comput. complex. **9** (2000), 1–15
1016-3328/00/010001-15 $ 1.50+0.20/0

computational complexity

EXPONENTIAL LOWER BOUNDS FOR DEPTH THREE BOOLEAN CIRCUITS

Ramamohan Paturi, Michael E. Saks,
and Francis Zane

Abstract. We consider the class Σ_3^k of unbounded fan-in depth three Boolean circuits, for which the bottom fan-in is limited by k and the top gate is an OR. It is known that the smallest such circuit computing the parity function has $\Omega(2^{\varepsilon n/k})$ gates (for $k = O(n^{1/2})$) for some $\varepsilon > 0$, and this was the best lower bound known for explicit (P-time computable) functions. In this paper, for $k = 2$, we exhibit functions in uniform NC^1 that require $2^{n-o(n)}$ size depth 3 circuits. The main tool is a theorem that shows that any Σ_3^2 circuit on n variables that accepts a inputs and has size s must be constant on a projection (subset defined by equations of the form $x_i = 0$, $x_i = 1$, $x_i = x_j$ or $x_i = \bar{x}_j$) of dimension at least $\log(a/s)/\log n$.

Key words. Circuit complexity, nonlinear lower bounds, constant depth circuits.

Subject classifications. 68Q99.

1. Introduction

Considerable progress has been made in understanding the limitations of unbounded fan-in Boolean circuits of bounded depth. The results of Ajtai (1983), Furst *et al.* (1981), Yao (1985), Håstad (1986), Razborov (1986), Smolensky (1987), among others, show that if the size of the circuit is not too large, then any function computed by such a circuit must be constant on a large subcube or can be approximated by a small degree polynomial. Such limitations of small size bounded depth circuits can be used to show that certain explicit functions such as parity and majority require a large number of gates. More precisely, a result of Håstad (1986) says that computing the parity function in depth d requires $\Omega(2^{\varepsilon n^{1/(d-1)}})$ gates for some $\varepsilon < 1$. Except for the constant ε this result is essentially tight.

Recently, Håstad *et al.* (1993) described a top down approach for proving lower bounds on depth 3 circuits. However, these and other techniques seem incapable of proving a lower bound on depth 3 circuits of the form $\Omega(2^{h(n)\sqrt{n}})$ with $h(n)$ unbounded, for any explicit Boolean function. Here, as usual, the term "explicit function" is a somewhat informal term, which is taken to mean "uniformly and efficiently computable", in, say P or NC.

To clarify the situation, it is useful to parameterize the lower bound in terms of the maximum fan-in of the bottom gates. Define Σ_d^k to be the set of depth d circuits with top gate OR such that each bottom gate has fan-in at most k. Then it follows from known results that there is a constant $\varepsilon \leq 1$ such that for any $k \geq 1$, any Σ_3^k circuit for the parity function or the majority function requires $\Omega(2^{\varepsilon n/k})$ gates at level 2, and such bounds are tight for $k = O(\sqrt{n})$.

As in Håstad *et al.* (1993), our motivation is to prove stronger lower bounds on depth 3 circuits that go beyond the above trade-off between bottom fan-in and size. We note that even for constant bottom fan-in $k \geq 2$, currently known lower bound techniques seem incapable of providing a lower bound better than $2^{n/k}$ on the number of gates at level 2. There is another independent compelling motivation for studying the depth 3 model with limited fan-in. Valiant (1977) showed that linear-size logarithmic-depth Boolean circuits with bounded fan-in can be computed by depth 3 unbounded fan-in circuits of size $O(2^{n/\log\log n})$ and bottom fan-in limited by n^{ε} for arbitrarily small ε. Also, if we consider linear-size logarithmic-depth circuits with the additional restriction that the graph of the connections is series-parallel, then such circuits can be computed by depth 3 unbounded fan-in circuits of size $2^{n/2}$ with bounded bottom fan-in. Thus, strong exponential lower bounds on depth 3 circuits would imply non-linear lower bounds on size of fan-in 2 Boolean circuits with logarithmic depth, an open problem proposed some twenty years ago in Valiant (1977).

In this paper, we take a modest step towards proving such strong bounds on depth 3 circuits. We show that for some explicit function, contained in logspace uniform NC^1, any Σ_3^2 circuit that computes it must have at least $2^{n-o(n)}$ gates. We obtain this result by showing that the function computed by a small Σ_3^2 circuit must be constant on a large "nicely structured" subset of the cube. These subsets, called projections, are defined by equating literals to each other or to constants.

The starting point for our argument is the top-down approach used in Håstad *et al.* (1993), which says that if the number of gates at level 2 of a Σ_3 circuit is small, there must be a depth 2 subcircuit that accepts a large number of inputs. We prove that such a depth 2 subcircuit (which in our case is a 2-CNF formula) must accept a projection of large size. We then give two con-

structions of functions such that any Σ_3^2 or Π_3^2 circuit computing them requires $2^{n-o(n)}$ size. For the first construction, we show that the set of codewords of an error-correcting code is not identically one on any large projection. Thus, any Σ_3^2 circuit accepting this set requires large size. It then follows that the $n+1$-variable function $g(x_1,\ldots,x_n,x_{n+1}) = x_{n+1}f(x_1,\ldots,x_n) + \bar{x}_{n+1}\bar{f}(x_1,\ldots,x_n)$ requires large size Σ_3^2 and Π_3^2 circuits. In the second construction, we construct a function which has a subfunction with the stronger property of not being *constant* on any large projection. To do so, we first show that, with high probability, a randomly chosen homogeneous multilinear n-variable polynomial of degree 2 over $GF(2)$ is nonconstant on every large projection. We then use derandomization techniques to construct a specific Boolean function with the property that it has a subfunction on a large enough set of variables which is not constant on any large projection. This property is stronger than what we needed to prove lower bounds on depth 3 fan-in 2 circuits, and may be useful in other settings.

The rest of the paper is organized as follows: In section 2, we review some basic definitions and results, including a proof that any symmetric function can be computed by a Σ_3^2 circuit of size at most $poly(n)2^{0.59n}$. In section 3, we show that any 2-CNF which accepts a large number of inputs must necessarily accept a projection with large dimension. Using this result, in sections 4 and 5 we construct functions which do not have depth 3 bottom fan-in 2 circuits of size less than $2^{n-o(n)}$.

2. Preliminaries

2.1. Boolean variables, literals and assignments. Let X denote the set $\{x_1, x_2, \ldots, x_n\}$ of variables and L denote the set $\{x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n\}$ of literals. If V is a subset of L, then $\overline{V}$ denotes the set $\{\bar{v} \mid v \in V\}$. An *assignment* of X is a function $\alpha : X \to \{0,1\}$, and a *partial assignment* is a function α from a subset of X to $\{0,1\}$. Associated to any partial assignment α is the subset $X(\alpha) \subseteq X$ of variables set to 1 by α and the set $L(\alpha) \subseteq L$ of literals set to 1 by that assignment.

2.2. 2-CNF formulae. We briefly review some basic facts about 2-CNF formulae. A 2-CNF formula Φ on a variable set X can be associated naturally with its implication digraph $D(\Phi)$, whose vertex set is L. Each clause $v \vee w$ (where v and w are literals) gives rise to two edges $\bar{v} \to w$ and $\bar{w} \to v$. Each singleton clause v gives rise to the edge $\bar{v} \to v$. Note that the map that exchanges each pair of complementary literals and reverses the direction of all edges is an isomorphism of $D(\Phi)$.

We say that literal v *implies* literal w if there is a directed path in $D(\Phi)$ from v to w. The *implies* relation is clearly transitive. The digraph $D(\Phi)$ defines a partition of L into *strong components*, i.e., maximal subsets V with the property that for any two vertices v and w in V, v implies w and w implies v. Note that $V \subseteq L$ is a strong component if and only if $\overline{V}$ is a strong component. A subset V of literals is said to be *initial* in $D(\Phi)$ if there is no edge entering V from outside V, and is said to be *final* if there is no edge from a vertex in V to a vertex outside V. Trivially each initial set and each final set is a union of strong components. If Φ is satisfiable, we say that the literal v is *fixed* by Φ if the value of v is the same for every satisfying assignment of Φ.

We state without proof the following facts, which are easy to prove and belong to the folklore about 2-CNF formulae.

PROPOSITION 2.1. *Let Φ be a 2-CNF formula on $\{x_1, \ldots, x_n\}$. Then:*

1. *An assignment α satisfies Φ if and only if $L(\alpha)$ is a final set in $D(\Phi)$.*
2. *If the relation "v implies w" holds in $D(\Phi)$ then in any satisfying assignment of Φ, $v = 0$ or $w = 1$.*
3. *Φ is satisfiable if and only if for each variable x_i, the literals x_i and $\bar{x}_i$ lie in different strong components.*
4. *If V is a strong component of D then in any satisfying assignment of Φ, either all literals in V are true or all literals in V are false.*
5. *If V is a strong component of D, and Φ is satisfiable, then one of the following two situations holds: either V consists entirely of fixed literals or there exist two satisfying assignments of Φ that differ precisely on the variables of V.*

A strong component consisting entirely of fixed literals is a *fixed component*; otherwise it is an *unfixed component*.

2.3. Circuits. As usual, for an integer d, Σ_d (resp. Π_d) denotes the class of layered unbounded fan-in Boolean circuits with d alternating levels of ANDs and ORs, and a single OR gate (resp. AND gate) at the top. The inputs are viewed as feeding into the first level, and the top gate is at the d-th level. Similar to Håstad *et al.* (1993), we define Σ_d^k (resp. Π_d^k) to be the class of circuits in Σ_d (resp. Π_d) such that all gates at the first level have fan-in at most k. For a Boolean function f, we define $s_d^k(f)$ to be the size (number of gates) of the

smallest Σ_d^k circuit computing f (here we assume $d \geq 3$, so that $s_d^k(f)$ is well defined). We are interested in computing lower bounds on $s_3^k(f)$ for explicit functions f, and will obtain such bounds for the case $k = 2$.

If C is a Σ_3 circuit having M AND gates at level 2, we write $C^1, C^2, \ldots, C^M$ for the Π_2 subcircuits at level 2. Each of the circuits C^i is equivalent to a CNF formula of the inputs. If C is a Σ_3^k circuit, then each of the C^i computes a k-CNF formula. If f is the function computed by C, then f is the OR of the functions $f_1, \ldots, f_M$ computed by the circuits $C^1, \ldots, C^M$. Let $\kappa(f)$ be the minimum number M such that f can be written as an OR of M 2-CNF functions. Trivially, $s_3^2(f) \geq \kappa(f)$ and since any 2-CNF on n variables can be expressed as a Π_2^2 circuit with at most $4n^2$ gates we have:

PROPOSITION 2.2. *Let f be a Boolean function on n variables. Then*

$$\kappa(f) \leq s_3^2(f) \leq \kappa(f)4n^2.$$

So to approximate $s_3^2(f)$ it suffices to analyze $\kappa(f)$. It is useful to think of the determination of $\kappa(f)$ as a cover problem: we want to cover the subset $A = f^{-1}(1)$ of $\{0,1\}^n$ by subsets of A each of which can be expressed as the accepting set of a 2-CNF.

As an example, consider $s_3^2(f)$ for symmetric Boolean functions. Consider first the slice functions: S_k^n is the n-variable function that is one on inputs of weight k (where the weight is the number of 1's in the input). It is easy to see that $\kappa(S_k^n) = \kappa(S_{n-k}^n)$ (given a circuit for S_k^n, replace all literals by their complements to get one for S_{n-k}^n), so assume $k \leq n/2$.

We want to cover the set of assignments of weight k by 2-CNFs. We can only use 2-CNFs whose accepting set consists of inputs of weight k.

To get an upper bound, consider the set $\mathcal{G}$ of Boolean formulas that can be constructed in the following way. Partition the variables arbitrarily into $k+1$ sets $V, P_1, \ldots, P_k$ where each of the P_i is of size 2 and V is of size $n-2k$. Define the formula Φ having clauses $\bar{x}_i$ for $x_i \in V$ and clauses $x_i \vee x_j$ and $\bar{x}_i \vee \bar{x}_j$ for each $P_r = \{x_i, x_j\}$. Then the assignment α satisfies Φ if and only if α is 0 on all variables in V and for each P_i, one variable is set to 1 and the other is set to 0. Hence each formula in $\mathcal{G}$ accepts only inputs of weight k.

We claim that there exists a set of such formulae having size $M \leq n2^{0.59n}$ that cover all assignments of weight k. Let α be an assignment of weight k. If Φ is a formula chosen uniformly at random from $\mathcal{G}$ then the probability that Φ covers α, i.e., that α satisfies Φ, is the probability that the k variables set to 1 by α belong to k distinct pairs P_i, which is easily shown to be $2^k/\binom{n}{k}$. Therefore, if we choose $\Phi_1, \ldots, \Phi_M$ independently and uniformly from $\mathcal{G}$, the

probability that none of them cover α is $(1-2^k/\binom{n}{k})^M \leq e^{-M2^k/\binom{n}{k}}$. Since there are $\binom{n}{k}$ such assignments, the probability that there is an assignment that is uncovered is at most

$$\binom{n}{k} e^{-M2^k/\binom{n}{k}}.$$

Thus, if M is at least $m(k) = 2^{-k}\binom{n}{k}\ln\binom{n}{k}$, then this probability is less than one, and some choice of $\Phi_1, \ldots, \Phi_M$ is a cover. $m(k)$ is maximized when $k = \frac{n}{3} - c$ for some constant c, and is at most $2^{0.59n}$.

Now any symmetric Boolean function is the OR of at most n slice functions and so if f is a symmetric Boolean function then $\kappa(f) \leq n^2 2^{0.59n}$ and $s_3^2(f) \leq 2^{0.59n+O(\log n)}$.

Our goal in this paper is to exhibit concrete functions which require circuits of much larger size S, that is, circuits of size S such that $(\log_2 S)/n$ approaches 1.

3. Projections

In this section, we prove that if a 2-CNF formula accepts many inputs, then it must accept a projection of large dimension.

A *projection* for a variable set X is a subset of the set of all assignments (or, equivalently, a subset of $\{0,1\}^n$), defined by equations of the form $v_i = 0$, $v_i = 1$, or $v_i = v_j$ where v_i and v_j are literals. Trivially, the condition $v_i = 0$ is equivalent to $\bar{v}_i = 1$ and the condition $v_i = v_j$ is equivalent to $\bar{v}_i = \bar{v}_j$. A projection is an affine subspace of $GF(2)^n$, and the *dimension* of a projection is its dimension as an affine subspace. A projection of dimension d can be specified by $2(d+1)$ sets $(A_0, B_0, A_1, B_1, A_2, B_2, \ldots, A_d, B_d)$ where $A_i \cup B_i$ are disjoint for $i \geq 0$, $\bigcup_{i\geq 0}(A_i \cup B_i) = X$, and $A_i \cup B_i$ are nonempty for $i \geq 1$. The projection P specified by such a sequence of sets consists of all assignments α which are 0 on the variables of A_0, 1 on the variables of B_0, and such that for each $j \geq 1$, all the variables in A_j are equal and all the variables in B_j are equal to the negation of the variables in A_j. When we say that a projection defines a partition, the partition defined is a partition of the variables not set to constants into the sets $A_i \cup B_i$ for $1 \leq i \leq d$. These sets are referred to as the *parts* of the partition. For a subset S of assignments, we define $\pi(S)$ to be the dimension of the largest projection P such that $P \subseteq S$. If f is a Boolean function, we write $\pi(f)$ for $\pi(f^{-1}(1))$.

The following result gives a lower bound on the number of gates at level 2, $\kappa(f)$ (and hence on the circuit size $s_3^2(f)$) in terms of $\pi(f)$:

THEOREM 3.1. *Let f be a Boolean function on n variables and suppose that $\pi(f) \leq d$. Then*

$$s_3^2(f) \geq \kappa(f) \geq \frac{|f^{-1}(1)|}{\sum_{i=0}^{d} \binom{n}{i}}.$$

Theorem 3.1 is an immediate consequence of the following:

LEMMA 3.2. *If Φ is a 2-CNF formula on n variables then Φ accepts at most $\sum_{i=0}^{\pi(\Phi)} \binom{n}{i}$ assignments.*

Theorem 3.1 follows since if f is covered by 2-CNFs $\Phi_1, \ldots, \Phi_M$, then $\pi(\Phi_i) \leq \pi(f)$, and so the lemma implies that each Φ_i accepts at most $\sum_{i=0}^{d} \binom{n}{i}$ assignments and hence M is at least $|f^{-1}(1)| / \sum_{i=0}^{d} \binom{n}{d}$.

So it suffices to prove the lemma. We begin with a definition. A set $Y = \{x_{j_1}, \ldots, x_{j_k}\}$ of variables is said to be *free* with respect to the set S of assignments if any assignment to the variables in Y can be extended to an assignment in S, i.e., for any assignment β to the variables in Y, there exists $\alpha \in S$ such that $\alpha(x_{j_i}) = \beta(x_{j_i})$ for $i \in [k]$. Define $\phi(S)$ to be the size of the largest set of free variables with respect to S.

If P is a projection of dimension d, and $V = \{x_{j_1}, \ldots, x_{j_d}\}$ is a set of representatives from the nonconstant classes of P, then it is easy to see that V is free with respect to P, and hence also free with respect to any superset of P. Hence we have:

PROPOSITION 3.3. *For any set $S \subseteq \{0,1\}^n$, $\phi(S) \geq \pi(S)$.*

In general $\phi(S)$ can be much larger than $\pi(S)$, but the following lemma shows that if S is the set of inputs accepted by a 2-CNF formula then equality holds:

LEMMA 3.4. *Let $S \subseteq \{0,1\}^n$ be the set of inputs accepted by a 2-CNF formula Φ. Then if V is a set of variables that is free with respect to S then there exists a projection $P \subseteq S$ for which the variables in V are in distinct nonconstant classes. Hence $\pi(S) = \phi(S)$.*

PROOF. We will call a literal *free* if the associated variable is free, and *nonfree* otherwise. Consider the implication digraph $D(\Phi)$. By definition, no free literal can imply another. Since the implies relation is transitive, we see that for each nonfree literal y exactly one of the following holds:

1. y is in the same strong component as some free literal.
2. y is implied by one or more free literals, but does not imply any free literals.
3. y implies one or more free literals, but is not implied by any free literals.
4. y neither implies nor is implied by a free literal.

We now construct a projection that satisfies all the clauses. Let α be any satisfying assignment. For each variable x_i of type (4), assign it according to α. For each variable of type (2), set it equal to 1. For each variable of type (3), set it equal to 0. Each remaining literal is set equal to the free literal to whose strong component it belongs. It is easily verified that every assignment consistent with this projection satisfies the formula Φ. □

To complete the proof of Theorem 1, observe that $\phi(S)$ is the VC-dimension (Vapnik & Chervonenkis 1971) of S when considered as a family of subsets of an n-element set. Lemma 3.2 now follows from $\phi(S) = \pi(S)$ and the following standard result from the theory of VC-dimension (see, e.g., Sauer 1972):

LEMMA 3.5. *If A is a family of subsets of an n-element set, and A has VC-dimension at most d, then*

$$|A| \leq \sum_{i=0}^{d} \binom{n}{i}.$$

4. Constructing hard functions: Codes

In this section, we give a simple construction of a function g in logspace uniform NC^1 which requires depth 3 circuits of size $\Omega(2^{n-o(n)})$. To do so, we first produce a function f on n bits that has the property that $f^{-1}(1)$ does not contain any large-dimensional projections. Then, by Theorem 3.1, f cannot be computed by small Σ_3^2 circuits. Finally, we use f to construct another function on $n+1$ bits which indexes f and $\bar{f}$. To do so, we define $g(x_1, \ldots, x_n, x_{n+1}) = x_{n+1} f(x_1, \ldots, x_n) + \bar{x}_{n+1} \bar{f}(x_1, \ldots, x_n)$. If f is hard for Σ_3^2 circuits, $\bar{f}$ is hard for Π_3^2 circuits, and g is hard for all depth 3 circuits.

To construct f, we start with a simple observation: If a set A contains a d-dimensional projection, then the set A has two points at a Hamming distance of at most n/d: If P is a d-dimensional projection, then the partition of the variables it creates must contain a part with at most n/d variables, and by fixing all the variables outside the part consistent with the projection we get

two points which are at a distance of at most n/d. If A is a set of codewords for a code with rate r and distance δ, then A has size 2^{rn} and cannot contain a projection of dimension larger than n/δ. We can use constructions of linear codes to come up with "dense" sets with no large projections (for examples, see Van Lint 1992). For example, one can construct binary BCH codes with codeword length n, dimension $n-1-t\log n$ and distance $2t+1$. Let f_t be the Boolean function which is 1 on the codewords of a BCH code with dimension $n-1-t\log n$. Then f_t is not identically 1 on any projection of dimension larger than $n/(2t+1)$. On the other hand, by Theorem 3.1, any Σ_3^2 circuit computing f_t in size S must accept a projection of dimension at least $\log(|f_t^{-1}(1)|/S)/\log n$. Hence, by taking $t=\sqrt{n/2}$, it follows that S must be at least $\Omega(2^{n-\sqrt{2n}\log n})$.

Summarizing, we have:

THEOREM 4.1. *The function g defined above requires depth 3 circuits of size $\Omega(2^{n-\sqrt{2n}\log n})$.*

5. Constructing hard functions: Low-degree polynomials

In this section, we will exhibit another explicit function in logspace uniform NC^1 for which $s_3^2(f) = 2^{n-o(n)}$. The lower bound on this function will be weaker than that for the function constructed in the previous section using codes. However, this function will have the property that it is not *constant* on any large projection, once certain index bits have been properly instantiated. By comparison, the function constructed in the previous section only has the property that it is not identically one on any large projection. Thus, this construction may be useful in other settings where the previous one is not.

The main idea is to consider the set $H_2(X)$ of multilinear $GF(2)$ polynomials in the variable set X that are homogeneous of degree 2. Each such polynomial is specified by a function a defined on the set $E(X)$ of edges of the complete graph on $\{1,2,\ldots,|X|\}$, where, for $e=\{i,j\}$, $a_e \in \{0,1\}$ is the coefficient of x_ix_j in the polynomial. First we will prove:

LEMMA 5.1. *Let $\epsilon > 0$ and X be sufficiently large (depending on ϵ). If f is a polynomial chosen uniformly at random from $H_2(X)$ then the probability that $\pi(f) \geq |X|^{1/2+\epsilon}$ is strictly less than* 1.

Now, this fact, Theorem 3.1, and the easily proved and well known fact that a nonzero degree 2 polynomial over $GF(2)$ is 1 on at least $2^{|X|-2}$ inputs implies

that for $|X|$ sufficiently large, there is a degree 2 $GF(2)$ polynomial f for which $s_3^2(f) \geq \kappa(f) \geq 2^{|X|-|X|^{1/2+\epsilon}\log_2|X|}$. In fact, the proof of Lemma 5.1 shows that for sufficiently large X, almost all functions in $H_2(X)$ satisfy this inequality. The problem, as usual, is to give a uniform construction of such polynomials, which we do not know how to do. Instead we proceed as follows. Lemma 5.1 can be strengthened to show that one can get good upper bounds on $\pi(f)$ if f is chosen from a k-wise independent distribution.

LEMMA 5.2. *Let $\epsilon > 0$ and k be sufficiently large (depending on ϵ). Let X be a set of size at least k and let D be a probability distribution over $H_2(X)$ such that for any set $\{e_1, \ldots, e_k\}$ of k edges in $E(X)$, the coefficients $a_{e_1}, \ldots, a_{e_k}$ are independent and unbiased. If f is a polynomial chosen from $H_2(X)$ according to D then the probability that $\pi(f) \geq |X|/k^{1/2-\epsilon}$ is strictly less than* 1.

It is well known (see, e.g., Alon *et al.* 1992) that for any integers $k \leq m$, there is an explicitly constructible set $S(m,k)$ of vectors in $\{0,1\}^m$ having size at most $(2m)^{\lceil (k+1)/2 \rceil}$ such that for a vector v chosen uniformly at random from $S(m,k)$, the coordinates of v are k-wise independent random variables. Furthermore, using the construction in Alon *et al.* (1992), the basis vectors which generate this set can be computed in logarithmic space. Noting that each function in $H_2(X)$ is specified by a vector in $\{0,1\}^m$ with $m = \binom{|X|}{2}$, we define $H_2(X,k)$ to be the subset of $H_2(X)$ consisting of those polynomials whose coefficient vector is chosen from $S(m,k)$. Each function in $H_2(X,k)$ can be explicitly indexed by a sequence of at most $b(X,k) = (k+2)\log|X|$ bits.

Again, by Theorem 3.1 and Lemma 5.2 we have:

COROLLARY 5.3. *Given $\epsilon > 0$ and k sufficiently large, for $|X| \geq k$ there exists a function g in $H_2(X,k)$ for which*

$$s_3^2(g) \geq 2^{|X|(1-k^{-1/2+\epsilon}\log_2|X|)}.$$

Now define the function $f_{X,k}$ on the variable set $X \cup Y$ where $|Y| = b(X,k)$ as follows: for an assignment α of X and β of Y, the assignment β of the variables in Y indexes a function g_β in $H_2(X,k)$, and $f_{X,k}(\alpha,\beta) = g_\beta(\alpha)$. Trivially, $s_3^2(f_{X,k}) \geq s_3^2(g)$ for any $g \in H_2(X,k)$. By the above corollary, for k sufficiently large, $s_3^2(f_{X,k}) \geq 2^{|X|(1-k^{-1/2+\epsilon}\log_2|X|)}$. For fixed $\delta > 0$ and all sufficiently large n, we define the Boolean function f_n on n variables as follows. View the first $n - n^{2/3+\delta/2}$ variables as X and the last $n^{2/3+\delta/2}$ variables as Y. Y is large enough to specify a function in $H_2(X,k)$ for $k = n^{2/3}$. Then we have:

COROLLARY 5.4. *For n sufficiently large and for any $\delta > 0$, f_n is logspace uniformly computable in NC^1 and*

$$s_3^2(f_n) \geq 2^{n-n^{2/3+\delta}}.$$

The fact that f_n is logspace uniformly computable in NC^1 follows from the observation that the basis for the space of vectors with limited independence can be generated by a logspace machine. So it remains to prove Lemma 5.1 and its generalization, Lemma 5.2.

PROOF OF LEMMA 5.1. We need to upper bound the probability that a random function in $H_2(X)$ has a large projection on which it is 1. Fix an integer d and let P be a projection of dimension d. As described in section 3, we can represent P by a sequence $(A_0, B_0, A_1, B_1, \ldots, A_d, B_d)$ of subsets of the variables. If f is a polynomial in $H_2(X)$ and G_f is the corresponding graph defined on X, let f_P be the function on variables $y_1, \ldots, y_d$ obtained from f by substituting 1 for each variable in A_0, 0 for each variable in B_0, and for $i \in [d]$ substituting y_i for each variable in A_i and $1 + y_i$ for each variable in B_i. Then f is constant on P if and only if f_P is a constant polynomial. We upper bound the probability that f_P is constant by upper bounding the probability that its degree is at most 1. Let $b_{i,j}$ be the coefficient of $y_i y_j$ in f_P. Then the event that f_P has degree at most 1 is the event that all of $b_{i,j}$ are 0. Now $b_{i,j}$ is just the number (mod 2) of edges in G_f between the sets $A_i \cup B_i$ and $A_j \cup B_j$. For a randomly chosen function in $H_2(X)$, $b_{i,j}$ is uniformly random and the $b_{i,j}$ are mutually independent. Hence the probability that f_P has degree at most 1 is $2^{-\binom{d}{2}}$. Note that the event that f_P has degree at most 1 only depends on the sequence of sets $(A_0 \cup B_0, A_1 \cup B_1, A_2 \cup B_2, \ldots, A_d \cup B_d)$ representing the projection. Since the number of ways to choose such a sequence is at most $(d+1)^n$ we can upper bound the probability that there exists a projection such that f_P has degree at most 1 by $2^{-\binom{d}{2}}(d+1)^{|X|}$. For $d = |X|^{1/2+\epsilon}$, this is less than 1. □

PROOF OF LEMMA 5.2. To show that the probability that $\pi(f) \geq |X|/k^{1/2-\epsilon}$ is strictly less than 1, we need the following:

CLAIM 5.5. *Let f be a Boolean function on a variable set X and $h, d \leq |X|$ be positive integers. If there is a projection of dimension d on which f is constant than there is a projection of dimension at least $dh/|X|-1$ on which f is constant and such that the number of unfixed variables is at most h.*

PROOF. To see the claim, consider a projection ϕ of dimension d on which f is constant, and let $P = (A_0, B_0, A_1, B_1, \ldots, A_d, B_d)$ be a sequence of sets representing the projection, with the parts ordered so that $|A_1 \cup B_1| \leq |A_2 \cup B_2| \leq \ldots \leq |A_d \cup B_d|$. Let j be the largest integer such that the number of the variables in the smallest j parts is at most h. Consider the projection ϕ' obtained from ϕ by fixing, for each $i > j$, all the variables in A_i to 1 and all variables in B_i to 0. Then ϕ' has at most h unfixed variables. Also it is a subset of ϕ, and so f is fixed on ϕ'. It can easily be seen that j, the dimension of ϕ', is at least $h/(|X|/d) - 1$ since $|X|/d$ is the average part size. □

Returning to the proof of Lemma 5.2, let D be a k-wise independent distribution on $H_2(X)$ and suppose f is selected according to D. By the claim, to upper bound the probability that f has a projection of dimension d it suffices to upper bound the probability that it has a projection with $h = \lceil k^{1/2} \rceil$ unfixed variables of dimension at least $d' = dh/|X| - 1$. Consider a projection P with h unfixed variables. Note that for such a projection, the number of pairs of unfixed variables is $\binom{h}{2} \leq k$. Hence, the random variables $a_{i,j}$ where x_i, x_j are unfixed are mutually independent. Thus we can now proceed exactly as in the previous lemma and say that the probability that f_P has degree at most 1 is at most $2^{-\binom{d'}{2}}$. As before we note that the event that f_P has degree at most 1 only depends on the d' parts $\{A_1 \cup B_1, \ldots, A_{d'} \cup B_{d'}\}$. Now we only need to count the d'-part partitions with at most h unfixed variables, and there are at most $(|X|d')^h$ of these and so the probability that for f chosen according to D, there exists a dimension d' projection P with h unfixed variables on which f is constant is at most $(|X|d')^h 2^{-\binom{d'}{2}}$. For $d' \geq |X|/k^{1/2-\epsilon}$ and k sufficiently large, this probability is less than 1. □

6. Conclusions and open problems

The obvious question that is suggested by this work is whether a large set accepted by a k-CNF ($k > 2$) must necessarily contain a projection of large dimension. However, it can be shown that there are large sets defined by even linear size 4-CNF which can only contain projections of dimension bounded by a constant. This follows from the existence of sparse parity check matrices which define codes with linear distance and constant rate in Gallager (1963) and Sipser & Spielman (1994). Results in Gallager (1963) show that there exist matrices with at most 4 1's in each row which are parity check matrices for codes with linear distance. The set of codewords defined by such a matrix is just the AND of many 4-variable parity constraints, and so can be accepted by

a 4-CNF. Because this set of codewords has linear distance, the same argument used in section 4 shows that this set is not 1 on any projection whose size is larger than some fixed constant. This implies that using the idea of projections to prove nonlinear lower bounds on circuit size using Valiant's reduction to depth 3 unbounded fan-in circuits cannot work.

However, it may still be possible to apply the technique directly to linear size and logarithmic depth circuits. In particular, we do not know the answer to the following question: Let $S \subseteq \{0,1\}^n$ be recognizable by a linear size and logarithmic depth (or just even linear size) circuit. Does S or $\bar{S}$ contain a projection of dimension $\Omega(n^\varepsilon)$ for some $\varepsilon > 3/4$? If we have an affirmative answer to the question, then it follows that the hard function constructed in section 5 would require nonlinear circuit size. The codes discussed in section 4 would not suffice since their complements contain large-dimensional projections.

One can also consider more general types of nice subsets of $\{0,1\}^n$. For instance: consider the set of subsets of $\{0,1\}^n$ that are affine subspaces. Is it true that for constant k, every Σ_3^k circuit is constant on an affine subspace of dimension $\Omega(n^\epsilon)$ for some ϵ (or even $\Omega(n)$)? Can one construct an explicit function which has no such subspace? A counting argument shows that almost all homogeneous multilinear polynomials of degree 3 over $GF(2)$ have the property that they are not constant on any affine subspace of dimension more than $\Omega(n^{2/3})$, but we do not yet know how to make this explicit.

Acknowledgements

The authors would like to thank Johan Håstad for pointing out an error in the example in section 2. The authors would also like to thank Russell Impagliazzo, Pavel Pudlák and Jiří Sgall for useful discussions, and the reviewers for helpful comments.

A version of this paper appeared previously as Paturi *et al.* (1997).

The second author would like to acknowledge support from NSF grant CCR-9215293 and from DIMACS (Center for Discrete Mathematics & Theoretical Computer Science), through NSF grant NSF-STC91-19999 and the New Jersey Commission on Science and Technology.

References

M. Ajtai (1983). σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic* **24**, 1–48.

N. Alon, J. Spencer, and P. Erdős (1992). *The Probabilistic Method*. Wiley.

M. Furst, J. B. Saxe, and M. Sipser (1981). Parity, circuits, and the polynomial-time hierarchy. In *22nd Annual Symposium on Foundations of Computer Science*, Nashville, TN, IEEE, 260–270.

R. G. Gallager (1963). *Low Density Parity–Check Codes*. MIT Press.

J. Håstad (1986). Almost optimal lower bounds for small depth circuits. In *Proc. Eighteenth Annual ACM Symposium on Theory of Computing*, Berkeley, CA, 6–20.

J. Håstad, S. Jukna, and P. Pudlák (1993). Top-down lower bounds for depth 3 circuits. In *34th Annual Symposium on Foundations of Computer Science*, Palo Alto, CA, IEEE, 124–129.

R. Paturi, M. E. Saks, and F. Zane (1997). Exponential lower bounds for depth 3 Boolean circuits. In *Proc. Twenty-Ninth Annual ACM Symposium on Theory of Computing*, El Paso, TX, 86–91.

A. A. Razborov (1986). Lower bounds on the size of bounded depth networks over a complete basis with logical addition. *Math. Zametki* **41**, 598–607 (in Russian). English translation in *Math. Notes*.

N. Sauer (1972). On the density of families of sets. *J. Combin. Theory Ser. A* **13**, 145–147.

M. Sipser and D. A. Spielman (1994). Expander codes. In *35th Annual Symposium on Foundations of Computer Science*, Santa Fe, NM, IEEE, 566–576.

R. Smolensky (1987). Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc. Nineteenth Annual ACM Symposium on Theory of Computing*, New York City, 77–82.

L. G. Valiant (1977). Graph-theoretic arguments in low-level complexity. In *Proc. 6th Symposium on Mathematical Foundations of Computer Science*, 162–176.

J. H. Van Lint (1992). *Introduction to Coding Theory*. Springer.

V.N. Vapnik and A. Ya. Chervonenkis (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16**, 264–280.

A. Yao (1985). Separating the polynomial-time hierarchy by oracles (preliminary version). In *26th Annual Symposium on Foundations of Computer Science*, Portland, OR, IEEE, 1–10.

Manuscript received 1 April 1997

Ramamohan Paturi
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093, U.S.A.
paturi@cs.ucsd.edu

Michael E. Saks
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903, U.S.A.
saks@math.rutgers.edu

Francis Zane
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093, U.S.A.
francis@cs.ucsd.edu