On the Power of Small-Depth Threshold Circuits

Johan Håstad Mikael Goldmann
Royal Institute of Technology
Stockholm, SWEDEN

Abstract

We investigate the power of threshold circuits of small depth. In particular we give functions which require exponential size unweigted threshold circuits of depth 3 when we restrict the bottom fanin. We also prove that there are monotone functions f_k which can be computed in depth k and linear size $\wedge, \vee$-circuits but require exponential size to compute by a depth $k-1$ monotone weighted threshold circuit.

1 Introduction

The study of circuit complexity has in one sense been successful and in another not so successful. While there are still no non-linear lower bounds on circuit-size for any function in NP, several interesting results have been shown for restricted circuit classes e.g. monotone circuits [And85, Raz85, AB87, KW88, RW89, RW90] and circuits of bounded depth [Ajt83, FSS84, Hås86, Raz, Smo87, Yao85].

The smallest natural circuit-class that is not known to be strictly contained in NP is TC^0, the set of functions computable by constant-depth polynomial-size circuits containing threshold gates. Threshold gates are quite powerful and many fairly complicated functions (like division, implicit in [BCH84]) are in TC^0.

It also seems like the techniques used for proving lower bounds for usual constant depth circuits (with or without modular gates) are not sufficient to prove lower bounds for threshold circuits. The best known results about small-depth threshold circuits are the results by Hajnal et al. [HMP+87] where, among other results, it is established that depth 3 threshold circuits of polynomial size are more powerful than corresponding circuits of depth 2.

To further understand the nature of threshold circuits, Yao [Yao89] studied monotone threshold circuits. In particular Yao was interested in the question whether for monotone circuits it is true that depth k and polynomial size is more powerful than depth $k-1$ and polynomial size. He proved that indeed this is the case by exhibiting a function f_{2k} computable by ordinary $\wedge$, $\vee$-circuits of depth $2k$ which required exponential size when the depth was restricted to k.

In this paper we generalize both the above results. First we give an explicit function which cannot be computed by small depth 3 unweighted threshold circuits of small bottom fanin. The proof of this is based on a communication game analyzed by Babai, Nisan and Szegedy [BNS89]. In this communication game $s+1$ players (which we call $P_j, j = 1, \ldots s+1$) participates and share some variables which are partitioned into $s+1$ groups (the j'th group being G_j). The player P_j knows all variables except the variables in G_j and the cost of evaluating a function is the number of bits the players have to exchange. Our proof is based on the observation that in this model it is quite cheap to evaluate an unweighted threshold circuit of depth 2 and bottom fanin bounded by s. Applying the correlation lemma of [HMP+87] and the lower bound for the communication game [BNS89] gives the desired lower bound.

We get nontrivial lower bounds as long as $s \leq c \log n$ for some constant c. It is very interesting to note that if these results could be extended to hold for s as large as $(\log n)^k$ for any constant k, then by a resent result by Yao [Yao] it would be possible to prove the given function could not be computed by constant depth polynomial size circuits contained modular gates with composite modulus. This would be the first such lower bound. One possible approach to this problem is to try to improve the lower bounds for the multiparty protocol communication games.

In the monotone case we sharpen Yao's result by proving an exponential difference between depth k and depth $k-1$. In this case we also extend the results by allowing arbitrary positive weights in the circuit. Let us outline the method of proof used.

We define a probability distribution p_k on minterms with a minimal number of 1's and a distribution q_k on maxterms with a maximal number of 1's. The key point is to prove that a threshold circuit of depth $k-2$ which has a good probability of outputting 1 on a random input picked according to the distribution p_k has a probability very close to 1 of outputting 1 on a random input picked according to the distribution q_k. This is done in Lemma 4,

This is exactly the same outline of proof as was used by Yao. We obtain sharper results by also arguing about the negation of f_k, $\bar{f}_k$ and the corresponding distributions $\bar{p}_k$ and $\bar{q}_k$.

It is very interesting to note that while the functions f_k cannot be computed by depth $k-1$ monotone threshold circuits of size $\leq 2^{n^{\frac{1}{2k}}}$, by a recent result of Allender [All89] they can be computed by depth 3 general threshold circuits of size $2^{O((\log n)^k)}$. This might be taken as a another piece of evidence that monotonicity is a severe restriction, and probably new techniques have to be developed to attack general threshold circuits of small depth.

2 Lower Bounds for Depth 3 Circuits

The function we will consider is the "generalized inner product function" considered in [BNS89]. We use doubly indexed variables $x_{i,j}$ where i ranges from 1 to n and j ranges from 1 to $s+1$. Our function f_s is defined by

$$f_s(x) = \sum_{i=1}^{n} \prod_{j=1}^{s+1} x_{i,j}$$

where the sum is calculated modulo 2. We will be interested in a communication game among $s+1$ players P_j, $j = 1, \ldots s+1$ where P_{j_0} knows the value of all variables $x_{i,j}$ except those with $j_0 = j$. Collectively the players want to evaluate $f_s(x)$ (or something close to it) and the measure of complexity is the number of bits exchanged. We call this game the s-communication game and we will say that a protocol ϵ-evluates a function f if the output agrees with f with probability $\frac{1}{2}+\epsilon$. Here the probability is taken over a random input (and not a probabilistic protocol). The following very powerful result is proved by Babai, Nisan and Szegedy.

Lemma 1 ([BNS89, Theorem 2]) *To ϵ-evaluate f_s in a s-communication game requires $\Omega(n2^{-2t} + \log \epsilon)$ bits communication.*

We will be interested in evaluating f_s by a depth 3 threshold circuit. In this section we assume that there are no weights in the circuit i.e. that each gate C in the circuit just counts the number of inputs to it that takes the value 1 and outputs 1 iff this number is at least t_C for a predetermined value t_C. We also assume that we have small bottom fanin i.e. that the number of inputs to any gate next to the input variables is small. The main result of the section is stated below:

Theorem 1 *To evaluate f_s by a depth 3 unweighted threshold circuit with bottom fanin at most s requires size $2^{\Omega(n/s4^s)}$.*

Proof: Assume that such a circuit exist, we will eventually obtain a contradiction to Lemma 1. We first need a lemma which is the same (although we phrase it differently and use that f_s is almost unbiased) as Lemma 3.3 of Hajnal et al.

Lemma 2 ([HMP+87, Lemma 3.3]) *Let f_s be computed by an unweighted threshold circuit where the top gate has fanin R. Then one of the inputs to the top gate $\frac{1}{R}$-evaluates f_s.*

In our case this input to the top gate will correspond to a depth 2 threshold circuit with bottom fanin bounded by s. For these we have the following interesting lemma.

Lemma 3 *Suppose f is computed by a depth 2 unweighted threshold circuit of size R and bottom fanin bounded by s. Then f can be evaluated in the s-communication game with $(s+1)\log R$ communicated bits.*

Proof: Since the bottom fanin is at most s every such gate can be evaluated by one of the players. Thus partition the gates between the players in such a way that that each player can evaluate all the gates given to him. Now each player need only tell the other players how many of his gates evaluated to 1. This he can do with $\log R$ bits. ∎

To finish the proof of the theorem we now just need to collect the pieces. By Lemmas 2 and 3 we see that by the circuit assumed to exist there is a protocol that $\frac{1}{R}$-evaluates f_s with $(s+1)\log R$ bits of communication. Combining this with Lemma 1 finishes the proof. ∎

Remark: We have used very little of the properties of the threshold gates in the proof. That the top gate was a threshold gate was only used to make sure that one of its inputs was correlated with f_s. On the third level we only needed that each gate only depended on s variables. The type of this dependence was unimportant. Thus only on level two did we use any real properties of the threshold gates.

If we instead limit the fanin on the middle level, then the situation gets even simpler.

Theorem 2 *Suppose an unweighted depth 3 threshold circuit computes the inner product (i.e. f_1) and that the fanin on the middle level is bounded by $\frac{cn}{\log n}$ for a small constant c. Then the size of the circuit is $2^{\Omega(n)}$.*

Let us just sketch the proof of this theorem. We know by Lemma 2 that we get a depth 2 threshold circuit with bounded top fanin that computes something correlated with the inner product function. Such a circuit can be evaluated by two players by one player telling the other player how many of his inputs to the various gates take the value 1. Lower bounds for the communicational complexity of evaluating the inner product by two players complete the proof.

3 Monotone Threshold Circuits

Let us start by formally defining our functions f_k. For technical reasons the circuit defining f_k will not be defined by a regular tree.

Definition 1 *The function f_k is a function of N^{2k-2} variables. It is defined by a depth k circuit which is a tree. At the leaves of the tree there are unnegated variables. The i'th level from the bottom consist of $\wedge$-gates if i is even and otherwise it consists of $\vee$-gates. The fanin at the top and bottom levels are N and at all other levels it is N^2.*

It will be convenient to also consider the functions $\bar{f}_k$, the negations of f_k. Clearly $\bar{f}_k$ is computed by a circuit very similar to the circuit computing f_k. The only difference being that $\wedge$ and $\vee$ change places and that the inputs are negated variables. To make $\bar{f}_k$ formally a monotone function we set $y_i = \bar{x}_i$ and let these be the input variables of $\bar{f}_k$.

There is a nice inductive definition of f_k which will be very useful to us. By the definition f_2 is a depth 2 circuit which is an $\wedge$ of size N where each input is a $\vee$ of size N. Now for even k, f_k is just f_2 with each input variables changed to an independent copy of f_{k-1}. Similarly for odd k, f_k is obtained from $\bar{f}_2$ by replacing each variable by an independent copy of f_{k-1}.

Using this construction it is natural to label each variable of f_k by a $2k-2$ tuple of numbers between 1 and N.

In this section we will be interested in weighted threshold circuits i.e. circuits containing gates which are weighted threshold gates. In a weighted threshold gate there are weights w_i on each input and the gate outputs 1 iff the sum $\sum w_i x_i$ is at least a predetermined threshold t. Such a circuit is monotone if all weights are positive and no variable is negated. We have:

Theorem 3 *A monotone weighted threshold circuit computing f_k which is of depth $k-1$ has size at least 2^{cN} for some constant $c > 0$ and $N > N_0$.*

We will look at inputs chosen according to two different probability distributions, p_k and q_k, for inputs in $f_k^{-1}(1)$ and inputs in $f_k^{-1}(0)$ respectively. They are the same distributions as those Yao constructed. Here p_k picks a random assignment to the variables which contains as many zeroes as possible under the condition that $f_k(x) = 1$, while q_k picks a random assignment to the variables which contains as many ones as possible under the condition that $f_k(x) = 0$.

The distributions p_k and q_k are defined inductively starting with $k = 2$. To get the feeling for the definition, please remember the inductive definition of f_k.

Index the variables in f_2 by two indices running from 1 to N and let x_{ij} be the j'th variable in the the i'th $\vee$ defining f_2.

An element from p_2 is chosen as follows: For each i, randomly and uniformly pick a $j(i)$, set $x_{ij(i)} = 1$ and set all other variables to 0.

An element from q_2 is chosen by randomly choosing i_0 and setting $x_{i_0 j} = 0$ for all j, while all other variables are set to 1.

Clearly for each x chosen by p_2, $f_2(x) = 1$, while for each x chosen by q_2, $f_2(x) = 0$. We will also need the corresponding distribution for $\bar{f}$. A random element from $\bar{p}_2$ ($\bar{q}_2$) is a random element from p_2 (q_2) which is changed by setting $y_{ij} = \bar{x}_{ij}$ for all i and j. Observe that the definition implies that if y is chosen according to $\bar{p}_2$ then $\bar{f}(y) = 0$ and if it chosen according to $\bar{q}_2$ then $\bar{f}(y) = 1$.

Now to chose an element from p_k proceed as follows:

1) For k even, choose a random element from p_2. Each variable now corresponds to an independent copy of f_{k-1} and the value of that variable will decide how to give values to the variables corresponding to that function. If a variable is given the value 1, then set the corresponding variables according to p_{k-1} and if the value is 0 then set all corresponding variables to 0.

2) For k odd, choose a random element from $\bar{q}_2$. If a variable is given the value 1, then set the corresponding variables according to p_{k-1} and if the value is 0 then set all corresponding variables to 0.

In a similar way an element from q_k is picked as follows:

1) For k even, choose a random element from q_2. If a variable is given the value 1, then set the corresponding variables to 1 and if the value is 0 then set the corresponding variables according to q_{k-1}.

2) For k odd, choose a random element from $\bar{p}_2$. If a variable is given the value 1, then set the corresponding variables to 1 and if the value is 0 then set the corresponding variables according to q_{k-1}.

Again we define $\bar{p}_k$ and $\bar{q}_k$ by negating variables.

To make formulas simple in the future, suppose that g is a Boolean function and let p be a probability distribution. Then we let $g(p)$ denote the probability that g takes the value 1 on a random input from the distribution p. Using this notation observe that the above definitions imply that:

$f_k(p_k) = 1,\ f_k(q_k) = 0,\ \bar{f}_k(\bar{p}_k) = 0,\ \bar{f}_k(\bar{q}_k) = 1$

Now we are ready to state the main lemma. Let $\epsilon = 0.01$.

Lemma 4 *Let g be a function which is computed by a monotone weighted threshold circuit of depth $k-2$ and size $\leq 2^{\epsilon N}$, then for $N > N_0$,*

1. $g(p_k) \geq 2^{-2\epsilon N} \Rightarrow g(q_k) > \frac{3}{5}$
2. $g(p_k) \geq \frac{2}{5} \Rightarrow g(q_k) > 1 - 2^{-2\epsilon N}$

Consider also the following lemma:

Lemma 5 *Let g be a function which is computed by a monotone weighted threshold circuit of depth $k-2$ and size $\leq 2^{\epsilon N}$, then for $N > N_0$,*

1. $g(\bar{q}_k) \geq 2^{-2\epsilon N} \Rightarrow g(\bar{p}_k) > \frac{3}{5}$

2. $g(\bar{q}_k) \geq \frac{2}{5} \Rightarrow g(\bar{p}_k) > 1 - 2^{-2\epsilon N}$

Let us prove that the two lemmas are equivalent. Suppose g is computed by a threshold circuit as described by the lemmas. We claim that its negation is computed by a circuit which is identical except that the inputs are negated and if a certain gate in the circuit computing g has fanin R, total weight $w = \sum_1^R w_i$ and threshold t then the corresponding gate in the circuit computing $\bar{g}$ also has the same weights w_i but threshold $w + 1 - t$. We leave it to the reader to verify this. The equivalence of the two lemmas is now obvious. In particular we see that the first part of Lemma 4 is equivalent to the second part of Lemma 5 and the other way around.

Let us see how Theorem 3 follows from Lemma 4. Suppose f_k was computed by a threshold circuit of size $2^{\epsilon N}$ and depth $k-1$. Suppose that the top gate has fanin R, weights w_i and threshold t. Let the function computed by the i'th input of the top gate be called g_i. This function is computed by a depth $k-2$ threshold circuit of size at most $2^{\epsilon N}$ and hence we can use Lemma 4.

Let

$$\begin{aligned} S_1 &= \{i \mid g_i(p_k) \leq 2^{-2\epsilon N}\}, \\ S_2 &= \{i \mid 2^{-2\epsilon N} < g_i(p_k) \leq \frac{2}{5}\}, \\ S_3 &= \{i \mid g_i(p_k) > \frac{2}{5}\}. \end{aligned}$$

We now modify the distributions p_k and q_k in the following manner. Let p be the distribution obtained by choosing an x according to p_k while requiring that $g_i(x) = 0$ for all $i \in S_1$. Similarly, we get q by choosing x according to q_k while requiring that $g_i(x) = 1$ for all $i \in S_3$.

The probabilities $g_i(p)$ and $g_i(q)$ are close to $g_i(p_k)$ and $g_i(q_k)$ respectively. We claim that

$$\begin{aligned} g_i(p) &\leq \frac{g_i(p_k)}{1 - 2^{-\epsilon N}}, \\ g_i(q) &\geq \frac{g_i(q_k) - 2^{-\epsilon N}}{1 - 2^{-\epsilon N}}. \end{aligned}$$

In the first case, the probability mass lost is at most $R2^{-2\epsilon N} \leq 2^{-\epsilon N}$. Normalizing now yields the inequality. In the second case, call the mass lost r. The worst case is when all the x:s removed make g_i one. Normalizing gives us

$$g_i(q) \geq \frac{g_i(q_k) - r}{1 - r}$$

and since $r \leq 2^{-\epsilon N}$ the second inequality follows.

Let P be the expected weight to the top gate when the input is chosen according to p and let Q be the corresponding weight when the input is chosen according to q. Then by the assumption that the circuit computes f_k we have $P \geq t$ and $Q \leq t - 1$. Now each input to the top gate is a function g_i which satisfies the hypothesis of Lemma 4. Then we get

$$P \leq \frac{2}{5(1 - 2^{-\epsilon N})} \sum_{i \in S_2} w_i + \sum_{i \in S_3} w_i$$

$$Q \geq \frac{\frac{3}{5} - 2^{-\epsilon N}}{1 - 2^{-\epsilon N}} \sum_{i \in S_2} w_i + \sum_{i \in S_3} w_i$$

which implies

$$P - Q \leq \sum_{i \in S_2} \frac{2^{-\epsilon N} - \frac{1}{5}}{1 - 2^{-\epsilon N}} \leq 0$$

and we have reached a contradiction.

All that remains is to prove the Lemma 4. We proceed by induction and since the four statements are pairwise equivalent, we need only prove the first part of each lemma for each k. On the other hand clearly we can use both parts of the induction hypothesis.

We start by proving Lemma 4 for the base case, $k = 2$. A threshold circuit of depth 0 is just a variable and thus we need only compute the probability that a variable is 1 in the two distributions. The distribution p_2 gives the value 1 to a variable with probability $\frac{1}{N}$ while the distribution q_2 gives it with probability $1 - \frac{1}{N}$. Thus Lemma 4 is true for $k = 2$ and sufficiently large N.

Now for the induction case assume that k is even (we will later see that odd k is almost the same, but we postpone this point) and that

$g(p_k) \geq 2^{-2\epsilon N}$. We want to prove that $g(q_k) \geq \frac{3}{5}$. Consider the depth $k-2$ circuit that computes g and look at its top gate. Assume that it has R inputs and threshold t. The ith input of the circuit corresponds to a function g_i which is computed by a depth $k-3$ threshold circuit of size $\leq 2^{\epsilon N}$.

First observe that we can erase all inputs corresponding to an i such that $g_i(p_k) \leq 2^{-4\epsilon N}$ without decreasing the value of $g(p_k)$ by more than $2^{-3\epsilon N}$. For notational convenience we ignore this tiny error term.

Now define distributions p_k^l, $l = 1, 2 \ldots n$ which look like p_k on the variables corresponding to the l'th input node to the top node of f_k but assigns ones to all other variables. Using that each variable is labelled by a $2k-2$ tuple of numbers between 1 and N, then a random setting of the variables according to the distribution p_k^l can be described as follows.

Any variables whose label does not start with l is set to 1. A random value m_0 between 1 and N is picked. The variables $x_{l,m_0,*}$ (we use this notation for set of variables whose first two labels are l and m_0) is set according to p_{k-1} while $x_{l,m,*}$ are set to 0 for $m \neq m_0$.

We have a very fruitful relationship between $g_i(p_k)$ and $g_i(p_k^l)$:

Lemma 6 $g_i(p_k) \leq \prod_{l=1}^N g_i(p_k^l)$.

Proof: An instance of p_k corresponds to an instance of each of p_k^l since p_k^l only really lives on variables with first label l. If any of these instance of p_k^l forces g_i to 0 so does p_k. The lemma now follows. ∎

Observe that from this lemma and the assumption that $g_i(p_k) \geq 2^{-4\epsilon N}$ it follows that there are at most $\frac{4\epsilon N}{\log 5 - 1} < 4\epsilon N$ different l such that $g_i(p_k^l) \leq \frac{2}{5}$.

Let q_k^l be the distribution which is similar to q_k but where we fix the random choice of where to put the copies of q_{k-1} to be the l'th branch out of the top $\wedge$. We have the following immediate observation:

Lemma 7 $g_i(q_k) = \frac{1}{N} \sum_{l=1}^N g_i(q_k^l)$.

Proof: This just follows from the fact that choosing an instance from q_k can be viewed as choosing a random l and then choosing a random instance from q_k^l. Let us next see the relation between $g_i(p_k^l)$ and $g_i(q_k^l)$. ∎

Lemma 8 *We have*

1. $g_i(p_k^l) \geq 2^{-2\epsilon N} \Rightarrow g_i(q_k^l) > \frac{3}{5}$
2. $g_i(p_k^l) \geq \frac{2}{5} \Rightarrow g_i(q_k^l) > 1 - 2^{-2\epsilon N}$.

Proof: Clearly we want to establish this by the induction hypothesis. First observe that g_i is computed by a depth $k-3$ threshold circuit of size $\leq 2^{\epsilon N}$ and thus the only problem to apply the induction hypothesis is that the probability distributions are not quite correct.

Observe first that both p_k^l and q_k^l give the value 1 to all variables whose first index is not l and thus we can disregard those variables. Now let m_0 be the random choice in the definition of p_k^l which maximizes $g_i(p_k^l)$. Define two distributions p_k^{l,m_0} and q_k^{l,m_0} where the first distribution is the part of p_k^l where the random choice is actually m_0 and the second distribution is obtained from q_k^l be setting $x_{l,m,*}$ to 0 for all $m \neq m_0$. Since g_i is monotone we get

$$g_i(q_k^l) \geq g_i(q_k^{l,m_0}).$$

Now clearly the induction hypothesis applies to the pair of distributions p_k^{l,m_0} and q_k^{l,m_0} (since they are just p_{k-1} and q_{k-1} on the variables with two first indices are l and m respectively, and all other variables are constants) and finally by the choice of m_0 we have

$$g_i(p_k^l) \leq g_i(p_k^{l,m_0})$$

and adding these facts together we have proved the lemma. ∎

Next, call i *small for* l if $g_i(p_k^l) < \frac{2}{5}$. Otherwise call i *large for* l. By the previous lemma we know that when i is large for l then $g_i(q_k^l) > 1 - 2^{-2\epsilon N}$. Let N_0 be some absolute constant and remember that the top gate of the circuit computing g has R inputs, weights w_i and threshold t.

We call l *strong* if

$$\sum_{\substack{i \text{ large} \\ \text{for } l}} w_i \geq t$$

otherwise l is *weak*.

Lemma 9 *If at least* $2N/3$ *different* l *are strong, then* $g(q_k) > \frac{3}{5}$ *for* $N \geq N_0$.

Proof: Take any strong l. We claim that $g(q_k^l) \geq 1 - R \cdot 2^{-2\epsilon N} \geq 1 - 2^{-\epsilon N}$. The reason is that if for all i large for l, the value of g_i is 1, then also g is 1 (since the large i have combined weight at least t for a strong l). The probability that an individual g_i is not 1 is bounded by $2^{-2\epsilon N}$ for a large i, and the claim follows. The lemma now follows by the same argument that gave Lemma 7. ∎

To complement the above lemma we have:

Lemma 10 *If at least* $N/3$ *different* l *are weak then* $g(p_k) < 2^{-2\epsilon N}$ *for* $N \geq N_0$.

Proof: We use Lemma 6 (mainly its proof). Let $w = \sum_1^R w_i$. By the remark after the lemma, for each i there are at most $4\epsilon N$ different l such that i is small for l.

We again use the characterization that an instance $x^{(k)}$ of p_k can be viewed as choosing independent copies $x^{(k),l}$ of p_k^l for $l = 1, 2 \ldots N$ and if one of the instances $x^{(k),l}$ forces g_i to 0, then so does $x^{(k)}$. We choose the pieces $x^{(k),l}$ in a specific order to facilitate the analysis of how many g_i are forced to 0. Say that an i is *alive* if g_i is not forced to 0 by the $x^{(k),l}$ chosen this far (otherwise it is *dead*). Let an l be *unused* if $x^{(k),l}$ is not chosen yet. Now consider the following procedure:

For $j = 1, 2 \ldots N/4$

Let l_j be the unused l with maximum combined weight on its small i that are alive. Choose $x^{(k),l_j}$.

Clearly $g(p_k)$ is bounded by the probability that the combined weight of the dead i is at most $w - t$ after the above procedure. We analyze this probability.

Claim: If the weight of the dead i is at most $w - t$ at stage j then the weight of live i that are small for l_j is at least $(w - t)/2$.

Just consider the unused part of the set of at least $N/3$ different l which initially were weak. Of these l, at least $N/3 - j$ are still unused and on these unused l we initially had weight at least $(N/3 - j)(w - t)$. We know that the weight of the dead i is at most $w - t$ and since each i is small for at most $4\epsilon N$ different l, at most weight $4\epsilon N(w - t)$ is lost by these casualties. Hence one of the unused l has weight at least

$$\frac{N/3 - j - 4\epsilon N}{N/3 - j}(w - t) \geq \frac{w - t}{2}$$

on its small i:s since $\epsilon < \frac{1}{96}$. This establishes the claim.

Let $s_j = \sum_{\substack{i \text{ live and} \\ \text{small for } l_j}} w_i$.

Consider the weight on the live i small for l_j which are killed by $x^{(k),l_j}$. The expected weight killed is at least $3s_j/5$ and the maximum is s_j. Thus by trivial reasoning, with probability at least $1/2$, at least $s_j/10$ are killed. By the claim this means that for the weight of killed i to remain below $w - t$ an event which happens with probability $1/2$ must happen at most 9 times in $N/4$ trials. The probability of this is 2^{-cN} for some $c > .02 = 2\epsilon$ and sufficiently large N. This completes the proof of Lemma 10. ∎

Obviously Lemmas 9 and 10 imply 1 of Lemma 4 and next let us complete the proof when k is even by establishing 1 of Lemma 5, i.e. we need to prove that if $g(\bar{q}_k) \geq 2^{-2\epsilon N}$ then $g(\bar{p}_k) > \frac{3}{5}$. Remember that $\bar{q}_k$ is defined by first picking an element from $\bar{q}_2$ and then changing each 0 to all zeroes and each 1 to a copy of $\bar{q}_{k-1}$. The $\bar{q}_2$ distribution is chosen by first choosing an input, l, to the top $\vee$ gate and then letting each input to that gate take the value 1 while all other inputs are given the value 0. If we fix the first choice in $\bar{q}_k$ to a given l we get a distribution which we call $\bar{q}_k^l$. Choose l to maximize $g(\bar{q}_k^l)$

and call it l_0. Now let $\bar{p}_k^{l_0}$ be the distribution that looks like $\bar{p}_k$ except that all variables with first index $\neq l_0$ are set to 0. Since g is monotone we have $g(\bar{p}_k) \geq g(\bar{p}_k^{l_0})$ and by the choice of l_0 we have $g(\bar{q}_k) \leq g(\bar{q}_k^{l_0})$ and we thus only need to establish the desired relation between $g(\bar{p}_k^{l_0})$ and $g(\bar{q}_k^{l_0})$, and from now on we only look at variables with first index l_0.

Let $\bar{q}_k^{l_0,m}$ be the distribution which looks like $\bar{q}_k^{l_0}$ except that all variable with second index not equal to m are given the value 1. Now we are basically in the same situation as before. First we have

Lemma 11 $g_i(\bar{q}_k^{l_0}) \leq \prod_{m=1}^{n} g_i(\bar{q}_k^{l_0,m})$.

Proof: As Lemma 6. ∎

Now let $\bar{p}_k^{l_0,m}$ be the distribution $\bar{p}_k^{l_0}$ conditioned upon the special choice in the second level chooses that branch m. Corresponding to Lemma 7 we have

Lemma 12 $g_i(\bar{p}_k^{l_0}) = \frac{1}{N}\sum_{m=1}^{N} g_i(\bar{p}_k^{l_0,m})$.

As before, we have by induction

1. $g_i(\bar{q}_k^{l_0,m}) \geq 2^{-2\epsilon N} \Rightarrow g_i(\bar{p}_k^{l_0,m}) > \frac{3}{5}$,
2. $g_i(\bar{q}_k^{l_0,m}) \geq \frac{2}{5} \Rightarrow g_i(\bar{p}_k^{l_0,m}) > 1 - 2^{-2\epsilon N}$.

Using a proof with the same outlines as those for Lemmas 9 and 10, we can now how the following

Lemma 13

$$g(\bar{q}_k^{l_0}) \geq 2^{-2\epsilon N} \Rightarrow g(\bar{p}_k^{l_0}) > \frac{3}{5}.$$

Now, by the choice of l_0 part 1 of Lemma 5 follows. This completes the proof for k even.

The proof when k is odd is now easy. We just need to observe that when k is odd the two cases just switch (i.e. the proof of 1 of Lemma 4 is now the proof of 1 of Lemma 5 and the other way around). The reason being that f_k now has a top $\vee$-gate and $\bar{f}_k$ has a top $\wedge$-gate while the recursive structure is the same. We have completed the proof of Lemma 4.

Acknowledgment: We would like to thank Andy Yao for sending an early draft of his paper and for encouraging us to publish the present paper. We are also grateful to Oded Goldreich for some helpful comments.

References

[AB87] N. Alon and R. B. Boppana. The monotone circuit complexity of boolean functions. *Combinatorica*, 7:1–22, 1987.

[Ajt83] M. Ajtai. $\sum_1^1$-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[All89] E. Allender. A note on the power of threshold circuits. *Proceedings 30'th Annual Symposium on Foundations of Computer Science*, pages 580–584, 1989.

[And85] A. E. Andreev. On a method for obtaining lower bounds for the complexity of individual monotone functions. *Dokl. Ak. Nauk. SSSR 282*, pages 1033–1037, 1985. English translation in *Sov. Math. Dokl.*, 31:530–534, 1985.

[BCH84] P. W. Beame, S. A. Cook, and H.J. Hoover. Log depth circuits for division and related problems. *Proceedings 25'th Annual Symposium on Foundations of Computer Science*, pages 1–6, 1984.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multipary protocols and logspace-hard pseudorandom sequences. *Proceedings of 21st Annual ACM Symposium on Theory of Computing*, pages 1–11, 1989.

[FSS84] M. Furst, J. Saxe, and M. Sipser. Parity, circuits, and the polynomial time hierarchy. *Math. System Theory*, 17:13–27, 1984.

[HMP+87] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan. Threshold circuits of bounded depth. *Proceedings 28th Annual IEEE Symposium on Foundation of computer science*, pages 99–110, 1987.

[Hås86] J. Håstad. *Computational Limitations of Small-Depth Circuits.* MIT PRESS, 1986.

[KW88] M. Karchmer and A. Wigderson. Monotone circuits for connectivity require super-logarithmic depth. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, 1988.

[Raz] A. A. Razborov. Lower bounds for the the size of circuits of bounded depth with basis $\wedge, \oplus$. preprint (in Russian). To appear in *Matem. Zam.*

[Raz85] A. A. Razborov. Lower bounds on monotone network complexity of the logical permanent. *Matem. Zam.*, 37(6):887–900, 1985. English translation in *Math. Notes of the Academy of Sciences of the USSR*, 37:485–493, 1985.

[RW89] R. Raz and A. Wigderson. Probabilistic communication complexity of boolean relations. In *Proceedings of the 30th Annual IEEE Symposium on Foundation of computer science*, 1989.

[RW90] R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. *22nd annual ACM Symposium on Theory of Computing*, pages 287–292, 1990.

[Smo87] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. *Proceedings of 19th Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[Yao] A. Yao. On acc and threshold circuits. manuscript.

[Yao85] A. Yao. Separating the polynomial-time hierarchy by oracles. *Proceedings 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.

[Yao89] A. Yao. Circuits and local computation. *Proceedings of 21st Annual ACM Symposium on Theory of Computing*, pages 186–196, 1989.