On the Computational Power of Depth 2 Circuits with Threshold and Modulo Gates*

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

Pavel Pudlák†
Mathematical Institute
Academy of Sciences
Žitná 25, 115 67 Praha 1
Czech Republic
e–mail pudlak@csearn.bitnet

June 5, 2000

Abstract

We investigate the computational power of depth two circuits consisting of MOD^r–gates at the bottom and a threshold gate with arbitrary weights at the top (for short, *threshold–*MOD^r circuits) and circuits with two levels of MOD gates (MOD^p-MOD^q circuits). In particular, we will show the following results.

(i) For all prime numbers p and integers q, r, it holds that if p divides r but not q then all *threshold–*MOD^q circuits for MOD^r have exponentially many nodes.

(ii) For all integers r, all problems computable by depth two $\{AND, OR, NOT\}$–circuits of polynomial size have *threshold–*MOD^r circuits with polynomially many edges.

(iii) There is a problem computable by depth 3 $\{AND, OR, NOT\}$–circuits of linear size and constant bottom fan–in which for all r needs *threshold–*MOD^r circuits with exponentially many nodes.

(iv) For p, r different primes, and $q \geq 2$, k positive integers, where r does not divide q, every MOD^{p^k}-MOD^q circuit for MOD^r has exponentially many nodes.

Results (i) and (iii) imply the first known exponential lower bounds on the number of nodes of *threshold–*MOD^r circuits, $r \neq 2$. They are based on a new method for estimating the minimum length of threshold realizations over predefined function bases, which, in contrast to previous related techniques [GHR92, BS90, KORS91,

*to appear in TCS, June 1997
†Supported by grant A119107 of the Academy of Sciences of the Czech Republic

G93, Gr94] works even if the weight of the realization is allowed to be unbounded, and if the bases are allowed to be nonorthogonal. The special importance of result (iii) consists of the fact that the known spectral–theoretically based lower bound methods for *threshold*–XOR circuits [BS90,KORS91] can provably not be applied to AC_0 functions. Thus, by (ii), result (iii) is sharp. It gives a partial negative answer to the open question whether there exist simulations of AC_0–circuits by small depth threshold circuits which are more efficient than that given by Yao's important result that ACC functions have depth 3 threshold circuits of quasipolynomial weight [Y90]. Finally we observe that our method works also for MOD^p-MOD^q circuits, if p is a power of a prime ((iv) above); see [KW91,YP94] for related results.

A preliminary version of this paper appeared in [KP94].

1 Introduction

1.1 Boolean Circuits and Threshold Circuits

Threshold circuits of small constant depth have become one of the most extensively studied computational models in circuit complexity theory. On the one hand, several interesting arithmetic and Boolean operations (addition, multiplication, division, ACC functions) have surprisingly efficient realizations by threshold circuits of depth not greater than 4 [AB91,BS90,Y90,BHKS93, HHK91]. On the other hand, a lot of quite elementary problems are open. For example, we do not know any superpolynomial lower bound on the number of nodes of depth 2 threshold circuits realizing explicit functions (if no restrictions on the weights are imposed).

In this paper we investigate depth two circuits consisting of a threshold gate at the top and MOD gates at the bottom, and circuits with two levels of MOD gates. We relate their computational power to that of constant depth circuits over the Boolean operations AND, OR, NOT, XOR, and MOD^r.

In our setting *the size of a circuit* is defined to be the number of the gates of the circuit. For a threshold circuit we define its *weight* as the sum of the absolute values of all weights, where we suppose that the weights are always integers. As usual, we denote by $AC_{0,k}$ and $AC_{0,k}[r]$ the classes induced by $\{AND, OR, NOT\}$–circuits, and $\{AND, OR, NOT, MOD^r\}$–circuits, respectively, of polynomial size and constant depth k, as well as $AC_0 = \bigcup_{k\in \mathbb{N}} AC_{0,k}$ and $AC_0[r] = \bigcup_{k\in \mathbb{N}} AC_{0,k}[r]$ and $ACC = \bigcup_{r\in \mathbb{N}} AC_0[r]$.

For convenience, we consider threshold circuits with respect to the $\{1,-1\}$–notation, i.e., the input gates produce $+1$ or -1 and each inner node v is performing a linear threshold operation t_v given by

$$t_v(y_1,\ldots,y_m) \;=\; sgn\left(\sum_{i=1}^{m} a_i y_i\right) \tag{1}$$

where m denotes the fan–in of v and $a_1,\ldots,a_m$ the weights (i.e., multiplicities) of ingoing edges.

It is well known that unbounded weight threshold gates are more powerful than gates with weights which are polynomially bounded in the input size. But large edge weights do not help too much. Any linear threshold function in n variables can be realized with weights in $\exp(O(n \log n))$ [MP68]. Moreover, for any (not necessarily constant) d, polynomial size unbounded weight threshold circuit of depth d can be simulated by depth $d+1$ threshold circuits of polynomial weight [GHR92,GK93].

Recent results on approximating AC_0 and ACC functions by low degree polynomials over $\mathbb{Z}$ and $\mathbb{Z}_p$, p prime, [R87,S87,A89,Y90,ABFR91] give depth efficient realizations of these functions by threshold circuits. In fact, all ACC functions have depth 3 threshold circuits of quasipolynomial weight [Y90]. It seems that AC_0 circuits can be efficiently simulated by even more restricted types of threshold circuits. Our paper partially answers this open question (studied, e.g., in [BS90,ABFR91,T91,B92]).

A lot of recent papers on circuit complexity considers depth 2 circuits over threshold gates and MOD gates [B90, BS90, ABFR91, KW91, GHR92, KORS91]. The problem of proving exponential lower bounds on the number of nodes has been solved for threshold circuits of depth two where the weights of the threshold functions are subexponential [HMPST87, K90, KW90], i.e., there are exponential lower bounds for the weight but not for the size. Another extensively studied type of circuits are *threshold*–MOD^2 circuits [B90, BS90, KORS91]. Functions having *threshold*–MOD^2 circuits with polynomially many nodes are called *polynomial threshold functions*. *Goldmann, Håstad* and *Razborov* have shown that there are quadratic threshold functions which need exponential weight depth 2 threshold circuits. *Bruck* and *Smolensky* found a method based on spectral analysis for proving exponential lower bounds on the number of nodes of unbounded weight *threshold*–MOD^2 circuits. In particular, they showed that there are even symmetric functions which need exponential size in this model.

As this method uses the orthogonality of MOD^2 functions, it can not be generalized to unbounded weight *threshold*–MOD^r circuits for $r \geq 3$. Using an approach based on probabilistic arguments we give the first lower bound method for circuits of this type. We prove that MOD^q needs exponential size *threshold*–MOD^r circuits if there is some prime p dividing q but not r. This improves a very recent result of *Goldmann* [G93] who used the discriminator method [HMPST87,GHR92] for showing the same under the restriction of polynomially bounded weights.

One initial question for this paper was whether all AC_0 functions can be computed by *threshold*–MOD^2 circuits of quasipolynomial size. We know that this is definitely not true for ACC functions. All spectral coefficients of the inner product mod 2 function are exponentially small and this gives an exponential lower bound [BS90]. But AC_0 functions always have large spectral coefficients [LMN89]. The lower bound method of [BS90] based on spectra can not be applied. *Bruck* and *Smolensky* [BS90] found AC_0 functions which are not polynomial threshold functions. But these functions have *threshold*–MOD^2 circuits of quasipolynomial weight. Do all AC_0 functions have quasipolynomial size or quasipolynomial weight *threshold*–MOD^r circuits for some $r \in \mathbb{N}$?

In section 3 we give a negative answer to this question. There exists a function computable by $AC_{0,3}$ circuits of constant bottom fan–in which needs exponential size *threshold–*MOD^r circuits for all $r \geq 2$. This is a sharp bound in the sense that for all integers r, $AC_{0,2}$ functions have polynomial weight *threshold–*MOD^r circuits. We show that in section 4.

In [R87,S87] the fundamental observation was made that $AC_0[p]$ functions can be represented by randomized $\mathbb{F}_p$–polynomials of polylogarithmic degree. This is crucial for the simulation results [A89,Y90] mentioned above. In [S87] it was used to prove, e.g. that $MOD^p \notin AC_0[q]$ for all different primes p, q. But apart from *Yao's* simulation result mentioned above very little is known about circuits over different (prime) MOD operations. The problem has also been studied in the context of non-uniform automata over groups [BST90], in particular ACC can be characterized using such automata.

For depth 2 circuits with a MOD^m gate on the top and arbitrary symmetric functions on the bottom, *Krause and Waack* [KW91] proved an exponential lower bound for the sequence identity function. The proof is based on communication complexity theory. Using a different technique, *Yan and Parberry* [YP94] proved an exponential lower bound for MOD^2–MOD^p, p prime, circuits computing conjunction. The lower bound problem for similar circuits of depth greater than 2 is open.

In section 5 we show that our lower bound technique works also in the case of circuits consisting of two layers of MOD gates (MOD^r–MOD^q circuits). In particular, we prove that for p, r different primes, and $q \geq 2$, k positive integers, where r does not divide q, every MOD^{p^k}-MOD^q circuit for MOD^r has exponential size.

1.2 Representing Boolean functions as a threshold function over given function bases

Most of the time we shall consider depth two circuits with a threshold gate at the top. Since we represent threshold functions as mappings vectors of $\{1, -1\}^n$ to $\{1, -1\}$, we shall represent the functions on the first level and the function computed by the circuit as mappings from $\{0, 1\}^n$ to $\{1, -1\}$. In section 5 we shall return to the more standard notation with both domain and range boolean.

Let $f : \{0, 1\}^n \longrightarrow \{1, -1\}$ be a Boolean function and H be a set of basis functions $h : \{0, 1\}^n \longrightarrow \{1, -1\}$. A threshold representation of f over H is a collection $\{w_h;\ h \in H\}$ of integers so that for all inputs $x \in \{0, 1\}^n$

$$f(x) \ = \ sgn\left(\sum_{h \in H} w_h \cdot h(x)\right).$$

The relevant cost measures of such representations are the size and the weight ($\sum_{h \in H} |w_h|$). It is more convenient to use just the number of all $h \in H$ with $w_h \neq 0$ which is one less than the size; we shall call it *the length*. In this paper we prove lower bounds on the length. Our techniques work also in the case when $f(x)$ is represented as the sign of a constant degree polynomial of the functions $h \in H$, which is, perhaps, not a particularly interesting generalization, but we need it in the proofs anyway.

Definition 1.1 *Let $l_H(f)$ and $T_H(f)$ denote the minimum length and minimum weight, respectively, of a representation of f as threshold function over H. Let $l_H(f) = \infty$ if there is no such representation.*

We will call H a *complete* basis if H generates all of $\mathbb{R}^{\{1,-1\}^n}$ as a real vector space. It is straightforward to check that if H is complete then each Boolean function has a threshold representation over H.

In this paper we investigate the basis of MOD^r functions over $\{0,1\}^n$, $r, n \in \mathbb{N}$, consisting of all functions of the type $MOD^r_{\vec{a},c}$, $\vec{a} \in \mathbb{Z}_r^n$ and $c \in \mathbb{Z}_r$. For all inputs $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ let

$$MOD^r_{\vec{a},c}(x) \;=\; \begin{cases} -1, & \text{if } \sum_{i=1}^n a_i x_i + c \equiv 0 \; mod \; r. \\ 1, & \text{otherwise.} \end{cases}$$

We will denote by MOD^r_n the function $MOD^r_{\vec{e},0}$, where $\vec{e}$ denotes the vector consisting only of ones. A quite straightforward calculation shows that for all natural $r \geq 2$ the basis of n–ary MOD^r functions is complete. Consequently, $l_{mod\ r}(f) \;\leq\; 2^n$ for each Boolean function f.

Further observe that $l_{mod\ r}(f)$ and $T_{mod\ r}(f)$ correspond to the minimal size and the minimal weight, respectively, of *threshold–MOD^r* circuits for f.

Clearly, MOD^2 functions are exactly the $\mathbb{F}_2$ linear functions. If the domain is supposed to be $\{1,-1\}^n$ then $MOD^2_{\alpha,0}$ equals the monomial $x^\alpha = \prod_{i,\alpha_i=1} x_i$. We write, for short, x^α instead of $MOD^2_{\alpha,0}$ and denote $l_\oplus = l_{mod\ 2}$.

Previous lower bound results on threshold representations are based on three different techniques, the *discriminator* method developed in [HPMST87], a geometric method for estimating probabilistic communication complexity [K91,KW91] and a *spectral* method for orthogonal bases developed in [BS90]. We give short descriptions of the first and the third method.

Proposition 1.1 *(Discriminator Lemma) Suppose that f can be represented as a threshold function over H. Then for all probability distributions R on the input set there is $h \in H$ so that $|E_R[f \cdot h]| \;\geq\; T_H(f)^{-1}$.* $\square$

Consequently, for proving exponential lower bounds on $T_H(f)$ one has to construct a probability distribution R on $\{0,1\}^n$ so that $\max\{|E_R[f \cdot h]|\}$ is exponentially small in n. This technique has been successfully applied in several interesting situations [HPMST87, GHR92, G93, MSS91], but as representations of large weight may have small length this method is not suited for estimating representation length.

Till now only one general method has been known for deriving exponential lower bounds on $l_H(f)$ for complete function bases H [BS90].

Proposition 1.2 *Consider the real vector space* $\mathbb{R}^{\{1,-1\}^n}$ *with respect to the positive definite scalar product*

$$\langle f,g\rangle = 2^{-n} \sum_{x\in\{1,-1\}^n} f(x)g(x)$$

and suppose that the functions in H *form an orthogonal system with respect to this scalar product. Then* $l_H \geq (\max\{|\langle f,h\rangle|\,,\ h\in H\})^{-1}$ *for all Boolean functions* $f:\{1,-1\}^n \longrightarrow \{1,-1\}$ $\square$.

This method is formulated in a more general fashion of generalized spectral coefficients and "nearly" orthogonal bases in [KORS91]. But the only natural example of an orthogonal basis is the set $\{x^\alpha,\ \alpha\in\{0,1\}^n\}$ of all $\mathbb{F}_2$–linear, respectively MOD^2 functions. The corresponding representation $f=\sum_\alpha \langle f,x^\alpha\rangle x^\alpha$ is called the spectral representation of the Boolean function f. Exponential lower bounds on $l_\oplus$ can be derived by showing that all spectral coefficients of f are exponentially small. This has been done, e.g., for the inner product mod 2 and the quadratic sum function [BS90].

However, AC_0 functions in n variables have always large spectral coefficients of norm at least $2^{-\log^c n}$ for some constant c. This can be easily derived from the following famous result of *Linial, Mansour, Nisan* [LMN90].

Proposition 1.3 *Let* f *be an* n*–ary Boolean function computable by an unbounded fan–in* $\{AND, OR, NOT\}$ *circuit of depth* d *and size* M*. Then for all* $t\leq n$

$$\sum_{\alpha\in\{0,1\}^n, |\alpha|>t} \langle f,x^\alpha\rangle^2 \ \leq\ M2^{-\frac{1}{4}t^{\frac{1}{d+2}}}. \quad \square$$

The resulting questions are how to prove exponential lower bounds on l_H for non–orthogonal function bases H such as for the MOD^r functions, $r\neq 2$, and how to prove exponential lower bounds on $l_\oplus$ for functions which have large spectral coefficients such as the AC_0 functions? Below we solve these problems by applying a new lower bound method which will be described in the next section. In particular, we prove the following two theorems.

Theorem 1 *Suppose a prime* p *does not divide an integer* $q>1$*. Then* $l_{mod\ q}(MOD_n^p) \geq c^n$ *for some* $c>1$.

Note that this implies a more general statement:

Corollary 1.1 *Suppose a prime* p *divides* r *but does not divide any of the numbers* $q_1,\ldots,q_k$*. Then, for some* $c>1$*, every depth two circuit for* MOD^r *consisting of* MOD^{q_i} *gates at the bottom,* $1\leq i\leq k$*, and a threshold gate at the top has size* $\geq c^n$.

<u>Proof:</u> This is because the MOD^{q_i} gates can be replaced by a MOD^q gate where q denote the smallest common multiple of the q_i. The number p obviously does not divide q. $\Box$

The Sipser function $S_{l,k,2}$ depending on $2lk$ variables $\{x_{i,j}, y_{i,j};\ 1 \le i \le l, 1 \le j \le k\}$ is defined as

$$S_{l,k,2}(x,y) \;=\; \wedge_{i=1}^{l} \vee_{j=1}^{k} (x_{ij} \wedge y_{ij}).$$

Theorem 2 *For every q there exists $c > 0$ so that* $l_{mod\ q}(S_{n,n,2}) \;\ge\; 2^{n^c}$.

In section 5 we apply the same technique to depth 2 circuits with both levels of counting modulo gates.

2 The General Method

Our results are based on two techniques: (1) random restrictions, and (2) estimating the *voting polynomial degree* of Boolean functions (see the definition below). By applying a random restriction we reduce the bottom fan-in and then apply results on the degree of the restricted boolean function.

As the bottom functions are counting modulo functions, we cannot apply the usual random restrictions, instead we randomly divide variables into two parts and choose a suitable subset of inputs for the second part. Then we use a specific property of threshold functions which enables us to average over this set of inputs. (Lemma 2.2, which formalizes this argument, is stated more generally: instead of a subset of inputs, it talks about probability distributions.) This destroys any influence of large fan-in gates on the first set of variables with high probability, if the circuit is small. Thus, if the original circuit was small, we are left with a circuit with a small bottom fan-in. Hence the function computed on these variables must have small degree. So, if a random restriction produces a function with large degree, it cannot be computed by a small circuit.

In this section we consider the simplest case, *threshold–MOD^2* circuits. There we can determine the partition by another set of variables z and we consider all inputs for the second half of the variables.

Definition 2.1 *For all Boolean functions f let* $\deg(f)$ *be the minimal number k for which f can be written as threshold function over functions depending on at most k variables. Equivalently,* $\deg(f)$ *denotes the minimal bottom fan–in for which f has a depth two realization with a threshold gate at the top.*

There are several important results on the voting polynomial degree. Clearly, $\deg(f) \le n$ for all n–ary Boolean functions f. In [ABFR90] it is shown that $\deg(XOR_n) = n$. In the next section we give a generalization: It holds that $\deg(MOD_p^n) \ge \lfloor \frac{n}{p-1} \rfloor$ for all primes p. Observe further the following result of *Minsky* and *Papert* [MP68].

Define, for all $l, k \in \mathbb{N}$, the function $P_{k,l}$, depending on $x_{i,j},\ 1 \leq i \leq k, 1 \leq j \leq l$, by

$$P_{l,k}(x) \ = \ \bigwedge_{i=1}^{l} \bigvee_{j=1}^{k} x_{i,j}.$$

Lemma 2.1 *For all $n \in \mathbb{N}$ it holds that* $\deg(P_{n,4n^2}) \ \geq \ n$. $\square$

Obviously, functions f of big voting polynomial degree may have very sparse realizations as thresholds of linear functions (take, e.g., MOD^2). In the following we give a procedure for constructing a hard function from f.

Definition 2.2 *For all n–ary Boolean functions $f = f(x)$ and all $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n), z = (z_1, \ldots, z_n) \in \{0,1\}^n$ let*

$$f^{op}(x, y, z) = f(u_1, \ldots, u_n)$$

where for all i, $1 \leq i \leq n$, $\quad u_i = (\bar{z}_i \wedge x_i) \vee (z_i \wedge y_i)$.

Proposition 2.1 *For all Boolean functions f, $l_{\oplus}(f^{op}) \ \geq \ 2^{\deg(f)}$.*

Proof:

Observe that for each assignment c to the z–variables (for short, z–assignment) $(f^{op})^c$ is equal to f applied to a subset of the inputs. Specifically, $(f^{op})^c$ depends on those x_i for which $c(z_i) = 0$ and on those y_i for which $c(z_i) = 1$, the remaining n x– and y–variables are redundant.

Let G be a minimal set of $3n$–ary linear functions so that $f^{op} \ = \ sgn\left(\sum_{g \in G} w_g \cdot g\right)$. Clearly, for each z–assignment c the induced threshold representation for $(f^{op})^c$ over $G^c = \{g^c;\ g \in G\}$ contains basic functions depending on redundant variables.

The crucial property is that those basic functions can be removed from the representation. This is due to the following lemma which is the key also for other lower bounds that we shall prove below.

Lemma 2.2 *Let $f : U \to \{-1, 1\}$ be a function, let $g_i : U \times V \to \{-1, 1\},\ i \in I$ be some gates, and let $w_i \in R,\ i \in I$ be arbitrary weights. Suppose*

$$f(u) = \operatorname{sgn} \sum_I w_i g_i(u, v),$$

for every $u \in U,\ v \in V$. Let P_u be a probability distribution on V for $u \in U$. Then

$$f(u) = \operatorname{sgn} \sum_I w_i \mathbf{E}_{P_u} g_i(u, v).$$

Proof: Let u be given. If $\sum_I w_i g(u,v)$ is positive (resp. negative) for every v, then

$$\mathbf{E}_{P_u}\left(\sum_I w_i g_i(u,v)\right) = \sum_I w_i \mathbf{E}_{P_u} g_i(u,v)$$

is positive (resp. negative). □

The proof of the following statement is straightforward and will be left to the reader.

Lemma 2.3 *Let $x_1,\ldots,x_n,y_1,\ldots,y_m$ be $\{1,-1\}$ variables and denote by U the uniform distribution on the set of all y–assignments b. Then for each linear function $x^\alpha y^\beta$, $\beta \neq 0^m$, and each x–assignment a it holds that $E_U[x^\alpha y^\beta(a,b)] = 0$.* □

We now prove Proposition 2.1 by a probabilistic argument.

We will call a basis function $g \in G$ *large* if it depends on at least $\deg(f)$ x– and y–variables. Consider the set of all 2^n assignments to the z–variables as a probability space with the uniform distribution.

We say that a z–assignment c *destroys* $g \in G$ if g depends on variables which are redundant for $(f^{op})^c$. It is straightforward to derive that for each large g the probability that c does not destroy g is at most $2^{-\deg(f)}$.

Consequently, with probability at least $1-|G|\,2^{-\deg(f)}$ there is some z–assignment which destroys all large $g \in G$.

Thus, if $|G| < 2^{\deg(f)}$ we can find some z–assignment c^* fulfilling this property. Following Lemma 2.2 and Lemma 2.3, all g^{c^*}, g large, can be removed from the induced representation of $(f^{op})^{c^*}$.

We obtain a threshold representation which guarantees $\deg((f^{op})^{c^*}) < \deg(f)$ and this is a contradiction. □

As $P^{op}_{n,4n^2}$ obviously belongs to $AC_{0,3}$ we obtain

Corollary 2.1 *There are $AC_{0,3}$ functions f such that all threshold–MOD^2–circuits for f have exponentially many nodes.* □

In the next section we apply this method in a more complicated way to more natural functions and obtain exponential lower bounds also on $l_{mod\ r}$, $r \neq 2$.

3 Lower bound proofs

In this section we shall prove Theorems 1, 2. First we need to estimate the voting polynomial degree of general MOD functions. We shall use complex valued functions. A similar method has been used in [BStr92] to obtain different results.

Lemma 3.1 *For all $n, k \in \mathbb{N}$, $1 < k < n$, it holds that*

$$\deg(MOD_n^p) \;\geq\; \left\lfloor \frac{n}{p-1} \right\rfloor.$$

Proof: It suffices to prove it for n with $(p-1)|n$. Fix a representation

$$MOD_n^p(x) \;=\; sgn(F(x)),$$

where each monomial in F has degree at most $\deg(MOD_n^p)$. Suppose that MOD_n^p depends on the $\{0,1\}$–variables $x_1, \ldots, x_n$ and take an arbitrary partition of $\{x_1, \ldots, x_n\}$ into blocks of size $p-1$, say $\{x_1, \ldots, x_{p-1}\}, \{x_p, \ldots, x_{2p-2}\}, \ldots, \{x_{n-p-1}, \ldots, x_n\}$. For each of these blocks choose p strings from $\{0,1\}^{p-1}$ with different sums mod$p$, say

$$000\ldots0,\; 100\ldots0,\; 110\ldots0,\; \ldots\;,\; 111\ldots1.$$

Now consider only the inputs from $\{0,1\}^n$ which have such a form on the blocks. Then we can think of the MOD function and F as functions defined on $[p]^m = \{0, 1, \ldots, p-1\}^m$, for $m = \frac{n}{p-1}$. Consider the monomials of F of degree less than m restricted to this domain and interpreted as mappings defined on $[p]^m$. They are functions of less than m of variables, hence can be represented as polynomials of degree less than m on $[p]^m$.

Now we can argue as in the case $p = 2$ in [BS90], we only need to choose suitable orthogonal basis of functions. Here we need to use functions $f : [p]^m \rightarrow \mathbb{C}$ with complex values in order to get a nice basis. The basis is

$$B = \left\{ \frac{1}{p^m} \mathrm{e}^{\frac{2\pi i}{p} \cdot \sum_{j=1}^m a_j y_j}; \vec{a} \in [p]^m \right\}.$$

Let us denote by $B' \subseteq B$ the subset of functions of the basis which do not depend on all variables $\{y_1, \ldots, y_m\}$, i.e. $a_i = 0$ for some i. Since F is a sum of functions which do not depend on all variables, we have

$$F(y) = \sum_{f \in B'} w_f f(y),$$

for some $w_f \in \mathbb{C}$. Denote by

$$g_k(y) = \mathrm{e}^{\frac{2\pi i}{p} \cdot \sum_{j=1}^m k y_j}.$$

Thus for $k \neq 0$, g_k is orthogonal to all functions in B'. Consider the following function

$$h(y) = \sum_{k=1}^{p-1} g_k(y).$$

Since

$$1 + \sum_{k=1}^{p-1} \mathrm{e}^{\frac{2\pi i}{p} \cdot k} = \sum_{k=0}^{p-1} \mathrm{e}^{\frac{2\pi i}{p} \cdot k} = 0,$$

we have

$$h(y) = \begin{cases} p-1 & \text{if} \quad \sum y_j \equiv 0 \bmod p \\ -1 & \text{if} \quad \sum y_j \not\equiv 0 \bmod p. \end{cases}.$$

Thus $h(y)$ has the same sign as $MOD_p^m(y)$, consequently the same sign as $F(y)$. Hence

$$0 < \sum_{y \in [p]^m} h(y)F(y) = \langle h, F \rangle = \sum_{k=1}^{p-1} \sum_{f \in B'} \bar{w}_f \langle g_k, f \rangle = 0,$$

which is a contradiction. $\square$

To prove Theorem 1 we have to transform the circuits into a special form. In particular, the transformed circuits will have an extra layer of constant fan-in gates. Therefore we state the the theorem in a slightly more general form given by the proof.

Theorem 3 *Suppose a prime p does not divide q and λ is a constant. Let C be a depth 3 circuit for MOD_n^p with a threshold gate on the top (unbounded weights), arbitrary gates of fan-in $\leq \lambda$ on the middle level and MOD^q gates on the bottom. Then the size of C is $\geq c^n$, where $c > 1$ is a constant which depends only on p, q and λ.*

First we shall show that such circuits can be reduced to a special form. We shall call a $MOD^q_{\vec{a},c}$–gate *simple*, if all coefficients a_i are either 0 or 1; the *domain* of such a gate is the set of variables where the coefficients are 1. Two gates $MOD^q_{\vec{a},c}$, $MOD^{q'}_{\vec{a}',c'}$ are *disjoint*, if they have disjoint domains.

Lemma 3.2 *Let C be a depth 3 circuit with a threshold gate on the top (unbounded weights), arbitrary gates of fan-in $\leq \lambda$, λ a constant, on the middle level and MOD^q–gates on the bottom. Then there exists a circuit C', of the same type such that $|C'| = O(|C|)$, the gates on the middle level are products of fan-in $\leq \lambda'$, where the constant λ' depends only on q and λ, and each product consists of disjoint simple MOD^q–gates.*

Put otherwise, C' is the sign of a polynomial of degree $\leq \lambda'$ of simple MOD^q–gates such that the gates in each monomial are disjoint. Thus it is clear that this lemma does not depend on the particular representation of boolean functions: we can use $1, -1$, or $0, 1$. Note that in the $0, 1$ representation the products are conjunctions.

Proof:We shall use $0, 1$ representation of boolean functions. Consider a gate g on the middle level and the MOD^q gates below. Each MOD^q gate defines a partition of the variables into $\leq q$ blocks according to the coefficients at the variables. Take the smallest common refinement of these partitions. Then the function computed at g can be represented as a function f of simple MOD^q gates whose domains are the blocks of this partition. Represent f as a polynomial, i.e. as a sum of conjunctions. The number of monomials is bounded by a constant, since we assume that g has constant fan-in. Note that in each conjunction there can be only one simple MOD^q gate on a given block, otherwise the conjunction is always false. Merge this sum with the sum in the threshold function on the top and we get a circuit of the required form. $\square$

Proof of Theorem 3: First we sketch the idea of the proof. We would like to use Lemma 2.2 as in the proof of Corollary 2.1. However, for $q > 2$, we cannot simply average over all inputs on a subset of variables, so we use the following idea. Let g be a simple MOD^q–gate and suppose that A is a subset of the domain of g of size $p(q-1)$. Consider q assignments to A, first with all zeros, second with p ones and the rest zeros, third with $2p$ ones and the rest zeros and so on. Then any of these restrictions will produce the same function from the MOD_n^p function. On the other hand, for any assignment to the variables outside of A, we get the same value when we average g over the chosen inputs for A. So in this way we can "kill" g. Now, since A has constant size, we can hit a large gate with high probability by choosing A at random. Of course, to kill all large gates, we have to repeat this several times.

Let us fix a representation of MOD_n^p as threshold function of products of MOD^q functions, each product of size $\leq \lambda$. I.e. fix sets G_t of MOD^q functions and integers w_t, $t \in I$ so that $|G_t| \leq \lambda$ and

$$MOD_n^p(x) = \mathrm{sgn}\left(\sum_{t\in I} w_t \prod_{g\in G_t} g(x)\right).$$

By Lemma 3.2 we can also assume that each product contains disjoint simple MOD^q–gates. We shall say that a simple MOD^q–gate is *large*, if its domain has size at least εn, ($\varepsilon > 0$ will be specified below). We shall say that a set $Z \subseteq \{1,\ldots,n\}$ is *good* for a simple MOD^q–gate, if all x_i, $i \in Z$ are in the domain of the gate.

Choose randomly independently m disjoint sets $A_1,\ldots,A_m \subseteq \{1,\ldots,n\}$ of size $r = p(q-1)$ where $m = \left\lfloor \frac{\varepsilon n}{2r} \right\rfloor$. Let $A = \bigcup_{j=1}^m A_j$; thus $|A| = mr \leq \frac{\varepsilon n}{2}$. We shall estimate the probability for a fixed large $g \in G$ that at least one A_j is good for g. Think of $A_1,\ldots,A_m$ as chosen one after another. Then in the $j-th$ step there remain still at least

$$\varepsilon n - |A| \geq \frac{\varepsilon n}{2}$$

variables in the domain of g. Thus

$$Pr[A_j \text{ is good for } g] \geq \left(\frac{\varepsilon n}{2n}\right)^r = \left(\frac{\varepsilon}{2}\right)^r.$$

Hence

$$Pr[\text{ some } A_j \text{ is good for } g] \geq 1 - \left(1 - \left(\frac{\varepsilon}{2}\right)^r\right)^m \geq 1 - c^{-n},$$

where $c > 0$ is some constant. Thus, if $\sum_t |G_t| < c^n$, we have nonzero probability that there exists a sequence $A_1,\ldots,A_m$ such that for each large gate $g \in \bigcup_t G_t$ there is A_j good for g.

Let $n_1 = n - mr$, $n_2 = mr$. We take an $\varepsilon > 0$ sufficiently small, namely such that

$$\varepsilon n \leq \frac{1}{\lambda}\left\lfloor \frac{n - \frac{\varepsilon n}{2}}{p-1} \right\rfloor \leq \frac{1}{\lambda}\left\lfloor \frac{n - mr}{p-1} \right\rfloor.$$

Then each small gate has size $< \frac{1}{\lambda}\left\lfloor \frac{n_1}{p-1} \right\rfloor$.

In order to apply Lemma 2.2, we split the variables $\{x_1, \ldots, x_n\}$ into two parts Y and Z : Y are those x_i's for which $i \notin A$, Z is the rest; $|Y| = n_1$, $|Z| = n_2$. Now we define a subset $V \subseteq \{0,1\}^{n_2}$. For each $j \in \{1, \ldots, m\}$ divide A_j into blocks $A_{j,1}, \ldots, A_{j,q-1}$ of size p. Think of vectors $v \in \{0,1\}^{n_2}$ as mappings $v : A \to \{0,1\}$. Let V consist of vectors v such that for every $1 \le j \le m$, there exists $1 \le k \le q$ such that

$$\begin{array}{lll} v_i = 1 & \text{for} & i \in A_{j,1} \cup \ldots \cup A_{j,k-1} \text{ and} \\ v_i = 0 & \text{for} & i \in A_{j,k} \cup \ldots \cup A_{j,q-1}. \end{array}$$

Thus for $v \in V$,

$$\sum_{i \in A} v_i \equiv 0 \mod p, \tag{2}$$

and for each A_j and k,

$$Pr\left[\sum_{i \in A_j} v_i \equiv k \bmod q\right] = \frac{1}{q},$$

where v is taken with uniform distribution on V, since p is coprime with q.

Hence, if $g(u,v)$ is a large gate, we have

$$\mathbf{E}_V g(u,v) = \frac{1 \cdot (q-1) + (-1) \cdot 1}{q} = \frac{q-2}{q},$$

where the expectation is taken over the uniform probability distribution on V. Thus $\mathbf{E}_V g(u,v)$ is constant for all u.

Now consider a product $\prod_{g \in G_t} g(u,v)$. Let G_t^L, resp. G_t^S, be the large, resp. small gates of G_t. Let ξ be the variables on which depend the small gates of G_t^S, i.e. the union of their domains. Fix a particular string ξ_0 of values for ξ. Since all gates in the product are disjoint, we can rewrite the conditional expectation as follows:

$$\mathbf{E}_V \left(\prod_{g \in G_t} g(u,v) \mid \xi = \xi_0 \right) =$$

$$= \prod_{g \in G_t^S} g(\xi_0) \cdot \mathbf{E}_V \left(\prod_{g \in G_t^L} g(u,v) \right).$$

We shall show that $\mathbf{E}_V \left(\prod_{g \in G_t^L} g(u,v) \right)$ is constant. Choose a good set from $A_1, \ldots, A_m$ for each gate $g \in G_t^L$ and let ζ be the variables on which these gates depend and which are not in the chosen good sets. Take a particular string ζ_0 of values for ζ and consider

$$\mathbf{E}_V \left(\prod_{g \in G_t^L} g(u,v) \mid \zeta = \zeta_0 \right).$$

Since the gates are disjoint and the probability distribution is independent on the domains (after fixing $\zeta = \zeta_0$), we can distribute the product to

$$\prod_{g \in G_t^L} \mathbf{E}_V \left(g(u,v) \mid \zeta = \zeta_0 \right).$$

By the above argument each term is the constant $\frac{q-2}{q}$, thus the product has the value $\left(\frac{q-2}{q}\right)^{|G_t^L|}$ independently of ζ_0. Thus

$$\mathbf{E}_V \left(\prod_{g \in G_t} g(u,v) \right) =$$

$$= \mathbf{E}_V \left(\prod_{g \in G_t^S} g(\xi) \cdot \left(\frac{q-2}{q} \right)^{|G_t^L|} \right) =$$

$$= \mathbf{E} \left(\prod_{g \in G_t^S} g(\xi) \right) \cdot \left(\frac{q-2}{q} \right)^{|G_t^L|}$$

depends only on small gates. In particular it depends on at most $\lambda \cdot \varepsilon n < \left\lfloor \frac{n_1}{p-1} \right\rfloor$ variables. On the other hand, by (2), $MOD_n^p(u,v) = MOD_{n_1}^p(u)$, for $v \in V$. Hence $MOD_{n_1}^p(u,v)$ is computed using gates of size smaller than $\left\lfloor \frac{n_1}{p-1} \right\rfloor$ which is a contradiction with Lemma 3.1. $\square$

Proof of Theorem 2: We shall use the same idea as in the proof above. The computation is more involved, since it is not so easy to ensure that after applying a restriction to $S_{l,k,2}$ we get a function of the same type.

Let $q \geq 2$ be given. Set

$$\varepsilon = \frac{1}{q+1},$$

and consider the Sipser function $S_{l,k,2}(x,y)$ where

$$k = 8n^2, \quad l = \lceil n^{\varepsilon} \rceil,$$

for some sufficiently large n. Thus we have kl variables x_{ij} and kl variables y_{ij}.

As in the proof of Theorem 1 we shall prove a stronger statement on circuits with an extra middle level of constant fan-in λ. By Lemma 3.2 we assume that we have products of size $\leq \lambda$ of disjoint simple MOD^q gates. Fix such a representation of $S_{l,k,2}$,

$$S_{l,k,2}(x,y) \;=\; \operatorname{sgn} \left(\sum_{\iota \in I} w_\iota \prod_{g \in G_\iota} g(x) \right),$$

where G_ι denote sets of $2kl$–ary simple MOD^q–gates.

We shall say that such a gate g is *large*, if its domain has size at least $\lambda^{-1}(kl)^{1-\varepsilon}$. Choose randomly independently $A_1, \ldots, A_m \subseteq \{1, \ldots, k\} \times \{1, \ldots, l\}$ of size $q-1$ where

$$m = \left\lfloor \frac{(kl)^{1-\varepsilon}}{2\lambda(q-1)} \right\rfloor .$$

Let $A = \bigcup_t A_t$, thus

$$|A| \leq \frac{(kl)^{1-\varepsilon}}{2\lambda}.$$

The meaning of a *good* set is the same as in the previous proof, but now we are only interested in variables x. We shall estimate the probability that at least one A_t is good for some large $g \in \bigcup_\iota G_\iota$:

$$Pr\left[\exists t \quad A_t \text{ is good for } g\right] \geq$$

$$\geq 1 - \left(1 - \left(\frac{\frac{(kl)^{1-\varepsilon}}{2\lambda}}{kl}\right)^{q-1}\right)^m =$$

$$= 1 - \left(1 - \frac{1}{(2\lambda)^{q-1}(kl)^{\varepsilon(q-1)}}\right)^{\left\lfloor \frac{(kl)^{1-\varepsilon}}{2(q-1)} \right\rfloor}.$$

Since

$$\varepsilon(q-1) = \frac{q-1}{q+1} = 1 - \frac{2}{q+1} < 1 - \varepsilon = 1 - \frac{1}{q+1},$$

The expression above is

$$\geq 1 - e^{-(kl)^{c_1}},$$

for some $c_1 > 0$. Thus, if $L = |\bigcup_\iota G_\iota|$ there is a good set A_t for each large gate with probability at least

$$1 - Le^{-(kl)^{c_1}}. \tag{3}$$

Let $p = 8l^2/k$. Consider a random assignment ρ of 0's and 1's to variables y, where 1 is assigned with probability p. We shall use the following Chernoff-type bound, cf. [HR89]:

Lemma 3.3 *Let $S = X_1 + \ldots + X_N$, where X_i are independent 0-1 random variables with $Pr[X_i = 1] = p$, let $M = pN$. Then*

$$Pr[|S - M| \geq \alpha M] \leq 2e^{-\alpha^2 M/3}.$$

□

First we observe that the number of 1's in ρ is at most $2pkl = 16l^3$ with probability

$$\geq 1 - 2e^{-pkl/3} = 1 - 2e^{-\frac{8}{3}l^3}. \tag{4}$$

The number of 1's among y_{ij}'s for a fixed i is at least $\frac{1}{2}pk = 4l^2$ with probability

$$\geq 1 - 2e^{-\frac{1}{4}pk/3} = 1 - 2e^{-\frac{2}{3}l^2},$$

hence this is true for all i with probability

$$\geq 1 - l \cdot 2e^{-\frac{2}{3}l^2}. \tag{5}$$

If $g \in \bigcup_\iota G_\iota$ is small, then the probability that there are fewer than $2p\lambda^{-1}(kl)^{1-\varepsilon}$ pairs (i,j) such that x_{ij} is in the domain of g and ρ fixes y_{ij} to 1 is at least

$$1 - 2e^{-p\lambda^{-1}(kl)^{1-\varepsilon}/3}.$$

Let us estimate the expression $p\lambda^{-1}(kl)^{1-\varepsilon}$:

$$p\lambda^{-1}(kl)^{1-\varepsilon} = \frac{8l^2}{k}\lambda^{-1}(kl)^{1-\varepsilon} = 8\lambda^{-1}l^{3-\varepsilon}k^{-\varepsilon} =$$

$$8\lambda^{-1}\left(\lceil n^\varepsilon \rceil\right)^{3-\varepsilon}\left(8n^2\right)^{-\varepsilon}.$$

This is asymptotically

$$8^{1-\varepsilon}\lambda^{-1}n^{\varepsilon-\varepsilon^2},$$

which is less than or equal to $\lambda^{-1}l$ for n sufficiently large. Thus we can conclude that the probability that there are $\leq \lambda^{-1}l$ such pairs is at least

$$1 - 2e^{-n^{c_2}},$$

for a constant $c_2 > 0$. The probability that this holds for all small $g \in \bigcup_\iota G_\iota$ is at least

$$1 - 2Le^{-n^{c_2}}. \tag{6}$$

Now we can put things together. Our goal is to reduce the circuit so that we can use Lemma 2.1:

We take the random assignment ρ of 0's and 1's to variables y. Then we get a $\bigwedge\bigvee x_{ij}$ over those pairs (i,j) for which $\rho_{ij} = 1$. The estimate (5) gives the probability that $\bigwedge_{i=1}^{l}\bigvee_{j=1}^{4l^2} x_{ij}$ will be a subfunction of it. The estimate (6) gives the probability that small $g \in \bigcup_\iota G_\iota$ will depend on $< \lambda^{-1}l$ remaining variables. Thus it remains to get rid of the large $g \in \bigcup_\iota G_\iota$. With probability estimated by (4), ρ will assign $kl - 16l^3 \geq m(q-1)$ zeros. Thus we can choose $A_1, \ldots, A_m$ so that

$$(i,j) \in A_t \quad \Rightarrow \quad \rho_{ij} = 0.$$

Since we are choosing from a randomly chosen set, this choice is also random, and we can use the estimate (3) plus (4) for a successful choice. Then we estimate the large $g \in \bigcup_\iota G_\iota$ in the same way as in the previous proof. Namely, we consider assignments to those x_{ij}'s for which $\rho_{ij} = 0$ (hence the restricted function does not depend on them) such that for each $A_t = \{(i_1,j_1), \ldots, (i_{q-1}, j_{q-1})\}$, $x_{i_1 j_1} = \ldots x_{i_h j_h} = 1$ and $x_{i_{h+1} j_{h+1}} = \ldots x_{i_{q-1} j_{q-1}} = 0$, for $h = 0, \ldots, q-1$.

The probability that all this can be arranged is given by (3)-(6):

$$1 - Le^{-(kl)^{c_1}} - 2e^{-\frac{8}{3}l^3} - 2le^{-\frac{2}{3}l^2} - 2Le^{-n^{c_2}}.$$

All the exponents are asymptotically $-n^{\alpha}$ for some $\alpha > 0$, thus the probability is positive, if $L < e^{n^{\alpha}}$, for a suitable $\alpha > 0$.

Using the same argument as in the proof of Theorem 3 we get that the restricted function is a threshold function of functions which depends on fewer than l variables which is a contradiction with Lemma 2.1. $\square$

4 An upper bound

In this section we show

Theorem 4 *For all natural r all $AC_{0,2}$ functions have threshold–MOD^r circuits of polynomial weight.*

Proof:

It is sufficient to show that for all primes p $P_{n,n} = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{n} x_{i,j}$ has *threshold–MOD^p* circuits of polynomial weight.

Denote by $A^n = \{0, \ldots, p-1\}^n \times \{1, \ldots, n\}$, and denote for all $(\alpha, i) \in A^n$ by $m^{\alpha,i}$ the MOD^p function defined by

$$m^{\alpha,i}(x_{1,1} \ldots, x_{n,n}) = 1 \quad \Longleftrightarrow \quad \sum_{j=1}^{n} \alpha_j x_{i,j} \equiv 1 \; mod \; p,$$

and $m^{\alpha,i}(x_{1,1} \ldots, x_{n,n}) = 0$ otherwise.

Observe that if for an input matrix x to $P_{n,n}$ the i-th row is nonzero then the vectors α fulfilling $m^{\alpha,i}(x) = 1$ form an affine hyperplane in $\mathbb{F}_p^n$.

Consequently, if we consider A^n as probability space with the uniform distribution then for any fixed input x the following holds.

If $P_{n,n}(x) = 1$ then for all $i_o \in \{1, \ldots, n\}$

$$Pr[m^{\alpha,i}(x) = 1 | i = i_o] = \frac{p^{n-1}}{p^n} = \frac{1}{p}$$

and, consequently, $Pr[m^{\alpha,i}(x) = 1] = \frac{1}{p}$.

On the other hand, if $P_{n,n}(x) = 0$ then there is an $i_o \in \{1, \ldots, n\}$ fulfilling

$$Pr[m^{\alpha,i}(x) = 1 | i = i_o] = 0,$$

i.e., $Pr[m^{\alpha,i}(x) = 1] \leq \frac{n-1}{np} = \frac{1}{p} - \frac{1}{np}$.

Using Lemma 3.3 it is straightforward to prove the existence of numbers $K, M \in O(n^4)$ such that for randomly, independently chosen $(\alpha_1, i_1), \ldots, (\alpha_M, i_M)$ from A^n and each input x to $P_{n,n}$ the following is true.

If $P_{n,n}(x) = 1$ then $Pr[\sum_{l=1}^{M} m^{\alpha_l, i_l}(x) < K] \ < 2^{-n^2}$,

if $P_{n,n}(x) = 0$ then $Pr[\sum_{l=1}^{M} m^{\alpha_l, i_l}(x) > K] \ < 2^{-n^2}$.

Now a standard argument shows the existence of $(\alpha_1, i_1), \ldots, (\alpha_M, i_M)$ in A^n so that for all inputs x

$$P_{n,n}(x) = 1 \quad \Longleftrightarrow \quad \sum_{l=1}^{M} m^{\alpha_l, i_l}(x) > K,$$

and, thus, that $T_{mod\ p}(P_{n,n}) \in O(n^4)$. $\square$

Note that at least in the case of $r = 2$ it is possible to "derandomize" the above construction i.e., to describe the circuit explicitly. What is needed is to compute disjunctions of n variables using a linear function of parities with precision greater than n^{-1}. This can be done using explicit matrices with small bias, see [AGHP92]. We conjecture that similar constructions are possible for all r.

5 Two levels of MOD gates

In this section we prove

Theorem 5 *Let p, r be two different primes, $q \geq 2$, $k \geq 1$ integers, $r \not| q$. Then for some $c > 1$ every MOD^{p^k}–MOD^q circuit for MOD_n^r has size $\geq c^n$.*

The method is based on the following version of Lemma 2.2. In contrast to previous sections we suppose here that all functions map into $\{0, 1\}$.

Lemma 5.1 *Let $f : U \to \{0, 1\}$, $g_i : U \times V \to \{0, 1\}$, $i \in I$, let m be an integer $m \geq 2$. Suppose*

$$f(u) \equiv \sum_{I} g_i(u, v) \mod m \tag{7}$$

for every u, v. Furthermore suppose that $|V|$ has an inverse modulo m. Then

$$f(u) \equiv \sum_{i \in I} \frac{1}{|V|} \sum_{v \in V} g_i(u, v) \mod m.$$

$\square$

The application of this lemma is the same as of Lemma 2.2: if g_i is a large MOD^q gate, we can make $\sum_{v \in V} g_i(u, v) \mod m$ constant and thus we reduce the bottom fan-in of the circuit. However the representation in (7) is not of the form that we have in a circuit with a *sum* modulo m on the top. In order to be able to use Lemma 5.1 we have to transform the circuit in a similar way as in the above proofs. (Note that we are replacing a MOD gate by summation.)

Lemma 5.2 *Suppose that*

$$f(x) = MOD^{p^k}(g_1, \ldots, g_m),$$

where each g_i is a MOD^q function of x. Then f can be represented as

$$f(x) \equiv \sum_{i=1}^{m'} h_i(x) \mod p,$$

where each h_i is a product of at most λ simple disjoint MOD^q gates, λ is a constant and m' is bounded by a polynomial of m.

Proof: It is well-known that

$$x \equiv 0 \mod p^k \quad \leftrightarrow \quad \forall i < k \binom{x}{p^i} \equiv 0 \mod p.$$

Thus counting modp^k can be reduced to counting mod$p$:

$$x \equiv 0 \mod p^k \quad \leftrightarrow \quad P_k(x) \equiv 0 \mod p,$$

where $P_k(x)$ is the polynomial

$$1 - \prod_{i<k} \left(1 - \binom{x}{p^i}^{p-1} \right).$$

Moreover this polynomial takes on only values 0 and 1 for x a nonnegative integer. Expanding the polynomial as a sum of monomials we get a representation as a sum mod$p$ of constant size conjunctions.

The rest is the same as in the proof of Lemma 3.2. $\Box$

Finally we need a lemma which is a weak form of a recent result of *Tsai* [Ts93] (note that the case of a prime m is due to *Smolensky* [S87]).

Lemma 5.3 *Suppose*

$$MOD_n^r(x) \equiv F(x) \mod m,$$

where r is a prime which does not divide m and $F(x)$ is a polynomial, then the degree of F is at least δn for some $\delta > 0$ depending only on r and q. $\Box$

Now the proof of Theorem 5 follows almost exactly the proof of Theorem 1. We choose the U and V in the same way as in the proof of Theorem 1, we have only to check that $|V|$ has inverse modulo m. But $|V| = r^t$ (where t is the number of blocks in the set A) and $m = p$ is a prime different from the prime r. Thus $|V|$ has the inverse. $\Box$

6 Open Problems

It remains open to prove exponential lower bounds on the size of general depth 2 threshold circuits. Another open problem is to answer the question whether AC_0 functions always have depth 2 threshold circuits of quasipolynomial size or weight. We conjecture that this is not the case. But to show an exponential lower bounds on the weight of depth 2 threshold circuits computing an explicit AC_0 functions is still open.

The next step in the ACC problem is to find a function which is not in $AC_{0,3}[m]$ for a composite m. This is open even for depth 3 circuits which use *only* MOD^m gates. A natural conjecture is that $MOD^p \notin AC_{0,3}[m]$, if p is a prime which does not divide m. This conjecture is open also for $AC_{0,2}[m]$, our result gives only a partial answer. In particular we still do not know whether MOD^5 can be computed by polynomial size depth 2 circuits with MOD^6 gates.

Acknowledgement

We would like to thank Mikael Goldmann for helpful discussion and suggesting a correction to an error in the earlier version of this paper. Further special thanks are devoted to Claudia Bertram–Kretzberg for fruitful discussion on the upper bound in section 4 and to anonymous referees for suggesting several improvements.

A major part of the results were obtained while Pavel Pudlák was supported by the Humboldt Fellowship at Universität Dortmund.

7 References

[A89] Allender,E.: *A note on the power of threshold circuits*, Proc. 30-th IEEE Symposium FOCS, 1989, 580–584.

[AB91] Alon,N.,J.Bruck: *Explicit constructions of depth–2 majority circuits for comparison and addition*, Technical Report RJ 8300 (75661) of the IBM Almaden Research Center, San Jose, 1991.

[AGHP92] Alon,N., O.Goldreich, J.Håstad, R.Peralta: *Simple constructions of almost k-wise independent random variables,* Random Structures and Algorithms 3/3, 1992, 289–304.

[ABFR91] Aspnes,J.,R.Beigel,M.Furst,S.Rudich: *The expressive power of voting polynomials*, Proc. 23-th ACM Symposium STOC, 1991, 402–409.

[Ba92] Barrington,D.A: *Quasipolynomial size circuits.* Proc. of IEEE Conf. SC'92, 1992, 86–93.

[BStr92] Barrington, M.D.A., Straubing, H.: *Complex polynomials and circuit lower bounds for modular counting,* Proc. of LATIN 92, 24-31.

[BST90] Barrington, M.D.A., Straubing, H., Thérien, D.: *Non-uniform automata over groups.* Information and Computation 89/2, 1990, 109-132.

[Be92] Beigel,R.: *Perceptrons, PP, and the Polynomial Hierarchy,* Proc. of IEEE Conf. SCT'92, 14–19.

[Be93] Beigel,R.: *The polynomial method in circuit complexity,* Proc. of IEEE Conf. SCT'93, 82-95.

[B90] Bruck,J.: *Harmonic analysis of polynomial threshold functions*, SIAM Journal of Discrete Mathematics, 3, Nr. 22, 1990, 168–177.

[BKS92] Bruck,J.,Th.Hofmeister,Th.Kailath,K.Y.Siu: *Depth efficient networks for division and related problems,* Technical Report 1992, to appear in 1993.

[BS90] Bruck,J.,R.Smolensky: *Polynomial threshold functions, AC^0 functions and spectral norms,* Proc. 31-th IEEE Symp. FOCS, 1990, 632–641.

[G93] Goldmann,M.: *A note on the power of majority gates and modular gates,* preprint, 1993.

[GHR92] Goldmann,M.,J.Håstad,A.A.Razborov: *Majority Gates versus general weighted threshold gates*, Computational Complexity 2 (1992), 277–300.

[GK93] Goldmann,M.,M.Karpinski: *Simulating Threshold Circuits by Majority Circuits,* Proc. of 25th ACM Symp. STOC, 1993, 551–560.

[Gr94] Grolmusz, V.: *A weight-size trade-off for circuits with MOD^m gates,* Proc. of 25-th ACM Symp. STOC, 1993, 68–74.

[HR89] Hagerup, T., Rüb, C.: *A guided tour of Chernoff bounds,* Inf. Process. Letters 33, 1989/90, 305-308.

[HMPST87] Hajnal,A.,W.Maass,P.Pudlák,M.Szegedy,G.Turán:*Threshold circuits of bounded depth,* Proc. 28-th IEEE Symp. FOCS, 1987, 99–110.

[HHK91] Hofmeister,Th.,W.Hohberg,S.Köhling: *Some notes on threshold circuits and multiplication in depth 4,* Inf. Process. Letters 39 (1991) 219–225

[KORS91] Kailath,Th., A.Orlitsky, V.Roychowdhury, K.Y.Siu: *A geometric approach to Threshold Circuit Complexity*, Proc. 4-th ACM Conf. COLT, 1991, 97–111.

[K90] Krause,M. *Geometric Arguments yield better bounds for threshold circuits and distributed computing,* Proc. of IEEE Conf. SCT'91, 314–322.

[KP94] Krause,M., P.Pudlák: *On the Computational Power of Depth 2 Circuits with Threshold and Modulo Gates,* Proc. 25-th ACM Symp. STOC, 1993, 48–57.

[KW91] Krause,M.,S.Waack: *Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in*, Proc. of 32-th IEEE Symp. FOCS, 1991, 777–787.

[LMN89] Linial,N.,Y.Mansour,N.Nisan: *Constant Depth Circuits, Fourier Transforms, and Learnability,* Proc. of 30-th IEEE Symp. FOCS, 1990, 574–579.

[MSS91] Maass,W.,G.Schnitger,E.Sonntag: *On the Computational Power of Sigmoid versus Boolean Threshold Circuits,* Proc. of 32-th IEEE Symp. FOCS, 1991, 767–776.

[MP68] Minsky,M.,S.Papert: *Perceptrons,* MIT Press, Cambridge 1988, Expanded edition. (First edition appeared in 1968.)

[R87] Razborov, A.A: *Lower bounds for the size of circuits of bounded depth with basis* $\{\oplus, \wedge\}$*,(Russian)* Matematiceskie Zametki 41, 1987, 598–607.

[S87] Smolensky,R.: *Algebraic methods in the theory of lower bounds for Boolean circuit complexity,* Proc. of 19-th ACM Symp. STOC, 1987, 77–82.

[T91] Tarui,J.: *Randomized polynomials, threshold circuits, and the polynomial hierarchy,* Proc. of Symp. STACS'91, 238–251.

[Ts93] Tsai,S.-C.: *Lower bounds on representing boolean functions as polynomials in* $\mathbb{Z}_m$*,* Proc. of IEEE Conf. SCT'93, 96-101.

[YP94] Yan, P.Y., Parberry, I.: *Exponential size lower bounds for some depth three circuits.* Information and Computation 112, (1994),117–130.

[Y90] Yao,A.C.: *On ACC and Threshold Circuits,* Proc. of 31-th IEEE Symp. FOCS, 1990, 619–628.