On Small Depth Threshold Circuits

Alexander A. Razborov*
Steklov Mathematical Institute
Vavilova 42, 117966, GSP–1, Moscow, RUSSIA

Abstract

In this talk we will consider various classes defined by small depth polynomial size circuits which contain threshold gates and parity gates. We will describe various inclusions between many classes defined in this way and also classes whose definitions rely upon spectral properties of Boolean functions.

1. Introduction

The main goal of the computational complexity theory is to be able to classify computational problems accordingly to their inherent complexity. At the first stage the problems are combined into large collections called complexity classes, each class consisting of problems which can be efficiently solved by an algorithm from a certain family. This allows one to unify many heterogeneous questions into only a few major problems about possible inclusions of one complexity class into another. Unfortunately, we are not even nearly close to solving the most important problems of this kind like the P vs. NP question or the NC vs. P question.

This talk will be devoted to a fragment of the complexity hierarchy (lying well below the class NC^1) where the existing machinery does allow us to answer questions on possible inclusions between complexity classes and in fact the result known at the moment give more or less complete picture of the fine structure within that fragment.

More precisely, we will be mostly interested in small depth circuits which contain threshold gates. There are two reasons for studying them.

*This paper was prepared while the author was visiting Department of Mathematics at MIT partially supported by the Sloan foundation.

to neural nets which is one of the most active areas in computer science. The basic element of a neural net is close to a threshold gate.

Another reason is that the complexity classes defined by small depth threshold circuits contain many interesting Boolean functions and are closely related to other complexity classes defined both in terms of small depth circuits and the spectral behavior of the function in question.

The paper is organized as follows. In Section 2 we introduce the necessary notation. Section 3 is devoted to (linear) threshold circuits of bounded depth. In Section 4 we consider complexity classes defined both in terms of the spectral representation and also in terms of polynomial thresholds. In Section 5 we merge together the two hierarchies of complexity classes considered in previous sections. The concluding section 6 contains some applications of the general theory to computing very concrete Boolean functions.

2. Notation

We will consider Boolean functions but for notational simplicity we will be working over $\{-1,1\}$ rather than $\{0,1\}$ where we let -1 correspond to 1 and 1 to 0. Thus variables will take the values $\{-1,1\}$ and a typical function will be from $\{-1,1\}^n$ to $\{-1,1\}$. In this notation the parity of a set of variables will be equal to their product and thus we will speak of a monomial rather than of the parity of a set of variables. If we have a vector x of variables (indexed like x_i or x_{ij}) then a monomial will be written in the form x^α where α is a $0,1$-vector of the same type.

A *threshold gate* with n inputs is determined by n integer *weights* $(w_1, w_2, \ldots, w_n)$ and a *threshold* T. On an input $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$ it takes the value $\mathrm{sign}(x_1 w_1 + \cdots + x_n w_n + T)$ (we will always assume w.l.o.g. that the linear form $x_1 w_1 + \cdots + x_n w_n + T$ never evaluates to 0). The parameter $\sum_{i=1}^n |w_i| + |T|$ is called the *total weight* of the corresponding threshold gate.

Circuits considered in this paper will be mostly assembled from threshold gates and gates which compute monomials (= parity gates in the $\{0,1\}$-terminology). We define the *size* of a circuit to be the number of gates.

In this section we will consider (linear) *threshold circuits* that is circuits consisting entirely of threshold gates. Let LT_d denote the class of functions computable by polynomial size depth d threshold circuits. Note that it is not quite clear a priori that functions computable even by, say, a single threshold gate have polynomial size circuits (of an arbitrary depth). The following well-known result (see e.g. [**?**]) takes care of this.

Theorem 3.1. *For each threshold gate with n inputs there exists a threshold gate which computes the same function and has the total weight at most* $\exp(O(n \log n))$.

A model which is more natural from the "polynomial" point of view is to have the restriction that absolute values of all (integer!) weights are bounded by a polynomial in the length of the input. We will refer to this restriction as the *small weights* restriction and let $\widehat{LT}_d$ denote the class of Boolean functions computable by polynomial size depth d small weights threshold circuits. It can be easily seen that $\widehat{LT}_d$-circuits can be further simplified to consist of MAJORITY gates and of negations which appear on input variables only.

Now we review lower bounds known for linear threshold circuits.

It is easy to see that LT_1 does not contain all Boolean functions in n variables for any $n \geq 2$. In fact, even such a simple function in just two variables as $x_1 x_2$ is outside of LT_1.

An example of a function in $LT_1 \setminus \widehat{LT}_1$ was first presented in [MK61]:

Theorem 3.2 (Myhill, Kautz). *Any linear threshold gate computing the LT_1-function*

$$\text{sign}\left(\sum_{i=1}^{q+1} 2^{i-1} x_i + \sum_{j=1}^{q} \left(2^q - 2^{j-1}\right) y_j - 2^q\right)$$

must have a coefficient which is at least as large as 2^q.

In fact, Myhill and Kautz gave also an example for which the better bound $\Omega\left(2^n/n\right)$ holds but the proof of this latter result is much harder. The separation between $\widehat{LT}_1$ and LT_1 also follows from more general Theorem 5.4 below.

In depth 2 the first lower bounds were proven in the seminal paper [HMP[+]87]. Namely, they established the following.

depth 2 *small weights threshold circuit computing the function INNER PRODUCT MOD* 2 *(which is defined as* $IP2_n(x_1, \dots, x_n, y_1, \dots, y_n) \rightleftharpoons (x_1 \wedge y_1) \oplus \dots \oplus (x_n \wedge y_n)$ *in* $\{0,1\}-$*notation) must have size* $\exp(\Omega(n))$.

Krause [Kra91] and Krause, Waack [KW91] slightly generalized and extended this result.

Note that $IP2_n \in \widehat{LT}_3$. Hence Theorem 3.3 gives the separation $\widehat{LT}_2 \neq \widehat{LT}_3$ (which, given Theorems 5.1 and 3.7, can be also deduced from Theorem 5.3 below).

No superpolynomial lower bounds are known for $\widehat{LT}_3$-circuits or even for LT_2-circuits. Maass, Schnitger and Sontag [MSS91] proved an $\Omega\left(\frac{\log\log n}{\log\log\log n}\right)$ bound on the size of depth 2 threshold circuits computing an explicitly given Boolean function. The following result was proved in [GHR92]:

Theorem 3.4 (Goldmann, Håstad, Razborov). *Any depth 2 threshold circuit computing INNER PRODUCT MOD* 2 *has size at least* $\Omega(n/\log n)$.

In fact, this is a direct consequence of Theorem 3.3 and the technique used for proving Theorem 3.7 below.

It seems that the only lower bound known for depth three comes from [HG90]. The *generalized inner product* mod 2, $GIP2_{n,s}$ is the Boolean function in ns variables defined (in $\{0,1\}$-notation) as follows:

$$GIP2_{n,s}(x_{ij}) \rightleftharpoons \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{s} x_{ij}.$$

In particular, $IP2_n \equiv GIP2_{n,2}$.

Theorem 3.5 (Håstad, Goldmann). *Any depth* 3 *small weights threshold circuit which computes* $GIP2_{n,s}$ *and has fan-in at most* $(s-1)$ *at the bottom level, must have size* $\exp\left(\Omega\left(\frac{n}{s4^s}\right)\right)$.

This result gives an exponential lower bound but only for circuits with fan-in at most $\epsilon \log n$ at the bottom level. It would be extremely interesting to strengthen Theorem 3.5 because of the following simulation discovered in [Yao90]. Let ACC be the class of functions computable by polynomial size bounded depth circuits over $\{\neg, \wedge, \vee, MOD_{m_1}, \dots, MOD_{m_k}\}$

is divisible by m.

Theorem 3.6 (Yao). *If $f_n \in ACC$ then f_n is also computable by depth 3 small weights threshold circuits of size* $\exp\left((\log n)^{O(1)}\right)$ *and with fan-in at most* $(\log n)^{O(1)}$ *at the bottom level.*

So far superpolynomial lower bounds for ACC-circuits are known only for the bases $\{\neg, \wedge, \vee, MOD_q\}$ where q is a power of a prime (Razborov [Raz87], Smolensky [Smo87], Barrington [Bar86]).

Let's now see how efficiently general threshold circuits can be simulated by threshold circuits with small weights. The results of Chandra, Stockmeyer, Vishkin [CSV84] and Pippenger [Pip87] imply that the function ITERATED ADDITION (that is addition of n n-bit numbers) is computable by constant depth polynomial size *small weights* circuits. A direct consequence of this is that $LT_1 \subseteq \widehat{LT}_d$ for some constant d which was estimated as $d = 13$ in [SB91]. A better construction (based in fact on the spectral technique to be discussed in the next section) was given by Siu and Bruck [SB91]. Namely, they showed that the ITERATED ADDITION is in $\widehat{LT}_3$ which implies $LT_1 \subseteq \widehat{LT}_3$ and, moreover, $LT_d \subseteq \widehat{LT}_{2d+1}$ for any d which in general may depend upon the number of variables. For fixed d this was further improved in [GHR92]:

Theorem 3.7 (Goldmann, Håstad, Razborov). $LT_d \subseteq \widehat{LT}_{d+1}$ *for any fixed* $d > 0$.

This implies that the classes defined by general threshold circuits and by small weights threshold circuits form the following alternating hierarchy:

$$\widehat{LT}_1 \subseteq LT_1 \subseteq \widehat{LT}_2 \subseteq LT_2 \subseteq \widehat{LT}_3 \subseteq \ldots \qquad (1)$$

Let me recall that the inclusion $\widehat{LT}_1 \subseteq LT_1$ is proper by Theorem 3.2 (or by Theorem 5.4), whereas LT_1 and $\widehat{LT}_2$ are trivially separated by the PARITY function. The inclusion $\widehat{LT}_2 \subseteq LT_2$ was shown to be proper by Goldmann, Håstad and Razborov [GHR92] (it is a consequence of Theorem 5.3 below). The question whether LT_2 is different from higher levels of the hierarchy (1) (or whether it contains NP) is open.

olds

Any Boolean function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ can be uniquely represented as a multilinear polynomial over reals:

$$f(x_1, \ldots, x_n) = \sum_{\alpha \in \{0,1\}^n} a_\alpha(f) x^\alpha. \tag{2}$$

This representation is called the *spectral representation* of f and its coefficients $\{a_\alpha(f) \mid \alpha \in \{0,1\}^n\}$ are *spectral coefficients* of f. We define

$$L_1(f) \rightleftharpoons \sum_{\alpha \in \{0,1\}^n} |a_\alpha(f)|$$

and

$$L_\infty(f) \rightleftharpoons \max_{\alpha \in \{0,1\}^n} |a_\alpha(f)|.$$

Similarly we might define the Euclidean norm $L_2(f)$ but it turns out that $L_2(f)$ equals 1 for any f. In fact, it implies that

$$L_1(f) \geq 1 \geq L_\infty(f), \; L_1(f) \cdot L_\infty(f) \geq 1. \tag{3}$$

In general, the spectral approach is a very useful tool in the study of Boolean functions (see e.g. [KKN88, LMN89, BOH90, KM91]). But in this survey we are exclusively interested in its applications to threshold circuits.

Along these lines Bruck and Smolensky [BS92] explicitly defined the class PL_1 which consists of all functions f_n with $L_1(f) \leq n^{O(1)}$ and the class $PL_\infty \rightleftharpoons \left\{ f_n \,\middle|\, L_\infty(f)^{-1} \leq n^{O(1)} \right\}$. Note that by (3), $PL_1 \subseteq PL_\infty$.

The classes which provide a strong link between threshold circuits and spectral properties of Boolean functions were defined by Bruck in [Bru90]. Namely, the class PT_1 consists of all functions f_n which allow a representation of the form

$$f_n(x_1, \ldots, x_n) = \mathrm{sign}\left(\sum_{\alpha \in A} w_\alpha x^\alpha \right) \tag{4}$$

where $A \subseteq \{0,1\}^n$, $|A| \leq n^{O(1)}$. Note that in the $\{0,1\}$-notation PT_1 equals the class of all functions computable by polynomial size depth 2 circuits with a (general) threshold gate at the top and parity gates at similarity with (2).

The class $\widehat{PT}_1$ is defined in the same way, only now we additionally require the weights w_α in (4) to be small ([Bru90]).

Bruck [Bru90] showed a general lower bound for PT_1-circuits which in our notation basically amounts to the following:

Theorem 4.1 (Bruck). $PT_1 \subseteq PL_\infty$.

Bruck and Smolensky [BS92] established the dual result:

Theorem 4.2 (Bruck, Smolensky). $PL_1 \subseteq \widehat{PT}_1$.

So we have the hierarchy

$$PL_1 \subseteq \widehat{PT}_1 \subseteq PT_1 \subseteq PL_\infty. \tag{5}$$

The inclusion $PL_1 \subseteq \widehat{PT}_1$ was shown to be proper in [BS92]:

Theorem 4.3 (Bruck, Smolensky). *The function*

$$EXACT_n(x_1, \ldots, x_n) = 1 \rightleftharpoons \sum_{i=1}^{n} x_i = n/2$$

is in $\widehat{PT}_1 \setminus PL_1$.

$\widehat{PT}_1 \subseteq PT_1$ was shown to be proper by Goldmann, Håstad and Razborov [GHR92] (see more general Theorem 5.4 below). The inclusion $PT_1 \subseteq PL_\infty$ is proper just because the class PL_∞ contains almost all functions; an explicit function separating those two classes was presented in [BS92].

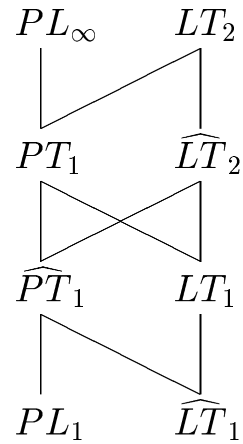
5. The Fine Structure

In this section we combine the two hierarchies (1) and (5) into one powerful picture.

It is clear that $\widehat{LT}_1 \subseteq \widehat{PT}_1$ and $LT_1 \subseteq PT_1$. Less obvious inclusions were established in [Bru90]:

Theorem 5.1 (Bruck). $\widehat{PT}_1 \subseteq \widehat{LT}_2$ *and* $PT_1 \subseteq LT_2$.

At the moment we have the following picture.



It turns out that this picture reflects *all* possible inclusions between the eight classes shown there. Let's review the reasons.

The following result was proved in [Bru90]:

Theorem 5.2 (Bruck). *The Complete Quadratic Function CQ_n which in the $\{0,1\}$-notation is given by $CQ_n(x_1,\ldots,x_n) \rightleftharpoons \oplus_{1\le i<j\le n}(x_i \wedge x_j)$, is in $\widehat{LT}_2 \setminus PL_\infty$.*

The next three easy separations were noticed by Goldmann, Håstad and Razborov in [GHR92].

1. Although we do not know any explicit superlinear lower bounds for LT_2-circuits, we still can claim that $PL_\infty \not\subseteq LT_2$ just because PL_∞ contains almost all functions.

2. If we consider MAJ_n instead of $EXACT_n$ in Theorem 4.3, then it can be improved to give $\widehat{LT}_1 \not\subseteq PL_1$.

3. The PARITY function is in $PL_1 \setminus LT_1$.

The two separations which are still needed to claim that our picture is complete, are $LT_1 \not\subseteq \widehat{PT}_1$ and $PT_1 \not\subseteq \widehat{LT}_2$. In other words, we need lower bounds analogous to those given by Theorem 3.3 but for simpler functions. Such bounds were proven in [GHR92].

Let

$$p_n(x,y) \rightleftharpoons \mathrm{sign}\left(1 + 2\sum_{i=0}^{n-1}\sum_{j=0}^{2n-1} 2^i y_j (x_{i,2j} + x_{i,2j+1})\right)$$

$$U_n(x) \rightleftharpoons \text{sign}\left(1 + 2\sum_{i=0}^{n-1}\sum_{j=0}^{4n-1} 2^i x_{i,j}\right).$$

Obviously, $p_n(x,y)$ and $U_n(x)$ are in PT_1 and LT_1 respectively.

Theorem 5.3 (Goldmann, Håstad, Razborov). *Any depth* 2 *small weights threshold circuit computing* $p_n(x,y)$ *must have size* $\exp(\Omega(n))$. *Hence* $PT_1 \not\subseteq \widehat{LT}_2$.

Theorem 5.4 (Goldmann, Håstad, Razborov). *For any representation*

$$U_n(x) = \text{sign}\left(\sum_{\alpha\in A} w_\alpha x^\alpha\right)$$

of $U_n(x)$ *in the form* (4) *we have* $\sum_{\alpha\in A} |w_\alpha| \geq \exp(\Omega(n))$. *Hence* $LT_1 \not\subseteq \widehat{PT}_1$.

As we noted before, Theorems 5.3 and 5.4 generalize and strengthen many of previous results.

A few words should be said about the method of proof of Theorems 3.3, 5.3, 5.4. Assume that $f_n(x_1,\ldots,x_n,y_1,\ldots,y_n)$ is a Boolean functions with its variables divided into two groups, x-variables and y-variables. Denote by $C_{1/2-\epsilon}(g;1\to 2)$ the *probabilistic one-way communication complexity of* g *with error* $1/2-\epsilon$ *i.e. with advantage* ϵ ([Yao79]). We consider the model in which the probability of being correct is at least $1/2+\epsilon$ for every pair of inputs, the random string is shared by both parties and the complexity is measured as the number of bits sent in the *worst* case (not the average). Let $C(g;1\to 2)$ be the corresponding deterministic measure.

The following lemma which was implicit in [HMP+87] is the key stone to proving Theorems 3.3, 5.3, 5.4:

Lemma 5.5. *Let* $w, d \geq 0$ *and* $f_n(x_1,\ldots,x_n,y_1,\ldots,y_n)$ *be computed by a depth* 2 *threshold circuit with a threshold gate of the total weight* w *at the top and arbitrary gates* g *satisfying* $C(g;1\to 2)\leq d$ *at the bottom. Then*

$$C_{1/2-1/(2w)}(f_n;1\to 2)\leq d.$$

In fact, the paper [HMP+87] dealt with the two-way communication complexity and also Krause [Kra91] and Krause, Waack [KW91] used similar arguments. It is not clear however whether the proof of Theorems 5.3, 5.4 can be carried over in the context of two-way complexity.

In this concluding section we will see that inclusions summarized in our main picture have been extremely useful for designing threshold circuits for very concrete Boolean functions.

We will be interested in such important functions as the ADDITION, MULTIPLICATION, DIVISION, COMPARISON (of two n-bit numbers), POWERING (computing x^n where x is an n-bit number), ITERATED ADDITION, ITERATED MULTIPLICATION, MAXIMUM and SORTING (of n n-bit numbers). In fact, some of these functions allow a naive implementation by constant depth polynomial size small weights circuits and the remaining functions can be implemented so using the reductions of Chandra, Stockmeyer and Vishkin [CSV84] and the results of Beame, Cook and Hoover [BCH86] and of Pippenger [Pip87]. This was observed in [HMP$^+$87] (see also [SB91]). However the depth of resulting circuits is far from optimal. We already noted in Section 3 that the circuits for the ITERATED ADDITION obtained in this way have depth 13; the MULTIPLICATION seems to require depth 10. We will present below more recent results many of which are based on the general theory from previous sections. They lead to much better (and in many cases tight) upper bounds in terms of depth.

Siu and Bruck [SB91] showed that the ADDITION and COMPARISON of two n-bit numbers are both in PL_1 and hence are doable in $\widehat{LT}_2$. A constructive version of Siu and Bruck's result was presented by Alon and Bruck [AB91]. Quite recently Siu and Roychowdhury [SR92] have used Theorem 3.7 to show that even the ITERATED ADDITION is in $\widehat{LT}_2$. All these results are optimal in depth since none of the three functions is in $\widehat{LT}_1$ (this is obvious for the ADDITION and ITERATED ADDITION; for the case of the COMPARISON see [SB91]).

Siu and Bruck [SB91] also showed that the MULTIPLICATION (of two n-bit numbers) is in $\widehat{LT}_4$ (this was later rediscovered in [HHK91] with a better bound on the circuit size). Siu and Roychowdhury [SR92] showed that in fact the MULTIPLICATION is doable in depth 3. This latter result is depth-optimal since [HMP$^+$87] proved before that the MULTIPLICATION is not in $\widehat{LT}_2$.

DIVISION and POWERING were shown to be in $\widehat{LT}_4$ by Siu, Bruck, Kailath and Hofmeister [SBKH91] and in $\widehat{LT}_3$ by Siu and Roychowdhury

[SR92]. To the best of my knowledge, it is open whether they are in $\widehat{LT}_2$ or not.

ITERATED MULTIPLICATION was shown to be in $\widehat{LT}_5$ by Siu, Bruck, Kailath and Hofmeister [SBKH91] and in $\widehat{LT}_4$ by Siu and Roychowdhury [SR92]. It is open whether it is doable in depth 3 or not.

Siu and Bruck [SB91] also considered the MAXIMUM and SORTING of n n-bit numbers and, using their method, placed them into $\widehat{LT}_3$ and $\widehat{LT}_4$ respectively. Siu, Bruck, Kailath and Hofmeister [SBKH91] improved on the second result showing that in fact the SORTING is also in $\widehat{LT}_3$. Whether these functions can be done in depth 2 seems to be open.

We summarize in the following table our knowledge on the depth-optimal constructions for the functions we've been discussing in this section.

Function	*upper bound*	*lower bound*
ADDITION	2 [SB91]	2
ITERATED ADDITION	2 [SR92]	2
MULTIPLICATION	3 [SR92]	3 [HMP+87]
ITERATED MULTIPLICATION	4 [SR92]	3 [HMP+87]
DIVISION	3 [SR92]	2
POWERING	3 [SR92]	2
COMPARISON	2 [SB91]	2 [SB91]
MAXIMUM	3 [SB91]	2
SORTING	3 [SBKH91]	2

7. Acknowledgement

I am highly indebted to Jehoshua Bruck and Michael Goldmann for valuable remarks promptly made on an earlier version of this paper.

References

[AB91] N. Alon and J. Bruck. Explicit constructions of depth-2 majority circuits for comparison and addition. Technical Report RJ 8300 (75661), IBM Research Division, August 1991.

report, University of Massachusetts, 1986.

[BCH86] P. Beame, S. Cook, and H. Hoover. Log depth circuits for division and related problems. *SIAM Journal on Computing*, 15:994–1003, 1986.

[BOH90] Y. Brandman, A. Orlitsky, and J. Hennesy. A spectral lower bound technique for the size of decision trees and two-level AND/OR circuits. *IEEE Transactions on Computers*, 39(2):282–287, February 1990.

[Bru90] J. Bruck. Harmonic analysis of polynomial threshold functions. *SIAM Journal on Discrete Mathematics*, 3(2):168–177, May 1990.

[BS92] J. Bruck and R. Smolensky. Polynomial threshold functions, AC^0 functions and spectral norms. *SIAM Journal on Computing*, 21(1):33–42, February 1992.

[CSV84] A. K. Chandra, L. Stockmeyer, and U. Vishkin. Constant depth reducibility. *SIAM Journal on Computing*, 13:423–439, 1984.

[GHR92] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. In *Proceedings of the 7th Structure in Complexity Theory Annual Conference*, 1992.

[HG90] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. In *Proceedings of the 31st IEEE FOCS*, pages 610–618, 1990.

[HHK91] T. Hofmeister, W. Honberg, and S. Köling. Some notes on threshold circuits and multiplication in depth 4. *Information Processing Letters*, 39:219–225, 1991.

[HKP91] J. Hertz, R. Krogh, and A. Palmer. *An Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.

[HMP+87] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. In *Proceedings of 28th IEEE FOCS*, pages 99–110, 1987.

on Boolean functions. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[KM91] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. In *Proceedings of the 23rd ACM STOC*, pages 455–464, 1991.

[Kra91] M. Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. In 6-*th Structure in Complexity Theory Conference*, pages 314–322, 1991.

[KW91] M. Krause and S. Waack. Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in. In *Proceedings of the 32th IEEE Symposium on Foundations of Computer Science*, pages 777–782, 1991.

[LMN89] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transforms and learnability. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 574–579, 1989.

[MK61] J. Myhill and W.H. Kautz. On the size of weights required for linear-input switching functions. *IRE Trans. on Electronic Computers*, EC10(2):288–290, June 1961.

[MSS91] W. Maass, G. Schnitger, and E. Sontag. On the computational power of sigmoid versus boolean threshold circuits. In *Proceedings of the 32nd IEEE Symposium on Foundations of Computer Science*, pages 767–776, 1991.

[Pip87] N. Pippenger. The complexity of computations by networks. *IBM J. Res. Develop.*, 31:235–243, 1987.

[Raz87] A. Razborov. Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):598–607, 1987. English translation in 41:4, pages 333-338.

with small weights. *SIAM Journal on Discrete Mathematics*, 4(3):423–435, 1991.

[SBKH91] K.-I. Siu, J. Bruck, T. Kailath, and T. Hofmeister. Depth-efficient neural networks for division and related problems. Technical Report RJ 7946, IBM Research, January 1991. To appear in IEEE Trans. Information Theory.

[Smo87] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[SR92] K.-Y. Siu and V. Roychowdhury. On optimal depth threshold circuits for multiplication and related problems. Manuscript, 1992.

[Yao79] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the* 11*th ACM STOC*, pages 209–213, 1979.

[Yao90] A. Yao. On *ACC* and threshold circuits. In *Proceedings of the* 31*th IEEE FOCS*, pages 619–627, 1990.