Depth Reduction for Circuits of Unbounded Fan-In*

Eric Allender†
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

Ulrich Hertrampf
Institut für Informatik
Universität Würzburg
D-8700 Würzburg
Federal Republic of Germany

December 12, 1991

*The results in this paper were originally announced in papers in Proc. 30th IEEE Symposium on Foundations of Computer Science (1989) and in Proc. 15th International Symposium on Mathematical Foundations of Computer Science (1990).

†Supported in part by National Science Foundation grants CCR-8810467 and CCR-9000045. Some of this research was performed while the first author was a visiting professor at Institut für Informatik, Universität Würzburg.

Depth Reduction for Circuits

Send proofs to the following address:
Eric Allender
Dept. of Computer Science
Rutgers University
New Brunswick, NJ 08903

Abstract. *We prove that constant depth circuits of size $n^{\log^{O(1)} n}$ over the basis AND, OR, PARITY are no more powerful than circuits of this size with depth four. Similar techniques are used to obtain several other depth reduction theorems; in particular, we show every set in AC^0 can be recognized by a family of depth-three threshold circuits of size $n^{\log^{O(1)} n}$. The size bound $n^{\log^{O(1)} n}$ is optimal when considering depth reduction over AND, OR, and PARITY. Most of our results hold both for the uniform and the nonuniform case.*

List of Symbols

$\in$ – element of
$\notin$ – not an element of
$\subseteq$ – contained in
$\not\subseteq$ – not contained in
$=$ – equal
$\neq$ – not equal to
$\leq$ – less than or equal to
$<$ – less than
$\geq$ – greater than or equal to
$>$ – greater than
$\cup$ – union
$\oplus$ – Exclusive OR, or PARITY
$\exists$ – logical "exists"
$\forall$ – logical "for all"
$\Longrightarrow$ – implies
$\Longleftrightarrow$ – if and only if
$\rightarrow$ – "into", used in the definition of functions
$\wedge$ – logical "and"
$\vee$ – logical "or"
$\neg$ – logical "not"
/ – a slanted bar, used in fractions.
$\{, \}$ – set brackets
$[,]$ – square brackets, used like parentheses
$'$– the "prime" sign, used in expressions such as $\sum'$
Σ – capital sigma, used in expressions such as Σ^n.
$\sum$ – the summation symbol, used in expressions such as $\sum_{j=1}^{m}$.
$\prod$ – the product symbol, used in expressions such as $\prod_{i=0}^{m-1}$
0 – zero
O – Capital Oh, sometimes the expression $O(\ldots)$ is used; note that this is "oh" and not "zero". Examples of usage follow: $O(\log n)$, $n^{O(1)}$.
1 – one
l – italics letter ell, used in expressions such as n^l
$\blacksquare$– used at the end of proofs.
| – a vertical bar, used in expressions such as $|x|$.

1 Introduction

In this paper, we present a number of depth-reduction theorems: theorems that state that certain classes of circuits of a given depth can be simulated efficiently by circuits of smaller depth. Before presenting our results, let us first contrast our work with earlier work showing that in a variety of settings, depth reduction cannot be achieved.

Some of the strongest lower bound results in complexity theory are bounds on the size required to compute functions on circuits of AND and OR gates of unbounded fan-in. (Throughout this paper, unless we explicitly say otherwise, negated input bits are available at the input level of the circuit, and thus NOT gates are not needed, by DeMorgan's Laws.) It was shown in Yao (1985) and Håstad (1987) that for all k there is a function computed by such circuits of linear size and depth k that cannot be computed on such circuits of depth $k-1$ and size less than exponential. That is, Yao (1985) and Håstad (1987) show that depth reduction is impossible in this setting.

Another circuit model that is of interest is the *threshold circuit* model; a threshold circuit is a circuit composed of MAJORITY gates. (A MAJORITY gate is a gate that takes the value 1 iff more than half of its inputs have the value 1. Note that in the literature, the term "threshold circuit" is used to refer to any of a number of closely-related models of computation. In particular, some work considers a model in which arbitrary real numbers are allowed as weights on the outputs of subcircuits; this model can be simulated efficiently by MAJORITY circuits by at most doubling the depth (Siu and Bruck (1991)). Throughout this paper, all results will be stated in terms of the model defined using MAJORITY gates.) Threshold circuits are studied in part because MAJORITY gates have roughly the same computational power as integer multiplication gates (Chandra et al. (1984)), and also because the "neural net" model of the brain is computationally equivalent to a threshold circuit (Parberry and Schnitger (1989), Parberry (1990)). Little is known about depth reduction in the context of threshold circuits; the best results in this direction are the results of Hajnal et al. (1987), where it was shown that there is a language recognized by a family of polynomial-size depth three majority circuits that cannot be recognized by polynomial-size depth two majority circuits.

In order to better understand the threshold circuit model, Yao (1989) initiated

a study of *monotone* threshold circuits. Yao's results were subsequently improved by Håstad and Goldmann (1990), who showed that for all k there is a function computed by monotone linear-size depth k circuits of AND and OR gates, that cannot be computed in less than exponential size by monotone threshold circuits of depth $k-1$.

In this paper, we show that the results of Yao, Håstad and Goldman fail in the non-monotone case; every language accepted by depth k polynomial-size circuits of AND and OR gates is accepted by depth three threshold circuits of size $n^{O(\log^k n)}$. That is, in contrast to the monotone case, depth k circuits can be simulated by depth 3 threshold circuits, with only a "modest" increase in size.

Another circuit model that has received considerable attention consists of circuits with AND and OR gates and MODp gates. Smolensky (1987), building on work by Razborov (1987), showed that constant-depth circuits of this type with MODp gates for a fixed prime modulus p can be approximated by low-degree polynomials, and then used this fact to derive lower bounds on the size of such circuits to compute MODq, for $q \neq p$. We show that for circuits of this type of size $2^{\log^{O(1)} n}$, constant depth is no more powerful than depth 4.

Our results are closely related to the important work of Toda (1991), showing that the polynomial hierarchy is contained in P^{PP}. Connections between the polynomial hierarchy and constant-depth families of circuits were established in Furst et al. (1984), and similar observations regarding threshold circuits and PP were made in Torán (1991). By making use of those connections, our results can be viewed as providing a circuit-based interpretation of some of Toda's work.

The paper is organized as follows. In Section 2 we present the basic definitions and notational conventions. In Section 3 we prove some elementary lemmas using algebraic properties allowing us to convert a circuit into a simpler form. In section 4 we prove our main depth reduction results in the framework of nonuniform circuit complexity. Section 5 contains the main lemmas needed to make these depth reduction results carry over into the uniform setting, and our uniform depth reduction results are presented in Sections 6 and 7. A summary is found in Section 8.

2 Definitions and Background

We assume familiarity with the basics of circuit complexity. For additional background, see Boppana and Sipser (1990), Barrington et al. (1990), and Ruzzo (1981).

A *family of circuits* is a set $\{C_n : n \geq 1\}$ where each C_n is a circuit for inputs of length n. In later sections of this paper, we will require that the function $n \mapsto C_n$ be easily computable in some sense (which will be made precise there). Such circuit families are called *uniform*. If no such restriction on constructibility is imposed, the circuit families are called *nonuniform*.

AC^0 is the class of languages accepted by families of circuits of polynomial size and depth $O(1)$, consisting of unbounded fan-in AND and OR gates. AC^0_k denotes the subclass of AC^0 accepted by circuits of depth k.

It is no loss of generality to consider only circuits that are "leveled" in the sense that each gate can be assigned to a "level" denoting the distance from the gate to the input level, where inputs to a gate at level i come only from gates at level $i-1$. Thus the inputs to a circuit are at level 0, the gates that directly process those inputs are at level 1, and the output gate of a depth k circuit is at level k.

We will also have cause to consider circuit complexity classes defined by other size bounds and using other types of gates. That leads to the following definition:

Definition 1

(i) SIZE($s(n)$)DEPTH($d(n)$) GATES(S) denotes the class of languages which can be recognized by circuit families of size $s(n)$ and depth $d(n)$ where the types of gates which can be used are in the set S.

(ii) BPSIZE($s(n)$)DEPTH($d(n)$) GATES(S) denotes the analogous class which is defined in terms of probabilistic circuits. That is, the circuits have some number of probabilistic bits as auxiliary inputs, drawn from the uniform distribution. A probabilistic circuit C with n inputs recognizes a set $L \subseteq \Sigma^n$ if for all $x \in L$, $\text{Prob}(C(x) = 1) > \frac{3}{4}$, and for all $x \notin L$, $\text{Prob}(C(x) = 1) < \frac{1}{4}$.

3 Depth Reduction and the Distributive Law

The main tool that we will use in later sections to carry out depth reduction is to view a circuit in a certain form as a polynomial of low degree, and then express that polynomial in standard form. (Of course, it is not a new observation that circuits correspond to polynomials over finite fields. In particular, this connection was used very effectively in Razborov (1987) and Smolensky (1987).) In this section, we derive the particular bounds that we will need later on.

As the $\oplus$-operation is exactly the same as addition in the field GF(2) and the $\wedge$-operation is exactly the same as multiplication, we can view a $\oplus$-gate of fan-in s as a sum over s summands, and similarly an $\wedge$-gate of fan-in r as a product of r factors. In this way, it is clear that a circuit of AND and PARITY gates may be viewed as a polynomial over GF(2). Next we note that similar observations hold for MODp gates for any prime p, and use this to show that depth two circuits of $\wedge$ and MODp gates can efficiently simulate circuits of greater depth.

Definition 2 A MODp gate is a gate that outputs 1 if and only if the number of inputs with value 1 is not divisible by p.

Lemma 3 Let $p \geq 2$ be prime. Then every four level circuit consisting of

- AND gates with fan-in r on level 1,
- MODp gates with fan-in s_2 on level 2,
- AND gates with fan-in t on level 3,
- one MODp gate with fan-in s_1 on level 4,

can be converted into a two level circuit consisting of AND gates with fan-in $r \cdot t \cdot (p-1)$ on level 1 and one MODp gate with fan-in $s_1 \cdot s_2^{t \cdot (p-1)}$ on level 2.

The conversion can be carried out using space logarithmic in the size of the resulting circuit.

Proof: A MODp gate does not correspond directly to a sum over the field GF(p). Instead using Fermat's theorem ($a^{p-1} \equiv 1 \mod p \iff a \not\equiv 0 \mod p$) one can see that a MOD$p$ gate with inputs $y_1, \ldots, y_s$ computes the function $\sum'(y_1, \ldots, y_s) := \left(\sum_{i \in \{1,\ldots,s\}} y_i \right)^{p-1}$ over GF(p). Obviously the $(p-1)$-th power

can be written as a product over $p-1$ (equal) factors. On the other hand, an AND gate computes multiplication over any field (if the inputs are only 0 or 1).

Our four level circuit may be described with an expression of the form

$$\sum_{(1\le i\le s_1)}{}' \prod_{(1\le j\le t)} \sum_{(1\le k\le s_2)}{}' \prod_{(1\le l\le r)} x_{i,j,k,l}.$$

Since we focus on the fan-in of the various gates, we will express this in the abbreviated form

$$\sum_{(s_1)}{}' \prod_{(t)} \sum_{(s_2)}{}' \prod_{(r)} (\ldots).$$

Thus over GF(p) we can rearrange our four level circuit as follows:

$$\begin{aligned}
\sum_{(s_1)}{}' \prod_{(t)} \sum_{(s_2)}{}' \prod_{(r)} (\ldots) &= \left(\sum_{(s_1)} \prod_{(t)} \prod_{(p-1)} \sum_{(s_2)} \prod_{(r)} (\ldots) \right)^{p-1} \\
&= \left(\sum_{(s_1)} \prod_{(t\cdot(p-1))} \sum_{(s_2)} \prod_{(r)} (\ldots) \right)^{p-1} \\
&= \left(\sum_{(s_1)} \sum_{(s_2^{t\cdot(p-1)})} \prod_{(t\cdot(p-1))} \prod_{(r)} (\ldots) \right)^{p-1} \\
&= \left(\sum_{(s_1\cdot s_2^{t\cdot(p-1)})} \prod_{(r\cdot t\cdot(p-1))} (\ldots) \right)^{p-1} \\
&= \sum_{(s_1\cdot s_2^{t\cdot(p-1)})}{}' \prod_{(r\cdot t\cdot(p-1))} (\ldots)
\end{aligned}$$

The final formula represents the desired circuit.

To complete the proof, we need only remark that each step in this transformation can easily be carried out by a Turing machine that needs only to store a constant number of indices (i.e., gate names) at any one time. Thus the entire transformation can be done using space logarithmic in the size of the resulting circuit. (Perhaps the only step where uniformity might not be entirely obvious is the step where the product is distributed over the sum. Thus consider a subcircuit of the form $\prod_{(r)} \sum_{(s)} (\ldots)$. The replacement subcircuit $\sum_{(s^r)} \prod_{(r)} (\ldots)$ is formed by taking the sum of all r-tuples formed by taking the Cartesian product of all the inputs to the original summation gates. A Turing machine can clearly carry out the required enumeration of r-tuples in the given space.) ∎

4 Depth Reduction for Nonuniform Circuits

In this section, we present our main depth reduction theorems. Unfortunately, the proofs presented in this section work only for nonuniform circuit families. In later sections, we show how to achieve depth reduction for uniform circuit families. Our reasons for including this section are:

- The proofs are much simpler in the nonuniform setting.
- Certain kinds of depth reduction are not yet known to hold in the uniform setting.
- We are able to achieve slightly better size bounds in the nonuniform setting.

The proof outline may be given quite simply:

- Show that constant depth circuits can be simulated by probabilistic depth two circuits. (This simulation is very similar to the result of Razborov (1987) (generalized by Barrington (1987) and Smolensky (1987)) showing that circuits with small depth can be approximated by polynomials of small degree. Our proof is also very similar to his. The difference is that we need a probabilistic circuit that works well on *all* inputs, as opposed to a deterministic circuit that works well on *most* inputs. It is possible to use the result of Smolensky (1987) to derive the result we need, but we feel it is simpler and more transparent to give a direct proof.)
- Use established techniques to make these probabilistic circuits deterministic. (Exactly which techniques are used depends on the type of deterministic circuit being constructed.)

Lemma 4 For any prime p and any constant k, there is a family of probabilistic depth-two circuits of size $n^{O(\log n)}$, computing the OR of n bits, with error less than $1/n^k$. The first level of this circuit consists of ANDs of fan-in $O(\log n)$, and the second level consists of a MODp gate.

Proof: In order to simplify the exposition, assume that $p = 2$. The generalization to other primes is straightforward.

In order to compute the OR of $b_1, b_2, \ldots b_n$, first consider the circuit B_n with one PARITY gate, where the inputs to the parity gate are

$$\{1\} \cup \{\mathrm{AND}(b_i, p_i) : 1 \leq i \leq n\},$$

where the p_i are probabilistic bits. It is easy to see that if $\mathrm{OR}(b_1 \ldots b_n) = 0$, then B_n outputs 1, and if $\mathrm{OR}(b_1 \ldots b_n) = 1$, then B_n outputs 0 with probability exactly $1/2$. (When carrying out the generalization to other primes p, this error will be $1/p$, requiring the expression "$k \log n$" in the next paragraph to be replaced by "$ck \log n$" for some constant c. The statement of the lemma does not depend on p, however.)

Now take $k \log n$ separate copies of B_n (with independent probabilistic inputs for each copy of B_n) and AND these $k \log n$ circuits together. Call this new circuit C_n. It is immediate that if $\mathrm{OR}(b_1 \ldots b_n) = 0$, then C_n outputs 1, and if $\mathrm{OR}(b_1 \ldots b_n) = 1$, then C_n outputs 0 with probability $1 - 1/n^k$.

Now by Lemma 3 (letting $s_1 = 1, t = k \log n, s_2 = n+1$, and $r = 2$), C_n can be converted into an equivalent circuit D_n consisting of a PARITY gate of $n^{O(\log n)}$ AND gates, where each AND gate has fan-in $O(\log n)$. Let E_n be the circuit that computes the negation of this D_n (i.e., the PARITY gate has an additional 1 input). Then E_n is a circuit with the properties claimed by the lemma. ∎

Corollary 5 For any prime p and any constant k, there is a family of probabilistic depth-two circuits of size $n^{O(\log n)}$, computing the AND of n bits, with error less than $1/n^k$. The first level of this circuit consists of ANDs of fan-in $O(\log n)$, and the second level consists of a MODp gate.

Proof: To compute the AND of $b_1, \ldots, b_n$, take the circuit D_n constructed in the proof of Lemma 4 and apply it to the inputs $\neg b_1, \ldots, \neg b_n$. This computes $\neg\mathrm{OR}(\neg b_1, \ldots, \neg b_n) = \mathrm{AND}(b_1, \ldots, b_n)$. ∎

Theorem 6 Let p be any prime, let q be any polynomial, and let L be accepted by a polynomial-size $\{\wedge, \vee, \mathrm{MOD}p\}$-circuit family of depth k. Then L is accepted by a probabilistic circuit family of depth two with error less than $1/q(n)$, where the first level of each circuit consists of $n^{O(\log^k n)}$ ANDs of fan-in $O(\log^k n)$, and the second level consists of a MODp gate.

Proof: Again, to simplify the exposition, we assume that $p = 2$. The generalization to arbitrary primes p is straightforward.

The proof proceeds by induction on k. The nontrivial parts of the basis case are proved in Lemma 4 and Corollary 5. For the induction step, let L be accepted

by a family of depth k circuits of polynomially-many unbounded-fan-in AND, OR, and PARITY gates. Consider the circuit C_n for inputs of length n. Assume without loss of generality that the output gate of C_n is an AND gate (the proof is entirely symmetric when it is an OR gate, it is trivial if it is a PARITY gate, and the slightly more general MODp case proceeds along similar lines, using Lemma 3.). Thus C_n is the AND of at most n^l circuits of depth $k-1$ for some l. By the inductive hypothesis, each of these n^l circuits may be replaced by a probabilistic depth 2 circuit of size $n^{O(\log^{k-1} n)}$, having error probability at most $1/n^a$ (where a may be any constant). The resulting circuit has error probability at most n^l/n^a. Also, the top-level AND in this circuit can be replaced by a probabilistic depth-two circuit of the sort guaranteed by Corollary 5; the resulting circuit is of the form

$$\sum_{(n^{O(\log n)})} \prod_{(O(\log n))} \sum_{(n^{O(\log^{k-1} n)})} \prod_{(O(\log^{k-1} n))} (\ldots)$$

and may be constructed to have error probability less than $1/q(n)$. (The only dependence on the polynomial q is in the choice of the constant a above, and in the constants in the various "$O(\log \ldots)$" terms in this expression.) The proof is completed with an appeal to Lemma 3. ∎

Theorem 6 shows how to simulate a deterministic circuit by a probabilistic circuit of depth two. Now it remains only to simulate the probabilistic circuits by deterministic circuits; this can be done using known techniques. In order to justify the precise bounds we claim, we present the details below in Theorems 7 and 9.

Theorem 7 Let p be any prime, and let L be accepted by a polynomial-size $\{\wedge,\vee,\text{MOD}p\}$-circuit family of depth k. Then L is accepted by a $\{\wedge,\vee,\text{MOD}p\}$-circuit family of depth four and size $n^{O(\log^k n)}$.

Proof: We use the technique of Ajtai and Ben-Or (1984).

By Theorem 6, L is accepted by a family of probabilistic $\{\wedge,\vee,\text{MOD}p\}$-circuits of depth two with error probability less than $\frac{1}{2n^2}$. Now we take n^2 of these circuits and AND them together on level 3. The error in the positive case (the input should be accepted) is still less than $\frac{1}{2}$ while the error in the negative case (the input should be rejected) is now less than $\frac{1}{(2n^2)^{n^2}}$ which is far less than $\frac{1}{2^{n^2}}$. Now take n of these depth 3 circuits and OR them together. This results in a depth

4 circuit with error less than $\frac{1}{2^n}$ in the positive case, and error less than $\frac{n}{2^{n^2}}$ in the negative case. The error probability being less than 2^{-n} in both directions one can use the argument of Adleman (1978) and Ajtai and Ben-Or (1984) that there must be values for the random bits that always lead to the correct output. Thus the four level circuit can be made deterministic (in a nonuniform way). ∎

Theorem 7 shows that depth four suffices for $\{\wedge,\vee,\text{MOD}p\}$-circuits. By using the more powerful MAJORITY gates, we can reduce the depth even further. First, however, we need an easy proposition showing that MODp gates can be replaced by MAJORITY gates in some situations, without increasing circuit depth. (This type of simulation is a more-or-less standard technique; see, for example, Bruck (1990). A more general result of this sort is proved by Bultman (1990).)

Proposition 8 If C is a depth two circuit with one MAJORITY gate as output and r MODp gates on level 1, where no MODp gate has more than pm inputs, then C is equivalent to a depth two threshold circuit with at most $2(p-1)mr+1$ MAJORITY gates.

Proof: We may assume without loss of generality that all the gates have exactly pm inputs. Let G be one of the MODp gates. For each integer $j, 1 \leq j < p$ and each integer $i \equiv j \bmod p$, $1 \leq i \leq pm$, build two MAJORITY gates, one which accepts if at least i of the pm inputs are 1, and one which accepts if at most i of the pm inputs are 1. It is immediate that if the number of the pm inputs that are 1 is equivalent to 0 mod p, then exactly $(p-1)m$ of the new MAJORITY gates will be 1, while otherwise exactly $(p-1)m+1$ of these gates will be 1. Replace G with the $2(p-1)m$ MAJORITY gates constructed in this way, and repeat this process for each of the other MODp gates in the circuit.

Now have the MAJORITY gate at level 2 accept iff at least $(p-1)rm+\frac{r}{2}$ of the MAJORITY gates on level 1 have value 1. ∎

Theorem 9 Let p be a prime. Then every set accepted by a polynomial-size family of $\{\wedge,\vee,\text{MOD}p\}$-circuits of depth k is accepted by a family of depth three threshold circuits of size $n^{O(\log^k n)}$.

Proof: The depth two probabilistic circuit constructed in Theorem 6 can be converted into a deterministic depth three circuit using the technique of Proposition 4.2 in Hajnal et al. (1987); this involves (1) building $2n$ independent copies of

the probabilistic depth two circuit and taking the MAJORITY of their outputs, and (2) noting that the resulting circuit has exponentially small error probability, and thus (nonuniformly) there exists a sequence of probabilistic bits that may be hardwired in, yielding a correct circuit. The resulting circuit is only polynomially larger than the original probabilistic circuit, and consists of AND gates on level 1, MODp gates on level 2, and a MAJORITY gate as the output gate.

The theorem now follows by Proposition 8 and the trivial reducibility of AND and OR to MAJORITY. ∎

The depth-reduction theorems are even more striking when stated in terms of circuits of size $2^{\log^{O(1)} n}$, instead of polynomial size. Note that using simple translational techniques (with the translation $x \mapsto x0^{2^{\log^l |x|}}$) one can use Theorem 6 to convert a depth k circuit of size $2^{\log^l n}$ into an equivalent probabilistic depth two circuit of size $2^{\log^{(k+1)l} n}$. This observation yields the following corollary.

Corollary 10

$$\begin{aligned} & \text{BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(O(1))\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\ = \;& \text{BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(2)\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\ = \;& \text{SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(4)\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\ \subseteq \;& \text{SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(3)\,\text{GATES}(\{\text{MAJORITY}\}) \end{aligned}$$

Proof: Immediate from Theorems 6, 7, and 9, and from the result of Ajtai and Ben-Or (1984) that any probabilistic circuit may be made deterministic (in the nonuniform setting) by increasing the size by a polynomial factor, and increasing the depth by an additive constant. ∎

Most, but not quite all, of these depth-reduction theorems are also known to hold in the setting of uniform circuit complexity. This is taken up in the following sections.

It is natural to wonder if these depth reduction results can be improved. For example, is every set in AC^0 accepted by polynomial-sized threshold circuits of depth three? Although that question is still open, we note that this sort of depth reduction cannot be achieved by polynomial-sized $\{\wedge,\vee,\text{MOD}p\}$-circuits; in fact, the bound $2^{\log^{O(1)} n}$ is optimal.

Proposition 11 For all $k, r \geq 1$ there exists an l such that

$$AC^0_l \not\subseteq \text{SIZE}(2^{O(\log^r n)})\text{DEPTH}(k)\,\text{GATES}(\wedge,\vee,\oplus)$$

Proof: By the result of Smolensky (1987), the computation of MOD_3 of n input bits by depth k circuits with AND, OR, and PARITY gates requires at least size $2^{O(n^{1/2k})}$. Thus computing the MOD_3 of $\log^{2kr+1} n$ bits in depth k needs size greater than $2^{O(\log^r n)}$. However, it is well-known that this function can be computed by AC^0 circuits (Fagin et al. (1985), Denenberg et al. (1986)). ∎

5 A Uniform Simulation

Although the proofs presented in Section 4 are quite simple, they suffer from the drawback that they are only suitable for *nonuniform* circuit complexity. That is, if L is accepted by a family of AC^0 circuits $\{C_n : n \in \mathbf{N}\}$ such that the function $n \mapsto C_n$ is efficiently computable, then the results of Section 4 tell us that there *exists* a family $\{D_n\}$ of depth-three threshold circuits of size $2^{\log^{O(1)} n}$ accepting L, but we have no guarantee that there is any efficient way to *construct* the circuits D_n.

The reason for this is that the proofs in the preceding section make use of probabilistic constructions. Although we were able to make use of established techniques for turning probabilistic circuits into deterministic circuits (as in Adleman (1978), Parberry and Schnitger (1988), and Hajnal et al. (1987)), these techniques seem to be inherently nonuniform.

In the literature on circuit complexity, a circuit family C_n is called "uniform" if the function $n \mapsto C_n$ is "easy to compute" in some sense. There are many notions of uniformity that are worthy of consideration, each with a different notion of "easy to compute." For example, in one of the first papers to consider uniform circuit complexity, Ruzzo (1981) considers a variety of uniformity notions, and P-uniform circuit complexity is discussed in Allender (1989a).

Even more relevant to this paper are the notions of uniformity discussed in Barrington et al. (1990). In that paper, Barrington et al. consider what version of uniform circuit complexity is most appropriate for use in defining classes of languages accepted by circuits of polynomial size and depth $O(1)$.

With some work, it would be possible to adapt the definitions of Barrington et al. (1990) to make them applicable to circuits of size $2^{\log^{O(1)} n}$. In the interest of simplicity, however, we do not choose to do that here. Instead, we use a rather generous notion of uniformity. By doing so, it will be obvious that the circuits we construct are uniform because of their regularity, whereas if we were to use

a more stringent notion of uniformity, it would be necessary to argue at length that the circuits satisfy the requirements of the uniformity condition.

Note that any machine that constructs a circuit of size $2^{\log^{O(1)} n}$ must use at least space $\log^{O(1)} n$. That leads us to the following

Convention: Throughout the rest of this paper, a family of circuits C_n will be said to be *uniform* if the function $n \mapsto C_n$ is computable in space $\log^{O(1)} n$.

Note that the composition of two functions computable in $\log^{O(1)} n$ space is also computable in $\log^{O(1)} n$ space. Thus when we show that a circuit C_n of one sort can be converted into an equivalent circuit D_n of another sort via a constant number of transformations, each of which is computable in $\log^{O(1)} n$ space, we can conclude that the circuit family $\{D_n\}$ is uniform if $\{C_n\}$ is uniform.

In order to achieve depth reduction in the setting of uniform circuits, we will drastically reduce the number of probabilistic bits used by depth two probabilistic circuits computing AND and OR. To do this we make use of the following result by Valiant and Vazirani (1986):

Theorem 12 Let $n \geq 1$ and let $S \subseteq \{0,1\}^n$ be a nonempty set. Suppose w_1, w_2, ..., w_n are randomly chosen from $\{0,1\}^n$. Let $S_0 = S$ and let $S_i = \{v \in S : v \cdot w_1 = v \cdot w_2 = \ldots = v \cdot w_i = 0\}$ for each $i \in \{1, \ldots, n\}$ (where the dot product of two vectors v, w of length m is $v \cdot w = \sum_{j=1}^{m} v_j w_j \bmod 2$). Let $P_n(S)$ be the probability that $|S_i| = 1$ for some $i \in \{0, \ldots, n\}$. Then $P_n(S) \geq \frac{1}{4}$.

Proof: See Theorem 2.4 in Valiant and Vazirani (1986).

Lemma 13 Let p be a prime number. Then for any constant c, there is a uniform family of probabilistic depth-two circuits of size $2^{O(\log^4 n)}$, with $O(\log^3 n)$ probabilistic bits, computing the OR (AND, MODp) of n bits with error less than $\frac{1}{n^c}$. The first level of this circuit consists of AND gates of fan-in $O(\log^4 n)$, and the second level consists of one MODp gate.

Proof: We give the proof only for the computation of OR, for the computation of AND essentially the same proof works, starting with the negated inputs and using de Morgan's laws. For MODp the statement holds trivially.

Let m be the least integer that is strictly greater than $\log_2 n$. We will construct a circuit $C_1^{(n)}$ with $n + m^2$ inputs, namely the original inputs $x_1, \ldots, x_n$ and

m^2 probabilistic inputs, arranged as m vectors of dimension m: $w_{1,1}, \ldots, w_{1,m}$, $w_{2,1}, \ldots, w_{2,m}$, ..., $w_{m,1}, \ldots, w_{m,m}$. An integer a in the range $1, \ldots, n$ can be viewed as a sequence $a_1, \ldots, a_m$ of m bits and thus it makes sense to define $a \cdot w_i = \sum_{j=1}^{m} a_j w_{i,j} \bmod 2$.

Consider the following five level circuit:

- Level 2 consists of $n \cdot m$ MODp gates $P_{a,k} (1 \leq a \leq n,\ 1 \leq k \leq m)$; these gates will compute the value $a \cdot w_k$. To see how to do this, note that the value of $a \cdot w_k$ is computed by taking the PARITY of $O(\log n)$ bits. (Exactly which bits of w_k take part in this computation depends on the constant a.) The DNF expression of this PARITY function can be expressed with $n^{O(1)}$ AND gates, with the property that on any given input, either none of the AND gates evaluates to true, or exactly one does. This can clearly be computed by a subcircuit with a MODp gate on level 2, with $n^{O(1)}$ AND gates on level 1, where the AND gates have fan-in $O(\log n)$. The collection of all these subcircuits can be produced in logarithmic space.

- Level 3 consists of $n(m+1)$ gates $D_{a,k} (1 \leq a \leq n,\ 0 \leq k \leq m)$ which shall take the value 1 if and only if $x_a = 1$ and $a \cdot w_i = 0$ for all $i \leq k$. These gates can obviously be realized as AND gates with fan-in $O(\log n)$, using the $P_{a,j}$ with $j \leq k$.

- Level 4 consists of $m+1$ gates $E_k (0 \leq k \leq m)$ which shall take the value 1 if and only if p does not divide $(p-1)$ plus the number of a's $(1 \leq a \leq n)$ such that $x_a = 1$ and $a \cdot w_i = 0$ for all $i \leq k$. These gates can obviously be realized as MODp gates with fan-in $n + (p-1)$, the inputs being the outputs of the $D_{a,k} (1 \leq a \leq n)$ and (p-1) constants 1. Note that in the case where $x_a = 0$ for all $a \in \{1, \ldots, n\}$, all E_k have value 1.

- Level 5 consists of one AND gate F of fan-in $m+1$, with the $E_k (0 \leq k \leq m)$ as inputs.

This circuit will output 1, if $OR(x_1, \ldots, x_n) = 0$. But if $OR(x_1, \ldots, x_n) = 1$, the set of all a, such that $x_a = 1$ is nonempty. Thus by Theorem 12, with probability at least $\frac{1}{4}$ there is a k, such that $D_{a,k}$ has value 1 for exactly one a; hence E_k has value 0, and consequently $C_1^{(n)}$ outputs 0 with probability at least $\frac{1}{4}$.

To amplify the probability in the case that the OR should be 1, we take $3c \log n$ independent copies of the circuit, each using the same inputs $x_1, \ldots, x_n$ but each with its own set of probabilistic bits. We combine the outputs of these circuits by one AND gate and call the resulting circuit $C_2^{(n)}$. Now it follows that

(i) $OR(x_1, \ldots, x_n) = 0 \implies C_2^{(n)}$ outputs 1

(ii) $OR(x_1, \ldots, x_n) = 1 \implies \text{Prob}(C_2^{(n)} \text{ outputs } 0) \geq 1 - (\frac{3}{4})^{3c \log n} > 1 - \frac{1}{n^c}$.

The resulting circuit is of the form

$$\prod_{(O(\log n))} \prod_{(O(\log n))} \sum_{(O(n))}{}' \prod_{(O(\log n))} \sum_{(n^{O(1)})}{}' \prod_{(O(\log n))} (\ldots)$$

which can be rearranged to the form

$$\sum_{(2^{O(\log^4 n)})}{}' \prod_{(O(\log^4 n))} (\ldots)$$

using Lemma 3.

This circuit computes the NOR of $x_1, \ldots, x_n$ with error less than $\frac{1}{n^c}$, using only $O(\log^3 n)$ probabilistic bits. Now note that it follows easily from Lemma 3 that the negation of this function can be computed by a circuit of essentially the same size, since $\neg \sum_{(r)}{}' (\ldots) = \sum{}'((p-1) + (\sum_{(r)}{}' (\ldots)))$.

All constructions can be carried out in space logarithmic in the size of the resulting circuit. ∎

Now we can inductively use Lemma 13 to show that all constant depth $\{\wedge, \vee, \oplus\}$-circuits of size at most $2^{\log^{O(1)} n}$ can be simulated in depth 2, using only a polylogarithmic number of probabilistic bits.

Lemma 14 Let p be a prime number, and let L be accepted by a uniform family of $\{\wedge, \vee, \text{MOD}p\}$-circuits of depth k and size $2^{O(\log^r n)}$, and let c be a constant. Then L is accepted by a uniform family of probabilistic circuits of depth two with error less than $\frac{1}{2^{c \log^r n}}$. The first level of the circuit consists of $2^{O(\log^{4kr} n)}$ AND gates of fan-in $O(\log^{4kr} n)$, the second level consists of one MODp gate, and only $O(\log^{3r} n)$ probabilistic bits are used.

Proof: The proof proceeds by induction on k. The basis case follows from Lemma 13.

For the induction step, let L be accepted by a uniform family of depth k circuits of size $2^{O(\log^r n)}$ with gates $\{\wedge,\vee,\mathrm{MOD}p\}$. Let c be any constant. Consider the circuit $C^{(n)}$ for inputs of length n. Assume without loss of generality that the output gate of $C^{(n)}$ is an AND gate (the proof is entirely symmetric when it is an OR gate, and it is trivial when it is a MODp gate). Thus $C^{(n)}$ is the AND of at most $2^{d\cdot\log^r n}$ circuits of depth $k-1$ for some constant d. By the inductive hypothesis, each of these circuits may be replaced by a probabilistic circuit of size $2^{O(\log^{4(k-1)r} n)}$, having error probability at most $\frac{1}{2^{3cd\cdot\log^r n}}$, and using $O(\log^{3r} n)$ probabilistic bits. If we use *one* sequence of $O(\log^{3r} n)$ probabilistic bits and use this sequence as the probabilistic input to *each* of these subcircuits, the resulting circuit has error probability at most $\frac{2^{d\cdot\log^r n}}{2^{3cd\cdot\log^r n}} \leq \frac{1}{2^{2c\cdot\log^r n}}$. Also, by Lemma 13, the top level AND in this circuit (with fan-in $2^{d\cdot\log^r n}$) can be computed by a formula of the form $\sum_{(2^{O(\log^{4r} n)})}' \prod_{(2^{O(\log^{4r} n)})}$ with error probability at most $\frac{1}{2^{2c\cdot\log^r n}}$, using at most $O(\log^{3r} n)$ probabilistic bits.

Putting these two parts together results in a circuit with error probability at most $\frac{1}{2^{c\cdot\log^r n}}$, of the form

$$\sum_{(2^{O(\log^{4r} n)})} \prod_{(O(\log^{4r} n))} \sum_{(2^{O(\log^{4(k-1)r} n)})} \prod_{(O(\log^{4(k-1)r} n))} (\ldots),$$

which can be rearranged to the form

$$\sum_{(2^{O(\log^{4kr} n)})} \prod_{(O(\log^{4kr} n))} (\ldots).$$

To complete the proof, one can show via an easy induction on k that the circuits produced by this construction are uniform. ∎

6 Depth Reduction for Uniform Deterministic Circuits

In this section we want to use the results of Section 5 to obtain depth reductions for deterministic circuits. Thus we first have to investigate how probabilistic circuits can be made deterministic using a constant number of levels of AND and OR gates. Our lemma accomplishing this is a circuit-based interpretation of the inclusion BPP $\subseteq \Sigma_2^p$ (Sipser (1983), Lautemann (1983)), and our proof proceeds along exactly those same lines.

Lemma 15 Let $\{C_n\}$ be a uniform family of probabilistic circuits accepting a language L, of size $s(n)$, depth $d(n)$, using $m(n)$ probabilistic bits, and having error probability less than $\frac{1}{m(n)}$. Then there is a uniform family of deterministic circuits accepting L having depth $d(n)+3$ and size $2^{m(n)}s(n)+2^{m(n)^2+1}+1$. The top level of this circuit consists of an AND gate of fan-in $2^{m(n)^2}$, whose inputs are OR gates of fan-in $2^{m(n)}$, the inputs to the OR gates are AND gates of fan-in $m(n)$, and the inputs to these AND gates come from $2^{m(n)}$ copies of the original circuit C_n, each with a different probabilistic sequence hardwired in.

Proof: Let x be any string of length n, let y be a sequence of $m = m(n)$ bits, and let c_y denote the output (zero or one) produced by C_n on input x with probabilistic sequence y. Thus

$$\begin{aligned} x \in L \quad &\text{implies} \quad |\{y : c_y = 0\}| < \tfrac{2^m}{m}, \\ \text{and} \quad x \notin L \quad &\text{implies} \quad |\{y : c_y = 1\}| < \tfrac{2^m}{m}. \end{aligned}$$

Let the letter S denote subsets of the set $\{0, \ldots, 2^m - 1\}$. We claim that

$$x \in L \iff \forall_{S:\ |S|=m}\ \exists_{z \in \{0,\ldots,2^m-1\}}\ \forall_{a \in S}\ \ c_{y(a,z)} = 1 \qquad (1)$$

where $y(a,z) = a + z \bmod 2^m$. Once this claim is proved, it is easy to see that the formula on the right hand side of (1) can be realized by a circuit with the desired requirements, and that the circuits can be constructed uniformly.

To prove equivalence (1) we first assume that there is an S such that for all z there is an a in S with $c_{y(a,z)} = 0$. This means that there are at least 2^m zeros among the $c_{y(a,z)}$'s (one for each z), but each zero can occur at most m times (once for each $a \in S$). Thus we in fact have at least $\frac{2^m}{m}$ zeros among the c_y's, which means that x is not in L.

For the converse, assume that for all S there exists a z_S such that for all a in S we have $c_{y(a,z_S)} = 1$. We will show that x is in L, by showing that the size of the set $G := \{y : c_y = 1\}$ is greater than $\frac{2^m}{m}$. Let $N = |G|$. Let $S+$ denote the set $\{y(a, z_S) : a \in S\}$; thus every set $S+$ is an m-element subset of G. Furthermore, each m-element set can occur at most 2^m times as $S+$ (once for each z); that is, $\binom{2^m}{m} \leq 2^m \binom{N}{m}$.

The argument is concluded by noting that for all $m \geq 2$,

$$2^m \binom{N}{m} \geq \binom{2^m}{m} \geq \tfrac{2^m}{m^m} \cdot \binom{2^m}{m} > \tfrac{2^m}{m^m m!} \cdot \left(\textstyle\prod_{i=0}^{m-1}(2^m - im)\right) = 2^m \cdot \binom{\frac{2^m}{m}}{m}$$

■

We use this result to obtain a depth 4 simulation of constant depth $\{\wedge,\vee,\oplus\}$-circuits:

Theorem 16 Let p be a prime number, and let $k, r \geq 1$. Then
Uniform SIZE($2^{O(\log^r n)}$)DEPTH(k) GATES($\wedge$,$\vee$,MODp)
$\subseteq$ Uniform SIZE($2^{O(\log^{4kr+3} n)}$)DEPTH(4) GATES($\wedge$,$\vee$,MODp)

Proof: Applying Lemma 15 to the depth two probabilistic simulation of Lemma 14, one obtains a depth 5 circuit, whose lower three levels are AND, MODp and again AND levels. As the third level AND gates have fan-in equal to the number of probabilistic bits in the probabilistic simulation, one can apply Lemma 3 to compress these three levels to two, achieving the desired size bound. ∎

Just as Theorem 16 is a uniform analogue of Theorem 7, the following theorem provides a result similar to Theorem 9 in the uniform setting.

Theorem 17 Let p be a prime number, and let $k, r \geq 1$. Then
Uniform SIZE($2^{O(\log^r n)}$)DEPTH(k) GATES($\wedge$,$\vee$,MODp)
$\subseteq$ Uniform SIZE($2^{O(\log^{4kr} n)}$)DEPTH(3) GATES(MAJORITY)

Proof: By Lemma 14 it suffices to simulate depth two probabilistic circuits of size $2^{O(\log^{4kr} n)}$ with $O(\log^{3r} n)$ probabilistic bits. This can be done by taking the majority over all sequences of random bits (i.e., a MAJORITY gate with inputs from $2^{O(\log^{3r} n)}$ copies of the depth two circuit – one copy for each probabilistic sequence). The resulting circuit accepts the correct language, and can be converted to a threshold circuit using Proposition 8. (It is easily observed that the conversion presented in the proof of Proposition 8 can be done uniformly.) ∎

Our main depth reduction results for uniform deterministic circuits are summarized in the next corollary, which follows immediately from Theorems 16 and 17:

Corollary 18 Let p be any prime number. Then

a) Uniform SIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\wedge$,$\vee$,MODp)
= Uniform SIZE($2^{\log^{O(1)} n}$)DEPTH(4) GATES($\wedge$,$\vee$,MODp)
$\subseteq$ Uniform SIZE($2^{\log^{O(1)} n}$)DEPTH(3) GATES(MAJORITY)

b) Uniform $AC^0_k \subseteq$ Uniform SIZE($2^{O(\log^{4k} n)}$)DEPTH(3) GATES(MAJORITY)
Uniform $AC^0_k \subseteq$ Uniform SIZE($2^{O(\log^{4k} n)}$)DEPTH(4) GATES($\wedge$,$\vee$,MODp)

7 Depth Reduction for Uniform Probabilistic Circuits

In this section we want to use the results of the previous sections to obtain depth reductions for probabilistic circuits. The first corollary follows from Lemma 14:

Corollary 19 Let p be any prime number. Then
Uniform BPSIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\wedge$,$\vee$,MODp)
= Uniform BPSIZE($2^{\log^{O(1)} n}$)DEPTH(2) GATES($\wedge$,$\vee$,MODp)

Proof: Starting with a probabilistic circuit of error less than $\frac{1}{4}$, use the technique of Ajtai and Ben-Or (1984) (which we also used in Theorem 7) to reduce the error probability to $\frac{1}{n}$, increasing the size by a polynomial factor, and adding only a constant number of levels to the depth. (Note that this transformation can be carried out uniformly.) Now we simulate that circuit (viewed as a deterministic circuit on the original inputs plus the random bits) via Lemma 14 by a depth two circuit introducing negligable additional error probability. Thus the error probability of the new circuit (now again viewed as a probabilistic circuit on the original inputs) is less than $\frac{1}{4}$. ∎

As we saw in the proof of Theorem 16, it is very easy to change probabilistic circuits into deterministic circuits in a uniform way if only $\log^{O(1)} n$ probabilistic bits are used. Fortunately, in many cases, a probabilistic circuit can be converted into an equivalent one using only a small number of probabilistic bits. The technique for doing this was developed by Nisan and Wigderson. The following paragraphs outline some of the results of Nisan and Wigderson (1988).

Nisan and Wigderson showed how to construct a certain type of pseudorandom generator. They have a very general construction that takes as its starting point a "hard" function. For example, it easily follows from Chapter 8 of Håstad (1987) that:

Theorem 20 Håstad (1987) $\forall d\ \exists k\ \forall$ polynomials p, and for all large n,
for any depth d circuit C of $n^{O(1)}$ AND and OR gates, taking inputs of size $\log^k n$

$$|Prob[C(x) = \text{PARITY}(x)] - \tfrac{1}{2}| \leq \tfrac{1}{p(n)}$$

where all strings of length $\log^k n$ are equally probable.

Using this "hardness" result for PARITY, Nisan and Wigderson then construct a pseudorandom generator that takes input of length $\log^k n$ and produces output that "looks random" to any circuit family in (nonuniform) AC^0_d. This pseudorandom generator is itself computable in (uniform) AC^0 (albeit with depth greater than d). Rephrasing the results of Nisan and Wigderson (1988) slightly, we obtain the following theorem:

Theorem 21 (Nisan and Wigderson (1988, Theorem 5))
Every set in Uniform BPSIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\{\wedge, \vee\}$) is accepted by a uniform circuit of type BPSIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\{\wedge, \vee\}$) that has only $\log^{O(1)} n$ probabilistic bits.

This allows us to obtain the following theorem:

Theorem 22 For any prime number p,
Uniform BPSIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\{\wedge, \vee\}$)
$\subseteq$ Uniform SIZE($2^{\log^{O(1)} n}$)DEPTH(4) GATES($\{\wedge, \vee, \text{MOD}p\}$)

Proof: Let L be in Uniform BPSIZE($2^{\log^{O(1)} n}$)DEPTH($O(1)$) GATES($\{\wedge, \vee\}$). By Theorem 21, we can assume that L is accepted by a uniform circuit family $\{C_n\}$ using only $\log^{O(1)} n$ probabilistic bits, and with error at most $\frac{1}{8}$. As in the proof of Corollary 19, note that the circuits themselves are deterministic, if we view the probabilistic bits as being part of the input; to be precise let this family be $\{D_m\}$, where for some constant c, $C_n = D_{n+\log^c n}$. By Lemma 14, the circuit family $\{D_m\}$ is simulated by a uniform family of probabilistic depth two circuits $\{E_m\}$ using $\log^d n$ probabilistic bits, for some constant d, with gates from $\{\wedge, \vee, \text{MOD}p\}$ and with error at most $\frac{1}{8}$. That is, L itself is accepted by a uniform family of depth two circuits using only $\log^c n + \log^d n$ probabilistic bits and gates from $\{\wedge, \vee, \text{MOD}p\}$, with error at most $\frac{1}{4}$.

By lemma 15, L is accepted by a uniform family of depth five circuits, and these can be converted into depth four circuits as in the proof of Theorem 16. ∎

Unfortunately, the pseudorandom generators constructed in Nisan and Wigderson (1988) do not allow us to prove anything about circuits that have PARITY gates. On the other hand, the technique of Nisan and Wigderson (1988) is quite general, and if we had an example of a suitable "hard" function, the proof strategy of Theorem 22 would carry over to the setting of circuits with PARITY gates. The following paragraphs make this precise.

Definition: Let us say that the function f is *suitably hard* if

- $f : \{0,1\}^m \to \{0,1\}$ is computable in space $m^{O(1)}$, and
- $\forall d\ \exists k\ \forall$ polynomials p, and for all large n,
  for any depth d circuit C of $n^{O(1)}$ AND, OR and PARITY gates, taking inputs of size $\log^k n$

$$|Prob[C(x) = f(x)] - \tfrac{1}{2}| \leq \tfrac{1}{p(n)}$$

  where all strings of length $\log^k n$ are equally probable.

It is reasonable to conjecture that suitably hard functions exist. In fact, the results of Razborov (1987) and Smolensky (1987) make it plausible that MAJORITY and MOD3 are suitably hard. Unfortunately, we do not see how to adapt the proof techniques of Razborov (1987) and Smolensky (1987) to show that there are any suitably hard functions.

Theorem 23 If suitably hard functions exist, then

$$\begin{aligned}
&\text{Uniform BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(O(1))\,\text{GATES}(\{\wedge,\vee,\oplus\}) \\
&= \text{Uniform BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(2)\,\text{GATES}(\{\wedge,\vee,\oplus\}) \\
&= \text{Uniform SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(4)\,\text{GATES}(\{\wedge,\vee,\oplus\}) \\
&\subseteq \text{Uniform SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(3)\,\text{GATES}(\{\text{MAJORITY}\})
\end{aligned}$$

8 Summary

Our main depth reduction theorems are:
In the nonuniform case:

$$\begin{aligned}
&\text{BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(O(1))\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\
= \;&\text{BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(2)\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\
= \;&\text{SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(4)\,\text{GATES}(\{\wedge,\vee,\text{MOD}p\}) \\
\subseteq \;&\text{SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(3)\,\text{GATES}(\{\text{MAJORITY}\})
\end{aligned}$$

and in the uniform case:

$$\begin{aligned}
&\text{Uniform SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(O(1))\,\text{GATES}(\wedge,\vee,\text{MOD}p) \\
= \;&\text{Uniform SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(4)\,\text{GATES}(\wedge,\vee,\text{MOD}p) \\
\subseteq \;&\text{Uniform SIZE}(2^{\log^{O(1)} n})\text{DEPTH}(3)\,\text{GATES}(\text{MAJORITY})
\end{aligned}$$

$$\begin{aligned}
&\text{Uniform BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(O(1))\,\text{GATES}(\wedge,\vee,\text{MOD}p) \\
= \;&\text{Uniform BPSIZE}(2^{\log^{O(1)} n})\text{DEPTH}(2)\,\text{GATES}(\wedge,\vee,\text{MOD}p)
\end{aligned}$$

We conjecture that for prime p, uniform probabilistic $\{\wedge,\vee,\mathrm{MOD}p\}$-circuits of constant depth can be simulated by uniform deterministic circuits of depth four using the same type of gates, but we were able to prove this only under unproven assumptions (Theorem 23).

In addition, we noted that these results are essentially optimal, in the sense that such a simulation in any constant depth cannot be achieved in size less than $2^{\log^{O(1)} n}$, even if we try only to simulate AC^0-circuits (Proposition 11). It remains open whether or not AC^0-circuits can be simulated by threshold circuits of fixed depth and polynomial size. A possible first step toward settling this question appears in Bruck and Smolensky (1990).

Recently, a number of other results have been proved concerning small depth threshold circuits of size $2^{\log^{O(1)} n}$. (Barrington has suggested calling this complexity class "Quasi-TC^0.") Yao (1990) improved our Theorems 9 and 17, removing the restriction that the modulus p be prime. We note, however, that it is still not known if depth reduction theorems such as Corollaries 10, 18, and 19 can be generalized to composite moduli.

Tarui (1991) has shown that AC^0 can be simulated by probabilistic depth two threshold circuits of size $2^{\log^{O(1)} n}$ with one-sided error. Related results may also be found in Beigel et al. (1990).

These developments in circuit complexity go hand-in-hand with significant progress being made in understanding the relationships that exist among various subclasses of PSPACE, such as the polynomial hierarchy, PP, and the counting hierarchy, starting with the seminal result of Toda (1991). These connections are surveyed in Allender and Wagner (1990), and similar connections are explored in Kannan et al. (1991).

Amid all of this recent work showing the surprising power of Quasi-TC^0 circuits, it has been suggested that even apparently larger complexity classes such as NC^1 might be contained in Quasi-TC^0. Note that, by standard translational techniques, this would imply that $\mathrm{DSPACE}(\log^{O(1)} n) \subseteq$ Quasi-TC^0, since for any set $L \in \mathrm{DSPACE}(\log^k n)$, the set $\{x10^{2^{\log^{2k}|x|}} : x \in L\}$ is in NC^1.

Acknowledgments

The first author acknowledges discussions with Ravi Boppana, Seinosuke Toda, David Barrington, Denis Thérien and Roman Smolensky; thanks are also due to

Pierre McKenzie and Denis Thérien for organizing the 1990 Barbados Workshop on Complexity Theory, where some of these discussions took place. We thank Ileana Streinu and Walter Hohberg for finding some errors in an earlier version, and we thank one of the anonymous referees for giving the paper a very thorough and careful reading. Finally, we thank Klaus Wagner for making our collaboration possible.

References

Adleman, L. (1978), Two theorems on random polynomial time, *in* "Proceedings, 19th IEEE Symposium on Foundations of Computer Science," pp. 75–83.

Ajtai, M. and Ben-Or, M. (1984), A theorem on probabilistic constant depth computations, *in* "Proceedings, 16th ACM Symposium on Theory of Computing," pp. 471–474.

Allender, E. (1989), A note on the power of threshold circuits, *in* "Proceedings, 30th IEEE Symposium on Foundations of Computer Science," pp. 580–584.

Allender, E. (1989a), P-uniform circuit complexity, J. ACM **36**, 912–928.

Allender, E. and Hertrampf, U. (1990), On the power of uniform families of constant depth threshold circuits, *in* "Proceedings, 15th International Symposium on Mathematical Foundations of Computer Science," 1990, Lecture Notes in Computer Science 452, pp. 158–164, Springer-Verlag, Berlin..

Allender, E. and Wagner, K. (1990), Counting hierarchies: polynomial time and constant depth circuits, *in* The Structural Complexity Column, (J. Hartmanis, Ed.), EATCS Bulletin 40.

Barrington, D. (1987), A note on the theorem of Razborov, unpublished manuscript.

Barrington, D. A. M., Immerman, N., and Straubing, H. (1990), On uniformity within NC^1, Journal of Computer and System Sciences **41**, 274–306.

Beigel, R., Reingold, N., and Spielman, D. (1990), The perceptron strikes back, *in* "Proceedings, 6th IEEE Structure in Complexity Theory Conference," pp. 286–291.

Boppana, R., and Sipser, M. (1990), The complexity of finite functions, *in* "Handbook of Theoretical Computer Science, vol. A: Algorithms and Complexity," (J. van Leeuwen, Ed.), pp. 757–804, MIT Press/Elsevier, Cambridge,MA/Amsterdam.

Bruck, J. (1990), Harmonic analysis of polynomial threshold functions, SIAM J. Disc. Math. **1**, 168–177.

Bruck, J., and Smolensky, R. (1990), Polynomial threshold functions, AC^0 functions, and spectral norms, *in* "Proceedings, 31st IEEE Symposium on Foundations of Computer Science," pp. 632–641.

Bultman, W. (1990), On the design of threshold circuits for some basic functions, manuscript, University of Illinois at Chicago.

Chandra, A., Stockmeyer, L., and Vishkin, U. (1984), Constant depth reducibility, SIAM J. Comput. **13**, 423–439.

Denenberg, L., Gurevich, Y., and Shelah, S. (1986), Definability by constant-depth polynomial-size circuits, Information and Control **70**, 216–240.

Fagin, R., Klawe, M., Pippenger, N., and Stockmeyer, L. (1985), Bounded-depth, polynomial-size circuits for symmetric functions, Theoretical Computer Science **36**, 239–250.

Furst, M., Saxe, J., and Sipser, M. (1984), Parity, circuits, and the polynomial-time hierarchy, Mathematical Systems Theory **17**, 13–27.

Hajnal, A., Maass, W., Pudlák, P., Szegedy, M., and Turán, G. (1987), Threshold circuits of bounded depth, to appear in Journal of Computer and System Sciences. *Preliminary version in* "Proceedings, 28th IEEE Symposium on Foundations of Computer Science," pp. 99–110.

Håstad, J. (1987), "Computational limitations for small-depth circuits," Doctoral Dissertation, MIT.

HÅSTAD, J., AND GOLDMANN, M. (1990), On the power of small-depth threshold circuits, *in* "Proceedings, 31st IEEE Symposium on Foundations of Computer Science," pp. 610–618.

KANNAN, R., VENKATESWARAN, H., VINAY, V., AND YAO, A. (1991), A circuit-based proof of Toda's theorem, to appear in Information and Computation.

LAUTEMANN, C. (1983), BPP and the polynomial hierarchy, Information Processing Letters **17**, 215–217.

NISAN, N., AND WIGDERSON, A. (1988), Hardness vs. randomness, *in* "Proceedings, 29th IEEE Symposium on Foundations of Computer Science," pp. 2–11.

PARBERRY, I. (1990), A primer on the complexity theory of neural networks, *in* "Formal Techniques in Artificial Intelligence: A Sourcebook" (R. Banerji, Ed.), Studies in Computer Science and Artificial Intelligence 6, pp. 217–268, North-Holland, Amsterdam.

PARBERRY, I., AND SCHNITGER, G. (1988), Parallel computation with threshold functions, J. Computer and System Science **36**, 278–302.

PARBERRY, I., AND SCHNITGER, G. (1989), Relating Boltzmann machines to conventional models of computation, Neural Networks **2**, 59–67.

RAZBOROV, A. (1987), Lower bounds on the size of bounded depth networks over a complete basis with logical addition, Mathematicheskie Zametki **41** (4), 598–607. English translation in Mathematical Notes of the Academy of Sciences of the USSR **41**:4, 333–338.

RUZZO, W. (1981), On Uniform Circuit Complexity, J. Comput. and System Sci. **21**, 365–383.

SIPSER, M. (1983), A complexity theoretic approach to randomness, *in* "Proceedings, 15th Annual ACM Symposium on Theory of Computing," pp. 330–335.

Siu, K., and Bruck, J. (1991), On the power of threshold circuits with small weights, SIAM J. Discrete Math. **4**, 423–435.

Smolensky, R. (1987), Algebraic methods in the theory of lower bounds for Boolean circuit complexity, *in* "Proceedings, 19th ACM Symposium on Theory of Computing," pp. 77–82.

Tarui, J. (1991), Randomized polynomials, threshold circuits, and the polynomial hierarchy, *in* "Proceedings, 8th Symposium on Theoretical Aspects of Computer Science," Lecture Notes in Computer Science 480, pp. 238–250, Springer-Verlag, Berlin..

Toda, S. (1991), PP is as hard as the polynomial-time hierarchy, SIAM J. Comput. **20**, 865–877.

Turán, J. (1991), Complexity classes defined by counting quantifiers, J. ACM **38**, 753–774.

Valiant, L., and Vazirani, V. (1986), NP is as easy as detecting unique solutions, Theoretical Computer Science **47** 85–93.

Yao, A. (1985), Separating the polynomial-time hierarchy by oracles, *in* "Proceedings, 26th IEEE Symposium on Foundations of Computer Science," pp. 1–10.

Yao, A. (1989), Circuits and local computation, *in* "Proceedings, 21st ACM Symposium on Theory of Computing," pp. 186–196.

Yao, A. (1990), On ACC and threshold circuits, *in* "Proceedings, 31st IEEE Symposium on Foundations of Computer Science," pp. 619–627.