Almost Optimal Lower Bounds for Small Depth Circuits

Johan Hastad*
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 goals of this branch is of course to show that $P \neq NP$. 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 resolving this question. To gain understanding of the problem of proving lower bounds and develop techniques, several restricted models of computation have been studied. Recently there has 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 an arbitrary number of 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 causes a function to be difficult to compute in this model 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 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

* 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.

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], Valiant [V], 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 be 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 two adjacent levels of ANDs and ORs, then by merging the two now adjacent levels with the same connective decrease the depth of the circuit to $k-1$. 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 the size not increasing too much was not proved to be exponentially small. 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.

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]. Here as everywhere else in the paper $\log n$ denotes logarithms to base 2.

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 polynomial 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 achieves this separation is still open.

1.4 Relations to PRAMs

The model of small depth circuits has relations to computation by parallel random access machines (PRAM). In particular, Stockmeyer and Vishkin [SV] proved that any function that can be computed on a slightly limited PRAM with a polynomial number of processors in time T can also be computed by polynomial size unbounded fanin circuits of depth $O(T)$. The limitations on the PRAM was a limitation of the instructionset to only contain relatively simple operations like addition, comparison, indirect addressing multiplication by $\log n$ size numbers, etc.

Thus our results imply among other things that parity requires time $\Omega(\frac{\log n}{\log \log n})$ to compute on such a PRAM. Interestingly enough the same bounds can be proved for more powerful PRAMs using extensions of the present techniques [BeH].

1.5 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. An earlier version of this paper appeared in [H1]. The paper is also part of my thesis [H2].

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 at most 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 are 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. *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, i.e., having variables as inputs.

2.2 Outline of Proof

Several 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. It is easily seen that depth 2 parity circuits require size 2^{n-1} [FSS]. The second step is much more difficult and contains 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 of ORs or as an OR of ANDs. Thus we can change this depth 2 circuit to and OR of ANDs which computes the same function, as below.

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 inputs of an OR gate we know that the output of that 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 many more gates than inputs and so we have to do something more efficient than forcing one gate per variable. 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 making 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.

We are looking for restrictions which simplify circuits efficiently. It seems hard to do this explicitly and we will use a probabilistic method. We will be working with random restrictions with distributions parameterized by a real number p which will usually 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. Thus the expected number of variables remaining 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 of 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 $\bar{x}_i$ 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 An elementary argument shows that $\alpha = \frac{2pt}{\ln \phi} + O(p^2 t) < 5pt$, for sufficiently small p, 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 identical except for notation. Our original proof was in terms of a labeling algorithm as in those 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 it facilitates 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 2 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

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

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

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 force 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 $(Pr[min(G) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1])$. 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 ρ is symmetric with respect to 0 and 1 and hence we can interchange x_i and $\bar{x}_i$ if necessary. 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,and we 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

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

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

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

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: *$Pr[A \mid B \wedge C] \leq Pr[A \mid C]$ 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: $Pr[\rho_1(Y) = * \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \leq (\frac{2p}{1+p})^{|Y|}$

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 try to estimate the other factor. Namely

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

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

(1) A part σ_1 which assigns values to the variables of Y.
(2) A part σ_2 which assigns values to some variables in the complement $\bar{T}$ 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

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

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

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 $\bar{T}$. Thus we can estimate the probability by $\alpha^{s-|Y|}$ using the induction hypothesis.

We have to be slightly careful when we use the induction hypothesis since $G\lceil_{\rho_1\sigma_1}$ might depend on variables in $T - Y$. These variables cannot, by the definition of Y, be in the minterm we are looking for. This implies that we can instead look at the formula $\wedge_{\sigma_T} G\lceil_{\rho_1\sigma_1\sigma_T}$ where σ_T ranges over all ways to give values to the remaining variables in T.

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. Since the term corresponding to $Y = \emptyset$ is 0 we can include it.

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

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

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 6 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 $\leq 2^{\frac{1}{10}n^{\frac{1}{k-1}}}$ subcircuits 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 represent 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 2 subcircuit can be written as an 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 $\frac{1}{2}$. If we choose $s = \frac{1}{10}n^{\frac{1}{k-1}}$ the probability that any of the $2^{\frac{1}{10}n^{\frac{1}{k-1}}}$ depth 2 circuits cannot be converted into a depth 2 circuit of the other type is bounded by $(2\alpha)^s$. Thus 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 parity of at least $n^{\frac{k-2}{k-1}} = m$ variables. Further it 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 because 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 similar corollaries could 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.

We prove that there are functions f_k^m which have linear size circuits of depth k but require exponential size circuits when the depth is restricted to $k-1$. To prove this we will introduce a new probability space of restrictions and reprove the Main Lemma for this space of restrictions.

5.1 The Sipser Functions f_k^m.

In [Si], Sipser defined a set of functions f_k^m which could be computed in depth k and linear size. He showed , however, that these functions require superpolynomial size when the depth is restricted to $k-1$. We will redefine f_k^m slightly and let it denote the function defined by the circuit in figure 6. To avoid confusion we will refer to the circuit in figure 6 as the defining circuit of f_k^m. The defining circuit is thus a tree with top fanin $\sqrt{\frac{m}{\log m}}$, bottom fanin $\sqrt{km\log m/2}$, while all the other fanouts are m. Each variable occurs at only one leaf.

Thus by definition f_k^m is a function of $m^{k-1}\sqrt{k/2}$ variables.

Figure 6

Yao has claimed exponential lower bounds for these functions, but the proof has not yet appeared. We have the following results for the functions f_k^m.

Theorem 4: *Depth $k-1$ circuits computing f_k^m are of size at least $2^{\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}}$ for $m > m_1$, where m_1 is some absolute constant.*

As an immediate corollary we get.

Corollary: *Polynomial size circuits of depth $f(n)$ are more powerful than polynomial size circuits of depth $f(n)-1$ if $f(n) < \frac{\log n}{3\log\log n} - \omega(\frac{\log n}{(\log\log n)^2})$.*

Proof: Follows from a computation using $n = m^{k-1}\sqrt{k/2}$ and $k = f(n)$.

5.2 New Random Restrictions.

One would like to prove Theorem 4 with the aid of the Main Lemma. Here, however, one runs into problems not encountered in the case of the parity function. If a restriction from R_p is applied to f_k^m the resulting function will be a constant function with very high probability. This happens since the gates at the bottom level are quite wide and with very high probability all gates will be forced. There is also a more philosophical reason why R_p destroys functions like f_k^m. R_p was designed to destroy any small-depth circuit, and will in particular destroy the circuits defining f_k^m. To get around this problem we will define another set of restrictions are designed not to destroy the circuits defining f_k^m.

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

For $\rho \in R_{q,B}^+$ and every B_i, $1 \leq i \leq r$ independently

1. With probability q let $s_i = *$ and else $s_i = 0$.
2. For every $x_k \in B_i$ let $\rho(x_k) = s_i$ with probability q and else $\rho(x_k) = 1$.

Similarly a $R^-_{q,B}$ probability space of restrictions is defined by interchanging the roles played by 0 and 1.

The idea behind these restrictions is that a block B_i will correspond to the variables leading into one of the ANDs in the bottom level in the circuit defining f^m_k. If the bottom level gates are ORs we use a restriction from R^-_q. These restrictions will, however, not be quite sufficient for our purposes and we need a complementary restriction.

Definition: For a restriction $\rho \in R^+_{q,B}$ let $g(\rho)$ be a restriction defined as follows: For all B_i with $s_i = *$, $g(\rho)$ gives the value 1 to all variables given the value $*$ by ρ except one to which it gives the value $*$. To make $g(\rho)$ deterministic we assume that it gives the value $*$ to the variable with the highest index given the value $*$ by ρ. If $\rho \in R^-_{q,B}$, then $g(\rho)$ is defined similarly but now takes the values 0 and $*$.

These probability spaces of restrictions do not assign values to variables independently as R_p did, but is nice enough so that the proof of our Main Lemma will go through with only minor modifications. Let $\rho g(\rho)$ denote the composition of the two restrictions. Observe that they are compatible since $g(\rho)$ assigns values to precisely the variables given the value $*$ by ρ.

Lemma 4: *Let G be an AND of ORs all of size $\leq t$ and ρ a random restriction from $R^+_{q,B}$. Then the probability that $G\lceil_{\rho g(\rho)}$ cannot be written as an OR of ANDs all of size $< s$ is bounded by α^s, where $\alpha = \frac{4q}{2^{\frac{1}{t}}-1} < \frac{4qt}{\log 2} < 6qt$.*

Remark 7 The same is true for $R^-_{q,B}$.

Remark 8 The probability of converting an OR of ANDs to an AND of ORs is the same.

As in the case of the Main Lemma, before proving Lemma 4, we prove a stronger lemma stating that we have the same estimate of the probability even when we condition upon an arbitrary function being forced to 1 by ρ. Define $AND(G\lceil_{\rho g(\rho)}) \geq s$ denote the event that $G\lceil_{\rho g(\rho)}$ cannot be written as an OR of ANDs of size $< s$.

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 $R^+_{q,B}$. Then*

$$Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_\rho \equiv 1] \leq \alpha^s$$

where $\alpha = \frac{4q}{2^{\frac{1}{t}}-1}$.

Remark 9 Recall that, if there is no restriction ρ satisfying the condition $F\lceil_\rho \equiv 1$ then the conditional probability in question is defined to be 0. Observe that we are only conditioning upon $F\lceil_\rho \equiv 1$ and not $F\lceil_{\rho g(\rho)} \equiv 1$.

Proof: We will only use the weaker bound $6qt$ for α and since the lemma is trivially true if $\alpha \geq 1$ we will assume $q < \frac{1}{6t}$ whenever convenient. The proof will be done in a similar way to the proof of the Stronger Main Lemma. We therefore only outline the proof, and give details only where the proofs differ.

As before

$$Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_\rho \equiv 1] \leq$$

$$\leq max\left(Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \equiv 1], Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1]\right)$$

The first term,

$$Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid (F \wedge G_1)\lceil_\rho \equiv 1]$$

is taken care of by the induction hypothesis.

We have to estimate the second term, $Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_{\rho} \equiv 1 \wedge G_1\lceil_{\rho} \not\equiv 1]$. We cannot assume that G_1 is an OR of only positive literals since the restrictions employed here assign 0 and 1 nonsymmetrically.

We denote the set of variables occurring in G_1 by T, and $|T| \leq t$. We do not know that G_1 must be made true by every minterm of $G\lceil_{\rho g(\rho)}$. This is because G_1 might be made true by $g(\rho)$. We do know, however, that for $G\lceil_{\rho}$ not to be the constant 0 some of the variables of T must be given the value $*$ by ρ. Suppose the variables of T belongs to r different blocks. Assume for notational convenience that these blocks are $B_i, i = 1 \ldots, r$. We call a block B *exposed* if there is a variable $x_i \in B$ such that $x_i \in T$ and $\rho(x_i) = *$. By the above remark there must be some exposed blocks for G not to be identically 0. Let Y denote the set of exposed blocks. Denote this event by $exp(Y)$ and let $[r]$ denote the set $\{1, 2, \ldots, r\}$.
We get

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

$$\sum_{Y \subset [r], Y \neq \emptyset} Pr[exp(Y) \mid F\lceil_{\rho} \equiv 1 \wedge G_1\lceil_{\rho} \not\equiv 1] \times Pr[AND(G\lceil_{\rho g(\rho)}) \geq s \mid F\lceil_{\rho} \equiv 1 \wedge G_1\lceil_{\rho} \not\equiv 1 \wedge exp(Y)]$$

The factors in the above sum can be estimated separately. Let us start with the first factor $Pr[exp(Y) \mid F\lceil_{\rho} \equiv 1 \wedge G_1\lceil_{\rho} \not\equiv 1]$. We need a little bit of extra notation. Let $P_i = \{j \mid x_j \in G_1 \wedge x_j \in B_i\}$ and let $N_i = \{j \mid \bar{x}_j \in G_1 \wedge x_j \in B_i\}$. Let us start with the simple case when Y consists of a single block B_i.

Lemma 6: $Pr[exp(B_i) \mid F\lceil_{\rho} \equiv 1 \wedge G_1\lceil_{\rho} \not\equiv 1] \leq 2q$.

Proof: By the definition of conditional probability we want to prove

$$\frac{\sum'_{exp(B_i)} Pr(\rho)}{\sum' Pr(\rho)} \leq 2q$$

Here the $'$ indicates 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 takes the form $\frac{0}{0}$ we have the convention that it takes the value 0. Now assume that ρ gives a nonzero contribution to the numerator. We define a restriction $\tilde{\rho} = H(\rho)$ which gives a larger contribution to the denominator. Let

1. $\tilde{\rho}(x_j) = \rho(x_j)$ for $x_j \notin B_i$
2. $\tilde{\rho}(x_j) = 0$ for $j \in P_i$
3. $\tilde{\rho}(x_j) = 1$ for $j \in N_i$
4. $\tilde{\rho}(x_j) = 1$ for $j \in B_i - N_i - P_i$ and $\rho(x_j) = 1$
5. $\tilde{\rho}(x_j) = 0$ for $j \in B_i - N_i - P_i$ and $\rho(x_j) = *$

To check that $\tilde{\rho}$ gives a contribution to the denominator we only have to check that $F\lceil_{\tilde{\rho}} \equiv 1$ and $G_1\lceil_{\tilde{\rho}} \not\equiv 1$. The first fact follows by noting that we only change values from $*$ to non-$*$ values. To see that the second condition is fulfilled we observe that rules 2 and 3 are tailored to this.

To get an estimate for the quotient we must compute the probability of ρ compared to $\tilde{\rho}$. We must also investigate what restrictions $\bar{\rho}$ satisfy $H(\bar{\rho}) = \tilde{\rho}$. Let us start with this second task.

Observe first that $s_i = *$ in the definition of $\bar{\rho}$ and hence that $\bar{\rho}$ only gives out the values $*$ and 1 on B_i. Obviously $\bar{\rho}(x_j) = \rho(x_j)$ for all x_j not in P_i or N_i. Furthermore $\bar{\rho}(x_j) = *$ for $x_j \in P_i$ since $G\lceil_{\bar{\rho}} \not\equiv 1$. Finally $\bar{\rho}$ can take any combination of 1 and $*$ on N_i provided it does not take the value of all 1 in the case when P_i is empty. Observe that all these $\bar{\rho}$ might not satisfy the condition

$F\lceil_{\bar{\rho}} \equiv 1$ but we are only trying to get an upper bound. Assume that $\bar{\rho}$ assigns l $*$'s on N_i and $|N_i| - l$ ones. Then

$$Pr(\bar{\rho}) = \frac{q}{1-q}\frac{q^l(1-q)^{|N_i|-l}}{(1-q)^{|N_i|}}Pr(\tilde{\rho}).$$

The first factor comes from the fact that $s_i = 0$ for $\tilde{\rho}$ while $s_i = *$ for $\bar{\rho}$. The second factor comes from the behavior on N_i. Observe that the probability that $\tilde{\rho}$ gives out only 0 on P_i is equal to the probability that $\bar{\rho}$ gives out only $*$. Summing up we get

$$\sum_{H(\bar{\rho})=\tilde{\rho}} Pr(\bar{\rho}) \leq \frac{q}{1-q}Pr(\tilde{\rho})\sum_{l=0}^{|N_i|}\binom{|N_i|}{l}(\frac{q}{1-q})^l = \frac{q}{1-q}Pr(\tilde{\rho})(1-q)^{-|N_i|}$$

since $|N_i| \leq t$ and $q < \frac{1}{6t}$ we have $(1-q)^{-|N_i|} < 2$. Using this we have

$$\frac{\sum'_{exp(B_i)} Pr(\rho)}{\sum' Pr(\rho)} \leq \frac{\sum_{\tilde{\rho}}\sum'_{\rho,H(\rho)=\tilde{\rho},exp(B_i)} Pr(\rho)}{\sum' Pr(\rho)} \leq$$

$$\frac{\sum'_{\tilde{\rho}} \frac{2q}{1-q}Pr(\tilde{\rho})}{\sum'_{\tilde{\rho}}(1+\frac{2q}{1-q})Pr(\tilde{\rho})} \leq 2q$$

and the proof is complete. ∎

Next we have

Lemma 7: $Pr[exp(Y) \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1] \leq (2q)^{|Y|}$.

Proof: This is proved in the same way as Lemma 6. We get restrictions contributing to the denominator by doing the changes in ρ on all the blocks simultaneously. ∎

Next we estimate the factor

$$Pr[AND(G\lceil_{\rho g(\rho)})^Y \geq s \mid F\lceil_\rho \equiv 1 \wedge G_1\lceil_\rho \not\equiv 1 \wedge exp(Y)]$$

We want to use induction and to do this we have to get rid of the condition $G_1\lceil_\rho \not\equiv 1$. In the blocks that are not exposed we know that ρ only takes the values 0 or 1. This conditioning can be incorporated in $F\lceil_\rho \equiv 1$.

In the exposed blocks we let the corresponding variables which are still alive after $\rho g(\rho)$ be in the ANDs of $G\lceil_{\rho g(\rho)}$. We try all possibilities of these variables and we estimate the probability that the remaining formula cannot be written as an OR of ANDs of size $s - |Y|$. This probability is taken over a restriction which does not include the blocks of Y. Thus we can use the induction hypothesis and we get the estimate $\alpha^{s-|Y|}$ for each setting of the variables corresponding to Y. Thus we get the total bound $2^{|Y|}\alpha^{s-|Y|}$.

Finally we evaluate the sum to get

$$\sum_{Y\subset[r],Y\neq\emptyset}(2q)^{|Y|}2^{|Y|}\alpha^{s-|Y|} = \alpha^s\sum_{i=1}^{r}\binom{r}{i}(\frac{4q}{\alpha})^i \leq$$

$$\alpha^s((1+\frac{4q}{\alpha})^r - 1) = \alpha^s(2^{r/t} - 1) \leq \alpha^s$$

This finishes the induction step and the proof of the Lemma 5. ∎

An interesting question is for what probability distributions on the space of restrictions is it possible to prove the lemma equivalent to the Main Lemma and Lemma 4. The general proof technique uses two crucial properties of the distribution.

(1) The condition $F\lceil_\rho \equiv 1$ for an arbitrary F does not bias the value of any variable too much towards $*$. This should also remain true even if we know that a the variable is not 1(0).
(2) It is possible to eliminate the variables of G_1 and use induction on a similar restriction over the remaining variables.

Condition (1) was taken care of by Lemmas 3 and 7. Condition (2) seems easier to satisfy and was so obviously satisfied that no formal lemma was needed. The verification was basically done where we claimed that induction could be used after eliminating G_1.

5.3 Back to the Proof of Theorem 4

Let us continue with the present restriction space $R^+_{q,B}$ and prove Theorem 4. We first prove a slightly stronger technical theorem.

Theorem 5: *There are no depth k circuits computing f^m_k with bottom fanin $\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}$ and $\leq 2^{\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}}$ gates of depth ≥ 2 for $m > m_0$ some absolute constant m_0.*

Note 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 is proved by induction over k. The base case for $k = 2$ is quite easy and is left to the reader.

For the induction step we use one of the restrictions defined above. Assume for definiteness that k is odd, so that 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. Recall that since the defining circuit of f^m_k is a tree the blocks are disjoint. Set $q = \sqrt{\frac{2k \log m}{m}}$ and apply a random restriction from $R^+_{q,B}$.

In the case of the parity function even after applying a restriction, it was trivial that the remaining circuit still computed parity or the negation of parity. In the case of f^m_k, we have tol prove that the new restrictions used transform f^m_k into something that is very close to f^m_{k-1}.

Lemma 8: *If k is odd then the circuit that defines $f^m_k\lceil_{\rho g(\rho)}$ for a random $\rho \in R^+_q$ will contain the circuit that defines f^m_{k-1} with probability at least $\frac{2}{3}$, for all m such that $\frac{m}{\log m} \geq 100k$, $m > m_1$, where m_1 is some absolute constant.*

Remark 10 Lemma 8 holds for even k when R^+ is replaced by R^-.
Proof: The fact that k is odd implies that the two lower levels look like:

Figure 7

We establish a series of facts.
Fact 1: The AND gate corresponding to block B_i takes the value s_i for all i, with probability at least $\frac{5}{6}$ for $m > m_0$.

The AND gate corresponding to block B_i takes the values s_i precisely when not only ones are given to the block. The probability of this happening is $(1-q)^{|B_i|} = (1-\sqrt{\frac{2k\log m}{m}})^{\sqrt{\frac{km}{2\log m}}} < e^{-k\log m} < \frac{1}{6}m^{-k}$ for $m > m_0$. Thus the probability that this happens for any block B_i is bounded by $\frac{1}{6}$ for $m > m_0$.

Fact 2 With probability at least $\frac{5}{6}$ at least $\sqrt{(k-1)m\log m/2}$ inputs given the value $*$ by $\rho g(\rho)$ to each OR gate at level $k-1$. Again this is true only for sufficiently large m.

The expected number of such inputs is $\sqrt{2km\log m}$ and the fact follows from known estimates using $\frac{m}{\log m} \geq 100k$. For completeness let us include a very elementary proof.

Let p_i be the probability that an OR gate has as input exactly i AND gates which take the value $*$. Then

$$p_i = \binom{m}{i}(\frac{2k\log m}{m})^{\frac{i}{2}}(1-\sqrt{\frac{2k\log m}{m}})^{m-i}.$$

Then for $i < \sqrt{km\log m}$ we have $p_i/p_{i-1} \geq \sqrt{2}$. Using $p_{\sqrt{mk\log m}} < 1$ we estimate $\sum_{i=1}^{\sqrt{mk\log m/2}} p_i$ by

$$\sum_{i=1}^{\sqrt{mk\log m/2}} p_i \leq p_{\sqrt{mk\log m/2}} \sum_{i=0}^{\infty} 2^{-\frac{i}{2}} \leq 4p_{\sqrt{mk\log m/2}} \leq$$

$$4\sqrt{2}^{-(1-\frac{1}{\sqrt{2}})\sqrt{mk\log m}} p_{\sqrt{mk\log m}} \leq 4\sqrt{2}^{-(1-\frac{1}{\sqrt{2}})10k\log m} \leq \frac{1}{6}m^{-k}$$

for $m > m_0$.

To sum up, with probability at least $\frac{2}{3}$ all OR gates at level $k-2$ will remain undetermined, and have at least $\sqrt{(k-1)m\log m/2}$ variables as inputs. This constitutes the defining circuit for f^m_{k-1}. The lemma is proved. ∎

Let us now finish the proof Theorem 5. We need to perform the induction step. This is done using the same argument as in the proof of Theorem 2. Apply a restriction from $R^+_{q,B}$ to the circuit. Observe first that if $\frac{m}{\log m} < 100k$ the result of the theorem is trivial and hence we can assume that the reverse inequality holds. By Lemma 8 the defining circuit still computes a function as difficult as f^m_{k-1} and setting some of the remaining variables the circuit can be made into the defining circuit of f^m_{k-1}.

On the other hand suppose that there existed a circuit of depth k, bottom fanin $\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}$ and size $2^{\frac{1}{12\sqrt{2k}}\sqrt{\frac{m}{\log m}}}$ which computed f^m_k. By using Lemma 4 and reasoning as in the proof of Theorem 2 we can interchange the ANDs and ORs on the last two levels without increasing the bottom fanin. Now it is possible to collapse two adjacent levels of OR gates and the resulting circuit will be of depth $k-1$. As in the proof of Theorem 2 the gates corresponding to subcircuits of depth 2 in this new circuit corresponds to gates of depth 3 in the old circuit. Thus we have obtained a circuit certified not to exist by induction. ∎

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 4 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 prove this result one needs to establish that a small circuit makes an error when trying to compute parity on a random input with a probability close to $\frac{1}{2}$. This problem and related problems are studied in [BoH].

To prove that a random oracle separates the different levels within the polynomial hierarchy one would have to strengthen Theorem 4 to say that no depth $k-1$ circuit computes a function which agrees with f_k^m for most inputs. This is not true in the case of f_k^m since if k is even(odd), the constant function 1(0) agrees with f_k^m for most inputs. However, perhaps it is possible to get around this by defining other functions more suited to this application.

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.

[BeH] Beame P. and Hastad J. "Optimal Bounds for Decision Problems on the CRCW PRAM", to appear in Symposium of 18th Annual ACM Symposium on Theory of Computing.

[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*.

[BoH] Boppana R. and Hastad J. "Approximation Properties of Constant Depth Circuits", manuscript in preparation.

[C] Cai J. "With Probability One, a Random Oracle Separates PSPACE from the Polynomial-Time Hierarchy", *Proceedings of 18th Annual ACM Symposium on Theory of Computing*, 1986, 21-29.

[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.

[H1] Hastad J., "Almost Optimal Lower Bounds for Small Depth Circuits", *Proceedings of 18th Annual ACM Symposium on Theory of Computing*, 1986, 6-20.

[H2] Hastad J., "Computational Limitations for Small-Depth Circuits", MIT Press, 1986.

[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.

[SV] Stockmeyer L.J and Vishkin U. "Simulation of Parallel Random Access Machines by Circuits",

SIAM J. on Computing, vol 13(2), 1984, pp. 404-422.

[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.