LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TM-286

IMPROVEMENTS OF YAO'S RESULTS ON PARITY CIRCUITS

JOHAN HASTAD

SEPTEMBER 1985

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

Improvements of Yao's Results on Parity Circuits

Johan Hastad*

September 13

1. Introduction

Proving lower bounds in various computational models is one of the most interesting branches of theoretical computer science. The number of good results is quite limited but at least in the case of bounded depth circuits there has been some progress. Much of the work has been done proving lower bounds for parity circuits, although other functions has been considered as well. The first results by Furst, Saxe and Sipser [FSS] gave among other things $\Omega(n^{\log^{(3(k-2))} n})$ lower bound for depth k parity circuits ($\log^{(i)} n$ denotes the logarithm function iterated i times). This was improved by Ajtai to give lower bound $n^{c_k \log n}$ [A]. Boppana considered the case of the majority function and was able to obtain exponential ($2^{\frac{1}{6} n^{\frac{1}{k-1}}}$) lower bounds. However he restricted the circuits to be monotone. Sipser in [S] was able to give explicit functions computable by depth k circuits of polynomial size but required superpolynomial size circuits if the depth was restricted to $k-1$.

Furst, Saxe and Sipser proved in [FSS] that exponential lower bounds for the parity function would imply the existence of an oracle separating polynomial space from the polynomial hierarchy. In [S] Sipser proved the corresponding theorem that exponential lower bounds would imply existence of oracles separating the different levels in the polynomial hierarchy. These bounds were finally obtained by Yao [Y] in his outstanding contribution. Yao obtained the size lower bound $\Omega(2^{n^{\frac{1}{4k}}})$ for depth k circuits computing parity. By his methods it is possible to decrease the number 4 to any constant larger than 2. It might be possible to even get the value 2. He gives the corresponding result for the functions considered in [S].

In this paper we will improve his lower bound to prove that there is no parity cicuits of depth k and size $2^{(\frac{1}{10})^{\frac{k}{k-1}} n^{\frac{1}{k-1}}}$. Our results also implies that polynomial size parity circuits must be of depth at least $\frac{\log n}{\log \log n}$.

The methods used in our paper are almost identical to the methods used by Yao [Y]. By a closer analysis we are able to prove that certain formula have only short minterms, while Yao was only able to prove that the formula was well approximated by its small minterms. This result gives significant simplifications to the rest of the proof. We are also able to get by with less severe restrictions than the restrictions Yao uses. Our improved bounds will also apply to functions considered in [S].

We note that the lower bounds given in this paper are quite tight as there are known constructions of parity circuits of size $2^{n^{\frac{1}{k-1}}}$ which has bottom fanin $n^{\frac{1}{k-1}}$.

---

* Supported by an IBM fellowship, partially supported by NSF grant DCR-8509905. The author is presently at MIT but some of the work was done while visiting AT&T Bell Laboratories.

2. Main Lemma

The main tool in proving all the above mentioned result is the concept of a random restriction first introduced in [FSS]. A random restriction assigns values to some of the variables and leaves the other variables alone. The fact that some variables will be fixed will allow us to simplify our circuits and we can get an induction going. The values of a restrictions will be 0,1 and *. 0 and 1 means that we give this value to the variable and * means that it stays a live variable. We will denote a random restriction by ρ. A restriction of type R_p gives a variable the value * with probability p and the values 0 and 1 with probability $\frac{1}{2} - \frac{p}{2}$. Given a restriction ρ and an arbitrary function G on n variables we will denote by $G|_\rho$ the induced function obtained by substituting 0 and 1 for the variables given these values by ρ. $G|_\rho$ will hence be a function on the variables given the value * by ρ.

We will be working with Boolean formulas. We will be writing AND's as products and OR's as sums. A minterm for a function is a minimal assignment that forces the function to be 1.

There are two versions of the proof of the main lemma which are almost identical except for notation. The original proof was in terms of a labeling algorithm. The present version of the proof avoiding the use of such an algorithm was proposed by Ravi Boppana.

For notational convenience let E_s denote the event that $G|_\rho$ has a minterm of size at least s.

Main Lemma: Let $G = \prod_{i=1}^{w} G_i$, where G_i are OR's of fanin $\leq t$. Let F be an arbitrary function. Let ρ be a random restriction in R_p. Then we have

$$Pr[E_s \mid F|_\rho \equiv 1] \leq \alpha^s$$

where α is the unique positive root of the equation

$$(1 + \frac{4p}{1+p}\frac{1}{\alpha})^t = (1 + \frac{2p}{1+p}\frac{1}{\alpha})^t + 1$$

Remark 1 An elementary argument shows that $\alpha \approx \frac{2pt}{\ln \phi} < 5pt$, where ϕ is the golden ratio.
Remark 2 If there is no assignment satisfying the condition $F|_\rho \equiv 1$ we will use the convention that the conditional probability in question is 0.

Proof: We will prove the main lemma by induction on w. If $w = 0$ the lemma is obvious ($G \equiv 1$). Suppose now that the statement is true for all values less than w. We will show that it is true for w. We have

$$Pr[E_s \mid F|_\rho \equiv 1] \leq max(Pr[E_s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \equiv 1], Pr[E_s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1])$$

The first term is

$$Pr[E_s|(F \wedge G_1)|_\rho \equiv 1]$$

However in this case E_s is equivalent to saying that $\prod_{i=2}^{w} G_i|_\rho$ has a minterm of size at least s. But this probability is $\leq \alpha^s$ by the inductive hypothesis since we are talking about a product of size $w - 1$.

Now consider the second term. By interchanging x_i and $\bar{x}_i$ we can assume that G_1 is an OR of only positive literals, i.e.

$$G_1 = \sum_{i \in T} x_i$$

where $|T| \leq t$. Let $\rho = \rho_1 \rho_2$, where ρ_1 is the restriction of the variables in T and ρ_2 is the restriction to the other variables. In the case corresponding to the second term we know that G_1 is not made true by the restriction. In this case G_1 has to be made true by some assignment of every minterm of $G|_\rho$. We will partition the minterms of $G|_\rho$ according to the subset of T is contained in the minterm. We will call a typical such subset Y. Observe that all the variables in Y must be given the value * by ρ_1. Now the second term in the max can be estimated by

$$\sum_{Y \subseteq T\, Y \neq \emptyset} Pr[\rho_1(Y) = * \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \times Pr[E_s^{Y,T} \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1 \wedge \rho_1(Y) = *]$$

Here the notation $E_s^{Y,T}$ means that $G|_\rho$ has a minterm of size at least s whose restriction to the variables in T assigns values to precisely those variables in Y.

Let us estimate the first factor (i.e. $Pr[\rho_1(Y) = *|\ldots]$). First we investigate which assignments satisfy the conditions $F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1$ and how this might effect the probability of a set of variables taking the value * under ρ. The important lemma is:

Lemma 1: Let $i \in Y$. Then if an assignment ρ satisfies the condition $F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1$ and has $\rho(x_i) = *$ then the corresponding assignment $\bar{\rho}$ where the only difference is that $\bar{\rho}(x_i) = 0$ also satisfies the condition.

Proof: The condition $G_1|_\rho \not\equiv 1$ obviously presents no problem. The other condition $F|_\rho \equiv 1$ is also easily verfied since the fact that $F|_\rho \equiv 1$ and $\rho(x_i) = *$ implies that changing the value of x_i cannot change the value of $F|_\rho$.

Next we prove

Lemma 2 $Pr[\rho(Y) = * \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \leq (\frac{2p}{1+p})^{|Y|}$

Proof: If we did not have the condition $F|_\rho \equiv 1$ the proof is quite straightforward. The condition that $G_1|_\rho \not\equiv 1$ implies that ρ does not assign the value 1 to any of the variables in T. Conditioning upon this, the probability that each individual variable is 0 is $\frac{1-p}{1+p}$ and the probability of being * is $\frac{2p}{1+p}$. The presence of the condition $F|\rho \equiv 1$ is dealt with by Lemma 1. Loosely speaking Lemma 1 tells us that this condition can only make the event we are interested in less probable. Let us make this formal. By definition of conditional probability we want to prove

$$\frac{\sum'_{\rho(Y)=*} Pr(\rho)}{\sum' Pr(\rho)} \leq (\frac{2p}{1+p})^{|Y|}$$

Here the $'$ indicate that we are only summing over ρ satisfying the condition $F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1$. Remember that if this quotient is of the form $\frac{0}{0}$ we have the convention that it takes the value

0. Now observe that if we have ρ giving a nonzero contribution to the numerator we have by Lemma 1 contributions in the denominator from all the possible ρ obtained by changing arbitrary stars of Y to zeros. Calculation now shows that this contribution is a factor $(\frac{1+p}{2p})^{|Y|}$ larger. This proves the lemma.

Next we try to estimate the other factor. Namely

$$Pr[E_s^{Y,T} \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1 \wedge \rho_1(Y) = *]$$

We partition this probability according to what values σ a potentially large minterm takes on the set Y. To get the probability into the right form to apply the induction hypothesis we need some further work because the second condition $G_1|_\rho \not\equiv 1$ is of the wrong form. To get rid of this condition we maximize over all possible ρ_1 satisfying the third condition. Thus we get:

$$\sum_{\sigma \in \{0,1\}^{|Y|}\, \sigma \neq 0^{|Y|}} [\max_{\rho_1 \in \{0,*\}^{T-Y}} Pr[E_s^{Y,T} \mid F|_{\rho_2 \sigma \rho_1}] \equiv 1]$$

The two last conditions have disappeared because σ and ρ_1 take care of them. The only thing to comment on is how to substitute the stars. In F we substitute the stars of ρ_1 by taking and of the two formulas resulting by substituting 0 and 1. In G we can just erase the stars because they will never help make $G|_\rho$ true. Now the probability is taken care by the induction hyptohesis. The set of variables is now restricted to be outside T and the substituted formula must have a minterm of size at least $s - |Y|$ on these variables. The probability for this is bounded by $\alpha^{s-|Y|}$. Thus we get the bound $(2^{|Y|} - 1)\alpha^{s-|Y|}$.

Finally we must evaluate the sum

$$\sum_{Y \subseteq T} (\frac{2p}{1+p})^{|Y|}(2^{|Y|} - 1)\alpha^{s-|Y|} = \alpha^s \sum_{i=0}^{|T|} \binom{|T|}{i} [(\frac{4p}{1+p}\frac{1}{\alpha})^i - (\frac{2p}{1+p}\frac{1}{\alpha})^i] =$$

$$\alpha^s((1 + \frac{4p}{1+p}\frac{1}{\alpha})^{|T|} - (1 + \frac{2p}{1+p}\frac{1}{\alpha})^{|T|}) \leq \alpha^s((1 + \frac{4p}{1+p}\frac{1}{\alpha})^t - (1 + \frac{2p}{1+p}\frac{1}{\alpha})^t = \alpha^s$$

The last equality follows by the definition of α. This finishes the induction step and the proof of the main Lemma.

3. Lower bounds for parity circuits

Using the main lemma we will prove:

Theorem 1 Parity cannot be computed by a depth k circuit containing $2^{\frac{1}{10}n^{\frac{1}{k-1}}}$ subcircuits of depth at least 2 and bottom fanin $\frac{1}{10}n^{\frac{1}{k-1}}$.

Remark: Observe that the theorem is optimal except for the constant $\frac{1}{10}$.

Proof: We will prove the theorem by induction over k. The base case $k = 2$ is well known. For the induction step we will use the normal way of applying a restriction and use our main Lemma.

Apply a random restriction from R_p with $p = n^{-\frac{1}{k-1}}$. Assume without loss of generality that the depth two circuits in our circuit are AND's of OR's. Then by our lemma that if we look at any such depth two subcircuit the corresponding function will now have no minterm of size $\geq s$ with probability $1 - 2^{-s}$. But this means that it can be written as an OR of AND's of size $\leq s$. Thus if we choose $s = \frac{1}{10}n^{\frac{1}{k-1}}$ we have a good probability that we can interchange the order of AND and OR in all depth 2 subcicuits and still have bottom fanin bounded by s. Observe that this gives us two adjacent levels of OR's which can be collapsed to decrease the depth of the circuit to $k-1$. The number of remaining variables is expected to be $n^{\frac{k-2}{k-1}}$ and with probability greater than $\frac{1}{3}$ we will get at least this number. If we denote this number by m we see that we get precisely a circuit which is certified not to exist by the induction hypothesis.

Theorem: There are no depth k parity circuits of size $2^{(\frac{1}{10})\frac{k}{k-1}n^{\frac{1}{k-1}}}$.

Treat the formula as a depth $k+1$ formula with bottom fanin 1. Hit it with a restriction from R_p using $p = \frac{1}{10}$ and we get a circuit which does not exist by the previos theorem.

Since there are no constants depending on k hidden in the theorem we get the following corollary

Corollary: Polynomial size parity circuits must have depth at least $\frac{\log n}{\log\log n + c}$ for some constant c.

It is not clear if this can be obtained from [Y].

Observe that we have used very little about parity. Only the lower bound for $k = 2$ and the fact that it behaves well with respect to restrictions. Thus we will be able to improve lower bounds for sizes of small depth circuits for other functions using our main lemma. Also observe that by using better bounds for α we can reduce the constant 10.

Acknowledgement I am very grateful to Ravi Boppana for reading an early draft of the paper and suggesting the version of the proof avoiding the labeling algorithm. I am also grateful to several people who have read and commented on drafts of this paper. These people include Ravi Boppana, Oded Goldreich, Shafi Goldwasser and David Schmoys.

References

[A] Ajtai M. "$\sum_1^1$-formulae on finite structures",
Annals of Pure and Applied Logic 24(1983) 1-48

[B] Boppana R. "Threshold functions and bounded depth monotone circuits"
Proceedings of 15th STOC,1984, 475-479.

[FSS] Furst M., Saxe J and Sipser M., "Parity, circuits, and the polynomial times hierarchy"
IEEE FOCS 22(1981) 260-270

[S] Sipser M. "Borel sets and circuit complexity", ACM STOC 16(1984).

[Y] Yao A. "Separating the Polynomial-Time Hierarchy by Oracles"
to appear in IEEE FOCS 26(1985)