Communication in Bounded Depth Circuits

Pavel Pudlák

Abstract: *We show that rigidity of matrices can be used to prove lower bounds on depth 2 circuits and communication graphs. We prove a general nonlinear lower bound on a certain type of circuits and thus, in particular, we determine the asymptotic size of depth d superconcentrators for all depths ≥ 4 (for even depths ≥ 4 it has been determined before).*

1. Introduction

This research is motivated by the problem of proving nonlinear lower bounds for circuits computing explicitly given boolean functions $\mathbf{f} : \{0,1\}^n \to \{0,1\}^n$. An important result of Valiant [V] reduces this problem (with an additional condition of depth $O(\log n)$) to proving lower bounds to certain depth 2 circuits where we allow arbitrary boolean functions as gates. If we can use arbitrary gates, then the phenomenon which causes complexity of the circuits is *information transfer* instead of *information processing* in the usual approach. Thus this research is closely related to the *complexity of communication networks*, with a prominent example of *superconcentrators*. On the other hand it is also related to *algebraic complexity* for two reasons: firstly it is natural to try to solve such questions first in the case when only linear gates over GF_2 are allowed; secondly some problems on communication networks can be treated as problems about computations of some linear mappings.

In the context of linear circuits Valiant's reduction leads to the concept of *the rigidity* of matrices. Valiant [V] has shown that sufficiently large bounds to the rigidity of the matrix corresponding to a linear function $\mathbf{f}$ give nonlinear lower bounds to the size of circuits of logarithmic depth. Unfortunately the known lower bounds on the rigidity of explicitly given matrices are too small. We shall show that the known lower bounds are, however, sufficient to prove nonlinear lower bounds on the size of depth 2 circuits. We shall show that this can be used to reprove the known lower bounds on superconcentrators and similar circuits of depth 2. Dolev, Dwork, Pippenger and Wigderson proved nonlinear lower bounds on the size of superconcentrators of even depths. We give a different and more general proof which gives asymptotically best values also for odd depths and uses weaker assumptions. This solves completely the question of the asymptotical size of bounded depth superconcentrators for depth ≥ 4; the remaining cases 2 and 3 will be considered in a forthcoming paper with Noga Alon [AP], where we determine the asymptotics for depth 3 and improve the lower bound for depth 2.

These results have also the following simple algebraic interpretation: there are simply defined matrices which cannot be decomposed into products of *sparse* matrices, where *sparse* means *linear number of nonzero elements.*

The paper is organized as follows. In section 2 we introduce basic concepts. In section 3 we prove the lower bound for depth 2 circuits based on rigidity. In section 4 we prove the lower bounds for all depths ≥ 3.

2. Algebraic and combinatorial properties

Let M be a matrix over a field F. We shall say that a *circuit computes the matrix* M, if it computes the function $\mathbf{y} = \mathbf{x}M$, (where $\mathbf{x}$ and $\mathbf{y}$ are row vectors). More generally, one can consider *affine* mappings of the form $\mathbf{y} = \mathbf{x}M + \mathbf{a}$, where $\mathbf{a}$ is a constant vector, however the constant term influences the complexity of the circuits very little, therefore we consider only the matrices. A *linear circuit over* F is a directed acyclic graph with input nodes labeled by variables $\mathbf{x}$, output nodes labelled by variables $\mathbf{y}$ and each node which is not an input labeled by *linear functions.* The *size* of a graph or a circuit is the number of *edges.* The *depth* is the length of the longest path (= number of edges in the path). As we are interested only in estimates up to a multiplicative constant, we may assume that the graph of the circuit is *leveled* (in the obvious meaning) whenever the depth is fixed.

We are primarily interested in boolean circuits. If we think of $\{0,1\}$ as GF_2, then boolean circuits are also algebraic circuits in a certain sense, since each boolean function is algebraic over GF_2. However this concept is different from the usual definition of algebraic circuits. In the usual definition we extend the field (here GF_2) by indeterminates $x_1, \ldots, x_n$ and compute *polynomials* i.e. we use only the string of indeterminates as an input, while, say in boolean circuits, we compute the values for all strings of the elements of GF_2. As easily seen these two concepts coincide if we use only linear gates, (or if we use the ring $GF_2[x_1, \ldots, x_n]/(x_1 = x_1^2, \ldots, x_n = x_n^2)$). In particular a linear GF_2 circuit for a 0-1 matrix is a circuit with *parity* gates.

It is well-known that in algebraic circuits multiplication does not help to compute linear functions. In fact, if we have an algebraic circuit computing a linear function which uses some polynomials as gates, then we can replace all nonlinear polynomials by linear ones while preserving the structure of the circuit. For boolean circuits it is an open problem whether nonlinear gates can reduce the size of circuits computing linear functions. However we know that in general we cannot change the nonlinear gates to linear ones while preserving the structure of the circuit and the computed linear function. The counterexample is obtained using a (12,32,5) binary code and the fact that there is no [12,5,5] linear code. (Take 5 input vertices, 12 intermediate ones and $5\binom{12}{4}$ output vertices; on the intermediate level compute the code, on the output level decode all bits in all possible ways using only 8 bits of the code word.) I am indebted to Petr Savický for the information about the existence of such a code.

The following is a well-known fact which relates the structure of a circuit with the rank of the computed matrix.

Lemma 1. (a) Let M be a matrix over a field F, let $r = \mathrm{rank_F M}$. Then in any linear circuit for M there is no $r-1$ vertex cut, hence there are r vertex disjoint paths from inputs to outputs.
(b) Let M be a 0-1 matrix with $\mathrm{rank_{GF_2} M = r}$. Then in any boolean circuit for M there are r vertex disjoint paths from inputs to outputs.

Note that if we know the rank of a *submatrix*, then we can conclude that the corresponding *subsets* of inputs and outputs are connected. This enables us to prove lower bounds on the size of linear and boolean circuits using lower bounds to the size of communication networks (=directed acyclic graphs with certain prescribed connections). However we have also a converse relation.

Construction. Let G be a directed acyclic graph with n inputs and m outputs, let F be an arbitrary field. Assign indeterminates over F to the edges in a one-to-one manner. Assign the product of the assigned values to every path in G. Finally assign the $n \times m$ matrix M to G where M_{ij} is the sum of the values of the paths connecting the $i-th$ input with the $j-th$ output.

Lemma 2. Suppose that there are r vertex disjoint paths connecting the inputs with the outputs in G. Then $\mathrm{rank_F M \geq r}$.

Proof: Let r such paths be chosen. Assign $1's$ to the indeterminates on the paths and $0's$ to the rest. After this substitution M has r 1's all in different rows and different columns; the remaining entries are $0's$. □

We shall interpret G with the labels as a linear circuit over the extension of F by the indeterminates. Let $v_i, i \in I$, be the predecessors of v in G, let z_i be the indeterminates assigned to (v_i, v), let f_i be the function computed at v_i, for $i \in I$. Then the function computed at v is $\sum z_i f_i$. The following is an obvious fact.

Lemma 3. This circuit computes M.

Lemmas 2 and 3 show that lower bounds on the linear circuits based on rank can be translated to lower bounds to communication circuits. Let us consider the following two concepts as an example. An *n-superconcentrator* is a directed acyclic graph with n sources and n sinks such that for every $k \leq n$ and every k element subset X of sources and every k element subset Y of sinks, there are k vertex disjoint paths connecting X with Y. The corresponding algebraic concept is an $n \times n$ matrix whose all square submatrices are regular. Thus proving lower bounds to superconcentrators is equivalent to proving lower bounds on circuits which compute such matrices. There are natural examples of such matrices, e.g. Vandermonde matrices with real positive entries.

We shall use the following weakening of the concept of a matrix with all square matrices regular.

Definition Let M be an $n \times n$ matrix, $0 < \varepsilon, \delta, 0 \leq \eta \leq 1$. We say that M *is $\varepsilon, \delta, \eta$-densely regular* or just $\varepsilon, \delta, \eta - DR$, if for every k, with $\eta n \leq k \leq n$, there are nonempty sets of k element subsets $\mathcal{X}, \mathcal{Y} \subseteq [1, n]^k$ such that for every $i, j = 1, \ldots, n$

$$\delta. \Pr(i \in X) \leq k/n \text{ and } \delta. \Pr(j \in Y) \leq k/n,$$

where $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ are chosen with some probability distributions. Furthermore we require that for random $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ the mean value of the rank of the matrix determined by X and Y is $\geq \varepsilon k$. We shall say that a directed acyclic graph with n inputs and n outputs is $\varepsilon, \delta, \eta - DR$ if the matrix assigned to it (over some field F) has this property, i.e. if for X and Y as above the mean value of the number of vertex disjoint paths connecting them is $\geq \varepsilon k$. We shall denote by $D(n, d, \varepsilon, \delta, \eta)$ the minimal size of a depth d linear circuit computing some $n \times n \quad \varepsilon, \delta, \eta - DR$ matrix over some field F. Equivalently (by Lemmas 1,2,3) $D(n, d, \varepsilon, \delta, \eta)$ is the minimal size of an $\varepsilon, \delta, \eta - DR$ directed acyclic graph with n inputs and n outputs and depth d.

This rather complicated definition has two reasons. Firstly it enables us to cover several interesting classes of matrices and graphs for which nonlinear lower bounds on the size of bounded depth circuits can be proved. Secondly the parameters enable us to perform a proof of the lower bound by induction on the depth. To obtain nonlinear lower bounds we shall need $\eta \to 0$ for $n \to \infty$ and ε, δ fixed.

We shall give several examples of interesting concepts which are covered by this concept. In our examples we shall have $\mathcal{X} = \mathcal{Y}$ equipped with the uniform probability distribution.

The first example is an n-superconcentrator which is clearly 1,1,0-DR.

Let us call a *full triangular matrix* any square matrix which has nonzero elements on the diagonal and above it and which has zeros below. A special case where the nonzero elements are $1's$ has been studied as the *parallel prefix matrix*.

An *n-shifter* is a graph with n sources $u_0, \dots, u_{n-1}$ and n sinks $v_0, \dots, v_{n-1}$ such that for every $0 \le k < n$ there are vertex disjoint paths $u_0 \to v_k, u_1 \to v_{k+1(\mathrm{mod}\, n)}, \dots, u_{n-1} \to v_{k-1(\mathrm{mod}\, n)}$.

An *n-parity shifter* is a graph with n sources $u_0, \dots, u_{n-1}$ and n sinks $v_0, \dots, v_{n-1}$ such that for every $0 \le k < n$ one can remove some edges so that in the remaining graph there is an odd number of paths between u_i and v_j iff $j \equiv i + k (\mathrm{mod}\ n)$.(Note that parity shifter is a generalization of shifter and that it can be equivalently defined as a graph which can be used to compute any shift using suitably chosen *parity* gates.)

Proposition 1. For any field F, (i) any full triangular $n \times n$ matrix is 1/2,1/2,0-DR, (ii) n-shifters and (iii) n-parity shifters are 1,1,0-DR, (iv) Vandermonde matrix is 1,1/2,0-DR.

Proof: (i) Let T be an $n \times n$ full triangular matrix. Let $1 \le k \le n$ be given. Take $l = \lfloor n/k \rfloor$. Define $\mathcal{X}(= \mathcal{Y})$ as the set of sets

$$\{a, a+l, a+2l, \dots, a+(k-1)l\}$$

where $1 \le a \le l$. The sets in $\mathcal{X}$ are disjoint, thus (taking the uniform distribution on $\mathcal{X}$) $\Pr(i \in X) \le l^{-1} \le 2k/n$. It is not hard to see that the submatrix of T determined by X, Y is also a full triangular matrix or a full triangular matrix with the main diagonal replaced by zeros. Hence the average rank is $\ge k-1$ which is $\ge k/2$, for $k \ge 2$, and for $k = 1$ the average rank is $\ge 1/2$ trivially.

(ii) Let G be an n-shifter. Let $k \le n$ be given. Take an arbitrary set $X_0 \subseteq \{1, 2, \dots, n\}$ of size k and put

$$\mathcal{X} = \mathcal{Y} = \{X_0 + l; l = 0, \dots, n-1\} (\text{modulo } n).$$

Then $\Pr(i \in X) = k/n$, and each $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ are connected by k vertex disjoint paths.

(iii) This argument works also for parity shifters we only have to show that if X is a set of inputs in a parity shifter and Y is a shifted copy of X of outputs, then there are $\mid X \mid$ vertex disjoint paths connecting the two sets. This property follows from the fact that parity shifters can be used *to compute* the shifts.

(iv) Use the same system $\mathcal{X}, \mathcal{Y}$ as in (i). Then any matrix determined by $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ is nonsingular. $\square$

Another example of $\varepsilon, \delta, \eta - DR$ graphs are *weak superconcentrators* of [DDPW] (we shall not define them here).

Let us also mention the relation of the computation of matrices by linear circuits of bounded depth to the decomposition of matrices into products. Let C_i be a circuit for a matrix $M_i, for i = 1, \ldots, d$. Connect outputs of C_1 with inputs of C_2 etc, then the circuit computes the product $M = M_1 \times \ldots \times M_d$. If C is a depth d leveled circuit for M, then C can be decomposed in such a way into depth 1 circuits; then the number of edges in the circuits is equal to the number of nonzero elements. Thus the matrices which cannot be computed by small circuits with bounded depth are those which cannot be decomposed into the product of sparse matrices (matrices with small number of nonzero elements).

3. Rigidity and bounded depth circuits

For a matrix M we denote by $|M|$ the number of nonzero elements of M. The *rigidity* of M is the function $R_M^F(r)$ defined by

$$R_M^F(r) = \min\{|A|; \mathrm{rank_F(M + A) \leq r}\}.$$

F denotes the field in question.

Theorem 1. For every positive ε there exists $\delta > 0$ such that for every $n \times n$ matrix M over a field F, $1 \leq m_1, m_2 \leq n$, if

$$R_M^F(r) \geq \varepsilon \frac{n^2}{r}, \text{ for } m_1 \leq r \leq m_2.$$

then for every two matrices A, B such that $M = A \times B$

$$|A| + |B| \geq \delta n \log \frac{m_2}{m_1}.$$

Recall the connection with bounded depth circuits: the conclusion is equivalent to the statement that every depth 2 circuit computing M with linear function as gates must have at least $\delta n \log \frac{m_2}{m_1}$ edges.

Lemma 4. There exists an $\varepsilon > 0$ such that for every $c_1 \geq c_2 \geq \ldots \geq c_n \geq 0, 1 \leq p \leq m \leq n$, if

$$\sum_{i=r}^{n} c_i^2 \geq \frac{1}{r} \text{ for } r = p, p+1, \ldots, m,$$

then

$$\sum_{i=1}^{n} c_i \geq \varepsilon(\log m - \log p).$$

Since $\sum_{i=x}^{\infty} i^{-2} = O(x^{-1})$ and $\sum_{i=x}^{\infty} i^{-1} = \Omega(\log x)$, the lemma is a consequence of the following claim.

Claim. Suppose $1 \leq p \leq m \leq n$, $c_p \geq c_{p+1} \geq \ldots \geq c_m \geq 0$, $c_{m+1}, \ldots, c_n \geq 0$, and

$$\sum_{i=r}^{n} c_i^2 \geq \sum_{i=r}^{m} i^{-2}, \text{ for } r = p, p+1, \ldots, m.$$

Then

$$\sum_{i=p}^{n} c_i \geq \sum_{i=p}^{m} i^{-1}. \tag{*}$$

Proof: We shall use induction on the number of intervals $[a, b]$ such that $p \leq a < b \leq m$,

$$c_a \geq a^{-1},\ c_i < i^{-1},\ \text{for } i = a+1, \ldots, b,$$

and either $b = m$, or $c_{b+1} \geq (b+1)^{-1}$.

(i) If there is no such interval, then $c_i \geq i^{-1}$, for $i = p, p+1, \ldots, m$, hence (*) holds true.

(ii) Suppose that claim holds for all sequences with k such intervals and let $\{c_i\}_{i=p}^n$ be a sequence with $k+1$ such intervals. Let $[a, b]$ be the last interval with this property, thus $c_i \geq i^{-1}$ for $i = b+1, \ldots, m$, or $b = m$. Let $\{d_i\}_{i=p}^n$ be the sequence obtained from $\{c_i\}_{i=p}^n$ using the following operations:

take a pair u, v such that $a+1 \leq u \leq b < v \leq n$ and increase c_u to u^{-1} or to a smaller value by decreasing by the same value c_v to v^{-1} or a larger value if $v \leq m$, or to 0 or a larger value if $v > m$.

More precisely it means that for $v \leq m$ we do the following:

if $u^{-1} - c_u \leq c_v - v^{-1}$, then we replace c_u by u^{-1} and c_v by $c_v - (u^{-1} - c_u)$ and
if $u^{-1} - c_u > c_v - v^{-1}$, then we replace c_v by v^{-1} and c_u by $c_u + (c_v - v^{-1})$;

and for $v > m$

if $u^{-1} - c_u \leq c_v$, then we replace c_u by u^{-1} and c_v by $c_v - (u^{-1} - c_u)$ and
if $u^{-1} - c_u > c_v$, then we replace c_v by 0 and c_u by $c_u + c_v$.

Let $\{c'_i\}_{i=p}^n$ be obtained by such an elementary operation. Then

(1) $c'_i = c_i$ for $i = p, \ldots, a$,

(2) $c'_i \geq i^{-1}$ for $i = b+1, \ldots, m$,

(3) $\sum_{i=a+1}^n c_i = \sum_{i=a+1}^n c'_i$,

(4) $\sum_{i=r}^n c_i^2 \geq \sum_{i=r}^n (c'_i)^2$, for $r = p, \ldots, a$.

The last fact follows easily from $c_u \geq c_v$. Repeat these elementary operations while it is possible; let $\{d_i\}_{i=p}^n$ be the resulting sequence. This means that either

(5) $d_i = i^{-1}$, for all $a+1 \leq i \leq b$,

or

(6) for some i, $a+1 \leq i \leq b$, $d_i < i^{-1}$ and $d_j = j^{-1}$, for all $j = b+1, \ldots, m$ and $d_j = 0$ for all $j = m+1, \ldots n$.

The second case is, however, not possible, since then we would have

$$\sum_{i=a+1}^{n} d_i = \sum_{i=a+1}^{n} c_i,$$

$$\sum_{i=a+1}^{m} d_i^2 = \sum_{i=a+1}^{n} d_i^2 \geq \sum_{i=a+1}^{m} c_i^2 \geq \sum_{i=a+1}^{m} i^{-2},$$

but $d_i \le i^{-1}$, for $i = a+1, \ldots, m$, and some $d_i < i^{-1}$, which would be a contradiction.

The sequence $\{d_i\}_{i=p}^n$ satisfies the assumptions of the claim, since the inequality

$$\sum_{i=r}^{n} d_i^2 \ge \sum_{i=r}^{m} i^{-2},$$

for $r = p, p+1, \ldots, a+1$, follows from (4), and for $r = a+2, \ldots, m$, it follows from the fact that $d_i \ge i^{-1}$ for $i = a+1, \ldots, n$, which is a consequence of (2) for $\{d_i\}_{i=p}^n$, and (5). Since $\{d_i\}_{i=p}^n$ has one interval $[a, b]$ less, we can apply the induction assumption to it, and thus we get

$$\sum_{i=p}^{n} c_i \ge \sum_{i=p}^{n} d_i \ge \sum_{i=p}^{m} i^{-1}.$$

This finishes the proof of the claim and thus also of the lemma. □

Proof of Theorem 1: Suppose $M = A \times B$, i.e. $M(i,j) = \sum_l A(i,l)B(l,j)$. If we replace the entries in the $l-th$ column of A by zeros (or the entries of the $l-th$ row of B by zeros), then the resulting product will differ from M in at most $a_l b_l$ entries, where a_l(resp. b_l) is the number of nonzero elements in the $l-th$ column of A (resp. $l-th$ row of B). Let $l_1, \ldots, l_m$ be a permutation of indices such that

$$a_{l_1} + b_{l_1} \ge \ldots \ge a_{l_m} + b_{l_m}.$$

If we set all columns $l_{r+1}, \ldots, l_m$ in A (or all rows $l_{r+1}, \ldots, l_m$ in B to zero), then the product will have rank at most r. Thus, by the assumption about the rigidity of M

$$\sum_{i=r+1}^{m} a_{l_i} b_{l_i} \ge \varepsilon \frac{n^2}{r} \quad \text{for} \quad m_1 \le r \le m_2.$$

Since $(a_{l_i} + b_{l_i})^2/2 \ge a_{l_i} b_{l_i}$, we can apply Lemma 4 with $c_i = (a_{l_i} + b_{l_i})/n$. Thus we obtain

$$|A| + |B| = \sum (a_{l_i} + b_{l_i}) = \Omega(n(\log m_2 - \log m_1)) = \Omega(n \log \frac{m_2}{m_1}).\Box$$

The simple property of having all square submatrices regular implies the bound on the rigidity needed in the theorem. The following theorem generalizes this observation.

Theorem 2. For every ε, δ positive, there exists a positive α such that for any field F every η, $0 < \eta \le 1$, and any $n \times n$ matrix M which is $\varepsilon, \delta, \eta - DR$,

$$R_M^F(r) \ge \alpha \frac{n^2}{r}, \quad \text{for} \quad \varepsilon\eta n/2 \le r \le \varepsilon n/2.$$

Proof: Let $\varepsilon\eta n/2 \le r \le \varepsilon n/2$ be given. Take $k = \lceil 2r/\varepsilon \rceil$. Let $Z \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ be a minimal set of positions to be changed in M in order to reduce the rank to r. Let $\mathcal{X}, \mathcal{Y}$ be the sets for k. Then

$$\Pr((i,j) \in X \times Y) \le \frac{k^2}{\delta^2 n^2}.$$

Let $\mathbf{z}$ be the random variable $\mid Z \cap (X \times Y) \mid$. Then

$$E\mathbf{z} = \sum_{(i,j)\in Z} \Pr((i,j) \in X \times Y) \leq \mid Z \mid \frac{k^2}{\delta^2 n^2}.$$

On the other hand $E\mathbf{z} \geq r$, since the mean value of the rank of the square submatrix determined by X, Y is $\geq \varepsilon k \geq 2r$, and at least r changes are necessary to reduce the rank from $2r$ to r. Hence

$$r \leq \mid Z \mid \frac{k^2}{\delta^2 n^2}.$$

$$\Rightarrow \qquad \mid Z \mid \geq \frac{\delta^2 r n^2}{k^2} \geq \frac{\varepsilon^2 \delta^2 n^2}{7r}. \quad \square$$

The simplest matrix for which we thus obtain a lower bound $\Omega(n^2/r)$ to the rigidity is the parallel prefix matrix. For this matrix the exact value has been determined in [PV].

Corollary 1. For every positive ε, δ and every $r \leq n$

$$D(n, 2, \varepsilon, \delta, \frac{1}{r}) = \Omega(n \log r).$$

Proof: By Theorems 1 and 2. $\square$

We do not know if the rigidity of $\Omega(n^2/r)$ implies nonlinear lower bounds for circuits of depth larger than 2. The next proposition shows that a larger lower bound would do it (unfortunately we cannot prove such a bound for any explicit matrix).

Proposition 2. Let F be a field and M an $n \times m$ matrix. Suppose $M = A_1 \times \ldots \times A_d$. Then for every r

$$|A_1| + \ldots + |A_d| > r\left(\frac{R_M^F(r)}{n}\right)^{\frac{1}{d}}$$

Thus we get a nonlinear bound, if for some $\varepsilon > 0$ and $\alpha(n) \to \infty$, $R_M^F(\varepsilon n) \geq n.\alpha(n)$.

Proof: Consider the circuit for M given by $M = A_1 \times \ldots \times A_d$, let S be its size ($S = |A_1| + \ldots + |A_d|$). Let r be given. There are at most r vertices in the directed graph of the circuit with outdegree $\geq S/r$. Let N be the $n \times m$ matrix where $N(i,j)$ is the sum of products of elements corresponding to the paths from i to j which hit at least one vertex with outdegree $\geq S/r$. Thus rank(N) $\leq$ r. Let $K = M - N$, thus K is determined by paths which go only through vertices with outdegree $< S/r$. As there are less than $(S/r)^d$ such paths, we have $|K| < n(S/r)^d$. On the other hand $|K| \geq R_M^F(r)$, whence the inequality follows. $\square$

4. A lower bound for depth ≥ 3

We shall show that the $\varepsilon, \delta, \eta - DR$ property implies nonlinear size of bounded depth circuits (and graphs) if $\varepsilon, \delta > 0$ are fixed and $\eta \to 0$ as $n \to \infty$.

Definition. Let f be a function defined on nonnegative integers such that $f(n) < n$ for $n > 0$. We shall denote by f^* the function defined by:

$$f^*(n) = \min\{i; f^{(i)}(n) \leq 1\},$$

where $^{(i)}$ denotes i-times iteration. We define

$$\lambda_1(n) = \lceil \log_2 n \rceil,$$

$$\lambda_{i+1} = \lambda_i^*.$$

(Pippenger [P2] uses $\log^{*\dots**(i-1)} n$ for λ_i.)

Theorem 3. Let $\varepsilon, \delta > 0$ be fixed.
(i) for every positive integers n and $r \leq n$,

$$D(n, 3, \varepsilon, \delta, 1/r) = \Omega(n \log \log r);$$

(ii) for every fixed $d \geq 2$, and every positive integers n and $r \leq n$,

$$D(n, 2d, \varepsilon, \delta, 1/r) = \Omega(n\lambda_d(r)),$$

$$D(n, 2d+1, \varepsilon, \delta, 1/r) = \Omega(n\lambda_d(r)).$$

We shall use the following auxiliary functions:

$$\kappa_0(0) = \kappa_0(1) = 0;\ \kappa_0(n) = \lfloor n^{1/2} \rfloor \text{ for } n > 1;$$

$$\kappa_{i+1} = \kappa_i^*.$$

We shall show lower bounds $\Omega(n\kappa_d(r))$ for depth $2d+1$ and $d \geq 1$, and lower bounds $\Omega(n\lambda_d(r))$ for depth $2d$ and $d \geq 2$. Then we shall use the fact that $\kappa_1(r) \approx \log_2 \log_2 r$ and $\kappa_d(r) \approx \lambda_d(r)$ for $d \geq 2 (f \approx g$ denotes $f = O(g)$ and $g = O(f))$. First we need some general properties of slowly growing functions obtained by iteration.

Lemma 5. Suppose $f(0) = f(1) = 0$ and $f(n) \leq \lfloor n^{1/2} \rfloor$, for every $n > 1$. Then for every $n \geq 0$:

$$f^*(n) \leq f(n) \leq \lfloor n^{1/2} \rfloor; \tag{i}$$

$$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq f^{(i+1)}(n) \text{ for every } \quad i > 0, \text{ provided the denominator is not 0;} \tag{ii}$$

$$f^{(i)}(n) \geq f^*(n)/2 \text{ for every } i \leq f^*(n)/2. \tag{iii}$$

Proof: (i) Clearly $f^*(0) = 0 \leq f(0)$. Let $n \geq 0$, and suppose (i) holds for all $m \leq n$. If $f(n+1) \leq 1$ then $f^*(n+1) = 0 \leq f(n+1)$. Hence assume $f(n+1) > 1$. Since $f(n+1) \leq \lfloor (n+1)^{1/2} \rfloor < n+1$, we have

$$f^*(n+1) = f^*(f(n+1)) + 1 \leq f(f(n+1)) + 1 \leq \lfloor f(n+1)^{1/2} \rfloor + 1 \leq f(n+1).$$

(ii)

$$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq \frac{f^{(i)}(n)}{\lfloor f^{(i)}(n)^{1/2}\rfloor} \geq f^{(i)}(n)^{1/2} \geq f^{(i+1)}(n).$$

(iii) Using the assumption $f(m) \leq \lfloor m^{1/2} \rfloor$ we get

$$f(n) > f^{(2)}(n) > \ldots > f^{(f^*(n))}(n).$$

Since the values of f are integers the gaps are at least 1. Thus we have (iii). □

Lemma 6. Suppose for some $\varepsilon > 0$ and every $n \geq 0$

$$\varepsilon f(n) \leq g(n) \leq f(n) \leq \lfloor n^{1/2} \rfloor.$$

Then for some $\delta > 0$ and every $n \geq 0$

$$\delta f^*(n) \leq g^*(n) \leq f^*(n).$$

Proof: Let n_0 be such that for all $n > n_0$

$$f(n) \leq \lfloor n^{1/2} \rfloor \leq \varepsilon n.$$

Hence if $f(n) > n_0$, then

$$ff(n) \leq \varepsilon f(n) \leq g(n).$$

Thus $f^*(n) \leq 2g^*(n) + n_0$, which implies $\delta f^*(n) \leq g^*(n)$ for some δ. The inequality $g^*(n) \leq f^*(n)$ is obvious. □

Lemma 7. $\kappa_1(n) \approx \log_2 \log_2 n$ and $\kappa_d(n) \approx \lambda_d(n)$ for $d \geq 2$.
Proof: For the first relation we have for $n \geq 2$,

$$\lfloor \log_2 \log_2 n \rfloor + 1 = \kappa_1(2^{2^{\lfloor \log_2 \log_2 n \rfloor}}) \leq \kappa_1(n) \leq \kappa_1(2^{2^{\lfloor \log_2 \log_2 n \rfloor + 1}}) =$$
$$= \lfloor \log_2 \log_2 n \rfloor + 2.$$

This implies that $\kappa_2(n) \approx \lambda_2(n)$, which implies the second relation by Lemma 6. □

Lemma 8. $D(n, 1, \varepsilon, \delta, 1/r) \geq \frac{\varepsilon\delta^2}{2n} r \geq \frac{\varepsilon\delta^2}{2} n r^{1/2}$.
Proof: Let s be the size of a graph with n inputs and n outputs having $\varepsilon, \delta, 1/r - DR$ property. Let $\mathcal{X}, \mathcal{Y}$ be some sets of subsets of cardinality $k =\lceil n/r \rceil$ given by the definition of $\varepsilon, \delta, 1/r - DR$ property. Let e be an edge of the graph. The probability that the end-points of e are in random $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ is at most $(k/\delta n)^2$, hence the mean value of the number of edges between X and Y is at most $s(k/\delta n)^2$. By the property of the graph we know that this number is larger or equal to εk, whence the bound. □

Lemma 9. Let $f(n) \leq \lfloor n^{1/2} \rfloor$, for every $n \geq 0$. Then for every $\alpha, \delta, \varepsilon$ positive reals there exists a positive real β such that if

$$\forall n \forall r \leq n \quad D(n, d, \varepsilon/2, \delta, 1/r) \geq \alpha n f(r), \tag{a}$$

then

$$\forall n \forall r \leq n \quad D(n, d+2, \varepsilon, \delta, 1/r) \geq \beta n f^*(r). \tag{b}$$

Proof: Suppose (a) holds true. Let C be a graph with n inputs and n outputs with depth $d+2$ and with $\varepsilon/2, \delta, 1/r - DR$ property. Let $V_i, i = 0, 1, \ldots, d+2$, be the levels of C. Put

$$A_0 = \{v \in V_1 \cup V_{d+1}; \deg(v) > f(r)\};$$

$$A_i = \{v \in V_1 \cup V_{d+1}; f^{(i+1)}(r) < \deg(v) \leq f^{(i)}(r)\}, \text{ for } i > 0.$$

Claim. For every $i, 1 \leq i \leq f^*(r)/2 - 3$ at least one of the following inequalities holds:

$$| A_0 \cup \ldots \cup A_{i-1} | \geq \frac{\varepsilon}{4} \cdot \frac{n}{f^{(i+1)}(r)}; \tag{1}$$

$$| \{(u,v); (u,v) \text{ incident with } A_i \cup A_{i+1} \cup A_{i+2}\} | \geq \frac{\varepsilon\delta}{4} n \ ; \tag{2}$$

$$| \{(u,v); (u,v) \text{ not incident with } A_0 \cup \ldots \cup A_{i+2}\} | \geq \alpha n \frac{f^{(i+2)}(r)}{f^{(i+3)}(r)}. \tag{3}$$

Proof of the claim: Let i be given. Suppose neither (1) nor (2) is true; we shall prove (3). Let k be an arbitrary integer such that $n/f^{(i+1)}(r) \leq k \leq n$. Let X, Y be random subsets of inputs, resp. outputs, of cardinality k (from the appropriate $\mathcal{X}$ and $\mathcal{Y}$). In the average there are at least εk vertex disjoint paths connecting them. Out of these paths there are at most $\frac{\varepsilon}{4} \cdot \frac{n}{f^{(i+1)}(r)} \leq \frac{\varepsilon}{4} \cdot k$ paths through $A_0 \cup \ldots \cup A_{i-1}$ by non (1), and there are at most $\frac{\varepsilon}{4} \cdot k$ paths through $A_i \cup A_{i+1} \cup A_{i+2}$ in *the average* by non (2), since the mean value of the number of vertices of $X \cup Y$ connected with $A_i \cup A_{i+1} \cup A_{i+2}$ is less than $\frac{k}{\delta n} \cdot \frac{\varepsilon\delta}{4} \cdot n = \frac{\varepsilon}{4} \cdot k$. Thus in the average at least $\frac{\varepsilon}{2} k$ such paths avoid $A_0 \cup \ldots \cup A_i \cup A_{i+1} \cup A_{i+2}$. Now we shall transform C into a depth d graph C': first omit the vertices of V_1 and V_{d+1}, then for each pair of edges $(u,v), (v,w)$ where $v \in A_{i+3} \cup A_{i+4} \cup \ldots$ add the edge (u,w). Thus the size of C' is at most $f^{(i+3)}(r)$-times larger than the size of C, since $f^{(i+3)}(r)$ is an upper bound on the degrees of vertices in $A_{i+3} \cup A_{i+4} \cup \ldots$. Above we have shown that C' is $\varepsilon/2, \delta, 1/f^{(i+1)}(r)$-DR. Thus by the assumption of the lemma C' has size at least

$$D(n, d, \varepsilon/2, \delta, (f^{(i+1)}(r))^{-1}) \geq \alpha n f(f^{(i+1)}(r)) = \alpha n f^{(i+2)}(r).$$

Hence

$$| C | \geq \frac{| C' |}{f^{(i+3)}(r)} \geq \frac{\alpha n f^{(i+2)}(r)}{f^{(i+3)}(r)}.$$

The last two inequalities follow by (ii) and (iii) of Lemma 5. ∎

Now we finish the proof of the lemma. We consider three cases corresponding to the conditions of the claim.

Case 1: for *some* $i \leq f^*(r)/2 - 3$ we have (1). Then, since each vertex in $A_0 \cup \ldots \cup A_{i-1}$ has degree at least $f^{(i)}(r)$,

$$| C | \geq \frac{\varepsilon}{4} \cdot \frac{n}{f^{(i+1)}(r)} \cdot f^{(i)}(r) \geq \frac{\varepsilon}{4} \cdot n \cdot f^{(i+1)}(r) \geq \frac{\varepsilon}{8} \cdot f^*(r).$$

Case 2: for *all* $i \leq f^*(r)/2 - 3$ we have (2). Then

$$| C | \geq \frac{1}{3} \cdot (\frac{f^*(r)}{2} - 3) \cdot \frac{\varepsilon\delta}{4} n = \Omega(n \cdot f^*(r)),$$

since the sets A_i are disjoint.

Case 3: for *some* $i \leq f^*(r)/2 - 3$ we have (3). Then

$$| C | \geq \alpha n \frac{f^{(i+2)}(r)}{f^{(i+3)}(r)} \geq \alpha n f^{(i+3)}(r) \geq \frac{\alpha}{2} n f^*(r).$$

□

Proof of Theorem 3 follows from Corollary 1, Lemmas 7,8 and 9. □

5. Applications and open problems

We shall state explicitly some applications of the lower bounds of Corollary 1 and Theorem 3 and mention related open problems.

By our results we get the following lower bounds to the bounded depth superconcentrators: $\Omega(n \log n)$ for depth 2; $\Omega(n \log\log n)$ for depth 3; $\Omega(n\lambda_d(n))$ for depth $2d$ and $2d+1, d \geq 2$. The same lower bound for depth 2 has been proved in [P1] where also an upper bound $O(n \log^2 n)$ is proved. In [AP] we shall prove $\Omega(n \log^{3/2} n)$ for depth 2. This shows a gap between week superconcentration properties (e.g. parallel prefix computations) and usual superconcentrators. A more direct proof of $\Omega(n \log\log n)$ for depth 3 (based on the same idea) and an upper bound of the same growth rate is shown in [AP]. Lower bounds $\Omega(n\lambda_d(n))$ (using a different proof) and upper bounds $O(n\lambda_d(n))$ have been shown in [DDPW] for depth $2d$. Since an upper bound for depth $2d$ is also an upper bound for depth $2d+1$, our lower bound for odd depths ≥ 5 is optimal. Thus the only open problem is to determine the size of depth 2 superconcentrators.

Next we consider boolean circuits computing the parallel prefix matrix. We get the same lower bounds as for superconcentrators: $\Omega(n \log n)$ for depth 2; $\Omega(n \log\log n)$ for depth 3; $\Omega(n\lambda_d(n))$ for depth $2d$ and $2d+1, d \geq 2$. (*Proof-sketch:* By Proposition 1, the $n \times n$ parallel prefix matrix is 1/2,1/2,0-DR. Let C be a boolean circuit for the matrix, let G be the underlying graph of C. By Lemma 1(b), G is $1/2, 1/2, 0-DR$, hence we can apply Corollary 1 and Theorem 3.) On the other hand it is not difficult to construct *linear* circuits for the parallel prefix matrix of sizes $O(n \log n), O(n \log\log n), O(n\lambda_d(n))$ respectively. Thus we get optimal bounds for *all* depths. These are the best lower bounds for explicitly defined 0-1 matrices even if we consider linear circuits over arbitrary fields. For matrices with large elements Shoup and Smolensky [SS] have shown a bound $\Omega(n^{1+1/d})$ for depth d. For nonlinear boolean functions we can get a little bit more for depth 2 boolean circuits with arbitrary gates using the lower bounds on depth 2 superconcentrators. Namely it is quite easy to design a boolean function with $O(n)$ input and output variables such that any circuit computing the function is an n-superconcentrator.

For shifters our bounds ($\Omega(n \log n)$ for depth 2; $\Omega(n \log\log n)$ for depth 3; $\Omega(n\lambda_d(n))$ for depth $2d$ and $2d+1, d \geq 2$) are rather small. The size of the smallest depth dn-shifters is $\approx n^{1+1/d}$ [PY]. It is possible that this is also the minimal size of *parity* shifters, but we only have the weak lower bounds presented above. The possibility of applying such methods to shifting circuits was first discovered in [PS].

Superconcentrators failed to produce nonlinear lower bounds to bounded fan-in and $O(\log n)$ depth circuits but it is still possible that this long-standing problem can be solved using, for instance, circuits computing shifts. Therefore we propose to study graphs such as the parity shifters.

References

[AP] N.Alon, P.Pudlák *Superconcentrators of depth 2 and 3,* to appear in Journ. of ACM.

[DDPW] D.Dolev, C.Dwork, N.Pippenger, A.Wigderson, *Superconcentrators, generalizers and generalized connectors (preliminary version)*, Proc. ACM STOC 1983, 42-51.

[P1] N.Pippenger, *Superconcentrators of depth 2,* J. Comput. System Sci. 24 (1982), 82-90.

[P2] N.Pippenger, *Communication networks,* in *Handbook of Theoretical Computer Science,* Ed. *J.* van Leeuwen, Elsevier, 1990, 806-833.

[PY] N.Pippenger, A.C.-C.Yao, *Rearrangeable networks with limited depth*, SIAM J. Alg. Disc. Meth. 3, (1981) 411-417.

[PS] P.Pudlák, P.Savický, *On shifting networks,* to appear in Theoretical Computer Science.

[PV] P.Pudlák, Z.Vavřín, *Computation of rigidity of order n^2/r for one simple matrix,* Comment. Math. Univ. Carolinae 32,2 (1991), 213-218.

[R] A.A.Razborov, *On rigid matrices* (in russian), unpublished.

[SS] V.Shoup, R.Smolensky, *Lower bounds for polynomial evaluation and interpolation,* Proc. IEEE FOCS 1991, 378-383.

[V] L.G.Valiant, *Graph-theoretic arguments in low level complexity*, Proc. MFCS 1977, Springer-Verlag LNCS, (1977) 162-176.

Author's address:

Matematický ústav AVČR
Žitná 25
11567 Praha 1
Czech Republic