spectral norm $\|\hat{f}\|_{1,\varepsilon}$ is related to the *quantum* communication complexity
 527 of XOR functions [34, 55]. In particular, it characterizes the bounded-error quantum
 528 communication complexity of XOR functions with constant \mathbb{F}_2 -degree [55]. (By \mathbb{F}_2 -degree,
 529 we mean the degree of a function when viewed as a polynomial over \mathbb{F}_2 .)

530 ► **Theorem 3.5.** *Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be a Boolean function. Then,*

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

532 (2) $\mathbb{H}_{\infty}(\hat{f}^2) \leq 2 \cdot C_{\min}^{\oplus}(f)$.

533 (3) $\mathbb{H}_\infty(\hat{f}^2) \leq 2(1 + \log_2 3) \cdot R_2^\oplus(f)$.⁵

534 The proof of Theorem 3.5 (1) expresses the quantity $\|\hat{f}\|_{1,\varepsilon}$ as a (minimization) linear
 535 program. We consider the dual linear program and exhibit a feasible solution that achieves
 536 an optimum of $(1 - \varepsilon) / \max_S |\hat{f}(S)|$. This proves the desired inequality. In order to prove
 537 part (2) and (3) of the theorem, the idea is to consider a “simple” function g that has “good”
 538 correlation with f , and then upper bound the correlation between f and g using Plancherel’s
 539 theorem (Fact 2.1) and the fact that g has a “simple” Fourier structure. For part (2), g is
 540 chosen to be the indicator function of an (affine) subspace where f is constant, whereas for
 541 part (3) the randomized parity-decision tree computing f itself plays the role of g .⁶

542 As a corollary of this theorem we also obtain upper bounds on the Rényi Fourier entropy
 543 $\mathbb{H}_{1+\delta}(\hat{f}^2)$ of order $1 + \delta$ for all $\delta > 0$. Recall that $\mathbb{H}_{1+\delta}(\hat{f}^2) \geq \mathbb{H}_\infty(\hat{f}^2)$ for every $\delta \geq 0$ and
 544 as $\delta \rightarrow \infty$, $\mathbb{H}_{1+\delta}(\hat{f}^2)$ converges to $\mathbb{H}_\infty(\hat{f}^2)$. Also $\mathbb{H}_1(\hat{f}^2)$ is the standard Shannon entropy of
 545 the Fourier distribution. We refer to the full version [5] for a detailed treatment of it.

546 We believe that these improved bounds on min-entropy of the Fourier distribution give a
 547 better understanding of Fourier coefficients of Boolean functions, and could be of independent
 548 interest. As a somewhat non-trivial application of Theorem 3.5 (in particular, part (2)) we
 549 verify the FMEI conjecture for read- k DNFs, for constant k . (A read- k DNF is a formula
 550 where each variable appears in at most k terms.)

551 ► **Theorem 3.6.** *For every Boolean function $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ that can be expressed*
 552 *as a read- k DNF, we have*

$$553 \quad \mathbb{H}_\infty(\hat{f}^2) \leq O(\log k) \cdot \text{Inf}(f).$$

554 This theorem improves upon a recent (and independent) result of Shalev [47] that
 555 establishes the FMEI conjecture for “regular” read- k DNFs (where regular means each term
 556 in the DNF has more or less the same number of variables, see [47] for a precise definition).
 557 In order to prove Theorem 3.6, we essentially show that $\text{Inf}(f)$ is at least as large as the
 558 minimum certificate complexity $C_{\min}(f)$ of f .

559 ► **Lemma 3.7.** *There exists a universal constant $c > 0$ such that for all $f: \{-1, 1\}^n \rightarrow$
 560 $\{-1, 1\}$ that can be expressed as a read- k DNF, we have*

$$561 \quad \text{Inf}(f) \geq c \cdot \text{Var}(f) \cdot (C_{\min}(f) - 1 - \log k).$$

562 The proof of this lemma is an application of the KKL theorem (Theorem 2.4). Now
 563 the proof of Theorem 3.6 follows with an application of the lemma in conjunction with
 564 Theorem 3.5 (2).

565 3.3 Implications of the FEI conjecture and connections to the 566 Bohnenblust-Hille inequality

567 Our final contribution is to understand better the structure of polynomials that ε -approximate
 568 Boolean functions on the Boolean cube. To be more specific, for simplicity we fix $\varepsilon = 1/3$
 569 and we consider polynomials p such that $|p(x) - f(x)| \leq 1/3$ for all $x \in \{-1, 1\}^n$, where

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 (1), along the lines of parts (2) and (3). 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.

23:14 Improved bounds on Fourier entropy and Min-entropy

570 f is a Boolean function. Such polynomials have proved to be powerful and found diverse
 571 applications in theoretical computer science. The single most important measure associated
 572 with such polynomials is its *degree*. The *least* degree of a polynomial that 1/3-approximates
 573 f is referred to as the *approximate degree* of f . Tight bounds on approximate degree have
 574 both algorithmic and complexity-theoretic implications, see for instance Sherstov’s recent
 575 paper [50] and references therein.

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

583 We first observe that if a flat polynomial p 1/3-approximates a Boolean function f , then
 584 the entropy of the Fourier distribution of f must be “large”. In particular, we show that
 585 $\mathbb{H}(\hat{f}^2)$ must be at least as large as the logarithm of the Fourier sparsity of p .

586 \triangleright **Claim 3.8.** If p is a flat polynomial with sparsity T that 1/3-approximates a Boolean
 587 function f , then

$$588 \quad \mathbb{H}(\hat{f}^2) = \Omega(\log T).$$

590 It then follows that assuming the FEI conjecture, a flat polynomial of degree d and
 591 sparsity $2^{\omega(d)}$ cannot 1/3-approximate a Boolean function. However, it is not clear to us
 592 how to obtain the same conclusion *unconditionally* (i.e., without assuming that the FEI
 593 conjecture is true) and, so we pose the following conjecture.

594 \triangleright **Conjecture 3.9.** No flat polynomial of degree d and sparsity $2^{\omega(d)}$ can 1/3-approximate a
 595 Boolean function.

596 \blacktriangleright **Remark 3.10.** We remark that there exists degree- d flat *Boolean* functions of sparsity 2^d .
 597 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
 598 $(d/2)$ -fold product of this Boolean function on disjoint variables, we obtain our remark.

599 Since we could not solve the problem as posed above, we make progress in understanding
 600 this conjecture by further restricting ourselves to the class of *block-multilinear* polynomials.
 601 An n -variate polynomial is said to be *block-multilinear* if the input variables can be *partitioned*
 602 into disjoint blocks such that every monomial in the polynomial has *at most* one variable
 603 from each block. Such polynomials have been well-studied in functional analysis since the
 604 work of Bohnenblust and Hille [9], but more recently have found applications in quantum
 605 computing [1, 38], classical and quantum XOR games [12], and polynomial decoupling [44].
 606 In the functional analysis literature block-multilinear polynomials are known as *multilinear*
 607 *forms*. In an ingenious work [9], Bohnenblust and Hille showed that for every degree- d
 608 multilinear form $p : (\mathbb{R}^n)^d \rightarrow \mathbb{R}$, we have

$$609 \quad \left(\sum_{i_1, \dots, i_d=1}^n |\hat{p}_{i_1, \dots, i_d}|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \leq C_d \cdot \max_{x^1, \dots, x^d \in [-1, 1]^n} |p(x^1, \dots, x^d)|, \quad (2)$$

611 where C_d is a constant that depends on d . In [9], they showed that it suffices to pick C_d to
 612 be exponential in d to satisfy the equation above. For $d = 2$, Eq. (2) generalizes Littlewood’s
 613 famous 4/3-inequality [36]. Eq. (2) is commonly referred to as the Bohnenblust-Hille (BH)

inequality and is known to have deep applications in various fields of analysis such as operator theory, complex analysis, etc. There has been a long line of work on improving the constant C_d in the BH inequality (to mention a few [22, 21, 3, 6, 45]). The best known upper bound on C_d (we are aware of) is polynomial in d . It is also conjectured that it suffices to let C_d be a *universal* constant (independent of d) in order to satisfy Eq. (2).

In our context, using the best known bound on C_d in the BH-inequality implies that a flat block-multilinear polynomial of degree d and sparsity $2^{\omega(d \log d)}$ cannot $1/3$ -approximate a Boolean function. However, from the discussion before Conjecture 3.9, we know that the FEI conjecture implies the following theorem.

► **Theorem 3.11.** *If p is a flat block-multilinear polynomial of degree d and sparsity $2^{\omega(d)}$, then p cannot $1/8$ -approximate a Boolean function.*

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

► **Lemma 3.12.** *Let p be a block-multilinear polynomial of degree- d that $1/8$ -approximates a Boolean function f . Then, $\deg(f) \leq d$.*

Now using the fact that Fourier entropy of f is at least as large as the logarithm of the sparsity of p (Claim 3.8), we obtain Theorem 3.11.

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

- (1) Does there exist a Fourier coefficient $S \subseteq [n]$ such that $|\widehat{f}(S)| \geq 2^{-O(\deg_{1/3}(f))}$? This would show $\mathbb{H}_\infty(\widehat{f}^2) \leq O(\deg_{1/3}(f))$.
- (2) Can we show $\mathbb{H}(\widehat{f}^2) \leq O(Q(f))$? Or, $\mathbb{H}_\infty(\widehat{f}^2) \leq O(Q(f))$? (where $Q(f)$ is the $1/3$ -error quantum query complexity of f , which Beals et al. [7] showed to be at least $\deg_{1/3}(f)/2$).
- (3) Does there exist a universal constant $\lambda > 0$ such that $\mathbb{H}(\widehat{f}^2) \leq \lambda \cdot \min\{C^1(f), C^0(f)\}$? This would resolve Mansour's conjecture.

In an earlier version of this manuscript we suggested that bounding the logarithm of the approximate spectral norm by $O(\deg_{1/3}(f))$ or $O(Q(f))$ might be an approach to answer Question (1) or (2) above. However, in a very recent work [14] 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.

657 — References

- 658 1 S. Aaronson and A. Ambainis. Forrelation: A problem that optimally separates quantum from
659 classical computing. *SIAM Journal of Computing*, 47(3):982–1038, 2018. Earlier in STOC’15.
660 arXiv:1411.5729.
- 661 2 A. Akavia, A. Bogdanov, S. Guo, A. Kamath, and A. Rosen. Candidate weak pseudorandom
662 functions in $AC^0 \circ MOD2$. In *Proceedings of the 5th Conference on Innovations in Theoretical
663 Computer Science*, ITCS’14, pages 251–260. ACM, 2014.
- 664 3 N. Albuquerque, F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. Sharp generalizations of
665 the multilinear Bohnenblust-Hille inequality. *Journal of Functional Analysis*, 266(6):3276–3740,
666 2014. arXiv:1306.3362.
- 667 4 A. Ambainis, M. Kokainis, and R. Kothari. Nearly optimal separations between communication
668 (or query) complexity and partitions. In *31st Conference on Computational Complexity, CCC
669 2016*, pages 4:1–4:14, 2016. Combines arXiv:1512.01210 and arXiv:1512.00661.
- 670 5 S. Arunachalam, S. Chakraborty, M. Koucký, N. Saurabh, and R. de Wolf. Improved bounds
671 on fourier entropy and min-entropy, 2018. arXiv:1809.09819.
- 672 6 F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda. The Bohr radius of the n -dimensional
673 polydisk is equivalent to $\sqrt{(\log n)/n}$. *Advances in Mathematics*, 264:726–746, 2014.
- 674 7 R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by
675 polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS’98. quant-
676 ph/9802049.
- 677 8 S. Ben-David, P. Hatami, and A. Tal. Low-sensitivity functions from unambiguous certificates.
678 In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017*, pages 28:1–28:23,
679 2017. arXiv:1605.07084.
- 680 9 H. F. Bohnenblust and E. Hille. On the absolute convergence of Dirichlet series. *Annals of
681 Mathematics*, pages 600–622, 1931.
- 682 10 R. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*,
683 63(5):257–261, 1997.
- 684 11 J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group
685 symmetries. *Geometric and Functional Analysis (GAFA)*, 7(3):438–461, 1997.
- 686 12 J. Briët, H. Buhrman, T. Lee, and T. Vidick. Multipartite entanglement in XOR games.
687 *Quantum Information & Computation*, 13(3-4):334–360, 2013. arXiv:0911.4007.
- 688 13 H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey.
689 *Theoretical Computer Science*, 288(1):21–43, 2002.
- 690 14 S. Chakraborty, A. Chattopadhyay, N. Mande, and M. Paraashar. Quantum Query-to-
691 Communication simulation needs a logarithmic overhead, 2019. arXiv:1909.10428.
- 692 15 S. Chakraborty, S. Karmalkar, S. Kundu, S. V. Lokam, and N. Saurabh. Fourier entropy-
693 influence conjecture for random linear threshold functions. In *LATIN 2018: Theoretical
694 Informatics - 13th Latin American Symposium, 2018*, pages 275–289, 2018.
- 695 16 S. Chakraborty, R. Kulkarni, S.V. Lokam, and N. Saurabh. Upper bounds on Fourier entropy.
696 *Theoretical Computer Science*, 654:92–112, 2016. The first version appeared as a technical
697 report TR13-052 on ECCG in 2013.
- 698 17 M. Cheraghchi, E. Grigorescu, B. Juba, K. Wimmer, and N. Xie. $AC^0 \circ MOD2$ lower bounds
699 for the Boolean inner product. *Journal of Computer and System Sciences*, 97:45 – 59, 2018.
- 700 18 G. Cohen and I. Shinkar. The complexity of DNF of parities. In *Proceedings of the 2016 ACM
701 Conference on Innovations in Theoretical Computer Science*, ITCS ’16, pages 47–58. ACM,
702 2016.
- 703 19 T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- 704 20 B. Das, M. Pal, and V. Visavaliya. The entropy influence conjecture revisited, 2011.
705 arxiv:1110.4301.
- 706 21 A. Defant, L. Frerick, J. Ortega-Cerdá, M. Ounaïes, and K. Seip. The Bohnenblust-Hille
707 inequality for homogeneous polynomials is hypercontractive. *Annals of Mathematics*, 174(1):485–
708 497, 2011. arXiv:0904.3540.

- 709 22 A. Defant, D. Popa, and U. Schwarting. Coordinatewise multiple summing operators in banach
710 spaces. *Journal of Functional Analysis*, 259(1):220–242, 2010.
- 711 23 E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates.
712 *Combinatorica*, 18(1):27–35, 1998.
- 713 24 E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proceedings*
714 *of the American Mathematical Society*, 124(10):2993–3002, 1996.
- 715 25 M. Göös. Lower bounds for clique vs. independent set. In *IEEE 56th Annual Symposium on*
716 *Foundations of Computer Science, FOCS 2015*, pages 1066–1076, 2015.
- 717 26 P. Gopalan, A. T. Kalai, and A. Klivans. Agnostically learning decision trees. In *Proceedings*
718 *of the 40th annual ACM symposium on Theory of computing, STOC '08*, pages 527–536, 2008.
- 719 27 P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson. Degree and sensitivity: Tails of two
720 distributions. In *31st Conference on Computational Complexity, CCC 2016*, pages 13:1–13:23,
721 2016. arxiv: 1604.07432.
- 722 28 L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083,
723 1975.
- 724 29 R. Hod. Improved lower bounds for the Fourier entropy/influence conjecture via lexicographic
725 functions, 2017. arxiv:1711.00762.
- 726 30 J. Kahn, G. Kalai, and Nathan Linial. The influence of variables on Boolean functions. In
727 *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages
728 68–80, 1988.
- 729 31 G. Kalai. The entropy/influence conjecture. Terence Tao’s blog: [https://terrytao.
730 wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/](https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/), 2007.
- 731 32 N. Keller, E. Mossel, and T. Schlam. A note on the entropy/influence conjecture. *Discrete*
732 *Mathematics*, 3(2):3364 – 3372, 2012. arXiv:1105.2651.
- 733 33 A. Klivans, H. Lee, and A. Wan. Mansour’s conjecture is true for random DNF formulas. In
734 *Proceedings of the 23rd Conference on Learning Theory*, pages 368–380, 2010.
- 735 34 T. Lee and A. Shraibman. Lower bounds in communication complexity. *Foundations and*
736 *Trends in Theoretical Computer Science*, 3(4):263–398, 2009.
- 737 35 N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform, and
738 learnability. *J. ACM*, 40(3):607–620, July 1993.
- 739 36 J. E. Littlewood. On bounded bilinear forms in an infinite number of variables. *The Quarterly*
740 *Journal of Mathematics*, 1:164–174, 1930.
- 741 37 Y. Mansour. An $n^{O(\log \log n)}$ learning algorithm for DNF under the uniform distribution.
742 *Journal of Computer and System Sciences*, 50(3):543–550, 1995.
- 743 38 A. Montanaro. Some applications of hypercontractive inequalities in quantum information
744 theory. *Journal of Mathematical Physics*, 53(12):122206, 2012. arXiv:1208.0161.
- 745 39 A. Montanaro and T. Osborne. On the communication complexity of XOR functions, 2009.
746 arXiv:0909.3392.
- 747 40 R. O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
- 748 41 R. O’Donnell and L-Y. Tan. A composition theorem for the Fourier entropy-influence conjecture.
749 In *Proceedings of Automata, Languages and Programming - 40th International Colloquium*,
750 pages 780–791, 2013. arXiv:1304.1347.
- 751 42 R. O’Donnell, J. Wright, Y. Zhao, X. Sun, and L-Y. Tan. A composition theorem for parity kill
752 number. In *IEEE 29th Conference on Computational Complexity, CCC 2014*, pages 144–154,
753 2014. arXiv:1312.2143.
- 754 43 R. O’Donnell, J. Wright, and Y. Zhou. The Fourier entropy-influence conjecture for certain
755 classes of Boolean functions. In *Proceedings of Automata, Languages and Programming - 38th*
756 *International Colloquium*, pages 330–341, 2011.
- 757 44 R. O’Donnell and Y. Zhao. Polynomial bounds for decoupling, with applications. In *31st Con-*
758 *ference on Computational Complexity, CCC 2016*, pages 24:1–24:18, 2016. arXiv:1512.01603.
- 759 45 D. Pellegrino and E. V. Teixeira. Towards sharp Bohnenblust-Hille constants. *Communications*
760 *in Contemporary Mathematics*, 20(3):1750029, 2018. arXiv:1604.07595.

- 761 **46** R. A. Servedio and E. Viola. On a special case of rigidity. Manuscript: <http://eccc.hpi-web.de/report/2012/144>, 2012.
- 762
- 763 **47** G. Shalev. On the Fourier Entropy Influence conjecture for extremal classes. arxiv:1806.03646, 2018.
- 764
- 765 **48** R. Shaltiel and E. Viola. Hardness amplification proofs require majority. *SIAM Journal on Computing*, 39(7):3122–3154, 2010.
- 766
- 767 **49** C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- 768
- 769 **50** Alexander A. Sherstov. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 311–324, 2018.
- 770
- 771
- 772 **51** A. Tal. Tight bounds on the Fourier spectrum of AC^0 . In *32nd Computational Complexity Conference, CCC 2017*, pages 15:1–15:31, 2017.
- 773
- 774 **52** H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 658–667, 2013. arXiv:1304.1245.
- 775
- 776
- 777 **53** A. Wan, J. Wright, and C. Wu. Decision trees, protocols and the entropy-influence conjecture. In *Innovations in Theoretical Computer Science, ITCS'14*, pages 67–80, 2014. arXiv:1312.3003.
- 778
- 779 **54** R. de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory of Computing*, 2008. ToC Library, Graduate Surveys 1.
- 780
- 781 **55** S. Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*, pages 1878–1885, 2014. arXiv:1307.6738.
- 782
- 783
- 784 **56** Z. Zhang and Y. Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.
- 785