Almost Optimal Lower Bounds for Small Depth Circuits

Johan Hastad *

Applied Mathematics department and Laboratory of Computer Science, MIT

Abstract: We give improved lower bounds for the size of small depth circuits computing several functions. In particular we prove almost optimal lower bounds for the size of parity circuits. Further we show that there are functions computable in polynomial size and depth k but requires exponential size when the depth is restricted to $k-1$. Our main lemma which is of independent interest states that by using a random restriction we can convert an AND of small ORs to an OR of small ANDs and conversely.

1. Introduction

Proving lower bounds for the resources needed to compute certain functions is one of the most interesting branches of theoretical computer science. One of the ultimate goal of this branch is of course to show that $NP \neq P$. However, it seems that we are yet quite far from achieving this goal and that new techniques have to be developed before we can make significant progress towards solving this question. To gain understanding of the problem of proving lower bounds and developing techniques, several restricted models of computation have been studied. Recently there have been significant progress in proving lower bounds in two circuit models. The first example is the the case of monotone circuits i.e. circuits just containing AND and OR gates and no negations. Superpolynomial lower bounds were proved for the clique function by Razborov [R] and these were improved to exponential lower bounds by Alon and Boppana [AB]. Andreev [An] independently obtained exponential lower bounds for other NP-functions.

The second model where interesting lower bounds have been proved is the model of ***small depth circuits***. These circuits have the full instruction set of AND, OR and negations and furthermore each AND and OR gate can have arbitrary many inputs. However the depth (the longest path from input to output) is restricted to be small e.g. constant. The unrestricted size of the AND gates is needed to make it possible to compute circuits depending on all inputs. In this paper we will prove exponential lower bounds for this model. Our technique enables us to prove lower bounds for several different functions. Thus we have at least partial understanding of what might cause a function to be difficult to compute in these models of computation.

Finally let us remark that even though the $P \neq NP$ question is one of the motivations to studying the problem of small depth circuits, we do not think that the techniques of this paper will

* Supported by an IBM fellowship, partially supported by NSF grant DCR-8509905. Some of the work was done while the author visited AT&T Bell Laboratories.

help in resolving that question. The results for small depth circuits and monotone circuits only show that it is possible to prove exponential lower bounds in nontrivial cases. This might be taken as a promising sign and encourage us to look for new techniques with renewed optimism.

1.1 Lower bounds for small depth circuits; A crucial Lemma.

The problem of proving lower bounds for small depth circuits has attracted the attention of several researchers in the field. Functions considered have been simple functions like parity and majority. The first superpolynomial lower bounds for the circuits computing parity was obtained by Furst, Saxe and Sipser [FSS]. Ajtai [Aj] independently gave slightly stronger bounds and Yao [Y] proved the first exponential lower bounds. (The case of monotone small depth circuits has been studied by Boppana [B] and Klawe, Paul, Pippenger and Yannakakis [KPPY].)

We will in this paper give almost optimal lower bounds for the size of circuits computing parity. However it is quite likely that the longer lasting contribution will our main lemma. The main lemma is the essential ingredient in the proof and it gives some insight why some problems require large circuits when the depth is small. The lemma tells us that given a depth two circuit, say an AND of small ORs (a gate is small if it has few inputs), then if one gives random values to a randomly selected subset of the variables then it is possible to write the resulting induced function as an OR of small ANDs with very high probability. Let us outline how this can be used to prove lower bounds for circuits computing parity.

Given a circuit of constant depth k computing parity we can give random values to some random inputs. The remaining circuit will still compute parity (or the negation of parity) of the remaining variables. By the virtue of the lemma it is possible to interchange to adjacent levels of ANDs and ORs and by merging the two adjacent levels with the same connective and this way decrease the depth of the circuit to $k-1$. And this can be done without increasing the size of the circuit significantly. An easy induction now gives the result.

The idea of giving random values to some of the variables was first introduced in [FSS] and weaker versions of our main lemma were used in [FSS] and [Y]. In [FSS] the probability of size not increasing to much was not proved to be exponentially small and Yao only proved that the resulting OR of small ANDs was in a technical sense a good approximation of the original function. This fact gave significant complications to the rest of the proof. Also, Yao did not obtain the sharp estimates for the probability of failure. Since we get almost optimal lower bounds for the size of parity circuits our estimates are sharp up to a constant.

1.2 Results obtained.

Our nearly optimal results for the size of parity circuits imply that a polynomial size circuit computing parity has to have depth essentially $\frac{\log n}{\log \log n}$. The best previous lower bounds for the depth of polynomial size parity circuits was $\sqrt{\log n}$ by Ajtai [Aj].

By similar methods it is possible to prove that there is a family of functions f_k^n of n inputs which have linear size circuits of depth k but require exponential size circuits when restricted to depth $k-1$. These functions f_k^n were introduced by Sipser in [S]. Sipser proved superpolynomial lower bounds for the size of the circuits when the depth was restricted to be $k-1$. Yao claimed exponential lower bounds for the same situation.

1.3 Small depth circuits and Relativized Complexity.

Lower bounds for small depth circuits have some interesting applications to relativized complexity. Furst, Saxe and Sipser proved in [FSS] that subexponential lower bounds (more precisely $\Omega(2^{(\log n)^i})$ for all i) for any constant depth k for the parity function would imply the existence of an oracle separating PSPACE from the polyno-

mial time hierarchy. Yao [Y] was the first to prove sufficiently good lower bounds to obtain the separation for an oracle A. Cai [C] extended his methods to prove that a random oracle separated the two complexity classes with probability 1.

In [S] Sipser proved the corresponding theorem that the same lower bounds for the functions f_k^n would imply the existence of oracles separating the different levels in the polynomial hierarchy. The lower bounds claimed by Yao gives the first oracle achieving this separation. Our bounds are of course also sufficient. The question whether a random oracle separates the levels is still open.

1.4 Outline of paper.

In section 3 we prove the main lemma. The necessary background and some motivation are given in section 2. The application to parity circuits is in section 4 and in section 5 we prove the lower bounds for the functions f_k^n and in section 6 we briefly mention some more details of the implications for relativized complexity. Finally in section 7 we mention some related results.

2. Background

2.1 Computational Model

We will be working with unbounded fanin circuits of small depth. A typical example looks like this.

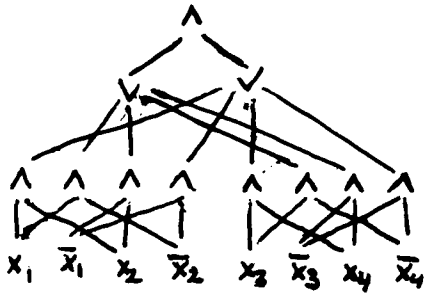

Figure 1

We can assume that the only negations occur as negated input variables. In general if there are negations higher up in the circuit we can move them down to the inputs using DeMorgan's laws. This procedure only doubles the size of the circuit. Observe that we have alternating levels of AND and OR gates since two adjacent gates of the same type can be collapsed into one gate.

The crucial parameters for a circuit is the depth and the size. Depth is defined as the length of the longest path from an input to the output and can also be thought of as the number of levels of gates. For instance the depth of the circuit in figure 1 is 3. Size is defined to be the total number of AND/OR gates and the circuit in figure 1 is of size 11. The fanin of a gate is defined as the number of inputs to it. We put no restriction on the fanin of the gates in our circuits. However we will be interested in the bottom fanin which is defined as the maximum fanin for any gate on the lowest level and hence has variables as inputs.

2.2 Outline of Proof

Many of the cited lower bounds proofs ([FSS],[Y] and the present paper) have the same outline. The proofs are by induction which proceeds as follows.

(1) Prove that parity circuits of depth 2 are large

(2) Prove that small depth k parity circuits can be converted to small depth $k-1$ parity circuits.

Of these two steps the first step is easy and tailored for the parity function. The result is that depth 2 parity circuits are of size 2^{n-1} [FSS]. The second step is much more difficult and here lies the difference between the papers. The basic idea for doing this lies in the fact that every function can be written either as an AND of ORs or as an OR and ANDs.

To give an idea of (2) assume that $k = 3$ and

we have the following depth 3 circuit.

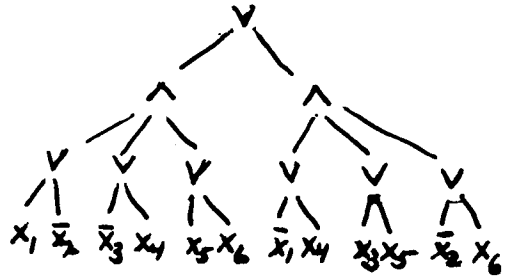

Figure 2

Take any gate at distance two from the inputs. It represents a subcircuit of depth 2. In this case this circuit will be an AND of ORs. Now observe that any function can be written either as an AND or ORs or as and OR of ANDs. Thus we can change this depth 2 circuit to and OR of ANDs which computes the same function. Thus we have the following circuit computing the same function.

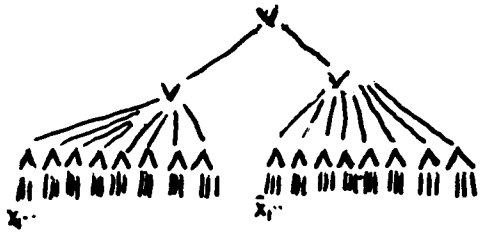

Figure 3

Observe that we have two adjacent levels consisting of OR gates. These two levels can be merged to one level and we get the following circuit of depth 2.

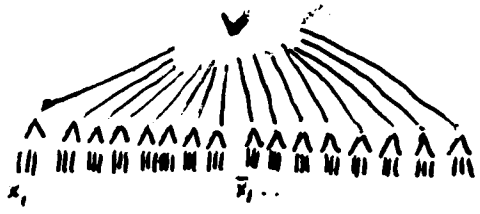

Figure 4

However doing this we run into one problem. When we convert and AND of ORs to an OR of ANDs the size of the circuit will in general increase considerably. Thus we have converted a *small* depth k circuit to a *large* depth $k-1$ circuit and hence we fail to achieve (2).

2.3 Restrictions

The way around this problem was introduced in [FSS] and works as follows. If we assign values to some of the variables we can simplify the circuit. In particular if we assign the value 1 to one of the input of an OR gate we know that the output of the OR gate will be 1 no matter what the other inputs are. In the same way we only need to know that one of the inputs to an AND gate is 0 to decide that it outputs 0. This means that for any specific gate on the bottom level we can force it by assigning a suitable value to one of its inputs. However there are much more gates than inputs and we have to do something more sophisticated. Let us first make formal what we mean by fixing some variables.

Definition: A restriction ρ is a mapping of the variables to the set $\{0, 1, *\}$.
$\rho(x_i) = 0$ means that we substitute the value 0 for x_i
$\rho(x_i) = 1$ means that we substitute 1
$\rho(x_i) = *$ means that x_i remains a variable.

Given a function F we will denote by $F\lceil_\rho$ the function we get by doing the substitutions prescribed by ρ. $F\lceil_\rho$ will be a function of the variables which were given the value *.

Example: Let $F(x_1, x_2, x_3, x_4, x_5)$ = majority of the variables and let $\rho(x_1) = 1, \rho(x_2) = *, \rho(x_3) = *, \rho(x_4) = 1$ and $\rho(x_5) = *$. Then $F\lceil_\rho(x_2, x_3, x_5)$ = at least one of x_2, x_3 and x_5 is 1.

A simple observation which is important to the proof of the result for parity is.

Observation: Parity$\lceil_\rho$= Parity or the negation of Parity.

The idea behind using restrictions is that they

should simplify the circuits we are working with. As pointed out above we could get rid of one gate by giving the value 0 or 1 to one of the variables. As we also noted that if we proceed this way we will run out of variables long before we run out of gates. The way to avoid this is to make more clever assignments serving many purposes simultaneously. To do this explicitly seems hard and our way of avoiding this is to rely on luck. We will pick a random restriction and it will do the job for us.

We will be working with random restrictions with distributions parameterized by a real number p which usually will be small.

Definition: A random restriction $\rho \in R_p$ satisfies
$\rho(x_i) = 0$ with probability $\frac{1}{2} - \frac{p}{2}$
$\rho(x_i) = 1$ with probability $\frac{1}{2} - \frac{p}{2}$
$\rho(x_i) = *$ with probability p.
independently for different x_i.

Observe that we have probability p of keeping a variable as a variable. Thus the expected number of variables remain is pn. Obviously the smaller p is the more we can simplify our circuits but on the other hand we have fewer remaining variables. We have to optimize this trade off when we make a choice of p.

The main improvement of the present paper over previous papers is that we analyze in a better way how much a restriction simplifies a circuit. We will prove a lemma which basically tells us that if we hit a depth two circuit with a random restriction then we can change an AND or ORs to an OR of ANDs without increasing the size. We prove that this fails with only exponentially small probability.

We will need some notation. A *minterm* is a minimal way to make a function 1. We will think of a minterm σ for a function F as a partial assignment with the following two properties.

(1) σ forces F to be true.
(2) No subassignment of σ forces F to be true.

Thus (2) says that σ is minimal satisfying (1).
Example Let $F(x_1, x_2, x_3)$ be the majority function. Then the minterms are σ_1, σ_2 and σ_3 where

$\sigma_1(x_1) = 1, \sigma_1(x_2) = 1, \sigma_1(x_3) = *$
$\sigma_2(x_1) = 1, \sigma_2(x_2) = *, \sigma_2(x_3) = 1$
$\sigma_3(x_1) = *, \sigma_3(x_2) = 1, \sigma_3(x_3) = 1$

The size of a minterm is defined as the number of variables to which it gives either the value 0 or the value 1. All three of the above minterms are of size 2. Observe that it is possible to write a function as an OR of ANDs where the ANDs precisely correspond to its minterms. The size of the ANDs will be the size of the minterms since x_i will be input precisely when $\sigma(x_i) = 1$ and $\overline{x}_i$ will be input precisely when $\sigma(x_i) = 0$.

3. Main Lemma

Our main lemma will tell us that if we apply a restriction we can with high probability convert an AND of ORs to an OR of ANDs. This will provide the tool for us to carry through the outline of the proof described in section 2.

Main Lemma: Let G be an AND of ORs all of size $\leq t$ and ρ a random restriction from R_p. Then the probability that $G\lceil_\rho$ cannot be written as an OR of ANDs all of size $< s$ is bounded by α^s where α is the unique positive root to the equation.

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

Remark 1 Provided that p is $o(1)$, an elementary argument shows that $\alpha \approx \frac{2pt}{\ln \phi} < 5pt$, where ϕ is the golden ratio.
Remark 2 By looking at $\neg G$ one can see that it is possible to convert an OR of ANDs to an AND or ORs with the same probability.
Remark 3 There are two versions of the proof of the main lemma which are almost identical except for notation. Our original proof was in terms of a labeling algorithm used by Yao [Y] in his proof. The present version of the proof, avoiding the use of such an algorithm was proposed by Ravi Boppana.

It turns out that it is easier to prove a slightly

stronger version of the main lemma. First we will require all minterms of $G\lceil_\rho$ to be small. By the remark above above this implies that $G\lceil_\rho$ can be written as an OR of small ANDs. A more significant difference between the main lemma and the stronger lemma we will prove is that we will estimate the probability conditioned upon any function being forced to be 1. The reason for this is that this makes the lemma provable by induction.

For notational convenience let $min(G) \geq s$ denote the event that $G\lceil_\rho$ has a minterm of size at least s.

Stronger Main Lemma Let $G = \wedge_{i=1}^{w} G_i$, 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

$$Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1] \leq \alpha^s$$

Remark 4: The stronger main lemma implies the main lemma by choosing $F \equiv 1$ and the fact that a function has a circuit which is an OR of ANDs corresponding to its minterms.

Remark 5 If there is no restriction ρ satisfying the condition $F\lceil_\rho \equiv 1$ we will use the convention that the conditional probability in question is 0.

Proof: We will prove the stronger main lemma by induction on w the number of ORs in our depth two circuit. A picture of G which is good to keep in mind is the following.

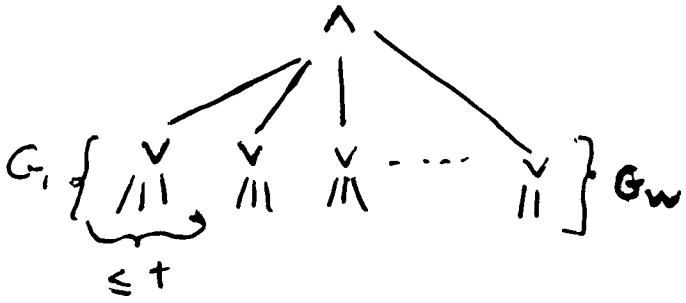

Figure 5

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 will first study what happens to G_1, the first OR in our circuit. We have two possibilities, either it is forced to be 1 or it is not. We will estimate these two probabilities separately. We have

$$Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1] \leq$$
$$\max(Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \equiv 1],$$
$$Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1])$$

The first term is

$$Pr[min(G) \geq s \mid (F \wedge G_1)\lceil_\rho \equiv 1]$$

However in this case $G\lceil_\rho = \wedge_{i=1}^{w} G_i\lceil_\rho = \wedge_{i=2}^{w} G_i\lceil_\rho$ since we are only concerned about ρ's which forces G_1 to be 1. Thus $min(G) \geq s$ is equivalent to saying that $\wedge_{i=2}^{w} G_i\lceil_\rho$ 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 $w-1$. We are conditioning upon another function being 1 but this is OK since we are assuming that the induction hypothesis is true for all F. It is precisely the fact that the conditioning keeps changing that "forced" us to introduce the stronger version of the main lemma.

Now consider the second term ($Pr[min(G) \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1]$). For notational convenience we will assume that G_1 is an OR of only positive literals, i.e.

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

where $|T| \leq t$. We do not lose generality by this since we can interchange x_i and $\bar{x}_i$. Let $\rho = \rho_1\rho_2$, where ρ_1 is the restriction of the variables in T and ρ_2 is the restriction of the other variables. Thus the condition that $G_1\lceil_\rho \not\equiv 1$ is equivalent to that ρ_1 does not take the value 1. Thus it is only a condition on ρ_1 and to remind us of this we will write the condition as $G_1\lceil_{\rho_1} \not\equiv 1$. Since we are now conditioning upon the fact that G_1 is not made true by the restriction, we know that G_1 has to be made true by every minterm of $G\lceil_\rho$ i.e. for every minterm σ there must be an $i \in T$ such that $\sigma(x_i) = 1$. Observe that σ might give values to some other variables in T and that these values might be both 0 and 1. We will partition

the minterms of $G\lceil_\rho$ according to what variables in T they give values to. We will call a typical such subset Y.

The fact that the minterm give values to the variables in Y implies in particular that the variables in Y were left as variables and hence were given the value * by ρ_1. We will denote this fact by the shorthand notation $\rho_1(Y) = *$. Further let $min(G)^Y \geq s$ denote the event that $G\lceil_\rho$ has a minterm of size at least s whose restriction to the variables in T assigns values to precisely those variables in Y. Using this notation we get

$$Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1] \leq$$

$$\sum_{Y \subseteq T, Y \neq \emptyset} Pr[min(G)^Y \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1] =$$

$$\sum_{Y \subseteq T, Y \neq \emptyset} Pr[min(G)^Y \geq s \wedge \rho_1(Y) = * \mid$$

$$F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1] =$$

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

$$\times Pr[min(G)^Y \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1 \wedge \rho_1(Y) = *]$$

The inequality and the first equality follows by the reasoning above and the last equality follows by the definition of conditional probability. Now we will estimate each of the two factors in each term of the above sum. Let us start with the the first factor (i.e. $Pr[\rho_1(Y) = *|\ldots]$).

To make life simpler we will start by ignoring the condition $F\lceil_\rho \equiv 1$.

Lemma 1. $Pr[\rho_1(Y) = * \mid G_1\lceil_{\rho_1} \not\equiv 1] = (\frac{2p}{1+p})^{|Y|}$

Proof: As remarked above the condition $G_1\lceil_{\rho_1} \not\equiv 1$ is precisely equivalent to $\rho_1(x_i) \in \{0, *\}$ for $i \in T$. The induced probabilities are $Pr[\rho(x_i) = 0] = \frac{1-p}{1+p}$ and $Pr[\rho(x_i) = *] = \frac{2p}{1+p}$. The lemma follows since the probabilities are independent. ∎

Now we have to take the condition $F\lceil_\rho \equiv 1$ into account. The intuition for doing this works as follows. The fact that something is determined to be 1 cannot make stars more likely since having a lot of stars is in a vague sense equivalent to making things undetermined. During a presentation of this material Mike Saks found a nice way to make this formal without looking at probabilities of individual restrictions. We first need an elementary fact from probability theory. Let A, B and C be three arbitrary events

Lemma 2. $Pr[A|B \wedge C] \leq Pr[A|C]$ is equivalent to $Pr[B \mid A \wedge C] \leq Pr[B \mid C]$.

This lemma follows from use of definition of conditional probability and trivial algebra. Our final estimate will be

Lemma 3. $Pr[\rho_1(Y) = * | F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \leq (\frac{2p}{1+p})^{|Y|}$

Proof: Let $A = (\rho_1(Y) = *), B = (F\lceil_\rho \equiv 1)$ and $C = (G_1\lceil_{\rho_1} \not\equiv 1)$. By the above lemmas we only have to verify that

$$Pr[F\lceil_\rho \equiv 1 \mid \rho_1(Y) = * \wedge G_1\lceil_{\rho_1} \not\equiv 1] \leq$$

$$Pr[F\lceil_\rho \equiv 1 \mid G_1\lceil_{\rho_1} \not\equiv 1]$$

This is clear from inspection since requiring that some variables are * cannot increase the probability that a function is determined. ∎

Next we estimate the other factor. Namely

$$Pr[min(G)^Y \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1 \wedge \rho_1(Y) = *]$$

To do this think of the minterm as consisting of two parts

(1) A part σ_1 which assign values to the variables of Y.

(2) A part σ_2 which assign values to some variables in the complement $\overline{T}$ of T.

This partition of the minterm is possible since we are assuming that it assign no values to variables in $T-Y$. Observe that σ_2 is a minterm of the function $G\lceil_{\rho\sigma_1}$. This obviously suggests that we can use the induction hypothesis. We only have to get rid of the unpleasant condition that $G_1\lceil_{\rho_1} \not\equiv 1$. This we do by maximizing over all ρ_1 satisfying this condition. We have

$$Pr[min(G)^Y \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1 \wedge \rho_1(Y) = *] \leq$$

$$\sum_{\sigma_1 \in \{0,1\}^{|Y|}\, \sigma_1 \neq 0^{|Y|}} \Big(\max_{\rho_1(Y)=*,\, \rho_1(T) \in \{0,*\}^{|T|}} Pr_{\rho_2}[min(G)^{Y,\sigma_1} \geq s \,|\, (F\lceil_{\rho_1\sigma_1})\lceil_{\rho_1} \equiv 1]\Big)$$

The two last conditions have disappeared because they involve only ρ_1 and we are now interested in a probability over ρ_2. By (2) above we know that $min(G)^{Y,\sigma_1} \geq s$ implies that $(G\lceil_{\rho_1\sigma_1})\lceil_{\rho_2}$ has a minterm of size at least $s - |Y|$ on the variables in $\overline{T}$. Thus we can estimate the probability by $\alpha^{s-|Y|}$ by the induction hypothesis. We need to comment on how to substitute the stars of ρ_1. This is done by taking and of the two formulas resulting by substituting 0 and 1.

To sum up each term in the sum is estimated by $\alpha^{s-|Y|}$ and we have $2^{|Y|}-1$ possible σ_1. This is because σ_1 must make G_1 true and hence cannot be all 0. Thus we get the total bound $(2^{|Y|} - 1)\alpha^{s-|Y|}$.

Finally we must evaluate the sum and since the term corresponding to $Y = \emptyset$ is 0 we can include it.

$$\sum_{Y \subseteq T} (\frac{2p}{1+p})^{|Y|}(2^{|Y|} - 1)\alpha^{s-|Y|} =$$

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

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

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

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

4. Lower bounds for small depth circuits

The first function we will prove lower bounds for is parity. We have

Theorem 1. There are no depth k parity circuits of size $2^{(\frac{1}{10})^{\frac{k}{k-1}} n^{\frac{1}{k-1}}}$ for $n > n_0^k$ for some absolute constant n_0.

Remark: Observe that this is quite close to optimal since it is known that parity can be computed by depth k circuits of size $n2^{n^{\frac{1}{k-1}}}$. The best previous lower bounds were $\Omega(2^{n^{\frac{1}{4k}}})$ by Yao [Y].

As in the case of the main lemma it will be more convenient to first prove something that is more suitable to induction.

Theorem 2. Parity cannot be computed by a depth k circuit containing $\leq 2^{\frac{1}{10}n^{\frac{1}{k-1}}}$ subcircuit of depth at least 2 and bottom fanin $\leq \frac{1}{10}n^{\frac{1}{k-1}}$ for $n > n_0^k$ for some absolute constant n_0.

Proof: We will prove the theorem by induction over k. The base case $k = 2$ follows from the well known fact that depth 2 parity circuits must have bottom fanin n. The induction step will be done as outlined in section 2. We can now with the help of the main lemma make sure that we convert a *small* depth k circuit to a *small* depth $k-1$ circuit.

Suppose without loss of generality that our depth k circuits are such that the gates at distance 2 from the inputs are AND gates and hence represents a depth 2 circuit with bottom fanin bounded by $\frac{1}{10}n^{\frac{1}{k-1}}$. Apply a random restriction from R_p with $p = n^{-\frac{1}{k-1}}$. Then by our lemma every individual depth two subcircuit can be written as and OR of ANDs of size bounded by s with probability $1 - \alpha^s$. By the chosen parameters α is bounded by a constant less than $\frac{1}{2}$. Thus if we

choose $s = \frac{1}{10}n^{\frac{1}{k-1}}$ it is true with probability at least $1 - (2\alpha)^s$ we can interchange the order of AND and OR in all depth 2 subcircuits 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. Thus with nonzero probability we can interchange the order of AND and OR in all depth 2 circuits and we also have at least $n^{\frac{k-2}{k-1}}$ remaining variables. In particular such a restriction exists. Applying this restriction to the circuit gives a depth $k-1$ circuit computing the parity of at least $n^{\frac{k-2}{k-1}} = m$ variables. Further is has bottom fanin bounded by $\frac{1}{10}n^{\frac{1}{k-1}} = \frac{1}{10}m^{\frac{1}{k-2}}$ and the number of gates of depth at least 2 is bounded by $2^{\frac{1}{10}n^{\frac{1}{k-1}}} = 2^{\frac{1}{10}m^{\frac{1}{k-2}}}$. The last fact follows from that a gate of depth at least 2 in the new circuit corresponds to a gate of depth at least three in the old depth k circuit. But this is precisely a circuit which is certified not to exist by the induction hypothesis. The proof of theorem 2 is complete. ∎

Let us now prove theorem 1. Consider the circuit as a depth $k+1$ circuit with bottom fanin 1. Hit it with a restriction from R_p using $p = \frac{1}{10}$ and by using our main lemma with $s = \frac{1}{10}n^{\frac{1}{k-1}}$ we see that we get a circuit which does not exist by theorem 2.

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 $\frac{\log n}{c+\log\log n}$ for some constant c.

Observe that this is tight since for every constant c there are such polynomial size circuits. Since Yao had constants in his theorems it is not clear if a similar corollary 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. Let us do majority

Theorem 3. Majority requires size $2^{(\frac{1}{10})^{\frac{k}{k-1}}n^{\frac{1}{k-1}}}$ depth k circuits for $n > n_0^k$ for some absolute constant n_0.

Proof: To make the proof go through we only need to make two observations. The base case $k = 2$ goes through. Secondly even if we require that the restriction gives out as many 1's as 0's we still have a nonzero probability that a random restriction satisfies all conditions. This requirement ensures that the smaller circuit also computes majority.

In general we do not need that we get back the same function but only that we get a function that is hard to compute. Loosely speaking we can prove the corresponding lower bounds as soon as the function even when hit by severe restriction still have large minterms. We leave the details to the interested reader.

5. Functions requiring depth k to have small circuits.

Sipser defined in [S] a set of functions f_k^m which could be computed in depth k and polynomial size. He also showed that these functions required superpolynomial size for depth $k-1$.

The functions were defined by a depth k circuit as follows:

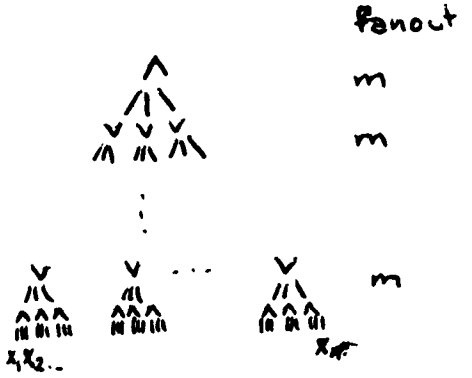

Figure 6

Thus the circuit is a tree with fanout m, depth k and each variable occurs only once. As mentioned in the introduction Yao has claimed exponential lower bounds for these functions. The proofs have not yet appeared but they are supposed to be as complicated as in the case of the parity function. Therefore we include our proofs even though they are not quite optimal. First redefine the functions slightly. Let g_k^m be defined by the following circuit:

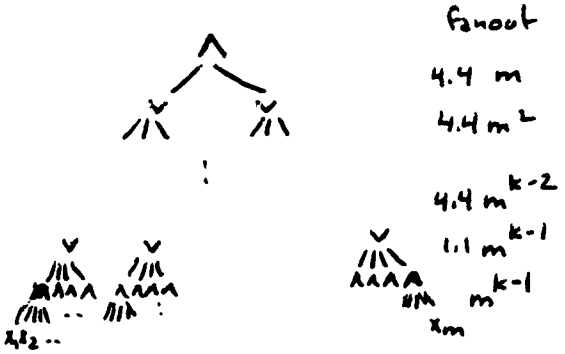

Figure 7

The only difference between f_k^m and g_k^m is thus that the fanouts in the defining tree varies for g_k^m. Observe that g_k^m is a function of $1.1^{k-1}4^{k-2}m^{\frac{k^2+k-2}{2}}$ variables. These functions might seem more complicated than f_k^m but they will simplify the notation in the proofs. Note that f_k^m can be viewed as a restriction of g_k^m and g_k^m appears as a restriction of $f_k^{2m^{k-1}}$ and thus the sizes of the $k-1$ depth circuit will be related. For the g_k^m we have the following theorem.

Theorem 4. Depth $k-1$ circuits computing g_k^m are of size at least $2^{\frac{1}{10}m}$ for $m > m_1$ where m_1 is some absolute constant.

One would like to prove Theorem 4 with the aid of the main lemma. However in this case we run into problems not encountered in the case of the parity function. If one applies a restriction from R_p to either f_k^m or g_k^m the resulting function will with very high probability be a constant function. The reason for this is that the gates at the bottom level are quite wide and with very high probability all gates will be forced. To get around this problem we will define another set of restrictions which will be more suitable to the present functions.

Definition: Let p_1, p_0 and p_* be real numbers satisfying $p_1 + p_0 + p_* = 1$ and $B = (B_i)_{i=1}^r$ a partition of the variables (The B_i are disjoint sets of variables and their union is the set of all variables). Let $R^+_{p_1,p_0,p_*,B}$ be the probability space of restrictions which takes values as follows.

For $\rho \in R^+_{p_1,p_0,p_*,B}$ and every B_i

$\rho(x_j) = 1$ for all $x_j \in B_i$ with probability p_1.

With probability $p_0 + p_*$ choose a random $x_k \in B_i$. Let $\rho(x_j) = 1$ for $j \neq k$ and $\rho(x_k) = 0, *$ with probability $\frac{p_0}{p_0+p_*}$ and $\frac{p_*}{p_0+p_*}$ respectively.

This is done independently for different B_i.

A $R^-_{p_0,p_1,p_*,B}$ probability space of restriction can be defined by interchanging the roles played by 0 and 1.

These sets of restrictions does not assign values independently but they are nice enough so that the proof of our main lemma will go through with some minor modifications. Define q to be $\max(\frac{p_*}{p_0+p_*}, \frac{p_*}{p_1|B_i|+p_*})$.

Lemma 4. Let G be an AND of ORs all of size $\leq t$ and ρ a random restriction from $R^+_{p_0,p_1,p_*,B}$.

Then the probability that $G\lceil_\rho$ cannot be written as an OR of ANDs all of size $< s$ is bounded by α^s where α is the unique positive root to the equation.

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

Remark 6 The same is true with $R^+_{p_0,p_1,p_*,B}$ replaced by $R^-_{p_0,p_1,p_*,B}$.

Remark 7 We have the same probability of converting an OR of ANDs to an AND of ORs.

As in the previous case we will prove a slightly stronger lemma stating that the same is true even conditioning upon something being forced to 1 by the restriction.

Lemma 5: Let $G = \wedge_{i=1}^w G_i$, where G_i are OR's of fanin $\leq t$. Let F be an arbitrary function. Let ρ be a random restriction in $R^+_{p_1,p_0,p_*}$. Then we have

$$Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1] \leq \alpha^s$$

Remark 8: Lemma 5 implies Lemma 4 as the strong main lemma implies the main lemma.

Remark 9: Remember that we have the convention that if there is no restriction ρ satisfying the condition $F\lceil_\rho \equiv 1$ then the conditional probability in question is 0.

Proof: The proof of Lemma 5 will be done the same way as the proof of the stronger main lemma. We therefore only give an outline and all details only in the case when they are different.

As before

$$Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1]$$
$$\leq max(Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \equiv 1],$$
$$Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1])$$

Also here the first term,

$$Pr[min(G\lceil_\rho) \geq s \mid (F \wedge G_1)\lceil_\rho \equiv 1]$$

is taken care of by the induction hypothesis.

The second term, $Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1]$ will be estimated the same way as before. However in this case we cannot assume that G_1 is an OR of only positive literals since the restrictions we are working with are nonsymmetric in 0 and 1. We will still denote the set of variables occurring in G_1 by T and as before $|T| \leq t$. As before we know that G_1 has to be made true by every minterm of $G\lceil_\rho$ and we will partition the minterms of $G\lceil_\rho$ according to what set of variables Y variables in T they give values to.

We get

$$Pr[min(G\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \leq$$
$$\sum_{Y \subseteq T, Y \neq \emptyset} Pr[\rho(Y) = * \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \times$$
$$Pr[min(G\lceil_\rho)^Y \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1 \wedge \rho(Y) = *]$$

We will again estimate the factors in the above sum separately. The way we do this will be slightly different and this is the main difference between the present proof and the proof of the stronger main lemma. Let us start with the first factor (i.e. $Pr[\rho_1(Y) = *|\ldots]$).

First we investigate which restrictions satisfy the conditions $F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1$ and how this might effect the probability of a set of variables taking the value * under ρ. The important lemma is:

Lemma 6: Let $i \in Y$. Then if an assignment ρ satisfies the condition $F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1$ and has $\rho(x_i) = *$ then the corresponding assignments $\tilde{\rho}$ where the only difference is that $\tilde{\rho}(x_i) = 0$ or 1 also satisfies the condition.

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

Next we prove

Lemma 7 $Pr[\rho(Y) = * \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \leq q^{|Y|}$

Proof: The proof of Lemma 7 will use Lemma 6. Loosely speaking Lemma 6 tells us that this condition can only make the event we are interested in less probable. Let us make this formal. By the definition of conditional probability we want to prove

$$\frac{\sum'_{\rho(Y)=*} Pr(\rho)}{\sum' Pr(\rho)} \leq q^{|Y|}$$

Here the $'$ indicate that we are only summing over ρ satisfying the condition $F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1$. 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 6 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 at least a factor $q^{-|Y|}$ larger. This proves Lemma 7.

Next we try to estimate the other factor. Namely

$$Pr[min(G\lceil_\rho)^Y \geq s | F\lceil_\rho \equiv 1 \wedge G_1\lceil_{\rho_1} \not\equiv 1 \wedge \rho_1(Y) = *]$$

As before think of the minterm as consisting of two parts

(1) A part σ_1 which assign values to the variables of Y.

(2) A part σ_2 which assign values to some variables in the complement $\overline{T}$ of T.

To use the induction hypothesis we have to get rid of the condition that $G_1\lceil_\rho \not\equiv 1$. Also here we maximize over the behavior of ρ on T. We have to be slightly careful since when we know the values of ρ on T the remaining part of the restriction is slightly different. We get around this as follows. If the specified values are 0 or 1 there is no problem since this can be incorporated in the condition $F\lceil_\rho \equiv 1$. If the value is * we have to do something else. We substitute all the other values in the the same block B_i (they are now known to be all 1) and in the future we take the probability over a restriction with one block less. In F and G we substitute as in the case of the stronger main lemma. Thus we can use the induction hypothesis and we get the estimate $(2^{|Y|} - 1)\alpha^{s-|Y|}$ for the second factor.

Finally we must evaluate the sum in the same way as before getting

$$\sum_{Y \subseteq T} q^{|Y|}(2^{|Y|} - 1)\alpha^{s-|Y|} \leq \alpha^s$$

This finishes the induction step and the proof of the lemma 5.

To prove theorem 4 we will first prove a slightly stronger technical theorem.

Theorem 5: There are no depth k circuit computing g_k^m with bottom fanin $\leq \frac{1}{10}m$ and size $\leq 2^{\frac{1}{10}m}$ for $m > m_0$ some absolute constant m_0.

First observe that that Theorem 5 implies Theorem 4 since a depth $k-1$ circuit can be considered as a depth k circuit with bottom fanin 1.

Theorem 5 will be proved by induction over k. The base case for $k = 2$ is quite easy and is left to the reader for verification.

For the induction step we will use one of the restrictions defined above. Assume for definiteness that that k is odd so the gates on the bottom level are AND gates. Define the sets B_i in the partition to be the set of variables leading into an AND gate. Now set $p_1 = m^{1-k}, p_0 = 1 - m^{-1} - m^{1-k}$ and $p_* = m^{-1}$ and apply a random restriction from $R^+_{p_1,p_0,p_*,B}$. Since the size of the blocks are all m^{k-1} we get $q = \frac{m^{-1}}{1-m^{1-k}}$.

In the case of the parity function we did not have to worry about what happened to the function we were trying to compute when we applied a restriction. The reason for this was that parity is very robust and always remains as parity. Here what happens to g_k^m when we apply a restriction

is of crucial importance. It was precisely the fact that the R_p restrictions simplified g_k^m to much that forced us to define the new probability space of restrictions. Thus we will first deal with this issue, namely to prove that the present restriction transforms g_k^m into something that is very close to g_{k-1}^m.

Lemma 8: If k is odd then the circuit that defines $g_k^m\lceil_\rho$ for a random $\rho \in R^+_{p_1,p_0,p_*,B}$ will contain the circuit that defines g_{k-1}^m with probability at least $\frac{2}{3}$ for $m > m_1$ for some absolute constant m_1.

Remark 10: For even k Lemma 8 holds with R^+ replaced by R^-.

Proof: The fact that k is odd implies that the two lower levels look like:

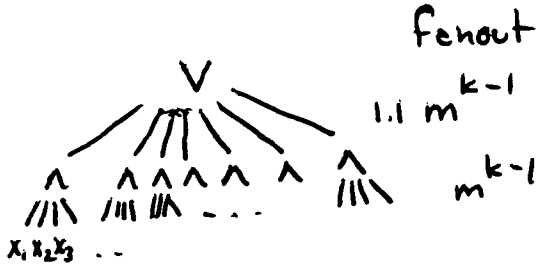

Figure 8

Observe that one can view the restriction as giving values to the AND gates. It gives the value 1 with probability p_1, the value 0 with probability p_0 and * with probability p_*. An OR gate can be forced to 1 by having one input with the value 1. The probability that an individual OR-gate will not be forced to 1 is $(1-m^{1-k})^{1.1m^{k-1}}$. For large m this is approximately $e^{-1.1}$ and thus the probability that the number of surviving ORs in an AND gate one level up is at least $1.1m^{k-2}$ i.e. at least a quarter is $1-2^{-cm^{k-2}}$ for some constant c for $m > m_0$ some absolute constant m_0. Thus the probability that this will be true for all AND gates is $\geq \frac{5}{6}$ if $m > m_1$ for some absolute constant m_1. If an OR-gate survives then the expected number of *'s in it is $1.1m^{k-2}$ and with probability $1-2^{-cm^{k-2}}$ it will be at least m^{k-2} for an individual OR gate and hence the probability that all OR gates will have at least this many inputs is $\geq \frac{5}{6}$ for $m > m_1$. The lemma is proved.

Let us now finish the proof Theorem 5. We need to do the induction step. This is done by the same argument as was used to prove Theorem 2 in section 4. We apply a restriction from $R^+_{p_1,p_0,p_*,B}$ by Lemma 8 the circuit still computes a function as difficult as g_{k-1}^m and setting some of the remaining variables we can make it into g_{k-1}^m. By Lemma 4 we can with high probability change the order of $ANDs$ and ORs in the last two levels and still maintain a small bottom fanin and we get a circuit certified not to exist by the induction hypothesis. ∎

6. Separation of Complexity classes by Oracles

As mentioned in the introduction lower bound results for small depth circuits can be used to construct oracles relative to which certain complexity classes are different [FSS],[S]. In particular the result for parity implies that there are oracles for which PSPACE is different from the polynomial time hierarchy. In the same way theorem 5 implies that there are oracles separating the different levels within the polynomial time hierarchy. As previously remarked, Yao's bounds [Y] were sufficient to obtain these separations. Cai [C] proved that PSPACE was different from the polynomial time hierarchy even for a random oracle. To obtain this result one has to strengthen the results and prove that no function computed by a small constant depth circuit can agree with parity on substantially more than half of the inputs. We discuss this type of results in the next section.

To prove that a random oracle separates the different levels within the polynomial hierarchy one would have to strengthen Theorem 5 to say that no depth $k-1$ circuit computes a function which agree with g_k^m for most inputs. This is not true in the case of g_k^m since if k is even the con-

stant function 1 agrees with g_k^n for most inputs. However perhaps it is possible to get around this by defining other functions more suited to this application.

7. Related results

As mentioned above the key to proving that $PSPACE^A \neq PH^A$ to a random oracle is to prove that small constant depth circuits accepts almost as many odd as even strings. In other words the output of the circuit agrees with parity for only slightly more than half of the inputs. A natural question is to make this statement precise. To this end define $h(s,k,n)$ to be the function such that any depth k circuit of size 2^s with n inputs agrees with parity for a fraction of the inputs which is at most $\frac{1}{2} + h(s,k,n)$. To obtain the separation it is sufficient to have $h(s,k,n) \leq c < \frac{1}{2}$ for $s = (\log n)^i$, all constants i and k and sufficiently large n. Cai obtained his result by showing that $h(n^{\frac{1}{4k}},k,n) = o(1)$. Ajtai had previously proved that $h(c\log n,k,n) \leq 2^{-n^{1-\epsilon}}$ for all constants c, k, and $\epsilon > 0$. It can be seen by construction that $h(s,k,n) \geq 2^{-\frac{n}{s^{k-1}}}$. Together with Ravi Boppana we can prove that $h(s,k,n) \leq 2^{-\Omega(\frac{n}{s^{k-1}})}$ for $k = 2$ and for general k and $s > n^{\frac{1}{k}}$. We get exponentially small but suboptimal results for general k and small s.

The constant $\frac{1}{10}$ in the theorems is clearly not the optimal constant. We have discarded information by only using $\alpha \leq 5pt$ and also the choice of making this quantity $\frac{1}{2}$ is not optimal. However there is a more significant way of improving the constant. It is possible to improve the Main Lemma to let α be a root of the equation

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

The way to get this is to observe that we have not used the full strength of Lemma 3. One way to get the better result which was observed by Ravi Boppana is to use partial summation.

One interesting question is what happens if we also allow the circuit to contain parity gates of arbitrary fanin. Clearly in this case parity has very small circuits but the interesting question is what happens with majority. We are able to prove that at least $\Omega(\log n)$ parity gates are required to have polynomial size constant depth circuits computing majority. Since parity can be computed by constant depth circuits given gates that compute majority, this is a weak piece of evidence that majority might be harder to compute in parallel than parity.

Acknowledgment 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. Mike Saks' observation which simplified the proof of lemma 3 was also helpful. I am also grateful to several people who have read and commented on drafts of this paper. These people include Ravi Boppana, Zvi Galil, Oded Goldreich, Shafi Goldwasser, Jeff Lagarias, Silvio Micali, Nick Pippenger and David Shmoys.

References

[Aj] Ajtai M. "$\sum_1^1$-Formulae on Finite Structures", *Annals of Pure and Applied Logic* 24(1983) 1-48

[AB] Alon N. and Boppana R. "The Monotone Circuit Complexity of Boolean Functions", Submitted to *Combinatorica*.

[An] Andreev A.E. "On one method of obtaining lower bounds of individual monotone function complexity" *Dokl. Ak. Nauk.* 282 (1985), pp 1033-1037.

[B] Boppana R. "Threshold Functions and Bounded Depth Monotone Circuits" *Proceedings of 16th Annual ACM Symposium on Theory of Computing*, 1984, 475-479. To appear in *Journal of Computer and System Sciences*, 1986.

[C] Cai J. "With Probability One, a Random Oracle Separates PSPACE from the Polynomial-Time Hierarchy" These proceedings.

[FSS] Furst M., Saxe J. and Sipser M., "Parity, Circuits, and the Polynomial Time Hierarchy", *Proceedings of 22nd Annual IEEE Symposium on Foundations of Computer Science*, 1981, 260-270.

[KPPY] Klawe M.,Paul W, Pippenger N. and Yannakakis M. "On Monotone Formulae with Restricted Depth" *Proceedings of 16th Annual ACM Symposium on Theory of Computing*, 1984, 480-487.

[R] Razborov A.A. "Lower Bounds for the Monotone Complexity of some Boolean Functions" *Dokl. Ak. Nauk.* 281 (1985), pp 798-801.

[S] Sipser M. "Borel Sets and Circuit Complexity", *Proceedings of 15th Annual ACM Symposium on Theory of Computing*, 1983, 61-69.

[V] Valiant L. "Exponential Lower Bounds for Restricted Monotone Circuits" *Proceedings 15th Annual ACM Symposium on Theory of Computing*, 1983, 110-117.

[Y] Yao A. "Separating the Polynomial-Time Hierarchy by Oracles" *Proceedings 26th Annual IEEE Symposium on Foundations of Computer Science*, 1985, 1-10.