INFORMATION AND COMPUTATION 112, 257–272 (1994)

Approximating Threshold Circuits by Rational Functions*

RAMAMOHAN PATURI AND MICHAEL E. SAKS[†]

Department of Computer Science and Engineering,
University of California at San Diego, La Jolla, California 92093

Motivated by the problem of understanding the limitations of threshold networks for representing boolean functions, we consider size–depth trade-offs for threshold circuits that compute the parity function. Using a fundamental result in the theory of rational approximation, we show how to approximate small threshold circuits by rational functions of low degree. We apply this result to establish an almost optimal lower bound of $\Omega(n^2/\ln^2 n)$ on the number of edges of any depth-2 threshold circuit with polynomially bounded weights that computes the parity function. We also prove that any depth-3 threshold circuit with polynomially bounded weights requires $\Omega(n^{1.2}/\ln^{5/3} n)$ edges to compute parity. On the other hand, we give a construction of a depth d threshold circuit that computes parity with $n^{1-1/\Theta(\phi^d)}$ edges where $\phi = (1+\sqrt{5})/2$ is the golden ratio. We conjecture that there are no linear size bounded depth threshold circuits for computing parity.

1. INTRODUCTION

Complexity theory of small depth circuits over various bases has received considerable attention in the last decade. The first complexity results on such circuits were obtained by Ajtai [1] and Furst *et al.* [11] who showed that the parity function on n variables cannot be computed by polynomial size, bounded depth, unbounded fan-in circuits over the standard basis {AND, OR, NOT}. Boppana and Sipser [7] is an excellent reference for the results in this area. Although much is now known about the limitations of bounded depth circuits over {AND, OR, NOT}, very little is known about the limitations of circuits over more powerful bases.

The basis consisting of *linear threshold* gates is of particular interest. A threshold gate produces a boolean output by comparing a weighted sum of its boolean inputs with a threshold value. Circuits over this basis are called threshold circuits. Bounded depth, polynomial size threshold circuits have considerably more computational power than similarly restricted circuits over the standard basis. For example, it has been shown that sorting,

* A preliminary version of this paper appeared in the Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science, 1990.

[†] Research supported in part by NSF Grant CCR-8911388.

0890-5401/94 $6.00

multiplication, and division can be performed by bounded depth polynomial size threshold circuits [10, 18, 5, 20, 9]. Such circuits also provide a fundamental computational model for neural networks [14, 13]. Despite the added power provided by the basis of threshold gates, there are still simple functions for which no bounded depth polynomial size threshold circuit is known. Indeed, it is generally believed that such circuits are not powerful enough to compute all functions in the class NC^1, that is, those computable by a logarithmic depth bounded fan-in circuit over the standard basis. At this point it is known that depth-2 polynomial size threshold circuits with polynomially bounded weights cannot compute all of NC^1 [12] and limitations have been established for other restricted classes of threshold circuits [14, 8, 25].

In this paper, we prove that the function computed by a depth-2 or depth-3 threshold circuit can be represented as the sign of a rational function of the inputs whose degree is bounded as a function of the number of edges in the circuit. As an application of these results, we prove lower bounds on the number of edges required to compute parity by such circuits. It is well known that the parity function on n variables can be computed by a depth-2 threshold circuit with a quadratic number of edges [15]. Our techniques yield an almost optimal lower bound $\Omega(n^2/\log^2 n)$ on the number of edges for computing the parity function with depth-2 threshold circuits whose weights are bounded by a polynomial in n. For depth-3 circuits, we obtain a lower bound of $\Omega(n^{1.2}/\ln^{5/3} n)$ edges, which, in particular, accomplishes the rather modest task of separating NC^1 from depth-3 threshold circuits with a linear number of edges and polynomially bounded weights.

For greater depths it is not hard to construct a circuit that computes parity with a linear number of edges in depth $O(\log \log n)$. In fact, we give a construction for each depth d between 2 and $O(\log \log n)$ (which we believe to be essentially optimal) with $O(n^{1+c^{-d}})$ edges for some constant $c > 1$ (see Section 3).

Our starting point was an attempt to show that the trade-off exhibited by this construction is the best possible. While our results fall short of this goal, our techniques involve an interesting application of *rational approximation theory* to complexity theory. While the representation of boolean functions by polynomials has proved to be very useful in the study of boolean function complexity, e.g., [19, 22, 23, 26, 3], our use of rational functions for approximating boolean functions seems to be the first. The technique holds promise towards a complete answer to this trade-off question and also seems to be a natural and a potentially useful approach to the analysis of threshold circuits in general. Such an approach might also be relevant to the solution of the question of whether TC^0 (the class of functions computable by bounded depth and polynomial size threshold circuits) is equal to NC^1. Recently, these techniques have been used to study the complexity class PP [6]. The rational approximation techniques in this paper have subsequently been applied in [21] to yield lower bounds on the number of gates in threshold circuits. It is also know that at least $\log_2 n$ threshold gates are required to compute parity even when there is no depth restriction [24].

In Section 2, we define threshold gates and circuits. In Section 3, we show how to construct parity circuits of depth d with $n^{1+\Theta(\phi^{-d})}$ edges. In Section 4, we present a lemma relating the edges and the nodes of depth-2 circuits that compute parity and discuss a related open problem concerning the minimum number of hyperplanes required to cut all the edges of the cube. In Section 5, we consider representations of boolean functions by real functions and prove that any polynomial representation of parity requires degree n. Using this, we give an alternate proof of a result of Alon *et al.* on the number of Hamming *stripes* required to cover the hypercube [2]. Finally, in Section 6, we present our approximation results and derive lower bounds on parity.

2. Threshold Circuits: Preliminaries

A threshold gate with fan-in n is an n-tuple $\vec{w} \in \mathbf{Z}^n$, where $\mathbf{Z}$ is the set of all integers. The function h: $\{0, 1\}^n \to \{0, 1\}$ computed by such a gate on input $\vec{b} \in \{0, 1\}^n$ is given by $\operatorname{sgn}(\sum_{i=1}^n w_i b_i)$ where $\operatorname{sgn}: \mathbf{R} \to \{0, 1\}$ is defined as

$$\operatorname{sgn}(\alpha) = \begin{cases} 1 & \text{if } \alpha > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We define the *norm* u of a threshold gate as $u = \sum_i |w_i|$.

A threshold circuit T on n inputs $(b_1, b_2, ..., b_n)$ is a directed acyclic graph with a designated node (output) and exactly $n+1$ *source* nodes, one for each variable and one for the constant 1. The semantics of such a circuit is given by interpreting each non-source node v as a threshold gate on its incoming edges. The *function* $y_v(b_1, ..., b_n)$ *computed by* v is obtained by functional composition in the obvious way. The function f_T: $\{0, 1\}^n \to \{0, 1\}$ computed by T is the function assigned to its designated output node.

The *node complexity* of T is defined as the number of non-source nodes of T. The *edge complexity* of T is defined as the number of edges in T.

We define the *level* of each node in the following way: The level of each source node is 0. The level of any other node i is 1 more than the maximum level of its immediate predecessors. The *depth* of T is the level of the output node.

The node (edge) complexity $N_f(n)$ ($E_f(n)$) of a boolean function f on n inputs is the node (edge) complexity of the minimal node (edge) threshold circuits that computes f.

For convenience, we assume that all our circuits are layered in that nodes at level i for $i \geqslant 1$ receive inputs only from level $i-1$. Such a restriction would increase the edge complexity by at most a factor of d, the depth of the circuit.

3. Threshold Circuits for Parity: Upper Bounds

A boolean vector $\vec{b}$ is *even* if $\sum_{i=1}^{n} b_i \equiv 0 \bmod 2$ and is *odd* otherwise. The *odd parity* function p_o on a sequence $\vec{b} = (b_1, b_2, ..., b_n)$ of boolean variables is defined to be the characteristic function of the set of odd vectors. The *even parity* function p_e is defined by $p_e(\vec{b}) = 1 - p_o(\vec{b})$. Both even and odd parity functions are referred to simply as *parity* functions.

We now consider the problem of constructing threshold circuits for parity with a minimal number of edges. The best construction that has been achieved for depth-2 threshold circuits uses $n^2/2 + O(n)$ edges [15]. At the first level, we have $\lfloor n/2 \rfloor$ gates with the ith gate G_i (which also denotes the output of the gate i) computing the function $\operatorname{sgn}(\sum_{j=1}^{n} 2b_j - 4i + 1)$ for $1 \leqslant i \leqslant \lfloor n/2 \rfloor$. In addition, we have $n+1$ "trivial" gates at the first level that will transmit the inputs $b_1, b_2, ..., b_n$ and the constant 1 to the gate at the second level. The output gate at the second level computes $\operatorname{sgn}(\sum_{j=1}^{\lfloor n/2 \rfloor} 4G_i - 2\sum_{i=1}^{n} b_i + 1)$. Note that the argument of the sign function is positive if the input $\vec{b}$ is odd and negative if the input $\vec{b}$ is even.

For greater depths, the above can be used as the basis in a recursive construction. To construct a circuit of even depth d, we partition the inputs into $\sqrt{n}$ sets of size $\sqrt{n}$. We recursively construct depth $d-2$ circuits to compute the parity of each of the sets and combine the outputs using a depth-2 circuit. For $d = o(\log \log n)$, the number of edges at the lowest level of recursion is $O(n^{1+1/2^{d/2}})$ which dominates the sum of all the edges at other levels. This is a special case of a construction by Beame *et al.* [4] which, for any symmetric function, yields a similar size–depth trade-off. In the case of the parity function, by a more careful construction, we get:

THEOREM 1. *For* $d \geqslant 2$ *and* $d = O(\log \log n)$, *there exists a depth-*d *threshold circuit with* $O(n^{1+1/\Theta(\phi^d)})$ *edges and* $O(n)$ *nodes that computes a parity function, where* $\phi = (1+\sqrt{5})/2$. (*All the weights in these circuits are* $O(\log n)$ *bits long.*)

Remark. In particular, from this theorem, we get an $O(\log \log n)$ depth threshold circuit for computing the parity function with node and edge complexities $O(n)$.

Proof. We first construct a depth-$(2d-2)$ circuit that computes parity on n variables using the divide-and-conquer technique. We then observe that this circuit can be collapsed to depth d. The depth-$(2d-2)$ circuit has $d-1$ stages of depth 2. The inputs to the first stage are the n boolean input variables. The outputs of stage i are the inputs for stage $i+1$. Stage $d-1$ has a single output which is the output of the circuit. Stage i has n_i inputs and n_{i+1} outputs. The inputs to stage i are partitioned into n_{i+1} groups of size n_i/n_{i+1}. The ith stage consists of n_{i+1} depth-2 circuits, each computing the parity of one of the groups using the standard construction. We will specify the integer parameters $n = n_1 > n_2 > \cdots > n_{d-1} > n_d = 1$ later.

The standard depth-2 construction has the key property that the weighted sum of the inputs to the second level is always ± 1, which can be converted to a boolean value by a simple linear transformation. Hence, no further thresholding is necessary. This enables us to eliminate all nodes at even levels except the input and output nodes in the following way. Consider a node u at an even level with inputs $y_{v_1}, ..., y_{v_k}$ coming from nodes $v_1, v_2, ..., v_k$ at the preceding level. Let $w_1, ..., w_k$ be the weights on these edges. Let $e = (u, u')$ be an edge from the node u to the node u' at the next higher level and let w be the weight on e. Since the weighted sum $\sum_{j=1}^{k} w_j y_{v_j}$ is always boolean, we can directly feed the inputs y_{v_j} to the gate u' with weights ww_j. If such a process is carried out for every outgoing edge of u, the node u can be eliminated. Thus we can eliminate all even levels and obtain a depth d circuit from our depth-$(2d-2)$ circuit.

We now select the parameters n_i to minimize the number of edges in the corresponding collapsed circuit. For notational convenience, we define $n_0 = n$ and $n_{d+1} = 1$. Note that, in the standard depth-2 construction for the parity of k variables, the number of edges feeding into the first level gates is at most k^2 and the number of edges feeding into the second level gate is at most $3k/2$. Thus in the uncollapsed circuit the number of edges from nodes at level $2i-2$ to nodes at level $2i-1$ is at most $n_{i+1}(n_i/n_{i+1})^2 = n_i^2/n_{i+1}$. In the collapsed circuit, each such edge is replaced by $3n_{i-1}/2n_i$ edges and level $2i-2$ is eliminated. The $(2i-1)$st level of the uncollapsed circuit becomes level i of the collapsed circuit and the total number of edges into that level is $3n_{i-1}n_i/2n_{i+1}$ for $i = 1, ..., d$. We now select n_i so that, for $i = 2, ..., d-1$, the number of edges feeding into level i gates from level $i-1$ gates is twice that of the number of edges feeding into level $i-1$ gates from level $i-2$ gates; that is,

$$2n_{i-1}n_i/n_{i+1} = n_{i-2}n_{i-1}/n_i.$$

This gives us the recurrence

$$n_i = 2n_{i-1}^2/n_{i-3}$$

for $i \geqslant 3$. After logarithms are taken, this recurrence becomes a Fibonacci recurrence. Using the boundary conditions, we obtain the total number of edges in the circuit to be $n^{1+\Theta(\phi^{-d})}$ where $\phi = (1+\sqrt{5})/2$, the golden ratio. ∎

Our inability to improve this construction suggests the following conjecture:

Conjecture 1. There exists a constant $c > 1$ such that any depth d circuit that computes parity requires $\Omega(n^{1+c^{-d}})$ edges.

This conjecture, if true, would imply that any depth $d = o(\log\log n)$ circuit that computes parity has a superlinear number of edges. As a consequence of the approximation results in Section 6, we obtain lower bounds on the edge complexity of depth-2 and depth-3 threshold circuits for parity. We begin with some particular properties of depth-2 threshold circuits.

4. Properties of Depth-2 Parity Circuits

We first establish a relationship between the node and the edge complexities of depth-2 threshold circuits that compute parity. Let p be any parity function on n inputs.

LEMMA 1. *For depth-2 threshold circuits,* $(2N_p(n) + nN_p(n-1) - n - 2)/2 \leqslant E_p(n) \leqslant (n+2)\,N_p(n) - n - 2$.

Proof. The upper bound follows easily from elementary graph-theoretic considerations. For the lower bound, we show that the outdegree of each input variable b_i is at least $N_p(n-1)/2$. Let T be a depth-2 threshold circuit that computes a parity function on n inputs. Fix an input b_i and let $\text{out}_T(i)$ denotes its outdegree. From T we derive a threshold circuit T' that computes a parity function on $n-1$ inputs whose node complexity is bounded from above by $2\,\text{out}_T(i) + 1$.

Let T_1 be the circuit obtained from T by merging the source node labelled b_i with the source labelled 1, i.e., assigning the value 1 to the input b_i. Let T_0 be the circuit obtained by deleting the source node labelled b_i, together with its incident edges, and negating the weight of each edge entering the output node. Now, let T'' be the circuit obtained by coalescing the output nodes and the corresponding source nodes of T_0 and T_1. Since

T_0 and T_1 compute the same parity function on $n-1$ inputs, it is not difficult to see that T'' computes that function as well. Note that each level-1 node of T that has no incoming edge from b_i corresponds to a pair of nodes in T'' that compute the same function but whose outgoing edges have weights summing to 0. Therefore, the contribution of each such pair to the output node of T'' is always 0. Hence we can remove all such nodes from T'' to obtain a circuit T' whose node complexity is bounded from above by $2\,\mathrm{out}_T(i)+1$. From this, we get the lower bound $(N_p(n-1)-1)/2$ on the outdegree of each non-constant source node, which in turn gives the lower bound on $E_p(n)$. ∎

A Geometric View. Each gate in a threshold circuit can be interpreted as an affine hyperplane. With this interpretation, it is not hard to prove:

LEMMA 2. *In a depth-2 circuit for parity, the set of hyperplanes associated with the nodes at level 1 must intersect every edge of the n-dimensional unit hypercube.*

This observation suggests the following problem.

Problem 1. What is the minimal number of hyperplanes required to intersect all the edges of the unit hypercube?

This problem appears in [17, 2] and it was conjectured that this number is n. An unpublished counterexample by M. Paterson is mentioned in [17]. By Lemma 1, a lower bound of $f(n)$ on the number of hyperplanes would imply an $\Omega(nf(n))$ lower bound on the edge complexity for computing parity with a depth-2 circuit.

In the construction of the previous section, observe that the circuits are *restricted* in the sense that the weights (except the ones that correspond to constant input 1) of all level 1 gates are nonnegative. Therefore the associated hyperplanes have nonnegative coefficients. It is not hard to see that at least n such hyperplanes are needed to cut all of the edges of the hypercube (no two of the n edges $(0^j1^{n-j}, 0^{j+1}1^{n-j-1})$ can be cut simultaneously by a hyperplane with nonnegative coefficients) and thus we have:

THEOREM 2. *Any restricted depth-2 threshold circuit that computes parity has edge complexity of* $n^2/2+\Omega(n)$.

In the case where the weights are not restricted, the best lower bound known on the number of hyperplanes needed to cut all edges of the hypercube is $\Omega(\sqrt{n})$, which gives an $\Omega(n^{3/2})$ bound on the edge complexity of parity with depth-2 circuits. The $\Omega(\sqrt{n})$ lower bound can be obtained by showing that any one single hyperplane can cut at most $O(\sqrt{n}2^n)$ edges of the cube [17]. We believe that this can be significantly improved but have been unable to do so by this technique. Instead, the lower bounds promised in the introduction are based on an analytic approach, which use approximation by rational functions together with a degree argument.

5. Real Representations of Boolean Functions

We consider the representation of boolean functions by real functions. In particular, we will be interested in the real functions that are polynomials or rational functions in their variables.

A real *polynomial* $f(x_1, x_2, ..., x_n)$ in $x_1, x_2, ..., x_n$ has the general form

$$f(x_1, x_2, ..., x_n) = \sum_{d_1 \geqslant 0, ..., d_n \geqslant 0} \alpha_{d_1 \cdots d_n} \prod_i x_i^{d_i}$$

such that only finitely many $\alpha_{d_1 \cdots d_n}$'s are non-zero. The degree of f is defined by

$$\deg(f) = \max_{d_1, ..., d_n} \left\{ \sum_i d_i : \alpha_{d_1, ..., d_n} \neq 0 \right\}.$$

A polynomial $f(x_1, x_2, ..., x_n)$ is *multilinear* if, for each non-zero coefficient $\alpha_{d_1, ..., d_n}$, $0 \leqslant d_i \leqslant 1$ for all i. A *rational* function is a ratio of two polynomials. The degree of a rational function is the maximum of the degrees of the corresponding polynomials.

We show that the boolean function computed by a threshold circuit can be represented by a polynomial or a rational function whose degree is a function of the size of the circuit. On the other hand, we show that any polynomial or rational representation of parity function requires high degree thus obtaining lower bounds on the size of threshold circuits. To carry out this program, we introduce three notions of representation. Let $g: \{0, 1\}^n \to \{0, 1\}$ be a boolean function and $f: \mathbf{R}^n \to \mathbf{R}$ be a real function.

We say f *exactly represents* g if, for all $\vec{b} \in \{0, 1\}^n$, $f(\vec{b}) = g(\vec{b})$. We have the following well-known fact regarding the exact representation of boolean functions by multilinear polynomials, which is easily proved (for example) by induction on the number of variables.

Fact 1. For every boolean function g on n variables, there is a multilinear real polynomial f of degree at most n which exactly represents g.

The other two notions of representation are weaker. We say that f *sign-represents* g if $g(\vec{b}) = \mathrm{sgn}(f(\vec{b}))$ for all boolean vectors b. For a real-valued function h on n variables, let $\|h\|_\infty = \max_{\vec{b} \in \{0, 1\}^n} |h(\vec{b})|$. We say f *approximates* g *within error* ε if $\|f - g\|_\infty \leqslant \varepsilon$. Observe that if f exactly represents g then f sign-represents g and if f approximates g with error $<1/2$, then $f-1/2$ sign-represents g.

We now show that the sign-representation of parity requires high degree.

LEMMA 3. *If a real polynomial* $f(x_1, x_2, ..., x_n)$ *sign-represents a parity function, then* $\deg(f) \geqslant n$.

Proof. Without loss of generality, we assume that f is multilinear since we are only concerned with the behavior of f on boolean inputs. We use induction on n to prove that $\deg(f)$ is at least n.

Clearly, the statement is true for $n=1$. For some $n \geqslant 2$, assume that $f(x_1, x_2, ..., x_n)$ sign-represents the function p_e on n variables. (The other case can be handled similarly.) Since f is multilinear, it can be written as

$$f(x_1, x_2, ..., x_n) = x_1 f_1(x_2, ..., x_n) + f_0(x_2, ..., x_n)$$

such that $f_0(x_2, x_3, ..., x_n)$ sign-represents p_e on $n-1$ variables and $f_1(x_2, x_3, ..., x_n) + f_0(x_2, x_3, ..., x_n)$ sign-represents p_0 on $n-1$ variables.

It follows that $f_1(x_2, x_3, ..., x_n)$ sign-represents p_o on $n-1$ variables. Observe that the degree of f_1 is one less than that of f. By the induction hypothesis, we have that $\deg(f_1) \geqslant n-1$. Hence, $\deg(f) \geqslant n$. ∎

COROLLARY 1. *If a rational function* $f(x_1, ..., x_n) = h_1(x_1, ..., x_n)/h_2(x_1, ..., x_n)$ *sign-represents a parity function on n variables, then the sum of the degrees of the polynomials* h_1 *and* h_2 *is at least n.*

Remark. This lemma can be used to give a simpler proof of the result of Alon *et al.* [2] concerning the minimal cardinality of a family of ± 1 vectors that is to *cover* all the vertices of a ± 1 n-dimensional hypercube. More precisely, they considered the following problem due to Knuth. Let $K(n, d)$ denote the minimal k for which there exists ± 1 vectors $\vec{v}_1, \vec{v}_2, ..., \vec{v}_k$ of dimension n such that for any ± 1 vector $\vec{u}$ of dimension n, there is an i, $1 \leqslant i \leqslant k$, such that $|\vec{v}_i \cdot \vec{u}| \leqslant d$, where $\vec{v} \cdot \vec{u}$ denotes the inner product of two vectors. Knuth's construction shows that $K(n, d) \leqslant \lceil n/(d+1) \rceil$ for $n \equiv d \pmod 2$. Alon and others proved a matching lower bound using elementary linear algebra. This lower bound can also be obtained from Lemma 3, by noting that a family of k vectors that satisfies Knuth's criterion can be used to obtain a polynomial of degree at most $k(d+1)$ that sign-represents parity. In the following, we give the proof.

Let $d \equiv n \pmod 2$. Assume that the ± 1 vectors $\vec{v}_1, ..., \vec{v}_k$ have the property that, for every point $\vec{x}$ of the ± 1 hypercube, there is some i such that $|\vec{v}_i \cdot \vec{x}| \leqslant d$. For $1 \leqslant i \leqslant k$ and $0 \leqslant j \leqslant d$, define the linear forms $h_{ij}(\vec{x}) = \vec{v}_i \cdot \vec{x} - d + 2j$. Since the dot product of two ± 1 vectors of length n is congruent to n (mod 2), it follows that every ± 1 vector $\vec{x}$ satisfies one of the equations $h_{ij}(\vec{x}) = 0$.

Let us define the parity of a ± 1 vector to be the parity of the number of -1's in it. It is easy to see that Lemma 3 applies also when the variables take the values ± 1. Observe that for each i and j, all ± 1 vectors $\vec{x}$ that satisfy $h_{ij}(\vec{x}) = 0$ have the same parity. Thus we can classify the h_{ij} as even or odd according to the parity of the vectors that satisfy the equation $h_{ij} = 0$. Let $f(\vec{x}) = [\prod h_{ij}]^2 - 1/2$, where the product is taken over all even h_{ij} and $g(\vec{x} = [\prod h_{ij}]^2 - 1/2$, where the product is taken over all odd h_{ij}. It is easy to see that both the polynomials f and g sign-represent the parity functions on n variables and so by Lemma 3 each has degree at least n. On the other hand, the sum of their degrees is $2k(d+1)$. Hence, $k(d+1) \geqslant n$.

6. Rational Approximation of Threshold Circuits

In this section, we prove our main approximation results for depth-2 and depth-3 circuits and apply them to obtain lower bounds on the edge complexity of such circuits that compute parity.

6.1. *Depth-2 Circuits*

THEOREM 3. *Let T be a depth-2 threshold circuit with n variables and at most E edges. Let $u_{\max}$ denote the maximum norm of its gates. Then*

1. *The boolean function computed by T can be sign-represented by a rational function whose degree is at most $4\sqrt{E}\ln(6u_{\max})$,*

2. *For any $\delta \in (0, 1/2)$, the boolean function computed by T can be approximated with error at most δ by a rational function whose degree is at most $4\sqrt{E}\ln^3(24u_{\max}/\delta)$.*

From the first part of this theorem and Corollary 1, we obtain the result claimed in the introduction.

COROLLARY 2. *Any depth-2 threshold circuit with polynomially bounded weights that computes parity has $\Omega(n^2/\ln^2 n)$ edges.*

In the rest of this section, we present a proof of Theorem 3. Our main tool is a construction due to Newman of low-degree rational functions that approximate the function $|x|$ on the interval $[-1, 1]$. We say that the function $r(x)$ approximates $|x|$ over the interval $[-1, 1]$ within error ε if $\sup_{x \in [-1,1]} |r(x) - |x|| \leqslant \varepsilon$.

For integer $k \geq 1$, let $\xi_k = e^{-1/\sqrt{k}}$, and

$$p_k(x) = \prod_{i=0}^{k-1} (x + \xi^i),$$

$$R_k(x) = x \frac{p(x) - p(-x)}{p(x) + p(-x)}$$

$$S_k(x) = \frac{p(x)}{p(x) + p(-x)}.$$

Note that $R_k(x)$ and $S_k(x)$ are rational functions of degree at most $k+1$.

THEOREM 4 (Newman [16]). *In the interval* $[-1, 1]$, $R_k(x)$ *approximates* $|x|$ *within error* $3e^{-\sqrt{k}}$.

This result is considered to be the starting point of the modern theory of rational approximation. It is worth mentioning that no real polynomial of degree k can approximate $|x|$ within error $o(1/k)$.

This theorem enables us to approximate the sgn function.

COROLLARY 3. *For* $0 < \delta < 1$, *the function* $\operatorname{sgn}(x)$ *can be approximated by* $S_k(x)$ *over* $H_\delta = [-1, -\delta] \cup [\delta, 1]$ *within error* $3e^{-\sqrt{k}}/2\delta$.

Proof. Note that $\operatorname{sgn}(x) = (1 + |x|/x)/2$ for $x \neq 0$. Therefore, $|S_k(x) - \operatorname{sgn}(x)| = |(R_k(x) - |x|)/2x| \leq 3e^{-\sqrt{k}}/2\delta$ over H_δ. ∎

We use this result to show that the function computed by any threshold gate can be well approximated by a rational function of small degree. Let G be a threshold gate on n inputs with weight vector $\vec{w}$ and let $f_G(\vec{b}) = \operatorname{sgn}(\sum_{i=1}^n w_i b_i)$ denote the boolean function computed by G. Let $u = \sum_{i=1}^n |w_i|$ be the norm of G. We then have the following approximation lemma for G.

LEMMA 4. *For each integer* $k \geq 1$, *there exists a real function* $g_k(\vec{b})$ *of degree* k *that approximates the boolean function* f_G *on* $\{0, 1\}^n$ *within error* $6ue^{-\sqrt{k}}$.

Proof. Let $s(\vec{b}) = (\sum_{i=1}^n w_i b_i - 1/2)/2u$. Since w_i are integers, we have $1/4u \leq |s(\vec{b})| \leq 1$, that is, $s(\vec{b}) \in H_{1/4u}$ for all $\vec{b} \in \{0, 1\}^n$. Thus by the corollary, $\operatorname{sgn}(s(\vec{b}))$ is approximated by the degree-k rational function $S_k(s(\vec{b}))$ within error $6ue^{-\sqrt{k}}$. ∎

We now prove the theorem stated in the beginning of the section.

Proof of Theorem 3. Let m be the number of level-1 gates and let $h_i(\vec{b})$ denote the boolean function computed by the ith gate for $1 \leqslant i \leqslant m$. Let $g(\vec{b}) = \sum_{i=1}^{m} w_i h_i(\vec{b})$, where $w_1, w_2, ..., w_m$ are the weights corresponding to the output gate at level-2. Then $f_T(\vec{b}) = \text{sgn}(g(\vec{b}))$.

To prove the first part of the theorem, we need to construct a low-degree rational function $r(x_1, ..., x_n)$ that sign-represents the function f_T. To do this we first construct a low-degree rational function approximation for g. Each of the boolean functions h_i can be exactly represented by a multilinear polynomial p_i whose degree is at most the fan-in of the associated gate. Also, by Lemma 4, each h_i can be approximated within error $6u_{\max} e^{-\sqrt{k}}$ by a degree k multivariate rational function h_i'. We select a positive real number d' and a positive integer k (to be specified later) and substitute into g for each h_i as follows: substitute p_i for h_i if h_i comes from a gate of fan-in at most d' and otherwise substitute h_i'. Call the resulting function $g'(\vec{x})$. Note that the number of terms for which h_i' is used is at most E/d' since the number of edges is bounded above by E. Thus if we express g' as a ratio of polynomials, the degree of the denominator is at most $(E/d')k$ and the degree of the numerator is at most $d' + (E/d')k$. Selecting $d' = \sqrt{Ek}$ guarantees that the degree of g' is at most $2\sqrt{Ek}$. Also, for $\vec{b} \in \{0, 1\}^n$,

$$\begin{aligned} |g'(\vec{b}) - g(\vec{b})| &\leqslant \sum_{i=1}^{m} |w_i| \, |h_i'(\vec{b}) - h_i(\vec{b})| \\ &\leqslant 6u_{\max} e^{-\sqrt{k}} \sum |w_i| \\ &\leqslant 6u_{\max}^2 e^{-\sqrt{k}}. \end{aligned}$$

If we select k so that this error is within 1/4 ($k = \lfloor 4 \ln^2(6u_{\max}) \rfloor$ will do), the function $r(\vec{x}) = g'(\vec{x}) - 1/2$ satisfies $g(\vec{b}) \geqslant r(\vec{b}) > g(\vec{b}) - 1$ and thus $\text{sgn}(r(\vec{b})) = \text{sgn}(g(\vec{b})) = f(\vec{b})$. The degree of r is then at most $4\sqrt{E}\ln(6u_{\max})$, as required to prove the first part of the theorem.

To prove the second part of the theorem, we note that on boolean inputs $g(\vec{b})$ is integer valued and bounded in absolute value by $u_{\max}$ and $g'(\vec{b})$ differs from $g(\vec{b})$ by at most 1/4, and thus the function $t(\vec{b}) = (g'(\vec{b}) - 1/2)/2u_{\max}$ has absolute value in the range $[1/(8u_{\max}), 1]$. Thus by Corollary 3, the function $S_j(t(\vec{x}))$ approximates $\text{sgn}(t(\vec{x}))$ with error at most $12u_{\max} e^{-\sqrt{j}}$. Choosing j to be $\lceil \ln^2(24u_{\max}/\delta) \rceil$ yields an error of approximation at most δ. The degree of the resulting function is j times the degree of t, which is bounded above by $4\sqrt{E}\ln^3(24u_{\max}/\delta)$. ∎

6.2. *Depth*-3 *Circuits*

In this section, we obtain a result for depth-3 circuits that is analogous to Theorem 3.

THEOREM 5. *Let T be a depth-3 threshold circuit with n variables and E edges. Let $u_{\max}$ denote the maximum norm of its gates. Then the boolean function computed by T can be sign-represented by a rational function whose degree is at most $24E^{5/6}\ln^{5/3}(7Eu_{\max})$.*

Once again we use Corollary 1 to derive the following consequence of Theorem 5.

COROLLARY 4. *Any depth-3 threshold circuit with polynomially bounded weights requires $\Omega(n^{1.2}/\ln^{5/3} n)$ edges for computing parity.*

Proof of Theorem 5. Let N_1 and N_2 denote the number of gates at levels 1 and 2 respectively. Let $x_1, ..., x_n$ denote the inputs to the circuit T (feeding into the level-1 gates). Let $y_1, ..., y_{N_1}$ and $z_1, ..., z_{N_2}$ be the outputs of level-1 and level-2 gates respectively. We can view each y_i as a function of $\{x_j\}$, and each z_i as a function either of $\{x_j\}$ or of $\{y_j\}$. Let $w_1, ..., w_{N_2}$ be the weights of the wires that feed into the gate at level-3 and let $g(x_1, ..., x_n) = \sum_{i=1}^{N_2} w_i z_i$. The output of the circuit is $f_T(x_1, ..., x_n) = \operatorname{sgn}(g(x_1, ..., x_n))$. We will construct a low-degree rational function $r(x_1, ..., x_n)$ to approximate g within error $<1/2$ over boolean inputs; since g is integer-valued, $r - 1/2$ will be a sign-representation of g and hence f_T.

To this end, we construct the rational functions $s_1, ..., s_{N_2}$ such that each s_i approximates z_i within error $1/2u_{\max}$ so that $r = \sum_{i=1}^{N_2} w_i s_i$ approximates $g = \sum_{i=1}^{N_2} w_i z_i$ within error $<1/2$. By part 2 of Theorem 3 with $\delta = 1/2u_{\max}$, we can approximate z_i within error δ by a rational function of degree at most $4\sqrt{E}\ln^3(48u_{\max}^2) \leq 32\sqrt{E}\ln^3(7u_{\max})$. If we were to take s_i to be this approximation, the resulting rational function $r(x_1, ..., x_n)$ would have degree $32N_2\sqrt{E}\ln^3(7u_{\max})$, which is too large for our purposes. To get a lower degree approximation, it will be useful to distinguish the z_i and the y_i according to the fan-in of their associated gates. For a positive integer d, define $I_d = \{i \mid z_i \text{ has fan-in at most } d\}$ and $J_d = \{j \mid y_j \text{ has fan-in at most } d\}$.

Let d_1 be a positive real number to be specified later. For those $i \notin I_{d_1}$, we will take s_i to be the rational function of degree at most $32\sqrt{E}\ln^3(7u_{\max})$ mentioned above. Since there are at most E/d_1 such i, the contribution to the degree of r is at most $32E^{3/2}\ln^3(7u_{\max})/d_1$. For $i \in I_{d_1}$, we use the following lemma to construct s_i.

LEMMA 5. *Let d be an arbitrary positive real number. There exist rational functions s_i for $i \in I_d$ such that*

1. *s_i approximates z_i within error $1/2u_{\max}$*
2. *all the s_i's have the same denominator*
3. *all s_i's have degree at most $6\sqrt{Ed}\ln(7Eu_{\max})$.*

Since all the s_i have the same denominator, the contribution to the degree of $r(x_1, ..., x_n)$ is at most $6\sqrt{Ed_1}\ln(7Eu_{\max})$. Hence the total degree of r is at most $32E^{3/2}\ln^3(7u_{\max})/d_1 + 6\sqrt{Ed_1}\ln(7Eu_{\max})$. We now select $d_1 = (32E/6)^{2/3}\ln^{4/3}(7u_{\max})$ to obtain an upper bound on the degree of r of $24E^{5/6}\ln^{5/3}(7Eu_{\max})$ as required to prove the theorem. ∎

It remains to prove Lemma 5.

Proof. For $i \in I_d$, each z_i can be viewed as a function of at most d of the boolean values y_j and thus can be represented exactly by a multilinear polynomial p_i in y_j of degree at most d. We next approximate each of the y_j by a rational function $q_j(x_1, ..., x_n)$ and obtain the approximation s_i to z_i by replacing each y_j by q_j in p_i.

Let d' be a positive real number and k be a positive integer (to be specified later). For $j \in J_{d'}$, we define q_j to be the multilinear polynomial exactly representing y_j; note that the degree of q_j is at most d'. For $j \notin J_{d'}$, we define q_j to be the degree k rational approximation to y_j obtained by applying Lemma 4; the error of this approximation is bounded above by $6u_{\max}e^{-\sqrt{k}}$.

We now need to analyze the degree and error of the s_i obtained in this way. Let q be the product of the denominators of the rational approximations q_j. The degree of q is at most kE/d' since there are at most E/d' y_j whose fan-in exceeds d' and the other y_j are exactly represented by polynomials. Since each term of p_i is a constant multiple of the product of at most d distinct y_j's, after q_j is substituted for y_j, the denominator of the term is a divisor of q and thus the term can be expressed with q as its denominator. The numerator of the result will have degree at most $dd' + kE/d'$ (which is at least the degree of the denominator). Choosing $d' = \sqrt{kE/d}$ to minimize this expression, we obtain an upper bound on the degree of s_i of $2\sqrt{kEd}$.

To analyze the error introduced by approximating z_i by s_i, we need to consider the error introduced by substituting q_j for y_j for $j \notin J_{d'}$ in the multilinear polynomial p_i. Note that for $j \in J_{d'}$, q_j is an exact representation of y_j. The following lemma gives a bound on the norm of f when its variables are perturbed. Recall that for any real function f on n variables, the norm of f is $\|f\|_\infty = \sup_{\vec{b} \in \{0,1\}^n} |f(\vec{b})|$.

LEMMA 6. *Let $f(\vec{x})$ be a multilinear polynomial on n variables. For any $\vec{b} \in \{0, 1\}^n$ and $\vec{a} \in \mathbf{R}^n$,*

$$|f(\vec{b}+\vec{a}) - f(\vec{b})| \leq \left[\prod_{i=1}^{n} (1 + 2|a_i|) - 1\right] \|f\|_\infty.$$

Proof. Define for $0 \leq i \leq n$ the functions $f_i(x_{i+1}, ..., x_n) = f(a_1 + b_1, ..., a_i + b_i, x_{i+1}, ..., x_n)$ and $g_i(x_{i+1}, ..., x_n) = f(b_1, ..., b_i, x_{i+1}, ..., x_n)$.

Note that each f_i and g_i are multilinear, $f_0 = g_0 = f$, $f_n = f(\vec{b} + \vec{a})$, and $g_n = f(\vec{b})$.

We prove that $\|f_i - g_i\|_\infty \leqslant \|f\|_\infty \left[\prod_{j=1}^{i} (1 + 2|a_j|) - 1\right]$ by induction on i. The basis case $i = 0$ is trivial. For the induction step, assume $0 \leqslant i < n$ and that $\|f_i - g_i\|_\infty \leqslant \|f\|_\infty \left[\prod_{j=1}^{i} (1 + 2|a_j|) - 1\right]$. It suffices to show that $\|f_{i+1} - g_{i+1}\|_\infty \leqslant (1 + 2|a_{i+1}|)\, \|f_i - g_i\|_\infty + 2|a_{i+1}|\, \|f\|_\infty$. By the triangle inequality of the norm, we have

$$\begin{aligned}\|f_{i+1} - g_{i+1}\|_\infty \leqslant\ & \|f_{i+1}(x_{i+2}, ..., x_n) - f_i(b_{i+1}, x_{i+2}, ..., x_n)\| \\ & + \|f_i(b_{i+1}, x_{i+2}, ..., x_n) - g_{i+1}(x_{i+2}, ..., x_n)\|\end{aligned}$$

Note that $g_{i+1}(x_{i+2}, ..., x_n) = g_i(b_{i+1}, x_{i+2}, ..., x_n)$ and thus the second summand is at most $\|f_i - g_i\|_\infty$. To bound the first summand, note that since f_i is linear in x_{i+1}, we can write $f_i(x_{i+1}, ..., x_n) = x_{i+1} h_i^1(x_{i+2}, ..., x_n) + h_i^2(x_{i+2}, ..., x_n)$.

Also $f_{i+1}(x_{i+2}, ..., x_n) = f_i(a_{i+1} + b_{i+1}, x_{i+2}, ..., x_n)$. Thus, the first summand is equal to $|a_{i+1}|\, \|h_i^1\|_\infty$, which is bounded above by $2|a_{i+1}|\, \|f_i\|_\infty$, since $h_i^1(x_{i+2}, ..., x_n) = f_i(1, x_{i+2}, ..., x_n) - f(0, x_{i+2}, ..., x_n)$. Summarizing these bounds, applying the triangle inequality once again, and noting that $\|g_i\|_\infty \leqslant \|f\|_\infty$, we get

$$\begin{aligned}\|f_{i+1} - g_{i+1}\|_\infty &\leqslant 2|a_{i+1}|\, \|f_i\|_\infty + \|f_i - g_i\|_\infty \\ &\leqslant 2|a_{i+1}|\, (\|g_i\|_\infty + \|f_i - g_i\|_\infty) + \|f_i - g_i\|_\infty \\ &\leqslant 2|a_{i+1}|\, \|f\|_\infty + (1 + 2|a_{i+1}|)\, \|f_i - g_i\|_\infty,\end{aligned}$$

as required to prove the induction step. ∎

When we substitute q_j for y_j into p_i to obtain s_i, all but at most $\sqrt{kE/d}$ of the q_j are exactly equal to y_j, and each remaining q_j differs from the corresponding y_j by at most $6u_{\max}\, e^{-\sqrt{k}}$. Thus, applying the above lemma, we obtain that s_i approximates z_i within error $[(1 + 6u_{\max}\, e^{-\sqrt{k}})^{\sqrt{kE/d}} - 1]$. We can select $k = \lfloor 9 \ln^2(7Eu_{\max}) \rfloor$ so that this is less than $1/2u_{\max}$. Thus we obtain an upper bound of $6\sqrt{Ed}\ln(7Eu_{\max})$ on the degree of s_i. This completes the proof of Lemma 5. ∎

RECEIVED June 5, 1991; FINAL MANUSCRIPT RECEIVED April 10, 1992

REFERENCES

1. AJTAI, M. (1983), Σ_1^1-formulae on finite structures, *Ann. Pure Appl. Logic* **24**, 1–48.
2. ALON, N., BERGMANN, E. E., COPPERSMITH, D., AND ODLYZKO, A. M. (1988), Balancing sets of vectors, *IEEE Trans. Inform. Theory* **34**, 128–130.

3. ASPNES, J., BEIGEL, R., FURST, M., AND RUDICH, S. (1991), The expressive power of voting polynomials, *Combinatorica*, to appear.
4. BEAME, P., BRISSON, E., AND LADNER, R. (1992), The Complexity of Computing Symmetric Functions using Threshold Circuits, *Theoretical Computer Science* **100**, 253–265.
5. BEAME, P., COOK, S. A., AND HOOVER, H. J. (1986), Log depth circuits for division and related problems, *SIAM J. Comput.* **15**, 994–1003.
6. BEIGEL, R., REINGOLD, N., AND SPIELMAN, D. (1994), *PP* is closed under intersection, *J. Comput. System Sci.*, to appear.
7. BOPPANA, R., AND SIPSER, M. (1990), The complexity of finite functions, *in* "The Handbook of Theoretical Computer Science," Vol. A, Elsevier, Amsterdam.
8. BRUCK, J. (1990), Harmonic analysis of polynomial threshold functions, *SIAM J. Discrete Math.* 168–177.
9. BRUCK, J., AND SMOLENSKY, R. (1992), Polynomial threshold functions, AC^0 functions and spectral norms, *SIAM J. Comput.* **21**, 33–42.
10. CHANDRA, A., AND STOCKMEYER, L. (1984), Constant depth reducibility, *SIAM J. Comput.* **13**, 423–439.
11. FURST, M., SAXE, J. B., AND SIPSER, M. (1984), Parity, circuits, and the polynomial time hierarchy, *Math. Systems Theory* **17**, 13–28.
12. HAJNAL, A., MAASS, W., PUDLÁK, P., SZEGEDY, M., AND TURÁN, G. (1993), Threshold circuits of bounded depth, *J. Comput. System Sci.* **46**(2), 129–154.
13. HINTON, G. E. (1987), "Connectionist learning procedures," CMU-CS-87-115.
14. MINSKY, M., AND PAPERT, S. A. (1969), "Perceptrons," MIT Press, Cambridge, MA.
15. MUROGA, S. (1971), "Threshold Logic and its Applications," Wiley–Interscience, New York.
16. NEWMAN, D. J. (1964), Rational approximation to $|x|$, *Michigan Math. J.* **11**, 11–14.
17. O'NEIL, P. E. (1971/1972), Hyperplanes cuts of an *n*-cube, *Discrete Math.* **1**, 193–195.
18. PARBERRY, I., AND SCHNITGER, G. (1988), Parallel computation with threshold functions, *J. Comput. System Sci.* **36**, 278–302.
19. RAZBOROV, A. A. (1986), Lower bounds on the size of bounded depth networks over a complete basis with logical addition, *Math. Zametki* **41**, 598–607 [in Russian]; English translation in *Math. Notes* **41**, 333–338.
20. REIF, J. (1992), On threshold circuits and polynomial computations, *SIAM J. Comput.* **21**, 896–908.
21. SIU, K., ROYCHOWDURY, V., AND KAILATH, T. (1994). Rational Approximation Techniques for Analysis of Neural Networks, *IEEE Trans. Inform. Theory* **40**, No. 2.
22. SMOLENSKY, R. (1987), Algebraic methods in the theory of lower bounds for Boolean circuit complexity, *in* "Proceedings of 19th Annual ACM Symposium on Theory of Computing," pp. 77–82.
23. SZEGEDY, M. (1989), "Algebraic Method in Lower Bounds for Computational Models with Limited Communication," Ph.D. Dissertation, University of Chicago.
24. WEGENER, I. (1990), "The Complexity of the Parity Function in Unbounded Fan-in Unbounded Depth Circuits," Technical Report, Universität Dortmund.
25. YAO, A. C-C. (1989), Circuits and local computation, *in* "Proceedings of the 21st Annual ACM Symposium on Theory of Computing," pp. 186–196.
26. YAO, A. C-C. (1990), On ACC and Threshold Circuits, *in* "31st IEEE Symposium on Foundations of Computer Science," pp. 619–627.