Pseudorandom Bits for Constant Depth Circuits with Few Arbitrary Symmetric Gates

Emanuele Viola*
Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138
viola@eecs.harvard.edu

April 2, 2005

Abstract

We exhibit an explicitly computable 'pseudorandom' generator stretching l bits into $m(l) = l^{\Omega(\log l)}$ bits that look random to constant-depth circuits of size $m(l)$ with $\log m(l)$ arbitrary symmetric gates (e.g. PARITY, MAJORITY). This improves on a generator by Luby, Velickovic and Wigderson (ISTCS '93) that achieves the same stretch but only fools circuits of depth 2 with one arbitrary symmetric gate at the top. Our generator fools a strictly richer class of circuits than Nisan's generator for constant depth circuits (Combinatorica '91) (but Nisan's generator has a much bigger stretch).

In particular, we conclude that every function computable by uniform poly(n)-size *probabilistic* constant depth circuits with $O(\log n)$ arbitrary symmetric gates is in $\mathit{TIME}\left(2^{n^{o(1)}}\right)$. This seems to be the richest probabilistic circuit class known to admit a subexponential derandomization.

Our generator is obtained by constructing an explicit function $f : \{0,1\}^n \to \{0,1\}$ that is very hard *on average* for constant-depth circuits of size $n^{\epsilon \cdot \log n}$ with $\epsilon \log^2 n$ arbitrary symmetric gates, and plugging it into the Nisan-Wigderson pseudorandom generator construction (FOCS '88). The proof of the average-case hardness of this function is a modification of arguments by Razborov and Wigderson (IPL '93), and Hansen and Miltersen (MFCS '04), and combines Håstad's switching lemma (STOC '86) with a multiparty communication complexity lower bound by Babai, Nisan and Szegedy (STOC '89).

*Research supported by NSF grant CCR-0133096, US-Israel BSF grant 2002246, ONR grant N-00014-04-1-0478.

1 Introduction

A *pseudorandom generator* $G : \{0,1\}^l \to \{0,1\}^m$ is an efficient procedure that stretches l input bits into $m \gg l$ output bits such that the output distribution of the generator *fools* small circuits. That is, for every circuit C of size m we have

$$\left| \Pr_{x \in \{0,1\}^l}[C(G(x)) = 1] - \Pr_{x \in \{0,1\}^m}[C(x) = 1] \right| \leq \frac{1}{m}.$$

Pseudorandom generators have found a striking variety of applications in Complexity Theory, most notably to *derandomize* probabilistic algorithms.

Starting with the seminal work of Nisan and Wigderson [NW], a series of results (e.g. [BFNW, STV, SU, Uma]) show how to construct pseudorandom generators starting from an explicit function that requires circuits of superpolynomial size. However, no such function is known to exist.

On the other hand, pseudorandom generators that fool *restricted* kinds of circuits, such as *constant-depth* circuits with unbounded fan-in, are already very interesting. They also have a large variety of applications (e.g. [NW, HVV]) and are central to understanding the power of randomness in restricted classes of algorithms. While there has been exciting progress in constructing explicit functions that require superpolynomial size constant-depth circuits with certain kinds of gates (e.g. [Hås, Raz, Smo, HG, RW, HM]), no explicit function is known to require superpolynomial size constant-depth circuits with MAJORITY gates (cf. [RR]). This is an obstacle to construct pseudorandom generators, as most constructions need such a function. This need is due to the fact that the reductions in the proofs of correctness of these constructions use (a polynomial number of) MAJORITY gates (cf. [Agr, Vio]).

But when starting from an *average-case* hard function, the reduction in the proof of correctness of the Nisan-Wigderson construction [NW] does not require MAJORITY gates (where a function $f : \{0,1\}^n \to \{0,1\}$ is average-case hard if polynomial-size circuits fail to compute f with probability at least $1/2 - 1/n^{\omega(1)}$ over random input). Thus, one can plug average-case lower bounds into the Nisan-Wigderson construction to get a generator that fools small constant-depth circuits. This approach is used in a celebrated work by Nisan [Nis] (that actually predates the more general construction in [NW]) where he exhibits a generator $G : \{0,1\}^l \to \{0,1\}^{2^{l^{\Omega(1)}}}$ that fools small AC^0 circuits (i.e. constant-depth circuits with AND and OR gates). This generator is based on the fact that PARITY is very average-case hard for small AC^0 circuits [Hås].

Subsequently, Luby, Velickovic and Wigderson (Theorem 2 in [LVW]) build a generator $G : \{0,1\}^l \to \{0,1\}^{l^{\Omega(\log l)}}$ that fools small SYM $\circ$ AND circuits, i.e. depth 2 circuits with one *arbitrary symmetric gate* at the top and AND gates at the bottom. By arbitrary symmetric gate we mean a gate that computes an arbitrary function whose value depends only on the number of input bits being 1, important examples being PARITY and MAJORITY. This generator is based on the fact that the 'generalized inner product' function is average-case hard for small SYM $\circ$ AND circuits with small bottom fan-in [BNS, HG].

The above two generators ([Nis] and Theorem 2 in [LVW]) fool two incomparable classes of circuits (i.e. small AC^0 circuits and small SYM $\circ$ AND circuits). In this work we exhibit a generator that fools a class of circuits strictly richer than both of them, namely small constant-depth circuits with few arbitrary symmetric gates.

1.1 Our Results

In this paper we exhibit the following generator.

Theorem 1. *For every constant d there is a constant $\epsilon > 0$ such that for every l there is a generator $G : \{0,1\}^l \to \{0,1\}^m$, where $m = m(l) := l^{\epsilon \log l}$, such that for every circuit C of size m and depth d with $\log m(l)$ arbitrary symmetric gates, we have:*

$$\left| \Pr_{x \in \{0,1\}^m}[C(x) = 1] - \Pr_{x \in \{0,1\}^l}[C(G(x)) = 1] \right| \leq \frac{1}{m},$$

and given $x \in \{0,1\}^l, i \leq m$, we can compute the i-th output bit of $G(x)$ in time $\mathrm{poly}(l)$.

The generator in Theorem 1 improves on the generator by Luby, Velickovic and Wigderson (Theorem 2 in [LVW]) that achieves the same stretch (up to a different constant ϵ) but only fools circuits of depth 2 (as opposed to any constant depth) with one symmetric gate at the top. (We elaborate more on the difference between the two generators in Section 6.) The generator in Theorem 1 also fools a strictly richer class of circuits than Nisan's generator that fools constant depth circuits [Nis]. (However, Nisan's generator has a much bigger stretch: it stretches l bits to $2^{l^{\Omega(1)}}$ bits.)

As a standard consequence of Theorem 1 we obtain the following subexponential derandomization of probabilistic constant depth circuits with a constant number of arbitrary symmetric gates. This seems to be the richest probabilistic circuit class known to admit a subexponential derandomization. (See, e.g., [NW] for the connection between generators and derandomization.)

Corollary 2. *Let a function f be computed by a uniform family of probabilistic* $\mathrm{poly}(n)$*-size constant depth circuits with $O(\log n)$ arbitrary symmetric gates. Then f can be computed in deterministic time $exp(2^{O(\sqrt{\log n})}) = 2^{n^{o(1)}}$.*

1.2 Techniques

The generator in Theorem 1 is obtained by plugging into the Nisan-Wigderson pseudorandom generator construction [NW] a function that is very hard on average for 'small' constant-depth circuits with 'few' arbitrary symmetric gates (cf. Theorem 3 below). Here a simple and crucial observation is that the reduction in the proof of correctness of the Nisan-Wigderson generator (essentially) does not increase the number of arbitrary symmetric gates.

Given our average-case hardness result (Theorem 3), the construction of our generator is simpler than the construction of the (weaker) generator by Luby, Velickovic and Wigderson (Theorem 2 in [LVW]) that uses more involved combinatorial arguments than those in [NW]. These more involved combinatorial arguments were probably used because the generator in [LVW] builds on a function that is hard on average for circuits of depth 2 (as opposed to any constant depth), and thus one cannot use directly the Nisan-Wigderson construction [NW] since the reduction in its proof of correctness increases the depth by 1.

We now state our average-case hardness result.

Theorem 3. *There is a function $f : \{0,1\}^* \to \{0,1\}$ computable in polynomial time such that for every constant d there is a constant $\epsilon > 0$ such that for every n and every circuit C of size $n^{\epsilon \cdot \log n}$, depth d and with $\epsilon \log^2 n$ arbitrary symmetric gates, the following holds:*

$$\Pr_{x \in \{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\epsilon \cdot \log n}.$$

We now explain the techniques involved in proving Theorem 3. To simplify the discussion we first focus on how to prove an average-case hardness result for 'small' constant-depth circuits with *one* arbitrary symmetric gate at the top, i.e. 'small' $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits (Theorem 4). The extension to circuits with more arbitrary symmetric gates is deferred to the paragraph "Circuits with more Arbitrary Symmetric Gates" below. We obtain our average-case hardness result for 'small' $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits through a modification of previous lower bounds. We now discuss these previous lower bounds, then we discuss why they are not sufficient for our purposes, and then we sketch the proof of our average-case hardness result for 'small' $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits.

Previous Lower Bounds: Babai, Nisan and Szegedy [BNS] prove that the "generalized inner product" function (i.e., $GIP_{n,s}(x) := \bigoplus_{i \leq n} \bigwedge_{j \leq s} x_{i,j}$) is very hard *on average* for multiparty communication complexity protocols among 'few' parties that communicate 'little'.

Håstad and Goldmann [HG] notice that any function computed by a 'small' depth 2 circuit with an arbitrary symmetric gate of unbounded fan-in at the top and (arbitrary) gates of 'small' fan-in at the bottom can be computed by a multiparty communication complexity protocol among 'few' parties communicating 'little'. Thus, by the above result [BNS], they obtain that *GIP* is average-case hard for that kind of circuits. Now, by the so-called "ϵ-discriminator lemma"[1] of Hajnal et. al. [HMP$^+$] they conclude that *GIP* cannot be computed, in the *worst-case*, by 'small' depth 3 circuits with one majority gate of unbounded fan-in at the top, arbitrary symmetric gates of unbounded fan-in in the middle, and (arbitrary) gates of 'small' fan-in at the bottom.

Razborov and Wigderson [RW] eliminate the constrain on the bottom fan-in: they exhibit a new function *RW* that cannot be computed, in the worst-case, by 'small' depth 3 circuits with one majority gate at the top, symmetric gates in the middle, and AND gates at the bottom, where all the gates have unbounded fan-in ($\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{AND}$ circuits). Their function *RW* is obtained from *GIP* by replacing each input variable with a parity function, i.e. $RW(x) := \bigoplus_{i \leq n} \bigwedge_{j \leq \log n} \bigoplus_{k \leq n} x_{i,j,k}$.

To explain their argument we introduce *restrictions* [FSS]. A restriction on m variables $x_1, x_2, \ldots, x_m$ is a map $\rho : \{x_1, x_2, \ldots, x_m\} \to \{0, 1, *\}$. For a circuit C we denote by $C|_\rho$ the circuit we get by doing the substitutions prescribed by ρ, followed by all obvious cancellations made possible by applying ρ. The input variables of $C|_\rho$ are the variables which were given the value * by ρ.

The argument in [RW] goes as follows: suppose that *RW* is computable by a 'small' $\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{AND}$ circuit C. Then there is a restriction ρ that accomplishes simultaneously two things: (1) $C|_\rho$ has 'small' bottom fan-in and (2) $C|_\rho$ is still computing *GIP* as a

[1] This lemma states that if a function is computed by a 'small' circuit with a MAJORITY gate at the top, then some input circuit to the MAJORITY gate computes the function 'well' on average.

subfunction. Note that, by definition of RW and by the nature of parity, (2) happens whenever for every i, j there is k such that $\rho(x_{i,j,k}) = *$. But (1) and (2) contradict the above result by Håstad and Goldmann.

Finally, Hansen and Miltersen [HM] observed that RW actually cannot be computed by 'small' circuits of *any* constant depth with one majority gate at the top, and one layer of arbitrary symmetric gates immediately below it, where all the gates have unbounded fan-in ($\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{AC}^0$ circuits). The argument in [HM] goes as follows: suppose that RW is computable by a 'small' $\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{AC}^0$ circuit C. Then there is a restriction ρ that accomplishes simultaneously two things: (1') $C|_\rho$ is equivalent to a 'small' $\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{AND}$ circuit and (2') $C|_\rho$ is still computing RW on an input of polynomially related size. (1') is obtained through Håstad's switching lemma [Hås], and for (2') they show that for every i, j there are 'many' k's such that $\rho(x_{i,j,k}) = *$. But (1') and (2') contradict the above result by Razborov and Wigderson.

Why Previous Lower Bounds Are Not Sufficient To Our Purposes: The main problem with these previous lower bounds is that they only give a function that is *worst-case* hard for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits, while as explained before we need a function that is *average-case* hard. In fact, the choice of parameters in the definition of RW implies that $\Pr_x[RW(x) = 0] = 1/2 + \Omega(1)$, and thus RW cannot be average-case hard (since the constant size circuit that always outputs '0' computes the function fairly well on average). Moreover the choice of parameters for the restrictions in [RW] does not guarantee that the reduction holds with high probability, which is needed to establish average-case hardness.

Proof Sketch of our Average-Case Hardness Result for $\mathrm{SYM} \circ \mathrm{AC}^0$ Circuits: We define a function f (similar to RW, but with a different choice of parameters), and we show that f is average-case hard for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits. Our argument simplifies the previous ones and goes as follows: Suppose that C is a small $\mathrm{SYM} \circ \mathrm{AC}^0$ circuit computing f. We argue that, with high probability $\left(1 - n^{-\Omega(\log n)}\right)$ over the choice of a random restriction ρ, both the following two events happen:

- Event $E_1 :=$ the function computed by $C|_\rho$ is computable by a multiparty communication complexity protocol among 'few' parties communicating 'little'.
- Event $E_2 := C|_\rho$ is computing GIP as a subfunction.

To show E_1 we use Håstad's switching lemma to argue that with high probability over ρ, $C|_\rho$ is equivalent to a 'small' depth-2 circuit with a symmetric gate at the top (of unbounded fan-in) and AND gates of 'small' fan-in at the bottom, and then use Håstad and Goldmann's connection [HG] between these circuits and multiparty communication complexity protocols (cf. paragraph "Previous Lower Bounds"). Now, when ρ satisfies both E_1 and E_2 we have that $\Pr_y[C|_\rho(y) \neq GIP(y)] \geq 1/2 - n^{-\Omega(\log n)}$ by the multiparty communication complexity lower bound by Babai, Nisan and Szegedy [BNS]. Since we can think of a random input x as being generated by first choosing a random restriction ρ and then a random input y for the *'s of ρ (so that $C(x) = C|_\rho(y)$), we have that

$$\begin{aligned}
&\Pr_x[C(x) \neq f(x)] \\
&\geq \Pr_y\left[C|_\rho(y) \neq GIP(y) \,\middle|\, \rho \text{ satisfies } E_1 \text{ and } E_2\right] \cdot \Pr_\rho\left[\rho \text{ satisfies } E_1 \text{ and } E_2\right] \\
&\geq \left(1/2 - n^{-\Omega(\log n)}\right) \cdot \left(1 - n^{-\Omega(\log n)}\right) \\
&= 1/2 - n^{-\Omega(\log n)}.
\end{aligned}$$

We show that the above argument goes through for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits C of size $n^{\Omega(\log n)}$ and this proves our average-case hardness result for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits.

Circuits with more Arbitrary Symmetric Gates: Before discussing how to extend our techniques to get an average-case hardness result for 'small' constant-depth circuits with $\epsilon \log^2 n$ arbitrary symmetric gates, we would like to mention two other approaches that give weaker bounds. Beigel (Theorem 5.1 in [Bei]) shows that for every circuit of size S and depth d with σ arbitrary symmetric gates there is another circuit of size $S^{2^\sigma+1}$ and depth $d+1$ with *one* arbitrary symmetric gate at the top computing the same function. Combining this with our average-case hardness result for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits one obtains an average-case hardness result for constant-depth circuits of size $n^{\epsilon \cdot \log n}$ with a constant number of arbitrary symmetric gates. But this approach gives weaker bounds (than $n^{\Omega(\log n)}$) if the circuits have $\sigma = \omega(1)$ arbitrary symmetric gates; and it gives nothing at all if the circuits have $\sigma = \log \log n$ arbitrary symmetric gates.

Chattopadhyay and Hansen [CH] prove a *worst-case* hardness result for constant-depth circuits of size $n^{\epsilon \cdot \log n}$ with $\epsilon \log^2 n$ arbitrary symmetric gates. They obtain this result independently from ours. Subsequently to our results for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits, they also prove an *average-case* hardness result for constant-depth circuits of size $n^{\epsilon \cdot \log n}$ with fewer arbitrary symmetric gates, namely $\epsilon \log n$.

Inspired by the work of Chattopadhyay and Hansen, we prove an average-case hardness result for constant-depth circuits of size $n^{\epsilon \cdot \log n}$ with $\epsilon \log^2 n$ arbitrary symmetric gates (Theorem 3). The proof of our result has the same structure of our result for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits discussed in the previous paragraph. The only difference is proving that, if C is a 'small' constant-depth circuit with $\epsilon \log^2 n$ arbitrary symmetric gates, then with high probability over a random restriction ρ the function computed by $C|_\rho$ is computable by a multiparty communication complexity protocol P among 'few' parties communicating 'little' (cf. event E_1 in the previous paragraph). The idea is to let the protocol P compute the outputs of each arbitrary symmetric gate in order. Specifically, first fix a topological order of the arbitrary symmetric gates (the simple order induced by reading the gates level by level from the inputs to the output node will do). Now consider the $\mathrm{SYM} \circ \mathrm{AC}^0$ subcircuit C_1 whose root is the first arbitrary symmetric gate in this order. We know that with high probability over the restriction ρ, the function computed by $C_1|_\rho$ is computable by a multiparty communication complexity protocol P_1 exchanging 'few' bits (cf. event E_1 in the previous paragraph). Our protocol P first simulates P_1 to determine the output b_1 of $C_1|_\rho$. Then it considers the $\mathrm{SYM} \circ \mathrm{AC}^0$ circuit C_2 whose root is the second arbitrary symmetric gate, and where the first arbitrary symmetric gate is replaced with the constant b_1. Again, we argue that the function computed by $C_2|_\rho$ is computable by a multiparty communication complexity protocol P_2 exchanging 'few' bits. Our protocol P now simulates P_2 to determine the output b_2 of $C_2|_\rho$. We continue in this way until all the arbitrary symmetric gates are computed. Assuming w.l.o.g. that the output gate of the circuit is included in the arbitrary symmetric gates, the protocol P computes $C|_\rho$.

1.3 Organization

This paper is organized as follows. In Section 2 we fix some notation. In Section 3 we show how our average-case hardness result (Theorem 3) implies our generator (Theorem 1). In Section 4 we prove our average-case hardness result for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits. In Section 5 we extend this to our average-case hardness result for constant-depth circuits with few arbitrary symmetric gates, thus proving Theorem 3. In Section 6 we elaborate on why our generator improves on the generator by Luby, Velickovic and Wigderson (Theorem 2 in [LVW]). The proof of a result in this last section is given in Appendix A. In Section 7 we discuss some open problems.

2 Preliminaries

An *arbitrary symmetric gate* is a gate that computes an arbitrary symmetric function, i.e. a function whose value depends only on the number of input bits being 1 (e.g. PARITY, MAJORITY). We use standard definitions of constant depth circuits, which we now briefly recall. Constant depth circuits consist of AND, OR and possibly other gates (e.g. one arbitrary symmetric gates). It is intended that all gates whose type is not specified are either AND or OR, and that AND and OR gates are not counted towards arbitrary symmetric gates. All circuit gates, unless specified otherwise, have unbounded fan-in. Circuits take both input variables and their negations as input. *Bottom* gates are the one adjacent to the input bits. The *top* gate is the output gate. Levels are numbered from the bottom. So the input bits are at level 0, the bottom gates at level 1 and so on. Gates at level i are connected to gates at levels $i-1$ and $i+1$ only. The *depth* of a circuit is the longest path from any input to the output. The *size* of a circuit is the number of gates in it. Multiple edges between pairs of nodes in the circuit are not allowed (otherwise an arbitrary symmetric gate can compute any function; this convention is standard in the literature, e.g. [HG]).

3 From Average-Case Hardness to Pseudorandomness

In this section we show how our average-case hardness result (Theorem 3) implies our generator (Theorem 1). We restate the theorems for the reader's convenience.

Theorem (1, restated). *For every constant d there is a constant $\epsilon > 0$ such that for every l there is a generator $G : \{0,1\}^l \to \{0,1\}^m$, where $m = m(l) := l^{\epsilon \log l}$, such that for every circuit C of size m and depth d with $\log m(l)$ arbitrary symmetric gates, we have:*

$$\left| \Pr_{x \in \{0,1\}^m}[C(x) = 1] - \Pr_{x \in \{0,1\}^l}[C(G(x)) = 1] \right| \leq \frac{1}{m},$$

and given $x \in \{0,1\}^l, i \leq m$, we can compute the i-th output bit of $G(x)$ in time poly(l).

Theorem (3, restated). *There is a function $f : \{0,1\}^* \to \{0,1\}$ computable in polynomial time such that for every constant d there is a constant $\epsilon > 0$ such that for every n and every circuit C of size $n^{\epsilon \cdot \log n}$, depth d and with $\epsilon \log^2 n$ arbitrary symmetric gates, the following holds:*

$$\Pr_{x \in \{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\epsilon \cdot \log n}.$$

Proof of Theorem 1, assuming Theorem 3. The generator is obtained by plugging the function from Theorem 3 into Nisan-Wigderson's pseudorandom generator construction [NW]. Specifically, they show how given a function $f : \{0,1\}^{\sqrt{l/2}} \to \{0,1\}$ and a parameter m (which we set to be $m(l) := l^{\epsilon \cdot \log l}$) to construct a generator $G : \{0,1\}^l \to \{0,1\}^m$ such that every circuit C for which

$$\left| \Pr_{x \in \{0,1\}^m}[C(x) = 1] - \Pr_{x \in \{0,1\}^l}[C(G(x)) = 1] \right| > 1/m$$

can be transformed into another circuit C' of size $|C| + \text{poly}(m)$ that computes the function f correctly with probability (over random input) greater than $1/2 + 1/m^2 = 1/2 + 1/l^{2\epsilon \log l}$.

As observed in [Nis, NW], C' is simply C with one more layer of AND (or OR) gates at the bottom, and possibly negating the output. Adding one layer of AND (or OR) gates at the bottom clearly does not increase the number of arbitrary symmetric gates in C, and we can think of negating the output by, say, including the top gate in the arbitrary symmetric gates and complementing it. Thus, if C is a circuit of size $m = m(l) = l^{\epsilon \log l}$ of depth d with $\log m(l) = \epsilon \log^2 l$ arbitrary symmetric gates we obtain another circuit C' of size $l^{O(\epsilon \log l)}$ of depth $d+1$ with $1 + \epsilon \log^2 l$ arbitrary symmetric gates that computes $f : \{0,1\}^{\sqrt{l/2}} \to \{0,1\}$ with probability greater than $1/2 + 1/l^{2\epsilon \log l}$. This contradicts Theorem 3 for sufficiently small ϵ.

The complexity of the generator follows from the arguments in [Nis, NW] and the fact that f is computable in time poly(l). □

4 Average-Case Hardness for $\mathrm{SYM} \circ \mathrm{AC}^0$ circuits

In this section we prove our average-case hardness result for 'small' constant-depth circuits with one arbitrary symmetric gate at the top.

Theorem 4. *There is a function $f : \{0,1\}^* \to \{0,1\}$ computable in polynomial time such that for every constant d there is a constant $\epsilon > 0$ such that for every n and every circuit C of size $n^{\epsilon \cdot \log n}$, depth d, with 1 arbitrary symmetric gate at the top, the following holds:*

$$\Pr_{x \in \{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\epsilon \cdot \log n}.$$

In the rest of this section we prove Theorem 4. In the proof we use two results which we describe in the following two subsections. The first is a version of Håstad's switching lemma [Hås] due to Beame [Bea], and the second is the multiparty communication complexity lower bound for *GIP* by Babai, Nisan and Szegedy [BNS].

4.1 Switching Lemma

We now describe the switching lemma we use in the proof of Theorem 4. As in [HM], the crucial property that we need is that the DNF obtained after applying the restriction is such that all the terms are mutually contradictory, i.e. no input satisfies more than one term. This allows us to merge the top OR gate of the DNF in the symmetric gate at the top (cf. Fact 6). The fact that this property holds for Håstad's switching lemma was already noted by Boppana and Håstad in [Hås] (inside the proof of Lemma 8.3). However, there does not seem to be a full proof of this fact in the literature. For this reason we use a slightly different version of the Håstad's switching lemma, due to Beame [Bea].

A *restriction* on m variables $x_1, x_2, \ldots, x_m$ is a map $\rho : \{x_1, x_2, \ldots, x_m\} \to \{0, 1, *\}$. For a function $f : \{0,1\}^m \to \{0,1\}$ we denote by $f|_\rho$ the function we get by doing the substitutions prescribed by ρ. $f|_\rho$ will be a function of the variables that were given the value * by ρ. Similar conventions hold for circuits. If ρ and ρ' are restrictions, and ρ' is defined on the variables mapped to * by ρ we write $\rho\rho'$ for the restriction obtained by combining ρ and ρ', so that $f|_{\rho\rho'} = (f|_\rho)|_{\rho'}$. Let $R_m^{\delta \cdot m}$ denote the uniform distribution on restrictions on m variables assigning exactly δm variables to $*$, and assigning random values to the others.

A *decision tree* on m variables is a labelled binary tree where edges and leaves are labelled with 0 or 1, and internal nodes with variables. A decision tree computes a function in the intuitive way, starting at the root and following the path according to the values of the input variables, and outputting the value at the reached leaf.

Lemma 5 ([Bea]). *Let φ be a DNF or a CNF formula in m variables with bottom fan-in at most r. For every $s \geq 0, p < 1/7$, the probability over $\rho \in R_m^{p \cdot m}$ that the function computed by $\varphi|_\rho$ is not computable by a decision tree of height strictly less than s is less than $(7pr)^s$.*

We will use Lemma 5 in combination with the following fact.

Fact 6. *Let f be a symmetric function of S decision trees of height h. Then f is computable by a depth 2 circuit of size $S \cdot 2^h + 1$ with a symmetric gate of unbounded fan-in at the top and AND gates of fan-in h at the bottom.*

Proof. Write each decision tree as a DNF with bottom fan-in h, where each term corresponds to a path leading to 1. The number of terms in each DNF is at most 2^h, i.e. at most the number of paths in a decision tree of height h. Because every input to a decision tree follows a unique path, each DNF we construct has the property that every input satisfies at most one term. Thus we can merge the top OR gate of all these DNFs with the top symmetric gate of the circuit. Specifically, if the original symmetric gate was $\psi(x_1, x_2, \ldots, x_S) = g(\sum_{i \leq S} x_i)$ for some arbitrary function $g : [S] \to \{0, 1\}$, the new symmetric gate is simply $\psi'(x_1, x_2, \ldots, x_{S \cdot 2^h}) := g(\sum_{i \leq S \cdot 2^h} x_i)$. □

4.2 Multiparty Communication Complexity

In this section we describe some results on communication complexity that will be used in the proof of our main results. The model of interest is the *multiparty communication complexity model.* In this model there are s parties, each having unlimited computational power, who wish to collaboratively compute a certain function. The input bits to the function are partitioned in s blocks, and the i-th party knows all the input bits except those corresponding to the i-th block in the partition. The communication between the parties is by "writing on a blackboard" (broadcast): any bit sent by any party is seen by all the others. The parties exchange messages according to a fixed protocol. The measure of interest is the number of bits exchanged by the parties. We refer the reader to the book by Kushilevitz and Nisan [KN] for background on this model.

Babi, Nisan and Szegedy [BNS] prove a multiparty communication complexity lower bound for the *generalized inner product* function $GIP_{n,s} : \{0,1\}^{n \cdot s} \rightarrow \{0,1\}$, which is defined as follows:

$$GIP_{n,s}(x) := \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{s} x_{i,j}.$$

Lemma 7 ([BNS]). *There is a partition of the inputs to $GIP_{n,s}$ in s blocks such that the following holds: Let P be a s-party communication complexity protocol exchanging at most $.1 \cdot (n/4^s - \log(1/\gamma))$ bits of communication, then*

$$\Pr_{x \in \{0,1\}^{n \cdot (.3 \log n)}} \Big[P(x) \neq GIP_{n,.3\log n}(x) \Big] \geq 1/2 - \gamma.$$

Håstad and Goldmann [HG] show that the function computed by a 'small' SYM $\circ$ AND circuit with 'small' bottom fan-in can be computed by a multiparty communication complexity protocol among 'few' parties exchanging 'few' bits.

Lemma 8 ([HG]). *Let C be a depth-2 circuit of size S with an arbitrary symmetric gate (of unbounded fan-in) at the top, and AND gates of fan-in strictly less than s at the bottom. Then the function computed by C can be computed (under any partition of the input) by a s-party communication complexity protocol exchanging $1 + s \log S$ bits.*

The idea in Lemma 8 is that since each bottom AND gate has fan-in strictly less than s then, for any partition of the input in s blocks, the input bits to each AND can lie in at most $s-1$ distinct blocks. Therefore we can assign each AND gate to some party that knows all the input bits necessary to compute it. Now each party broadcasts the number of AND gates assigned to him that evaluate to 1, which takes at most $\log S$ bits. Since the top gate is symmetric this information is sufficient to compute the output of the circuit.

Our next lemma combines the above observation by Håstad and Goldmann with the "switching lemma" results from the previous section to argue the following: for every small SYM $\circ$ AC^0 circuit, w.h.p. over a suitable restriction ρ, the function computed by $C|_\rho$ can be computed by a multiparty communication complexity protocol among 'few' parties exchanging 'few' bits.

Lemma 9. *For every constant d there is a constant $\epsilon > 0$ such that the following holds. Let $C : \{0,1\}^n \rightarrow \{0,1\}$ be a circuit of size $n^{\epsilon \cdot \log n}$, depth d, with 1 arbitrary symmetric gate at the top. Let ρ be a random restriction on the n input variables that assigns * to a subset of the variables of relative size $1/n^{.1}$, i.e. let $\rho \in R_n^{n/n^{.1}}$. Then with probability at least $1 - n^{-\Omega(\log n)}$ over ρ, the function computed by $C|_\rho$ is computable (under any partition of the input) by a $.3 \log n$-party communication complexity protocol exchanging $\log^3 n$ bits of communication.*

Proof. The proof amounts to a combination of the previous lemmas for some specific setting of parameters.

Claim 10. *With probability* $1 - n^{-\Omega(\log n)}$ *over* $\rho \in R_n^{n/n^{.1}}$, *the function computed by* $C|_\rho$ *is computable by a depth-2 circuit of size* $|C| \cdot 2^{.3 \log n}$ *with a symmetric gate (of unbounded fan-in) at the top and AND gates of fan-in strictly less than* $.3 \log n$ *at the bottom.*

The lemma follows by the above claim using Lemma 8, which implies that the function computed by a depth-2 circuit of size $S = |C| \cdot 2^{.3 \log n} \leq n^{\log n}$ with a symmetric gate (of unbounded fan-in) at the top and AND gates of fan-in strictly less than $.3 \log n$ at the bottom is computable by a $.3 \log n$-party communication complexity protocol exchanging $1 + (.3 \log n) \log S \leq \log^3 n$ bits.

We now prove Claim 10. Similar calculations have already been done elsewhere (e.g., Lemma 2 in [LMN]). However, we have not found the exact claim we need in the literature.

Proof of Claim 10. We see the restriction ρ as $d-1$ successive applications of restrictions $\rho_1, \rho_2, \ldots, \rho_{d-1}$ each mapping to * a subset of variables of relative size $1/n^\alpha$ of the (remaining) variables. Taking $\alpha = .1/(d-1)$ we have that, after applying all $d-1$ restrictions, the total number of variables mapped to * is $n \cdot (1/n^\alpha)^{d-1} = n/n^{.1}$, and so this distribution on restrictions is exactly $R_n^{n/n^{.1}}$.

For every $i \in [d-1]$ let DT_i be the event that, after applying the first i restrictions $\rho_1, \rho_2, \ldots, \rho_i$, the function computed by every gate at level i is computable by a decision tree of height strictly less than $.3 \log n$. We now bound $\Pr_\rho[\text{ not } DT_{d-1}]$. Note that it is at most

$$\Pr_{\rho_1}[\text{ not } DT_1] + \Pr_{\rho_1,\rho_2}[\text{ not } DT_2 | DT_1] + \ldots + \Pr_{\rho_1,\rho_2,\ldots,\rho_{d-1}}[\text{ not } DT_{d-1} | DT_{d-2}].$$

We now bound each term. Fix any $i \leq d-1$ and consider $\Pr_{\rho_1,\rho_2,\ldots,\rho_i}[\text{ not } DT_i | DT_{i-1}]$ (if $i = 1$, think of the input variables as functions computed by decision trees of depth 1, and define $DT_0 := \text{TRUE}$). Fix any gate φ at level i. Without loss of generality assume φ is an OR gate (otherwise we can consider its negation, apply the same reasoning, and then negate again). Since we are conditioning over DT_{i-1}, all the functions computed by gates at level $i-1$ can be computed by decision trees of height (strictly) less than $.3 \log n$. Write each such function as a DNF with terms of size at most $.3 \log n$ (where each term corresponds to a path in the decision tree leading to '1'). Merging the top OR gates of all these DNFs with φ we see that, given DT_{i-1}, the function computed by φ is a DNF with terms of size at most $r = .3 \log n$. By Lemma 5 the probability over the choice of the i-th restriction ρ_i that the function computed by $\varphi|_{\rho_1 \rho_2 \cdots \rho_i}$ cannot be computed by a decision tree of depth strictly less than $s = .3 \log n$ is at most

$$(7pr)^s = (7 \cdot (1/n^\alpha) \cdot (.3 \log n))^{.3 \log n} = n^{-\Omega(\log n)}.$$

Thus by a union bound we have that

$$\Pr_{\rho_1,\rho_2,\ldots,\rho_i}[\text{ not } DT_i | DT_{i-1}]$$

is at most $n^{-\Omega(\log n)}$ times the number of gates at level i. Therefore, if the circuit C has size $n^{\epsilon \log n}$ for sufficiently small ϵ we have

$$\Pr_\rho[\text{ not } DT_{d-1}] \leq n^{-\Omega(\log n)} \cdot |C| = n^{-\Omega(\log n)}.$$

We have shown that with probability $1 - n^{-\Omega(\log n)}$ (over ρ) the function computed by $C|_\rho$ is computable by a symmetric function of $|C|$ decision trees of height strictly less than $.3 \log n$. By Fact 6 we can write each decision tree as a DNF and merge the top OR gates of these DNFs into the top symmetric gate of C, thus proving the claim. $\square$

$\square$

4.3 Proof of Theorem 4

We now prove Theorem 4. We restate the theorem for the reader's convenience.

Theorem (4, restated). *There is a function $f : \{0,1\}^* \to \{0,1\}$ computable in polynomial time such that for every constant d there is a constant $\epsilon > 0$ such that for every n and every circuit C of size $n^{\epsilon \cdot \log n}$, depth d, with 1 arbitrary symmetric gate at the top, the following holds:*

$$\Pr_{x \in \{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\epsilon \cdot \log n}.$$

Proof of Theorem 4. Similarly to [RW], we consider the function obtained by attaching PARITY gates on n bits at the bottom of $GIP_{n,.3\log n}$. That is, let $f_n : \{0,1\}^{n^2(.3\log n)} \to \{0,1\}$ be defined as

$$f_n(x) := \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{.3\log n} \bigoplus_{k=1}^{n} x_{i,j,k}.$$

We will prove Theorem 3 with f_n as hard function. While f_n is a function on $m = m(n) := n^2(.3\log n)$ bits, it will be convenient to parameterize it by n. Since we will prove $n^{\Omega(\log n)}$ lower bounds for f_n and the input length of f_n is $m = \text{poly}(n)$, we also obtain $m^{\Omega(\log m)}$ lower bounds for f_n (for a different hidden constant in the $\Omega(\cdot)$).

It is easy to see that f_n is computable in polynomial time.

Let $C : \{0,1\}^m \to \{0,1\}$ be a circuit as in the statement of Theorem 4, for a sufficiently small constant ϵ. Let ρ be a random restriction on the m input variables that assigns * to a subset of the variables of relative size $1/m^{.1}$, i.e. let $\rho \in R_m^{m/m^{.1}}$.

Consider the following two events.

- Event $E_1 :=$ the function computed by $C|_\rho$ is computable (under any partition of the input) by a $.3\log n$-party communication complexity protocol exchanging $n^{.2}$ bits.
- Event $E_2 :=$ for every $i \in [n], j \in [.3\log n]$ there is $k \in [n]$ such that $\rho(x_{i,j,k}) = *$. (In other words, for each of the $n \cdot (.3\log n)$ bottom parity functions of f_n, ρ maps some of its input variable to *.)

Claim 11. $\Pr_{\rho \in R_m^{m/m^{.1}}}[E_1 \wedge E_2] \geq 1 - n^{-\Omega(\log n)}$.

Before proving Claim 11 let us see how we can use it to prove Theorem 4. Suppose that some $\rho \in R_m^{m/m^{.1}}$ satisfies both E_1 and E_2. Then

$$\Pr_{y \in \{0,1\}^{m/m^{.1}}}[C|_\rho(y) \neq f_n|_\rho(y)] \geq 1/2 - n^{-\Omega(\log n)}. \quad (1)$$

This holds by Lemma 7. Specifically, fix any restriction ρ' taken on the variables mapped to * by ρ, such that for every $i \in [n], j \in [.3 \log n]$ there is *exactly one* $k \in [n]$ such that $\rho\rho'(x_{i,j,k}) = *$. We then have that $f_n|_{\rho\rho'}$ equals $GIP_{n,.3 \log n}$ (up to possibly negating some input variables). If the function computed by $C|_\rho$ is computable by a s-party communication complexity protocol exchanging $n^{.2}$ bits then clearly the same holds for the function computed by $C|_{\rho\rho'}$. Therefore by the multiparty communication complexity lower bound for *GIP* (Lemma 7) we obtain (noticing that for $s = .3 \log n, \gamma = 2^{-n^{.3}}$ we have $.1 \cdot (n/4^s - \log(1/\gamma)) = \Omega(n^{.4} - n^{.3}) > n^{.2}$):

$$\Pr_{z \in \{0,1\}^{n(.3 \log n)}}[C|_{\rho\rho'}(z) \neq f_n|_{\rho\rho'}(y)] \geq 1/2 - 1/2^{n^{\Omega(1)}} \geq 1/2 - n^{-\Omega(\log n)}.$$

Equation 1 follows noticing that we can think of a random y as choosing first a random ρ' as above and then a random $z \in \{0,1\}^{n(.3 \log n)}$ for the *'s of ρ' (so that $C|_\rho(y) = C|_{\rho\rho'}(z)$).

Thus we have:

$$\begin{aligned}
&\Pr_x[C(x) \neq f_n(x)] \\
&= \Pr_{\rho \in R_m^{m/m^{.1}}, y \in \{0,1\}^{m/m^{.1}}}[C|_\rho(y) \neq f_n|_\rho(y)] \\
&\geq \Pr_{\rho \in R_m^{m/m^{.1}}, y \in \{0,1\}^{m/m^{.1}}}[C|_\rho(x) \neq f_n|_\rho(x) | E_1 \wedge E_2] \cdot \Pr[E_1 \wedge E_2] \\
&\geq \left(1/2 - n^{-\Omega(\log n)}\right) \cdot \left(1 - n^{-\Omega(\log n)}\right) \qquad \text{(by Equation 1 and Claim 11)} \\
&= 1/2 - n^{-\Omega(\log n)},
\end{aligned}$$

which proves Theorem 3.

It is only left to prove Claim 11.

Proof of Claim 11. We show that E_1 and E_2 each do not happen with probability at most $n^{-\Omega(\log n)}$.

The bound on $\Pr_\rho[\text{not } E_1]$ is given by Lemma 9. (The direct application of Lemma 9 gives communication complexity $\text{poly} \log(n) \ll n^{.2}$ for circuits of size $m^{\epsilon \log m} \geq n^{\epsilon \log n}$).

We now bound $\Pr_\rho[\text{not } E_2]$. Fix $i \in [n], j \in [.3 \log n]$. The probability that for every $k \in [n]$ we have $\rho(x_{i,j,k}) \neq *$ is the probability that a random subset $A \subseteq [m]$ of size $m/m^{.1} = m^{.9}$ does not intersect a fixed subset $B \subseteq [m]$ of size n. This probability is at most the probability that $m^{.9}$ *independent* random elements uniformly distributed in $[m]$ all fall outside B (to see this, think of choosing the random subset A one element at the time, and note that when an element falls outside B it is more likely for the next element to fall inside B). This latter probability is

$$\left(1 - \frac{n}{m}\right)^{m^{.9}} \leq exp(-m^{.9}n/m) \leq exp(-m^{\Omega(1)}) \ll n^{-\Omega(\log n)}$$

where we used that $m = n^2 \cdot (.3 \log n)$. By a union bound we have

$$\Pr[\text{not } E_2] \leq n \cdot (.3 \log n) \cdot n^{-\Omega(\log n)} = n^{-\Omega(\log n)}.$$

□

$\square$

We point out that Theorem 4 is tight for the particular choice of

$$f_n(x) = \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{.3\log n} \bigoplus_{k=1}^{n} x_{i,j,k}.$$

Namely, f_n is computable by PARITY $\circ$ AND circuits of size $n^{O(\log n)}$. This can be seen by writing the function computed by each AND as a PARITY of $n^{O(\log n)}$ AND's (cf. [RW]).

5 Fooling Circuits with more Arbitrary Symmetric Gates

In this section we prove our average-case hardness result for constant-depth circuits of size $n^{\epsilon \log n}$ with $\epsilon \log^2 n$ arbitrary symmetric gates (Theorem 3). The proof has the same structure as the proof of our average-case hardness result for circuits with *one* arbitrary symmetric gate (Theorem 4). The only difference is that now we want to argue that event E_1 happens with high probability even for circuits with $\epsilon \log^2 n$ arbitrary symmetric gates, i.e. we want to show that with high probability over the restriction ρ, the function computed by $C|_\rho$ is computable by a multiparty communication complexity protocol among 'few' parties exchanging 'few' bits. Thus the proof of Theorem 3 follows from the next lemma.

Lemma 12. *For every constant d there is a constant $\epsilon > 0$ such that the following holds. Let $C : \{0,1\}^n \to \{0,1\}$ be a circuit of size $n^{\epsilon \cdot \log n}$, depth d, with $\epsilon \log^2 n$ arbitrary symmetric gates. Let ρ be a random restriction on the n input variables that assigns * to a subset of the variables of relative size $1/n^{.1}$, i.e. let $\rho \in R_n^{n/n^{.1}}$. Then with probability at least $1 - n^{-\Omega(\log n)}$ over ρ, the function computed by $C|_\rho$ is computable (under any partition of the input) by a $.3\log n$-party communication complexity protocol exchanging $\log^5 n$ bits of communication.*

Proof. Assume without loss of generality that the output gate of the circuit C is included in the arbitrary symmetric gates. Fix a topological order of the arbitrary symmetric gates (the simple order induced by reading the gates level by level from the inputs to the output node will do). For every $i \in \{1, \dots, \epsilon \log^2 n\}, z \in \{0,1\}^{i-1}$, define $C_{i,z}$ as the subcircuit of C whose output gate is the i-th arbitrary symmetric gate but where the previous arbitrary symmetric gates are replaced with z (i.e., the j-th gate is replaced with the j-th bit in z). Note $C_{i,z}$ is a SYM $\circ$ AC^0 circuit.

Claim 13. *For a sufficiently small constant $\epsilon > 0$, with probability $1 - n^{-\Omega(\log n)}$ over $\rho \in R_n^{n/n^{.1}}$ we have that for every $i \in \{1, \dots, \epsilon \log^2 n\}$ and $z \in \{0,1\}^{i-1}$ the function computed by $C_{i,z}|_\rho$ is computable (under any partition of the input) by a $.3\log n$-party communication complexity protocol $P_{i,z}$ exchanging $\log^3 n$ bits of communication.*

Proof. The claim follows by noting that the number of SYM $\circ$ AC^0 circuits $C_{i,z}$ is at most

$$\left(\epsilon \log^2 n\right) \cdot 2^{\epsilon \log^2 n} \leq n^{1+\epsilon \log n}$$

and then using a union bound and Lemma 9, which states that for each fixed circuit $C_{i,z}$, with probability $1 - n^{-\Omega(\log n)}$ over ρ, the function computed by $C_{i,z}|_\rho$ is computable by a $.3\log n$-party communication complexity protocol exchanging $\log^3 n$ bits. $\square$

The lemma follows by noting that whenever ρ satisfies the conclusion of the above claim we have (under any partition of the input bits) the following $.3 \log n$-party communication complexity protocol P for $C|_\rho$: On input x compute $C|_\rho(x)$ as follows. Simulate P_1 to compute $b_1 = C_1|_\rho(x)$. Then simulate P_{2,b_1} to compute $b_2 = C_{2,b_1}|_\rho(x)$. Then simulate $P_{3,b_1 \circ b_2}$ to compute $b_3 = C_{3,b_1 \circ b_2}|_\rho(x)$. Continue in this way until $C_{\epsilon \log^2 n, z}(x) = C|_\rho(x)$ (this last equality is easy to verify).

Since each protocol $P_{i,z}$ exchanges at most $\log^3 n$ bits of communication, and we simulate $\epsilon \log^2 n$ of these protocols, the total number of bits exchanged by the protocol P is at most $\log^5 n$. $\square$

It is perhaps interesting to note that, unlike the corresponding protocol in the proof of Theorem 4, the protocol in the above lemma is not simultaneous, i.e. the bits sent by a party in general depend on the bits previously sent by other parties (cf. [KN] for background on simultaneous protocols). Thus in our proof we are taking advantage of the fact that the lower bound for *GIP* (Lemma 7) holds even for non-simultaneous protocols. We do not know how to prove the same result starting from a multiparty communication complexity lower bound for simultaneous protocols.

6 Our Generator vs. Luby, Velickovic and Wigderson's

In this section we elaborate on why our generator (Theorem 1) improves on the generator by Luby, Velickovic and Wigderson (Theorem 2 in [LVW]). Recall that the generator in [LVW] fools 'small' depth 2 circuits with one arbitrary symmetric gate at the top (SYM $\circ$ AND circuits). On the other hand our generator fools 'small' circuits of any constant depth with 'few' arbitrary symmetric gates.

We note that there are several results (e.g. [Raz, Smo, All, Yao, BRS, BT]) showing that 'small' circuits in certain 'rich' constant-depth circuit classes can be converted into 'not-too-big' SYM $\circ$ AND circuits. Thus one may wonder whether we can use these results to deduce that the generator in [LVW] is already powerful enough to give our main result (Theorem 1), i.e. whether it can fool 'small' constant-depth circuits with 'few' arbitrary symmetric gates.

The problem with this idea is that *in all these conversion results the blow-up in the circuit size is bigger than the saving of the generator.* More specifically, these conversion results show how to convert, say, a AC^0 circuit of size S into a SYM $\circ$ AND circuit of size *quasi*-polynomial, i.e. $S^{\log^{O(1)} S}$, where the constant in the $O(1)$ depends on the depth of the original circuit. However, to fool a circuit of size $S^{\log^{O(1)} S}$, the generator in [LVW] needs a seed of length at least S, and therefore it is of no use in this particular setting.

It seems natural to ask whether the known conversion results are the best possible, i.e. if the quasi-polynomial blow-up is inherent in the conversion. There are works (e.g. [BRS, RW]) suggesting that this is indeed the case. We give another result of this flavor.

Specifically, we show how to modify the lower bound in Theorem 4 to get a function computable by polynomial size PARITY $\circ$ AC^0 circuits that is average-case hard for super-polynomial size SYM $\circ$ AND circuits. The idea is to change the fan-in of the bottom parities of f so that they are computable by polynomial size AC^0 circuits (specifically we change their fan-in from n to $\log^3 n$). While our lower bound is only 'slightly' superpolynomial (i.e. $n^{\Omega(\log\log n)}$), it shows that the parameters of our generator (Theorem 1) *cannot* be obtained combining a conversion result with Theorem 2 in [LVW], even if we only want to fool PARITY $\circ$ AC^0 circuits.

Theorem 14. *There is a function $f : \{0,1\}^* \to \{0,1\}$ computable by uniform polynomial size* PARITY $\circ$ AC^0 *circuits and a constant $\epsilon > 0$ such that for every n and every* SYM $\circ$ AND *circuit C of size $n^{\epsilon \cdot \log\log n}$, the following holds:*

$$\Pr_{x\in\{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\epsilon\cdot\log\log n}.$$

The proof of Theorem 14 is given in Appendix A.

7 Open Problems

Can the techniques in this paper be used to prove (average-case) hardness results for constant-depth circuits with $\omega(\log^2 n)$ arbitrary symmetric gates? Such a hardness result would follow from a positive answer to the following open question: Let C be constant-depth circuit of size $n^{\epsilon \log n}$ with $\omega(\log^2 n)$ arbitrary symmetric gates, and let ρ be a restriction as in the statement of Lemma 9. Is it true that with high probability over ρ the function computed by $C|_\rho$ is computable by a $.9\log n$-party communication complexity protocol exchanging $n^{.9}$ bits?

8 Acknowledgments

We thank Salil Vadhan for his helpful reading of this paper. We thank Chattopadhyay and Hansen for sending us their paper [CH], and the anonymous referees for helpful comments.

References

[Agr] M. Agrawal. Hard Sets and Pseudo-random Generators for Constant Depth Circuits. In *Twenty First Foundations of Software Technology and Theoretical Computer Science, December 13-15, Bangalore, India*, pages 58–69. Springer-Verlag, 2001.

[All] E. Allender. A Note on the Power of Threshold Circuits. In *30th Annual Symposium on Foundations of Computer Science*, pages 580–584, Research Triangle Park, North Carolina, 30 Oct.–1 Nov. 1989. IEEE.

[BFNW] L. Babai, L. Fortnow, N. Nisan, and A. Wigderson. BPP Has Subexponential Time Simulations Unless EXPTIME has Publishable Proofs. *Computational Complexity*, 3(4):307–318, 1993.

[BNS] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. Twenty-first Symposium on the Theory of Computing (Seattle, WA, 1989).

[Bea] P. Beame. A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, November 1994. Available from `http://www.cs.washington.edu/homes/beame/`.

[Bei] R. Beigel. When do extra majority gates help? polylog(N) majority gates are equivalent to one. *Comput. Complexity*, 4(4):314–324, 1994. Special issue devoted to the 4th Annual McGill Workshop on Complexity Theory.

[BRS] R. Beigel, N. Reingold, and D. A. Spielman. The Perceptron Strikes Back. In *Structure in Complexity Theory Conference*, pages 286–291, 1991.

[BT] R. Beigel and J. Tarui. On ACC. *Comput. Complexity*, 4(4):350–3–66, 1994. Special issue devoted to the 4th Annual McGill Workshop on Complexity Theory. Preliminary version in FOCS '91.

[CH] A. Chattopadhyay and K. A. Hansen. Lower Bounds for Circuits With Few Modular and Symmetric Gates. Manuscript, 2005.

[FSS] M. L. Furst, J. B. Saxe, and M. Sipser. Parity, Circuits, and the Polynomial-Time Hierarchy. *Mathematical Systems Theory*, 17(1):13–27, April 1984.

[HMP$^+$] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993.

[HM] K. A. Hansen and P. B. Miltersen. Some Meet-in-the-Middle Circuit Lower Bounds. In *Proceedings of the 29th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, Lecture Notes in Computer Science, Volume 3153, pages 334 – 345, August 22–27 2004.

[Hås] J. Håstad. *Computational limitations of small-depth circuits.* MIT Press, 1987.

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

[HVV] A. Healy, S. Vadhan, and E. Viola. Using nondeterminism to amplify hardness. In *Proceedings of the Thirty-Six Annual ACM Symposium on the Theory of Computing*, pages 192–201, Chicago, IL, 13–15 June 2004. Invited to *SIAM Journal of Computing*, STOC Special Issue.

[KN] E. Kushilevitz and N. Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[LMN] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *J. Assoc. Comput. Mach.*, 40(3):607–620, 1993.

[LVW] M. Luby, B. Velickovic, and A. Wigderson. Deterministic Approximate Counting of Depth-2 Circuits. In *In Proceedings of the 2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993.

[Nis] N. Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[NW] N. Nisan and A. Wigderson. Hardness vs Randomness. *J. Computer & Systems Sciences*, 49(2):149–167, Oct. 1994.

[RW] A. Razborov and A. Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993.

[Raz] A. A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987.

[RR] A. A. Razborov and S. Rudich. Natural Proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, Aug. 1997.

[SU] R. Shaltiel and C. Umans. Simple Extractors for All Min-Entropies and a New Pseudo-Random Generator. In *42nd Annual Symposium on Foundations of Computer Science*. IEEE, 14–17 Oct. 2001.

[Smo] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 77–82, New York City, 25–27 May 1987.

[STV] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. System Sci.*, 62(2):236–266, 2001. Special issue on the Fourteenth Annual IEEE Conference on Computational Complexity (Atlanta, GA, 1999).

[Uma] C. Umans. Pseudo-random generators for all hardnesses. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 627–634. ACM Press, 2002.

[Vio] E. Viola. The Complexity of Constructing Pseudorandom Generators from Hard Functions. *Comput. Complexity*, 13(3-4):147–188, 2004.

[Yao] A. C. Yao. On ACC and threshold circuits. In *In Proc. 31st Ann. IEEE Symp. Found. Comput. Sci.*, pages 619–627, 1990.

A Proof of Theorem 14

Proof of Theorem 14. The proof follows closely the proof of Theorem 4. Let

$$g_n : \{0,1\}^{n(.3\log n)\log^3 n} \to \{0,1\}$$

be defined as

$$g_n(x) := \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{.3\log n} \bigoplus_{k=1}^{\log^3 n} x_{i,j,k}.$$

We will prove Theorem 14 with g_n as hard function. While g_n is a function on $m = m(n) := n \cdot (.3 \log n) \cdot (\log^3 n)$ bits, it will be convenient to parameterize it by n. Since we will prove $n^{\Omega(\log \log n)}$ lower bounds for g_n and the input length of g_n is $m = n \cdot \mathrm{poly} \log n$, we also obtain $m^{\Omega(\log \log m)}$ lower bounds for g_n (for a different hidden constant in the $\Omega(\cdot)$).

Note that g_n is computable by a (uniform) polynomial size circuit of depth 5 with one PARITY gate at the top. To see this note that each of the bottom parities in the definition of g_n is only on $\log^3 n$ bits, and therefore it can be computed by a (uniform) circuit of size $\mathrm{poly}(n)$ and depth 4 (see e.g. [Hås], Theorem 2.2).

Let $C : \{0,1\}^m \to \{0,1\}$ be a circuit as in the statement of Theorem 14, for a sufficiently small constant ϵ. Let ρ be a random restriction on the m input variables that assigns * to a subset of the variables of relative size $1/\log(n)$, i.e. let $\rho \in R_m^{m/\log(n)}$.

Consider the following two events.

- Event $E_1' :=$ the function computed by $C|_\rho$ is computable (under any partition of the input) by a $.3 \log n$-party communication complexity protocol exchanging $n^{.2}$ bits.
- Event $E_2' :=$ for every $i \in [n], j \in [.3 \log n]$ there is $k \in [\log^3 n]$ such that $\rho(x_{i,j,k}) = *$. (In other words, for each of the $n \cdot (.3 \log n)$ bottom parity functions of f_n, ρ maps some of its input variable to *.)

As before, Theorem 14 follows from the next claim (cf. the proof of Theorem 4).

Claim 15. $\Pr_{\rho \in R_m^{m/\log(n)}}[E_1' \wedge E_2'] \geq 1 - n^{-\Omega(\log \log n)}$.

Proof. We show that E_1' and E_2' each do not happen with probability at most $n^{-\Omega(\log \log n)}$.

We now bound $\Pr_\rho[\text{not } E_1']$. Analogously to the proof of Lemma 9, the main step is proving the following claim: with high probability $\left(1 - n^{-\Omega(\log \log n)}\right)$ over $\rho \in R_m^{m/\log(n)}$, the function computed by $C|_\rho$ is computable by a depth 2 circuit of size $|C| \cdot 2^{.3 \log n} = n^{\epsilon \cdot \log \log n} \cdot 2^{.3 \log n}$ with a single symmetric gate (of unbounded fan-in) at the top and AND gates of fan-in strictly less than $.3 \log n$ at the bottom.

While this probability can be bound directly, similarly to what is done in [RW], it seems simpler to use again the Switching Lemma. Fix a bottom AND gate φ of C, and think of the input variables to φ as clauses of size $r = 1$. By Lemma 5 the probability over the choice of ρ that the function computed by $\varphi|_\rho$ cannot be computed by a decision tree of depth strictly less than $s = .3 \log n$ is at most

$$(7pr)^s = (7 \cdot (1/\log(n)) \cdot 1)^{.3 \log n} = n^{-\Omega(\log \log n)}.$$

Therefore if the circuit C has size $n^{\epsilon \log \log n}$ for sufficiently small ϵ we have, by a union bound, that with probability $1 - n^{-\Omega(\log \log n)}$ (over ρ) the function computed by $C|_\rho$ is computable by a symmetric function of $|C|$ decision trees of height strictly less than $.3 \log n$. By Fact 6 we can write each decision tree as a DNF and merge the top OR gates of these DNFs into the top symmetric gate of C, and thus E_1' holds.

We now bound $\Pr_\rho[\text{not } E_2']$. Fix $i \in [n], j \in [.3 \log n]$. The probability that for every $k \in [\log^3 n]$ we have $\rho(x_{i,j,k}) \neq *$ is is the probability that a random subset $A \subseteq [m]$ of size $m/\log(n)$ does not intersect a fixed subset $B \subseteq [m]$ of size $\log^3 n$. This probability is at most the probability that $m/\log(n)$ *independent* random elements uniformly distributed in $[m]$ all fall outside B (to see this, think of choosing the random subset A one element at the time, and note that when an element falls outside B it is more likely for the next element to fall inside B). This latter probability is

$$\left(1 - \frac{\log^3 n}{m}\right)^{m/\log(n)} \leq exp(-\Omega(\log^2 n)) \ll n^{-\Omega(\log\log n)}.$$

By a union bound we have

$$\Pr[\text{not } E_2'] \leq n \cdot (.3 \log n) \cdot n^{-\Omega(\log\log n)} = n^{-\Omega(\log\log n)}.$$

$\square$

$\square$