Theoretical Computer Science 36 (1985) 239-250
North-Holland

BOUNDED-DEPTH, POLYNOMIAL-SIZE CIRCUITS FOR SYMMETRIC FUNCTIONS

Ronald FAGIN, Maria M. KLAWE, Nicholas J. PIPPENGER and Larry STOCKMEYER
IBM Research Laboratory, San Jose, CA 95193, U.S.A.

Communicated by M.S. Paterson
Received October 1983
Revised June 1984

Abstract. Let $\mathcal{F} = \{f_1, f_2, \ldots\}$ be a family of symmetric Boolean functions, where f_n has n Boolean variables, for each $n \geq 1$. Let $\mu_{\mathcal{F}}(n)$ be the minimum number of variables of f_n that each have to be set to constant values so that the resulting function is a constant function. We show that the growth rate of $\mu_{\mathcal{F}}(n)$ completely determines whether or not the family $\mathcal{F}$ is 'good', that is, can be realized by a family of constant-depth, polynomial-size circuits (with unbounded fan-in). Furthermore, if $\mu_{\mathcal{F}}(n) \leq (\log n)^k$ for some k, then the family $\mathcal{F}$ is good. However, if $\mu_{\mathcal{F}}(n) \geq n^{\varepsilon}$ for some $\varepsilon > 0$, then the family is not good.

1. Introduction

Several papers have recently appeared about families of Boolean (0-1 valued) functions that can be realized by bounded-depth, polynomial-size circuits with $\neg$, $\wedge$, and $\vee$ gates, with unbounded fan-in [1, 4, 5, 7, 11]. Let us call such a family *good*, and a family that is not good *bad*. Thus, let f_n be a Boolean function of n Boolean variables, $n = 1, 2, \ldots$. The family $\{f_1, f_2, \ldots\}$ is good if there is a constant d and a polynomial p such that, for each n, the function f_n can be realized by a circuit with depth d and at most $p(n)$ nodes.

Furst, Saxe and Sipser [7] showed that certain simple families are bad. Such families include parity (where $f_n(x_1, \ldots, x_n) = 1$ if an even number of the x_i's are 1) and majority (where $f_n(x_1, \ldots, x_n) = 1$ if at least half of the x_i's are 1).

Let f be a Boolean function of n Boolean variables. The function f is *symmetric* if $f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every $x_1, \ldots, x_n$ and for every permutation π of $\{1, \ldots, n\}$. Thus, if f is symmetric, then $f(x_1, \ldots, x_n)$ is completely determined by the number of x_i's which are equal to 1. In this paper we restrict our attention to *symmetric* functions. We give both positive results (which guarantee that a family is good), and negative results (which guarantee that a family is bad.) Our negative results generalize those of Furst, Saxe and Sipser [7], since the functions they consider (parity and majority) are symmetric. However, we do not give an independent proof of their results, since our results are based on theirs. Our results provide

a unifying framework to show why such families of functions as parity and majority are bad. Thus, Furst, Saxe, and Sipser prove that majority is bad by proving that parity is bad, and by providing a reduction of parity to majority; in this paper, we extend this result to a large class of families of functions.

In Section 2 we provide definitions and give some basic facts. Let $\mathcal{F} = \{f_1, f_2, \ldots\}$ be a family of symmetric functions, where f_n has n variables, for each $n \geq 1$. Let $\mu_{\mathcal{F}}(n)$ be the minimum number of variables of f_n that each have to be set to constant values so that the resulting function is a constant function. In Section 3 we prove that if $\mu_{\mathcal{F}}(n) \geq n^{\varepsilon}$ for some $\varepsilon > 0$, then the family is not good. In Section 4 we prove a combinatorial result which we call the Group Translation Lemma, which is helpful to us later. In Section 5 we prove that if $\mu_{\mathcal{F}'} = O(\mu_{\mathcal{F}})$, and if $\mathcal{F}$ is good, then so is $\mathcal{F}'$. In Section 6 we show that if $\mu_{\mathcal{F}}(n) \leq (\log n)^k$ for some k, then $\mathcal{F}$ is good. In Section 7 we discuss improvements, and possible improvements, to our results. The results of Sections 3 and 6 were obtained independently by Denenberg, Gurevich and Shelah [6]. The result of Section 6 was obtained independently by Ajtai and Ben-Or [2] and Mayr [9] using different methods.

2. Definitions, and facts about spectra

We define the *spectrum* of a symmetric function f to be a word w in $\{0, 1\}^{n+1}$, where, for $0 \leq i \leq n$, the ith character w_i in w is equal to the value of f when i variables are set to 1 and the other variables are set to 0. Clearly, a symmetric function and its spectrum uniquely determine each other. If a circuit C realizes a symmetric function f, then by the *spectrum of* C we mean the spectrum of f.

For a word w, let $|w|$ denote the length of w. Assume that $w = w_0 w_1 \ldots w_n \in \{0, 1\}^{n+1}$. A *subword* of w is a consecutive substring $w_i w_{i+1} \ldots w_j$. A *constant* word is one in which all of the bits are 0 or all of the bits are 1. Let $\Gamma(w)$ be the length of the longest constant subword of w, and let $M(w)$ be $n+1-\Gamma(w)$. If w is the spectrum of f, then we define the *measure* of f (and the measure of w) to be $M(w)$. It is easy to see that the measure of f is the minimum number of variables of f that each have to be set to constant values so that the resulting function is a constant function. Hence, what we are calling the measure is a natural complexity measure, since a function with a small measure is 'close to being a constant function'.

Example. Let f be a symmetric function of 8 variables with spectrum $w = 010111101$. Then $\Gamma(w) = 4$ (because of the constant subword 1111), and $M(w) = 5$. Thus, the measure of f is 5. This corresponds to the fact that if 5 variables are set to constants, then the resulting function is a constant function. In particular, if we set 3 variables to 1 (where 3 is the length of the prefix 010), and 2 variables to 0 (where 2 is the length of the suffix 01), then the resulting function is identically 1. That is, $f(x_1, x_2, x_3, 1, 1, 1, 0, 0)$ is identically 1.

Let $\mathcal{F} = \{f_1, f_2, \ldots\}$ be a family of symmetric functions, where f_n has n variables, for each $n \geq 1$. We define the *measure function* $\mu_{\mathcal{F}}$ (of the family $\mathcal{F}$) by letting $\mu_{\mathcal{F}}(n)$ be the measure of f_n, for $n \geq 1$. If $\mathcal{F}$ is clear from context, then we may write simply μ for $\mu_{\mathcal{F}}$.

A *circuit* is an acyclic, directed graph, with arbitrary fan-in (i.e., in-degree). Each node with fan-in zero is called an *input node* and is identified with a *literal* (a variable or its negation); for convenience, we assume that we do not have two distinct nodes identified with the same literal. The (*input*) *variables* of the circuit are those variables x such that either x or $\neg x$ (or both) is identified with an input node. Each node with fan-in greater than zero is labeled as either an $\wedge$-gate or an $\vee$-gate. There is exactly one node with fan-out zero; this node is the *output node.* Each node computes a Boolean function of the input variables, in the obvious way. When a node of the circuit takes on the value 1 for a given assignment to the input variables, we say that the node *accepts* under that assignment. The Boolean function that is computed by the output node is said to be *realized* by the circuit. The *size* of a circuit is the number of edges. The *depth* of a circuit is the length of a longest path from some input node to the output node. We may refer to a circuit of size s and depth d as an (s, d) circuit. Assume that p is a function, and that C is a circuit with n input variables. We say that C is a (p, d) *circuit* if C is a $(p(n), d)$ circuit. Assume that $w \in \{0, 1\}^{n+1}$; thus, w is the spectrum of a symmetric function with n variables. We say that w is an (s, d) *spectrum* (respectively, a (p, d) *spectrum*) if the associated symmetric function is realized by an (s, d) circuit (respectively, a (p, d) circuit.) (Our definitions of circuits and circuit size differ from the ones used by Furst, Saxe and Sipser [7] in several inconsequential ways.)

We shall frequently make use of the following simple propositions about spectra.

Proposition 2.1. *Let p be a monotone-increasing function from $\mathbb{Z}^+$ (the positive integers) into $\mathbb{Z}^+$. Assume that $w \in \{0, 1\}^{n+1}$ is a (p, d) spectrum.*

(a) *The* ***complement*** *$\bar{w}$ of w (that is, the result of replacing every 0 by a 1 and vice versa) is a (p, d) spectrum.*

(b) *The* ***reverse*** *w^R of w (that is, the result of writing w backwards) is a (p, d) spectrum.*

(c) *Let $g: \mathbb{Z}^+ \to \mathbb{Z}^+$ be an unbounded, monotone-increasing function with $g(1) = 1$. Let $g^{-1}: \mathbb{Z}^+ \to \mathbb{Z}^+$ be defined for each $i \in \mathbb{Z}^+$ by letting $g^{-1}(i)$ be the greatest integer j for which $g(j) \leq i$. Each subword of w of length at least $g(n)+1$ is a $(p \circ g^{-1}, d)$ spectrum.*

Note. In (c) above, we shall make use of the following cases, where $k \in \mathbb{Z}^+$: (i) $g(n) = \lceil n/k \rceil$ (where $g^{-1}(n)$, as defined above, is kn), and (ii) $g(n) = \lfloor n^{1/k} \rfloor$ (where $g^{-1}(n) = (n+1)^k - 1$).

Proof. (a) This follows easily by duality, that is, by interchanging $\wedge$ with $\vee$, interchanging 0 with 1, and interchanging each literal with its negation.

(b) Here we simply take the original circuit and replace every literal by its negation.

(c) Let C be a (p, d) circuit with spectrum w. Let w' be a subword of w with $|w'| = m+1$ and $m \geq g(n)$. By setting $n-m$ variables of C to the appropriate constants, we obtain a circuit C' with m input variables and spectrum w'. The depth of C' is at most d. By definition of g^{-1}, it follows easily that g^{-1} is monotone-increasing and that $g^{-1}(m) \geq g^{-1}(g(n)) \geq n$. Since p is increasing, $p(n) \leq p(g^{-1}(m))$. Therefore, the size of C', expressed as a function of m, is at most $p(g^{-1}(m))$. □

Proposition 2.2. *Let p and q be polynomials, and d a constant. Assume that $w^{(i)} \in \{0,1\}^{n+1}$ is a (p, d) spectrum, for $1 \leq i \leq q(n)$. Then $w^{(1)} \wedge \cdots \wedge w^{(q(n))}$ (respectively $w^{(1)} \vee \cdots \vee w^{(q(n))}$) is a $((p(n)+1)q(n), d+1)$ spectrum.*

Proof. The result of and-ing (respectively, or-ing) together the circuits for $w^{(i)}$ $(1 \leq i \leq q(n))$ has size at most $(p(n)+1)q(n)$. □

3. Negative results

In this section we show that if $\mu_{\mathscr{F}}(n) \geq n^{\varepsilon}$ for some $\varepsilon > 0$, then the family $\mathscr{F}$ is bad, that is, cannot be realized by bounded-depth, polynomial-size circuits.

We shall make use of the following two theorems, both of which are implicit results of Furst, Saxe and Sipser [7].

Theorem 3.1. *Let p be a polynomial, and $d \geq 3$ a positive integer. There is a polynomial p', which depends only on p, such that for each sufficiently large n and each (p, d) spectrum $w \in \{0,1\}^{n+1}$ there is a subword w' of w of length at least $\frac{1}{4}n^{1/4}+1$ that is a $(p', d-1)$ spectrum.*

Theorem 3.2. *Let p be a polynomial, and d a constant. If n is sufficiently large, then the spectrum of the majority function with n variables is not a (p, d) spectrum.*

Before we prove the main result of this section, we need the following lemma.

Lemma 3.3. *Assume that $w = w_0 w_1 \ldots w_n \in \{0,1\}^{n+1}$ is a $(p, 2)$ spectrum, where p is a polynomial of degree k. If n is sufficiently large, then $w_{k+1} = w_{k+2} = \cdots = w_{n-k-1}$.*

Proof. By duality, as in the proof of Proposition 2.1(a), we can assume that the output node is an $\vee$-gate. Assume now that the conclusion to the lemma is false; we shall derive a contradiction. Let r be minimal such that $w_r \neq w_{r+1}$, and $k+1 \leq r < n-k-1$. There are two cases.

Case 1: $w_r = 1$. Thus, $w_{r+1} = 0$. Let $\{u_1, \ldots, u_r\}$ be an arbitrary subset of exactly r of the variables of the circuit, and let $u_{r+1}, \ldots, u_n$ be the remaining variables.

Since $w_r = 1$, we know that the output node (an $\vee$-gate) takes on the value 1 when $u_1 = \cdots = u_r = 1$ and $u_{r+1} = \cdots = u_n = 0$. Therefore, there is an $\wedge$-gate, which we shall denote by v, that takes on the value 1 when $u_1 = \cdots = u_r = 1$ and $u_{r+1} = \cdots = u_n = 0$. We now show that the literal $\neg u_{r+1}$ is an input node that connects to v in the circuit. For, if not, then it is easy to see that the node v (and hence, the output node) takes on the value 1 when $u_1 = \cdots = u_{r+1} = 1$ and $u_{r+2} = \cdots = u_n = 0$. This is a contradiction, since by assumption $w_{r+1} = 0$. Therefore, $\neg u_{r+1}$ connects to v. Similarly, $\neg u_i$ connects to v for each i where $r+1 \leq i \leq n$. Therefore, the node v takes on the value 1 whenever precisely $u_1, \ldots, u_r$ take on the value 1, but takes on the value 0 whenever any r-sized subset of the variables takes on the value 1. Similarly, for each set Y of exactly r variables, there is a node v_Y that takes on the value 1 whenever precisely the variables in Y take on the value 1, but takes on the value 0 whenever any r-sized subset $X \neq Y$ of the variables takes on the value 1. It follows that $v_{Y_1} \neq v_{Y_2}$ whenever $Y_1 \neq Y_2$. Hence, the number of $\wedge$-gates is at least $\binom{n}{r}$. Therefore, the size of the circuit is at least $\binom{n}{r}$. Hence, $p(n) \geq \binom{n}{r}$.

It is well known that the function mapping i to $\binom{n}{i}$ is first monotone-increasing, and then monotone decreasing. Therefore, if $i \leq r \leq j$, then $\binom{n}{r} \geq \min\{\binom{n}{i}, \binom{n}{j}\}$. In our case, we know that $k+1 \leq r \leq n-k-2$, and so $\binom{n}{r} \geq \min\{\binom{n}{k+1}, \binom{n}{k+2}\} = \binom{n}{k+1}$ (we assume that n is large enough that $k+2 \leq \frac{1}{2}n$). Therefore, since we showed that $p(n) \geq \binom{n}{r}$, it follows that $p(n) \geq \binom{n}{k+1}$. But if n is sufficiently large, then this is impossible, since $p(n)$ is a polynomial of degree k, while $\binom{n}{k+1}$ is of degree $k+1$.

Case 2: $w_r = 0$. Thus, $w_{r+1} = 1$. The proof is similar to that of Case 1, except that we consider subsets $\{u_1, \ldots, u_{r+1}\}$ of exactly $r+1$ variables. We let v be an $\wedge$-gate that takes on the value 1 when $u_1 = \cdots = u_{r+1} = 1$ and $u_{r+2} = \cdots = u_n = 0$. Since $w_r = 0$, the literals u_i (for $1 \leq i \leq r+1$) are input nodes that each connect to v in the circuit. An argument similar to that in Case 1 then shows that the number of $\wedge$-gates (and hence, the size of the circuit) is at least $\binom{n}{r+1}$. Hence, $p(n) \geq \binom{n}{r+1}$. Since $k+2 \leq r+1 \leq n-k-1$, it follows as before that $p(n) \geq \binom{n}{k+1}$. The proof then concludes as in Case 1. □

We are now ready to prove a result (Theorem 3.4 below) which immediately implies our promised negative result that if $\mu_{\mathcal{F}}(n) \geq n^{\varepsilon}$ for some $\varepsilon > 0$, then $\mathcal{F}$ is bad.

Theorem 3.4. *Assume that there is some $\varepsilon > 0$ such that $\mu_{\mathcal{F}}(n) \geq n^{\varepsilon}$ for infinitely many n. Then the family $\mathcal{F}$ is bad, that is, cannot be realized by bounded-depth, polynomial-size circuits.*

Proof. Assume that the family $\mathcal{F} = \{f_1, f_2, \ldots\}$ is good; we shall derive a contradiction. Since the family $\mathcal{F}$ is good, there is a polynomial p and an integer d such that f_n has a (p, d) spectrum, for each n. By applying Theorem 3.1 $d-2$ times, we see that there is a polynomial q such that if n is sufficiently large, the spectrum of f_n contains a subword z of length at least $(1/4)^{d-2} n^{(1/4)^{d-2}} + 1$ that is a $(q, 2)$ spectrum. We can assume without loss of generality that q is monotone increasing. Let k be

the degree of q. By Lemma 3.3 we see all of the bits except possibly the first $k+1$ and last $k+1$ bits of the subword z are the same. Let δ be arbitrary such that $0<\delta<(1/4)^{d-2}$. It follows that if n is sufficiently large, then the spectrum of f_n contains a constant subword of length at least n^δ.

For ease in discussion, let us fix for now a sufficiently large n such that $\mu_{\mathcal{F}}(n) \geq n^\varepsilon$, and let w be the spectrum of f_n. Let us write w as tuv (the concatenation of t, u, and v), where u is the longest constant subword of w. Since u is the longest constant subword of w, and since, as we showed, w contains a constant subword of length at least n^δ, it follows that $|u| \geq n^\delta$. Now $|t|+|v| = \mu(n) \geq n^\varepsilon$. Therefore, either $|t| \geq \frac{1}{2}n^\varepsilon$ or $|v| \geq \frac{1}{2}n^\varepsilon$; by considering the reverse of w if necessary (as in Proposition 2.1(b)), we can assume that $|t| \geq \frac{1}{2}n^\varepsilon$. By considering the complement of w if necessary, we can assume that the constant word u consists of all 1's. Hence, the last bit of t is 0. Therefore, w is of the form $t'0uv$, where $|t'| \geq \frac{1}{2}n^\varepsilon - 1$, and where u is a constant word of at least n^δ 1's. Let j be an integer such that $1/j$ is smaller than the minimum of ε and δ. So, if n is sufficiently large, then $|t'| > \lfloor n^{1/j} \rfloor$ and $|u| > \lfloor n^{1/j} \rfloor$. Define $N = \lfloor n^{1/j} \rfloor$. It is easy to see that the spectrum of the majority function of N variables can be obtained by and-ing together appropriate length N subwords of $w = t'0uv$. By Propositions 2.1(c) and 2.2, it follows that the majority function is $(p', d+1)$ spectrum for some polynomial p'. Since $N = \lfloor n^{1/j} \rfloor$ can be taken to be arbitrarily large, this is a contradiction of Theorem 3.2. □

We shall refer to the following immediate corollary later.

Corollary 3.5. *If $\mu_{\mathcal{F}}(n) \geq n^\varepsilon$ for some $\varepsilon > 0$ and every n, then $\mathcal{F}$ is bad.*

Remark. Theorem 3.4 cannot be generalized to nonsymmetric Boolean functions. As a simple example, let $n = k^2$, partition the variables into k blocks of size k, let f_i be the $\wedge$ of the variables in the ith block, and let the (nonsymmetric) function f be the $\vee$ of f_i over $1 \leq i \leq k$. It is easy to see that at least $k = n^{1/2}$ variables of f must be set to constants in order to make f a constant function. A different example is given by Ajtai and Ben-Or [2].

4. Group Translation Lemma

In this section we prove a combinatorial result, which we call the Group Translation Lemma, which will be helpful later.

Let G be a group and let X be a set. Assume that for each $g \in G$ there is an associated function $f_g : X \to X$. If $g \in G$ and $x \in X$, then we shall simply write gx for $f_g x$. The group G is said to *act transitively* on X if

(a) $ex = x$ for the identity element $e \in G$;

(b) $(gh)x = g(hx)$, when $g, h \in G$ and $x \in X$; and

(c) for each $x, y \in X$ there is $g \in G$ such that $gx = y$.

If $S \subseteq X$ and $g \in G$, then by gS, we mean $\{gs : s \in S\}$. If $H \subseteq G$ and $S \subseteq X$, then by HS we mean $\{gx : g \in H \text{ and } x \in S\}$.

Lemma 4.1 (Group Translation Lemma). *Assume that the finite group G acts transitively on the finite set X, and that $S \subseteq X$ is nonempty. Then $HS = X$ for some $H \subseteq G$ with $|H| \leq (|X|/|S|)(1+\ln|S|)$.*

Note. Intuitively, the lemma says that there is a small number of translations of S that cover X. Clearly, the smallest number of translations of S that cover X that we could possibly hope for is $|X|/|S|$; the lemma says that we can almost attain this number.

Proof. We shall define a random subset H of G such that $HS = X$ and show that

$$\mathscr{E}(|H|) \leq (|X|/|S|)(1+\ln|S|),$$

where $\mathscr{E}(\ldots)$ denotes 'the expectation of ...'. This clearly implies the theorem.

Let $0 < p < 1$ be a real number to be chosen later. Let H_1 be a random subset of G obtained by taking each element of G independently with probability p. Let H_2 be a subset of G obtained by taking, for each x in X but not in H_1S, some $g \in G$ such that $x \in gS$. Let $H = H_1 \cup H_2$. Clearly, $HS = X$.

Let $G(x, y) = \{g \in G : x = gy\}$. For each $y \in X$, $\{G(x, y)\}_{x \in X}$ is a partition of G, so

$$\sum_{x \in X} |G(x, y)| = |G|.$$

Since G acts transitively on X, it follows that $G(x, y)$ is a left coset of $G(y, y)$, and so

$$|G(x, y)| = |G(y, y)| = |G|/|X|.$$

Let $G(x, S) = \{g \in G : x \in gS\}$. For each $x \in X$, $\{G(x, y)\}_{y \in S}$ is a partition of $G(x, S)$ into equal-sized sets, and so

$$|G(x, S)| = |G||S|/|X|.$$

Each element of G appears in H_1 with probability p, so

$$\mathscr{E}(|H_1|) = |G|p.$$

An element $x \in X$ fails to appear in H_1S only if each element of $G(x, S)$ fails to appear in H_1. This event occurs with probability

$$(1-p)^{|G(x,S)|} = (1-p)^{|G||S|/|X|}$$

Thus,

$$\mathscr{E}(|H_2|) \leq |X|(1-p)^{|G||S|/|X|}$$

and

$$\mathscr{E}(|H|) \leq |G|p + |X|(1-p)^{|G||S|/|X|}.$$

Using the inequality $1-p \leq e^{-p}$ and setting

$$p = (|X| \ln |S|)/(|G||S|)$$

completes the proof. □

Remark. A simpler counting argument gives $|H| \leq \lceil (|X|/|S|) \ln |X| \rceil$, which would suffice for the purposes of this paper. We prefer to prove the sharper bound, since it might be of independent interest and useful in other applications.

5. The order of $\mu_{\mathcal{F}}$ characterizes goodness

In this section we show that the growth rate of the measure function $\mu_{\mathcal{F}}$ completely determines whether or not the family $\mathcal{F}$ is good. In particular, if $\mu_{\mathcal{F}'} = O(\mu_{\mathcal{F}})$, and if $\mathcal{F}$ is good, then so is $\mathcal{F}'$.

Let us denote by $\theta_{m,n}$ the symmetric function which takes on the value 1 precisely if at least m of the n variables take on the value 1. Thus, the spectrum of $\theta_{m,n}$ is the word $0^{\lceil m \rceil}1^{n+1-\lceil m \rceil}$. Functions $\theta_{m,n}$ are called *threshold functions.* We may refer to the family $\{\theta_{f(n),n} : n = 1, 2, \ldots\}$ where f is a function, as *threshold f.* We begin with some lemmas about the circuits of threshold functions.

Let $T_{m,n}(k)$ be the minimum possible size of a circuit of depth k for $\theta_{m,n}$.

Lemma 5.1

$$T_{lm,ln}(k+2) \leq (8m)^{l/2}(ln+1)l(T_{m,n}(k)+1).$$

Proof. Let the variables $x_1, \ldots, x_{ln}$ be partitioned into l blocks $B_1, \ldots, B_l$, each containing n variables. Let C be a $(T_{m,n}(k), k)$ circuit for $\theta_{m,n}$. For $1 \leq j \leq l$, let C_j be the result of substituting the variables of B_j for the variables of C. Let D be the $(l(T_{m,n}(k)+1), k+1)$ circuit obtained by and-ing together the outputs of $C_1, \ldots, C_l$.

Let X denote the set of prime implicants of $\theta_{lm,ln}$, so $|X| = \binom{ln}{lm}$. (The *prime implicants* of $\theta_{lm,ln}$ are the functions obtained by and-ing together lm distinct variables.) Let $S \subseteq X$ denote the set of prime implicants accepted by D (in the obvious sense), so $|S| = \binom{n}{m}^l$. Let G be the symmetric group on the variables $x_1, \ldots, x_{ln}$. The group G acts on circuits in an obvious way. It also acts transitively on X in such a way that gD accepts the prime implicants in gS.

By the Group Translation Lemma, there is a set $H \subseteq G$ such that $HS = X$ and

$$|H| \leq (\binom{ln}{lm}/\binom{n}{m}^l)(1 + l \ln \binom{n}{m}).$$

Using the inequalities [10]

$$2^{nH(m/n)}/(8m)^{1/2} \leq \binom{n}{m} \leq 2^{nH(m/n)}$$

(where $H(\xi) = -\xi \log_2 \xi - (1-\xi)\log_2(1-\xi) \leq 1$), we have

$$|H| \leq (8m)^{l/2}(ln+1).$$

Clearly, the $((8m)^{l/2}(ln+1)l(T_{m,n}(k)+1), k+2)$ circuit obtained by or-ing together the outputs of the circuits $\{gD\}_{g\in H}$ accepts $\theta_{lm,ln}$, which completes the proof. □

Corollary 5.2. *Let p be a polynomial, and let d, m, n and s be positive integers. Assume that $m+s\leqslant n$, and that $\theta_{m,n}$ is a threshold function with a (p, d) circuit. Then $\theta_{m,n+s}$ has a $(p', d+2)$ circuit, where p' is a polynomial that depends only on p.*

Proof. By Lemma 5.1, $\theta_{2m,2n}$ has a $(q, d+2)$ circuit, where q is a polynomial that depends only on p. The corollary follows by substituting 1's for m variables and 0's for $n-m-s$ variables. □

Corollary 5.3. *Let p be a polynomial, and let c, d, m and n be positive integers. Assume that $\theta_{m,n}$ is a threshold function with a (p, d) circuit. Then $\theta_{cm,n}$ has a $(p', d+2)$ circuit, where p' is a polynomial that depends only on p and c.*

Proof. By Lemma 5.1, $\theta_{cm,cn}$ has a $(q, d+2)$ circuit, where q is a polynomial that depends only on p and c. The corollary follows by substituting 0's for $(c-1)n$ variables. □

Lemma 5.4. *Let p be a polynomial, and let d, m and n be positive integers. Assume that $m<\frac{1}{2}n$, and that $\theta_{m,n}$ is a threshold function with a (p, d) circuit. Let f be an arbitrary symmetric function with n variables, with measure at most m. Then f is realized by a $(p', d+4)$ circuit, where p' is a polynomial that depends only on p.*

Proof. Let $w^{(i)}\in\{0, 1\}^{n+1}$ be the word with 1 in the ith position and 0's elsewhere $(0\leqslant i\leqslant n)$. Assume that $0\leqslant i<m$ or $n-m<i\leqslant n$. We shall now discuss the size and depth of a circuit with spectrum $w^{(i)}$. By Corollary 5.2 we know that $\theta_{m,n+m}$ has a $(q, d+2)$ circuit, where q is a polynomial that depends only on p. Let u be the spectrum of $\theta_{m,n+m}$; thus, $u=0^m1^{n+1}$. Assume now that $0\leqslant i<m$. Let $u'=0^i1^{n+1-i}$, and let $u''=0^{i+1}1^{n-i}$. Clearly, u' and u'' are each subwords of u. Since $m<\frac{1}{2}n$, the length of each of u' and u'' is greater than $g(n)+1$, where $g(i)=\lceil\frac{1}{2}i\rceil$. Therefore, by Proposition 2.1(c), we see that each of u' and u'' are $(q\circ g^{-1}, d+2)$ spectra. (Here g^{-1} maps i onto $2i$.) It is easy to see that $w^{(i)}$ is the result of and-ing together u' and the complement of u''. Therefore, when $1\leqslant i<m$, we see by the above that $w^{(i)}$ is a $(q', d+3)$ spectrum, where q' is a polynomial that depends only on p. By taking the reverse of $w^{(i)}$, we see by Proposition 2.1(b) that the same is true when $n-m<i\leqslant n$.

Let $v=0^m1^{n+1-2m}0^m$. Let $w=0^m1^{n+1-m}$ be the spectrum of $\theta_{m,n}$. Clearly $v=w\wedge w^R$; hence, v is a $(2p+2, d+1)$ spectrum.

Let f be an arbitrary symmetric function of n variables, with measure m. Let x be the spectrum of f. Define t by setting $t=x$, if the longest constant subword of x contains all 1's, and otherwise setting $t=\bar{x}$, the complement of x. By or-ing together v with appropriate choices of $w^{(i)}$, we can obtain t. Thus, by Proposition 2.1(a) and

2.2, we see that x is a $(p', d+4)$ spectrum, where p' is a polynomial that depends only on p. □

Lemma 5.5. *Let p be a polynomial, and let d, m, and n be positive integers. Assume that $w \in \{0,1\}^{n+1}$ is a (p, d) spectrum with measure m. Then $\theta_{m/2,n-m}$ has a $(p', d+1)$ circuit, where p' is a polynomial that depends only on p.*

Proof. If $m=0$, then the result is immediate. So, assume that $m>0$. Since w has measure m, we can assume, by reversing or complementing if necessary, that $w = u01^j v$, where $|u| \geq \lceil m/2 \rceil - 1$ and $j \geq n-m$. As in the proof of Theorem 3.4, we obtain the spectrum of $\theta_{m/2,n-m}$ by and-ing together appropriate length $n-m+1$ subwords of w. □

We are now ready to prove the main result of this section.

Theorem 5.6. *If $\mu_{\mathscr{F}'} = O(\mu_{\mathscr{F}})$, and if $\mathscr{F}$ is good, then so is $\mathscr{F}'$.*

Proof. Let $\mathscr{F} = \{f_1, f_2, \ldots\}$, and let $\mathscr{F}' = \{f_1', f_2', \ldots\}$, where f_n and f_n' each have n variables ($n = 1, 2, \ldots$). Assume that f_n has a (p, d) circuit for each n, and let c be an integer such that $\mu_{\mathscr{F}'}(n) \leq c\mu_{\mathscr{F}}(n)$ for each n. Let us write μ for $\mu_{\mathscr{F}}$, and μ' for $\mu_{\mathscr{F}'}$. Take n sufficiently large that $\mu(n) < n/(2c)$ (this is possible by Theorem 3.4). By Lemma 5.5 we know that $\theta_{\mu(n)/2,n-\mu(n)}$ has a $(p_1, d+1)$ circuit, for some polynomial p_1. Therefore, by Corollary 5.2, we know that $\theta_{\mu(n)/2,n}$ has a $(p_2, d+3)$ spectrum, for some polynomial p_2. Let $j = 2c\lceil \frac{1}{2}\mu(n) \rceil$. By Corollary 5.3 it follows that $\theta_{j,n}$ has a $(p_3, d+5)$ spectrum, for some polynomial p_3. Clearly $j \geq c\mu(n) \geq \mu'(n)$, and $\mu'(n) \leq c\mu(n) < \frac{1}{2}n$. So, by Lemma 5.4 we know that f_n' is realized by a $(p_4, d+9)$ circuit, for some polynomial p_4. This concludes the proof. □

6. Positive results

In this section we show that if $\mu_{\mathscr{F}}(n) \leq (\log n)^k$ for some constant k, then the family $\mathscr{F}$ is good. We note that this paper arose when one of the authors conjectured that if $\mathscr{F}$ is good, then $\mu_{\mathscr{F}}$ is bounded by a constant k. Of course, our results in this section show that this conjecture is false.

Theorem 6.1. *Let k be a positive integer. Assume that $m = O((\log n)^k/(\log\log n)^{k-1})$. Then there is a polynomial p such that $\theta_{m,n}$ has a $(p, 2k+1)$ circuit.*

Note. The case $k=1$ of this theorem was proved by Khasin [8].

Proof. We shall proceed by induction on k. If $k=1$, let $l=m$; otherwise, let $l = \lceil \log n / \log\log n \rceil$. Let $m' = \lceil m/l \rceil$ and $n' = \lceil (n + lm' - m)/l \rceil$. We claim that there

is a polynomial p' such that $\theta_{m',n'}$ has a $(p', 2k-1)$ circuit. If $k=1$, then $m'=1$ and the claim is satisfied by or-ing together the variables. If $k \geq 2$, then $m' = O((\log n)^{k-1}/(\log\log n)^{k-2})$, and the claim is satisfied by inductive hypothesis. Thus, $T_{m',n'}(2k-1) \leq p'(n')$. Lemma 5.1 shows that $T_{lm',ln'}(2k+1) \leq q(ln')$ for some polynomial q. Substituting 1's for $lm'-m$ variables and 0's for $(ln'-n)-(lm'-m)$ variables completes the proof. □

Corollary 6.2. *If $\mu_{\mathcal{F}}(n) \leq (\log n)^k$ for some k, then the family $\mathcal{F}$ is good.*

Proof. Since $(\log n)^k = O((\log n)^{k+1}/(\log\log n)^k)$, the corollary is an immediate consequence of Theorem 6.1 and Lemma 5.4. □

It is interesting to note that recent work of Boppana [3] easily implies that if the family threshold $(\log n)^k$ is realized by a family of monotone (p, d) circuits for some polynomial p, then $d = \Omega(k)$. Therefore, restricting attention to monotone circuits of polynomial size, the depth bound of Theorem 6.1, when expressed as a function of k, is the best possible to within a constant factor.

7. Improvements and possible improvements

In this section, we discuss issues related to improving our results. We begin with a theorem that relates to improving both our positive and our negative results.

Theorem 7.1. *Assume that $\mathcal{F}$ is a good family, that $g: \mathbb{Z}^+ \to \mathbb{Z}^+$ is a surjective, monotone-increasing function, and that $\mu_{\mathcal{F}}(n) \geq g(n)$ for every n. Let $g^{-1}: \mathbb{Z}^+ \to \mathbb{Z}^+$ be defined as in Proposition* 2.1(c). *Then there are constants c, d, and m such that every symmetric function with $n \geq m$ variables can be realized by a $((g^{-1}(n))^c, d)$ circuit.*

Proof. Let $w \in \{0,1\}^{n+1}$ be an arbitrary nonempty word. By surjectiveness of g, we can find N such that $g(N) = n$. Since $\mathcal{F}$ is good, Theorem 3.4 implies that $N \geq 2n$ for all sufficiently large n. By Theorem 5.6 we see that threshold g is a good family. Thus, there is a polynomial p and a constant d' such that if N is sufficiently large, then the word $x = 0^n 1^{N+1-n}$ is a (p, d') spectrum. Let $v^{(i)} = 0^i 1^{n+1-i}$ (for $0 \leq i \leq n+1$). If $0 \leq i \leq n$, then $v^{(i)}$ is a subword of x of length $n+1$ which, by Proposition 2.1(c), is a $(p \circ g^{-1}, d')$ spectrum. Of course, $v^{(n+1)} = 0^{n+1}$ is also a $(p \circ g^{-1}, d')$ spectrum. Letting $w^{(i)} \in \{0,1\}^{n+1}$ be the word with 1 in ith position and 0's elsewhere (for $0 \leq i \leq n$), it follows easily from Propositions 2.1(a) and 2.2 that $w^{(i)}$ is a $(2p \circ g^{-1} + 2, d'+1)$ spectrum, and that w is an $((n+1)(2p \circ g^{-1}(n)+3), d'+2)$ spectrum. Since g is surjective and monotone-increasing, we know that $n \leq g^{-1}(n)$. Thus, there is a constant c such that $(n+1)(2p \circ g^{-1}(n)+3) \leq (g^{-1}(n))^c$ for n sufficiently large. □

Note that (the contrapositive of) Theorem 7.1, along with Furst, Saxe, and Sipser's result that parity is a bad family, immediately implies our negative result of Corollary 3.5. This is because if $g(n) = \lfloor n^{\varepsilon} \rfloor$ for each n, then g^{-1} is bounded by a polynomial. Let us now see what Theorem 7.1 says about improving Corollary 3.5. Ajtai [1] has shown that, for some constant c, the family of parity functions on n variables cannot be realized by bounded depth circuits of size $n^{c(\log n)}$. From this fact, along with Theorem 7.1, we can conclude that if $\mu_{\mathscr{F}}(n) \geq 2^{(\log n)^{0.5+\varepsilon}}$ for some $\varepsilon > 0$, then the family $\mathscr{F}$ is bad. This is a stronger statement than Corollary 3.5. Further improvements of Ajtai's result automatically lead (via Theorem 7.1) to further strengthenings of Corollary 3.5.

Let us now consider our positive results in the light of Theorem 7.1. By considering the functions g where $g(n) = (\log n)^k$, as in Corollary 6.2, we see from Theorem 7.1 that, for each $\varepsilon > 0$, there is a constant d so that every symmetric function with n variables can be realized by a $(2^{n^{\varepsilon}}, d)$ circuit. This is the best known size for constant depth [5]. Again, this result could be improved if we could improve our positive result in Corollary 6.2.

References

[1] M. Ajtai, Σ_1^1-formulae on finite structures, *Annals Pure and Applied Logic* **24** (1983) 1-48.

[2] M. Ajtai and M. Ben-Or, A theorem on probabilistic constant depth computations, *Proc. 16th ACM SIGACT Symp. on the Theory of Computing* (1984) 471-474.

[3] R.B. Boppana, Threshold functions and bounded depth monotone circuits, *Proc. 16th ACM SIGACT Symp. on the Theory of Computing* (1984) 475-479.

[4] A. Chandra, S. Fortune and R. Lipton, Unbounded fan-in circuits and associative functions, *Proc. 15th ACM Symp. on Theory of Computing* (1983) 52-60.

[5] A. Chandra, L. Stockmeyer and U. Vishkin, Constant depth reducibility, *SIAM J. Computing* **13** (1984) 423-439.

[6] L. Denenberg, Y. Gurevich and S. Shelah, Cardinalities definable by constant-depth, polynomial-size circuits, Report TR-26-83, Aiken Computation Laboratory, Harvard University, Cambridge, MA, 1983.

[7] M. Furst, J.B. Saxe and M. Sipser, Parity, circuits, and the polynomial time hierarchy, *Proc. 22nd IEEE Symp. on Foundations of Computer Science* (1981) 260-270.

[8] L.S. Khasin, Complexity bounds for the realization of monotonic symmetrical functions by means of formulas in the basis $\vee$, &, $\neg$, *Sov. Phys. Dokl.* **14** (1970) 1149-1151.

[9] E. Mayr, Personal communication, 1984.

[10] W.W. Peterson, *Error Correcting Codes* (MIT Press, Cambridge, MA, 1961) Appendix 1.

[11] M. Sipser, Borel sets and circuit complexity, *Proc. 15th ACM Symp. on Theory of Computing* (1983) 61-69.