$n^{\Omega(\log n)}$ Lower Bounds on the Size of Depth 3 Threshold Circuits with AND Gates at the Bottom

Alexander Razborov*
Steklov Mathematical Institute
Moscow, RUSSIA

Avi Wigderson†
Hebrew University
Jerusalem, ISRAEL

to appear in *Information Processing Letters*

Abstract

We present a function in ACC^0 such that any depth 3 threshold circuit which computes this function and has AND gates at the bottom must have size $n^{\Omega(\log n)}$.

Key words. computational complexity; threshold circuits; lower bounds

1. Introduction

Constant depth threshold circuits have recently gained a considerable attention (see [Raz92] for a survey of known results). There are at least two reasons for this.

The first reason is that threshold circuits are closely connected to neural nets which is one of the most active areas in computer science.

Another reason is that the complexity classes defined by constant depth threshold circuits contain many interesting Boolean functions and are closely related to other complexity classes. Perhaps the most impressive result along these lines is due to Yao [Yao90] who showed that the whole ACC^0 (which consists of functions computable by polynomial size constant depth circuits over the basis $\{\wedge, \vee, \neg, \bmod m\}$ for an arbitrary but fixed m) is

*The research was done while this author was visiting Hebrew University.

†Partially supported by the Wolfson Research Awards, administered by the Israel Academy of Sciences and Humanities

doable by depth 3 threshold circuits of size $2^{(\log n)^{O(1)}}$ which have AND gates of fan-in at most $(\log n)^{O(1)}$ at the bottom.

Currently there are only two techniques for proving lower bounds on the size of constant depth threshold circuits, both are based upon reductions to certain problems in communication complexity.

The first technique uses (at least implicitly) two-party communication model. It appeared first in the paper [HMP*87] and then, in different contexts, in
[Kra91, KW91, GHR92]. This method allows one to obtain exponential lower bounds for depth 2 threshold circuits [HMP*87, GHR92] and, more generally, for depth 2 circuits with symmetric gates of a certain restricted kind [Kra91, KW91].

The second technique which was suggested by Håstad and Goldmann [HG91], uses the powerful result in the multiparty communication complexity due to Babai, Nisan and Szegedy [BNS89]. Their method gives exponential lower bounds for depth 3 threshold circuits *with bottom fan-in at most* $\frac{1}{3}\log n$. The bounds are achieved for the ACC^0 function

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

Håstad–Goldmann's result can be viewed as a sort of lower bound to Yao's above mentioned reduction: the restriction $(\log n)^{O(1)}$ on the fan-in can not be weakened to anything less than $\frac{1}{3}\log n$.

In this note we are primarily interested in a slightly different model. Namely, we allow at the bottom level ANDs only but do not place any restrictions on their fan-in. Any polynomial size depth 3 threshold circuit with logarithmic fan-in at the bottom level can be easily converted to a polynomial size depth 3 threshold circuit with ANDs at the bottom (see Theorem 6 below). Hence our model is stronger than that considered by Håstad and Goldmann. Our bounds, however, are weaker.

More precisely, we show that the function

$$f_n(x) \rightleftharpoons \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{\log n} \bigoplus_{k=1}^{n} x_{ijk}$$

which is clearly in ACC^0, requires size $n^{\Omega(\log n)}$ when computed by depth 3 threshold circuits with ANDs at the bottom (Theorem 3). This bound is easily seen to be tight (Theorem

5). It follows that the size bound $2^{(\log n)^{O(1)}}$ in Yao's reduction can not be relaxed to merely polynomial size, even if we allow AND gates of an arbitrary fan-in at the bottom.

Johan Håstad noticed that our machinery can also be used for proving $N^{\Omega(\log N)}$ lower bounds on the size of depth 3 threshold circuits with bottom fan-in at most $N^{1-\epsilon}$ for an ACC^0 function similar to f_n, where N is the total number of variables and $\epsilon > 0$ is an arbitrary constant. With his kind permission, this result is included into the paper (see Theorem 8 and Corollary 9).

The proofs are obtained by combining the Håstad-Goldmann bound with random restrictions in style of [FSS84, Yao85, Has86, And87].

2. The results

All notation used in this note is standard (see e.g. [HG91] or [Raz92]). We will be considering only *unweighted* threshold circuits which can be alternatively described as circuits consisting of (monotone) *Boolean* threshold functions with negations allowed at inputs only.

The following function originally introduced in [BNS89] is called the *generalized inner product* mod 2:

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

Håstad and Goldmann proved for it the following bound:

Proposition 1 (Håstad, Goldmann). *Any depth* 3 *threshold circuit which computes* $GIP2_{n,s}$ *and has bottom fan-in at most* $(s-1)$*, must be of size* $\exp\Bigl(\Omega\Bigl(\frac{n}{s4^s}\Bigr)\Bigr)$.

Corollary 2. *Any depth* 3 *threshold circuit which computes* $GIP2_{n,\log n}$ *and has bottom fan-in at most* $\frac{1}{3}\log n$*, must be of size* $\exp\Bigl(n^{\Omega(1)}\Bigr)$.

We define the following function:

$$f_n(x) \rightleftharpoons \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{\log n} \bigoplus_{k=1}^{n} x_{ijk}. \tag{1}$$

Note that in the notation of [KRW91], f_n is the composition of $GIP2$ and $PARITY$ functions, namely $f_n \equiv GIP2_{n,\log n} \circ PARITY_n$.

The main result of this note is the following

Theorem 3. *Any depth* 3 *threshold circuit which computes* f_n *and has AND gates at the bottom must be of size* $n^{\Omega(\log n)}$.

Remark 4. In fact, the middle level of the circuit in Proposition 1 may have *arbitrary* symmetric gates. It will become clear from the proof that the same holds for our theorem 3.

Proof of Theorem 3: Let C_n be a depth 3 threshold circuits which has ANDs at the bottom and computes f_n. The strategy of the proof is to hit C_n with a random restriction in order to reduce fan-in at the bottom level. Then we will apply Corollary 2.

Set $p \rightleftharpoons \frac{2\ln n}{n}$. Let $\boldsymbol{\rho}$ be the random restriction which assigns independently each variable to $*$ with probability p and to $0, 1$ with equal probabilities $\frac{1-p}{2}$. Given a Boolean function g in n variables and a restriction ρ, we will denote by $\rho(g)$ the function we get by doing the substitutions prescribed by ρ. Similarly, $\rho(C_n)$ is defined as the result of doing the same substitution with the circuit C_n followed by all obvious cancellations at the bottom level made possible by applying ρ.

Let $\mathcal{K}$ be a conjunction of literals. Denote by $|\mathcal{K}|$ the number of literals in $\mathcal{K}$. We are going to show that for each $\mathcal{K}$ we have

$$\mathbf{P}\left[\boldsymbol{\rho}(\mathcal{K}) \text{ has fan-in} \geq \frac{1}{3}\log n\right] \leq n^{-\Omega(\log n)}. \tag{2}$$

Consider two cases.

Case 1. $|\mathcal{K}| \leq (\log n)^2$.

We have

$$\mathbf{P}\left[\boldsymbol{\rho}(\mathcal{K}) \text{ has fan-in} \geq \frac{1}{3}\log n\right] \leq \left(\begin{array}{c}(\log n)^2 \\ \frac{1}{3}\log n\end{array}\right) \cdot p^{\frac{1}{3}\log n} \leq O(p\log n)^{\frac{1}{3}\log n} \leq n^{-\Omega(\log n)}.$$

Case 2. $|\mathcal{K}| \geq (\log n)^2$.

In this case

$$\mathbf{P}\left[\boldsymbol{\rho}(\mathcal{K}) \text{ has fan-in} \geq \frac{1}{3}\log n\right] \leq \mathbf{P}[\boldsymbol{\rho}(\mathcal{K}) \not\equiv 0] = \left(\frac{1+p}{2}\right)^{|\mathcal{K}|} \leq n^{-\Omega(\log n)}.$$

Now, when we have (2), the reduction to Corollary 2 becomes easy. Namely, if C_n were of size $\leq n^{\epsilon\log n}$ for a sufficiently small ϵ, (2) would imply

$$\mathbf{P}\left[\boldsymbol{\rho}(C_n) \text{ has fan-in} \leq \frac{1}{3}\log n \text{ at the bottom level}\right] \geq 1 - o(1). \tag{3}$$

On the other hand, for each fixed pair (i,j),

$$\mathbf{P}\left[\boldsymbol{\rho}\left(\bigoplus_{k=1}^{n} x_{ijk}\right) \text{ is a constant}\right] = (1-p)^n \leq n^{-2}$$

hence

$$\mathbf{P}\left[\text{for all } (i,j),\ \boldsymbol{\rho}\left(\bigoplus_{k=1}^{n} x_{ijk}\right) \text{ is not a constant}\right] \geq 1 - o(1). \tag{4}$$

Pick ρ so that the two events in the left-hand sides of (3) and (4) are fulfilled. Then we have the desired contradiction with Corollary 2 since the event in (4) implies that $GIP2_{n,\log n}$ is a subfunction of $\rho(f_n)$ (possibly up to negating some variables).

The proof is complete.■

The following shows that the bound $n^{\Omega(\log n)}$ in Theorem 3 is tight.

Theorem 5. *f_n is computable by depth 3 threshold circuits of size $n^{O(\log n)}$ with ANDs at the bottom.*

Proof: Using distributivity, we can rewrite f_n as

$$f_n \equiv \bigoplus_{i=1}^{n} \bigoplus_{\pi:[\log n]\longrightarrow[n]} \bigwedge_{j=1}^{\log n} x_{i,j,\pi(j)}.$$

Now we just have to merge the two $\oplus$-levels together and replace the resulting PARITY gate by a depth 2 threshold circuit of size $n^{O(\log n)}$ (see e.g. [HMP*87]).■

The following simple observation shows that our model is at least as strong as the model from [HG91].

Theorem 6. *Any polynomial size depth 3 threshold circuit with fan-in $O(\log n)$ at the bottom level can be simulated by a polynomial size depth 3 threshold circuit with ANDs at the bottom level.*

Proof: As each bottom gate has fan-in $O(\log n)$, the function it computes can be written as a polynomial size sum (over the integers) of conjunctions of length $O(\log n)$ (in fact, the standard DNF of this function has the required form). These summation gates can be merged (in a standard fashion) with the middle layer threshold gates to yield the result.■

Remark 7. Just as in the proof of Theorem 5, $GIP2_{n,s}$ itself can be done by threshold circuits with ANDs at the bottom within polynomial size and depth 3. In view of Proposition 1, this suggests that the simulation opposite to Theorem 6 is rather unlikely.

We conclude this note with another application of our technique suggested by Johan Håstad. In order to state the corresponding results, it will be convenient to slightly generalize definition (1) of the function f_n. Namely, let

$$f_{m,n}(x) \rightleftharpoons \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{\log n} \bigoplus_{k=1}^{m} x_{ijk}.$$

Theorem 8. *Any depth* 3 *threshold circuit which computes* $f_{m,n}$ *and has fan-in at most* s *at the bottom, must be of size* $\min\Big(\Omega(m/s)^{\Omega(\log n)},\ \exp\Big(n^{\Omega(1)}\Big)\Big)$.

Proof: We follow the proof of Theorem 3 with $p \rightleftharpoons \frac{2\ln n}{m}$. The same analysis as in the case 1 shows that for any function f in at most s variables,

$$\mathbf{P}\Big[\boldsymbol{\rho}(f) \text{ depends on } \geq \frac{1}{3}\log n \text{ variables}\Big] \leq \left(\begin{array}{c} s \\ \frac{1}{3}\log n \end{array} \right) \cdot p^{\frac{1}{3}\log n} \leq O\left(\frac{ps}{\log n}\right)^{\frac{1}{3}\log n} \leq$$
$$O(s/m)^{\Omega(\log n)},$$

which proves the analogue of (3). The analogue of (4) is proved exactly as before.■

Corollary 9. *For any fixed* $\epsilon > 0$, *let* $n = n(N) \rightleftharpoons N^{\epsilon/2}$ *and* $m = m(N) \rightleftharpoons \frac{N}{n\log n}$. *Then* $f_{m,n}$ *is a function in* N *variables such that any depth* 3 *threshold circuit computing this function and having fan-in at most* $N^{1-\epsilon}$ *at the bottom, must be of size* $N^{\Omega(\log N)}$.

3. Acknowledgement

We are grateful to Johan Håstad for his permission to include here Theorem 8.

References

[And87] А.Е. Андреев. О методе получения более чем квадратичных нижних оценок для сложности π-схем. *Вестник МГУ, сер. матем и механ.*,

42(1):63–66, 1987. A.E. Andreev, On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.* 42(1):63-66, 1987.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and logspace-hard pseudorandom sequences. In *Proceedings of the* 21*st ACM STOC*, pages 1–11, 1989.

[FSS84] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits and the polynomial time hierarchy. *Math. Syst. Theory*, 17:13–27, 1984.

[GHR92] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. In *Proceedings of the* 7*th Structure in Complexity Theory Annual Conference*, pages 2–13, 1992. To appear in *Computational Complexity*.

[Has86] J. Håstad. *Computational limitations on Small Depth Circuits*. PhD thesis, Massachusetts Institute of Technology, 1986.

[HG91] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1(2):113–129, 1991.

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

[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.

[KRW91] M. Karchmer, R. Raz, and A. Wigderson. Super-logarithmic depth lower bounds via direct sum in communication complexity. In 6*th Structure in Complexity Theory Conference*, pages 299–304, 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* 32*th IEEE FOCS*, pages 777–782, 1991.

[Raz92] A. Razborov. On small depth threshold circuits. In *Proceedings of the SWAT 92, Lecture Notes in Computer Science,* 621, pages 42–52, Springer-Verlag, New York/Berlin, 1992.

[Yao85] A. Yao. Separating the polynomial-time hierarchy by oracles. In *Proceedings of the* 26*th IEEE FOCS*, pages 1–10, 1985.

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