More on Computing Boolean Functions by Sparse Real Polynomials and related types of Threshold Circuits

Matthias Krause
Lehrstuhl Informatik II
Universität Dortmund
D 44221 Dortmund
e–mail krause@daedalus.
informatik.uni-dortmund.de

Pavel Pudlák*
Mathematical Institute
Academy of Sciences
of Czech Republic
Žitná 25
CZ 115 67 Praha 1
e–mail pudlak@mbox.cesnet.cz

Abstract

In this paper we investigate the computational power of threshold–AND circuits versus threshold–XOR circuits. Starting from the observation that small weight threshold–AND circuits can be simulated by small weight threshold–XOR circuits we pose the question whether a

*Supported by the grant A1019602 of the Academy of Sciences of the Czech Republic. A major part of the results were obtained while the author was supported by the Humboldt–Fellowship at the University Dortmund.

similar simulation exists for small size unbounded weight circuits. The answer to this question, and this is the main result of this paper, is no. We present a function with small threshold–AND circuits for which all threshold–XOR circuits have exponentially many nodes. This answers a fundamental question on separating subsets of the hypercube by hyperplanes induced by sparse real polynomials: Is it generally better to choose domain $\{1,-1\}^n$ or are there functions having more compact polynomials over $\{0,1\}^n$? We prove our result by a new lower bound argument which, we hope, contributes to a better understanding of the computational limitations of small depth threshold circuits. As a further result we show that there are $AC_{0,3}$–functions which do not have small threshold–AND circuits. This completes the picture obtained from a series of results [1][4][7][14] on computing AC_0–functions by small depth threshold circuits. The result can be generalized to any level of the AC_0–hierarchy, and, thus, illustrates the power of alternation with respect to unbounded weight threshold operations.

1 Introduction

In the last decade a lot of research has been done on the computational power of small depth circuits of unbounded fan-in over AND-, OR–, XOR–, modular and threshold gates (e.g. [21][22][8][23][12][1] to give at least a sample of the most important results). A large number of results achieved in this area shows that attempts to analyze the computational power of various architectures of small depth circuits often lead to another fundamental question in Boolean complexity theory: which resources (degree, number of monomials) do we need, if we want to represent a given Boolean function f by a multivariate polynomial over some appropriate domain? For example, an efficient way to prove a bound on the complexity of AC_0–functions is based on showing that their subfunctions with high probability can be computed by low degree polynomials over the Boolean semiring [13]. Another example is that all functions having polynomial size constant depth circuits over AND–, OR– and MOD_p–gates, p–prime (for short: $AC_0[p]$–functions) can be computed by randomized low degree $\mathbb{Z}_p$–polynomials with bounded error [21][22]. This leads to the important result that for all $q \neq p$ prime, MOD_q is not an $AC_0[p]$–function [22]. Another interesting characterization of the complexity of AC_0–functions is that they can be efficiently approximated by low degree integer polynomials [4]. Moreover, using *Yao*'s result on simulating circuits over AND, OR, MOD_m by depth 3 threshold circuits [23], it has been shown in [11]that, for all natural m, $AC_0[m]$–functions can be written as the middle bit of some low degree integer polynomial.

In this paper we will show several results characterizing the computational power of threshold–AND– versus threshold–XOR circuits. Further we will answer some open questions concerning the computability of AC_0–functions by these types of circuits.

There is an obvious relation between threshold–AND–, resp. threshold–XOR circuits and the representation of Boolean functions by real polynomials with respect to the following mode of computation.

1.1 The Model

Let $f : X \longrightarrow \{0,1\}$ be a Boolean function over the domain $X \subseteq \mathbb{R}^n$ and $p : \mathbb{R}^n \longrightarrow \mathbb{R}$ be an n–variate real polynomial. We say that p is a *voting polynomial* for f w.r.t. domain X if for all inputs $x \in X$

$$f(x) = 1 \implies p(x) < 0 \text{ and } f(x) = 0 \implies p(x) > 0.$$

In other words, the hyperplane defined by $p(z) = 0$ has to separate the sets $f^{-1}(0)$ and $f^{-1}(1)$.

Usually, n-ary Boolean functions are defined over the domain $\{0,1\}^n$. However, in many cases there is a chance to get more compact voting polynomials if one switches to the domain $\{1,-1\}^n$ by replacing 0 by 1, and 1 by -1. One obvious witness for this phenomenon is the XOR function $p_{\bar{1}}(x_1, \ldots, x_n) = x_1 \oplus \ldots \oplus x_n$. The single monomial $z_1 \cdot \ldots \cdot z_n$ is obviously a voting polynomial for XOR_n w.r.t. the domain $\{1,-1\}^n$ and, in fact, it computes XOR exactly. On the other hand, *Goldmann* [9] showed:

Theorem 1 *Each polynomial computing the parity function in n variables over $\{0,1\}^n$ has exponentially many monomials.*

If the domain is $\{0,1\}^n$ then monomials correspond to AND–functions. But also over $\{1,-1\}^n$ we can find a sparse voting polynomial for AND, namely $z_1 + \ldots + z_n + n - \frac{1}{2}$.

The initial question for this paper is whether the expressive power of polynomials with respect to domain $\{1,-1\}^n$ is generally greater than with respect to $\{0,1\}^n$. Or, are there functions having compact voting polynomials w.r.t. $\{0,1\}^n$which need exponentially many monomials over $\{1,-1\}^n$? We will give an affirmative answer.

In order to be able to switch easily between the ranges $\{0,1\}$ and $\{1,-1\}$, we introduce the notation $r^* = (-1)^r$ for $r \in \{0,1\}$; we shall use it also for Boolean functions f and 0-1-vectors x (thus, e.g. f^* is defined by $f^*(x) = (-1)^{f(x)}$ and x^* by $(x_1^*, \ldots, x_n^*)$). The sign function is defined as usual, namely, $sgn(r)$ of a real number r is 1 if r is positive, it is -1 if r negative and $sgn(0) = 0$.

Consequently, a polynomial p is a voting polynomial for $f : \{0,1\}^n \longrightarrow \{0,1\}$ w.r.t. $\{0,1\}^n$ if for all $x \in \{0,1\}^n$

$$f^*(x) = sgn(p(x)),$$

and w.r.t. $\{1,-1\}^n$, if for all $x \in \{0,1\}^n$

$$f^*(x) = sgn(p(x^*)).$$

Clearly, on these domains every polynomial is equal to a multilinear polynomial. Thus we shall consider only multilinear polynomials. We will write them in the form

$$p(z_1, \ldots, z_n) \;=\; \sum_{\alpha \in \{0,1\}^n} c_\alpha z^\alpha,$$

where the monomials z^α are defined as $z^\alpha \;=\; \prod_{i,\alpha_i=1} z_i$.

The canonical cost measures for voting polynomial representations are the *degree* (the length of a longest monomial) and the *length* (the number of monomials in the polynomial). Other relevant cost measures are the *advantage* and, for integer polynomials, the *weight* of a representation, which we will define in a bit more general setting.

Suppose that basis functions $f_1, \ldots, f_q : X \longrightarrow \{0,1\}^n$ and real numbers $v_0, v_1, \ldots, v_q$ form a *threshold representation* of a given Boolean function $f : X \longrightarrow \{0,1\}^n$, i.e., for all inputs $x \in X$

$$f^*(x) = sgn(h(x)),$$

where $h(x)$ is defined to be $h(x) = v_0 + \sum_{k=1}^q v_k f_k(x)$. If all coefficients v_k are integral then $\sum_{k=1}^q |v_k|$ is called the *weight* of the representation.

In general, the advantage δ of such an representation is define to be

$$\delta = \min_{x \in X} \frac{h(x)}{\sum_{k=1}^{q} |v_k|}.$$

The advantage measures the robustness of the representation. If we assume that $\sum_{k=1}^{q} |v_k| = 1$, what we can always do w.l.o.g., then the weights can be changed by any amount of $\epsilon\delta q^{-1}$, $\epsilon < 1$ without producing errors on X. Consequently, all weights can be supposed to be some integer multiples of some number $\frac{1}{K}$ for some integer $K \in O(q\delta^{-1})$. By multiplying all coefficients by K we get integral coefficients. We obtain:

Lemma 1.1 *Each threshold representation $f^* = sgn(v_0 + \sum_{k=1}^{q} v_k f_k)$ of advantage δ can be transformed into a threshold representation with integer coefficients for f over the same basis functions with weight $O(q^2\delta^{-1})$.* $\square$

Representing Boolean functions by voting polynomials is equivalent to computing them by depth 2 threshold circuits of a special structure.

A Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$ is called a *threshold function* if there are integers $a_0, a_1, \ldots, a_n$ such that $f(x) = 1$ iff $\sum_{i=1}^{n} a_i x_i \geq a_0$ for all $x \in \{0,1\}^n$. A threshold gate is a gate performing a threshold function. The numbers a_i can be thought of as the multiplicities of the ingoing edges with x_i negated, if $a_i < 0$. The value $\sum_{i=1}^{n} |a_i|$ will be called the *weight* of the threshold gate.

Threshold circuits have been investigated very intensively in the last years, e.g. [12][23][15][7]. They form a basic data structure for neural computing and play an important role in circuit complexity theory. Compared with circuits over AND, OR, MOD operations they seem to be more powerful. NC–operations like addition, division, and sorting of n–bit numbers have small weight threshold circuits of surprisingly small depth ≤ 3, and all $AC_0[m]$–functions can be computed by depth–3 threshold circuits of nearly polynomial weight [23]. On the other hand, it is always a serious problem to prove lower bounds on even restricted types of small depth threshold circuits. The best lower bound results on threshold circuits so far are exponential lower bounds on the size of bounded weight depth 2 threshold circuits in [12][15][10] and on the size of unbounded weight depth 2 threshold circuits which have restricted types of operations at the bottom level, in [5][7][14]. We hope that the new lower bound argument underlying the proof of our main result will contribute to a better understanding what the computational limitations of small depth threshold circuits are.

A *threshold–AND* circuit is an unbounded fan-in circuit of depth 2 with AND gates at the bottom level and one threshold gate as output gate at the top. *Threshold–XOR* circuits are defined in a similar way. For circuits $C = C(x_1, \ldots, x_n)$ of this type, we define the *weight* of C to be the weight of the threshold top gate. Note that in the relevant literature, circuits of this type are sometimes called perceptrons.

There is a close connection between threshold-XOR and threshold-AND circuits on the one hand and voting polynomial on the other. Namely,

1. Let $f : \{1, -1\}^n \longrightarrow \{1, -1\}$. Then f has a threshold–XOR circuit with q XOR–gates iff there exists a voting polynomial p for f w.r.t. domain $\{1, -1\}^n$ which has q monomials.

2. Let $f : \{0, 1\}^n \longrightarrow \{0, 1\}$. Then f has a threshold–AND circuit with q AND–gates containing only positive literals iff there exists a voting polynomial p for f w.r.t. domain $\{0, 1\}^n$ which has q monomials.

Thus in case of the domain $\{1, -1\}^n$ threshold–XOR are equivalent to voting polynomials. In case of the domain $\{0, 1\}^n$ it is true with the restriction that we use only positive literals in the conjunctions. We do not know, how essential this restriction is.

1.2 Previous Results and our Results

There are a large number of results on estimating the minimal voting polynomial degree, weight or length for particular Boolean functions. One of the first such bounds is an $\Omega(n)$ lower bound on the minimal voting polynomial degree of the function $\bigwedge_i \bigvee_j x_{i,j}$, which was proved by *Minsky* and *Papert* [17]. In [14] we used this result for investigating the complexity of AC_0–functions with respect to voting polynomial length, and we will use it again in this paper. Another important result is that voting polynomials for the parity function in n variables always have maximal possible degree n [7][4]. *Aspnes, Beigel, Furst* and *Rudich* have shown that all AC_0–functions can be approximated by real polynomials of low degree while this is not true for parity [4]. Such results are very useful for proving lower bounds on the size of bounded depth circuits. Among others, the last mentioned result has been used to give a nice new proof that parity is not in AC_0 [4]. Further related results are due to *Paturi* [20], and *Nisan* and *Szegedy* [19] who showed that most symmetric functions, including the OR function, need large degree for being realized with advantage bounded by a constant.

Observe that the minimal voting polynomial degree, $\deg(f)$, of any Boolean function f does not depend on whether the domain is $\{0,1\}^n$ or $\{1,-1\}^n$. By replacing the variables x_i, resp. x_i^*, by $\frac{1}{2}(1-x_i^*)$, resp. $1-2x_i$ we can transform a voting polynomial computing f over $\{0,1\}^n$, rep. $\{1,-1\}^n$ into one of the same degree computing f over $\{1,-1\}^n$, resp. $\{0,1\}^n$. But the length, clearly, can increase exponentially in the degree. Theorem 1 shows that there are functions for which voting polynomial length over $\{0,1\}^n$ can really be exponentially larger then over $\{1,-1\}^n$.

Our first result implies that if a Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$ has a small length, moderately bounded advantage voting polynomial w.r.t. domain $\{0,1\}^n$, then it has it w.r.t. $\{1,-1\}^n$, too.

We formulate this result in a bit more general fashion in terms of threshold circuits.

Theorem 2 *If $f : \{0,1\}^n \longrightarrow \{0,1\}$ has threshold–AND circuits of weight w then it has threshold–XOR circuits of weight $(nw)^{O(1)}$.*

We will prove this result using probabilistic arguments. In particular, we exhibit the existence of polynomial weight approximation schemes, consisting of parity functions, for the logical OR function. Using methods described in papers of *Naor* and *Naor* [18] and *Alon, Goldreich, Håstad* and *Peralta* [3] such approximation schemes can be efficiently constructed. This implies an efficient procedure for obtaining the simulating threshold parity circuits. Very similar techniques were used already by *Alon* and *Bruck* [2] for constructing small weight polynomials for DNF's of polynomial size as well as for comparison and addition of two n–bit numbers.

A similar simulation of threshold–XOR by threshold–AND circuits cannot exist because Goldmann, actually, showed a more general result in [9]: We need exponentially many gates if we want to compute parity with constantly many *AND–, OR–* levels and one threshold gate as the output gate.

Consequently, polynomial weight threshold–XOR circuits are strictly more powerful than polynomial weight threshold–AND circuits. The resulting question is whether this remains true if we look at unbounded weight circuits of polynomial size. Is there some simulation transforming threshold–AND circuits with polynomially many nodes (but, eventually, exponential weight) into threshold–XOR circuits with polynomially many nodes? Or, in other words, does the existence of a sparse voting polynomial w.r.t. $\{0,1\}^n$ always imply the existence of a sparse voting polynomial w.r.t. $\{1,-1\}^n$ for the same function? The answer to this question, and this is the main result of the paper, is no.

Theorem 3 *There is a function $f : \{0,1\}^n \longrightarrow \{0,1\}$ with the following two properties.*

(i) *f has threshold–AND circuits with $\sqrt{n}$ AND–nodes (in fact, it has a voting polynomial over $\{0,1\}^n$ of length $\sqrt{n}$);*

(ii) *all threshold–XOR circuits for f have $\exp(n^{\Omega(1)})$ XOR–gates (thus all voting polynomials over $\{1,-1\}^n$ for f have exponential length).*

Theorems 2 and 3 can be seen in a line with results from [7][5][10][14] as a step towards better understanding the computational power of polynomial weight versus unbounded weight threshold gates.

Several nontrivial bounds for threshold–XOR circuits have been proved before. *Bruck* and *Smolensky* [7] used the method of spectral analysis for estimating the minimal voting polynomial length over $\{1,-1\}^n$ for particular functions. They showed the existence of polynomial weight threshold–XOR circuits for *comparison* and *addition* of n–bit numbers. On the other hand they proved that functions with a "flat" spectrum need always exponential voting polynomial length over $\{1,-1\}$. This gave exponential lower bounds for the *inner product mod 2* ($\bigoplus_{i=1}^{n} x_i y_i$) and the *quadratic sum function* ($\bigoplus_{i<j} x_i y_j$). It was impossible for us to use spectral methods for proving Theorem 3. All functions known to us with a "flat" spectrum have large XOR's as subfunction, and thus, Theorem 1, cannot have small threshold–AND circuits. Moreover, it seems that the spectrum of functions having small size threshold–AND circuits is not flat.

We will use a mixture of geometric and probabilistic arguments for showing that the existence of small size threshold–XOR circuit for $f \circ OR$ implies the existence of small weight threshold–AND circuits for a large variety of subfunctions of f, for every Boolean function f. Hereby, $f \circ OR$ is defined by taking OR's of pairwise disjoint sets of n variables as inputs for f. Clearly, if f is a threshold function then $f \circ OR$ has small threshold–AND circuits. *Goldmann, Håstad, Razborov* proved an exponential lower bound on the weight of threshold–XOR circuits computing a particular threshold function. Putting all these things together we obtain the proof of Theorem 3.

Another result of this paper concerns the computability of AC_0–functions. It is known that $AC_{0,2}$–functions, i.e., functions having depth 2 circuits over AND,OR,$\neg$ of polynomial size, always have polynomial weight threshold–AND circuits as well as polynomial weight threshold–MOD_m circuits for all $m \in \mathbb{N}$ [2][14].

In [14] we have developed a probabilistic method for proving lower bounds on the number of nodes of threshold–MOD circuits, which, in contrast to spectral lower bound methods, is suited to prove exponential lower bounds for $AC_{0,3}$–functions. But it remained open how $AC_{0,k}$–functions for $k > 2$ behave with respect to threshold–AND circuits, as our argument from [14] can definitely not be applied to threshold–AND circuits. Using *Håstad*'s *Switching Lemma* together with the lower bound on the voting polynomial degree of $\bigwedge_i \bigvee_j x_{i,j}$ [17] mentioned above we will derive

Theorem 4 *All threshold–AND circuits computing the function $S_{l,m,n}$ on lmn variables defined by*

$$S_{l,m,n}(x_{i,j,k}) \;=\; \bigwedge_{i=1}^{l} \bigvee_{j=1}^{m} \bigwedge_{k=1}^{n} x_{i,j,k},$$

where $l = 4.4\sqrt{n},\ m = 1.1n^2$, have $\exp(n^{\Omega(1)})$ many AND–gates.

As in [13] we can extend Theorem 4 in an inductive way to get a similar separation result for higher levels of the $AC_{0,k}$–hierarchy. In particular, for all $k > 1$ one can inductively define $AC_{0,k}$–functions f_n^k which need threshold–$\{AND, OR\}^{k-2}$–circuits of size $\exp(n^{\Omega(1)})$. Theorem 4 proves this for $k = 3$. By threshold–$\{AND, OR\}^{k-2}$–circuits we mean circuits with one threshold output gate connected with $k-2$ levels of AND and OR gates.

The rest of the paper is organized as follows. In section 2 we prove all our results on threshold–AND versus threshold–XOR circuits. In section 3 we present our observations on AC_0–functions.

2 On Threshold–AND versus Threshold–XOR Circuits

2.1 Circuits of bounded weight

We will call a Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$ an OR–function, if there is some $\alpha \in \{0,1\}^n$ such that for all inputs x

$$f(x) \;=\; \bigvee_{i,\alpha_i=1} x_i.$$

XOR–functions and AND–functions are defined in a similar way by replacing $\vee$ by $\oplus$ and $\wedge$, respectively. For proving Theorem 2 we will show the following slightly stronger result.

Proposition 2.1 *Suppose that we are given OR–functions $f_1,\ldots,f_q$ over n variables, and real numbers $v_0, v_1, \ldots, v_q$ such that $v_0+\sum_{i=1}^q v_i f_i$ is a threshold representation for a given function $f : \{0,1\}^n \longrightarrow \{0,1\}$ with advantage δ. Then, for $s = (\lceil nq^2\delta^{-2}\rceil + 1)$ there are XOR–functions $p_1,\ldots,p_s$ and reals $u_0, u_1, \ldots, u_s$ such that $u_0+\sum_{j=1}^q u_j p_j$ is a threshold representation for f with advantage $\frac{\delta}{2}$.*

Observe that using de Morgan's rules it is straightforward to transform a threshold–AND circuit with q AND nodes and weight w into an equivalent threshold representation of length q over the OR–functions with advantage w^{-1}. Using Proposition 2.1 and Lemma 1.1 we obtain the existence of equivalent threshold–XOR circuits of weight $(nqw)^{O(1)}$.

The proof of Proposition 2.1:

It is sufficient to prove that for each $\epsilon > 0$ and $t > \lceil \frac{n}{4\epsilon^2} \rceil$ there are parity functions $p_1,\ldots,p_t$ such that for all $x \in \{0,1\}^n$

$$\left| \bigvee_{i=1}^{n} x_i - \frac{2}{t}\sum_{i=1}^{t} p_i(x) \right| \;\le\; \epsilon. \tag{1}$$

We can then take $\epsilon = \frac{\delta}{2q}$ and $t = \lceil nq^2\delta^{-2}\rceil + 1$ and replace the OR–functions by a corresponding weighted sum of at most t parity functions and get the desired threshold representation for f over the parity functions.

For deriving (1) it is sufficient to show that for $t > \frac{n}{4\epsilon^2}$ and all $x \in \{0,1\}^n,\ x \neq \vec{0}$ it holds that

$$Pr\left[\left|1 - \frac{2}{t}\sum_{i=1}^{t} \mathbf{p}_i(x)\right| > \epsilon\right] \ < \ e^{-n} \ < \ 2^{-n}, \tag{2}$$

where $\mathbf{p}_1, \ldots, \mathbf{p}_t$ denote random XOR–functions which are mutually independently chosen w.r.t. the uniform distribution from the set

$$\{ \bigoplus_{i,\alpha_i=1} x_i; \alpha \in \{0,1\}^n\}$$

of all XOR–functions. Observe that for all $x \in \{0,1\}^n,\ x \neq \vec{0}$ and a random XOR–function $\mathbf{p}$

$$Pr[\mathbf{p}(x) = 1] = Pr[\mathbf{p}(x) = 0] = \frac{1}{2}.$$

We use the following well-known Chernoff's bound: For $i = 1, \ldots, t$ let X_i denote mutually independent random variables with $Pr[X_i = 1] = Pr[X_i = -1] = \frac{1}{2}$ for all $i = 1, \ldots, t$. Then for all $a > 0$ we have:

$$Pr\left[\left|\sum_{i=1}^{t} X_i\right| > a\right] \ < \ 2e^{-\frac{a^2}{2n}}. \tag{3}$$

By setting $a = 2\epsilon t$ and $\mathbf{p}_i = \frac{1}{2}(X_i + 1)$ we can derive (2) from (3) by standard calculations. $\Box$

2.2 Unbounded Weight Circuits – Introduction into the proof of Theorem 3

Let us consider Boolean functions of the following type $t_n \circ OR_n^m : (\{0,1\}^m)^n \longrightarrow \{0,1\}$, where t_n is an n–ary threshold function and $OR_n^m : (\{0,1\}^m)^n \longrightarrow \{0,1\}^n$ is defined as

$$OR_n^m(x_{1,1}, \ldots, x_{m,1}, \ldots, x_{n,1}, \ldots, x_{m,n}) \ = \ \left(\bigvee_{i=1}^{m} x_{i,1}, \ldots, \bigvee_{i=1}^{m} x_{i,n}\right).$$

Using de Morgan rules it is straightforward to see that functions of this type will always have threshold–AND circuits with n AND–nodes. So, for showing Theorem 3 it is sufficient to prove the following

Proposition 2.2 *There is a threshold function t_n such that all threshold–XOR circuits for $t_n \circ OR_n^n$ have* $\exp(n^{\Omega(1)})$ *XOR gates.*

The rest of this section is devoted to the proof of Proposition 2.2.

Let us consider the threshold functions $T_{l,k} : \left(\{0,1\}^k\right)^l \longrightarrow \{0,1\}$ defined by

$$T^*_{l,k}(x_{1,1},\ldots,x_{1,k},\ldots,x_{l,1},\ldots,x_{l,k}) \;=\; sgn\left(\sum_{i=1}^{k} 2^i \sum_{j=1}^{l} x^*_{i,j} + \frac{1}{2}\right).$$

$T_{l,k}$ has a natural interpretation as the first bit of the sum of l k–bit numbers. From the general upper bound of $\exp(n \log n)$ [17] on the weight of n–nary threshold functions it follows that functions of the above type are *universal* in the sense that each n–ary threshold function occurs as subfunction in $T_{l,k}$ for some appropriate $l, k \in O(n \log n)$.

We will work with the following threshold function $t_{4m,m}$ defined as

$$t^*_{4m,m}(x_{1,1},\ldots,x_{1,4m},\ldots,x_{m,1},\ldots,x_{l,4m}) \;=\; sgn\left(\sum_{i=1}^{m} 2^i \sum_{j=1}^{4m} x^*_{i,j} - 3n2^{n+1} + \frac{1}{2}\right).$$

We will show that the existence of small threshold–XOR circuits for $t_{4m,m} \circ OR^m_{4m^2}$ implies the existence of small weight threshold–AND circuits for $T_{m,m}$. By Theorem 2, this would prove the existence of small weight threshold–XOR circuits for $T_{m,m}$. However, *Goldmann, Håstad* and *Razborov* have shown in [10], Corollary 4.4, that such circuits do not exist:

Lemma 2.1 *Threshold–XOR circuits computing $T_{m,m}$ have weight $2^{\Omega(m)}$.* $\square$

We will use the following statement which, by Lemma **??**, is equivalent to Lemma 2.1.

Lemma 2.2 *For all threshold realizations for $T_{m,m}$ over the XOR–functions of length q and advantage δ it holds $\delta^{-1} q \geq 2^{\Omega(m)}$.* $\square$

From now on let $t_n = T_{4m,m}$, i.e. $n = 4m^2$. Further let us fix a threshold–XOR circuit for $t_n \circ OR_n^m$. In particular, let

$$t_n^* \circ OR_n^m(x) \;=\; sgn(h(x)),$$

for all $x \in (\{0,1\}^m)^n$, where

$$h(x) \;=\; v_0 + \sum_{k=1}^{q} v_k p_k(x) \;=\; u_0 + \sum_{k=1}^{q} u_k p_k^*(x).$$

(Observe the relations $u_0 = v_0 + \frac{1}{2}\sum_k v_k$ and $u_k = -\frac{1}{2}v_k$.) Let us write the parity functions p_k as

$$p_k \;=\; \bigoplus_{(i,j)\in I_k} x_{i,j},$$

where $I_1, \ldots, I_q$ are pairwise distinct and nonempty subsets of $[n] \times [m]$. Further suppose that the representation is normalized, i.e.,

$$\sum_{k=0}^{q} |v_k| = 1.$$

Observe that $\max\{|v_k|,\ 1 \le k \le q\} \ge \frac{1}{2q}$. Otherwise, our target function would be constant $sgn(v_0)$. Consequently,

$$\max\{|u_k|,\ 1 \le k \le q\} \;\ge\; \frac{1}{4q}. \tag{4}$$

2.3 Two technical Lemmas

Fix $y \in \{0,1\}^n$. For a function $f : (\{0,1\}^m)^n \longrightarrow \mathbb{R}$ we denote by $\mathbf{E}_y[f(x)]$ the expected value of $f(x)$ with respect to the uniform distribution over all x for which $OR_n^m(x) = y$. Observe that for such x, $t_n^*(y) = sgn(h(x))$, hence

$$t_n^*(y) \;=\; sgn(\mathbf{E}_y[h(x)]) \;=\; sgn\left(v_0 + \sum_{k=1}^{q} v_k \mathbf{E}_y[\bigoplus_{(i,j)\in I_k} x_{i,j}]\right).$$

For all $k = 1, \ldots, q$ let us denote by J_k the projection of I_k to $[n]$, i.e., $J_k \subseteq [n]$ is defined by

$$j \in J_k \quad\Longleftrightarrow\quad \exists i,\ x_{i,j} \in I_k$$

Calculating the values $\mathbf{E}_y[p_k]$ we obtain the following result which is of basic importance for our proof.

Lemma 2.3 *For all $y \in \{0,1\}^n$,*

$$t_n^*(y) \;=\; sgn\left(v_0 + \sum_{k=1}^{q} \frac{v_k}{2} \bigvee_{j \in J_k} y_j + err(y)\right),$$

where $|err(y)| \;\le\; qn2^{-(m-1)}$.

Proof:
It is sufficient to show that for all $k = 1, \ldots, q$ and $y \in \{0,1\}^n$,

$$\left| \mathbf{E}_y \left[\bigoplus_{(i,j) \in I_k} x_{i,j} \right] - \frac{1}{2} \bigvee_{j \in J_k} y_j \right| \;\le\; n2^{m-1}.$$

First observe that for all x satisfying $OR_n^m(x) = y$

$$\bigoplus_{(i,j) \in I_k} x_{i,j} \;=\; \bigoplus_{(i,j) \in I_k, y_j = 1} x_{i,j} =_{df} p_y(x).$$

Let $A = \{j;\ y_j = 1\}$. If $A \cap J_k = \emptyset$ then the statement of the lemma is obviously true as

$$\mathbf{E}_y \left[\bigoplus_{(i,j) \in I_k} x_{i,j} \right] \;=\; \frac{1}{2} \bigvee_{j \in J_k} y_j \;=\; 0.$$

If $A \cap J_k \ne \emptyset$ then observe that if we denote

$$B = (\{0,1\}^n)^{A \cap J_k} \quad \text{and} \quad B^* = (\{0,1\}^n \setminus \{\vec{0}\})^{A \cap J_k}$$

then

$$\mathbf{E}_B[p_y(x)] \;=\; \frac{1}{2} \quad \text{and} \quad \mathbf{E}_{B^*}[p_y(x)] \;=\; \mathbf{E}_y \left[\bigoplus_{(i,j) \in I_k} x_{i,j} \right],$$

where $\mathbf{E}_B$ and $\mathbf{E}_{B^*}$ denote the expected values w.r.t. the uniform distribution on B and B^*, respectively. Consequently,

$$\frac{1}{2} \;=\; \frac{|B^*|}{|B|} \mathbf{E}_{B^*}[p_y(x)] + \frac{\sum_{x \in B \setminus B^*} p_y(x)}{|B|},$$

i.e.,

$$\frac{|B|}{2\,|B^*|} - \frac{|B\setminus B^*|}{|B^*|} \;\le\; \mathbf{E}_y[\bigoplus_{(i,j)\in I_k} x_{i,j}] \;\le\; \frac{|B|}{2\,|B^*|}.$$

Observe that $\frac{|B|}{2|B^*|} > \frac{1}{2}$ and $\frac{|B|}{2|B^*|} - \frac{|B\setminus B^*|}{|B^*|} < \frac{1}{2}$. Thus

$$\left|\mathbf{E}_y\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right] - \frac{1}{2}\bigvee_{j\in J_k} y_j\right| \;=\; \left|\mathbf{E}_y\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right] - \frac{1}{2}\right| \;\le\; \frac{|B\setminus B^*|}{|B^*|}.$$

Let us denote by $a = |A\cap J_k|$, then we can estimate the last term by

$$= \frac{2^{ma} - (2^m-1)^a}{(2^m-1)^a} = \frac{a(2^m-1)^{a-1} + \binom{a}{2}(2^m-1)^{a-2} + \ldots}{(2^m-1)^a} \le$$

$$\frac{a}{2^m-1} \le n2^{-(m-1)}. \quad \Box$$

Now our aim is to show the existence of a subcube W of $\{0,1\}^n$ which is *"good"* in the following sense. On the one hand, W should have the property that

$$t_n^*(y) \;=\; sgn\left(v_0 + \sum_{k=1}^{q}\frac{v_k}{2}\bigvee_{j\in J_k} y_j\right)$$

for all $y\in W$ with advantage polynomially bounded in n and q. On the other hand, t_n restricted to W should be equal to some $T_{l,k}$ with $l,k\in n^{\Omega(1)}$. This two things together yield a contradiction in the way described above, if q is subexponentially small.

Let us investigate the deviation of $h(x)$ from the mean $\mathbf{E}_y[h(x)]$ when x is uniformly chosen from $\{x;\ OR_n^m(x) = y\}$. In particular, we will estimate the number $\delta(y),\ y\in\{0,1\}^n$, which is defined to be the maximal number $\delta > 0$ for which there exists $x^+, x^-\in(\{0,1\}^m)^n$ with $OR_n^m(x^+) = OR_n^m(x^-) = y$ such that

$$h(x^+) \ge \mathbf{E}_y[h(x)] + \delta \quad \text{and} \quad h(x^-) \le \mathbf{E}_y[h(x)] - \delta.$$

As, by definition, $t_n^*(y) = sgn(h(x))$ for all x with $OR_n^m(x) = y$ it must hold that

$$|\mathbf{E}_y[h(x)]| \;>\; \delta(y).$$

Consequently, by Lemma 2.3, if for some subcube W of $\{0,1\}^n$ it holds that $min\{\delta(y);\ y \in W\} > qn2^{-(m-1)}$ then

$$t_n^*(y) = sgn\left(v_0 + \sum_{k=1}^{q} \frac{v_k}{2} \bigvee_{j \in J_k} y_j\right)$$

on W with advantage at least $min\{\delta(y), y \in W\} - qn2^{-(m-1)}$.

For $A \subseteq [n] = \{1, \ldots, n\}$ we define a subcube of $\{0,1\}^n$

$$One_A^n = \{y \in \{0,1\}^n;\ \forall i \in A\ y_i = 1\}.$$

The main ingredience of the proof of our theorem is the following technical lemma.

Lemma 2.4 *Let U denote a system of subsets of $[n]$, $|U| < 2^{n-2}$, (U will define the subsystem of "bad" subcubes). Then there exists a subset $A \subseteq [n]$, $A \notin U$ such that for all y in One_A^n*

$$\delta(y) \geq \frac{1}{4q^3} - \frac{1}{q}\frac{|U|}{2^{n-2} - |U|} - n2^{-(m-2)}.$$

2.4 Proving the Theorem using Lemma 2.4

Suppose that $t_{4m,m} \circ OR_m^{4m^2}$ can be realized by a threshold–parity circuit with q parity nodes.

For subsets $A \subseteq ([m] \times [4m])$ and indices $i, 1 \leq i \leq m$, we denote $A_i =_{df} A \cap (\{i\} \times [4m])$.

We define the set of forbidden restrictions $U \subseteq P([m] \times [4m])$ as

$$U \;=\; \{A \subseteq [m] \times [4m];\ \max\{|A_i|\,;\ 1 \leq i \leq m\} > 3m\}.$$

Lemma 2.4 and Lemma 2.3 provide the existence of $A \subseteq [m] \times [4m]$ fulfilling $|A_i| = 3m$ for all $i = 1, \ldots, m$ such that $t_{4m,m}|_{One_{A'}^{4m^2}}$ can be computed by a threshold–AND circuit with q AND–nodes with advantage $\delta \geq \frac{1}{4q^3} - \frac{2}{q}\delta_1 - q\delta_2$, where,

$$\delta_1 = \frac{|U|}{2^{4m^2-2} - |U|} \quad \text{and} \quad \delta_2 = 4m^2 2^{-(m-1)}.$$

By estimating the size of U in a standard way (e.g., using Stirling's formula) we obtain the existence of some constant $\beta > 0$ such that

$$\delta \geq \frac{1}{4q^3} - q2^{-\beta m}.$$

On the other hand observe that $t_{4m,m}|_{One_A^{4m^2}}$ equals to our universal threshold function $T_{m,m}$. Together with Theorem 2 we obtain the existence of some threshold–XOR circuit with $s \leq (mq\delta^{-1})^{O(1)}$ XOR nodes computing $T_{m,m}$ with advantage $\frac{\delta}{2}$.

From Lemma 2.2 we know that there must hold the relation $s\frac{2}{\delta} \geq 2^{\Omega(m)}$. This can only be achieved if $q \in 2^{\Omega(m)}$, we are done. □

It remains to prove Lemma 2.4. Our main tool for doing this is the following technical lemma, based on the orthogonality of the parity functions.

2.5 The Deviation Lemma

Lemma 2.5 *Let $r \geq 2$, $U \subseteq \{0,1\}^r$, $|U| < 2^{r-1}$, and $g = w_1p_1^* + \ldots w_sp_s^*$, where $w_i \in \mathbb{R}$ and $p_1, \ldots, p_s$ are pairwise distinct and nonconstant parity functions over $\{0,1\}^r$. Then there exist inputs $y^+, y^- \in \{0,1\}^r \setminus U$ such that*

$$g(y^+) \;\geq\; \delta \quad \textit{and} \quad g(y^-) \;\leq\; -\delta$$

where $\delta = \max_i |w_i| - \frac{|U|}{2^{r-1}-|U|} \sum_{i=1}^s |w_i|$.

First observe that due to the orthogonality of the parity functions we have that for every pair p_i, p_j of distinct parity functions

$$\mathbf{E}(p_j^* \mid p_i^*(x) = 1) = 0. \tag{5}$$

Hereby, $\mathbf{E}$ denotes the expected value w.r.t. the uniform distribution on $\{0,1\}^r$.

Hence, for all subsets $U \subseteq \{0,1\}^r$, $|U| < 2^{r-1}$ we have

$$\left|\mathbf{E}(p_j^* \mid p_i^*(x) = 1, x \notin U)\right| \leq \frac{|U|}{2^{r-1} - |U|}.$$

Fix an i and suppose that $w_i \geq 0$. Then

$$\mathbf{E}(\sum w_jp_j^* \mid p_i^*(x) = 1, x \notin U) = \sum w_j\mathbf{E}(p_j^* \mid p_i^*(x) = 1, x \notin U) =$$

$$w_i + \sum_{j \neq i} w_j \mathbf{E}(p_j^* \mid p_i^*(x) = 1, x \notin U).$$

By (5), we get

$$\mathbf{E}(\sum w_j p_j^* \mid p_i^*(x) = 1, x \notin U) \geq w_i - \frac{|U|}{2^{r-1} - |U|} \sum_{j \neq i} |w_i|,$$

thus there is at least one $y^+ \in \{0,1\}^r \setminus U$ for which we have this inequality.

For proving the existence of y^- take the expectation over all y outside U for which $p_i^*(y) = -1$. If $w_i \leq 0$ then we consider $p_i^*(y) = -1$ for y^+ and $p_i^*(y) = 1$ for y^-. □

2.6 Completing the proof by showing Lemma 2.4 using the Deviation Lemma

Let us fix $y \in \{0,1\}^n \setminus \{\vec{0}\}$ and let $A = \{j;\ y_j = 1\}$. In the sense of the Lemma 2.5, $\delta(y)$ plays the role of δ if we consider as g the following function $h_y : (\{0,1\}^m)^A \longrightarrow \mathbb{R}$ which is induced from h by y.

$$h_y(x) \ = \sum_{D \subseteq [m] \times A, D \neq \emptyset} u_{D,A} \left(\bigoplus_{(i,j) \in D} x_{i,j} \right)^*,$$

where

$$u_{D,A} = \sum_{k, I_k \cap [m] \times A = D} u_k.$$

Observe that for all x with $OR_n^m(x) = y$ it holds $h(x) = u_0 + h_y(x)$, and that all parity functions occurring in the definition of h_y are pairwise distinct and not constant. By Lemma 2.5 we obtain that

$$\delta(y) \geq \max\{|u_{D,A}|;\ D \subseteq A \times [m], D \neq \emptyset\} - \frac{|W|}{2^{|A|m-1} - |W|}, \qquad (6)$$

where $W = (\{0,1\}^m)^A \setminus \left(\{0,1\}^m \setminus \vec{0}\right)^A$. This is because $\sum_D |u_{D,A}| \leq \sum_k |u_k| \leq 1$. Consequently,

$$\delta(y) \geq \max\{|u_{D,A}|;\ D \subseteq A \times [m], D \neq \emptyset\} - n2^{-(m-2)}. \qquad (7)$$

Here we used that for $t > 2a > 1$

$$\frac{t^a - (t-1)^a}{(t-1)^a - \frac{1}{2}t^a} \le \frac{a}{\frac{1}{2}t - a},$$

with $a = |A|$ and $t = 2^m$ and taking into account that $a \le n$. This inequality can be easily derived by exhibiting the relation $t^a - at^{a-1} > (t-1)^a$.

Hence, for proving Lemma 2.4 we have to find some *good* set $A \subseteq [n]$ such that for all A', $A \subseteq A'$, we can find some $D \subseteq A' \times [m]$, $D \neq \emptyset$, such that $u_{D,A'}$ is large. This is not trivial. We certainly know that some of the $u_k's$ are large. That means that for $A = [n]$ there is nothing to prove. But for "smaller" A, all coefficients $u_{D,A}$ may become exponentially small.

We proceed in two steps. First we prove the existence of $A' \subseteq [n]$, $A' \notin U$ such that for all A'', $A' \subseteq A''$ there is some $D' \subseteq A'' \times [m]$, which is not necessarily nonempty, such that

$$|u_{D',A''}| \ge \frac{1}{4q^3} - \frac{2}{q}\frac{|U|}{2^{n-1} - |U|},$$

(note the change in the exponent at 2). Then, in a second stage, using the ideas of the first step, we prove the existence of some $A \notin U$ such that for all A', $A \subseteq A'$ there is some $D \subseteq A' \times [m]$, $D \neq \emptyset$ such that

$$|u_{D,A'}| \ge \frac{1}{4q^3} - \frac{2}{q}\frac{|U|}{2^{n-2} - |U|}.$$

Together with inequality (7) this proves Lemma 2.4.

By Lemma 2.5, (with $U = \emptyset$), and (4) we get some $c \in \{0,1\}^{n\times m}$ such that

$$\sum_k u_k p_k^*(c) \ge \max_k |u_k| \ge \frac{1}{4q}.$$

Let us define

$$\tilde{p}_k(y_1, \ldots, y_n) =_{df} \bigoplus_i \left(y_i \cdot \bigoplus_{(i,j)\in I_k} c_{i,j} \right).$$

Observe that we have

$$\sum_k u_k \tilde{p}_k^*(\vec{1}) = \sum_k u_k p_k^*(c) \ge \frac{1}{4q}. \tag{8}$$

The parity functions $\tilde{p}_k$ are not necessarily distinct, so we join the identical ones and add their weights; let p'_j, u'_j be the resulting system of distinct parity functions and their weights. So we have, by (8), $\sum_j u'_j p'^*_j(\vec{1}) \geq \frac{1}{4q}$, hence

$$\max |u'_j| \geq \frac{1}{4q^2}.$$

By Lemma 2.5, there is some $b \in \{0,1\}^n$ such that for $A' = \{i;\ b_i = 1\}$ it holds $A' \notin U$, and

$$\sum_j u'_j p'^*_j(b) \geq \max_j |u'_j| - \frac{|U|}{2^{n-1} - |U|} \sum_j |u'_j| \geq \frac{1}{4q^2} - \frac{|U|}{2^{n-1} - |U|}. \qquad (9)$$

Now define $a \in (\{0,1\}^m)^n$ by $a_{i,j} = b_i c_{i,j}$ and observe that

$$\sum_k u_k p^*_k(a) =$$

$$\sum_k u_k \tilde{p}^*_k(b) = \sum_j u'_j p'^*_j(b).$$

Now let $A'' \subseteq [n]$ with $A' \subseteq A''$ be arbitrarily fixed. Observe that

$$\sum_k u_k p^*_k(a) = \sum_{D \subseteq (A'' \times [m])} \ \sum_{I_k \cap A'' \times [m] = D} u_k p^*_k(a).$$

Since $a_{i,j} = 0$ for all i, j outside $A'' \times [m]$ we have that for all $D \subseteq A'' \times [m]$ the parity functions $p^*_k(a)$ in the term

$$\sum_{I_k \cap (A'' \times [m] = D)} u_k p^*_k(a)$$

get the same value. This means that for all $D \subseteq (A'' \times [m])$

$$\left| \sum_{I_k \cap (A'' \times [m] = D)} u_k p^*_k(a) \right| = \left| \sum_{I_k \cap (A'' \times [m] = D)} u_k \right| = |u_{D,A''}|.$$

Thus, there must be some $D' \subseteq (A'' \times [m])$, not necessarily empty, such that

$$|u_{D,A''}| \ \geq \ \frac{1}{4q^3} - \frac{1}{q} \frac{|U|}{2^{n-1} - |U|}.$$

In a second stage we extend the above argument to get some $D \neq \emptyset$. Let k_0 be such that $|u_{k_0}| = \max_k |u_k| \geq \frac{1}{4q}$. Let i_0 be such that $D_0 =_{df} I_{k_0} \cap (\{i_0\} \times [m]) \neq \emptyset$. Consider the weighted set system on $([n] \setminus \{i_0\}) \times [m]$ consisting of sets $I'_k = I_k \setminus (\{i_0\} \times [m])$ for all those k for which $I_k \cap (\{i_0\} \times [m]) = D_0$, and take the same weights u_k.

Let us define a system of "bad" restrictions U' of subsets of $[n]/\{i_0\}$ by

$$E \in U' \iff E \cup \{i_0\} \in U.$$

By the first part, we get some $A_1 \subseteq [n] \setminus \{i_0\}$, $A_1 \notin U'$ such that for all A_2, $A_1 \subseteq A_2 \subseteq [n] \setminus \{i_0\}$ there is some $D_1 \subseteq A_2 \times [m]$ such that

$$\left| \sum_{k, I_k \cap (\{i_0\} \times [m]) = D_0, I_k \cap (A_1 \times [m]) = D_1} u_k \right| \geq \frac{1}{4q^3} - \frac{1}{q} \frac{|U|}{2^{n-2} - |U|}.$$

Now look at $A = A_1 \cup \{i_0\}$ and let A', $A \subseteq A' \subseteq [n]$ be arbitrarily fixed. Observe that $A \notin U$. Let $A_2 =_{df} A' \setminus \{i_0\}$ and find some D_1 as above. Then for $D = D_0 \cup D_1$ it holds $D \neq \emptyset$ and

$$\sum_{k, I_k \cap (\{i_0\} \times [m]) = D_0, I_k \cap (A_1 \times [m]) = D_1} u_k = u_{D,A}. \ \square$$

3 Alternation versus unbounded weight threshold operations

This section is devoted to the proof of Theorem 4. Let us first recall the theorem.

Theorem 4: *All threshold–AND circuits computing the function* $S_{l,m,n} : \{0,1\}^{lmn} \longrightarrow \{0,1\}$ *defined by*

$$S_{l,m,n}(x_{i,j,k}) = \bigwedge_{i=1}^{l} \bigvee_{j=1}^{m} \bigwedge_{k=1}^{n} x_{i,j,k},$$

where $l = 4.4\sqrt{n}$, $m = 1.1n^2$, *have exponentially in* n *many AND–gates.*

Proof:

Note that essential parts of the proof are very closely related to [13], section 5. Suppose that the set $X_{l,m,n} = \{x_{i,j,k}\}$ of input variables for $S_{l,m,n}$ is divided into blocks $X_{i,j}$ containing the variables $x_{i,j,1}, \ldots, x_{i,j,n}$ for all $i = 1, \ldots, l,\ j = 1, \ldots, m$. As in [13] we consider random restrictions of the set $X_{l,m,n}$ which are distributed according to $R(p^*, p_1, p_0)$, where $p^*, p_1, p_0 \in [0,1],\ p^* + p_1 + p_0 = 1$. The distribution $R(p^*, p_1, p_0)$ is defined by doing the following experiment independently for each block $X_{i,j}$.

- With probability p_1 set $\rho_{i,j,k} = 1$ for all $k = 1, \ldots, n$.
- With probability $p_0 + p^*$ we do the following: we choose uniformly some $k,\ 1 \leq k \leq n$, and set $\rho_{i,j,k} = 0$ with probability $\frac{p_0}{p_0+p^*}$, and $\rho_{i,j,k} = *$ with probability $\frac{p^*}{p_0+p^*}$. All $\rho_{i,j,k'}$, for $k' \neq k$, are set to be one.

We have to put together the following three technical results. The first is the so called Switching Lemma of *Håstad* [13] (Lemma 4).

Lemma 3.1 *Let G be a Boolean function over $X_{l,m,n}$ which can be written as AND over OR's of fan-in at most t. Then*

$$Prob_{R(p^*,p_1,p_0)}[G^\rho \textit{ has a minterm of size } > s] \leq \alpha^s,$$

where α is the unique positive root of the equation $\left(1 + \frac{2q}{\alpha}\right)^t = \left(1 + \frac{q}{\alpha}\right)^t + 1$, and q is defined to be $\max\left\{\frac{p_0}{p_0+p^*}, \frac{p_0}{p_1 lm+p^*}\right\}$, □

Furthermore we need the following result (e.g. [6], a nice proof can be found in [16], Lemma 3).

Lemma 3.2 *If all the minterms of a function F have size at most k, and all the maxterms have size at most l, then f can be computed by a decision tree of size kl.* □

Finally, we will use the following result, which is equivalent to Lemma 4 in [16].

Lemma 3.3 *If a boolean function $H : \{0,1\}^n \longrightarrow \{0,1\}$ can be realized by a decision tree of depth s, then there is an n–ary real polynomial of degree s computing H exactly.* □

Now suppose that a function f over $X_{l,m,n}$ can be realized by a real polynomial p of length d over $\{0,1\}$. Consider the monomials of p as to be AND over OR's of fan-in 1. Using the terminology of Lemma 3.1 we can conclude from Lemma 3.2 and Lemma 3.3 that each particular monomial of p reduces to an n–ary real polynomial of degree at most s with probability at least $1-\alpha^s$, where α is defined as above with $t=1$. Consequently, for all $s \leq n$,

$$\Pr_{R(p^*,p_1,p_0)}[\deg(f^\rho) > s] \ \leq \ d\alpha^s, \tag{10}$$

where $\alpha = \max\left\{\frac{p_0}{p_0+p^*}, \frac{p_0}{p_1 lm+p^*}\right\}$.

Now, for all $n \in \mathbb{N}$ denote $T_n = S_{4.4n,1.1n^2,n^2}$, and $U_n = S_{1.1n,n,1}$, and let furthermore $R_n = R(\frac{1}{n}, \frac{1}{n^2}, 1-\frac{1}{n}-\frac{1}{n^2})$. A straightforward application of Lemma 8 from [13] gives:

$$Prob_{R_n}[T_n^\rho \text{ contains } U_n \text{ as a subfunction}] \ \geq \frac{2}{3}.$$

Putting this together with the well known result of *Minsky* and *Papert* [17] that

$$\deg(S_{n,4n^2,1}) \geq n$$

for all $n \in \mathbb{N}$, we obtain

$$Prob_{\rho\in_{R_n}\{0,1,*\}^n}[\deg(T_n^\rho) \geq \frac{1}{2}\sqrt{n}] \ \geq \ \frac{2}{3}. \tag{11}$$

From (10) and (11), we obtain the relation $d\alpha^{\frac{1}{2}\sqrt{n}} \ \geq \ \frac{2}{3}$, where α is determined by R_n in the way described in (10). As in [13], it can be shown that $\alpha < \frac{2}{n}$ for n large enough. We obtain $d = 2^{\Omega(n\log n)} = 2^{N^{\Omega(1)}}$ where N denotes the number of variables of T_n. This proves the theorem. □

As in [13] we can extend Theorem 4 in an inductive way to get a similar separation result for higher levels of the $AC_{0,k}$–hierarchy. In particular, for all $k>1$ we can inductively define $AC_{0,k}$–functions f_n^k which need threshold–$\{AND, OR\}^{k-2}$–circuits of size $\exp(n^{\Omega(1)})$. Theorem 4 proves this for $k=3$. The functions f_n^k will have the property that with high probability a random restriction distributed according to some appropriate $R(p^*,p_1,p_0)$ will do the following two things. It converts f_n^k into $f_{n'}^{k-1}$ and it reduces a small threshold–$\{AND, OR\}^{k-2}$–circuit into a circuit with small bottom fan-in. By the induction hypothesis, such circuits do not exist for $f_{n'}^{k-1}$. We have disproved the existence of small threshold–$\{AND, OR\}^{k-2}$ for f_n^k.

4 Conclusions and open problems

If we look at our main result from the point of view of voting polynomials, it closes a gap in our knowledge about the role of the domains $\{0, 1\}$ and $\{1, -1\}$. However in context of bounded depth circuits there are still related open problems. In particular, we do not know, if we can replace the parity functions by gates counting modulo a number different from 2:

Problem 1 Is there a boolean function which can be computed by a polynomial size *threshold-AND* circuit, but cannot be computed by a polynomial size *threshold-MOD_p* circuit, for p an odd prime? The same is open for p composite.

The obstacle to generalization of our result is that the MOD_p functions are not orthogonal for $p \neq 2$.

Also we do not quite understand the role of the negation in the context of voting polynomials.

Problem 2 What is the relation of threshold-AND circuits, containing negated variables, to voting polynomials over the domain $\{0, 1\}^n$?

An interesting combinatorial problem concerns the proof of the theorem. Our proof shows that for a function t, if $t \circ OR_n^m$ can be computed by a *threshold-parity* circuit of polynomial size (with m sufficiently large), a lot of restrictions of t can be computed with polynomial advantage. In the proof we have to admit restrictions which fix more than one half of the input variables to 1. Our feeling is that this is not necessary and, perhaps, it is enough to fix only $O(\log n)$ variables. This would give a more elegant proof.

Our Theorem 4 and the large number of other results on AC^0 functions might give the impression that we know all about computing AC^0 functions by threshold circuits. Nevertheless, there is a basic open problem:

Problem 3 Is there an AC^0 function which cannot be computed by depth 2 threshold circuits of subexponential size? It is open also for depth 2 threshold circuits with polynomially bounded weights.

The best lower bound for depth 2 threshold circuits with polynomially bounded weights can be proved for *inner product modulo 2* padded with dummy variables so that it is an AC^0 function. It gives only $2^{\log^{O(1)} n}$ [7].

References

[1] Allender, E.: *A note on the power of threshold circuits*, Proc. 30th IEEE Symposium FOCS, 1989, 580–584.

[2] Alon, N., J. Bruck: *Explicit constructions of depth two majority circuits for comparison and addition,* Tech. Report RJ 8300 (75661) IBM San Jose, 1991, to appear in SIAM Journal of Discrete Mathematics

[3] Alon, N., O. Goldreich, J. Håstad, R. Peralta: *Simple constructions of almost k–wise independent random variables*, Random Structures and Algorithms 3, (1992 9, 289–304.

[4] Aspnes, J., R. Beigel, M. Furst, S. Rudich: *The expressive power of voting polynomials,* Proc. 23rd ACM Conference STOC, 1991, 402–409.

[5] Bruck, J.: *Harmonic analysis of polynomial threshold functions*, SIAM Journal of Discrete Mathematics, 3, No. 22, (1990), 168–177.

[6] Blum, M., R. Impagliazzo: *Generic oracles and oracle classes*, 28th IEEE Symposium FOCS, 1987, 118–126.

[7] Bruck, J., R. Smolensky: *Polynomial threshold functions, AC^0–functions and spectral norms,* Proc. 31st IEEE Symposium FOCS, 1990, 632–641.

[8] Furst, M., J. Saxe, M. Sipser: *Parity, Circuits, and the Polynomial Time Hierarchy,* Proc. 22nd IEEE Symposium FOCS, 1981, 260–270.

[9] Goldmann, M.: *personal communication* December 1993

[10] Goldmann, M., J. Håstad, A.A. Razborov: *Majority Gates versus general weighted threshold gates*, Computational Complexity 2 (1992), 277–300.

[11] Green, F., J. Köbler, J. Toran: *The power of the middle bit,* Proc. of 7th IEEE Conference on Structures in Complexity Theory, Boston, 1992, 111–117.

[12] Hajnal, A., W. Maass, P. Pudlák, M. Szegedy, G. Turán: *Threshold circuits of bounded depth,* Proc. 28th IEEE Symposium FOCS, 1987, 99–110.

[13] Håstad, J.: *Almost optimal lower bounds for small depth circuits* Proc. 18th ACM Conference STOC, 1986, 6–20.

[14] Krause, M., P. Pudlák: *On the computational power of depth 2 circuits with threshold– and MOD gates,* Proc. 26th ACM Conference STOC, 1994, pp. 49–59.

[15] Krause, M., S. Waack: *Variation ranks of communication matrices and lower bounds for depth 2 circuits having symmetric gates with unbounded fanin,* Proc. 32th IEEE Symposium FOCS, 1991, pp. 777–786.

[16] Linial, N., Y. Mansour, N. Nisan: *Constant Depth Circuits, Fourier Transforms, and Learnability,* Proc. 30th IEEE Symposium FOCS, 1989, pp.574–579, also Journal of the ACM 40(3), (1993), pp. 607–620.

[17] Minsky, M., S. Papert: *Perceptrons,* MIT Press, Cambridge 1988 (expanded edition, first edition appeared in 1968).

[18] Naor, J., M. Naor: *Small bias probability spaces: Efficient constructions and applications,* Proc. 22nd ACM Conference STOC, 1990, pp. 213–223.

[19] Nisan, N., M. Szegedy: *On the degree of Boolean functions as Real Polynomials,* Proc. 24th ACM Conference STOC, 1992, 462–467.

[20] R. Paturi: *On the degree of polynomials that approximate symmetric Boolean functions* Proc. 24th ACM Conference STOC, 1992, 468–474.

[21] Razborov, A.A: *Lower bounds for the size of circuits of bounded depth with basis* $\{\oplus, \wedge\}$, Journal Math. Zametki. 41, 1987, 598–607.

[22] Smolensky, R.: *Algebraic methods in the theory of lower bounds for Boolean circuit complexity,* Proc. 19th ACM Conference STOC, 1987, 77–82.

[23] Yao, A.C.: *On ACC and Threshold Circuits,* Proc. 31st IEEE Symposium FOCS, 1990, 619–628.