Electronic Colloquium on Computational Complexity, Report No. 137 (2012)

On the correlation of parity and small-depth circuits

Johan Håstad[*]
KTH - Royal Institute of Technology

November 1, 2012

Abstract

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

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 [3], Sipser [8], Ajtai [1] Yao [10], Håstad, [4], Smolensky [9] and Razborov [6] seemed to be a promising road of establishing such lower but came to a almost complete halt in late 1980'ies. One possible explanation for this situation is given by the concept of "Natural proofs" introduced by Razborov and Rudish [7] but a new generation of researchers might find different techniques making what now seems impossible be feasible.

Possibly the simplest non-trivial 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 [3] 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 [10] and Håstad [4] 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

[*]Research supported by ERC Advanced grant 226 203.

ISSN 1433-8092

possible 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 chose a restriction such that one can remove one level of the circuit. This is achieved by applying the switching lemma of Håstad [4] 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 than half of the inputs. If this fraction of agreement 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 rather 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 non-trivial 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 one but whenever it outputs zero 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 [5]. The paper by Impagliazzo et al has as 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, considerably simpler argument. The bounds obtained by [5] are in most cases identical (apart from some involved constants) to ours but for linear size circuits 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 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 as if we have two layers of the same type we can directly move the inputs of the lower gate 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 fanout is arbitrary.

We study the sub-circuits 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.

We let $p \in [0,1]$ be a real number and a random restriction from the space R_p is defined by, for each variable x_i independently, keeping it as a variable with probability p and otherwise setting it to one of the 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 set of restrictions that are monotone in ρ in the sense that fixing the value of more variables can only make ρ more likely belong to the set.

Definition 2.1 *A set $\mathcal{F}$ of restriction 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 [4] 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 zero or one).

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 $(g_i)_{i=1}^m$ 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 τ that gives values to the queried variables. The claim is that $g_i\lceil_\tau$ 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 and which contains at most S gates at distance at least 2 from the inputs. Then the correlation of f with parity is bounded by*

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

Note that if $t = \omega(\log S)$ then replacing d by $d + 1$ and setting $t = 1$ gives a stronger bound as compared to applying the theorem directly.

As discussed in the introduction, the proof is very much based on the proof of the switching lemma of [4] and let us start be 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 was not stated.

Lemma 3.2 *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 the decision tree computing $f\lceil_\rho$. Then*

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

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.

We divide the analysis into two cases depending on whether C_1 is forced to one or not. 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

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

and

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

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 term (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 time being 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 τ be an assignment to the variables in Y. We can now bound (1) as

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

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

Lemma 3.3 *If Y is a set of set 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. If ρ' is changed to a 0 for k different inputs then

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

and as

$$\sum_{i=0}^{r} \binom{r}{i} x^i = (1+x)^r, \tag{2}$$

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

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

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 τ). Changing the value of ρ on any variable not contained in C_1 from $*$ to a non-$*$ value cannot violate any of the conditions.

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

$$\sum_{Y} (2^{|Y|} - 1)(5pt)^{s-|Y|} \left(\frac{2p}{1+p}\right)^{|Y|}. \tag{3}$$

Before ending the proof let us observe that this bound is correct also in the case $r \geq s$, 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 3.3, bounded by $\left(\frac{2p}{1+p}\right)^r$, which is smaller than the corresponding term in (3).

Using (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. ■

In the current proof we will need an extension of the previous lemma where the failure probability is much smaller but of course the conclusion is weaker.

Lemma 3.4 *Let $(f_i)_{i=1}^S$ be a collection of depth-2 circuit each of bottom fanin t and let s be a parameter satisfying $2^s \geq S$. 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 $S(20pt)^m$. This statement is true conditioned on $\rho \in \mathcal{F}$.*

Proof: The idea of the proof is to follow the proof of the previous lemma and whenever we run into trouble, we query the offending variables in the decision tree. Let us start the formal proof.

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 3.2 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 $(S-1)(20pt)^m$.

Now let us look at the more interesting case that $f_1\lceil_\rho$ cannot be computed by a decision tree of depth s. In particular, the construction of the decision tree as given in the proof of Lemma 3.2 must come to a point when the number of variables queried is at least s.

Inside that proof we have a partially constructed decision tree where some variables T are queried on the current path (and assigned the values given by some restriction τ) and we have identified a set of variables Y given the value $*$ by ρ. Let us set $T_0 = Y \cup T$ and define s' as the size of this set, i.e., $|T_0| = s' \geq s$.

We query the variables in T_0 in our common s-partial decision tree that we are constructing. In this way we get $2^{s'}$ nodes in that partially constructed tree, indexed by assignments τ_j to the variables in T_0. We want to apply the induction hypothesis to each of the families $(f_i\lceil_{\tau_j})_{i=1}^S$, when considering restrictions ρ on the variables outside T_0. The key for the proof is the verify that the obtained conditioning is of the correct form and hence that we can apply the induction hypothesis.

The conditioning we want to apply is the original $\rho \in \mathcal{F}$, that $\rho(T_0) = *$, and that during the creation of the decision tree in the proof of Lemma 3.2 we reach the node given by a set T, values τ on this set, and a set Y.

We claim that these conditions give a downward closed set when considering restrictions taking values outside T_0. This is obvious for the two first conditions and let us be careful with the third condition. Suppose ρ satisfies this condition and let i_0 the the index such that $\rho'(x_i) = \rho(x_i)$ when $i \neq i_o$ and $\rho(x_{i_0}) = *$.

Remember that i_0 does not belong to T_0. In the proof of Lemma 3.2 we first encounter a clause that is not forced to 1 by ρ and we query these variables in the decision tree (and these variables are put into T and hence, in particular none of these variables is x_{i_0}). Consider the path of the decision tree giving the values defined by τ (both in the decision trees constructed under ρ and ρ'). These constructions continue on parallel paths finding the next clause which is not forced to one. It is easy to see that also this clause does not contain x_{i_0} and the process continues in identical ways for ρ and ρ' until we have found all of T and the set Y. We conclude that also this third conditioning is of the correct form.

We have thus queried s' variables in our common s-partial decision tree and the probability that any path in this tree needs to query m variables is bounded by the probability that in any of the $2^{s'}$ nodes, defined by the assignments τ_j, we need to query $m - s'$ additional variables. As we have just verified that we can apply induction (on restrictions outside the set T_0) we conclude that the probability of this happening, at any fixed node, is bounded by $S(20pt)^{m-s'}$.

Be Lemma 3.2 the probability that $depth(f\lceil_\rho) \geq s'$ is at most $(5pt)^{s'}$ and thus the contribution to the overall failure probability of this case is bounded by

$$(5pt)^{s'} 2^{s'} S(20pt)^{m-s'} \leq 2^{-s'} S(20pt)^m \leq (20pt)^m,$$

where we used the assumption that $S \leq 2^s$. Adding this probability to the probability $(S-1)(20pt)^m$ obtained in the first case, finishes the proof. ■

Remark. The argument in the above proof might sound robust but let us point out a rather fragile point. A conditioning of the form "The process in proof of Lemma 3.2 gives a decision tree of depth at least 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 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. After this remark let us return to 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 3.2. Take a depth two 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 sub-circuits of depth 2 appearing in C and apply a restriction with $p = \frac{1}{40t}$. We see that except with probability $2^{-pn/4}$ this collection of functions can be computed by a common $(\log S)$-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 $\log S$-partial decision tree, the restriction of f can be computed by a depth $d-1$ circuit of bottom fanin at most $\log S$. 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. ■

Acknowledgement. I am grateful to Andrej Bogdanov for pointing reminding my that one has to be careful with conditioning. I am also grateful to Schloss Dagstuhl for the great environment and the possibility to do this research.

References

[1] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[2] P. Beame, R. Impagliazzo, and S. Srinivasen. Approximating AC^0 by small height decision tress and a deterministic algorithm for $\#AC^0SAT$. In *Proceedings Computational Complexity*, 2012.

[3] M. Furst, J.B. Saxe, and M. Sipser. Parity, circuits and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17:13–27, 1984.

[4] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, STOC '86, pages 6–20, New York, NY, USA, 1986. ACM.

[5] R. Impagliazzo, W. Matthews, and R. Paturi. A satisfiability algorithm for AC^0. In *SODA*, pages 961–972, 2012.

[6] A. Razborov. Bounded-depth formulae over the basis { AND,XOR} and some combintorial problems (in russian). *Problems of Cybernetics. Complexity Theory and Applied Mathematical Logic*, pages 149–166, 1988.

[7] A. Razborov and S. Rudich. Natural proofs. *Journal of Computer and System Science*, 55:24–35, 1997.

[8] M. Sipser. Borel sets and circuit complexity. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, STOC '83, pages 61–69, New York, NY, USA, 1983. ACM.

[9] 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*, STOC '87, pages 77–82, New York, NY, USA, 1987. ACM.

[10] A. C-C. Yao. Separating the polynomial-time hierarchy by oracles. In *Foundations of Computer Science, 1985., 26th Annual Symposium on*, pages 1 –10, oct. 1985.