Upper Bounds on Fourier Entropy

Sourav Chakraborty[1], Raghav Kulkarni[2], Satyanarayana V. Lokam[3], and Nitin Saurabh[4]

[1] Chennai Mathematical Institute, Chennai, India
sourav@cmi.ac.in
[2] Centre for Quantum Technologies, Singapore
kulraghav@gmail.com
[3] Microsoft Research India, Bangalore, India
satya@microsoft.com
[4] The Institute of Mathematical Sciences, Chennai, India
nitin@imsc.res.in

Abstract. Given a function $f : \{0,1\}^n \to \mathbb{R}$, its *Fourier Entropy* is defined to be $-\sum_S \widehat{f}^2(S) \log \widehat{f}^2(S)$, where $\hat{f}$ denotes the Fourier transform of f. This quantity arises in a number of applications, especially in the study of Boolean functions. An outstanding open question is a conjecture of Friedgut and Kalai (1996), called the Fourier Entropy Influence (FEI) Conjecture, asserting that the Fourier Entropy of any Boolean function f is bounded above, up to a constant factor, by the total influence (= average sensitivity) of f.

In this paper we give several upper bounds on the Fourier Entropy of Boolean as well as real valued functions. We first give upper bounds on the Fourier Entropy of Boolean functions in terms of several complexity measures that are known to be bigger than the influence. These complexity measures include, among others, the logarithm of the number of leaves and the average depth of a parity decision tree. We then show that for the class of Linear Threshold Functions (LTF), the Fourier Entropy is at most $O(\sqrt{n})$. It is known that the average sensitivity for the class of LTF is bounded by $\Theta(\sqrt{n})$. We also establish a bound of $O_d(n^{1-\frac{1}{4d+6}})$ for general degree-d polynomial threshold functions. Our proof is based on a new upper bound on the *derivative of noise sensitivity*. Next we proceed to show that the FEI Conjecture holds for read-once formulas that use AND, OR, XOR, and NOT gates. The last result is independent of a recent result due to O'Donnell and Tan [15] for read-once formulas with *arbitrary* gates of bounded fan-in, but our proof is completely elementary and very different from theirs. Finally, we give a general bound involving the first and second moments of sensitivities of a function (average sensitivity being the first moment), which holds for real valued functions as well.

1 Introduction

Fourier transforms are extensively used in a number of fields such as engineering, mathematics, and computer science. Within theoretical computer science, Fourier analysis of Boolean functions evolved into one of the most useful and versatile tools; see the book [14] for a comprehensive survey of this area and pointers to numerous other sources of literature on this subject. In particular, it plays an important role in several results in complexity theory, learning theory, social choice, inapproximability, metric spaces, etc. If $\hat{f}$ denotes the Fourier transform of a Boolean function f, then $\sum_{S\subseteq[n]} \widehat{f}^2(S) = 1$ and hence we can define an entropy of the distribution given by $\widehat{f}^2(S)$:

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

The Fourier Entropy-Influence (FEI) Conjecture, made by Friedgut and Kalai [6] in 1996, states that for every Boolean function, its Fourier entropy is bounded above by its total influence :

Fourier Entropy-Influence Conjecture There exists a universal constant C such that for all $f : \{0,1\}^n \to \{+1,-1\}$,

$$\mathbb{H}(f) \leq C \cdot \mathsf{Inf}(f) \quad , \tag{2}$$

where $\mathsf{Inf}(f)$ is the total influence of f which is the same as the average sensitivity $\mathsf{as}(f)$ of f. The latter quantity may be intuitively viewed as the expected number of coordinates of an input which, when flipped, will cause the value of f to be changed, where the expectation is w.r.t. the uniform distribution on the input assignments of f.

1.1 Motivation

Resolving the FEI conjecture is one of the most important open problems in the Fourier analysis of Boolean functions. The conjecture intuitively asserts that if the Fourier coefficients of a Boolean function are "smeared out," then its influence must be large, i.e., at a typical input, the value of f changes in several different directions. The original motivation for the conjecture in [6] stems from a study of threshold phenomena in random graphs.

The FEI conjecture has numerous applications. It implies a variant of *Mansour's Conjecture* [12] stating that for a Boolean function computable by a DNF formula with m terms, most of its Fourier mass is concentrated on $\mathsf{poly}(m)$-many coefficients. A proof of Mansour's conjecture would imply a polynomial time *agnostic* learning algorithm for DNF's [7] answering a major open question in computational learning theory.

The FEI conjecture also implies that for any n-vertex graph property, the influence is at least $c(\log n)^2$. The best known lower bound, by Bourgain and Kalai [1], is $\Omega((\log n)^{2-\epsilon})$, for any $\epsilon > 0$. See [10], [16] and [11] for a detailed explanation on these and other consequences of the conjecture.

1.2 Prior Work

The first progress on the FEI conjecture was made in 2010 in [11] showing that the conjecture holds for random DNFs. O'Donnell et al. [16] proved that the conjecture holds for symmetric functions and more generally for any d-part symmetric functions for constant d. They also established the conjecture for functions computable by read-once decision trees. Keller et al. [10] studied a generalization of the conjecture to biased product measures on the Boolean cube and proved a variant of the conjecture for function with extremely low Fourier weight on the high levels. O'Donnell and Tan [15] verified the conjecture for read-once formulas using a composition theorem for the FEI conjecture. Wan et al. [17] studies the conjecture from the point of view of existence of efficient prefix-free codes for the random variable, $\mathcal{X} \sim \hat{f}^2$, that is distributed according to $\hat{f}^2$. Using this interpretation they verify the conjecture for bounded read decision trees. It is also relatively easy to show that the FEI conjecture holds for a random Boolean function, e.g., see [3] for a proof. By direct calculation, one can verify the conjecture for simple functions like AND, OR, $\mathsf{Majority}$, Tribes etc.

1.3 Our results

We report here various upper bounds on Fourier entropy that may be viewed as progress toward the FEI conjecture.

Upper bounds by Complexity Measures. The $\mathsf{Inf}(f)$ of a Boolean function f is used to derive lower bounds on a number of complexity parameters of f such as the number of leaves or the average depth of a decision tree computing f. Hence a natural weakening of the FEI conjecture is to prove upper bounds on the Fourier entropy in terms of such complexity measures of Boolean functions. By a relatively easy argument, we show that

$$\mathbb{H}(f) = O(\log \mathsf{L}(f)), \tag{3}$$

where $\mathsf{L}(f)$ denotes the minimum number of leaves in a decision tree that computes f. If $\mathsf{DNF}(f)$ denotes the minimum size of a DNF for the function f, note that $\mathsf{DNF}(f) \leq \mathsf{L}(f)$. Thus improving (3) with $O(\log \mathsf{DNF}(f))$ on the right hand side would resolve *Mansour's conjecture* – a long-standing open question about sparse Fourier approximations to DNF formulas motivated by applications to learning theory – and a special case of the FEI conjecture for DNF's. We note that (3) also holds when the queries made by the decision tree involve parities of subsets of variables, conjunctions of variables, etc. It also holds when $\mathsf{L}(f)$ is generalized to the number of subcubes in a *subcube partition* that represents f. Note that for a Boolean function

$$\mathsf{Inf}(f) \leq \log(L_c(f)) \leq \log(\mathsf{L}(f)) \leq D(f),$$

where $L_c(f)$ is number of subcubes in a *subcube partition* that represents f and $D(f)$ is the depth of the decision tree computing f.

We also prove the following strengthening of (3):

$$\mathbb{H}(f) = O(\bar{\mathsf{d}}(f)), \tag{4}$$

where $\bar{\mathsf{d}}(f)$ denotes the *average depth* of a decision tree computing f (observe that $\bar{\mathsf{d}}(f) \leq \log(\mathsf{L}(f)))$. Note that the average depth of a decision tree is also a kind of entropy: it is given by the distribution induced on the leaves of a decision tree when an input is drawn uniformly at random. Thus (4) relates the two kinds of entropy, but only up to a constant factor. We further strengthen (4) by improving the right-hand side in (4) to *average depth* of a *parity* decision tree computing f, that is, queries made by the decision tree are parities of a subset of variables.

Upper bounds on the Fourier Entropy of Polynomial Threshold Functions. The Fourier Entropy-Influence conjecture is known to be true for unweighted threshold functions, i.e., when $f(x) = \mathsf{sign}(x_1 + \cdots + x_n - \theta)$ for some integer $\theta \in [0..n]$. This follows as a corollary of the result due to O'Donnell et al. [16] that the FEI conjecture holds for all symmetric Boolean functions. It is known that the influence for the class of linear threshold functions is bounded by $\Theta(\sqrt{n})$ (where the lower bound is witnessed by Majority [13]). Recently Harsha et al. [8] studied average sensitivity of polynomial threshold function (see also [5]). They proved that average sensitivity of degree-d polynomial threshold functions is bounded by $O_d(n^{1-(1/4d+6)})$, where $O_d(\cdot)$ denotes that the constant depends on degree d. This suggests a natural and important weakening of the FEI conjecture: *Is Fourier Entropy of polynomial threshold functions bounded by a similar function of n as their average sensitivity?* In this paper we answer this question in the positive. An important ingredient in our proof is a bound on the derivative of noise sensitivity in terms of the noise parameter.

FEI inequality for Read-Once Formulas. We also prove that the FEI conjecture holds for a special class of Boolean functions: Read-Once Formulas over {AND, OR and XOR}, i.e., functions computable by a tree with AND, OR and XOR gates at internal nodes and each variable (or its negation) occurring *at most once* at the leaves. Our result is independent of a very recent result by O'Donnell and Tan [15] that proves the FEI conjecture holds for read-once formulas that allow *arbitrary* gates of bounded fan-in. However, our proof is completely elementary and very different from theirs. Prior to these results, O'Donnell et al. [16] proved that the FEI conjecture holds for read-once *decision trees*. Our result for read-once formulas is a strict generalization of their result. For instance, the tribes function is computable by read-once formulas but not by read-once decision trees. Our proof for read-once formulas is a consequence of a kind of tensorizability for $\{0, 1\}$-valued Boolean functions. In particular, we show that an inequality similar to the FEI inequality is preserved when functions depending on *disjoint* sets of variables are combined by AND, OR and XOR operators.

A Bound for Real valued Functions via Second Moment. Recall [9] that total influence $\mathsf{Inf}(f)$ or average sensitivity $\mathsf{as}(f)$ is related to $\hat{f}$ by the well-known identity: $\mathsf{as}(f) = \mathsf{Inf}(f) = \sum_S |S| \widehat{f}^2(S)$. Hence, an equivalent way to state the FEI conjecture is that there is an absolute constant C such that for all Boolean f,

$$\mathbb{H}(f) \leq C \cdot \sum_S |S| \widehat{f}^2(S) \ . \tag{5}$$

Here, we prove that for all δ, $0 \leq \delta \leq 1$, and for all f with $\sum_S \widehat{f}^2(S) = 1$, and hence for Boolean f in particular,

$$\mathbb{H}(f) \leq \sum_S |S|^{1+\delta} \widehat{f}^2(S) \ + \ (\log n)^{O(1/\delta)} \ . \tag{6}$$

An alternative interpretation of the above theorem states

$$\mathbb{H}(f) \leq \mathsf{as}(f)^{1-\delta} \cdot \mathsf{as}_2(f)^{\delta} \ + \ (\log n)^{O(1/\delta)} \ , \tag{7}$$

where $\mathsf{as}_2(f) := \sum_S |S|^2 \widehat{f}^2(S)$. We also mention that $\mathsf{as}_2(f) \leq \mathsf{s}(f)^2$ (see [2]), where $\mathsf{s}(f)$ is the *maximum* sensitivity of f.

It is important to note that (6) holds for *arbitrary*, i.e., even non-Boolean, f such that (without loss of generality) $\sum_S \widehat{f}^2(S) = 1$. On the other hand, there are examples of non-Boolean f for which the FEI conjecture (5) is *false*. Combining (7) with a "tensorizability" property [16] of $\mathbb{H}(f)$ and $\mathsf{as}(f)$, it is possible to show that for *all* f, $\mathbb{H}(f) = O(\mathsf{as}(f) \log n)$. Hence proving the FEI conjecture should involve removing the "extra" log factor while exploiting the Boolean nature of f.

Remainder of the paper. We give basic definitions in Section 2. Section 3 contains upper bounds in terms of complexity measures. In Section 4 and Section 5 we consider special classes of Boolean functions namely, the *polynomial threshold functions* and *Read-Once formulas*. We then provide bounds for real valued functions in Section 6. Due to space limitations, proofs had to be omitted from this extended abstract. For a more complete version, please see [2].

2 Preliminaries

We recall here some basic facts of Fourier analysis. For a detailed treatment please refer to [4, 14]. Consider the space of all functions from $\{0,1\}^n$ to $\mathbb{R}$, endowed with the inner product $\langle f, g \rangle = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)g(x)$. The character functions $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$ for $S \subseteq [n]$ form an orthonormal basis for this space of functions w.r.t. the above inner product. Thus, every function $f : \{0,1\}^n \longrightarrow \mathbb{C}$ of n boolean variables has the *unique Fourier* expansion: $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x)$. The vector $\hat{f} = (\hat{f}(S))_{S \subseteq [n]}$ is called the Fourier transform of the function f. The Fourier coefficient $\hat{f}(S)$ of f at S is then given by, $\hat{f}(S) = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x)$. The norm of a function f is defined to be $\|f\| = \sqrt{\langle f, f \rangle}$. Orthonormality of $\{\chi_S\}$ implies *Parseval's identity*: $\|f\|^2 = \sum_S \hat{f}^2(S)$.

We only consider *finite probability distributions* in this paper. The *entropy* of a distribution $\mathcal{D}$ is given by, $\mathbb{H}(\mathcal{D}) := \sum_{i \in \mathcal{D}} p_i \log \frac{1}{p_i}$. In particular, the *binary* entropy function, denoted by $\mathrm{H}(p)$, equals $-p \log p - (1-p)\log(1-p)$. All logarithms in the paper are base 2, unless otherwise stated.

We consider Boolean functions with range $\{-1, +1\}$. For an $f : \{0,1\}^n \to \{-1, +1\}$, $\|f\|$ is clearly 1 and hence Parseval's identity shows that for Boolean functions $\sum_S \hat{f}^2(S) = 1$. This implies that the squared Fourier coefficients can be thought of as a probability distribution and the notion of Fourier Entropy (1) is well-defined.

The *influence of f in the i-th direction*, denoted $\mathsf{Inf}_i(f)$, is the fraction of inputs at which the value of f gets flipped if we flip the i-th bit:

$$\mathsf{Inf}_i(f) = 2^{-n} |\{x \in \{0,1\}^n : f(x) \neq f(x \oplus e_i)\}| \ ,$$

where $x \oplus e_i$ is obtained from x by flipping the i-th bit of x.

The (total) *influence* of f, denoted by $\mathsf{Inf}(f)$, is $\sum_{i=1}^n \mathsf{Inf}_i(f)$. The influence of i on f can be shown, e.g., [9], to be $\mathsf{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2$ and hence it follows that $\mathsf{Inf}(f) = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2$.

For $x \in \{0,1\}^n$, the *sensitivity of f at x*, denoted $\mathsf{s}_f(x)$, is defined to be $\mathsf{s}_f(x) := |\{i : f(x) \neq f(x \oplus e_i), 1 \leq i \leq n\}|$, i.e., the number of coordinates of x, which when flipped, will flip the value of f. The (maximum) *sensitivity of the function f*, denoted $\mathsf{s}(f)$, is defined to be the largest sensitivity of f at x over all $x \in \{0,1\}^n$: $\mathsf{s}(f) := \max\{\mathsf{s}_f(x) : x \in \{0,1\}^n\}$. The *average sensitivity of f*, denoted $\mathsf{as}(f)$, is defined to be $\mathsf{as}(f) := 2^{-n} \sum_{x \in \{0,1\}^n} s_f(x)$. It is easy to see that $\mathsf{Inf}(f) = \mathsf{as}(f)$ and hence we also have $\mathsf{as}(f) = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2$.

The noise sensitivity of f at ϵ, $0 \leq \epsilon \leq 1$, denoted $\mathsf{NS}_\epsilon(f)$, is given by $\Pr_{x, y \sim_\epsilon x} [f(x) \neq f(y)]$ where $y \sim_\epsilon x$ denotes that y is obtained by flipping each bit of x independently with probability ϵ. It is easy to see that $\mathsf{NS}_\epsilon(f) = \frac{1}{2} - \frac{1}{2} \sum_S (1 - 2\epsilon)^{|S|} \hat{f}(S)^2$. Hence the derivative of $\mathsf{NS}_\epsilon(f)$ with respect to ϵ, denoted $\mathsf{NS}_\epsilon{}'(f)$, equals $\sum_{S \neq \emptyset} |S|(1-2\epsilon)^{|S|-1} \hat{f}(S)^2$.

3 Bounding Entropy using Complexity Measures

In this section, we prove upper bounds on Fourier entropy in terms of some complexity parameters associated to decision trees and subcube partitions.

3.1 *via* leaf entropy : Average Decision Tree Depth

Let T be a decision tree computing $f : \{0,1\}^n \to \{+1, -1\}$ on variable set $X = \{x_1, \ldots, x_n\}$. If $A_1, \ldots, A_L$ are the sets (with repetitions) of variables queried along the root-to-leaf paths in the tree T, then the average depth (w.r.t. the uniform distribution on inputs) of T is defined to be $\bar{d} := \sum_{i=1}^{L} |A_i| 2^{-|A_i|}$. Note that the average depth of a decision tree is also a kind of entropy: if each leaf λ_i is chosen with the probability $p_i = 2^{-|A_i|}$ that a uniformly chosen random input reaches it, then the entropy of the distribution induced on the λ_i is $\mathbb{H}(\lambda_i) = -\sum_i p_i \log p_i = \sum_i |A_i| 2^{-|A_i|}$. Here, we will show that *the Fourier entropy is at most twice the leaf entropy of a decision tree.*

W.l.o.g., let x_1 be the variable queried by the root node of T and let T_1 and T_2 be the subtrees reached by the branches $x_1 = +1$ and $x_1 = -1$ respectively and let g_1 and g_2 be the corresponding functions computed on variable set $Y = X \setminus \{x_1\}$. Let $\bar{d}$ be the average depth of T and $\bar{d_1}$ and $\bar{d_2}$ be the average depths of T_1 and T_2 respectively. We first observe a fairly straightforward lemma relating Fourier coefficients of f to the Fourier coefficients of restrictions of f.

Lemma 1. *Let $S \subseteq \{2, \ldots, n\}$.*

(i) $\widehat{f}(S) = (\widehat{g_1}(S) + \widehat{g_2}(S))/2.$
(ii) $\widehat{f}(S \cup \{1\}) = (\widehat{g_1}(S) - \widehat{g_2}(S))/2.$
(iii) $\bar{d} = (\bar{d_1} + \bar{d_2})/2 + 1.$

Using Lemma 1 and concavity of entropy we establish the following technical lemma, which relates the entropy of f to entropies of restrictions of f.

Lemma 2. *Let g_1 and g_2 be defined as before in Lemma 1. Then,*

$$\mathbb{H}(f) \leq \frac{1}{2}\,\mathbb{H}(g_1) + \frac{1}{2}\,\mathbb{H}(g_2) + 2 \ . \tag{8}$$

Let $\bar{\mathsf{d}}(f)$ denote the minimum *average* depth of a decision tree computing f. As a consequence of Lemma 2 we obtain the following bound.

Theorem 1. *For every Boolean function f, $\mathbb{H}(f) \leq 2 \cdot \bar{\mathsf{d}}(f)$.*

Remark 1. The constant 2 in the bound of Theorem 1 cannot be replaced by 1. Indeed, let $f(x, y) = x_1 y_1 + \cdots + x_{n/2} y_{n/2} \bmod 2$ be the inner product mod 2 function. Then because $\widehat{f}^2(S) = 2^{-n}$ for all $S \subseteq [n]$, $\mathbb{H}(f) = n$. On the other hand, it can be shown that $\bar{\mathsf{d}}(f) = \frac{3}{4}n - o(n)$. Hence, the constant must be at least 4/3.

Average Parity Decision Tree Depth Let L be a linear transformation. Applying the linear transformation on a Boolean function f we obtain another Boolean function Lf which is defined as $Lf(x) := f(Lx)$, for all $x \in \{0,1\}^n$. Before proceeding further, we note down a useful observation.

Proposition 2 *Let $f : \{0,1\}^n \to \{+1, -1\}$ be a Boolean function. For an invertible linear transformation $L \in \mathsf{GL}_n(\mathbb{F}_2)$, $\mathbb{H}(f) = \mathbb{H}(Lf)$.*

Let T be a parity decision tree computing $f : \{0,1\}^n \to \{+1,-1\}$ on variable set $X = \{x_1, \ldots, x_n\}$. Note that a parity decision tree computing f also computes Lf and vice versa. This implies that we can always ensure that a variable is queried at the root node of T via a linear transformation. Let us denote the new variable set, after applying the linear transformation, by $Y = \{y_1, \ldots, y_n\}$. W.l.o.g, let y_1 be the variable queried at the root. Let T_1 and T_2 be the subtrees reached by the branches $y_1 = 0$ and $y_1 = 1$ respectively and let g_1 and g_2 be the corresponding functions computed on variable set $Y \setminus \{y_1\}$. Using Proposition 2 we see that the proof of Lemma 1 and Lemma 2 goes through in the setting of parity decision trees too. Hence, we get the following strengthening of Theorem 1.

Theorem 3. *For every Boolean function f, $\mathbb{H}(f) \leq 2 \cdot \oplus\text{-}\bar{\mathsf{d}}(f)$, where $\oplus\text{-}\bar{\mathsf{d}}(f)$ denotes the minimum average depth of a parity decision tree computing f.*

3.2 *via* L_1-norm (or Concentration) : Decision Trees and Subcube Partitions

Note that a decision tree computing a Boolean function f induces a partition of the cube $\{0,1\}^n$ into monochromatic subcubes, i.e., f has the same value on all points in a given subcube, with one subcube corresponding to each leaf. But there exist monochromatic subcube partitions that are not induced by any decision tree. Consider any subcube partition $\mathcal{C}$ computing f (see [2]).There is a natural way to associate a probability distribution with $\mathcal{C}$: C_i has probability mass $2^{-(\text{number of co-ordinates fixed by } C_i)}$. Let us call the entropy associated with this probability distribution *partition entropy.* Based on the results of the previous subsection, a natural direction would be to prove that the Fourier entropy is bounded by the partition entropy. Unfortunately we were not quite able to show that but, interestingly, there is a very simple proof to see that the Fourier entropy is bounded by the logarithm of the number of partitions in $\mathcal{C}$. In fact, the proof gives a slightly better upper bound of the logarithm of the spectral-norm of f. For completeness sake, we note this observation [2] but we remark that it should be considered folklore. Our goal in presenting the generalization to subcube partitions is also to illustrate a different approach. The approach uses the concentration property of the Fourier transform and uses a general, potentially powerful, technique. One way to do this is to use a result due to Bourgain and Kalai (Theorem 3.2 in [10]). However, we give a more direct proof (see [2]) for the special case of subcube partitions.

4 Upper bound on Fourier Entropy of Threshold Functions

In this section, we establish a better upper bound on the Fourier entropy of polynomial threshold functions. We show that the Fourier entropy of a linear threshold function is bounded by $O(\sqrt{n})$, and for a degree-d threshold function it is bounded by $O_d(n^{1-\frac{1}{4d+6}})$. We remark that the bound is significant because the average sensitivity of a linear threshold function on n variables is bounded by $O(\sqrt{n})$, and this is tight. Also the bound on the Fourier entropy of degree-d threshold functions is the best known bound on their average sensitivity [8, 5].

For $f : \{0,1\}^n \to \{+1,-1\}$, let $W^k[f] := \sum_{|S|=k} \hat{f}(S)^2$ and $W^{\geq k}[f] := \sum_{|S|\geq k} \hat{f}(S)^2$. We now state our main technical lemma which translates a bound on noise sensitivity to a bound on the derivative of noise sensitivity.

Lemma 3. *Let $f : \{0,1\}^n \to \{+1,-1\}$ be such that $\mathsf{NS}_\epsilon(f) \leq \alpha \cdot \epsilon^\beta$, where α is independent of ϵ and $\beta < 1$. Then, $\mathsf{NS}_\epsilon{}'(f) \leq \frac{3}{1-e^{-2}} \cdot \frac{\alpha}{1-\beta} \cdot (1/\epsilon)^{1-\beta}$.*

From [16] we have the following bound on entropy.

Lemma 4. *[16] Let $f : \{0,1\}^n \to \{+1,-1\}$ be a Boolean function. Then, $\mathbb{H}(f) \leq \frac{1}{\ln 2}\mathsf{Inf}[f] + \frac{1}{\ln 2}\sum_{k=1}^n W^k[f] k \ln \frac{n}{k} + 3 \cdot \mathsf{Inf}[f]$.*

Using Lemma 3 we prove a technical lemma that provides a bound on $\sum_{k=1}^n W^k[f] k \ln \frac{n}{k}$.

Lemma 5. *Let $f : \{0,1\}^n \to \{+1,-1\}$ be a Boolean function. Then, $\sum_{k=1}^n W^k[f] k \ln \frac{n}{k} \leq exp(1/2) \cdot \frac{3}{1-e^{-2}} \cdot \frac{4^{1-\beta}}{(1-\beta)^2} \cdot \alpha \cdot n^{1-\beta}$.*

Using Lemma 5 and Lemma 4 we obtain the following theorem which bounds the Fourier entropy of a Boolean function.

Theorem 4. *Let $f : \{0,1\}^n \to \{+1,-1\}$ be a Boolean function such that $\mathsf{NS}_\epsilon(f) \leq \alpha \cdot \epsilon^\beta$. Then*

$$\mathbb{H}(f) \leq C \cdot \left(\mathsf{Inf}[f] + \frac{4^{1-\beta}}{(1-\beta)^2} \cdot \alpha \cdot n^{1-\beta} \right) ,$$

where C is a universal constant.

In particular, for polynomial threshold functions there exist non-trivial bounds on their noise sensitivity.

Theorem 5 (Peres's Theorem). *[13] Let $f : \{0,1\}^n \to \{+1,-1\}$ be a linear threshold function. Then $\mathsf{NS}_\epsilon(f) \leq O(\sqrt{\epsilon})$.*

Theorem 6. *[8] For any degree-d polynomial threshold function $f : \{0,1\}^n \to \{+1,-1\}$ and $0 < \epsilon < 1$, $\mathsf{NS}_\epsilon(f) \leq 2^{O(d)} \cdot \epsilon^{1/(4d+6)}$.*

As corollaries of Theorem 4, using Theorem 5 and Theorem 6, we obtain the following bounds on the Fourier entropy of polynomial threshold functions.

Corollary 1. *Let $f : \{0,1\}^n \to \{+1,-1\}$ be a linear threshold function. Then, $\mathbb{H}(f) \leq C \cdot \sqrt{n}$, where C is a universal constant.*

Corollary 2. *Let $f : \{0,1\}^n \to \{+1,-1\}$ be a degree-d polynomial threshold function. Then, $\mathbb{H}(f) \leq C \cdot 2^{O(d)} \cdot n^{1-\frac{1}{4d+6}}$, where C is a universal constant.*

5 Entropy-Influence Inequality for Read-Once Formulas

In this section, we will prove the Fourier Entropy-Influence conjecture for read-once formulas using AND, OR, XOR, and NOT gates. We note that a recent (and independent) result of O'Donnell and Tan [15] proves the conjecture for read-once formulas with *arbitrary* gates of bounded fan-in. But since our proof is completely elementary and very different from theirs, we choose to present it here.

It is well-known that both Fourier entropy and average sensitivity add up when two functions on disjoint sets of variables are added modulo 2. Our main result here is to show that somewhat analogous "tensorizability" properties hold when composing functions on disjoint sets of variables using AND and OR operations.

For $f : \{0,1\}^n \to \{+1,-1\}$, let $f_{\mathbb{B}}$ denote its 0-1 counterpart: $f_{\mathbb{B}} \equiv \frac{1-f}{2}$.

$$\text{Let's define:} \qquad \mathbf{H}(f_{\mathbb{B}}) := \sum_S \widehat{f_{\mathbb{B}}}^2(S) \log \frac{1}{\widehat{f_{\mathbb{B}}}^2(S)}. \tag{9}$$

An easy relation enables translation between $\mathbb{H}(f)$ and $\mathbf{H}(f_{\mathbb{B}})$:

Lemma 6. *Let $p = \Pr[f_{\mathbb{B}} = 1] = \widehat{f_{\mathbb{B}}}(\emptyset) = \sum_S \widehat{f_{\mathbb{B}}}^2(S)$ and $q := 1 - p$. Then,*

$$\mathbb{H}(f) = 4 \cdot \mathbf{H}(f_{\mathbb{B}}) + \varphi(p), \quad \textit{where} \tag{10}$$

$$\varphi(p) := \mathrm{H}(4pq) - 4p(\mathrm{H}(p) - \log p). \tag{11}$$

$$\text{For } 0 \le p \le 1, \text{ let's also define:} \qquad \psi(p) := p^2 \log \frac{1}{p^2} - 2\,\mathrm{H}(p). \tag{12}$$

▶ **Intuition:** Before going on, we pause to give some intuition about the choice of the function ψ and the function κ below (15). In the FEI conjecture (2), the right hand side, $\mathsf{Inf}(f)$, does not depend on whether we take the range of f to be $\{-1,+1\}$ or $\{0,1\}$. In contrast, the left hand side, $\mathbb{H}(f)$, depends on the range being $\{-1,+1\}$. Just as the usual entropy-influence inequality composes w.r.t. the parity operation (over disjoint variables) with $\{-1,+1\}$ range, we expect a corresponding inequality with $\{0,1\}$ range to hold for the AND operation (and by symmetry for the OR operation). However, Lemma 6 shows the translation to $\{0,1\}$-valued functions results in the annoying additive "error" term $\varphi(p)$. Such additive terms that depend on p create technical difficulties in the inductive proofs below and we need to choose the appropriate functions of p carefully.

For example, we know $4\,\mathbf{H}(f_{\mathbb{B}}) + \varphi(p) = \mathbb{H}(f) = 4\,\mathbf{H}(1 - f_{\mathbb{B}}) + \varphi(q)$ from Lemma 6. If the conjectured inequality for the $\{0,1\}$-valued entropy-influence inequality has an additive error term $\psi(p)$ (see (13) below), then we must have $\mathbf{H}(f_{\mathbb{B}}) - \mathbf{H}(1 - f_{\mathbb{B}}) = \psi(p) - \psi(q) = (\varphi(q) - \varphi(p))/4 = p^2 \log \frac{1}{p^2} - q^2 \log \frac{1}{q^2}$, using (11). Hence, we may conjecture that $\psi(p) = p^2 \log \frac{1}{p^2} +$ (an additive term symmetric w.r.t. p and q). Given this and the other required properties, e.g., Lemma 7 below, for the composition to go through, lead us to the definition of ψ in (12). Similar considerations w.r.t. composition by parity operation (in addition to those by AND, OR, and NOT) leads us to the definition of κ in (15). ◀

Let us define the **FEI01 Inequality** (the 0-1 version of FEI) as follows:

$$\mathbf{H}(f_{\mathbb{B}}) \leq c \cdot \mathsf{as}(f) + \psi(p), \tag{13}$$

where $p = \widehat{f_{\mathbb{B}}}(\emptyset) = \Pr_x[f_{\mathbb{B}}(x) = 1]$ and c is a constant to be fixed later.

The following technical lemma gives us the crucial property of ψ:

Lemma 7. *For ψ as above and $p_1, p_2 \in [0,1]$, $p_1 \cdot \psi(p_2) + p_2 \cdot \psi(p_1) \leq \psi(p_1 p_2)$.*

Given this lemma, an inductive proof yields our theorem for read-once formulas over the complete basis of $\{\mathsf{AND}, \mathsf{OR}, \mathsf{NOT}\}$.

Theorem 7. *The FEI01 inequality* (13) *holds for all read-once Boolean formulas using* AND*,* OR*, and* NOT *gates, with constant $c = 5/2$.*

To switch to the usual FEI inequality (in the $\{-1,+1\}$ notation), we combine (13) and (10) to obtain

$$\mathbb{H}(f) \leq 10 \cdot \mathsf{as}(f) + \kappa(p), \quad \text{where} \tag{14}$$

$$\kappa(p) := 4\psi(p) + \varphi(p) = -8\,\mathrm{H}(p) - 8pq - (1 - 4pq)\log(1 - 4pq). \tag{15}$$

Since it uses the $\{-1,+1\}$ range, we expect that (14) should be preserved by parity composition of functions. The only technical detail is to show that the function κ also behaves well w.r.t. parity composition. We show that this indeed happens. This leads us to the main theorem of this section:

Theorem 8. *If f is computed by a read-once formula using* AND*,* OR*,* XOR*, and* NOT *gates, then $\mathbb{H}(f) \leq 10\,\mathsf{Inf}(f) + \kappa(p)$.*

Remark 2. The parity function on n variables shows that the bound in Theorem 8 is tight; it is not tight without the additive term $\kappa(p)$. It is easy to verify that $-10 \leq \kappa(p) \leq 0$ for $p \in [0,1]$. Hence the theorem implies $\mathbb{H}(f) \leq 10\,\mathsf{Inf}(f)$ for all read-once formulas f using AND, OR, XOR, and NOT gates.

6 A Bound for Real valued Functions via Second Moment

Due to space constraints we only state the theorem here, the full proof appears in [2].

Theorem 9. *If $f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S$ is a real-valued function on the domain $\{0,1\}^n$ such that $\sum_S |\hat{f}(S)^2| = 1$ then for any $\delta > 0$,*

$$\sum_{S \subseteq [n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) = \sum_S |S|^{1+\delta} \hat{f}(S)^2 + 2\log_{1+\delta} n + 2(2\log n)^{1+\delta/\delta}(\log n)^2 \ .$$

As a corollary to Theorem 9, we also obtain the bound (7) in terms of the first and second moments of sensitivities of a function.

Acknowledgements. N. Saurabh thanks Ryan O'Donnell for helpful discussions on Lemma 3.

References

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

[2] Sourav Chakraborty, Raghav Kulkarni, Satyanarayana V. Lokam, and Nitin Saurabh. Upper bounds on fourier entropy. Technical report, Electronic Colloquium on Computational Complexity (ECCC), TR13-052, 2013.

[3] Bireswar Das, Manjish Pal, and Vijay Visavaliya. The entropy influence conjecture revisited. Technical report, arXiv:1110.4301, 2011.

[4] Ronald de Wolf. A brief introduction to fourier analysis on the boolean cube. *Theory of Computing, Graduate Surveys*, 1:1–20, 2008.

[5] I. Diakonikolas, P. Raghavendra, R. Servedio, and L. Tan. Average sensitivity and noise sensitivity of polynomial threshold functions. *SIAM Journal on Computing*, 43(1):231–253, 2014.

[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 Tauman Kalai, and Adam R. Klivans. Agnostically learning decision trees. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 527–536, 2008.

[8] Prahladh Harsha, Adam Klivans, and Raghu Meka. Bounding the sensitivity of polynomial threshold functions. *Theory of Computing*, 10(1):1–26, 2014.

[9] 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.

[10] Nathan Keller, Elchanan Mossel, and Tomer Schlank. A note on the entropy/influence conjecture. Technical report, arXiv:1105.2651, 2011.

[11] Adam Klivans, Homin Lee, and Andrew Wan. Mansour's conjecture is true for random dnf formulas. In *Proceedings of the 23rd Conference on Learning Theory*, pages 368–380, 2010.

[12] 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.

[13] Ryan O'Donnell. *Computational applications of noise sensitivity.* PhD thesis, MIT, 2003.

[14] Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.

[15] Ryan O'Donnell and Li-Yang Tan. A composition theorem for the fourier entropy-influence conjecture. In *Proceedings of Automata, Languages and Programming - 40th International Colloquium*, 2013.

[16] Ryan O'Donnell, John Wright, and Yuan Zhou. The fourier entropy-influence conjecture for certain classes of boolean functions. In *Proceedings of Automata, Languages and Programming - 38th International Colloquium*, pages 330–341, 2011.

[17] Andrew Wan, John Wright, and Chenggang Wu. Decision trees, protocols and the entropy-influence conjecture. In *Innovations in Theoretical Computer Science, ITCS'14*, pages 67–80, 2014.