Electronic Colloquium on Computational Complexity
ECCC Technical Reports Series 1994 Report Nr: TR94-023

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

Matthias Krause[†] Pavel Pudlák[‡]

Received *December 22, 1994*

Abstract. We investigate the computational power of depth two circuits consisting of MOD^r–gates at the bottom and a threshold gate 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 (quasi) polynomial size can be represented by threshold–MOD^r circuits with (quasi)polynomially many edges.
(iii) There is a problem computable by depth three $\{AND, OR, NOT\}$–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 p 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 lower bound method for the representation length of functions as threshold functions over predefined function bases, which, in contrast to previous related techniques works even if the edge weights are allowed to be unbounded and if the bases are nonorthogonal. The special importance of result (iii) consists in the fact that the known spectral–theoretically based lower bound methods for threshold–XOR circuits can provably not be applied to AC_0–functions. Thus, by (ii), result (iii) is quite sharp and 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 $ACC^* \subseteq TC^*_{0,3}$ [Y90]. Finally we observe that our method works also for MOD^p-MOD^q circuits, if p is a power of a prime.

[†] Lehrstuhl Informatik II, Universität Dortmund, D 44221 Dortmund. E–mail krause@daedalus.informatik.uni-dortmund.de
[‡] Mathematical Institute, Academy of Sciences of Czech Republic, Žitná 25, CZ 115 67 Praha 1. E–mail pudlak@csearn.bitnet

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 arithmetic and Boolean operations of practical interest (addition, multiplication, division, ACC–functions) have surprisingly efficient realizations by threshold circuits of depth smaller than 4 [AB91,BS90,Y90,BHKS93, HHK91]. On the other hand, a lot of quite elementary problems are open. For example, it is open to prove an explicite superpolynomial lower bounds on the number of nodes of depth two threshold circuits.

In this paper we analyze the computational power of 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 do that by relating them to constant depth circuits over the Boolean operations AND, OR, NOT, XOR, MOD^r.

In our setting the size of a circuit is defined to be the number of nodes, where the weight denotes the number of edges. We denote by $AC^*_{0,k}$ and $AC^*_{0,k}[r]$ the classes induced by $\{AND, OR, NOT\}$–circuits, $\{AND, OR, NOT, MOD^r\}$–circuits, respectively, of quasipolynomial size, i.e. size $\exp((\log n)^{O(1)})$ and constant depth k. (The star indicates quasipolynomiality, see e.g. [Ba92] for a motivation why in the present context *quasi*polynomial size classes are considered.) The classes AC^*_0, $AC^*_0[r]$ and ACC^* are defined, correspondingly.

Following [B90,BS90,GHR92] we denote by $\overline{LT}^*_k$ and LT^*_k the classes induced by depth k threshold circuits of quasipolynomial weight, quasipolynomial size, respectively. 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

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

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

Unbounded fan–in circuits of (quasi)polynomial weight and constant depth k with symmetric gate operations can be simulated by (quasi)polynomial weight threshold circuits of depth $k+1$. In contrast to circuits over AND, OR and constant MOD's the computational power of threshold circuits, relative to a fixed architecture, is growing significantly if "large" (i.e. exponential) edge weights are allowed. For $k \in \{1, 2\}$ it is proved that $\overline{LT}^*_k$ is properly contained in LT^*_k [GHR92]. But edge weights don't help too much. It is known that 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 depth d threshold circuit of (quasi)polynomially size can be simulated by depth $d+1$ threshold circuits of (quasi)polynomial weight [GHR92,GK93], i.e., $\overline{LT}^*_k \subseteq LT^*_k \subseteq \overline{LT}^*_{k+1}$, for all $k \geq 1$.

Recent very nontrivial results on realizing and approximating AC_0– and ACC–functions by low degree integer polynomials [R87,S87,A89,Y90,ABFR91] imply very depth efficient realizations of these functions by threshold circuits. In fact, it holds $ACC^* \subseteq \overline{LT}_3^*$ [Y90]. Our paper partially answers the open question (studied, e.g., in [BS90,ABFR91,T91,B92]) whether AC_0^*–circuits can be efficiently simulated by even more restricted types of threshold circuits.

A lot of recent papers on circuit complexity are dealing with depth two circuits over threshold– and MOD–gates [B90,BS90,ABFR91,KW91,GHR92,KORS91]. Our special interest is devoted to the following related complexity classes.

Definition 1.1 *For all natural $r \geq 2$ let $QT[r]$ and $\overline{QT}[r]$ contain all sequences of Boolean functions computable by threshold-MOD^r-circuits of quasipolynomial size, quasipolynomial weight, respectively. Following [B90,BS90] the classes $QT[2]$ and $\overline{QT}[2]$, $r \in \mathbb{N}$ are shortly denoted by QT and $\overline{QT}$, respectively. QT-functions are sometimes called quasipolynomial threshold functions.*

Using the described basic results it is straightforward to see that $QT[r] \subseteq \overline{LT}_3^*$ for all $r \in \mathbb{N}$.

The problem of proving exponential lower bounds on the number of nodes is solved for threshold circuits of depth two if the number of edges is bounded [HPMST87,K90,KW90], and for threshold-MOD^2–circuits [B90,BS90,KORS91].

The underlying methods enabled to prove a number of nontrivial relations between QT–, LT– and $AC_0[r]$–classes. It has been shown that $\overline{LT}_2^* \not\subseteq QT$ [BS90] and $QT \not\subseteq \overline{LT}_2^*$ [GHR92], i.e., both $\overline{LT}_2^*$ and QT are properly contained in $\overline{LT}_3^*$.

Till now the problem of proving exponential lower bounds on the size of threshold-MOD^r–circuits, $r \neq 2$, was open. This has to do with natural limitations of the spectral–theoretic approach used in [B90,BS90,KORS91] which will be briefly discussed in the next subsection. Using a more straightforward approach (section 2) we establish the first effective lower bound method for threshold-MOD^r–circuits, $r \in \mathbb{N}$. This allows to prove that $MOD^r \not\in QT[q]$ if there is some prime p dividing r but not q (section 3). Consequently, $QT[r]$ is a proper subclass of $\overline{LT}_3^*$ for all $r \in \mathbb{N}$. Our result also improves a very recent result of *Goldmann* [G93] saying that if there is a prime number p dividing r but not q then $MOD^r \not\in \overline{QT}[q]$.

Exhibiting the orthogonal structure of the inner product mod 2 function it has been shown that $AC_0[2] \not\subseteq QT$ [BS90] and $AC_0[2] \not\subseteq \overline{LT}_2^*$ [HPMST87]. Using an alternative lower bound method the last relation has been generalized in [KW90], i.e., $AC_0[r] \not\subseteq \overline{LT}_2^*$ for all $r \geq 2$. Observe that our result mentioned above yields $AC_0[r] \not\subseteq QT$ for all $r \geq 2$. In other words, both $\overline{LT}_2^*$– and QT–circuits are not strong enough for efficiently simulating ACC–circuits.

However, what about AC_0? The L_∞–norm of AC_0^*–functions is at least $\exp(-\log^{O(1)} n)$ (subsection 1.2) and this prevents a successful application of the known lower bound methods for $\overline{LT}_2^*$ and QT to AC_0^*–functions. The resulting question is whether AC_0 is even contained in one of the classes LT_2^*, $\overline{LT}_2^*$, or $QT[r]$, $r \geq 2$.

In section 3 we give a partial negative answer. There exists a problem computable by $AC_{0,3}$–circuits of constant bottom fan–in which does not belong to $QT[r]$ for all $r \geq 2$. This is a sharp bound. In section 4 we will show that for all integers r $AC_{0,2}^* \subseteq QT[r]$.

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 one main basis for the simulation results [A89,Y90] mentioned above and allows to prove, e.g., that $MOD^p \not\in AC_0^*[q]$ for all different primes p, q. But apart from *Yao's* result $ACC^* \subset \overline{LT}_3^*$ very little is known about circuits over different (prime) MOD–operations.

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 circuits of depth greater two 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 threshold functions over given function bases

A representation of a Boolean function f as threshold function over a basis H of Boolean functions over $\{0,1\}^n$ is given by a collection $\{w_h;\ h \in H\}$ of integers so that for all inputs x

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

The relevant cost measures of such representations are the length, given by the number of all $h \in H$ with $w_h \neq 0$, and the weight, given by $\sum_{h\in H} |w_h|$. In this paper we prove lower bounds on the length; let us note that our techniques work also in the case when $f(x)$ is represented as the sign of a small (sublogarithmic) degree polynomial of the functions $h \in H$.

Important: If not stated otherwise, all Boolean functions are supposed to map into $\{1, -1\}$. We are switching from the usual $\{0,1\}$– to the $\{1,-1\}$–notation by replacing 0 by 1 and 1 by -1.

Definition 1.2 *Let $l_H(f)$ and $T_H(f)$ denote the minimum length, 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 to be a *complete* basis if H generates the whole $\mathbb{R}^{\{1,-1\}^n}$ as 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 $MOD^r_{\vec{a},c}$, $\vec{a}, c \in \{0, \ldots, r-1\}^{n+1}$. 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 theoretic 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* $|E_R[f \cdot h]| \;\geq\; T_H(f)^{-1}$. □

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 linear functions. The corresponding representation $f = \sum_\alpha \langle f, x^\alpha\rangle x^\alpha$ is called the spectral representation of the Boolean function f, where the maximum over all values $|\langle f, x^\alpha\rangle|$ is called the L_∞–norm of f. Proposition 2.1 yields

Corollary 1.1 *[BS90]* $l_\oplus(f) \geq L_\infty(f)^{-1}$. $\square$

Using this method exponential lower bounds on $l_\oplus$ are shown, e.g., for the inner product mod 2 function and for the quadratic sum function [BS90]. However, the following lemma shows that Proposition 2.1 can not be applied to AC_0^*–functions.

Lemma 1.1 *For all* $(f_n)_{n\,\in\,\mathbb{N}} \in AC_0^*$ *it holds* $L_\infty(f_n)^{-1} \in \exp\left(\log^{O(1)} n\right)$.

This is a quite straightforward corollary of the following result of *Lineal, Mansour, Nisan* [LMN90].

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

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

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

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

Note that this implies a more general statement:

Corollary 1.2 *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 $\geq\ c^n$.*

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

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

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

2 The General Method

Our results are based on estimating the *voting polynomial degree* of Boolean functions.

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, $\deg(f) \leq n$ 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 $\deg(MOD_p^n) \geq \lfloor \frac{n}{p-1} \rfloor$ for all primes p. Observe further the following result of *Minsky* and *Papert* [MP68].

Lemma 2.1 *For all $n \in$ IN it holds* $\deg(P_{n,4n^2}) \geq n$, □

where for all $l, k \in$ IN the function $P_{k,l}$, depending on $x_{i,j}$, $1 \leq i \leq k, 1 \leq j \leq l$, is defined as

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

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

Definition 2.2 *For all n–nary 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, $1 \leq i \leq n$, $u_i = (\bar{z}_i \wedge x_i) \vee (z_i \wedge y_i)$.

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

Proof:

Observe that for each assignment c to the z–variables (for short, z–assignment) $(f^{op})^c$ and f are equivalent. $(f^{op})^c$ 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*

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

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

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

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

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

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 Boolean 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 $E_U[x^\alpha y^\beta(a, b)] = 0$.* □

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

Corollary 2.1 *$AC_{0,3} \not\subseteq QT$.* □

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.

Lemma 3.1 *For all $n \in \mathbb{N}$ and prime numbers p it holds* $\deg(MOD_n^p) \geq \left\lfloor \frac{n}{p-1} \right\rfloor$. □

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 $m = \frac{n}{p-1}$. The monomials of degree $< m$ of F are some general functions, but can be represented as polynomials of degree $< m$ 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 $f : [p]^m \to \mathbb{C}$ with complex values in order to get a nice basis. The basis is

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

Let us denote by $B' \subseteq B$ the subset of functions of the basis which do not depend on all variables $\{x_1, \ldots, x_m\}$, i.e. $a_i = 0$ for some i. Since F is a sum of functions which do not depend on all variables, we have

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

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

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

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

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

Since

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

we have

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

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

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

which is a contradiction. □

Instead of proving Theorem 1 we prove a more general theorem.

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 $\leq \lambda$ 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. □

Proof of Theorem 3: Let us fix a representation of MOD^p_n 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

$$MOD^p_n(x) = \operatorname{sgn}\left(\sum_{t\in I} w_t \prod_{g\in G_t} g(x)\right).$$

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 $Z \subseteq \{1, \dots, n\}$ 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 $A_1, \dots, A_m \subseteq \{1, \dots, n\}$ 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 $A_1, \dots, A_m$ 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

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

where $c > 0$ is some constant. Thus, if $\sum_t |G_t| < c^n$, we have nonzero probability that there exists a sequence $A_1, \dots, A_m$ 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 $\varepsilon > 0$ 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 $\{x_1, \dots, x_n\}$ 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 $j \in \{1, \dots, m\}$ divide A_j into blocks $A_{j,1}, \dots, A_{j,q-1}$ 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 $1 \leq j \leq m$, there exists $1 \leq k \leq q$ 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 the small gates of G_t^S depend, 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: 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

$$\geq 1 - 2e^{-4pkl/3} = 1 - 2e^{-\frac{16}{3}l^3}. \tag{4}$$

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 is at least

$$1 - 2e^{-4p\lambda^{-1}(kl)^{1-\varepsilon}/3}.$$

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 n to the power

$$\varepsilon(3-\varepsilon)-2\varepsilon=\varepsilon-\varepsilon^{2}<\varepsilon.$$

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 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 $A_t=\{(i_1,j_1),\ldots,(i_{q-1},j_{q-1})\}$, $x_{i_1j_1}=\ldots x_{i_hj_h}=1$ 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):

$$1-Le^{-(kl)^{c_1}}-2e^{-\frac{16}{3}l^3}-2le^{-\frac{2}{3}l^2}-2Le^{-n^{c_2}}.$$

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 1 we get that the restricted function is a threshold function of functions which depends on $<l$ variables which is a contradiction with Lemma 2.1. □

4 An upper bound

In this section we show

Theorem 4 *For all natural r* $AC^{*}_{0,2}\subseteq\overline{QT}[r]$.

Proof:

It is sufficient to show that for all primes p $P_{n,n} = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{n} x_{i,j}$ is in $\overline{QT}[p]$.

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

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

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

Observe that if for an input matrix x to $P_{n,n}$ the i-th row is nonzero then for all $b \in \{1.\dots,p-1\}$ the vectors α fulfilling $m^{\alpha,b,i}(x) = 1$ 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 it holds the following.

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

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

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

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

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

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

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 $(\alpha_1, b_1, i_1), \dots, (\alpha_M, b_M, i_M)$ from A^n and each input x to $P_{n,n}$ the following is true.

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

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

Now an standard argument shows the existence of $(\alpha_1, b_1, i_1), \dots, (\alpha_M, b_M, i_M)$ in A^n so that for all inputs x

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

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

5 Two levels of MOD gates

In this section we prove

Theorem 5 *Let p, r be primes, $q \geq 2$, $k \geq 1$ integers. Then for some $c > 1$ every MOD^{p^k}-MOD^q circuit for MOD_r^n 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 the inverse mod$m$. *Then*

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

□

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.

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. □

Finally we need the following result of *Smolensky* [S87] which was recently extended to composite numbers m by *Tsai* [Ts93]

Lemma 5.3 *Suppose*

$$MOD_r^n(x) \equiv F(x) \mod m,$$

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. □

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. □

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 show that $AC_0 \not\subseteq \overline{LT_2}$. We conjecture that even $AC_{0,3} \not\subseteq LT_2$.

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.

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. A further special thank is devoted to Claudia Bertram–Kretzberg for fruitful discussion on the upper bound section 4.

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*, Proceedings der 30. 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.

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

[Ba92] Barrington,D.A: *Quasipolynomial size circuits.* Proc. IEEE Conference Structure in Complexity, 1992, 86–93.

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

[Be93] Beigel,R.: *The polynomial method in circuit complexity* 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. 31th IEEE Conference 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*, J. of Computational Complexity 2 (1992), 277–300.

[GK93] Goldmann,M.,M.Karpinski: *Simulating Threshold Circuits by Majority Circuits.* Proc. 25th ACM Conference STOC, 1993.

[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. 28th IEEE Conf. FOCS, 1987, 99–110.

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

[LMN89] Linial,N.,Y.Mansour,N.Nisan: *Constant Depth Circuits, Fourier Transforms, and Learnability.* Proc. 30th IEEE Conference FOCS, 1990.

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

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

[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. 32th IEEE Conference FOCS, 1991, 777–787.

[MSS91] Maass,W.,G.Schnitger,E.Sonntag *On the Computational Power of Sigmoid versus Boolean Threshold Circuits* Proc. of FOCS'91, 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* $\{\oplus, \wedge\}$, Journal Math. Zametki. 41, 1987, 598–607.

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

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

[Ts93] Tsai,S.-C.: *Lower bounds on representing boolean functions as polynomials.* SCT'93, 96-101.

[YP94] Yan, P.Y., Parberry, I.: *Exponential size lower bounds for some depth three circuits.* Information and Computation, to appear.

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