The Fourier Entropy–Influence Conjecture for certain classes of Boolean functions

Ryan O'Donnell*, John Wright*, and Yuan Zhou*

Department of Computer Science, Carnegie Mellon University

Abstract. In 1996, Friedgut and Kalai made the *Fourier Entropy–Influence Conjecture*: For every Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ it holds that $\boldsymbol{H}[\widehat{f}^2] \leq C \cdot \mathbf{I}[f]$, where $\boldsymbol{H}[\widehat{f}^2]$ is the spectral entropy of f, $\mathbf{I}[f]$ is the total influence of f, and C is a universal constant. In this work we verify the conjecture for symmetric functions. More generally, we verify it for functions with symmetry group $S_{n_1} \times \cdots \times S_{n_d}$ where d is constant. We also verify the conjecture for functions computable by read-once decision trees.

* {odonnell,jswright,yuanzhou}@cs.cmu.edu. This research performed while the first author was a member of the School of Mathematics, Institute for Advanced Study. Supported by NSF grants CCF-0747250 and CCF-0915893, BSF grant 2008477, and Sloan and Okawa fellowships.

1 Introduction

The field of *Fourier analysis of Boolean functions* $f\colon \{-1,1\}^n \to \{-1,1\}$ plays an important role in many areas of mathematics and computer science, including complexity theory, learning theory, random graphs, social choice, inapproximability, arithmetic combinatorics, coding theory, metric spaces, etc. For a survey, see e.g. [17]. One of the most longstanding and important open problems in the field is the *Fourier Entropy–Influence (FEI) Conjecture* made by Friedgut and Kalai in 1996 [6]:

Fourier Entropy–Influence (FEI) Conjecture. $\exists C\ \forall f,\quad \boldsymbol{H}[\widehat{f}^2] \le C \cdot \mathbf{I}[f].$ *That is,*

$$\sum_{S\subseteq[n]} \widehat{f}(S)^2 \log_2 \frac{1}{\widehat{f}(S)^2} \le C \cdot \sum_{S\subseteq[n]} \widehat{f}(S)^2 |S|.$$

The quantity $\boldsymbol{H}[\widehat{f}^2] = \sum \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}$ on the left is the *spectral entropy* or *Fourier entropy* of f. It ranges between 0 and n and measures how "spread out" f's Fourier spectrum is. The quantity $\mathbf{I}[f] = \sum \widehat{f}(S)^2|S|$ appearing on the right is the *total influence* or *average sensitivity* of f. It also ranges between 0 and n and measures how "high up" f's Fourier spectrum is. (For definitions of the terms used in this introduction, see Section 2.)

The FEI Conjecture is superficially similar to the well-known *Logarithmic Sobolev Inequality* [9] for the Boolean cube which states that $\mathbf{Ent}[f^2] \le 2 \cdot \mathbf{I}[f]$ holds for any $f\colon \{-1,1\}^n \to \mathbb{R}$, where $\mathbf{Ent}[g] = \mathbf{E}[g \ln g] - \mathbf{E}[g] \ln \mathbf{E}[g]$. However note that the FEI Conjecture requires $f\colon \{-1,1\}^n \to \{-1,1\}$ to be Boolean-*valued*, and it definitely fails for real-valued f.

1.1 Applications of the conjecture

Friedgut and Kalai's original motivation for the FEI Conjecture came from the theory of random graphs. Suppose f represents a monotone graph property with $\mathbf{Pr}[f(G) = 1] = 1/2$ when $G \sim G(v, 1/2)$ (here $n = \binom{v}{2}$). If we also consider $\mathbf{Pr}[f(G) = 1]$ for $G \sim G(v,p)$, then the total influence $\mathbf{I}[f]$ equals the reciprocal of the derivative of this quantity at $p = 1/2$. Hence the property has "sharp threshold" if and only if $\mathbf{I}[f]$ is large. Friedgut and Kalai sought general conditions on f which would force $\mathbf{I}[f]$ to be large. They conjectured that having significant symmetry — and hence, a spread-out Fourier spectrum — was such a property.

The FEI Conjecture also easily implies the famed KKL Theorem [11]. To see this, first note that $\boldsymbol{H}[\widehat{f}^2] \ge \boldsymbol{H}_\infty[\widehat{f}^2] = \min_S\{\log \frac{1}{\widehat{f}(S)^2}\}$, the *min-entropy* of $\widehat{f}^2$. Thus the FEI Conjecture is strictly stronger than the following:

Fourier Min-Entropy–Influence Conjecture $\exists C\ \forall f\ \ \boldsymbol{H}_\infty[\widehat{f}^2] \le C \cdot \mathbf{I}[f]$. *That is,* $\exists S \subseteq [n]$ *such that* $\widehat{f}(S)^2 \ge 2^{-C\cdot\mathbf{I}[f]}$.

In particular, for balanced f (i.e., $\mathbf{E}[f] = 0 = \widehat{f}(\emptyset)^2$) the above conjecture implies there is a nonempty S with $\widehat{f}(S)^2 \ge 2^{-C\cdot\mathbf{I}[f]}$. Since $\mathbf{Inf}_j[f] \ge \widehat{f}(S)^2$ for each $j \in S$ we conclude $\max\{\mathbf{Inf}_i[f]\} \ge 2^{-C\cdot n\cdot\max\{\mathbf{Inf}_i[f]\}}$ whence $\max\{\mathbf{Inf}_i[f]\} \ge \Omega(\frac{1}{C}) \cdot \frac{\log n}{n}$, which is KKL's conclusion. Indeed, by applying the above deduction just to the *nonempty* Fourier coefficients it is straightforward to deduce $\mathbf{I}[f] \ge \frac{1}{C}\mathbf{Var}[f] \log \frac{1}{\max_i\{\mathbf{Inf}_i[f]\}}$, a strengthening of the KKL Theorem due to Talagrand [20]. We also remark that since $\mathbf{Inf}_i[f] = \widehat{f}(\{i\})$ for monotone f, the KKL Theorem implies the Fourier Min-Entropy–Influence Conjecture holds for monotone functions.

Finally, as emphasized by Klivans and coauthors [8, 14], the FEI Conjecture is also important because it implies a version of *Mansour's Conjecture* from 1994.

Mansour's Conjecture [15]. $\forall \epsilon > 0\ \exists K$ *such that if* $f\colon \{-1,1\}^n \to \{-1,1\}$ *is computable by a DNF formula with* $m \geq 2$ *terms then by taking* $\mathcal{S}$ *to be the* m^K *sets* S *for which* $\widehat{f}(S)^2$ *is largest,* $\sum_{S \notin \mathcal{S}} \widehat{f}(S)^2 \leq \epsilon$.

In fact, Mansour conjectured more strongly that one may take $K = O(\log \frac{1}{\epsilon})$. It is well known [3] that if f is computed by an m-term DNF then $\mathbf{I}[f] \leq O(\log m)$. Thus the Fourier Entropy–Influence Conjecture would imply $\boldsymbol{H}[\widehat{f}^2] \leq C \cdot O(\log m)$, from which it easily follows that one may take $K = O(C/\epsilon)$ in Mansour's Conjecture. Mansour's Conjecture is important because if it is true then the query algorithm of Gopalan, Kalai, and Klivans [7] would agnostically learn DNF formulas under the uniform distribution to any constant accuracy in polynomial time. Establishing such a result is a major open problem in computational learning theory [8]. Further, sufficiently strong versions of Mansour's Conjecture would yield improved pseudorandom generators for DNF formulas; see, e.g., [4, 14] for more on this important open problem in pseudorandomness.

1.2 Prior work

As far as we are aware, the result in [14] showing that the FEI Conjecture holds for random DNFs is the only published progress on the FEI Conjecture since it was posed. In this subsection we collect some observations related to the conjecture, all of which were presumably known to Friedgut and Kalai and should be considered folklore. See also [12] for additional recent discussion of the conjecture.

The FEI Conjecture holds for "the usual examples" that arise in analysis of Boolean functions — Parities (for which the conjecture is trivial), ANDs and ORs, Majority, Tribes [1], and Inner-Product-mod-2. This may be established by direct calculation based on the known Fourier coefficient formulas for these functions (see [21] for Majority and [16] for Tribes). By considering the AND and OR functions it is easy to show that the constant C must be at least 4. We can show that $C = 4$ is necessary and sufficient for the Tribes functions as well; smaller constants suffice for Inner-Product-mod-2 and Majority. The authors are also aware of an explicit family of functions which show the necessity of $C \geq 60/13 \approx 4.615$. For a gross upper bound, it is not too hard to show that $\boldsymbol{H}[\widehat{f}^2] \leq (1 + \log n) \cdot \mathbf{I}[f] + 1$; indeed, this will be shown in the course of the present paper.

The FEI Conjecture "tensorizes" in the following sense: For $f\colon \{-1,1\}^n \to \{-1,1\}$ and $M \in \mathbb{Z}_+$, define $f^{\oplus M} \to \{-1,1\}^{Mn} \to \{-1,1\}$ by $f(x^{(1)}, \ldots, x^{(M)}) = f(x^{(1)})f(x^{(2)}) \cdots f(x^{(M)})$. Then it's easy to check that $\boldsymbol{H}[\widehat{f^{\oplus M}}^2] = M \cdot \boldsymbol{H}[\widehat{f}^2]$ and $\mathbf{I}[f] = M \cdot \mathbf{I}[f]$. This is of course consistent with the FEI Conjecture; it also implies that the following weaker-looking conjecture is actually equivalent to FEI:

$$\text{for all } f\colon \{-1,1\}^n \to \{-1,1\}, \quad \boldsymbol{H}[\widehat{f}^2] \leq C \cdot \mathbf{I}[f] + o(n).$$

To see the equivalence, given $f\colon \{-1,1\}^n \to \{-1,1\}$, apply the above to $f^{\oplus M}$, divide by M, and take the limit as $M \to \infty$.

1.3 Our results and approach

In this work, we prove the FEI Conjecture for some classes of Boolean functions.

Theorem 1. *The FEI Conjecture holds for symmetric functions, with* $C = 12.04$.

Although the class of symmetric functions is fairly small, there was sentiment that it might be a difficult case for FEI: for symmetric functions, $\widehat{f}(S) = \widehat{f}(S')$ whenever $|S| = |S'|$ and hence their Fourier spectrum is maximally spread out on each level.

Our proof of Theorem 1 uses the same high-level idea as in the well-known KKL Theorem [11]: namely, prove a certain inequality for the *discrete derivatives* $D_i f$ of f, and then sum over i. In our case, the key inequality we need for the derivatives is that they are very noise-sensitive:

Theorem. Let g be a discrete derivative of a symmetric function $f\colon \{-1,1\}^{n+1} \to \{-1,1\}$. Then for all real $1 \le c \le n$ it holds that $\mathbf{Stab}_{1-\frac{c}{n}}[g] \le \frac{2/\sqrt{\pi}}{\sqrt{c}}\,\mathbf{E}[g^2]$.

(For the notation used here, see Section 2.) Having established Theorem 1, it is not too hard to generalize it as follows:

Theorem 2. *The FEI Conjecture holds for* d-part-symmetric *functions, with $C = 12.04 + \log_2 d$.*

A d-part-symmetric function is essentially a function with symmetry group of the form $S_{n_1} \times \cdots \times S_{n_d}$. This theorem also generalizes the folklore fact that FEI holds (up to an additive constant) with $C = O(\log n)$, since *every* function is n-part-symmetric.

Finally, with an unrelated, direct inductive argument we show the following:

Theorem 3. *The FEI Conjecture holds for functions computable by read-once decision trees, with $C = 4.88$.*

Remark: In independent and concurrent work, the FEI Conjecture was verified for monotone symmetric functions (a special case of Theorem 1) and decision lists (a special case of Theorem 3) using different methods of proof [22].

2 Definitions and notation

We use the notation $\mathbb{N} = \{0,1,2,\dots\}$, $\mathbb{Z}_+ = \mathbb{N} \setminus \{0\}$, and $[n] = \{1,2,\cdots,n\}$. Throughout we write log for $\log_2$; for the natural logarithm we write ln. The expressions $0\log 0$ and $0\log\frac{1}{0}$ are to be interpreted as 0.

2.1 Basics of Boolean Fourier analysis

This paper is concerned with *Boolean functions* $f\colon \{-1,1\}^n \to \mathbb{R}$, especially *Boolean-valued functions* $f\colon \{-1,1\}^n \to \{-1,1\}$. Every Boolean function f has a unique multilinear polynomial representation over $\mathbb{R}$,

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S) \prod_{i\in S} x_i.$$

This is known as the *Fourier expansion* of f, and the real numbers $\widehat{f}(S)$ are the *Fourier coefficients* of f. We have the formula $\widehat{f}(S) = \mathbf{E}[f(x)\prod_{i\in S} x_i]$. (Here and throughout, expectation $\mathbf{E}[\cdot]$ is with respect to the uniform distribution of x on $\{-1,1\}^n$, unless otherwise specified.) In particular, $\widehat{f}(\emptyset) = \mathbf{E}[f]$. An important basic fact about Fourier coefficients is *Parseval's identity*: $\sum_{S\subseteq[n]} \widehat{f}(S)^2 = \mathbf{E}[f(x)^2]$. A consequence of Parseval is that $\sum_{S\subseteq[n]} \widehat{f}(S)^2 = 1$ for Boolean-valued f. Thus the numbers $\widehat{f}(S)^2$ can be thought of as a probability distribution on the subsets of $[n]$.

Given $f\colon \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$, we define the *discrete derivative* $D_i f\colon \{-1,1\}^n \to \mathbb{R}$ by $D_i f(x) = \frac{f(x^{(i=1)}) - f(x^{(i=-1)})}{2}$, where $x^{(i=b)}$ denotes $(x_1,\dots,x_{i-1},b,x_{i+1},\dots,x_n)$. It holds that

$$\widehat{D_i f}(S) = \begin{cases} 0 & \text{if } i \in S,\\ \widehat{f}(S\cup\{i\}) & \text{if } i \notin S; \end{cases}$$

i.e,. D_i acts on the Fourier expansion as formal differentiation. The *influence of i on f* is

$$\mathbf{Inf}_i[f] = \mathbf{E}[(D_i f)^2] = \sum_{S \ni i} \widehat{f}(S)^2.$$

In the particular case that f is Boolean-valued, the derivative $D_i f$ is $\{-1, 0, 1\}$-valued and we have the combinatorial interpretation $\mathbf{Inf}_i[f] = \mathbf{Pr}[f(x^{(i=1)}) \neq f(x^{(i=-1)})]$. The *total influence* of $f\colon \{-1,1\}^n \to \mathbb{R}$ is

$$\mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] = \sum_{S \subseteq [n]} \widehat{f}(S)^2 |S|.$$

For $0 \leq \rho \leq 1$, we say that $x, y \in \{-1,1\}^n$ are a pair of *ρ-correlated random strings* if x is distributed uniformly randomly on $\{-1,1\}^n$ and y is formed by setting $y_i = x_i$ with probability $\frac{1}{2} + \frac{1}{2}\rho$, $y_i = -x_i$ with probability $\frac{1}{2} - \frac{1}{2}\rho$, independently for each $i \in [n]$. We may now define the *noise stability of f at ρ* and give its Fourier formula:

$$\mathbf{Stab}_\rho[f] = \mathop{\mathbf{E}}_{x,y\ \rho\text{-correlated}}[f(x)f(y)] = \sum_{S \subseteq [n]} \widehat{f}(S)^2 \rho^{|S|}.$$

We often stratify the Fourier coefficients into *levels*; the level of S is simply $|S|$. We define the *weight of f at level k* to be $\mathbf{W}^k[f] = \sum_{|S|=k} \widehat{f}(S)^2$. Thus

$$\mathbf{I}[f] = \sum_{k=0}^{n} \mathbf{W}^k[f] \cdot k, \quad \mathbf{Stab}_\rho[f] = \sum_{k=0}^{n} \mathbf{W}^k[f]\rho^k.$$

Finally, given a random variable or probability distribution we write $\boldsymbol{H}[\cdot]$ for its (binary) Shannon entropy. Hence for $f\colon \{-1,1\}^n \to \{-1,1\}$, $\boldsymbol{H}[\widehat{f}^2] = \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2}$, called the *Fourier entropy*, or *spectral entropy*, of f. Thus the Fourier Entropy–Influence Conjecture may be stated as follows: there is a universal constant C such that $\boldsymbol{H}[\widehat{f}^2] \leq C \cdot \mathbf{I}[f]$ holds for all Boolean-valued functions f.

2.2 Some Boolean function classes

We will call a function $f\colon \{-1,1\}^n \to \mathbb{R}$ *symmetric* if it is invariant under any permutation of the coordinates $[n]$. Equivalently, f is symmetric if the value of $f(x)$ depends only the *Hamming weight* of x, defined to be $\#\{i \in [n] : x_i = -1\}$. In this case we may identify f with the function $f\colon \{0, 1, \ldots, n\} \to \mathbb{R}$ whose value at s equals $f(x)$ for any x of Hamming weight s.

We generalize the notion to that of *d-part-symmetric* functions, $d \in \mathbb{Z}_+$. We say the function $f\colon \{-1,1\}^n \to \mathbb{R}$ is d-part-symmetric if there is a partition $[n] = V_1 \cupdot V_2 \cupdot \cdots \cupdot V_d$ such that f is invariant under any permutation of the coordinates in any part V_i. Equivalently, f is d-part-symmetric if its symmetry group is isomorphic to $S_{n_1} \times \cdots \times S_{n_d}$ for numbers $n_1 + \cdots + n_d = n$. Note that a symmetric function is 1-part-symmetric, and every function $f\colon \{-1,1\}^n \to \mathbb{R}$ is n-part-symmetric.

We also generalize the notion of Fourier "levels" for d-part-symmetric functions. Suppose f is d-part-symmetric with respect to the partition $[n] = V_1 \cupdot \cdots \cupdot V_d$, where $|V_i| = n_i$. Then $\widehat{f}(S)$ depends only on the numbers $|S \cap V_1|$, ..., $|S \cap V_d|$. We consider all possible such sequences

$$\boldsymbol{k} \in \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\} \times \cdots \times \{0, 1, \ldots, n_d\},$$

and say that $S \subseteq [n]$ is at *multi-level* $\boldsymbol{k}$ if $|S \cap V_i| = \boldsymbol{k}_i$ for each $i \in [d]$. We also use the notation

$$|\boldsymbol{k}| = \boldsymbol{k}_1 + \boldsymbol{k}_2 + \cdots + \boldsymbol{k}_d, \qquad \mathbf{W}^{\boldsymbol{k}}[f] = \sum_{S \text{ at multi-level } \boldsymbol{k}} \widehat{f}(S)^2,$$

so $\mathbf{I}[f] = \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \cdot |\boldsymbol{k}|$.

Finally, we recall the definition of *decision trees*. We may define the notion as follows. We say that $f\colon \{-1,1\}^n \to \{-1,1\}$ is computable as a depth-0 decision tree if it is constantly -1 or 1. We inductively say that it is computable as a depth-d decision tree if there is a coordinate $i \in [n]$ such that

$$f(x) = \begin{cases} f_0(x) & \text{if } x_i = 1, \\ f_1(x) & \text{if } x_i = -1, \end{cases}$$

where f_0 and f_1 are computable by depth-$(d-1)$ decision trees. We further say that the decision-tree computation is *read-once* if f_0 and f_1 depend on disjoint sets of coordinates and are themselves inductively read-once.

3 Symmetric and d-part-symmetric functions

In this section we prove Theorems 1 and 2, establishing the FEI Conjecture for symmetric and $O(1)$-part-symmetric functions. Although Theorem 2 strictly generalizes Theorem 1, we prefer to prove Theorem 1 separately and then generalize it afterward.

When $f\colon \{-1,1\}^n \to \{-1,1\}$ is symmetric we have $\widehat{f}(S)^2 = \mathbf{W}^k[f]/\binom{n}{k}$ whenever $|S| = k$. Hence

$$\boldsymbol{H}[\widehat{f}^2] \quad = \quad \sum_{k=0}^{n} \mathbf{W}^k[f] \log \tfrac{\binom{n}{k}}{\mathbf{W}^k[f]} \quad = \quad \sum_{k=0}^{n} \mathbf{W}^k[f] \log \binom{n}{k} \;+\; \sum_{k=0}^{n} \mathbf{W}^k[f] \log \tfrac{1}{\mathbf{W}^k[f]}. \tag{1}$$

Thus Theorem 1 is an immediate consequence of the following two theorems:

Theorem 4. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a symmetric function. Then $\sum_{k=0}^{n} \mathbf{W}^k[f] \log \binom{n}{k} \le C_1 \cdot \mathbf{I}[f]$, where $C_1 = \frac{1}{\ln 2}(1 + \frac{4\sqrt{2e}}{\sqrt{\pi}}) \le 9.04$.*

Theorem 5. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be any function, not necessarily symmetric. Then $\sum_{k=0}^{n} \mathbf{W}^k[f] \log \frac{1}{\mathbf{W}^k[f]} \le 3 \cdot \mathbf{I}[f]$.*

We prove these theorems in the subsequent subsections of the paper, following which we give the extension to d-part-symmetric functions.

3.1 Theorem 4: Derivatives of symmetric functions are noise-sensitive

We begin with an easy lemma.

Lemma 1. *Let $p_1, \dots, p_m$ be a nonnegative unimodal sequence; i.e., there exists $k \in [m]$ such that $p_1, \dots, p_k$ is a nondecreasing sequence and $p_k, \dots, p_m$ is a nonincreasing sequence. Let $g : [m] \to \{-1, 0, 1\}$ have the property that the sets $g^{-1}(-1)$ and $g^{-1}(1)$ are interleaving. Then $|\sum_{i=1}^{m} p_i g(i)| \le \max\{p_i\}$.*

Proof. It is without loss of generality to assume that g is never 0; for otherwise, we can restrict attention to the indices $i_1 \le i_2 \le \cdots \le i_\ell$ where g has value ± 1, noting that $p_{i_1}, \dots, p_{i_\ell}$ is still a unimodal sequence. Now if $m \le 1$ the result is trivial. Otherwise, choose $k \in [m-1]$ such that at least one of p_k, p_{k+1} equals $\max\{p_i\}$. By negating g if necessary we may assume that $g(k) = 1$, $g(k+1) = -1$. We complete the proof by showing $-p_{k+1} \le \sum_{i=1}^m p_i g(i) \le p_k$. For the upper bound, unimodality implies $p_{k+1} \ge p_{k+2} \ge p_{k+3} \cdots$; hence $\sum_{i=k+1}^m p_i g(i) = (-p_{k+1} + p_{k+2}) + (-p_{k+3} + p_{k+4}) + \cdots \le 0$ (regardless of whether this ends on $-p_m$ or $+p_m$). Similarly, we must have $\sum_{i=0}^{k-1} p_i g(i) \le 0$. Combining these, we obtain that $\sum_{i=1}^m p_i g(i) \le p_k$ as claimed. The proof of the lower bound $\sum_{i=1}^m p_i g(i) \ge -p_{k+1}$ is similar.

We now show the key theorem stated in Section 1.3 on the noise sensitivity of symmetric derivatives.

Theorem 6. *Let g be a discrete derivative of a symmetric function $f\colon \{-1,1\}^{n+1} \to \{-1,1\}$. Then for all real $1 \le c \le n$ it holds that $\mathbf{Stab}_{1-\frac{c}{n}}[g] \le \frac{2/\sqrt{\pi}}{\sqrt{c}} \mathbf{E}[g^2]$.*

Proof. Let (x, y) be a $(1-\frac{c}{n})$-correlated pair of random strings in $\{-1,1\}^n$. We will show that

$$\Big|\mathbf{E}[g(y) \mid x]\Big| \le \frac{2/\sqrt{\pi}}{\sqrt{c}}, \quad \text{independent of } x. \tag{2}$$

Since g is $\{-1,0,1\}$-valued, it will follow from (2) that

$$\mathbf{Stab}_{1-\frac{c}{n}}[g] = \mathbf{E}[g(x)g(y)] \le \frac{2/\sqrt{\pi}}{\sqrt{c}} \mathbf{E}[|g(x)|] = \frac{2/\sqrt{\pi}}{\sqrt{c}} \mathbf{E}[g^2],$$

as required. To show (2) we first observe that given x of Hamming weight s, the Hamming weight t of y is distributed as the sum of two independent binomial random variables, $t_1 \sim \mathrm{Bin}(s, 1-\frac{c}{2n})$ and $t_2 \sim \mathrm{Bin}(n-s, \frac{c}{2n})$. Being the sum of independent Bernoulli random variables, it is well known [13] that t has a unimodal probability distribution. Since g is a derivative of a symmetric Boolean-valued function, it is itself symmetric; hence we may identify it with a function $g : \{0, 1, \dots, n\} \to \{-1, 0, 1\}$. Further, because f is $\{-1,1\}$-valued, g must have the property that $g^{-1}(-1)$ and $g^{-1}(1)$ are interleaving subsets of $\{0, 1, \dots, n\}$. Thus (2) follows from Lemma 1 assuming that $\max_i\{\mathbf{Pr}[t = i]\} \le \frac{2/\sqrt{\pi}}{\sqrt{c}}$. To show this, we may assume without loss of generality that $s \ge n/2$. Then

$$\max_i\{\mathbf{Pr}[t = i]\} \le \max_i\{\mathbf{Pr}[t_1 = i]\} \le \frac{1}{\sqrt{2\pi s(1-\frac{c}{2n})\frac{c}{2n}}} \le \frac{1}{\sqrt{2\pi\frac{n}{2}\frac{1}{2}\frac{c}{2n}}} = \frac{2/\sqrt{\pi}}{\sqrt{c}}$$

where the second inequality is the basic estimate $\max_i\{\mathbf{Pr}[\mathrm{Bin}(m,p)] = i\} \le \frac{1}{\sqrt{2\pi m p(1-p)}}$ which uses Stirling's formula (see [5, Ch. VII.3]).

We can now give the proof of Theorem 4.

Proof. (Theorem 4.) Using $\binom{n}{k} \le (\frac{en}{k})^k$ for all $1 \le k \le n$ we have

$$\sum_{k=0}^{n} \mathbf{W}^k[f] \log\binom{n}{k} \le \sum_{k=1}^{n} \mathbf{W}^k[f] k \log(e) + \sum_{k=1}^{n} \mathbf{W}^k[f] k \log\tfrac{n}{k} = \frac{1}{\ln 2}\left(\mathbf{I}[f] + \sum_{k=1}^{n} \mathbf{W}^k[f] k \ln\tfrac{n}{k}\right)$$

and hence it suffices to show

$$\sum_{k=1}^{n} \mathbf{W}^k[f] k \ln\tfrac{n}{k} \le \frac{4\sqrt{2e}}{\sqrt{\pi}} \cdot \mathbf{I}[f]. \tag{3}$$

Let g be any derivative of f; say $g = D_n f$. By symmetry of f, the right side of (3) is $\frac{4\sqrt{2e}}{\sqrt{\pi}} \cdot n\,\mathbf{E}[g^2]$. As for the left side, for $k \in [n]$ we have

$$\mathbf{W}^{k-1}[g] = \sum_{\substack{S\subseteq[n-1]\\|S|=k-1}} \widehat{g}(S)^2 = \sum_{\substack{S\subseteq[n]\\|S|=k, S\ni n}} \widehat{f}(S)^2 = \frac{k}{n}\mathbf{W}^k[f].$$

Hence the left side of (3) is $\sum_{k=1}^n n\mathbf{W}^{k-1}[g]\ln\frac{n}{k}$. Thus after dividing by n we see that (3) is equivalent to

$$\sum_{k=0}^{n-1} \mathbf{W}^k[g]\ln\frac{n}{k+1} \leq \frac{4\sqrt{2e}}{\sqrt{\pi}} \cdot \mathbf{E}[g^2]. \tag{4}$$

Using the approximation $\ln m + \gamma + \frac{1}{2m} - \frac{1}{12m^2} \leq \sum_{j=1}^m \frac{1}{j} \leq \ln m + \gamma + \frac{1}{2m}$ one may obtain

$$\sum_{k=0}^{n-1} \mathbf{W}^k[g]\ln\frac{n}{k+1} \leq \sum_{k=0}^{n-1}\mathbf{W}^k[g]\sum_{j=k+1}^{n}\frac{1}{j} = \sum_{j=1}^{n}\frac{1}{j}\sum_{k=0}^{j-1}\mathbf{W}^k[g] \leq \exp(\tfrac{1}{2})\sum_{j=1}^{n}\frac{1}{j}\mathbf{Stab}_{1-\frac{1}{2j}}[g],$$

where in the last step we used that $\exp(\frac{1}{2})(1-\frac{1}{2j})^k \geq 1$ for all $k \leq j-1$. We may now apply Theorem 6 with $c = \frac{n}{2j}$ to obtain

$$\sum_{k=0}^{n} \mathbf{W}^k[g]\ln\frac{n}{k+1} \leq \exp(\tfrac{1}{2})\sum_{j=1}^{n}\frac{1}{j}\frac{2\sqrt{2}}{\sqrt{\pi}}\frac{\sqrt{j}}{\sqrt{n}} \cdot \mathbf{E}[g^2] \leq \frac{4\sqrt{2e}}{\sqrt{\pi}} \cdot \mathbf{E}[g^2],$$

using $\sum_{j=1}^n \frac{1}{\sqrt{j}} \leq 2\sqrt{n}$. Thus we have verified (4) and completed the proof.

3.2 Theorem 5: Spectral level entropy versus influence

In this section we establish Theorem 5. We begin with a well-known fact about "maximum entropy" which we prove for completeness:

Proposition 1. *Let K be a random variable supported on $\mathbb{Z}_+$. Then $\boldsymbol{H}[K] \leq \mathbf{E}[K]$.*

Proof. For $k \in \mathbb{Z}_+$ write $p_k = \mathbf{Pr}[K = k]$, and let G be a Geometric($\frac{1}{2}$) random variable. Then by the nonnegativity of binary relative entropy,

$$0 \leq \mathbf{D}(K||G) = \sum_{k=1}^{n} p_k \log\tfrac{p_k}{(1/2)^k} = -\boldsymbol{H}[P] + \sum_{k=1}^{n} p_k\log(2^k) = \mathbf{E}[K] - \boldsymbol{H}[K].$$

We will also need a simple corollary of the edge-isoperimetric inequality for $\{-1,1\}^n$.

Proposition 2. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ and write $\mathbf{W}^0[f] = 1-\epsilon$. Then $\mathbf{I}[f] \geq \frac{1}{2}\epsilon\log(1/\epsilon) + \epsilon$.*

Proof. By negating f if necessary we may assume $\mathbf{E}[f] \leq 0$. We think of f as the indicator of a subset $A \subseteq \{0,1\}^n$; the density of this set is $p = \frac{1}{2} + \frac{1}{2}\mathbf{E}[f] = \frac{1}{2} - \frac{1}{2}\sqrt{1-\epsilon} \leq 1/2$. It follows from the edge-isoperimetric inequality [10, 19, 2] that $\mathbf{I}[f]$, which is 2^{-n} times the size of A's edge-boundary, is at least $2p\log(1/p)$. It is then elementary to check that $(1-\sqrt{1-\epsilon})\log(\frac{2}{1-\sqrt{1-\epsilon}}) \geq \frac{1}{2}\epsilon\log(1/\epsilon)+\epsilon$, as required.

We can now establish Theorem 5.

Proof. (Theorem 5.) Write $\mathbf{W}^0[f] = 1-\epsilon$. We may assume $\epsilon > 0$ else the result is trivial. Then

$$\sum_{k=0}^{n} \mathbf{W}^{\boldsymbol{k}}[f] \log \tfrac{1}{\mathbf{W}^{k}[f]} \le \left(\sum_{k \neq 0} \mathbf{W}^{k}[f] \log \tfrac{1}{\mathbf{W}^{k}[f]} \right) + 2\epsilon \qquad \text{(as } (1-\epsilon)\log\tfrac{1}{1-\epsilon} \le \tfrac{1}{\ln 2}\epsilon \le 2\epsilon)$$

$$= \epsilon \left(\sum_{k \neq 0} \frac{\mathbf{W}^{k}[f]}{\epsilon} \log \frac{\epsilon}{\mathbf{W}^{k}[f]} \right) + \epsilon \log(1/\epsilon) + 2\epsilon$$

$$\le \epsilon \left(\sum_{k \neq 0} \frac{\mathbf{W}^{k}[f]}{\epsilon} \cdot k \right) + 2 \cdot \mathbf{I}[f] \quad = \quad 3 \cdot \mathbf{I}[f],$$

where the last inequality used Propositions 1 and 2.

3.3 Theorem 2: extension to d-part-symmetric functions

We show now how to extend the proof of Theorem 1 to obtain Theorem 2, the FEI Conjecture for d-part-symmetric functions. Suppose then that f is d-part-symmetric with respect to the partition $[n] = V_1 \cup\!\!\!\cdot \cdots \cup\!\!\!\cdot V_d$, where $|V_i| = n_i$, and recall the multi-level index notation $\boldsymbol{k}$ from Section 2.2. Since $\widehat{f}(S)^2 = \mathbf{W}^{\boldsymbol{k}}[f] / \prod_{i=1}^{d} \binom{n_i}{\boldsymbol{k}_i}$ whenever S is at multi-level $\boldsymbol{k}$, we have

$$\boldsymbol{H}[\widehat{f}^2] \quad = \quad \sum_{i=1}^{d} \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \log \binom{n_i}{\boldsymbol{k}_i} \; + \; \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \log \tfrac{1}{\mathbf{W}^{\boldsymbol{k}}[f]},$$

similarly to (1). Since

$$\mathbf{I}[f] = \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \cdot |\boldsymbol{k}| = \sum_{i=1}^{d} \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \cdot \boldsymbol{k}_i$$

we can prove Theorem 2 by establishing the following two generalizations of Theorems 4 and 5:

Theorem 7. *Let* $f \colon \{-1,1\}^{V_1 \cup\!\!\!\cdot \cdots \cup\!\!\!\cdot V_d} \to \{-1,1\}$ *be invariant under permutations of the coordinates in* V_i*. Then* $\sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \log \binom{n_i}{\boldsymbol{k}_i} \le C_1 \cdot \sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \cdot \boldsymbol{k}_i$*, where* $C_1 = \frac{1}{\ln 2}(1 + \frac{4\sqrt{2e}}{\sqrt{\pi}}) \le 9.04$.

Theorem 8. *Let* $f \colon \{-1,1\}^{V_1 \cup\!\!\!\cdot \cdots \cup\!\!\!\cdot V_d} \to \{-1,1\}$ *be any function, not necessarily with any symmetries. Then* $\sum_{\boldsymbol{k}} \mathbf{W}^{\boldsymbol{k}}[f] \log \frac{1}{\mathbf{W}^{\boldsymbol{k}}[f]} \le (3 + \log d) \cdot \mathbf{I}[f]$.

We begin by proving Theorem 7.

Proof. (Theorem 7.) We assume $i = d$ without loss of generality and write $\overline{V_d} = V_1 \cup\!\!\!\cdot \cdots \cup\!\!\!\cdot V_{d-1}$. For $y \in \{-1,1\}^{\overline{V_d}}$ we define $f_y \colon \{-1,1\}^{V_d} \to \{-1,1\}$ by $f_y(z) = f(y,z)$; the function f_y is symmetric for each y by assumption. Applying Theorem 4 to each f_y and then taking expectations we obtain

$$\sum_{k'=0}^{n_d} \mathop{\mathbf{E}}_{y}\big[\mathbf{W}^{k'}[f_y]\big] \log \binom{n_d}{k'} \le C_1 \cdot \sum_{k'=0}^{n_d} \mathop{\mathbf{E}}_{y}\big[\mathbf{W}^{k'}[f_y]\big] \cdot k'. \tag{5}$$

Now

$$\mathop{\mathbf{E}}_{y}\big[\mathbf{W}^{k'}[f_y]\big] = \sum_{\substack{S \subseteq V_d \\ |S| = k'}} \mathop{\mathbf{E}}_{y}\big[\widehat{f_y}(S)^2\big] = \sum_{\substack{S \subseteq V_d \\ |S| = k'}} \sum_{T \subseteq \overline{V_d}} \widehat{f}(T \cup S)^2 = \sum_{\boldsymbol{k} \colon \boldsymbol{k}_d = k'} \mathbf{W}^{\boldsymbol{k}}[f]$$

where the middle equality is an easy exercise using Parseval. Substituting this into (5) yields the desired inequality.

As for Theorem 8, its proof is essentially identical to that of Theorem 5, using the following generalization of Proposition 3:

Proposition 3. *Let $\boldsymbol{K}$ be a random variable supported on $\mathbb{N}^d \setminus \{\mathbf{0}\}$ and write $L = |\boldsymbol{K}|$. Then $\boldsymbol{H}[\boldsymbol{K}] \le (1+\log d)\,\mathbf{E}[L]$.*

Proof. Using the chain rule for entropy as well as Proposition 1, we have

$$\boldsymbol{H}[\boldsymbol{K}] = \boldsymbol{H}[L] + \boldsymbol{H}[\boldsymbol{K} \mid L] \le \mathbf{E}[L] + \sum_{\ell=1}^{\infty} \mathbf{Pr}[L=\ell] \cdot \boldsymbol{H}[\boldsymbol{K} \mid L=\ell]. \tag{6}$$

Given $L=\ell$ there are precisely $\binom{\ell+d-1}{d-1}$ possibilities for $\boldsymbol{K}$; hence

$$\boldsymbol{H}[\boldsymbol{K} \mid L=\ell] \le \log \binom{\ell+d-1}{d-1} = \ell \cdot \frac{\log\binom{\ell+d-1}{d-1}}{\ell} \le \ell \cdot \frac{\log\binom{1+d-1}{d-1}}{1} = \ell \cdot \log d. \tag{7}$$

The second inequality here follows from the technical Lemma 2 below. The proof is completed by substituting (7) into (6).

Here we give the technical lemma used in the previous proof.

Lemma 2. *For each $c \in \mathbb{N}$ the function $\frac{1}{\ell}\log\binom{\ell+c}{c}$ is a decreasing function of ℓ on $\mathbb{Z}_+$.*

Proof. We wish to show that for all $\ell \in \mathbb{Z}_+$,

$$\begin{aligned} \tfrac{1}{\ell}\ln\tbinom{\ell+c}{c} \ge \tfrac{1}{\ell+1}\ln\tbinom{\ell+1+c}{c} \quad &\Leftrightarrow \quad (\ell+1)\ln\tbinom{\ell+c}{c} \ge \ell\ln\tbinom{\ell+1+c}{c} \quad \Leftrightarrow \quad \tbinom{\ell+c}{c}^{\ell+1} \ge \tbinom{\ell+1+c}{c}^{\ell} \\ &\Leftrightarrow \quad \tbinom{\ell+c}{c} \ge \left[\tbinom{\ell+1+c}{c}/\tbinom{\ell+c}{c}\right]^{\ell} \quad \Leftrightarrow \quad \tbinom{\ell+c}{c} \ge (1+\tfrac{c}{\ell+1})^{\ell}. \end{aligned}$$

This last inequality is easily shown by induction on ℓ: if one multiplies both sides by $(1+\frac{c}{\ell+1})$ one obtains $\binom{\ell+1+c}{c} \ge (1+\frac{c}{\ell+1})^{\ell+1}$ which exceeds $(1+\frac{c}{\ell+2})^{\ell+1}$ as required for the induction step.

4 Proving the conjecture for read-once decision trees

In this section we prove Theorem 3, establishing the Fourier Entropy–Influence Conjecture with constant $C \le 4.88$ for functions computable by read-once decision trees. We begin with a technical lemma.

Lemma 3. *Let $\mu_0, \mu_1 \in [-1,1]$. Define*

$$g(\mu_0,\mu_1) = \left(\tfrac{\mu_0+\mu_1}{2}\right)^2 \log\left(\tfrac{2}{\mu_0+\mu_1}\right)^2 + \left(\tfrac{\mu_0-\mu_1}{2}\right)^2 \log\left(\tfrac{2}{\mu_0-\mu_1}\right)^2 + 2 - \tfrac{\mu_0^2}{2}\log\tfrac{4}{\mu_0^2} - \tfrac{\mu_1^2}{2}\log\tfrac{4}{\mu_1^2}.$$

Then

$$g(\mu_0,\mu_1) \le C(\tfrac{1}{2} - \tfrac{1}{2}\mu_0\mu_1) \tag{8}$$

holds for some universal constant C. The constant $C = 4.88$ suffices.

Proof. Note that g is unchanged when either of its arguments is negated. Since the right side of (8) only decreases when μ_0 and μ_1 have the same sign, we may assume both are nonnegative. By symmetry we may further assume $0 \le \mu_1 \le \mu_0 \le 1$. Write $\mu_1 = r\mu_0$ where $r \in [0,1]$. Simple manipulation shows that

$$g(\mu_0,\mu_1) = g(\mu_0, r\mu_0) = 2 - h(r)\mu_0^2,$$

where

$$h(r) = r^2 \log \tfrac{1}{r} - \tfrac{1}{2}(1-r)^2 \log \tfrac{1}{1-r} + \tfrac{1}{2}(1+r)^2 \log(1+r).$$

Thus it remains to show that

$$2 - h(r)\mu_0^2 \le C(\tfrac{1}{2} - \tfrac{1}{2}r\mu_0^2) \quad \Leftrightarrow \quad (\tfrac{1}{2}rC - h(r))\mu_0^2 \le \tfrac{1}{2}C - 2 \tag{9}$$

holds for all $\mu_0, r \in [0,1]$ for some C. Assuming we take $C \ge 4$, the right side of the second inequality in (9) will be nonnegative; hence, it holds if and only it holds for $\mu_0 = 1$. Putting $\mu_0 = 1$ into the first inequality in (9), we see that it suffices for $C \ge 4$ to be an upper bound on

$$\ell(r) = \frac{2 - h(r)}{\frac{1}{2} - \frac{1}{2}r}.$$

Plainly, ℓ is a continuous function on $(0,1)$ with $\ell(0+) = 0$. It is also easy to check that $\ell(1-) = 0$. It follows that ℓ is uniformly bounded on $[0,1]$; i.e., there is a universal C such that $\ell(r) \le C$ for all $r \in [0,1]$. Numerically, one may check that $\ell(r) \le 4.88$.

We may now prove Theorem 3 with C equal to the constant from Lemma 3.

Proof. (Theorem 3.) The proof is by induction on the depth d of the decision tree. In the base case, $d = 0$, the function f is constant and so $\boldsymbol{H}[\widehat{f}^2] \le C \cdot \mathbf{I}[f]$ holds trivially. For the general case, assume without loss of generality that the root of the tree queries coordinate n. Let f_0 and f_1 be the subfunctions computed by the $x_n = 1$ and $x_n = -1$ subtrees, respectively. Since the decision tree is read-once, the functions f_0 and f_1 depend on disjoint sets of the coordinates from $[n-1]$, say J_0 and J_1. In particular, if we write $\mu_i = \mathbf{E}[f_i] = \widehat{f}_i(\emptyset)$ for $i = 0,1$, then $\mathbf{E}[f_0 f_1] = \mu_0\mu_1$. Also, f_0 and f_1 are computable by read-once decision trees of depth $d-1$; hence we may later apply the induction hypothesis to them. Our task is now to prove $\boldsymbol{H}[\widehat{f}^2] \le C \cdot \mathbf{I}[f]$ for f.

For any Boolean function f whose subfunctions based on x_n are f_0 and f_1, it holds that

$$\mathbf{I}[f] = \mathbf{Inf}_n[f] + \tfrac{1}{2}\mathbf{I}[f_0] + \tfrac{1}{2}\mathbf{I}[f_1].$$

In our case, $\mathbf{Inf}_n[f] = \mathbf{Pr}[f_0 \ne f_1] = \frac{1}{2} - \frac{1}{2}\,\mathbf{E}[f_0 f_1] = \frac{1}{2} - \frac{1}{2}\mu_0\mu_1$. Thus

$$\mathbf{I}[f] = \tfrac{1}{2} - \tfrac{1}{2}\mu_0\mu_1 + \tfrac{1}{2}\mathbf{I}[f_0] + \tfrac{1}{2}\mathbf{I}[f_1]. \tag{10}$$

As for $\boldsymbol{H}[\widehat{f}^2]$, we have

$$f(x) = \frac{1+x_n}{2} f_0(x) + \frac{1-x_n}{2} f_1(x) = \tfrac{1}{2}(f_0(x) + f_1(x)) + \tfrac{1}{2}x_n(f_0(x) - f_1(x)).$$

It follows that the Fourier coefficients of f, which are indexed by $S \subseteq J_0 \cup J_1 \cup \{n\}$, are given by

$$\widehat{f}(S) = \begin{cases} \tfrac{1}{2}(\mu_0 + \mu_1) & \text{if } S = \emptyset, \\ \tfrac{1}{2}(\mu_0 - \mu_1) & \text{if } S = \{n\}, \\ \tfrac{1}{2}\widehat{f}_i(S) & \text{if } \emptyset \ne S \subseteq J_i, \\ (-1)^i \tfrac{1}{2}\widehat{f}_i(T) & \text{if } S = T \cup \{n\},\ \emptyset \ne T \subseteq J_i. \end{cases}$$

Note that for each $\emptyset \ne S \subseteq J_i$, we have the quantity $\frac{\widehat{f}_i(S)^2}{4}$ occurring twice in the squared Fourier spectrum of f. We may therefore compute

$$\boldsymbol{H}[\widehat{f}^2] = \left(\tfrac{\mu_0+\mu_1}{2}\right)^2 \log\left(\tfrac{2}{\mu_0+\mu_1}\right)^2 + \left(\tfrac{\mu_0-\mu_1}{2}\right)^2 \log\left(\tfrac{2}{\mu_0-\mu_1}\right)^2 + 2\sum_{i=0,1}\sum_{\emptyset \ne S \subseteq J_i} \tfrac{\widehat{f}_i(S)^2}{4} \log \tfrac{4}{\widehat{f}_i(S)^2}. \tag{11}$$

We simplify the expression on the right in (11):

$$
\begin{aligned}
2\sum_{i=0,1}\sum_{\emptyset\neq S\subseteq J_i}\tfrac{\widehat{f}_i(S)^2}{4}\log\tfrac{4}{\widehat{f}_i(S)^2} &= 2\sum_{i=0,1}\left(-\tfrac{\mu_i^2}{4}\log\tfrac{4}{\mu_i^2}+\sum_{S\subseteq J_i}\tfrac{\widehat{f}_i(S)^2}{4}\log\tfrac{4}{\widehat{f}_i(S)^2}\right)\\
&= 2\sum_{i=0,1}\left(-\tfrac{\mu_i^2}{4}\log\tfrac{4}{\mu_i^2}+\tfrac{1}{4}\boldsymbol{H}[\widehat{f_i}^2]+\tfrac{1}{4}\log 4\right)\\
&= 2-2\sum_{i=0,1}\tfrac{\mu_i^2}{4}\log\tfrac{4}{\mu_i^2}+\tfrac{1}{2}\sum_{i=0,1}\boldsymbol{H}[\widehat{f_i}^2]. \qquad (12)
\end{aligned}
$$

Substituting (12) into (11) yields

$$\boldsymbol{H}[\widehat{f}^2] = g(\mu_0,\mu_1)+\tfrac{1}{2}\boldsymbol{H}[\widehat{f_0}^2]+\tfrac{1}{2}\boldsymbol{H}[\widehat{f_1}^2], \qquad (13)$$

where $g(\mu_0,\mu_1)$ is the function from Lemma 3. Using this lemma along with the induction hypothesis on f_0 and f_1, we obtain from (13) that

$$\boldsymbol{H}[\widehat{f}^2] \leq C(\tfrac{1}{2}-\tfrac{1}{2}\mu_0\mu_1)+\tfrac{1}{2}C\cdot\mathbf{I}[f_0]+\tfrac{1}{2}C\cdot\mathbf{I}[f_1] = C\cdot\mathbf{I}[f],$$

where we used (10). The induction is complete.

Remark: The above proof technique can be extended to handle the more general class of "recursively read-once functions" (as defined in [18]). It seems that one can show $C = 19$ suffices for this class; however, a formal proof would require proving a numerical lemma much more difficult than Lemma 3. Details will appear elsewhere.

5 Closing remarks

As the general Fourier Entropy–Influence Conjecture seems difficult to resolve, one may try to prove it for additional interesting classes of Boolean functions. For example, Wan has suggested linear threshold functions as a possibly tractable case [22]. One may also try tackling the Fourier Min-Entropy–Influence Conjecture first (or the slightly stronger consequence that $\exists S\neq\emptyset$ s.t. $\widehat{f}(S)^2 \geq 2^{-C\cdot\mathbf{I}[f]/\mathbf{Var}[f]}$). However this will already likely require stronger tools than the ones used in this paper. The reason is that as mentioned in Section 1.1, this conjecture implies the KKL Theorem, and there is no known proof of KKL which avoids the Hypercontractive or Logarithmic Sobolev Inequalities.

Acknowledgments. The authors would like to thank Rocco Servedio, Li-Yang Tan, and Andrew Wan for sharing their insights on the Entropy–Influence Conjecture.

References

1. Michael Ben-Or and Nathan Linial. Collective coin flipping. In Silvio Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990.
2. Arthur Bernstein. Maximally connected arrays on the n-cube. *SIAM Journal on Applied Mathematics*, pages 1485–1489, 1967.
3. Ravi Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.
4. Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *Proceedings of the 14th Annual International Workshop on Randomized Techniques in Computation*, pages 504–517, 2010.

5. William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, 3rd edition, 1968.
6. Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996.
7. Parikshit Gopalan, Adam Kalai, and Adam Klivans. Agnostically learning decision trees. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 527–536, 2008.
8. Parikshit Gopalan, Adam Kalai, and Adam Klivans. A query algorithm for agnostically learning DNF? In *Proceedings of the 21st Annual Conference on Learning Theory*, pages 515–516, 2008.
9. Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
10. Lawrence Harper. Optimal assignments of numbers to vertices. *Journal of the Society for Industrial and Applied Mathematics*, 12(1):131–135, 1964.
11. Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.
12. Gil Kalai. The entropy/influence conjecture. Posted on Terence Tao's *What's new* blog, http://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/, 2007.
13. Julian Keilson and Hans Gerber. Some results for discrete unimodality. *Journal of the American Statistical Association*, 66(334):386–389, 1971.
14. Adam Klivans, Homin Lee, and Andrew Wan. Mansours Conjecture is true for random DNF formulas. In *Electronic Colloquium on Computational Complexity TR10-023*, 2010.
15. Yishay Mansour. Learning boolean functions via the Fourier Transform. In Vwani Roychowdhury, Kai-Yeung Siu, and Alon Orlitsky, editors, *Theoretical Advances in Neural Computation and Learning*, chapter 11, pages 391–424. Kluwer Academic Publishers, 1994.
16. Yishay Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995.
17. Ryan O'Donnell. Some topics in analysis of boolean functions. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 569–578, 2008.
18. Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco Servedio. Every decision tree has an influential variable. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 31–39, 2005.
19. Kenneth Steiglitz and Arthur Bernstein. Optimal binary coding of ordered numbers. *Journal of the Society for Industrial and Applied Mathematics*, 13(2):441–443, 1965.
20. Michel Talagrand. On Russo's approximate zero-one law. *Annals of Probability*, 22(3):1576–1587, 1994.
21. Robert Titsworth. *Correlation properties of cyclic sequences*. PhD thesis, California Institute of Technology, 1962.
22. Andrew Wan, 2010. Talk at the Center for Computational Intractability.