SIAM J. COMPUT.
Vol. 43, No. 5, pp. 1699–1708
© 2014 Society for Industrial and Applied Mathematics

ON THE CORRELATION OF PARITY AND SMALL-DEPTH CIRCUITS*

JOHAN HÅSTAD†

Abstract. We prove that the correlation of a depth-d unbounded fanin circuit of size S with parity of n variables is at most $2^{-\Omega(n/(\log S)^{d-1})}$.

Key words. circuits complexity, small-depth circuits, parity, switching lemma

AMS subject classification. 68Q17

DOI. 10.1137/120897432

1. Introduction. Proving absolute lower bounds for concrete computational problems in realistic models of computation is a holy grail for the area of computational complexity. The general program of proving lower bounds for simple computational models with important early contributions by Furst, Saxe, and Sipser [4], Sipser [9], Ajtai [1], Yao [11], Håstad [5], Smolensky [10], and Razborov [7] seemed to be a promising road of establishing such lower bounds but came to an almost complete halt in late 1980s. One possible explanation for the lack of progress might be that lower bounds for stronger models of computation need other methods. This is made formal by the concept of "natural proofs" introduced by Razborov and Rudich [8].

Possibly the simplest nontrivial model of computation is that of bounded-depth Boolean circuits of unbounded fanin. Such a circuit contains AND-gates and OR-gates of unbounded fanin, takes as inputs literals (i.e., variables in positive or negated form), and has a depth that is bounded by a constant independent of the number of variables. It was a major step forward when Ajtai [1] and Furst, Saxe, and Sipser [4] independently proved that the simple function of parity requires circuits of superpolynomial size to be computed in this model.

These bounds were later improved by Yao [11] and Håstad [5], establishing that size $\exp(n^{\Theta(\frac{1}{d-1})})$ was necessary and sufficient to compute the parity of n variables by a depth-d circuit.

These results were based on the concept of random restrictions where most inputs of a circuit are fixed to constants in order to simplify the circuit. The key observation is that the simplified circuit should compute the parity (or possibly the negation of this function) on the remaining variables and thus if these simplifications are substantial enough a contradiction is obtained. To be more concrete, by tuning parameters, it is possible to choose a restriction such that one can remove one level of the circuit. This is achieved by applying the switching lemma of Håstad [5], which says that it possible to switch an and-of-ors into an or-of-ands keeping, with high probability, the bottom fanin small.

It is not difficult to see that all the proofs based on restrictions in fact established that any small circuit of constant depth only agrees with parity for marginally more

*Received by the editors November 11, 2012; accepted for publication (in revised form) June 24, 2014; published electronically September 25, 2014. This research was supported by ERC advanced grant 226 203.

http://www.siam.org/journals/sicomp/43-5/89743.html

†KTH Royal Institute of Technology, Stockholm S100 44, Sweden (johanh@kth.se).

than half of the inputs. Apart from being an interesting result in itself, this fact is central to Cai's result [3] that a random oracle separates PSPACE from the polynomial time hierarchy.

In this paper we are interested in obtaining exact bounds on the best possible agreement, and if this fraction is $(1+c)/2$ let us call c "the correlation." It follows, more or less immediately, from the proof of Håstad that a circuit of size 2^s and depth d only has correlation $\exp(-\Omega(n^{\frac{1}{d-1}}))$ with parity provided that $s \leq o(n^{\frac{1}{d-1}})$. The bottleneck in this argument is the estimate for the probability that we are not able to do the required switching. It is curious that the estimate of the correlation gets only marginally better with decreasing values of s and obtaining a bound better than $\exp(-\Omega(n^{\frac{1}{d-2}}))$ for this correlation seems to require some new idea.

Somewhat surprisingly, Ajtai [1], who did not get as strong bounds for the size of depth-d circuits computing parity exactly, proved the stronger bound $\exp(-\Omega(n^{1-\epsilon}))$ for the correlation of parity and the output of polynomial size circuits of depth d.

The correlation was recently proved to be much smaller by Beame, Impagliazzo, and Srinivasan [2], who established the unusual bound of $\exp(-\Omega(n/(2^{2ds^{4/5}})))$ for the correlation of circuits of size 2^s and depth d with parity. Motivated by this result and the techniques used we now revisit the old techniques based on the switching lemma with the aim of strengthening the bounds for the correlation of small circuits with parity.

Before turning to discussing our methods and results, let us describe how to construct a set of depth-d circuits of reasonably small size that have a nontrivial correlation with parity. Dividing the inputs into groups of size s^{d-1} it is possible to compute the parity exactly of each such group of inputs with a circuit of size roughly 2^s and depth d. Circuits with this property can be chosen to have either an $\wedge$-gate or $\vee$-gate as their output gate and let us assume the latter. Taking the disjunction of all these circuits we maintain depth d and get a circuit that mostly outputs 1 but whenever it outputs 0, it always agrees with the parity of the inputs. The correlation of this circuit with parity is easily seen to be 2^{-g}, where $g \approx n/s^{d-1}$ is the number of groups. The purpose of this paper is to prove that this construction is, up to the constants involved, optimal.

The proof is very much based on an extension of the switching lemma, where we do not allow the failure to do the required switching but instead, with small probability, fix some additional variables not set by the original randomized restriction.

Similar bounds, by related but not identical methods, have been obtained independently by Impagliazzo, Matthews, and Paturi [6]. That paper has as its main goal to obtain a satisfiability algorithm for AC^0, and the bound for the correlation of parity and small-depth circuits is obtained as a corollary. Our more direct approach leads to, in our eyes, a considerably simpler argument. The bounds obtained by [6] are in most cases identical (apart from some involved constants) to ours but for circuits of size $n^{1+o(1)}$ their results are stronger.

2. Preliminaries. We study circuits consisting of unbounded fanin $\wedge$- and $\vee$-gates. We reserve the letter n for the number of inputs to this circuit. We denote by x_i, $1 \leq i \leq n$, the inputs to this circuit and we assume that we have a unique output gate as we are interested in circuits computing Boolean functions.

The size of the circuit is defined to be the number of gates it contains, not counting the inputs (and in fact most of the time we do not even count the gates next to the inputs). The depth is defined to be longest path from any input to the output. We assume that the gates appear in alternating layers of $\wedge$-gates and $\vee$-gates because if

we have two layers of the same type we can directly move the inputs of the lower gate to the higher gate. By introducing dummy gates of fanin one we can assume that the circuit is layered with gates at level i getting inputs from layer $i-1$. This causes at most a constant blow-up in the size of the circuit and this small change is not important for us. As stated above we allow arbitrary fanin of the gates and as we are dealing with circuits, also the fan-out is arbitrary. We use 1 to denote "true" and 0 to denote "false."

We study the subcircuits of depth 2 given by the two layers closest to the inputs. We sometimes treat these as disjoint circuits, but in these situations we only include the gates of distance at least two away from the input in the size count and thus we can duplicate any common gates to make the circuits disjoint.

Let $p \in [0,1]$ be a real number, and a random restriction from the space R_p is defined as follows. For each variable x independently keep it as a variable with probability p and otherwise set it to one of the two constants 0 and 1 with equal probability. A typical restriction is denoted by ρ and the notation for keeping a variable is $\rho(x_i) = *$, while the other two outputs of ρ are 0 and 1, interpreted in the natural way. For a function f we let $f\lceil_\rho$ be the function in the untouched variables obtained by making the substitutions described by ρ.

We analyze many conditional probabilities and in particular we are interested in sets of restrictions that are monotone in ρ in the sense that fixing the value of more variables can only make ρ more likely to belong to the set.

DEFINITION 2.1. *A set $\mathcal{F}$ of restrictions is* downward closed *if whenever $\rho \in \mathcal{F}$ and $\rho'(x_i) = \rho(x_i)$ for all x_i such that $\rho(x_i) \in \{0,1\}$, then $\rho' \in F$.*

An equivalent definition is to say that for any $\rho \in \mathcal{F}$ changing the value on any input from the value $*$ to either non-$*$ value, the resulting restriction is also an element of $\mathcal{F}$.

The classical conditioning for the switching lemma of [5] is to focus on conditions of the form $F\lceil_\rho \equiv 1$ for a Boolean function F. It is easy to see that the set of restrictions that satisfy such a condition is downward closed but there are many downward closed sets that are not on this form. One example would be the set of restrictions such that the value of $F\lceil_\rho$ is independent of the remaining variables (but can be either 0 or 1).

The following simple lemma follows immediately from the definition.

LEMMA 2.2. *Let $\mathcal{F}$ and $\mathcal{F}'$ be two downward closed sets of restrictions. Then the set $\mathcal{F} \cap \mathcal{F}'$ is also downward closed.*

We assume that the reader is familiar with the concept of a decision tree and we need the following extension.

DEFINITION 2.3. *A set of functions $(g_i)_{i=1}^S$ has a* common s-partial decision tree *of depth d if there is a decision tree of depth d such that at each leaf of this decision tree, each function g_i is computable by an ordinary decision tree of depth s of the variables not queried on the given path.*

Said differently, each path of the decision tree defines a restriction π that gives values to the queried variables. The claim is that $g_i\lceil_\pi$ can be computed by a decision tree of depth s for each i. Finally, let us formally define correlation.

DEFINITION 2.4. *A function f has correlation c with a function g iff*

$$\Pr[f(x) = g(x)] = (1+c)/2,$$

where the probability is taken over a uniformly chosen x.

3. The main argument. Let us first state our main theorem.

THEOREM 3.1. *Let $f : \{0,1\}^n \mapsto \{0,1\}$ be computed by a depth d circuit of bottom fanin t which contains at most S gates of distance at least 2 from the inputs. Then the correlation of f with parity is bounded by*

$$2^{-c_d n/t(\log S)^{d-2}},$$

where c_d is a positive constant depending only on d.

We have an immediate corollary.

COROLLARY 3.2. *Let f, $\{0,1\}^n \mapsto \{0,1\}$ be computed by a depth-d circuit of size S. Then the correlation of f with parity is bounded by*

$$2^{-c_d n/(\log S)^{d-1}},$$

where c_d is a positive constant depending only on d.

Proof. This follows from Theorem 3.1 as we can consider a depth d circuit as a depth $d+1$ circuit with bottom fanin one. Gates at distance two from the inputs in this new circuit correspond to gates in the original circuit. □

We now turn to the proof of Theorem 3.1. As discussed in the introduction, the proof is very much based on the proof of the switching lemma of [5], and let us start by stating this lemma and recalling its proof. In the process we slightly generalize the lemma in that we allow a conditioning of the form that the restriction belongs to an arbitrary set that is downward closed. This generalization follows from the original proof but we are not aware that this strengthening has been stated explicitly anywhere.

LEMMA 3.3. *Let f be computed by a depth-2 circuit of bottom fanin t. Let $\mathcal{F}$ be a downward closed set of restrictions and ρ a random restriction from R_p. Let depth$(f\lceil_\rho)$ be the minimal depth of a decision tree computing $f\lceil_\rho$. Then*

$$\Pr[\mathit{depth}(f\lceil_\rho) \geq s \mid \rho \in \mathcal{F}] \leq (5pt)^s.$$

Proof. Suppose, without loss of generality, that f is a CNF, i.e., that it can be written as

$$f = \wedge_{i=1}^m C_i,$$

where each C_i is a disjunction of at most t literals. The proof is by induction over m and the base case is when $m = 0$, in which case $f\lceil_\rho$ is always computable by a decision tree of depth 0. Whenever needed we can clearly assume that $5pt \leq 1$ as otherwise the lemma is meaningless.

We divide the analysis into two cases depending on whether C_1 is forced to 1. Let us for notational convenience assume that C_1 is the disjunction of $x_1, x_2 \ldots x_{t_0}$ for some $t_0 \leq t$. Clearly we can bound the probability of the lemma as the maximum of

$$\Pr[\mathrm{depth}(f\lceil_\rho) \geq s \mid \rho \in \mathcal{F} \wedge C_1\lceil_\rho \equiv 1]$$

and

$$\Pr[\mathrm{depth}(f\lceil_\rho) \geq s \mid \rho \in \mathcal{F} \wedge C_1\lceil_\rho \not\equiv 1]. \tag{3.1}$$

The first term is taken care of by induction applied to f without its first conjunction (and thus having size at most $m-1$) using Lemma 2.2 to ensure that the conditioning is of the correct form. We need to consider the second case given by (3.1).

Since (3.1) implies that $C_1\lceil_\rho \not\equiv 1$, to avoid that $f\lceil_\rho \equiv 0$ there must be some nonempty set Y of size $r > 0$ of variables appearing in C_1 which are given the value $*$ by ρ. Let us for the moment assume that $r < s$ and we later verify that the resulting estimate is valid also when $r \geq s$. We construct a decision tree by first querying the variables in Y. One of the 2^r answers forces f to 0 and this will not result in a decision tree of depth at least s. Let π be an assignment to the variables in Y. We can now bound (3.1) as

$$\sum_{\pi,Y} \Pr[\mathrm{depth}(f\lceil_{\pi\rho}) \geq s - r \wedge \rho(Y) = * \wedge \rho(C_1/Y) = 0 \mid \rho \in \mathcal{F} \wedge C_1\lceil_\rho \not\equiv 1],$$

where Y is a nonempty set of size r. A key lemma is the following.

LEMMA 3.4. *If Y is a set of size r containing variables from C_1, then*

$$\Pr[\rho(Y) = * \mid \rho \in \mathcal{F} \wedge C_1\lceil_\rho \not\equiv 1] \leq \left(\frac{2p}{1+p}\right)^r.$$

Proof. Assume that a restriction ρ contributes to the probability in question. Consider all the ways of constructing restrictions ρ' by changing, in all possible ways, the values taken by the restriction on the set Y, taking values only 0 and $*$. (Remember that ρ gives the value $*$ to all variables in Y.) Note that any constructed ρ' still satisfies the conditioning and that each ρ' is constructed from a unique ρ as the two restrictions agree outside Y. If ρ' is changed to take the value 0 for k different inputs, then

$$\Pr[\rho'] = \left(\frac{1-p}{2p}\right)^k \Pr[\rho],$$

and as

$$\sum_{i=0}^{r} \binom{r}{i} x^i = (1+x)^r, \tag{3.2}$$

we conclude that the total probability of all restrictions constructed from ρ is at least

$$\left(\frac{1+p}{2p}\right)^r \Pr[\rho]$$

and the lemma follows. □

Finally we estimate

$$\Pr[\mathrm{depth}(f\lceil_{\pi\rho}) \geq s - r \mid \rho(Y) = * \wedge \rho(C_1/Y) = 0 \wedge \rho \in \mathcal{F} \wedge C_1\lceil_\rho \not\equiv 1]$$

by induction. We need to check that the conditioning defines a downward closed set but this is more or less obvious as we are considering restrictions on variables outside Y (as these are already fixed by π). Changing the value of ρ on any variable not contained in C_1 from $*$ to a non-$*$ value cannot violate any of the four conditions.

Thus we can conclude that the probability of obtaining a decision tree of depth at least s is at most

$$\sum_{Y} (2^{|Y|} - 1)(5pt)^{s-|Y|} \left(\frac{2p}{1+p}\right)^{|Y|}. \tag{3.3}$$

Before ending the proof let us observe that this bound is correct also in the case $r = |Y| \geq s$, as in this case the depth of the decision tree is always at least s and the probability of this case happening is, by Lemma 3.4, bounded by $\left(\frac{2p}{1+p}\right)^r$, which is smaller than the corresponding term in (3.3) as we are assuming that $5pt \leq 1$.

Using (3.2) again we see that we get the final estimate

$$(5pt)^s \left(\left(1 + \frac{4}{5t(1+p)} \right)^{|C_1|} - \left(1 + \frac{2}{5t(1+p)} \right)^{|C_1|} \right).$$

This is an increasing function of $|C_1|$ and thus we can assume that this number equals t. The second factor is of the form $(1+2x)^t - (1+x)^t$ and, as this is an increasing function of x and $(1+2x) \leq (1+x)^2$, it is bounded by $y^2 - y$, where $y = (1 + \frac{2}{5t})^t$. Finally, as $1 + x \leq e^x$, this can be upper bounded by $e^{4/5} - e^{2/5} < 1$ and this finishes the inductive step. □

The above proof is an induction but it is not difficult to unravel the induction and given a circuit for f and a restriction ρ to explicitly give a recursive procedure that constructs a decision tree for $f\lceil_\rho$. As this process is important for us, let us describe it in more detail, as follows.

Procedure DT-create(f, ρ).

1. Find the first clause, C, not forced to 1 by ρ.
2. Let Y be the set of variables appearing in C given the value $*$ by ρ. Suppose $|Y| = r$ and let Y be the first r variables queried in the decision tree. For each of the 2^r assignments, π, to these variables we recurse by looking at $f\lceil_\pi$.

We can note that with good probability (about $1 - 2pt$) the clause C does not contain any variable given the value $*$ by ρ and in this situation the construction of the decision tree is completed. The proof also uses that one of the assignments on Y forces f to be false and no recursion is needed. This gives a better constant in the lemma but this is not essential for the proof to work. To make it easier to reason about conditioning in the proof below we highlight the assignments used in the above construction.

Let τ be a binary string of length s used for the choices in DT-create. To be more precise, whenever a set Y of size r is found we use the next r unused bits of τ as values on this set. If fewer than r unused bits remain in τ the procedure terminates. We call τ "the advice." Let us say that "τ suggests that $f\lceil_\rho$ requires a decision tree of size at least s" if all the bits of τ are used or we do not have a sufficient number of unused bits and terminate. We use the shorthand $\mathrm{Sugg}(\tau, f, \rho) = s'$ to denote the fact that the size of the union of all the sets Y seen in DT-create(f, ρ) is s' when τ is used as the advice. We denote a typical union of Y's by T and use the notation $\mathrm{Prod}(\tau, f, \rho) = T$. Note that it might be the case that $s' > s$, which happens in the case when a discovered Y has size larger than the number of unused bits remaining in τ.

We use the term "suggests" as it is not a strict implication. Indeed if some assignment suggests that $f\lceil_\rho$ requires a deep decision tree, this is not necessarily true as some other process might find a small depth decision tree for $f\lceil_\rho$. On the other hand it is not difficult to see that if $f\lceil_\rho$ requires a decision tree of depth s, then this is suggested by some advice.

LEMMA 3.5. *If depth$(f\lceil_\rho) \geq s$, then Sugg$(\tau, f, \rho) \geq s$ for at least one $\tau \in \{0,1\}^s$.*

Proof. Indeed if $f\lceil_\rho$ requires a decision tree of depth s, then any process trying to create a decision of smaller depth fails. In particular DT-create(f, ρ) must result in some path of length at least s and this implies the existence of the advice τ. □

We have also given the essential part of the proof of the following lemma.

LEMMA 3.6. *Let $\tau \in \{0,1\}^s$. Then*

$$Pr[Sugg(\tau, f, \rho) \geq s'] \leq (3pt)^{s'}$$

for any $s' \geq 0$. This remains true conditioned on $\rho \in F$ for any downward closed set F.

Proof. The proof is essentially a repeat of the proof of Lemma 3.3 up to (3.3), where the factor $2^{|Y|-1}$ is not needed as τ gives a set of unique values to the variables of Y. Clearly we also get a different bound from the induction case and thus the final calculation turns into

$$\sum_{Y \neq \emptyset} (3pt)^{s-|Y|} \left(\frac{2p}{1+p}\right)^{|Y|} = (3pt)^s \left(\left(1 + \frac{2}{3t(1+p)}\right)^{|Y|} - 1\right) \tag{3.4}$$

$$\leq (3pt)^s \left(e^{2/3} - 1\right) \leq (3pt)^s. \tag{3.5}$$

□

As there are 2^s different τ, Lemmas 3.5 and 3.6 together give the upper bound $(6pt)^s$ for the probability bounded by $(5pt)^s$ by Lemma 3.3. The main reason for the slightly worse bound is that we did not use the fact that giving the values 0 to all variables in Y forces f to zero and hence cannot suggest that f has a decision tree of depth s when $s > |Y|$.

Let us state a simple lemma that will be useful in handling conditioning. It essentially states that Prod has a downward closure property.

LEMMA 3.7. *Suppose τ suggests that $f\lceil_\rho$ requires a decision tree of depth s and that $Prod(\rho, f, \tau) = T$. Let ρ' and $i_0 \notin T$ be such that $\rho'(x_i) = \rho(x_i)$ for $i \neq i_0$ and that $\rho(x_{i_o}) = *$. Then τ suggests that $f\lceil_{\rho'}$ requires a decision tree of depth s and $Prod(\rho', f, \tau) = T$.*

Proof. Any clause forced to 1 by ρ is also forced to 1 by ρ'. The clauses not forced to 1 by ρ and inspected during DT-create(f, ρ) using advice τ cannot contain x_{i_0} as i_0 does not belong to T and all the other variables occurring in these clauses are fixed by ρ, while $\rho(x_{i_0}) = *$. From this it follows that T is produced also by DT-create(f, ρ') under advice τ. □

Remark. The argument in the above proof might sound robust but let us point out a rather fragile point. A conditioning of the form "DT-create(f, ρ) does not find a τ that suggests that $f\lceil_\rho$ requires a decision tree of depth at most s" does not give a downward closed set. To see this, look at the depth-2 function

$$(x_1 \vee x_2) \wedge (x_2 \vee x_3 \vee x_4) \wedge (\bar{x}_2 \vee x_5 \vee x_6)$$

and let ρ be the restriction that gives $*$ to all variables while ρ' sets $\rho'(x_1) = 1$. Then while processing ρ one constructs a decision tree of depth 4 (first x_1 and x_2, and then two more variables depending on the value of x_2). On the other hand, while processing ρ' there is no need to query x_2 at first since the first clause is true independent of the value of x_2 and thus one ends up with a decision tree of depth 5. This is not in violation of the above lemma as ρ and ρ' take different values on x_1 and 1 appears in any T produced on any advice τ.

After this detour let us return to the main path and state our main lemma.

LEMMA 3.8. *Let $(f_i)_{i=1}^S$ be a collection of depth-2 circuits each of bottom fanin t and let s be a parameter satisfying $2^s \geq 2S$. Let $\mathcal{F}$ be a downward closed set of restrictions and ρ a random restriction from R_p. Then the probability that $(f_i\lceil_\rho)_{i=1}^S$*

is not computable by a common s-partial decision tree of depth m is at most $S(24pt)^m$. This statement is true conditioned on $\rho \in \mathcal{F}$.

Proof. The idea of the proof is to follow the proof of Lemma 3.3 (or to be more precise, its variant proving Lemma 3.6 where we also have an advice τ), and whenever we run into trouble, we query the offending variables in the decision tree. Let us start the formal proof. All probabilities below are conditioned on $\rho \in F$ but we leave this conditioning implicit to give shorter formulas.

We prove the lemma by induction over S and the number of variables n. Clearly the lemma is true if either of these numbers is 0. Similarly to the proof of Lemma 3.3 we divide the analysis into two cases depending on whether $f_1\lceil_\rho$ is computable by a decision tree of depth s. Let us first discuss the case when this is indeed true.

The set of restrictions such that $f_1\lceil_\rho$ is computable by a decision tree of depth at most s is obviously downward closed as changing ρ from $*$ to a non-$*$ value on any input maintains the property that $f_1\lceil_\rho$ can be computed by a decision tree of a given depth. Now we apply the inductive version of the lemma to $(f_i)_{i=2}^S$ and $\mathcal{F}$ replaced with the subset of $\mathcal{F}$ which has the property that $depth(f_1\lceil_\rho) \leq s$ (which is a downward closed set by Lemma 2.2). A common s-partial decision tree for $(f_i\lceil_\rho)_{i=2}^S$ is clearly such a decision tree even if we include $f_1\lceil_\rho$ and thus the probability of needing more than depth m and being in this case is, by induction, bounded by $(S-1)(24pt)^m$.

Now let us look at the more interesting case that $f_1\lceil_\rho$ cannot be computed by a decision tree of depth s. Let us use f as shorthand for the collection of functions $(f_i)_{i=1}^S$ and let $s\text{-cdt}(f) \geq m$ denote the event that the common s-partial decision tree of the collection requires depth at least m.

In this case, by Lemma 3.5, there exists at least one $\tau \in \{0,1\}^s$ that suggests that f_1 requires a decision tree of depth s' for some $s' \geq s$. Thus in this case we can estimate the probability of the event of the lemma from above by

$$\sum_{s' \geq s} \sum_{\tau \in \{0,1\}^s} \Pr[s\text{-cdt}(f\lceil_\rho) \geq m \mid \text{Sugg}(\tau, f_1, \rho) = s'] \Pr[\text{Sugg}(\tau, f_1, \rho) = s'].$$

If $\text{Sugg}(\tau, f_1, \rho) = s'$, then there is a unique T of size s' such that $\text{Prod}(\tau, f_1, \rho) = T$. Let us first assume that $s' \leq m$ and later verify that the estimate obtained is true also when $s' > m$. If $f\lceil_\rho$ does not have a common s-partial decision tree of depth m, then there must be some assignment π on T such that $f\lceil_{\rho\pi}$ does not have a common s-partial decision tree of depth $m - s'$. This implies that we can estimate the probability as

$$\sum_{s' \geq s} \sum_{\tau, \pi} \Pr[s\text{-cdt}(f\lceil_{\rho\pi}) \geq m - s' \mid \text{Sugg}(\tau, f_1, \rho) = s'] \Pr[\text{Sugg}(\tau, f, \rho) = s'],$$

where the sum is over $\tau \in \{0,1\}^s$ and $\pi \in \{0,1\}^{s'}$. Now, by Lemma 3.6, we know that the second probability is bounded by $(3pt)^{s'}$. We claim that the first probability is, by induction, bounded by $(24pt)^{m-s'}$, and this follows if we can establish that conditioning is of the correct form, i.e., if it gives a downward closed set.

By Lemma 2.2 we only need to check that the newly introduced conditioning is of the correct form. As π sets all variables in T we only need to check downward closure when changing variables outside T. For those variables, this property follows by Lemma 3.7 and thus we can apply induction to each term. As we have $2^{s+s'}$ terms we get the overall bound

$$S(24pt)^{m-s'} 2^{s+s'} (3pt)^{s'} = S 2^{s-2s'} (24pt)^m.$$

Let us first note that this estimate is also true in the case when $s' > m$ as we then have the total estimate $2^{s+s'}(3pt)^{s'} = 2^{s-2s'}(24pt)^{s'}$ using only Lemma 3.5. Now, summing over $s' \geq s$ and using $2^s \geq 2S$ we see that the probability of this case is bounded by $(24pt)^m$, and adding this probability to the probability $(S-1)(24pt)^m$ obtained in the first case finishes the proof. □

Let us finally wrap up the proof of Theorem 3.1.

Proof of Theorem 3.1. We prove the result by induction over d. Let us see how to establish the base case $d = 2$ directly from Lemma 3.3. Take a depth-2 circuit of bottom fanin t and apply a restriction with $p = \frac{1}{10t}$. In this situation, except with probability $2^{-\Omega(n/t)}$ we have $pn/2$ variables remaining and the resulting function is computed by a decision tree of depth strictly less than $pn/2$. In this case the restricted function has no correlation with parity. In other cases the correlation is at most 1 and the result follows.

For the induction step, let $(f_i)_{i=1}^k$ with $k \leq S$ be the subcircuits of depth 2 appearing in C and apply a restriction with $p = \frac{1}{48t}$. We see that, except with probability $2^{-pn/4}$, this collection of functions can be computed by a common $(1 + \log S)$-partial decision tree of depth at most $pn/4$. It is also the case that except with probability $2^{-\Omega(pn)}$, at least $pn/2$ variables are given the value $*$ by ρ.

This implies that, with probability $1-2^{-\Omega(pn)}$, at any leaf of this common $1+\log S$-partial decision tree, the restriction of f can be computed by a depth $d-1$ circuit of bottom fanin at most $1 + \log S$. By the induction hypothesis the correlation of such a function with parity of the remaining variables (which are at least $pn/4$) is bounded by

$$2^{\Omega(-pn/\log S(\log S)^{d-3})},$$

and the theorem follows. □

Acknowledgments. I thank Andrej Bogdanov for reminding me that one has to be careful with conditioning. The input from two anonymous referees was also essential in helping me make the proof crisper. As this research was initiated and partially completed during a workshop at Schloss Dagstuhl, I want to thank Schloss Dagstuhl for creating a great environment for research and discussions.

REFERENCES

[1] M. Ajtai, Σ_1^1-*formulae on finite structures*, Ann. Pure Appl. Logic, 24 (1983), pp. 1–48.

[2] P. Beame, R. Impagliazzo, and S. Srinivasan, *Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#AC^0SAT$*, in Proceedings of the IEEE Conference on Computational Complexity, 2012, pp. 117–125.

[3] J.-Y. Cai, *With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy*, J. Comput. System Sci., 38 (1989), pp. 68–85.

[4] M. Furst, J. B. Saxe, and M. Sipser, *Parity, circuits and the polynomial-time hierarchy*, Math. Systems Theory, 17 (1984), pp. 13–27.

[5] J. Håstad, *Almost optimal lower bounds for small depth circuits*, in Proceedings of the 18th Annual ACM Symposium on Theory of Computing, STOC '86, New York, ACM, 1986, pp. 6–20.

[6] R. Impagliazzo, W. Matthews, and R. Paturi, *A satisfiability algorithm for AC0*, in Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12, 2012, pp. 961–972.

[7] A. Razborov, *Bounded-depth formulae over the basis {AND,XOR} and some combinatorial problems* (in Russian), Probl. Cybernetics Complexity Theory Appl. Math. Logic (1988), pp. 149–166.

[8] A. Razborov and S. Rudich, *Natural proofs*, J. Comput. System Sci., 55 (1997), pp. 24–35.

[9] M. Sipser, *Borel sets and circuit complexity*, in Proceedings of the 15th Annual ACM Symposium on Theory of Computing, STOC '83, New York, ACM, 1983, pp. 61–69.
[10] R. Smolensky, *Algebraic methods in the theory of lower bounds for boolean circuit complexity*, in Proceedings of the 19th Annual ACM Symposium on Theory of Computing, STOC '87, New York, ACM, 1987, pp. 77–82.
[11] A. C-C. Yao, *Separating the polynomial-time hierarchy by oracles*, in Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science, FOCS '85, 1985, pp. 1–10.