MAJORITY GATES VS. GENERAL WEIGHTED THRESHOLD GATES

Mikael Goldmann, Johan Håstad
and Alexander Razborov

Abstract. In this paper we study small depth circuits that contain threshold gates (with or without weights) and parity gates. All circuits we consider are of polynomial size. We prove several results which complete the work on characterizing possible inclusions between many classes defined by small depth circuits. These results are the following:

1. A single threshold gate with weights cannot in general be replaced by a polynomial fan-in unweighted threshold gate of parity gates.
2. On the other hand it can be replaced by a depth 2 unweighted threshold circuit of polynomial size. An extension of this construction is used to prove that whatever can be computed by a depth d polynomial size threshold circuit with weights can be computed by a depth $d+1$ polynomial size unweighted threshold circuit, where d is an arbitrary fixed integer.
3. A polynomial fan-in threshold gate (with weights) of parity gates cannot in general be replaced by a depth 2 unweighted threshold circuit of polynomial size.

Key words. circuit complexity; majority circuits; threshold circuits; lower bounds.
Subject classifications. 68Q15.

1. Introduction

In this paper we study small depth circuits that contain threshold gates. We will be working in the discrete model of computation i.e., all variables and values of intermediate results will take Boolean values and, in particular, we will not deal with real numbers. A threshold gate with m inputs is determined by m weights $(w_1, w_2 \dots w_m)$ and a threshold T. On inputs $y_1, y_2 \dots y_m$ it takes value 1 if $\sum_{i=1}^{m} w_i y_i \geq T$ and 0 otherwise. It is easy to see that a Boolean function can be computed by a threshold gate with integer coefficients (that is with integer weights and threshold) if and only if it can be computed by a threshold gate with arbitrary real coefficients. So in what follows we will consider only threshold gates with integer coefficients.

A model of computation, more realistic than general threshold gates (at least from a physical point of view) is to have the restriction that the absolute values of the (integer) weights are bounded by a polynomial in the length of the input. If we do not care whether the number of wires is increased by a polynomial factor, this model is equivalent to eliminating the weights totally. In this case a gate will output 1 iff the number of inputs that take the value 1 exceeds T, for a given threshold T. It is not hard to see that in this case we only need majority gates, and thus we get very simple circuits. However, for notational convenience, most of the time we will not eliminate the weights and thus we will denote this type of circuits small weight threshold circuits.

In this paper all circuits we consider will be of polynomial size. This will sometimes influence our language. In particular we will say "Depth 2 small weight threshold circuits cannot compute the inner product" when we actually mean "Polynomial size, depth 2 threshold circuits where all weights are bounded by a polynomial cannot compute the inner product".

There are two reasons for studying threshold circuits. The first reason is that threshold circuits are very closely connected 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. However, in many instances one prefers to have a continuous model where variables and intermediate results take real values. This change does not increase the computational power significantly as proved by Maass, Schnitger and Sontag [12]. Another, maybe more fundamental, difference is that neural nets frequently contain feedback edges i.e., the underlying graph is not acyclic. On the other hand a strong point of similarity is the restriction to small depth. Many neural nets considered have depth 2 or 3. The computational capacity of neural nets is far from clear, and thus any information about their power would be very interesting. Even if we do not consider exactly the same model, we think that our results will be useful to this end. For more information about neural nets see [8].

The second reason is that small depth threshold circuits is one of the simplest natural classes of circuits for which no superpolynomial lower bounds are known for an explicit function. Thus in particular it is not known whether all functions in NP can be computed by depth 3 threshold circuits without weights (or by depth 2 threshold circuits with weights). All lower bounds known so far are for very limited classes. In particular there are good lower bounds for depth 2 circuits with small weights by Hajnal *et al.* [6] and more recently by Krause [10] and Krause, Waack [11].

The techniques of Hajnal *et al.* were extended by Håstad and Goldmann [7] to deal with depth 3 circuits with small weights and small bottom fanin. These lower bounds agreed very well with our intuition and also the results about monotone threshold circuits by Yao [21] (extended in [7]) contained no surprises.

The first surprise was presented by Allender [1] who, inspired by the results of Toda [18], proved that depth 3 threshold circuits of subexponential size could do all of AC^0. Yao [22] extended this to ACC^0 which consists of all functions computable by polynomial size constant depth circuits over the basis $\{\wedge, \vee, \neg, \bmod m\}$ for an arbitrary fixed m (note that this last class could actually contain all of NP!). Siu and Bruck [16] showed that even as simple circuits as an unweighted threshold of parity gates could do much more than expected. Our results follow the same pattern, by establishing some not very surprising lower bounds and proving that the lower bounds are in fact optimal by showing a surprising construction.

It is not difficult to see that there are functions computed by a general threshold gate that cannot be computed by a threshold with small (or no) weights [14]. Siu and Bruck in [16] show that such functions are doable by depth 3 unweighted threshold circuits, and in general a depth d weighted threshold circuit can be simulated by a depth $2d+1$ unweighted threshold circuit. They conjecture that depth $d+2$ unweighted threshold circuits would be sufficient and pose the question whether depth 3 unweighted threshold circuits are required to simulate a single threshold gate with arbitrary weights. What we first thought was a stepping stone to answering this question in the affirmative, is our result that there is an explicit threshold gate which cannot be replaced by an unweighted threshold of parity gates. This is a weaker statement since it is easy to see that any function that is an unweighted threshold of parity gates is also an unweighted threshold of unweighted thresholds [2].

Using a similar proof we prove that there is a function which can be written as a general threshold gate of parity gates, but which cannot be computed by depth 2 majority circuits. Both these facts seem to speak in favor to the affirmative solution of the question of Bruck and Siu, but the intuition is wrong and we prove that the answer is negative. In fact anything that can be computed by an arbitrary threshold gate can be computed in polynomial size and depth 2 with unweighted threshold gates. This construction extends to prove that whatever can be computed by depth d general threshold circuits can be computed in depth $d+1$ with unweighted threshold gates. This in particular proves the above mentioned conjecture of Siu and Bruck.

An outline of the paper is as follows. Sections 2 and 3 introduce the notation and recall some basic facts. In Section 3 we also establish a connection between threshold circuits and communication complexity used in later sections.

Sections 4 through 7 contain the results of this article, and they are fairly independent of each other. In Section 4 we prove the lower bounds. The basic tool for proving these lower bounds is a communication analogue of the *correlation lemma* from [6] which says that if $f(x, y)$ is a small threshold of "simple" functions then there is an efficient probabilistic one-way communication protocol which predicts the value $f(x, y)$ with considerable advantage. In Section 5 we prove a converse of the correlation lemma. In Section 6 we establish an upper bound on the one-way communication complexity of an arbitrary function computable by a single threshold gate.

In Section 7 we show how to simulate large weights by small weights without losing much in depth. The bulk of this section is devoted to showing how a certain function, $U_{n,m}(x)$, which is universal for the class of all functions computable by a single threshold gate with weights, can be computed by a depth 2 polynomial size unweighted threshold circuit. This result is then extended to show that a polynomial size depth d threshold circuit with arbitrary weights can be simulated by a polynomial size depth $d+1$ unweighted threshold circuit.

In Section 8 we sum up the relations between the considered complexity classes. In the final section we mention some recent results related to this work.

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 monomials rather than 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 x^α where α is a $0,1$-vector of the same type. Observe that using $\{-1, 1\}$ rather than $\{0, 1\}$ does not change anything when considering threshold gates. This allows us to write the function computed by a threshold gate as the sign of a linear form. Since we are dealing with functions that take the values $\{-1, 1\}$, we require the argument to the sign function to be nonzero.

In our circuits we allow no multiple edges and we define the *size* to be the number of gates. We will be interested in the following classes:

- LT_d The set of functions computable by circuits consisting of general threshold gates which have polynomial size and depth d.
- $\widehat{LT}_d$ The set of functions computable by circuits consisting of small weight threshold gates which have polynomial size and depth d.
- PT_1 [2] The set of functions that can be represented as the sign of a sparse polynomial. This corresponds to a general threshold gate with monomials at the inputs.
- $\widehat{PT}_1$[2] The set of functions that can be represented as the sign of a sparse polynomial with integer coefficients that have absolute value bounded by a polynomial. This corresponds to a small weight threshold gate with monomials at the inputs.
- PL_1 [3] The set of functions whose Fourier transform has L_1-norm bounded from above by a polynomial.
- PL_∞ [3] The set of functions whose Fourier transform has L_∞-norm that is at least an inverse polynomial.

We will not discuss relations between these classes here but defer this to the discussion at the end of the paper.

3. Preliminaries

We will frequently use the following well known fact:

LEMMA 1. *Any threshold gate with n inputs can be represented as a threshold gate with integer weights w_i and threshold T such that $|w_i|, |T| \leq 2^{-n}(n+1)^{(n+1)/2}$.*

The proof can be found e.g., in [13].

We will be interested in functions that can be written as a small threshold of members of some set of functions. Let $\mathcal{H}$ be a class of functions $h : \{-1,1\}^n \longrightarrow \{-1,1\}$ e.g., the class of monomials, the class of all functions representable by a threshold gate of weight $\leq b$ etc. Assume for the sake of simplicity that $\mathcal{H}$ is closed under negation.

Definition 2. *Let $T_{\mathcal{H}}(f)$ be the minimal possible w for which f allows a representation*

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{s} a_i h_i(x)\right)$$

where $h_i \in \mathcal{H}$ and $w = \sum_{i=1}^{s} |a_i|$, $a_i \in \mathbf{Z}$.

The parameter w will be called the *total weight* of the corresponding threshold gate.

Let the *correlation* $D^R_{\mathcal{H}}(f)$ of the family $\mathcal{H}$ with the function f with respect to a distribution R on $\{-1,1\}^n$ be the value

$$\max_{h \in \mathcal{H}} \mathbf{E}_R\left[f(x)h(x)\right].$$

This leads to the following definition:

Definition 3. *The correlation of f with respect to a family of functions $\mathcal{H}$ is $D_{\mathcal{H}}(f) = \min_R D^R_{\mathcal{H}}(f)$ where the minimum is taken over all distributions R.*

Lemma 3.3 from [6] can now be stated as follows:

Lemma 4. *$T_{\mathcal{H}}(f) \geq D^{-1}_{\mathcal{H}}(f)$.*

For completeness let us give its proof.

Proof. We just have to prove that $T_{\mathcal{H}}(f) \cdot D^R_{\mathcal{H}}(f) \geq 1$ for any distribution R. Let f be any Boolean function with the representation

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{s} a_i h_i(x)\right),$$

where this representation has minimum total weight, i.e.,

$$T_{\mathcal{H}}(f) = \sum_{i=1}^{s} |a_i|.$$

By the definition of $D^R_{\mathcal{H}}(f)$ we now just need to show that for some $h \in \mathcal{H}$ we have

$$\sum_{i=1}^{s} |a_i| \cdot \mathbf{E}_R\left[f(x)h(x)\right] \geq 1.$$

We will show that in fact such h can be chosen from the set $\{\pm h_i | 1 \leq \ i \leq s\}$ (which is a subset of $\mathcal{H}$ since $\mathcal{H}$ is closed under negation). This readily follows from

$$1 = \mathbf{E}_R\left[f(x)f(x)\right] \leq \mathbf{E}_R\left[f(x)\sum_{i=1}^{s} a_i h_i(x)\right] \leq$$
$$\sum_{i=1}^{s} |a_i| \left|\mathbf{E}_R\left[f(x)h_i(x)\right]\right| \leq \sum_{i=1}^{s} |a_i| \cdot \max_{1\leq i\leq s} \left|\mathbf{E}_R\left[f(x)h_i(x)\right]\right|$$

which completes the proof of the lemma. □

It will be convenient for us to use an analogue of this lemma stated in terms of communication complexity. 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* (see [19, 20]). For the purposes of this paper 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. We have the following lemma:

Lemma 5. *Let $d = \max_{h\in\mathcal{H}} C(h; 1 \to 2)$. Then*

$$C_{1/2-1/(2T_{\mathcal{H}}(f))}(f; 1 \to 2) \leq d.$$

In other words, there exists a one-way probabilistic protocol of complexity $\leq d$ which guarantees advantage at least $(2T_{\mathcal{H}}(f))^{-1}$ for every pair of inputs.

Proof. Let

$$f(x, y) = \text{sign}\left(\sum_{i=1}^{s} a_i h_i(x, y)\right) \tag{1}$$

where $\sum_{i=1}^{s} |a_i| = w = T_{\mathcal{H}}(f)$. The players use the common random string to choose h_i and then they compute and answer $\text{sign}(a_i)h_i(x, y)$. They choose h_i with probability $\frac{|a_i|}{w}$. The communication complexity is clearly bounded by d. To take care of the advantage, note that the output is correct iff $f(x, y) = \text{sign}(a_i)h_i(x, y)$ or, in other words, $f(x, y)\text{sign}(a_i)h_i(x, y) = 1$. Hence for each particular input (x, y) the advantage equals

$$\frac{1}{2}\mathbf{E}\left[f(x, y)\text{sign}(a_i)h_i(x, y)\right] = \frac{1}{2}\sum_{i=1}^{s} \frac{|a_i|}{w} \cdot f(x, y)\text{sign}(a_i)h_i(x, y) =$$
$$\frac{f(x, y)}{2w}\sum_{i=1}^{s} a_i h_i(x, y) = \frac{1}{2w}\left|\sum_{i=1}^{s} a_i h_i(x, y)\right| \geq \frac{1}{2w},$$

where the last equality follows from (1). □

Several previous proofs of lower bounds are implicitly based on Lemma 5 or similar statements. [6, 11] use the lower bound for the communication complexity of "INNER PRODUCT MOD 2" (this bound holds even for the two-way case), while [7] uses a straightforward generalization to a multi-party communication game.

Let us just remark here that if $\mathcal{H}$ is the set of all monomials then $d = 1$ and if $\mathcal{H}$ is the set of all threshold gates with total weight bounded by S then $d \leq \lceil \log(2S+1) \rceil$.

4. Lower bounds

In this section we will prove that there is a function which can be computed by a threshold gate with large weights with variables as inputs but not by a threshold gate with small weights and monomials as inputs. We will also prove that there is a function which can be computed by a threshold gate with large weights with monomials as inputs but not by a depth 2 threshold circuit with small weights. Using the notation from the introduction this will show that $LT_1 \not\subseteq \widehat{PT}_1$ and $PT_1 \not\subseteq \widehat{LT}_2$ respectively.

The proofs go as follows. First we define a function $p(x,y)$ in PT_1 wich will be shown to be "hard". More precisely, Theorem 6 establishes a tradeoff between the advantage, ϵ, achieved, and the number of bits, d, sent by a randomized one-way communication protocol for $p(x,y)$. On the other hand, Lemma 5, when applied to depth 2 small weight threshold circuits, shows that any function $f(x,y)$ in $\widehat{LT}_2$ has a randomized one-way protocol that uses little communication but still computes $f(x,y)$ correctly with considerable advantage. Combining this with Theorem 6 shows that $p(x,y) \notin \widehat{LT}_2$.

To show that $LT_1 \not\subseteq \widehat{PT}_1$ we define a function $U_{n,m}(x)$ in LT_1 such that if $U_{n,m}(x) \in \widehat{PT}_1$ then $p(x,y) \in \widehat{PT}_1$. Since $\widehat{PT}_1 \subseteq \widehat{LT}_2$ we have $p(x,y) \notin \widehat{PT}_1$ and thus $U_{n,m}(x) \in LT_1 \setminus \widehat{PT}_1$.

The function $p(x,y)$ is defined as follows:

$$p(x,y) = \mathrm{sign}(2P(x,y)+1)$$

where

$$P(x,y) = \sum_{i=0}^{n-1} \sum_{j=0}^{2n-1} 2^i y_j (x_{i,2j} + x_{i,2j+1}).$$

We will now show the following theorem:

THEOREM 6. *For any $\epsilon > 0$ possibly depending on n we have*

$$2^d \geq \Omega\left(\frac{\epsilon \cdot 2^{n/2}}{\sqrt{n}}\right)$$

where $d = C_{1/2-\epsilon}(p; 1 \to 2)$.

Before we prove the theorem, let us apply it to circuits. First we show that $PT_1 \not\subseteq \widehat{LT}_2$ by establishing that $p(x,y) \in PT_1 \setminus \widehat{LT}_2$. Clearly $p(x,y)$ is in PT_1. The following corollary to Theorem 6 shows that $p(x,y) \notin \widehat{LT}_2$:

COROLLARY 7. *If $p(x,y)$ is computed by a depth 2 threshold circuit with weights bounded by w and size s then*

$$sw^2 \geq \Omega\left(\frac{2^{n/2}}{n^{5/2}}\right).$$

PROOF. The gates at the bottom level all have one-way complexity at most $\lceil \log(4wn^2+1) \rceil$ since player 1 just sends the weight contributed by his inputs. This weight is an integer in the range $[-2wn^2, 2wn^2]$. Now Lemma 5 gives us

$$C_{1/2-1/(2sw)}(p; 1 \to 2) \leq \lceil \log(4wn^2+1) \rceil.$$

The corollary now follows from Theorem 6. $\square$

Since $\widehat{PT}_1 \subseteq \widehat{LT}_2$ [2], we clearly have that $p(x,y) \notin \widehat{PT}_1$. Using the fact that the one-way complexity of a monomial is constant, an argument analogous to the proof of Corollary 7 shows the following:

COROLLARY 8. *If $p(x,y)$ is computed by a threshold gate of monomials then the total weight w of this gate satisfies*

$$w \geq \Omega\left(\frac{2^{n/2}}{\sqrt{n}}\right).$$

Next we show that $LT_1 \not\subseteq \widehat{PT}_1$.

Let $r_{n,m}(x) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} 2^i x_{ij}$ and let $U_{n,m}(x) = \text{sign}(2r_{n,m}(x)+1)$.

Clearly, $U_{n,m}(x) \in LT_1$. Assume that $U_{n,4n}(x) \in \widehat{PT}_1$. Let C be the $\widehat{PT}_1$–circuit computing $U_{n,4n}(x)$. Make the following variable substitution:

$$x_{i,2k} \leftarrow x_{i,2k} y_k, \quad x_{i,2k+1} \leftarrow x_{i,2k+1} y_k.$$

We then get a circuit C' with the following properties: C' has the same weights and the same number of monomials as C. Thus C' computes a function in $\widehat{PT}_1$. On the other hand it is easy to see that the substitution transforms $U_{n,4n}(x)$ into $p(x,y)$, and since $p(x,y) \notin \widehat{PT}_1$ we have a contradiction. This argument actually shows that the bound established in Corollary 8 holds for $U_{n,4n}(x)$ as well.

COROLLARY 9. *If $U_{n,4n}(x)$ is computed by a threshold gate of monomials then the total weight w of this gate satisfies*

$$w \geq \Omega\left(\frac{2^{n/2}}{\sqrt{n}}\right).$$

In general, Theorem 6 shows that $p(x,y)$ cannot be written as a small depth 2 circuit with a bounded weight threshold gate at the top and simple gates at the inputs, where simple means "having small one-way communication complexity" (i.e., majority, mod m for constant m etc.).

In the remainder of this section we will prove the theorem.

PROOF OF THEOREM 6. Take a probabilistic protocol for p. If we take some distribution R on inputs, then by a standard argument there is a deterministic protocol where player 1 sends d bits which is ϵ–biased with respect to p on the distribution R. That is to say

$$|\mathbf{E}_R\left[p(x,y)k(x,y)\right]| \geq 2\epsilon, \tag{2}$$

where $k(x,y)$ is the output of the protocol. Let us define the distribution on inputs that will allow us to derive the lower bound on d. Let $B(M)$ be the distribution that is obtained as the sum of $2M$ Bernoulli variables, where each variable takes the value $1/2$ and $-1/2$, each with probability $1/2$. Let $A_j = \frac{1}{2}\sum_{i=0}^{n-1} 2^i(x_{i,2j} + x_{i,2j+1})$. It is easy to see that A_j can take any integer value in $[-2^n+1, 2^n-1]$.

Let R_x be a distribution on x that makes the A_j independent and $B\left(2^n-1\right)$–distributed. Let U be the uniform distribution on y.

We choose a pair (x,y) by picking y according to U and x according to R_x under the condition that $|P(x,y)| = 2$. We call this distribution R.

Now let us look at a protocol $k(x,y)$ that is ϵ–biased with respect to R. Player 1, who has x, sends a d–bit message, $m = m(x)$, to player 2 after which player 2 gives the output of the protocol. What player 2 says can only depend on m and y. We can write the output as $k_m(y)$, that is to say $k(x,y) = k_{m(x)}(y)$.

By assumption we have that (2) holds for $k(x,y)$. We will now give the following upper bound for the left hand side of (2):

$$|\mathbf{E}_R\left[p(x,y)k(x,y)\right]| \leq O\left(\frac{2^d\sqrt{n}}{2^{n/2}}\right). \tag{3}$$

This along with (2) will give us the statement of the theorem.

Let $q(x,y)$ be the following function:

$$q(x,y) = \begin{cases} P(x,y)/2 & \text{if } |P(x,y)| = 2 \\ 0 & \text{otherwise.} \end{cases}$$

This means that if (x,y) is chosen according to $R_x \times U$ then $q(x,y) = p(x,y)$ on the domain of R, and $q(x,y) = 0$ otherwise.

Now, $\Pr_{R_x,U}\left[|P(x,y)| = 2\right] \geq \Omega\left(\frac{1}{\sqrt{n}2^{n/2}}\right)$. This is true since for any fixed y, if we take x according to R_x then $P(x,y)/2$ is $B(2n(2^n-1))$–distrubuted. Hence

$$|\mathbf{E}_R\left[p(x,y)k(x,y)\right]| \leq |\mathbf{E}_{R_x,U}\left[q(x,y)k(x,y)\right]| \cdot O\left(\sqrt{n}2^{n/2}\right).$$

It therefore suffices to show that

$$|\mathbf{E}_{R_x,U}\left[q(x,y)k(x,y)\right]| \leq O\left(2^{d-n}\right). \tag{4}$$

It is useful to make the following observation: There are 2^d possible messages that player 1 might send. We can enumerate them as $m_1, \ldots, m_{2^d}$. So for every x there is an l such that $\mathbf{E}_U\left[q(x,y)k(x,y)\right] = \mathbf{E}_U\left[q(x,y)k_{m_l}(y)\right]$. This gives us

$$|\mathbf{E}_U\left[q(x,y)k(x,y)\right]| \leq \sum_{l=1}^{2^d} |\mathbf{E}_U\left[q(x,y)k_{m_l}(y)\right]| \tag{5}$$

for any fixed x. Let us now show that (4) holds. First we use (5) to get

$$|\mathbf{E}_{R_x,U}\left[q(x,y)k(x,y)\right]| \leq \mathbf{E}_{R_x}\left[|\mathbf{E}_U\left[q(x,y)k(x,y)\right]|\right] \leq \\ \sum_l \mathbf{E}_{R_x}\left[|\mathbf{E}_U\left[q(x,y)k_{m_l}(y)\right]|\right].$$

Then we use Cauchy-Schwartz inequality and simple manipulation.

$$\sum_l \mathbf{E}_{R_x}\left[|\mathbf{E}_U\left[q(x,y)k_{m_l}(y)\right]|\right] \leq \sum_l \mathbf{E}_{R_x}\left[\mathbf{E}_U\left[q(x,y)k_{m_l}(y)\right]^2\right]^{1/2} = \\ \sum_l \mathbf{E}_{U,U}\left[k_{m_l}(y)k_{m_l}(y')\mathbf{E}_{R_x}\left[q(x,y)q(x,y')\right]\right]^{1/2} \leq \\ 2^d\mathbf{E}_{U,U}\left[|\mathbf{E}_{R_x}\left[q(x,y)q(x,y')\right]|\right]^{1/2} \leq \\ 2^d\left(2^{1-2n} + |\mathbf{E}_{R_x}\left[q(x,y)q(x,y') \mid y \neq \pm y'\right]|\right)^{1/2}.$$

Thus in order to complete the proof it is sufficient to show that for all y and y' such that $y \neq \pm y'$ we have

$$|\mathbf{E}_{R_x}[q(x,y)q(x,y')]| \leq O\left(2^{-2n}\right). \tag{6}$$

Let $W = \sum_{\{j \mid y_j = y'_j\}} A_j y_j$ and $Z = \sum_{\{j \mid y_j = -y'_j\}} A_j y_j$.

Then W is $B(k(2^n-1))$–distributed and Z is $B((2n-k)(2^n-1))$–distributed for some k where $0 < k < 2n$. Moreover, W and Z are independent. We have $P(x,y) = 2(W+Z)$ and $P(x,y') = 2(W-Z)$. This gives us the following:

$$\begin{aligned}
&|\mathbf{E}_{R_x}[q(x,y)q(x,y')]| = \\
&|\mathbf{P}[P(x,y)P(x,y') = 4] - \mathbf{P}[P(x,y)P(x,y') = -4]| = \\
&\left|\mathbf{P}\left[W^2 - Z^2 = 1\right] - \mathbf{P}\left[W^2 - Z^2 = -1\right]\right| = \\
&|\mathbf{P}[|W| = 1]\,\mathbf{P}[Z = 0] - \mathbf{P}[W = 0]\,\mathbf{P}[|Z| = 1]| \leq O\left(2^{-2n}\right).
\end{aligned}$$

The last inequality follows from the fact that

$$|\mathbf{P}[|W| = 1] - 2\cdot\mathbf{P}[W = 0]| \leq O(2^{-3n/2}),$$

$$|\mathbf{P}[|Z| = 1] - 2\cdot\mathbf{P}[Z = 0]| \leq O(2^{-3n/2})$$

and that

$$\mathbf{P}[Z = 0] \leq O(2^{-n/2}),$$
$$\mathbf{P}[W = 0] \leq O(2^{-n/2}).$$

We have proved (6) and thereby (4) and (3). This completes the proof of the theorem. □

5. Sufficiency of the correlation lemma

We saw in the last section that Lemma 4 and its communication analogue Lemma 5 are very useful. In this section we will explain this by proving that Lemma 4 can be partially reversed. Namely, the condition that for every distribution on inputs there is a function in $\mathcal{H}$ which is polynomially correlated with a function f implies that f can be written as a small threshold of functions in $\mathcal{H}$.

This result implicitly follows from a general statement proved by Freund [4, Theorem 1]. Since our proof is simpler we include it here. We have:

THEOREM 10. $T_{\mathcal{H}}(f) \leq 2nD_{\mathcal{H}}^{-2}(f)$.

PROOF. Consider a two person game where player 1 chooses an input x and player 2 chooses a function h which belongs to $\mathcal{H}$. The result of the game is that player 2 wins $h(x)f(x)$ from player 1. By definition $D_{\mathcal{H}}^{R}(f)$ is the expected gain of player 2 when player 1 plays with the mixed strategy defined by R and player 2 plays optimally and knows player 1's strategy. Thus for any mixed strategy of player 1, player 2 can always win $D_{\mathcal{H}}(f)$ on the average. By the minmax theorem for zero-sum two person games [15], there is a mixed strategy for player 2 which guarantees him this gain. In our case this means that there is a probability distribution E on elements of $\mathcal{H}$ such that for any x

$$\mathbf{E}_E\left[f(x)h(x)\right] \geq D_{\mathcal{H}}(f). \tag{7}$$

Now let $r = 2nD_{\mathcal{H}}^{-2}(f)$. Consider r independent copies $h_1, \ldots, h_r$ of the distribution E and denote their sum by H. By Chernoff's bound we have from (7):

$$\mathbf{P}\left[f(x)H(x) \leq 0\right] < \exp\left(-rD_{\mathcal{H}}^2(f)/2\right) < 2^{-n}.$$

Hence for at least one possible tuple $h_1, \ldots, h_r$ we have $H(x)f(x) > 0$ for all x which means $f(x) = \mathrm{sign}\,(H(x))$. This completes the proof of Theorem 10. □

Unfortunately, we can not hope to reverse Lemma 5 in a similar way. The reason is that a probabilistic communication protocol might "use" arbitrary functions h for which $C(h; 1 \to 2) \leq d$, not only those from $\mathcal{H}$. The conversion becomes possible only if $\mathcal{H}$ consists of *all* such functions, but this class is not particularly interesting for applications.

6. An upper bound on the communication complexity of threshold functions

In the next section we will show the surprising result $LT_1 \subseteq \widehat{LT}_2$. By Lemma 5 this implies that for any $f_n \in LT_1$ there exists $k > 0$ such that $C_{1/2-n^{-k}}(f_n; 1 \to 2) \leq O(\log n)$. However we can prove much better upper bound using the spectral norm technique from [3]. This easy result, interesting in its own right, explains at least why the "expected" separation $LT_1 \not\subseteq \widehat{LT}_2$ cannot be proved via Lemma 5 and serves as a prelude to the next section.

For a Boolean function $f : \{-1,1\}^n \longrightarrow \{-1,1\}$ its L_1-*norm* is defined by $L_1(f) = \sum_{\alpha \in \{0,1\}^n} |a_\alpha|$ where $f(x) = \sum_{\alpha \in \{0,1\}^n} a_\alpha x^\alpha$ is the uniquely determined polynomial representation of f. We have the following general result:

Theorem 11. *For each function f, we have that*

$$C_{1/2-(2L_1(f))^{-1}}(f;1\to 2)\le 1.$$

Proof. The following fact was used in [3, proof of Lemma 1]:

Lemma 12. ([3]) *For any $f(x)$ there exists a probability distribution α over $\{0,1\}^n$ such that for each $x\in\{-1,1\}^n$,*

$$\mathbf{E}\left[\mathrm{sign}\,(a_\alpha)\,x^\alpha\right]=\frac{f(x)}{L_1(f)}. \tag{8}$$

Now, the two players pick α in accordance with this distribution and output $\mathrm{sign}\,(a_\alpha)\,x^\alpha$ using the fact that the one-way communication complexity of a monomial is ≤ 1. By (8), the advantage achieved by this protocol at each input x is equal to $(2L_1(f))^{-1}$. □

Let $g_n:[2^n]\times[2^n]\longrightarrow\{-1,1\}$ be the *ordering* function defined by $g_n(x,y)=1$ iff $x\le y$. Siu and Bruck [16] showed that $L_1(g_n)\le O(n)$. Along with Theorem 11 and obvious amplification this gives the following result:

Theorem 13. $C_{1/2-1/n}(g_n;1\to 2)\le O(1)$.

This should be compared with the result of Yao [20] saying that $C_{1/2-\epsilon}(g_n;1\to 2)\ge\Omega(n)$ for each *fixed* $\epsilon>0$.

Theorem 13 is easily extended to arbitrary functions from LT_1:

Theorem 14. *For each $f_n(x,y)\in LT_1$,*

$$C_{1/2-(n\log_2 n)^{-1}}(f_n;1\to 2)\le O(1).$$

Proof. By Lemma 1, $f_n(x,y)$ can be represented in the form

$$f_n(x,y)=\mathrm{sign}\left(\sum_{i=1}^n a_ix_i\ +\ \sum_{j=1}^n b_jy_j\ +\ c\right) \tag{9}$$

where $|a_i|,|b_j|\le\exp(O(n\log n))$. Now the two parties wishing to compute $f_n(x,y)$ proceed as follows. They compute $-\sum_{i=1}^n a_ix_i$ and $c+\sum_{j=1}^n b_jy_j$ respectively and apply the protocol for the ordering function from Theorem 13 to compute $g_n\left(-\sum_{i=1}^n a_ix_i,\ c+\sum_{j=1}^n b_jy_j\right)$. This gives $C_{1/2-(Cn\log_2 n)^{-1}}(f_n;1\to 2)\le O(1)$ for some $C>0$; again a straightforward amplification allows us to get rid of the constant C. □

REMARK 15. *The proof of Theorem 14 reveals that the ordering function is universal for LT_1 in the sense of communication complexity.*

7. Replacing large weights by small weights

In this section we will show how to simulate threshold circuits which have large weights by threshold circuits with small weights. Let us start with the basic construction. It will be convenient for us to slightly change the definition of the sign function in this section by putting $\mathrm{sign}(0) = 1$.

Recall the function $U_{n,m}$ defined in Section 4. As observed in that section, $U_{n,m}$ can be computed by a threshold gate. On the other hand, it is universal in a very strong sense. Namely take any function that is computed by a threshold gate of n inputs. Then Lemma 1 implies that this function is a subfunction of $U_{\frac{1}{2}(n+1)\log(n+1),2(n+1)}$. Thus to achieve our goals it will be sufficient to compute $U_{n,m}$ by a depth 2 polynomial size threshold circuit with small weights.

In what follows $s > 0$ will be considered as a fixed integer and l will be a parameter. To avoid confusion with other usage of n and m we will for the next couple of paragraphs discuss how to compute $U_{a,b}$ for some parameters a and b which later will be chosen as functions of n and m. For notational simplicity we will also assume that a and b are powers of 2.

One of the building blocks will be the following function $M_l(y)$ of one integer variable:

$$\begin{aligned} M_l(y) \quad = \quad \sum_{i=-2b}^{2b} & \mathrm{sign}\left(y - i \cdot 2^{l+s\log a} - 2^l + a^{-s}2^l\right) \\ & -\mathrm{sign}\left(y - i \cdot 2^{l+s\log a} - 2^{l+1} - a^{-s}2^l\right) \\ & +\mathrm{sign}\left(y - i \cdot 2^{l+s\log a} + 2^l - a^{-s}2^l\right) \\ & -\mathrm{sign}\left(y - i \cdot 2^{l+s\log a} + 2^{l+1} + a^{-s}2^l\right) \end{aligned}$$

where the summation extends over all four terms. Let us establish some properties of M_l.

Let $y = j \cdot 2^{l+s\log a} + \delta$ where j is an integer and $|\delta| \leq 2^{l+s\log a-1}$. Observe that all $i \neq j$ contribute 0 in the sum defining $M_l(y)$ since the terms cancel in pairs. It is straightforward to obtain:

LEMMA 16. *For $|y| < 2b2^{l+s\log a} + 2^{l+1} + a^{-s}2^l$ we have the following:*

If $2^{l+1} + a^{-s}2^l > |\delta| > 2^l - a^{-s}2^l$ *then* $M_l(y) = 2\text{sign}(\delta)$.
If $|\delta| > 2^{l+1} + a^{-s}2^l$ *or* $|\delta| < 2^l - a^{-s}2^l$ *then* $M_l(y) = 0$.
If $|\delta| = 2^{l+1} + a^{-s}2^l$ *or* $|\delta| = 2^l - a^{-s}2^l$ *then* $|M_l(y)| \leq 2$.

We use the following shorthand: Let $\Sigma_{t_1}^{t_2}(x) = \sum_{i=t_1}^{t_2} \sum_{j=0}^{b-1} 2^i x_{ij}$ and let $\Sigma_{ls}(x) = \Sigma_{\max(l-s\log a-\log b,\ 0)}^{\min(l+s\log a,\ a-1)}(x)$ $(0 \leq l \leq a + \log b)$. Now consider $N_l(x) = M_l(\Sigma_{ls}(x))$. The total weight of each threshold gate in the definition of $N_l(x)$ is $\leq O(a^{2s}b^2)$ since we can cancel the common factor $2^{l-s\log a-\log b}$. Observe that this is polynomial in a and b. Observe also that $|\Sigma_{ls}(x)| < 2b2^{l+s\log a}$ and hence Lemma 16 applies. Furthermore

$$\Sigma_{ls}(x) \equiv r_{a,b}(x) - \Sigma_0^{l-s\log a-\log b-1}(x) \mod 2^{l+s\log a}.$$

Clearly $|\Sigma_0^{l-s\log a-\log b-1}(x)| < a^{-s}2^l$. Using this bound a straightforward application of Lemma 16 yields

LEMMA 17. *If* $2^l \leq |r_{a,b}(x)| < 2^{l+1}$ *then* $N_l(x) = 2U_{a,b}(x)$

and

LEMMA 18. *Let* $q_l = |r_{a,b}(x) \bmod 2^{l+s\log a}|$. *If* $q_l \geq (2 + \frac{2}{a^s})2^l$ *or* $q_l \leq (1 - \frac{2}{a^s})2^l$ *then* $N_l(x) = 0$.

REMARK 19. *Note that Lemmas 17 and 18 are still valid for x ranging over $\{-1,0,1\}^{ab}$ rather than $\{-1,1\}^{ab}$.*

Define $N_{a,b}(x) = \sum_{l=0}^{a+\log b} N_l(x)$. For every non-zero value of $r_{a,b}(x)$ we have that the premise of Lemma 17 holds for exactly one l. Intuitively it is clear (we will prove something similar later) that for most x the condition of Lemma 18 holds for all other l. This implies that for a random x we have $2U_{a,b}(x) = N_{a,b}(x)$. Now observe that $N_{a,b}$ can be computed by a depth 2 threshold circuit with small weights and furthermore we only need a sum at the top gate. It is, however, easy to see that this equation for $U_{a,b}$ does not hold for all x. To remedy this we use some randomization. Let us now return to the question of computing $U_{n,m}$. We assume that m and n are also powers of 2. Let

$$\begin{aligned} z &= \{z_{ijk}, z_k \mid 0 \leq i \leq n-1,\ 0 \leq j \leq m-1,\ 0 \leq k \leq 2n-1\}, \\ r(z) &= \sum_{i,j,k} 2^{i+k+1} z_{ijk} + \sum_k 2^k z_k, \\ U'(z) &= \text{sign}(r(z)). \end{aligned}$$

Then $U'(z)$ can be obtained from $U_{4n,2mn}$ by substituting 0 for some of the variables. Let $N'(z)$ be obtained from $N_{4n,2mn}$ by the same substitution.

Now let α be an integer in $[1, 2^{2n} - 1]$ with binary representation

$$\alpha_{2n-1}\alpha_{2n-2}\ldots\alpha_1\alpha_0.$$

By substituting $x_{ij}\alpha_k$ for z_{ijk} and α_k for z_k we transform $U'(z)$ to

$$\text{sign}\,(2\alpha r_{n,m}(x) + \alpha) = U_{n,m}(x)$$

since α is positive. Transform with the same substitution $N'(z)$ to some $N_\alpha(x)$. Observe that $N_\alpha(x)$ can be written as a sum of threshold functions. For the record let us note that the total weight of any gate at the bottom level is $O\left(m^2 n^{2s+2}\right)$.

Let $r = 2r_{n,m}(x)+1$. By Remark 19 we know that $N_\alpha(x) = 2U_{n,m}(x)$ except when αr for some l does not fall under the conditions prescribed by Lemmas 17 or 18 for $a = 4n$ and $b = 2mn$. We will pick α at random from $[1, 2^{2n} - 1]$ and we need to analyze the probability of this event.

Fix the value of l, then we have the following bad events:

1. We have that $\left(1 - \frac{2}{(4n)^s}\right) 2^l \leq |r\alpha| \leq 2^l$.
2. We have that $2^{l+1} \leq |r\alpha| \leq \left(1 + \frac{1}{(4n)^s}\right) 2^{l+1}$.
3. We have $|r\alpha| \geq 2^l \left((4n)^s - 2 - \frac{2}{(4n)^s}\right)$ and it does not satisfy the hypothesis of Lemma 18 (with $a = 4n$ and $b = 2mn$.)

We have

LEMMA 20. *If $m < 2^{n/2}$ then for a fixed l the probability of either of these bad events happening is $O(n^{-s})$.*

PROOF. The first bad event is equivalent to $(1 - \frac{2}{(4n)^s})K \leq \alpha \leq K$ for $K = 2^l / |r|$. For any K the probability of this event is clearly $O(n^{-s})$. The second bad event is handled in the same way. To analyze the probability of the third bad event let us divide the analysis according to whether $2n$ is greater than $l+s\log n$. If $2n \geq l+s\log n$ then, since r is odd, $r\alpha \bmod 2^{l+s\log n}$ is almost uniformly distributed modulo $2^{l+s\log n}$ (0 is slightly underrepresented) and the bound is then obvious since we are looking at a subset of density $O(n^{-s})$. On the other hand if $2n < l + s\log n$ we argue as follows. The bad intervals for $r\alpha$ are of length $\Omega(2^l)$ and $|r| < 4m2^n = O\left(2^{3n/2}\right)$. Hence the length of the corresponding intervals for α is $\Omega\left(2^l/|r|\right) = \Omega(1)$. Now, since bad and good intervals alternate, the length of each good interval is a factor $\Omega(n^s)$ longer than the length of each bad interval and the first interval $\left[0, 2^l\left((4n)^s - 2 - \frac{2}{(4n)^s}\right)\right]$ is good, the lemma follows also in this case. □

Take n^{2s} random independent α^i and let $V(x) = \sum_{i=1}^{n^{2s}} N_{\alpha^i}(x)$. We will need the following elementary inequality (see e.g., [9]):

LEMMA 21. (HOEFFDING'S INEQUALITY) *Let $X_1, ..., X_k$ be independent random variables with values in the interval $[0,1]$ and $S = \sum_{i=1}^k X_i$. Let $\mu = \mathbf{E}[S/k]$. Then*

$$\mathbf{P}[S - k\mu \geq kt] \leq \exp\left(-\Omega(kt^2)\right).$$

Now we have

LEMMA 22. *If $m < 2^{n/2}$ and n is sufficiently large, then for any fixed r the probability that there exists x with $r = 2r_{n,m}(x) + 1$ and such that $|2n^{2s}U_{n,m}(x) - V(x)| \geq n^{s+2}$ does not exceed $\exp(-\Omega(n^2))$.*

PROOF. Denote by B the number of those pairs (l,i) $(l \leq 4n + \log(2mn), i \leq n^{2s})$ for which $r\alpha^i$ falls into an interval which is bad with respect to l. Note that $|2n^{2s}U_{n,m}(x) - V(x)| \leq 2B$ for each x with $r = 2r_{n,m}(x) + 1$. Hence we only have to check that $\mathbf{P}[B \geq n^{s+2}/2] \leq \exp(-\Omega(n^2))$.

Let B_i be the contribution of pairs (l,i) for a fixed i. Then $B = \sum_{i=1}^{n^{2s}} B_i$, where the B_i are independent. Note also that $0 \leq B_i \leq O(n)$ and that $\mathbf{E}[B_i] \leq O(n^{1-s})$. We apply now Hoeffding's inequality with $X_i = \frac{B_i}{Cn}$, $k = n^{2s}$ and $t = \frac{n^{1-s}}{3C}$ (where C is a sufficiently large constant) to get the result. □

This implies

COROLLARY 23. *If $m < 2^{n/2}$ and n is sufficiently large then there is a choice of α^i, $i = 1, 2, \ldots n^{2s}$ such that $|2n^{2s}U_{n,m}(x) - V(x)| < n^{s+2}$ for all x.*

PROOF. There are only $\exp(O(n))$ different r's so for at least one choise of α^i the inequality in Lemma 22 is fulfilled for all r's and hence for all x's. □

Please observe that $\mathrm{sign}(V(x))$ can be computed by a depth 2 threshold circuit with small weights. Using $s = 2$ in Corollary 23 together with the universality of U we get

THEOREM 24. *Suppose f can be computed by a threshold gate with arbitrary weights. Then f can be computed by a small weight threshold circuit of polynomial size and depth 2.*

In the standard terminology the above theorem says that $LT_1 \subseteq \widehat{LT}_2$. This immediately generalizes to

$$LT_d \subseteq \widehat{LT}_{2d}. \tag{10}$$

In fact, when the depth d is fixed more is true, and the easiest way to prove this is to introduce a new complexity class where we mix small and large weights.

Definition 25. *Let $\widetilde{LT}_d$ be the set of functions that can be computed by depth d threshold circuits of polynomial size, where the top gate has total weight bounded by a polynomial while we have no restrictions on the weights of the other gates.*

Clearly both $\widehat{LT}_d$ and LT_{d-1} are contained in $\widetilde{LT}_d$. For the converse we have the following striking theorem:

Theorem 26. *For any fixed $d \geq 1$, $\widetilde{LT}_d = \widehat{LT}_d$.*

Proof. For $d = 1$ there is nothing to prove. Let us first prove the theorem for $d = 2$. Take any $f \in \widetilde{LT}_2$ and a circuit which computes f. Suppose the total weight at the top level is bounded by n^t and assume for notational simplicity that each of the bottom gates computes a subfunction of $U_{n,n}$. We can also assume that the sum in the top gate never evaluates to 0. The inputs to the top gate are threshold gates with unbounded weights and n inputs. We can assume (by introducing dummy gates) that there are no direct wires from input variables to the top gate. Now for each gate G_i on the second layer pick a corresponding function $V_i(x)$ which satisfies Corollary 23 with $s = t+2$. Now instead of inputting the value of G_i to the top gate, input $V_i(x)$. Since the top gate in the circuit defining V_i is a sum, the resulting circuit after replacing the G_i can be converted into a depth 2 circuit. Suppose that the weights of the original top gate were w_i with $\sum |w_i| \leq n^t$. Then the output of this new circuit is

$$\operatorname{sign}\left(\sum w_i V_i(x)\right).$$

But now we have

$$\sum w_i V_i(x) = 2n^{2t+4} \sum w_i G_i(x) + \sum w_i (V_i(x) - 2n^{2t+4} G_i(x))$$

and by Corollary 23

$$\left|\sum w_i (V_i(x) - 2n^{2t+4} G_i(x))\right| \leq n^{t+4} \sum |w_i| \leq n^{2t+4}$$

while $|\sum w_i G_i(x)| \geq 1$. This implies that

$$\operatorname{sign}\left(\sum w_i V_i(x)\right) = \operatorname{sign}\left(\sum w_i G_i(x)\right) = f(x).$$

Thus the converted circuit computes the correct function and we have proved the theorem in the case of $d = 2$.

To prove the theorem for general d, we just need to replace the gates with unbounded weights by gates with small weights one level at the time. □

We get the following immediate corollary:

COROLLARY 27. *For any fixed* $d \geq 2$, $LT_{d-1} \subseteq \widehat{LT}_d$.

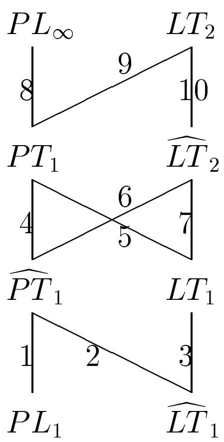
The construction in the proof of Theorem 26 gives an enormous blowup in size. For instance if we start with a circuit of size n then the resulting circuit will be of size $O(n^{c(d)})$ where $c(d)$ is exponential in the depth d. In particular, we do not know whether Corollary 27 holds for the case of d growing with n or not. Equation (10) however is true for arbitrary d and in fact we can do better:

THEOREM 28. *For any fixed* $\epsilon > 0$ *and any* $d \geq 1/\epsilon$ *possibly depending on* n, *we have* $LT_d \subseteq \widehat{LT}_{(1+\epsilon)d}$.

PROOF. Cut the LT_d-circuit into $\lfloor \epsilon d \rfloor$ slices, each slice being of constant depth, and apply Corollary 27 for each slice separately. □

8. Summary

Using our current results we can now establish *all* possible relations between the most basic complexity classes defined by small depth threshold circuits. These relations are summarized in the following picture:



Let us first comment on the inclusions: 2,3,4,5 and 10 are obvious. 6,8,9 were proved in [2]. The inclusion 1 was proved in [3] and 7 is Theorem 24.

Let us point out that all inclusions that do not follow from the above diagram are strict. The reasons are the following:

- $PL_1 \not\subseteq LT_1$ – separated by “PARITY”,
- $\widehat{LT}_2 \not\subseteq PL_\infty$ – separated by “COMPLETE QUADRATIC” [2],
- $LT_1 \not\subseteq \widehat{PT}_1$ – Corollary 9,
- $PT_1 \not\subseteq \widehat{LT}_2$ – Corollary 7,
- $\widehat{LT}_1 \not\subseteq PL_1$ – separated by “MAJORITY”,
- $PL_\infty \not\subseteq LT_2$ – seperated by counting arguments.

9. Related results

We conclude by mentioning some recent results related to this work, that answer open questions asked in a preliminary version of this paper.

Siu and Roychowdhury [17] have used Theorem 26 to get small depth polynomial size majority circuits for several problems. In particular they show that iterated addition can be done in depth 2 and multiplication in depth 3, which answers two open questions asked by us. Both these results are optimal in depth.

Also, Goldmann and Karpinski [5] improve on Theorem 26 in two respects. They get an explicit construction that does not rely on randomized existence arguments, and the blow-up in size is independent of the depth of the circuit.

Acknowledgements

We are indebted to Marek Karpinski for several interesting discussions on the topic of this paper. We are grateful to Jehoshua Bruck, Roman Smolensky, Wolfgang Maass and Yoav Freund for many useful comments, and to Pekka Orponen for pointing out to us the sources [8, 13]. We also thank the anonymous referees for valuable remarks and suggestions.

The authors are grateful for the financial support from DIMACS during a stay at Princeton University when part of this work was done.

A preliminary version of this paper appeared in *Proceedings of the 7th Structure in Complexity Theory Annual Conference.*

References

[1] E. Allender. A note on the power of threshold circuts. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 580–584, 1989.

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

[3] J. Bruck and R. Smolensky. Polynomial threshold functions, AC^0 functions and spectral norms. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 632–641, 1990.

[4] Y. Freund. Boosting a weak learning algorithm by majority. In *Workshop on Computational Learning Theory*, pages 202–216, 1990.

[5] M. Goldmann and M. Karpinski. Constructing depth $d+1$ majority circuits that simulate depth d threshold circuits. Manuscript, 1992.

[6] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. In *Proceedings of 28 IEEE Symposium on Foundations of Computer Science*, pages 99–110, 1987.

[7] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 610–618, 1990.

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

[9] M Hofri. *Probabilistic Analysis of Algorithms*. Springer-Verlag, 1987.

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

[11] 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 32nd IEEE Symposium on Foundations of Computer Science*, pages 777–782, 1991.

[12] 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.

[13] S. Muroga. *Threshold logic and its applications.* Wiley-Interscience, 1971.

[14] 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.

[15] G. Owen. *Game theory.* Academic press, second edition, 1982.

[16] K.-Y. Siu and J. Bruck. On the power of threshold circuits with small weights. *SIAM Journal on Discrete Mathematics*, 4(3):423–435, 1991.

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

[18] S. Toda. On the computational power of PP and $\oplus P$. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 514–519, 1989.

[19] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[20] A. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, pages 420–428, 1983.

[21] A. Yao. Circuits and local computation. In *Proceedings of the 21st ACM Symposium on Theory of Computing*, pages 186–196, 1989.

[22] A. Yao. On ACC and threshold circuits. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 619–627, 1990.

Manuscript received 7 December 1991

Mikael Goldmann
Department of Numerical Analysis
and Computing Science
Royal Institute of Technology
S-100 44 Stockholm, SWEDEN
migo@nada.kth.se

Johan Håstad
Department of Numerical Analysis
and Computing Science
Royal Institute of Technology
S-100 44 Stockholm, SWEDEN
johanh@nada.kth.se

Alexander A. Razborov
Steklov Mathematical Institute
Vavilova, 42, 117966, GSP-1,
Moscow, RUSSIA
razb@log.mian.su