A Universal Lower-Bound Technique for the Size of Decision Trees and Two-Level AND/OR Circuits

Yigal Brandman[1], *Alon Orlitsky*[2], *and John Hennessy*[1]

Abstract

A universal lower-bound technique for the size and other implementation characteristics of an arbitrary Boolean function as a decision tree and as a two-level AND/OR circuit is derived. The technique is based on the power spectrum coefficients of the n-dimensional Fourier transform of the function. The bounds vary from constant to exponential and are tight in many cases. Several examples are presented.

1 Introduction

The *two-level AND/OR* implementation of a Boolean function f (also called *sum of products*, *conjunctive normal form*, or *single-output PLA*) is a logic circuit in which all the input variables or their complements are connected to AND gates; the outputs of all these AND gates are connected to a single OR gate whose output is the value of f.

A *decision tree* is a rooted binary tree where each internal node (including the root) is labeled by a variable, and each leaf and each edge is labeled by a Boolean value (0 or 1). One of the two edges emanating from each internal node is labeled 0 and the other is labeled 1. Given an input (i.e., an assignment of Boolean values to the variables labeling the internal nodes of the tree), there is exactly one path $v_1, \ldots, v_k$ such that v_1 is the root, v_k is a leaf, and for $i = 1, \ldots, k-1$, the label of the edge (v_i, v_{i+1}) is the value assigned to the variable labeling v_i. The computed value of the tree for that input, is the Boolean label of the leaf v_k. A *decision tree for f* is one whose computed value is f for every input.

In this paper we establish a relationship between the *power spectrum* coefficients of the *n-dimensional Fourier transform* of f and several characteristics of its covers (these terms

[1]Department of Electrical Engineering, Stanford University, Stanford, CA 94305
[2]AT&T Bell Laboratories, Murray Hill, NJ 07974

are defined in the next section). Using this relationship, we derive a universal technique that, for any Boolean function f, lower bounds:

1) The number of leaves in any decision tree for f.

2) The average number of decisions in any decision tree for f.

3) The number of AND gates in any AND/OR circuit that computes f.

The first two bounds are derived in Section 3, the third is proven in Section 4.

Spectral techniques in digital logic were investigated by many. Lechner [1] used the Fourier transform to classify symmetries of boolean functions. Karpovsky [2] suggested using the number of nonvanishing coefficients in the Fourier transform of a function as a measure for its complexity. Hurst, Miller, and Muzio [3] related the circuit complexity of a function to its power spectrum coefficients. A review of spectral techniques in logic synthesis and analysis can be found in Karpovsky [4] and Hurst, Miller, and Muzio [5]. An extended view of the results presented in this paper appears in [6].

The results presented here are theoretical yet they might be applicable to logic optimization. Consider for example the following heuristic algorithm for generating a small (few leaves) decision tree for a function f. To decide which variable will label the root consider, for each variable x_i, the restricted functions: $f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n)$ and $f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$. Compute the lower bound of Corollary 2 for both restrictions and add them together. Choose the variable that minimizes this sum. Once you determine which variable labels the root, continue recursively with each subtree. This algorithm has time complexity polynomial in 2^n, the number of bits needed to describe f. It would be interesting to compare it with other techniques for optimizing various decision trees some of which described in [7].

2 Definitions

2.1 Covers

A *cube* in $\{0,1\}^n$ (a *cube* when n is understood) is a Cartesian product $\prod_{i=1}^{n} \pi_i$ where each π_i is either $\{0\}$, $\{1\}$, or $\{0,1\}$. Any cube C has a unique representation as such a product;

let $\pi_i(C)$ denote the i'th set in the product and call it the i'th *projection* of C.

Let S be a subset of $\{0,1\}^n$. A set $\mathcal{C}$ of cubes is a *cover over* S if $\bigcup\{C \in \mathcal{C}\} = S$. It is a *nonoverlapping cover over* S if, in addition, the cubes in $\mathcal{C}$ are disjoint: $C_1 \cap C_2 = \emptyset$ for all distinct $C_1, C_2 \in \mathcal{C}$.

Let $f : \{0,1\}^n \rightarrow \{-1, 0, 1\}$. A cube C is *f-constant* if f is constant over it: $f(\overline{x}) = f(\overline{y})$ for all $\overline{x}, \overline{y} \in C$. A (nonoverlapping) cover $\mathcal{C}$ over a subset $S \subseteq \{0,1\}^n$ is a *(nonoverlapping) cover for f over S* if every cube in $\mathcal{C}$ is f-constant. The set

$$S_f \stackrel{\text{def}}{=} \{\overline{x} \in \{0,1\}^n : f(\overline{x}) \neq 0\}$$

is the *support set* of f. A cover for f over S_f is also called a *cover for f*. The technique presented in this paper provides bounds on covers for f. By setting the value of f to 0 for certain inputs, it is possible to derive bounds on covers for f over any subset of $\{0,1\}^n$. This option, which is used in the sequel, is the reason for choosing $\{-1, 0, 1\}$ as the range of f instead of the standard $\{0,1\}$.

Example 1 Let $\overline{x} = (x_1, \ldots, x_n)$ and define the *OR* function:

$$f(\overline{x}) \stackrel{\text{def}}{=} x_1 \vee x_2 \vee \ldots \vee x_n = \begin{cases} 1 & \text{if } x_i = 1 \text{ for some } i \in \{1,\ldots,n\}, \\ 0 & \text{if } x_i = 0 \text{ for all } i \in \{1,\ldots,n\}. \end{cases}$$

The support set of f is $\{0,1\}^n - \{(0,\ldots,0)\}$. A cube is f-constant if and only if at least one of its projections is $\{1\}$ (then $f(\overline{x}) = 1$ for all $\overline{x}$ in the cube), or if it is the cube $\{(0,\ldots,0)\}$. For $j = 1,\ldots,n$ let C_j and C'_j be the cubes defined by:

$$\pi_i(C_j) \stackrel{\text{def}}{=} \begin{cases} \{1\} & \text{if } i = j, \\ \{0,1\} & \text{if } i \neq j, \end{cases} \qquad \text{and} \qquad \pi_i(C'_j) \stackrel{\text{def}}{=} \begin{cases} \{0\} & \text{if } i < j, \\ \{1\} & \text{if } i = j, \\ \{0,1\} & \text{if } i > j. \end{cases}$$

Then both $\mathcal{C} = \{C_1, \ldots, C_n\}$ and $\mathcal{C}' = \{C'_1, \ldots, C'_n\}$ are covers for f but only $\mathcal{C}'$ is a nonoverlapping cover. $\square$

The minimum, over all covers for f, of the number of cubes in the cover is the *cover size* of f, denoted $C(f)$. The least number of nonoverlapping cubes needed to cover f is $C_N(f)$, the *nonoverlapping cover size* of f. Clearly,

$$C_N(f) \geq C(f).$$

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. It is well known that each AND gate in an AND/OR implementation of f corresponds to an f-constant cube in $\{0,1\}^n$ and these cubes cover f over its support set. Hence $C(f)$ is the minimum number of AND gates in any AND/OR implementation of f. Also, if $f : \{0,1\}^n \to \{-1,1\}$, then each leaf in a decision tree for f corresponds to an f-constant cube in $\{0,1\}^n$ and these cubes form a nonoverlapping cover for f over its support set. Hence $C_N(f)$ lower bounds the number of leaves in any decision tree for f.

The *size* $|C|$ of a cube C is the number of elements in the cube. The *dimension* of C is

$$\dim(C) \stackrel{\text{def}}{=} \log|C| = |\{i : \pi_i(C) = \{0,1\}\}|,$$

the number of its projections that are $\{0,1\}$. The *information content* of C is

$$\text{info}(C) \stackrel{\text{def}}{=} n - \dim(C),$$

the number of fixed coordinates in the cube; it is the "amount of information" about x provided by knowing that $x \in C$. A measure for covers that takes into account the relative sizes of the cubes is the *entropy*. Let $\mathcal{C}$ be a cover over a set $S \subseteq \{0,1\}^n$ and let $\overline{x} \in S$. Define

$$\text{info}_{\mathcal{C}}(\overline{x}) \stackrel{\text{def}}{=} \max\{\text{info}(C) : \overline{x} \in C \in \mathcal{C}\}$$

to be the information content of the smallest cube in $\mathcal{C}$ that contains $\overline{x}$. The *entropy* of the cover $\mathcal{C}$ is:

$$H(\mathcal{C}) \stackrel{\text{def}}{=} \frac{1}{2^n} \sum_{\overline{x} \in S} \text{info}_{\mathcal{C}}(\overline{x}).$$

For a nonoverlapping cover $\mathcal{C}$, summation over inputs can be substituted by summation over cubes, yielding:

$$H(\mathcal{C}) = \frac{1}{2^n} \sum_{C \in \mathcal{C}} |C| \cdot \text{info}(C).$$

This formulation explains why $H(\mathcal{C})$ is called the entropy of $\mathcal{C}$: Suppose that $\overline{x} \in \{0,1\}^n$ is chosen uniformly at random. The probability $P(C)$ that $\overline{x}$ belongs to a cube C is $|C|/2^n$. Hence

$$\text{info}(C) = n - \dim(C) = \log\left(\frac{2^n}{|C|}\right) = \log\left(\frac{1}{P(C)}\right)$$

and

$$H(\mathcal{C}) = \sum_{C\in\mathcal{C}} P(C)\cdot\log\left(\frac{1}{P(C)}\right).$$

is the standard Shannon entropy of $\mathcal{C}$. The *cover entropy* of f,

$$H(f) \stackrel{\text{def}}{=} \min\{H(\mathcal{C}) : \mathcal{C} \text{ is a cover for } f\},$$

is the minimum entropy of all (possibly overlapping) covers for f. The *nonoverlapping entropy* of f, denoted $H_N(f)$, is the minimum entropy of all nonoverlapping covers for f.

It can be easily shown that if a decision tree induces a cover $\mathcal{C}$ (necessarily nonoverlapping), then $\text{info}_{\mathcal{C}}(\overline{x})$ decisions are made to determine the value of f for the input $\overline{x}$. Therefore, $H_N(f)$ is a lower bound on the average number of decisions made in any decision tree for f. In Subsection 4.3 we use $H(f)$ to lower bound $C(f)$, the minimal number of AND gates in an AND/OR implementation of f.

Example 2 In Example 1, $\mathcal{C}'$ contains n cubes and

$$H(\mathcal{C}') = \frac{1}{2^n}\sum_{i=1}^{n} i2^{n-i} = 2 - \frac{n}{2^n} - \frac{2}{2^n}.$$

We will see in example 5 that $C_N(f) = n$ and $H_N(f) = 2 - \frac{n}{2^n} - \frac{2}{2^n}$. Thus, the cover mentioned is optimal in both respects. □

The techniques derived in this paper lower bound $C_N(f)$, $C(f)$, $H_N(f)$, and $H(f)$. They do so via the n-dimensional Fourier transform defined next.

2.2 The n-Dimensional Fourier Transform

One way to describe the *n-dimensional Fourier transform over* $\{0,1\}^n$ (also known as the *Hadamard* or *Walsh transform*) is (cf., [8]) via its basis functions. For each $\overline{X} = (X_1,\ldots,X_n) \in \{0,1\}^n$, the $\overline{X}$'th *basis function* is the *Walsh function*

$$W_{\overline{X}}(\overline{x}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2^n}}\Big(-1\Big)^{\sum_{i=1}^{n} X_i x_i}.$$

Let

$$\text{ones}(\overline{X}) \stackrel{\text{def}}{=} \sum_{i=0}^{n} X_i$$

denote the*Hamming weight* of $\overline{X}$: the number of components X_i in $\overline{X}$ that are 1, and let

$$\text{zer}(\overline{X}) \stackrel{\text{def}}{=} n - \text{ones}(\overline{X})$$

denote the number of zero components. Then $W_{\overline{X}}(\overline{x})$ partitions $\{0,1\}^n$ into $2^{\text{ones}(\overline{X})}$ maximal $W_{\overline{X}}$-constant cubes. Each cube is of size $2^{\text{zer}(\overline{X})}$ and $W_{\overline{X}}$ is constant over it: either $+1/\sqrt{2^n}$ or $-1/\sqrt{2^n}$. The Fourier transform $\mathcal{F}(f)$ of f is defined as

$$[\mathcal{F}(f)](\overline{X}) \stackrel{\text{def}}{=} \sum_{\overline{x}\in\{0,1\}^n} f(\overline{x}) W_{\overline{X}}(\overline{x}).$$

Used as the argument of the transform $\mathcal{F}(f)$, the variable $\overline{X}$ is known as *frequency* or *sequency.* Some authors use ω instead of $\overline{X}$; we chose $\overline{X}$ to emphasize the symmetry between the function and transform domains.

Two properties of the transform will be extensively used in this paper:

1) Linearity:

$$\mathcal{F}(f+g) = \mathcal{F}(f) + \mathcal{F}(g).$$

2) Parseval's Theorem:

$$\sum_{\overline{x}\in\{0,1\}^n} f^2(\overline{x}) = \sum_{\overline{X}\in\{0,1\}^n} [\mathcal{F}(f)]^2(\overline{X}).$$

The function $[\mathcal{F}(f)]^2$ is called the *power spectrum* of f and in the rest of the paper it is related to the cover sizes and entropies of f.

3 Nonoverlapping Covers

In this section we consider nonoverlapping covers. In Subsection 3.1 we derive a family of inequalities relating the Fourier transform of f and the sizes of cubes in any nonoverlapping cover for f. In subsection 3.2 we use two of these inequalities to derive lower bounds on $C_N(f)$ and $H_N(f)$. Throughout this section we assume that $f : \{0,1\}^n \to \{-1,0,1\}$ and that $\mathcal{C}$ is a nonoverlapping cover for f (over its support set).

3.1 Relating the Fourier Transform and Nonoverlapping Covers

If $\overline{X}$ is such that $X_i = 1$ then the definition of the Fourier transform implies that $W_{\overline{X}}$ changes along the i'th dimension: $W_{\overline{X}}(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n) = -W_{\overline{X}}(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$

for all $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. If $\pi_i(C)$, the i'th projection of an f-constant cube C, is $\{0,1\}$ then, over C, f is constant along the i'th dimension. Hence, $f(\overline{x})W_{\overline{X}}(\overline{x})$ is antisymmetric over C with respect to the i'th dimension and in the sum $\sum_{\overline{x}\in C} f(\overline{x})W_{\overline{X}}(\overline{x})$, every term cancels out with one equal in magnitude but of opposite sign. Thus:

Lemma 1 For all $\overline{X} \in \{0,1\}^n$ and all $C \in \mathcal{C}$,

$$\{i : X_i = 1\} \bigcap \{i : \pi_i(C) = \{0,1\}\} \neq \emptyset \quad \Rightarrow \quad \sum_{\overline{x}\in C} f(\overline{x})W_{\overline{X}}(\overline{x}) = 0. \quad \square$$

The size of a cube C is $2^{\dim(C)}$ and its intersection with *each* of the cubes induced by $W_{\overline{X}}$ is a cube of size 2^k for some $k \leq \mathrm{zer}(\overline{X})$. Therefore, if $\dim(C) > \mathrm{zer}(\overline{X})$, then $2^{\dim(C)-k}$ such intersections fit in C. By the structure of $W_{\overline{X}}$, half of them are 'plus' and half are 'minus'. Yielding:

Corollary 1 For any $C \in \mathcal{C}$ and $\overline{X} \in \{0,1\}^n$,

$$\mathrm{info}(C) < \mathrm{ones}(\overline{X}) \rightarrow \sum_{\overline{x}\in C} f(\overline{x})W_{\overline{X}}(\overline{x}) = 0.$$

Proof: $|\{i : X_i = 1\}| = \mathrm{ones}(\overline{X})$ while $|\{i : \pi_i(C) = \{0,1\}\}| = n - \mathrm{info}(C) > n - \mathrm{ones}(\overline{X})$. Hence, $\{i : X_i = 1\} \bigcap \{i : \pi_i(C) = \{0,1\}\} \neq \emptyset$. $\square$

This suggests two simple hierarchies on the vectors in $\{0,1\}^n$. The first hierarchy is based on the Hamming weight of the vector. Define the k'th *band* of $\{0,1\}^n$ to be

$$B_k \stackrel{\text{def}}{=} \{\overline{x} : \mathrm{ones}(\overline{x}) = k\},$$

the set of vectors in $\{0,1\}^n$ lying on the boundary of the Hamming sphere of radius k. The second hierarchy is cover-dependent and is based on the information content of the cube containing the vector:

$$\mathcal{C}_i \stackrel{\text{def}}{=} \bigcup \{C \in \mathcal{C} : \mathrm{info}(C) = i\}.$$

Recall that throughout this section $\mathcal{C}$ is a nonoverlapping cover for f hence the sets $\mathcal{C}_i$ are always disjoint. Note also that $\mathcal{C}_0$ is either $\emptyset$ or $\{0,1\}^n$ whereas $\mathcal{C}_1$ is either $\emptyset$ or $\{0,1\}^n$ or one of the n "half spaces" and $\mathcal{C}_n$ can be any subset of $\{0,1\}^n$. Define

$$\mathcal{C}_k^l \stackrel{\text{def}}{=} \begin{cases} \bigcup_{i=k}^{l} \mathcal{C}_i & \text{if } 0 \leq k \leq l \leq n, \\ \emptyset & \text{otherwise.} \end{cases}$$

to be the set of inputs contained in cubes of size at least 2^{n-l} and at most 2^{n-k}. Corollary 1 said that only cubes in $\mathcal{C}_k^n$ contribute to the k'th-band coefficients of the transform. To further explore this relationship, we look at the restriction of f to sets of cubes.

If S is a subset of $\{0,1\}^n$, let

$$f|_S(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} f(x) & \text{if } \overline{x} \in S \\ 0 & \text{if } \overline{x} \notin S. \end{cases}$$

Thus, $f|_{\mathcal{C}_0^n}$ is f and $f|_{\mathcal{C}_k^n}$ is f restricted to the cubes of $\mathcal{C}$ whose information content is at least k. For every $k \in \{0, \ldots, n\}$,

$$f(\overline{x}) = f|_{\mathcal{C}_k^n}(\overline{x}) + f|_{\mathcal{C}_0^{k-1}}(\overline{x}).$$

It follows from Corollary 1 that if $\text{ones}(\overline{X}) \geq k$ then $[\mathcal{F}(f|_{\mathcal{C}_0^{k-1}})](\overline{X}) = 0$. Using the linearity of the transform, we get

$$\text{ones}(\overline{X}) \geq k \Rightarrow [\mathcal{F}(f)](\overline{X}) = [\mathcal{F}(f|_{\mathcal{C}_k^n})](\overline{X}). \tag{1}$$

This is the first clue for the bound. It enables us to know $[\mathcal{F}(f|_{\mathcal{C}_k^n})](\overline{X})$ for $\overline{X}$'s with at least k ones without knowing what $\mathcal{C}_k^n$ is. In turn, $\mathcal{C}_k^n$ relates to the number of cubes in $\mathcal{C}$ because all cubes in $\mathcal{C}_k^n$ are of size at most 2^{n-k}. Next, we need to relate the Fourier coefficients in $B_k, \ldots, B_n$ to the size of $\mathcal{C}_k^n$. Because of cancellations, we do so via the *power spectrum function* of f defined by:

$$[\mathcal{E}(f)](\overline{X}) = [\mathcal{F}(f)]^2(\overline{X}).$$

The *energy* in the i'th band is

$$E_i(f) \stackrel{\text{def}}{=} \sum_{\overline{X} \in B_i} [\mathcal{E}(f)](\overline{X}).$$

Squaring Equation (1) and summing over all vectors in B_i yields

Lemma 2 Let $0 \leq k \leq n$. For all $k \leq i \leq n$,

$$E_i(f) = E_i(f|_{\mathcal{C}_k^n}). \quad \square$$

Parseval's theorem, can now be used to relate the energies in $B_k, \ldots, B_n$ to the size of $\mathcal{C}_k^n$:

Lemma 3 For every $0 \leq k \leq n$,

$$\sum_{i=k}^{n} |\mathcal{C}_i| \geq \sum_{i=k}^{n} E_i(f)$$

with equality for $k = 0$.

Proof: Consider $f|_{\mathcal{C}_k^n}$. By Lemma 2, $E_i(f|_{\mathcal{C}_k^n}) = E_i(f)$ for all $i \geq k$. Hence,

$$\begin{aligned} \sum_{i=k}^{n} E_i(f) &= \sum_{i=k}^{n} E_i(f|_{\mathcal{C}_k^n}) \leq \sum_{i=0}^{n} E_i(f|_{\mathcal{C}_k^n}) \\ &= \sum_{\overline{x} \in \{0,1\}^n} (f|_{\mathcal{C}_k^n}(\overline{x}))^2 = \sum_{\overline{x} \in \mathcal{C}_k^n} 1 = |\mathcal{C}_k^n| = \sum_{i=k}^{n} |\mathcal{C}_i| \end{aligned}$$

with equality for $k = 0$. $\square$

Spelled out, the lemma reads as follows:

$$\begin{aligned} |\mathcal{C}_n| &\geq E_n(f) \\ |\mathcal{C}_{n-1}| + |\mathcal{C}_n| &\geq E_n(f) + E_{n-1}(f) \\ &\vdots \\ |\mathcal{C}_0| + \cdots + |\mathcal{C}_{n-1}| + |\mathcal{C}_n| &= E_n(f) + E_{n-1}(f) + \cdots + E_0(f). \end{aligned} \tag{2}$$

Multiplying each equation by some constant (positive for the inequalities) and adding them up we get:

Theorem 1 Let $\{a_i\}_{i=0}^{n}$ be a nondecreasing sequence of reals (not necessarily positive). Then, for all covers $\mathcal{C}$ of f,

$$\sum_{i=0}^{n} a_i |\mathcal{C}_i| \geq \sum_{i=0}^{n} a_i E_i(f). \quad \square$$

3.2 Lower Bounds on Nonoverlapping Covers

Choosing different values for the a_i's it is possible to derive various inequalities. Here, we demonstrate two: one involves $C_N(f)$ which is a lower bound on the number of leaves in any decision tree for f. The other, involves $H_N(f)$, a lower bound on the average number of decisions made in any decision tree for f.

Corollary 2 **(Lower bound on the nonoverlapping cover size of f)**

$$C_N(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} 2^i E_i(f).$$

Proof: By the last Theorem, for all covers $\mathcal{C}$ of f,

$$|\mathcal{C}| = \sum_{i=0}^{n} \frac{|\mathcal{C}_i|}{2^{n-i}} \geq \sum_{i=0}^{n} \frac{E_i(f)}{2^{n-i}}$$

and $C_N(f)$ is the minimum of all these $|\mathcal{C}|$'s. $\square$

A similar proof will show that:

Corollary 3 **(Lower bound on the nonoverlapping entropy of f)**

$$H_N(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} i E_i(f). \quad \square$$

At the beginning of the section we assumed that the range of f was $\{-1,0,1\}$ and that the covers where over the support set of f, but both bounds can be applied to covers over *any subset* of $\{0,1\}^n$. All that needs to be done is redefine the values of f to be $+1$ or -1 inside the subset (by letting $f(\overline{x}) = -1$ wherever it was 0) and identically 0 outside the subset. In particular, if a cover is needed for the 0's then f should be redefined to be -1 wherever it was previously 0 and to be 0 wherever it was 1. In the next example we want a cover for both the 0's and the 1's so we redefine f to have the range $\{-1,1\}$.

Example 3 For even n let

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_1x_2 \oplus x_3x_4 \oplus \cdots \oplus x_{n-1}x_n = 1 \\ -1 & \text{if } x_1x_2 \oplus x_3x_4 \oplus \cdots \oplus x_{n-1}x_n = 0. \end{cases}$$

From the Hadamard structure of the function, we get (cf., [2, 9]):

$$[\mathcal{F}(f)]^2(\overline{X}) = 1 \text{ for all } \overline{X} \in \{0,1\}^n.$$

Hence, $E_i(f) = \binom{n}{i}$ and,

$$C_N(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} 2^i \binom{n}{i} = \left(\frac{3}{2}\right)^n,$$

$$H_N(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} i \binom{n}{i} = \frac{1}{2} n.$$

In fact, the optimal nonoverlapping cover in terms of both size and entropy has

$$|\mathcal{C}_i| = 2^{n/2} \binom{n/2}{n-i}.$$

Thus,

$$C_N(f) = \frac{1}{2^{n/2}} \sum_{i=n/2}^{n} 2^i \binom{n/2}{n-i} = 2^{n/2} \sum_{i=0}^{n/2} \binom{n/2}{i} / 2^i = (\sqrt{3})^n$$

and,

$$H_N(f) = \frac{1}{2^{n/2}} \sum_{i=n/2}^{n} i \binom{n/2}{n-i} = \frac{1}{2^{n/2}} \sum_{i=0}^{n/2} \binom{n/2}{i} (n-i) = \frac{3}{4} n. \quad \square$$

This function has the remarkable property that its power spectrum is flat: $[\mathcal{F}(f)]^2(\overline{X}) = 1$ for all $\overline{X}$. Such functions are known (cf., [9]) as *bent* functions and the same lower bound applies to all of them. For some functions, though, we can get an even better bound:

Example 4 The majority function. For odd n, define:

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} -1 & \text{if } \sum_{i=1}^{n} x_i > n/2 \\ 1 & \text{if } \sum_{i=1}^{n} x_i < n/2. \end{cases}$$

Then (cf., [2]), for all $\overline{X} \in B_i$,

$$[\mathcal{F}(f)](\overline{X}) = \begin{cases} \frac{2}{\sqrt{2^n}} (-1)^{\frac{i-1}{2}} \frac{\binom{i}{i/2}\binom{n-i-1}{\frac{n-i-1}{2}}}{\binom{\frac{n-1}{2}}{i/2}} & \text{if } i \text{ is odd} \\ 0 & \text{if } i \text{ is even.} \end{cases}$$

Note that for every i, the transform has the same absolute value in all entries of the i'th and $(n-i-1)$'th bands. Hence, for the entropy, we get:

$$H_N(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} i E_i = \frac{4}{2^{2n}} \sum_{i=0,2,\ldots,n-1} i \binom{n}{i} \left(\frac{\binom{i}{i/2}\binom{n-i-1}{\frac{n-i-1}{2}}}{\binom{\frac{n-1}{2}}{i/2}} \right)^2$$

$$= \frac{4n}{2^{2n}} \binom{n-1}{\frac{n-1}{2}} \sum_{i=0,2,\ldots,n-1} \binom{i}{i/2} \binom{n-1-i}{\frac{n-1-i}{2}} = \frac{4n}{2^{2n}} \binom{n-1}{\frac{n-1}{2}} 2^{n-1} \geq \sqrt{\frac{n}{2}}.$$

As for the cover size, considering E_n alone, the lower bound yields:

$$C_N(f) \geq E_n(f) = \frac{1}{2^n} \left[2 \binom{n-1}{\frac{n-1}{2}} \right]^2 \approx \frac{2^n}{n}. \quad \square$$

3.3 A slight improvement of the bound

Consider set (2) of equations once again. Subtracting each equation from the last yields:

$$\begin{aligned} |\mathcal{C}_0| &\leq E_0(f) \\ |\mathcal{C}_1| + |\mathcal{C}_0| &\leq E_0(f) + E_1(f) \\ &\vdots \\ |\mathcal{C}_n| + \cdots + |\mathcal{C}_1| + |\mathcal{C}_0| &= E_0(f) + E_1(f) + \cdots + E_n(f). \end{aligned}$$

Each $\mathcal{C}_i$ is consists of cubes of dimension $n-i$. Therefore, $|\mathcal{C}_i|$ is a multiple of 2^{n-i}. The above set of equations can thus be rewritten as:

$$\begin{aligned}
|\mathcal{C}_0| &\leq 2^n \lfloor E_0(f)/2^n \rfloor \\
|\mathcal{C}_1| + |\mathcal{C}_0| &\leq 2^{n-1} \lfloor (E_0(f) + E_1(f))/2^{n-1} \rfloor \\
&\vdots \\
|\mathcal{C}_n| + \cdots + |\mathcal{C}_1| + |\mathcal{C}_0| &= 1 \lfloor (E_0(f) + E_1(f) + \cdots + E_n(f))/1 \rfloor.
\end{aligned} \tag{3}$$

We can now subtract each equation from the last one again to get a set of equations equivalent to (2) which, in some cases, provides a better bound than before.

Example 5 Consider the function $f(\overline{x}) \stackrel{\text{def}}{=} x_1 \vee \cdots \vee x_n$ of example 1. If the lower bound of Corollaries 2 and 3 is applied, we get

$$C_N(f) \geq \frac{1}{2^n} \sum 2^i E_i(f) = 1 + \left(\frac{3}{4}\right)^n - 2\left(\frac{1}{2}\right)^n < 2,$$

while,

$$H_N(f) \geq \frac{1}{2^n} \sum i \cdot E_i(f) = \frac{n}{2}/2^n.$$

Both bounds are not very good so we try the new bound. computing equations (3), and subtracting each from the last, we get:

$$\begin{aligned}
|\mathcal{C}_n| &\geq 1 \\
|\mathcal{C}_{n-1}| + |\mathcal{C}_n| &\geq 3 \\
&\vdots \\
|\mathcal{C}_1| + \cdots + |\mathcal{C}_{n-1}| + |\mathcal{C}_n| &\geq 2^n - 1 \\
|\mathcal{C}_0| + |\mathcal{C}_1| + \cdots + |\mathcal{C}_{n-1}| + |\mathcal{C}_n| &= 2^n - 1.
\end{aligned}$$

And proceeding with these equations instead of (2), we get:

$$C_N(f) \geq n \quad \text{and} \quad H_N(f) \geq 2 - \frac{n}{2^n} - \frac{2}{2^n}.$$

Example 2, shows that these bounds are achievable, hence equality holds in both formulas.

□

4 Overlapping Covers

This section parallels the last one except that the derivation applies to overlapping covers. However, to show why a separate bound for overlapping covers is needed, we start with a short subsection that uses the lower-bound technique of the last section to demonstrate an exponential gap between the minimal size of an overlapping and a nonoverlapping cover for a function. Again, $f : \{0,1\}^n \rightarrow \{-1,0,1\}$ and $\mathcal{C}$ is a (now possibly overlapping) cover for the support set of f.

4.1 Exponential Gap Between Overlapping and Nonoverlapping Covers

For even n, let

$$f(\overline{x}) \stackrel{\text{def}}{=} x_1x_2 \vee x_3x_4 \vee \cdots \vee x_{n-1}x_n.$$

Then f has an overlapping cover of size $n/2$. Yet, using the lower bound, it can be shown that

$$C(f) \geq 1 + 4\left(\left(\frac{17}{16}\right)^{n/2} - \left(\frac{3}{4}\right)^{n/2}\right).$$

Thus, the size of the overlapping cover for f grows linearly with n while the size of any nonoverlapping cover is exponential in n. Note that the gap in this example exists only in the covers for f over $\{\overline{x} : f(\overline{x}) = 1\}$. Any cover for f over $\{\overline{x} : f(\overline{x}) = 1\}$ (including overlapping covers) has an exponential size.

4.2 Relating the Fourier Transform and Overlapping Covers

We write $\overline{Y} \leq \overline{X}$ if $Y_i \leq X_i$ for all $i \in \{1, \ldots, n\}$ and define the *accumulative power spectrum* of $\overline{X}$ as

$$A(\overline{X}) \stackrel{\text{def}}{=} \sum_{\overline{Y} \leq \overline{X}} [\mathcal{E}(f)](\overline{Y}). \tag{4}$$

The *accumulative power spectrum* of the i'th band is then defined as

$$A_i(f) \stackrel{\text{def}}{=} \sum_{\overline{X} \in B_i} A(\overline{X}).$$

It follows from (4) that each $[\mathcal{E}(f)](\overline{X})$ contributes to the accumulative energy of $\binom{n-\text{ones}(\overline{X})}{n-i}$ frequencies in the i'th band. Therefore,

$$A_i(f) = \sum_{j=0}^{i} \binom{n-j}{n-i} E_j(f). \tag{5}$$

Let

$$V_i(f) \stackrel{\text{def}}{=} \sum_{\{C \,:\, C \text{ is } f\text{-constant and } \text{info}(C)=i\}} |C|$$

denote the total volume of f-constant cubes with information content i. For example, if $n = 2$ and $f(x_1, x_2) \stackrel{\text{def}}{=} 1$, then $V_0(f) = V_2(f) = 4$ and, since there are four f-constant cubes of size 2, their total volume is $V_1(f) = 8$. The *complement* $\tilde{C}$ of a cube C is defined by

$$\pi_i(\tilde{C}) \stackrel{\text{def}}{=} \begin{cases} \{0\} & \text{if } \pi_i(C) = \{0,1\}, \\ \{0,1\} & \text{if } \pi_i(C) = \{0\} \text{ or } \{1\}. \end{cases}$$

If C is f-constant then the power spectrum of $f|_C$ is 0 outside $\tilde{C}$ and constant over $\tilde{C}$. This property can be used to show:

Lemma 4 For all $i \in \{0,\ldots,n\}$, $\qquad V_i(f) \leq A_i(f)$. $\square$

Let $\text{dmax}(\overline{x})$ be the dimension of the largest f-constant cube that contains $\overline{x}$. Then $\overline{x}$ is contained in at least $\binom{\text{dmax}(\overline{x})}{d}$ f-constant cubes of dimension d. The average, over $\overline{x}$, of the number of f-constant cubes of dimension d containing $\overline{x}$ is $V_{n-d}(f)/2^n$. Setting $d = 1$ and using (5), we get

$$\frac{1}{2^n} \sum_{\overline{x} \in \{0,1\}^n} \text{dmax}(\overline{x}) \leq \frac{V_{n-1}(f)}{2^n} \leq \frac{A_{n-1}(f)}{2^n} \leq n - \frac{1}{2^n} \sum_{i=0}^{n} i E_i(f).$$

Since, $\text{info}_{\mathcal{C}}(\overline{x}) \geq n - \text{dmax}(\overline{x})$ for all covers $\mathcal{C}$, this proves

Theorem 2 **(Lower bound on the cover entropy of f)**

$$H(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} i E_i(f). \quad \square$$

4.3 Lower Bound on Overlapping Covers

Next, we use Theorem 2 to prove a lower bound for $C(f)$, the minimal number of (possibly overlapping) cubes needed to cover f. Let $\mathcal{P}_{\mathcal{C}} = \{P_1, \ldots, P_{|\mathcal{C}|}\}$ be a partition of S_f into $|\mathcal{C}|$ sets, not necessarily cubes, such that $P_i \subseteq C_i$ for $i = 1, \cdots, |\mathcal{C}|$. Extending the definition of entropy to partitions:

$$H(\mathcal{P}_{\mathcal{C}}) \stackrel{\text{def}}{=} \sum_{i=1}^{|\mathcal{C}|} \frac{|P_i|}{2^n} \log\left(\frac{2^n}{|P_i|}\right) = \frac{1}{2^n}\sum_{i=1}^{|\mathcal{C}|} |P_i|\,(n - \log|P_i|)$$

and since $|\mathcal{P}_{\mathcal{C}}| = |\mathcal{C}|$, we obtain

$$H(f) \leq H(\mathcal{C}) \leq H(\mathcal{P}_C) \leq \log|\mathcal{C}|.$$

Hence,

Lemma 5 **(Lower bound on the cover size of f)**

$$C(f) \geq 2^{H(f)}. \quad \square$$

Combined with Theorem 2, we get

$$C(f) \geq 2^{H(f)} \geq 2^{\frac{1}{2^n}\sum_{i=0}^{n} iE_i(f)}.$$

We conclude the paper by applying the lower bound technique to the functions of examples 1-4. In all the examples the range of the function is defined as $\{-1, 1\}$ thus the bounds obtained are for covers of $S_f = \{0,1\}^n$.

Example 6 For all the 'EXCLUSIVE OR functions',

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_k} = 1, \\ -1 & \text{if } x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_k} = 0, \end{cases}$$

the technique yields the bound $H(f) \geq k$ and therefore $C(f) \geq 2^k$. Clearly, both are exact. $\square$

Example 7 For the 'OR function',

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_1 \vee x_2 \vee \ldots \vee x_n = 1, \\ -1 & \text{if } x_1 \vee x_2 \vee \ldots \vee x_n = 0, \end{cases}$$

(modified from Example 1 to have ± 1 range), the technique yields very poor bounds: $H(f) \geq \frac{2n}{2^n}$ hence, $C(f) \geq 2^{\frac{2n}{2^n}}$ while direct computation shows: $H(f) \approx 2$ and $C(f) = n + 1$. $\square$

Example 8 For the 'flat-spectrum' function of Example 3,

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_1x_2 \oplus x_3x_4 \oplus \cdots \oplus x_{n-1}x_n = 1, \\ -1 & \text{if } x_1x_2 \oplus x_3x_4 \oplus \cdots \oplus x_{n-1}x_n = 0, \end{cases}$$

(defined for even n), the bound yields:

$$H(f) \geq \frac{1}{2^n} \sum_{i=0}^{n} i \binom{n}{i} = \frac{1}{2} n \quad \text{thus,} \quad C(f) \geq 2^{\frac{n}{2}}.$$

The exact values are:

$$H(f) = \frac{1}{2^{n/2}} \sum_{i=n/2}^{n} i \binom{n/2}{n-i} = \frac{3}{4} n \quad \text{and} \quad C(f) = 2^{n/2} \sum_{i=0}^{n/2} \binom{n/2}{i} / 2^i = (\sqrt{3})^n.$$

□

Example 9 For the majority function of Example 4,

$$f(\overline{x}) \stackrel{\text{def}}{=} \begin{cases} -1 & \text{if } \sum_{i=1}^{n} x_i > n/2, \\ 1 & \text{if } \sum_{i=1}^{n} x_i < n/2, \end{cases}$$

(defined for odd n), the bound yields: $H(f) \geq \frac{2n}{2^n} \binom{n-1}{\frac{(n-1)}{2}}$ which, for large n, can be approximated by:

$$H(f) \geq \sqrt{n} \quad \text{and therefore,} \quad C(f) \geq 2^{\sqrt{n}}.$$

The actual values are:

$$H(f) = \frac{n}{2} \quad \text{and} \quad C(f) \approx \frac{2^n}{\sqrt{n}}. \quad \square$$

Acknowledgments

We thank Abbas El Gamal for helpful discussions.

References

References

[1] R.J. Lechner, 'Harmonic Analysis of Switching Functions,' in *Recent Developments in Switching Theory*, A. Mukhopdhyay, ed., Academic Press, New York, 1971.

[2] M.G. Karpovsky, *Finite Orthogonal Series in the Design of Digital Devices*, John Wiley, New York, 1976.

[3] S.L.Hurst, D.M. Miller, and J.C. Muzio, 'Spectral Method of Boolean Function Complexity,' *IEE Electron. Lett.* 18, 572-574 (1982).

[4] M.G. Karpovsky (editor), *Spectral Techniques and Fault Detection*, Academic Press, Florida, 1985.

[5] S.L.Hurst, D.M. Miller, and J.C. Muzio, *Spectral Techniques in Digital Logic*, Academic Press, Florida, 1985.

[6] Y. Brandman, *Spectral Lower-Bounds Techniques for Logic Circuits*, Ph.D. dissertation, Stanford University, Stanford, California, 1987.

[7] R.E. Bryant, 'Graph-Based Algorithms for Boolean Function Manipulation,' *IEEE Transactions on Computer*, Vol. C-35, 677-691, 1986.

[8] R.N. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, 1978.

[9] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error Correcting Codes*, North Holland Publishing Company, New York, 1977.