Some Meet-in-the-Middle Circuit Lower Bounds

Kristoffer Arnsfelt Hansen and Peter Bro Miltersen

Department of Computer Science, University of Aarhus, Denmark
{arnsfelt,bromille}@daimi.au.dk

Abstract. We observe that a combination of known top-down and bottom-up lower bound techniques of circuit complexity may yield new circuit lower bounds.

An important example is this: Razborov and Wigderson showed that a certain function f in $\mathbf{ACC}^0$ cannot be computed by polynomial size circuits consisting of two layers of MAJORITY gates at the top and a layer of AND gates at the bottom. We observe that a simple combination of their result with the Håstad switching lemma yields the following seemingly much stronger result: The same function f cannot be computed by polynomial size circuits consisting of two layers of MAJORITY gates at the top and an arbitrary $\mathbf{AC}^0$ circuit feeding the MAJORITY gates.

1 Introduction

During the 1980's and 1990's significant progress was made understanding the power of *bounded depth* computation, i.e. computation performed by unbounded fan-in circuits of constant depth and polynomial size containing layers of gates of various kinds. Such research was motivated by its fundamental nature. Indeed, it can be argued that understanding such restricted computation must be a prerequisite to solving problems such as **P** vs. **NP**.

Two methodologies for proving lower bounds on the power of constant depth computation were developed in the 1980's:

1. *top-down* methods, exemplified by lower bounds obtained using two- or multi-party communication complexity [1, 2] and
2. *bottom-up* methods, exemplified by lower bounds obtained using the Håstad switching lemma [3–9].

Roughly speaking, a top-down argument proves that a problem cannot be solved by a certain kind of circuit by assuming to the contrary that it can, looking at the output gate of the circuit and working its way downwards to an input wire where a contradiction is reached. In contrast, a bottom-up argument starts at the input and successively eliminates gates at the bottom of the circuit, until the circuit is gone and a contradiction is reached.

Progress on the understanding of bounded depth computation in the last decade has been less encouraging than the rate of progress in the 1980s. One reason for this is indicated by the Razborov-Rudich *natural proof* concept [10] that shows, under certain cryptographic assumptions, that the kind of arguments we are currently using will not be strong enough to, say, separate $\mathbf{TC}^0$ from classes above it.

Still, there seems to be plenty of possible results not ruled out by the natural-proof obstacle. In this paper, we show a surprisingly simple way of *combining* top-down and bottom-up arguments yielding some new circuit lower bounds.

The general principle of the technique we describe is as follows: Suppose we start out with a theorem (typically obtained using top-down methods) that a certain boolean function f cannot be computed by a small circuit of type $\mathcal{C} \circ \text{AND}$, i.e., a layered Boolean circuits of which the bottom layer consists of AND-gates feeding into a circuit of some class $\mathcal{C}$. Furthermore assume that the bottom layer of each circuit in $\mathcal{C}$ consists of gates computing *symmetric* functions (say, MAJORITY). Further, assume that we have a function f' and that hitting f' with random restrictions of certain parameters results in a function of which f is a restriction (for instance, f' and f could both be the parity function). Then, using the Beame/Håstad switching lemma, we may obtain a new result stating that f' cannot be computed by $\mathcal{C} \circ \mathbf{AC}^0$ circuits, i.e., circuits consisting a multi-output $\mathbf{AC}^0$ circuit feeding its results into a $\mathcal{C}$-circuit. Thus, the switching lemma allows us to replace the bottom AND gate in the original lower bound with an entire $\mathbf{AC}^0$ circuit. As the technique combines a top-down argument with a bottom-up argument (i.e., random restrictions and the switching lemma) we call the resulting bounds *meet-in-the-middle lower bounds.* This kind of argument has been used sporadically before: Zhang, Barrington and Tarui [11] characterized the symmetric functions computable by quasi-polynomial size perceptrons. Goldmann [5] proved that computing the MOD_2-function requires $\mathbf{Th} \circ \mathbf{AC}^0$ circuits of size $2^{n^{\Omega(1)}}$. Furthermore, Berg and Ulfberg [4] and also Krause and Pudlák [6] gave $\mathbf{\Sigma}_k$ functions which require $\mathbf{Th} \circ \mathbf{AC}^0_{k-2}$ circuits of size $2^{n^{\Omega(1)}}$. These proofs all used ideas along the lines of what we describe above. However, the approach does not seem to have been investigated systematically before. In this paper we use it to prove some, to the best of our knowledge, previously unstated lower bounds.

We find that the most interesting application of the technique in this paper is the following result: $\mathbf{ACC}^0$ cannot be computed by $\mathbf{MAJ} \circ \mathbf{MAJ} \circ \mathbf{AC}^0$ circuits, i.e., circuits consisting of two layers of majority gates above an $\mathbf{AC}^0$ circuit (for a full explanation of our notation for circuit classes, see Sect. 2.1). The previous best result along these lines and the result we base our proof upon is the result by Razborov and Wigderson [9], showing that $\mathbf{ACC}^0$ cannot be simulated by $\mathbf{MAJ} \circ \mathbf{MAJ} \circ \mathbf{AND}$ circuits, i.e., circuits consisting of two layers of majority gates above a layer of $\mathbf{AND}$ gates. As $\mathbf{MAJ}$ is symmetric, the meet-in-the-middle technique applies. More precisely, we prove the following theorem.

Theorem 1. *Any* $\mathbf{MAJ} \circ \mathbf{SYM} \circ \mathbf{AC}^0$ *circuit computing* $(\text{MOD}_2)_n \circ \text{AND}_{\log n} \circ (\text{MOD}_2)_n$ *has size* $n^{\Omega(\log n)}$.

Tsai [12] proved strong lower bounds on the degree of the MOD_l function viewed as a polynomial over $\mathbf{Z}_m$, with l having a prime factor not dividing m. As a polynomial over $\mathbf{Z}_m$ may be viewed as a $\mathbf{MOD}_m \circ \mathbf{AND}$ circuit, and MOD_m

is a symmetric function, the meet-in-the-middle technique applies. Using this, as well as a well known lower bound for the MAJORITY function, we can prove the following theorems.

Theorem 2. *Let m and l be positive integers. There is a constant c such that, if l has a prime factor not dividing m, then depth $d+1$ $\mathbf{MOD}_m \circ \mathbf{AC}^0_d$ circuits require more than size $2^{\frac{1}{14}(cn)^{\frac{1}{d}}}$ to compute either of* MOD_l *and* $\neg\mathrm{MOD}_l$, *for sufficiently large n, and if m is not a prime power, this size is also required to compute* $\neg\mathrm{MOD}_m$.

Theorem 3. *Let m be any integer. Then for any integer d, any $\epsilon > 0$ and $c < \frac{1}{2}$, depth $d+1$ $\mathbf{MOD}_m \circ \mathbf{AC}^0_d$ circuits require more than size $2^{\frac{1-\epsilon}{14}(cn)^{\frac{1}{d}}}$ to compute* MAJORITY *for sufficiently large n.*

Finally, we consider a slightly modified version of the Sipser functions f_k^m used by Håstad [8], computed by $\mathbf{\Pi}_k$ circuits of polynomial size but shown there to require exponential size $\mathbf{AC}^0_{k-1}$ circuits. We define the function f_k^m on $m^{k-1}\sqrt{\frac{k}{2}m\log m}$ variables by a $\mathbf{\Pi}_k$ formula, with all different unnegated variables at the leaves. The fanin on all levels is m except at the bottom level where the fanin is $\sqrt{\frac{k}{2}m\log m}$. We prove the following theorem.

Theorem 4. *For $k > 1$ and sufficiently large m, there are no depth $k+1$ $\mathbf{MOD}_m \circ \mathbf{AC}^0_k$ circuits computing f_k^m with bottom fanin $\frac{1}{12\sqrt{2k}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}$ and size $2^{\frac{1}{12\sqrt{2k}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}}$.*

Using a weaker notion of representation of boolean functions by polynomials and using a similar proof technique as well as extending an idea of Håstad ([8] Theorem 9.1), we can obtain the following lower bounds for $\mathbf{AC}^0$ circuits with a single MOD_m gate (allowed to appear anywhere in the circuit), thus generalizing Theorem 2.

Theorem 5. *Let m and l be positive integers. There is a constant c such that, if l has a prime factor not dividing m, then depth d $\mathbf{AC}^0_d$ circuits containing a single* MOD_m *gate, require more than size $2^{\frac{1}{28}(cn)^{\frac{1}{d}}}$ to compute either of* MOD_l *and* $\neg\mathrm{MOD}_l$ *for sufficiently large n.*

This is somewhat similar to a result by Aspnes et al [13] (later generalized by Beigel [14] and Barrington and Straubing [15]) concerning $\mathbf{AC}^0$ circuits with a single MAJORITY gate.

This paper is organized as follows: In Sect. 2 we make precise the notation we use for circuit classes and state the previous results we use as bottom-up and top-down ingredients in order to derive the theorems above. In Sect. 3, we derive our five main results stated above. Using the approach of Furst, Saxe and Sipser [7] one can use super-quasi-polynomial lower bounds for constant depth circuits to derive oracle separations of corresponding subclasses of **PSPACE**. In Sect. 4, we describe the oracle consequences of the three first exponential lower bounds above.

2 Preliminaries

2.1 Circuit Classes

The circuits we consider will be built from various families of unbounded fanin gates. The inputs are allowed to be both unnegated and negated variables as well as the constants 0 and 1. Let $x_1, \ldots, x_n$ be boolean inputs. For a positive integer m, the MOD_m function outputs 1 if and only if $\sum_{i=1}^{n} x_i \not\equiv 0 \pmod{m}$. The MAJORITY function outputs 1 if and only if $\sum_{i=1}^{n} x_i \geq \frac{n}{2}$. Given integer weights $w = (w_1, \ldots, w_n)$ and an integer threshold t, a threshold gate with weights w and threshold t outputs 1 if and only if $\sum_{i=1}^{n} w_i x_i \geq t$.

Let **AND** and **OR** denote the families of unbounded fanin AND and OR gates. Let $\mathbf{MOD}_m$, **MAJ** and **Th** denote the families of MOD_m, MAJORITY and threshold gates. Finally let **SYM** be the family of all symmetric gates.

If G is a family of boolean gates and $\mathcal{C}$ is a family of circuits we let $G \circ \mathcal{C}$ denote the class of circuits consisting of a G gate taking circuits from $\mathcal{C}$ as inputs. If we need to specify a specific bound on the fanin of some of the gates, this will be specified by a subscript.

$\mathbf{\Pi}_k$ is the class of functions computed by depth k circuits of alternating AND and OR gates with an AND gate as output. $\mathbf{\Sigma}_k$ is the analogous class with an OR gate as output instead. Let $\mathbf{AC}^0_k$ be the union of $\mathbf{\Pi}_k$ and $\mathbf{\Sigma}_k$ and $\mathbf{AC}^0$ be the union of $\mathbf{AC}^0_k$ for all k.

Theorem 5 consider $\mathbf{AC}^0$ circuits containing just a single MOD_m gate. Under this special restriction, our assumption that NOT gates occur only as negations on the input variables, cannot in general be assumed without loss of generality: the standard way of moving the negations to the bottom of the circuits may increase the number of MOD_m gates. The theorem can, however, be proved with NOT gates occurring anywhere in the circuits and we do *not* have to charge the NOT gates when defining the depth of the circuits.

Usually we would by our notation above implicitly consider the circuit classes defined to be of polynomial size. It will be convenient however, to use the same notation, but giving the size explicitly in the statement of our lower bounds.

2.2 Bottom-up Ingredients

A *restriction* on a set V of boolean variables is a map $\rho : V \to \{0, 1, \star\}$. It acts on a boolean function $f : V \to \{0, 1\}$, creating a new boolean function f_ρ on the set of variables for which $\rho(x) = \star$, obtained by substituting $\rho(x)$ for $x \in V$ whenever $\rho(x) \neq \star$. The variables x for which $\rho(x) = \star$ are called *free*; the other variables *set*.

Let R^l_n denote the set of all restriction ρ leaving l variables free. We will use the following version of Håstads Switching Lemma due to Beame [3].

Lemma 6. *Let f be a DNF formula in n variables with terms of length at most r. Let $l = pn$ and pick ρ uniformly at random from R^l_n. Then*

$$\Pr[f_\rho \text{ does not have a decision tree of depth } d] < (7pr)^d .$$

The main feature of Beame's switching lemma crucial to us is that it gives us a decision tree. A decision tree may be viewed as a DNF *where all terms are mutually contradictory.* This means that instead of viewing it as an OR of AND's, we can also view it as a sum of products, over any ring, representing false as 0 and true as 1. As this property is explicit in Beame's lemma, we shall us it whenever we can.

In order to prove Theorem 4, we will need a special kind of restrictions, however. The corresponding switching lemma is not stated in terms of decision trees, yet we will use it similarly obtaining the crucial property in a different way.

Let q be a real number and $(B_i)_{i=1}^r$ a partition of the variables. Let $R_{q,B}^+$ be the probability distribution on restrictions ρ given as follows. For every B_i, we let $s_i = \star$ independently with probability q and $s_i = 0$ otherwise. Then for every $x_k \in B_i$, we let $\rho(x_k) = s_i$ independently with probability q and $\rho(x_k) = 1$ otherwise.

Additionally we define, for $\rho \in R_{q,B}^+$ the restriction $g(\rho)$ as follows: For every B_i with $s_i = \star$, $g(\rho)$ assigns the value 1 to every variable except the variable with the highest index, which is assigned $\star$.

Similarly we define the distribution $R_{q,B}^-$, and for $\rho \in R_{q,B}^-$ also $g(\rho)$ by interchanging 0 and 1 in the definitions.

Lemma 7 ([8], Lemma 6.3). *Let f be a DNF formula with terms of length at most r. Pick ρ at random from $R_{q,B}^+$. Then*

$$\Pr[f_{\rho g(\rho)} \textit{ can not be written as a CNF with clauses of size at most } s] < (6qr)^s \quad .$$

The same statement holds for converting a CNF into a DNF, and both of these also holds for $\rho \in R_{q,B}^-$ instead.

Berg and Ulfberg showed, that by closely inspecting the proof of Lemma 7, one actually already obtains a DNF with small mutually contradictory terms, when applying a restriction to a CNF with small clauses [4].

Let in the following $q = \sqrt{\frac{2k \log m}{m}}$ and let B_i be the partition of the variables given by the inputs to the gates at the bottom of f_k^m. We will then have that applying a restriction from $R_{q,B}^+$ does not destroy f_k^m too much.

Lemma 8 ([8], Lemma 6.8). *For odd k and for random $\rho \in R_{q,B}^+$, then the circuit defining $(f_k^m)_{\rho g(\rho)}$ will contain the circuit defining f_{k-1}^m with probability at least $\frac{2}{3}$, when $\frac{m}{\log m} \geq 100k$ for $m > m_1$ where m_1 is some absolute constant.*

The same statement holds for even k, using $\rho \in R_{q,B}^-$ instead.

2.3 Top-down Ingredients

Håstad and Goldmann gave exponenential lower bounds for $\mathbf{TC}_3^0$ circuits with small bottom fanin [2].

Theorem 9. *Any depth 3 circuit with a* MAJORITY *gate at the top,* **SYM** *gates below and any kind of gates with fanin at most* $s-1$ *at the bottom which computes* $(\mathrm{MOD}_2)_n \circ \mathrm{AND}_s$ *must be of size* $2^{\Omega(\frac{n}{s4^s})}$.

Corollary 10. *Any depth 3 circuit with a* MAJORITY *gate at the top,* **SYM** *gates below and any kind of gates with fanin at most* $\frac{1}{3}\log n$ *at the bottom which computes* $(\mathrm{MOD}_2)_n \circ \mathrm{AND}_{\log n}$ *must be of size* $2^{n^{\Omega(1)}}$.

The results of Håstad and Goldmann were used by Razborov and Wigderson [9] to show the following lower bound.

Theorem 11. *Any* **MAJ**$\circ$**SYM**$\circ$**AND** *circuit computing* $(\mathrm{MOD}_2)_n \circ \mathrm{AND}_{\log n} \circ (\mathrm{MOD}_2)_n$ *must be of size* $n^{\Omega(\log n)}$.

Their proof involved applying a random restriction to the circuit, reducing the fanin of the AND gates to $\frac{1}{3}\log n$, while still having a circuit computing $(\mathrm{MOD}_2)_n \circ \mathrm{AND}_{\log n}$ as a subfunction.

The MOD_m-degree of a boolean function f on n variables is defined as the minimum degree of a polynomial p over $\mathbf{Z}_m$ such that $f(x) = 0 \Leftrightarrow p(x) = 0$ for all $x \in \{0,1\}^n$. Barrington, Beigel and Rudich gave $n^{\Omega(1)}$ lower bounds on the MOD_m-degree of the MOD_l and $\neg\mathrm{MOD}_l$ function when l has a prime factor not dividing m and also on the $\neg\mathrm{MOD}_m$ function when m is not a prime power[16]. These lower bounds was improved to $\Omega(n)$ by Tsai [12].

Green studied a weaker notion of MOD_m-degree [17], and generalized the lower bound of Tsai. The *weak* MOD_m-degree of a boolean function f (which is not the zero function) is the minimum degree of a polynomial p over $\mathbf{Z}_m$, which is not the zero polynomial, such that $p(x) \not\equiv 0 \pmod m \Rightarrow f(x) = 1$ for all $x \in \{0,1\}^n$.

Using this definition the the following lower bounds was obtained.

Theorem 12. *Let* m *and* l *be positive integers. If* l *has a prime factor not dividing* m *then the weak* MOD_m*-degree of the* MOD_l *and* $\neg\mathrm{MOD}_l$ *functions on* n *variables is* $\Omega(n)$*. If* m *is not a prime power the same holds for* $\neg\mathrm{MOD}_m$.

We will also need the following well known (weak) MOD_m-degree lower bounds for the MAJORITY function and the AND function ([8] Lemma 9.8, [18] Proposition 1).

Lemma 13. *Let* m *be any positive integer. The weak* MOD_m*-degree of the* MAJORITY *and* AND *functions on* n *variables are* $\geq \frac{n}{2}$ *and* n*, respectively.*

3 Lower Bound Proofs

In all proofs, we shall without loss of generality assume that the various real quantities appearing are integers.

3.1 Proof of Theorem 1

Let d be any constant and let C be a $\mathbf{MAJ} \circ \mathbf{SYM} \circ \mathbf{\Pi}_d$ circuit computing $f_n = (\mathrm{MOD}_2)_n \circ \mathrm{AND}_{\log n} \circ (\mathrm{MOD}_2)_n$, i.e., the circuit has $m = n^2 \log n$ Boolean inputs. Assume, to the contrary that the size S of C is $n^{o(\log n)}$. Also assume that n is sufficiently large.

We shall hit the circuit C and the function f_n with a random restriction ρ taken from $R_m^{m^{0.99}}$ and argue that

1. With overwhelming probability, $(f_n)_\rho$ contains as a subfunction (i.e., can be further restricted so as to obtain) the function $f_{\sqrt{n}}$.
2. With probability at least $\frac{1}{2}$, the function computed by $(C_n)_\rho$ can also be computed by a $\mathbf{MAJ} \circ \mathbf{SYM} \circ \mathbf{AND}$ circuit of size $n^{o(\log n)}$.

The desired lower bound then follows immediately from Theorem 11.

– Proof of item 1: The variables of f_n are divided into $n \log n$ group of n variables, each group corresponding to a (bottom) MOD_2 operation in the definition of f_n. If the random restriction ρ leaves $\sqrt{n}$ variables in each and all of these groups free, we have the desired subfunction, so we just have to show that this is the case with overwhelming probability. For a fixed set A of $n \geq \frac{\sqrt{m}}{\log m}$ variables, the probability that a random subset B of size $m^{0.99}$ of the $m = n^2 \log n$, variables (B being the free variables of the random restriction), has $|A \cap B| < n^{1/2}$ is negligible. Thus, the probability that ρ leaves less than $n^{\frac{1}{2}}$ variables free in any of the $n \log n$ groups is also negligible, as desired.
– Proof of item 2: We may view ρ as a composition of several random restrictions ρ_i, i=1, ..., d, with ρ_i taken at random from $R_m^{l_i}$ with $l_i = \frac{n}{(28 \log S)^d}$ and one last random restriction making the final number of free variables the correct number.
Look at the Π_d circuit at the bottom of C. Assume the gates at level $i+1$ from the bottom are OR gates and that after having applied $\rho_1, \ldots, \rho_i$ that gates at level $i+1$ represent decision trees of depth at most $\log S$ and hence DNFs of term length at most $\log S$. By Lemma 6, the probability that the function computed at a gate in level $i+1$ can not be written as a decision tree of depth at most $\log S$ is bounded by $\left(\frac{7}{28}\right)^{-\log S} = 2^{-2\log S}$. A similar argument holds for the levels of AND gates. (Also, we view the AND or OR gates at the bottom level, as taking OR or AND gates of fanin 1 as input). Thus, the probability that ρ does not turn *all* the gates of the Π_d circuit into decision trees of depth at most $\log S$ is at most $2^{-2\log S} S \leq \frac{1}{2}$. Thus, with probability at least $\frac{1}{2}$, C_ρ can be computed as the majority of a number of symmetric gates, each taking as input a function computed by a decision tree of depth $\log S$. Decision trees are also DNFs *with the additional property that any input assignment makes at most one term true.* Thus, a symmetric gate G that takes as input a number of such trees $T_1, T_2, \ldots, T_k$ can be replaced with another symmetric gate G' that takes as inputs all terms (i.e.,

conjunctions of variables) of all the DNFs corresponding to all the trees T_j. Thus, with probability at least $\frac{1}{2}$ we have that C_ρ can be computed by a $\mathbf{MAJ} \circ \mathbf{SYM} \circ \mathbf{AND}$ circuit of size polynomial in S and $2^{\log S}$ (the latter quantity being the size of a decision tree of depth $\log S$), i.e., polynomial in s. Since $S = n^{o(\log n)}$, the size of the new circuit is also $n^{o(\log n)}$ and we are done. □

3.2 Proof of Theorem 2

First observe that after applying a restriction to MOD_l or $\neg\mathrm{MOD}_l$, by further setting at most $l-1$ extra variables, it results in the same function on fewer variables. Thus we can, using Theorem 12 choose c such that none of the functions on n variables that can appear after applying a restriction, has MOD_m-degree less than cn.

Now assume that C is a $\mathrm{MOD}_m \circ \mathbf{AC}^0_d$ circuit of depth $d+1$ and size $S = 2^{\frac{1}{14}(cn)^{\frac{1}{d}}}$ that computes a function with that property.

For $0 \le i \le d-1$ let $n_{i+1} = \frac{n}{14}\left(14 \log_2 S\right)^{-i} = \frac{n}{14}\left((cn)^{\frac{1}{d}}\right)^{-i}$. Pick a restriction ρ at random from $R_n^{n_d}$ by picking it as a composition of randomly chosen restrictions $\rho_1, \ldots, \rho_d$, where $\rho_1 \in R_n^{n_1}$ and $\rho_{i+1} \in R_{n_i}^{n_{i+1}}$ for $0 \le i \le d-1$.

Viewing a gate at level 1 as a DNF with terms of length 1 or a CNF with clauses of size 1, the probability that the function computed does not have a decision tree of height at most $\log_2 S$ after applying ρ_1, is less than $(7\frac{n_1}{n})^{\log_2 S} = \frac{1}{S}$ by Lemma 6.

Now assuming that after applying the restrictions $\rho_1, \ldots, \rho_i$, every function computed at a gate at level i has a decision tree of height at most $\log_2 S$, consider a gate at level $i+1$. If the gate is an OR gate we rewrite all it's inputs as CNFs with clauses of size at most $\log_2 S$, and merge them with the gate obtaining a CNF with clauses of size at most $\log_2 S$. If the gate is an AND gate we similarly obtain a DNF with terms of size at most $\log_2 S$.

Then using Lemma 6 again, the probability that the function computed at the gate at level $i+1$ does not have a decision of height at most $\log_2 S$ after applying ρ_{i+1}, is less than $(7\frac{n_{i+1}}{n_i} \log_2 S)^{\log_2 S} = \frac{1}{S}$.

If in fact $i = d$, that is the gate at level $i+1$ is the top MOD_m gate, we also rewrite all it's inputs as DNFs with terms of size at most $\log_2 S$. Since these are constructed from decision trees, we can argue as in the proof of Theorem 1 to simply absorb the top OR gates into the MOD_m gate, resulting in a $\mathbf{MOD}_m \circ \mathbf{AND}$ circuit with bottom fanin less than $\log_2 S$.

Since the probability of failure at every gate is less than $\frac{1}{S}$, there exists an appropriate restriction ρ, making the restricted function on $n_d = \frac{(cn)^{\frac{1}{d}}}{14c}$ variables be computed by a $\mathbf{MOD}_m \circ \mathbf{AND}$ circuit with bottom fanin less than $\log_2 S$. This implies that the restricted function has MOD_m-degree at most $\log_2 S = \frac{(cn)^{\frac{1}{d}}}{14}$, thus contradicting the existence of C. □

3.3 Proof of Theorem 3

The proof will closely follow the outline of the proof of Theorem 2. Let S and n_{i+1} be as before, where $c < \frac{1}{2}$, and assume assume that C is a $\mathbf{MOD}_m \circ \mathbf{AC}_d^0$ circuit of depth $d+1$ but of size $S^{1-\epsilon}$ computing MAJORITY.

Now since the size of C is only $S^{1-\epsilon}$ the probability that a random restriction $\rho \in R_n^{n_d}$ does not create a $\mathbf{MOD}_m \circ \mathbf{AND}$ circuit with bottom fanin at most $\log_2 S$ is less than $S^{-\epsilon} = o(n^{-\frac{1}{2}})$.

The probability that the number of variables assigned 0 and 1 by ρ differ by at most 1 is at least $\frac{1}{2^{n-n_d}}\binom{n-n_d}{\frac{n-n_d}{2}}$. Using Stirling's approximations this is $\Omega\left((n-n_d)^{-\frac{1}{2}}\right) = \Omega\left(n^{-\frac{1}{2}}\right)$.

Thus for all large enough n there exists a restriction ρ implying the existence of a $\mathbf{MOD} \circ \mathbf{AND}$ circuit with bottom fanin less than $\log_2 S$ computing MAJORITY on n_d variables, contradicting the existence of C as before, now using the MOD_m-degree lower bound for MAJORITY in Lemma 13. $\square$

3.4 Proof of Theorem 4

The proof consists of several cases. We will only give the proof for $\mathbf{MOD}_m \circ \mathbf{\Sigma}_k$ circuits and for $k = 2$ and for odd $k > 2$.

For $k = 2$, assume that C is a $\mathrm{MOD}_m \circ \mathbf{\Sigma}_2$ circuit of size $S = 2^{\frac{1}{24}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}}$ with bottom fanin $r = \frac{1}{24}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}$ computing f_2^m. Pick at random ρ from $R_{q,B}^{-}$, and apply the restriction $\rho g(\rho)$. Using Lemma 7 the probability that a single $\mathbf{\Sigma}_2$ subcircuit cannot be represented as a CNF with clauses of size at most $s = \frac{1}{12\sqrt{2}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}$ after applying the restriction is at most $(6qr)^s = \left(\frac{1}{2}\left(\frac{\log m}{m}\right)^{\frac{1}{4}}\right)^s \leq 2^{-s}$. Thus the probability of failure at any $\mathbf{\Sigma}_2$ subcircuit is at most $S2^{-s} = 2^{\frac{1}{12}\left(\frac{1}{2}-\frac{1}{\sqrt{2}}\right)\left(\frac{m}{\log m}\right)^{\frac{1}{4}}} = o(1)$.

Thus for m large enough, we can using using Lemma 8 pick a restriction where simultaneously all $\mathbf{\Sigma}_2$ subcircuits can be converted to CNFs with clauses of size at most s and the circuit still computes f_1^m. Now, after the restriction the function computed at every $\mathbf{\Sigma}_2$ subcircuit has *both* a CNF with clauses of size at most s and a DNF with terms of size at most r (since it already had that before applying the restriction). It then follows that they also have decision trees of depth at most sr. These are then converted into DNFs with mutually contradictory terms of size at most sr and the top OR gates can be absorbed into the MOD_m gate, resulting in a $\mathrm{MOD}_m \circ \mathrm{AND}$ circuit with bottom fanin $sr = \frac{1}{288\sqrt{2}}\sqrt{\frac{m}{\log m}}$ computing f_1^m, which is the AND function on $\sqrt{\frac{1}{2}m\log m}$ variables, which contradicts Lemma 13.

For general $k > 2$ the proof goes by induction. Assume C is a $\mathrm{MOD}_m \circ \mathbf{\Sigma}_k$ circuit of size $S = 2^{\frac{1}{12\sqrt{2k}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}}$ with bottom fanin $r = \frac{1}{12\sqrt{2k}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}$

computing f_k^m. Assuming k is odd means that the gates at the bottom are OR gates.

Pick at random ρ from $R_{q,B}^+$, and apply the restriction $\rho g(\rho)$. Then the $\mathbf{\Pi}_2$ subcircuits of C at the two bottom levels can all be converted to $\mathbf{\Sigma}_2$ circuits with bottom fanin $s = \frac{1}{12\sqrt{2(k-1)}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}$ except with probability at most $S2^{-s} = 2^{\left(\frac{1}{\sqrt{k}}-\frac{1}{\sqrt{k-1}}\right)\frac{1}{12\sqrt{2}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}} = o(1)$ and the top OR gates can then be merged with OR gates just above. Now as before, for m large enough, we can pick a restriction obtaining a $\mathrm{MOD}_m \circ \mathbf{\Sigma}_{k-1}$ circuit of size $S \le 2^{\frac{1}{12\sqrt{2(k-1)}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}}$ and bottom fanin s computing f_{k-1}^m, thereby obtaining a contradiction. $\square$

By adding gates of fanin 1 at the bottom of depth k $\mathbf{MOD}_m \circ \mathbf{AC}^0$ circuits we obtain the following.

Corollary 14. *For sufficiently large m, $\mathbf{MOD}_m \circ \mathbf{AC}_{k-1}^0$ circuits computing f_k^m require size at least $2^{\frac{1}{12\sqrt{2k}}\left(\frac{m}{\log m}\right)^{\frac{1}{4}}}$.*

One can obtain the stronger lower bound $2^{\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}}$ by using the observation of Berg and Ulfberg mentioned in Sect. 2.2.

3.5 Proof of Theorem 5

Much of the proof is the same as the proof of Theorem 2. In particular choose c in the same way, and assume that C is a $\mathbf{AC}_d^0$ circuit, containing a single MOD_m gate G, of depth d and size $\frac{1}{3}S^{\frac{1}{2}}$, where $S = 2^{\frac{1}{14}(cn)^{\frac{1}{d}}}$, that computes a function f with the stated property. Define n_{i+1} and pick the restriction ρ in the same way.

Let C_G be the subcircuit of C with G as its output. Let C_0 and C_1 be the subcircuits obtained from C by substituting the constants 0 and 1 instead of G. We will apply the restriction to all three circuits simultaneously and obtain decision trees of height at most $\frac{1}{2}\log_2 S$ for every gate except G.

At any time, when encountering a NOT gate (which are not included in the counting of levels), by negating the leaves of the decision tree obtained for its input, we can obtain a similar decision tree for itself, and in this way ignore them in this proof.

By the choice of n_{i+1}, we obtain the desired decision trees at every gate except with probability at most $\frac{1}{S^{\frac{1}{2}}}$. When we in fact we encounter the MOD_m gate G, we can as in the proof of Theorem 2 obtain a $\mathbf{MOD}_m \circ \mathbf{AND}$ circuit with bottom fanin less than $\frac{1}{2}\log_2 S$ for the function computed.

Since the probability of failure at every gate is less than $\frac{1}{S^{\frac{1}{2}}}$, and the total number of gates in the 3 circuits is less than $S^{\frac{1}{2}}$, there exists an appropriate restriction ρ, such that we in this way obtain polynomials p, q_0 and q_1 of degree $\frac{1}{2}\log_2 S = \frac{(cn)^{\frac{1}{d}}}{28}$ over $\mathbf{Z}_m$ on $n_d = \frac{(cn)^{\frac{1}{d}}}{14c}$ variables variables, such that $p(x) \not\equiv 0 \pmod m \Leftrightarrow C_{G,\rho}(x) = 1$, $q_0(x) = C_{0,\rho}(x)$ and $q_1(x) = C_{1,\rho}(x)$ for all x.

From these polynomials we will construct a *weak* representation, in the sense of Lemma 12, of either f_ρ or $\neg f_\rho$. We will consider 3 cases.

1. There exists x such that $C_{G,\rho} = 1$ and $C_{1,\rho}(x) = 1$. In this case $p(x)q_1(x)$ weakly represents f_ρ, because q_1 is an *exact* representation of $C_{1,\rho}$.
2. There exists x such that $C_{G,\rho}(x) = 1$ and for every x it holds that $C_{G,\rho}(x) = 1$ implies $C_{1,\rho}(x) = 0$. In this case $p(x)$ weakly represents $\neg f_\rho$.
3. For every x it holds that $C_{G,\rho}(x) = 0$. In this case $q_0(x)$ weakly represents f_ρ, since f_ρ is not identically zero.

Thus we always obtain a weak representation of degree at most $2\frac{1}{2}\log_2 S = \frac{(cn)^{\frac{1}{d}}}{14}$ of either f_ρ or $\neg f_\rho$ on $n_d = \frac{(cn)^{\frac{1}{d}}}{14c}$ variables contradicting the existence of C. Finally, by picking c sufficiently small we obtain the result as stated.

4 Oracle Consequences

Using the approach of Furst, Saxe and Sipser [7] one can use super-quasi-polynomial lower bounds for constant depth circuits to derive oracle separations of corresponding subclasses of PSPACE. Vollmer showed how to obtain such separations in a uniform way under several conditions, obtaining oracle separations for subclasses of PSPACE defined in terms of generalized operators [19].

More precisely, we will use the following part of a more general theorem.

Theorem 15 ([19], Theorem 4.4). *Let $G_1, \ldots, G_k$ and $G'_1, \ldots, G'_l$ be families of boolean gates such that a given function computed by polynomial size $G_1 \circ \cdots \circ G_k$ circuits require $G'_1 \circ \cdots \circ G'_l \circ \Sigma \circ \mathrm{AND}_{\log^{O(1)} n}$ circuits of super-quasi-polynomial size. Then relative to an oracle $G_1 \cdots G_k \cdot \mathrm{P} \not\subseteq G'_1 \cdots G'_l \cdot \mathrm{P}$.*

Here the Σ denotes gates which takes inputs of which at most one of them is 1 at any time, and outputs 1 if in fact any of its inputs is 1. In all our applications these can be removed. Indeed we always have that G'_l is symmetric, and thus they can absorb the sums.

The super-quasi-polynomial lower bounds of this paper would give oracle separations for the classes $\mathrm{MOD}_m\mathrm{P} \cdot \boldsymbol{\Pi}^p_k$ and $\mathrm{MOD}_m\mathrm{P} \cdot \boldsymbol{\Sigma}^p_k$. Using the approach of Torán [20] showing that $\mathrm{PP}^K = C{\cdot}K$ for any class K in the counting hierarchy, one can similarly show that $\mathrm{MOD}_m\mathrm{P}^{\boldsymbol{\Sigma}^p_k} = \mathrm{MOD}_m\mathrm{P} \cdot \boldsymbol{\Pi}^p_k$. Now from Theorem 2 we then directly obtain the following, using the theorem of Vollmer.

Corollary 16. *Let m and l be positive integers. If l has a prime factor not dividing m then relative to some oracle, $\mathrm{MOD}_l\mathrm{P} \not\subseteq \mathrm{MOD}_m\mathrm{P}^{\mathrm{PH}}$ and $\mathrm{coMOD}_l\mathrm{P} \not\subseteq \mathrm{MOD}_m\mathrm{P}^{\mathrm{PH}}$. If m is not a prime power then relative to some oracle, $\mathrm{coMOD}_m\mathrm{P} \not\subseteq \mathrm{MOD}_m\mathrm{P}^{\mathrm{PH}}$.*

From Theorem 3 and finally from Theorem 4 we obtain

Corollary 17. *For any integer m, relative to some oracle, $\mathrm{PP} \not\subseteq \mathrm{MOD}_m\mathrm{P}^{\mathrm{PH}}$.*

Corollary 18. *For any m and every $k > 1$, relative to some oracle, $\boldsymbol{\Pi}^p_k \not\subseteq \mathrm{MOD}_m\mathrm{P}^{\boldsymbol{\Sigma}^p_{k-1}}$.*

The case $k = 1$ follows by using Lemma 13 instead.

Acknowledgements The authors are supported by BRICS, Basic Research in Computer Science, a Centre of the Danish National Research Foundation.

References

1. Karchmer, M., Wigderson, A.: Monotone circuits for connectivity require super-logarithmic depth. SIAM Journal on Discrete Mathematics **3** (1990) 255–265
2. Håstad, J., Goldmann, M.: On the power of small-depth threshold circuits. Computational Complexity **1** (1991) 113–129
3. Beame, P.: A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington (1994) Availible online at www.cs.washington.edu/homes/beame.
4. Berg, C., Ulfberg, S.: A lower bound for perceptrons and an oracle sepration of the $\mathrm{PP}^{\mathrm{PH}}$ hierarchy. Journal of Computer and System Sciences **56** (1998) 263–271
5. Goldmann, M.: On the power of a threshold gate at the top. Information Processing Letters **63** (1997) 287–293
6. Krause, M., Pudlák, P.: Computing boolean functions by polynomials and threshold circuits. Computational Complexity **7** (1998) 346–370
7. Furst, M., Saxe, J.B., Sipser, M.: Parity, circuits, and the polynomial-time hierarchy. Mathematical Systems Theory **17** (1984) 13–27
8. Håstad, J.: Computational limitations of small-depth circuits. MIT Press (1987)
9. Razborov, A., Wigderson, A.: $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. Information Processing Letters **45** (1993) 303–307
10. Razborov, A.A., Rudich, S.: Natural proofs. Journal of Computer and System Sciences **55** (1997) 24–35
11. Zhang, Z.L., Barrington, D.A.M., Tarui, J.: Computing symmetric functions with AND/OR circuits and a single MAJORITY gate. In: Proceedings of the 10th Annual Symposium on Theoretical Aspects of Computer Science. Volume 665 of Lecture Notes in Computer Science., Springer (1993) 535–544
12. Tsai, S.C.: Lower bounds on representing Boolean functions as polynomials in Z_m. SIAM Journal on Discrete Mathematics **9** (1996) 55–62
13. Aspnes, J., Beigel, R., Furst, M., Rudich, S.: The expressive power of voting polynomials. Combinatorica **14** (1994) 1–14
14. Beigel, R.: When do extra majority gates help? Polylog(n) majority gates are equivalent to one. Computational Complexity **4** (1994) 314–324
15. Barrington, D.A.M., Straubing, H.: Complex polynomials and circuit lower bounds for modular counting. Computational Complexity **4** (1994) 325–338
16. Barrington, D.A.M., Beigel, R., Rudich, S.: Representing Boolean functions as polynomials modulo composite numbers. Computational Complexity **4** (1994) 367–382
17. Green, F.: A complex-number fourier technique for lower bounds on the mod-m degre. Computational Complexity **9** (2000) 16–38
18. Tarui, J.: Degree complexity of Boolean functions and its applications to relativized separations. In: Proceedings of the Sixth Annual Structure in Complexity Theory Conference, IEEE Computer Society Press, (1991) 382–390
19. Vollmer, H.: Relating polynomial time to constant depth. Theoretical Computer Science **207** (1998) 159–170
20. Torán, J.: Complexity classes defined by counting quantifiers. Journal of the ACM **38** (1991) 753–774