ENTROPY OF OPERATORS OR WHY MATRIX MULTIPLICATION IS HARD FOR DEPTH-TWO CIRCUITS

STASYS JUKNA

Abstract. We consider unbounded fanin depth-2 circuits with *arbitrary* boolean functions as gates. We define the entropy of an operator $f : \{0,1\}^n \to \{0,1\}^m$ as the logarithm of the maximum number of vectors distinguishable by at least one special subfunction of f.

Our main result is that every depth-2 circuit for f requires at least *entropy*(f) wires. This is reminiscent of a classical lower bound of Nechiporuk on the formula size, and gives an information-theoretic explanation of *why* some operators require many wires. We use this to prove a tight estimate $\Theta(n^3)$ of the smallest number of wires in any depth-2 circuit computing the product of two n by n matrices over any finite field. Previously known lower bound for this operator was $\Omega(n^2 \log n)$.

1. Introduction

One of the challenges in circuit complexity is to prove a nonlinear lower bound for *logarithmic depth* circuits computing an explicitly given boolean operator $f : \{0,1\}^n \to \{0,1\}^n$. This corresponds to simultaneous computation of the sequence of boolean functions $f_j : \{0,1\}^n \to \{0,1\}$, where $f_j(\vec{x})$ is the j-th coordinate of the vector $f(\vec{x})$. An important result of Valiant [25] reduces this problem to proving lower bounds for certain depth-2 circuits, where we allow arbitrary boolean functions as gates. Note that in this case the phenomenon which causes complexity of circuits is *information transfer* instead of *information processing* in the case of single functions. It is therefore important to understand what properties of operators do force high information transfer in their depth-2 circuits.

A *depth-2 circuit* for $f : \{0,1\}^n \to \{0,1\}^m$ is a directed acyclic graph of depth 2 with n input nodes $x_1, \ldots, x_n$, and m output nodes $z_1, \ldots, z_m$. Every noninput node computes an *arbitrary* boolean function of its inputs, and there is no bound on the fanin or on the fanout. The *size* of a circuit is the total number of wires in it. Without loss of generality, we may assume that there are no direct wires from inputs to outputs: this can be easily achieved by adding at most n new wires. Thus, in depth-2 circuits we have three layers of nodes: the input, the middle and the output layer.

Let $s_2(f)$ denote the minimum size of a depth-2 circuit computing f. Note that $s_2(f) \leq n^2$ holds for every operator $f : \{0,1\}^n \to \{0,1\}^n$.

Superlinear lower bounds of the form $\Omega(n \log n)$ for depth-2 circuits were obtained using graph-theoretic arguments by analyzing some superconcentration properties of the circuit as a graph [8, 14, 18, 16, 22]. Lower bounds of the form $\Omega(n \log^{3/2} n)$ were proved in [2, 20]. Unfortunately, the approach based on superconcentrators cannot lead to lower bounds

To appear in *Theory of Computing Systems*.

Research supported by the DFG grant SCHN 503/4-1.

Institute of Mathematics and Informatics, Vilnius, Lithuania.

Email: jukna@thi.informatik.uni-frankfurt.de.

for depth-2 circuits larger than $\Omega(n \log^2 n)$, since there are depth-2 superconcentrators with $O(n \log^2 n)$ [15], and even $O(n \log^2 n / \log \log n)$ [21] edges.

The (numerical) limitation of the graph-theoretic lower bounds comes from their power: they show much more than that the number of wires must be large—they also provide an information about the structure of the underlying graphs. It is therefore natural to expect to prove larger lower bounds, if we only care about the number of wires in a circuit, not about its structure. And indeed, such an approach has already led to a lower bound of the form $s_2(f) = \Omega(n^{3/2})$ for a bilinear operator of cyclic convolution [6].

In this paper we use an even more direct argument to prove a general lower bound $s_2(f) \geq$ *entropy*(f), where the entropy of $f : \{0,1\}^n \to \{0,1\}^m$ is just the logarithm of the maximum number of vectors distinguishable by at least one special subfunction of f. This gives a simple explanation of what operators and, more importantly, *why* require many wires. The amazing simplicity of the proof indicates that hight entropy of operators is a fundamental reason for a complicated information transfer in their circuits. The bound itself is reminiscent of a classical lower bound of Nechiporuk [13] on the formula size of a boolean function as the logarithm of the number of its subfunctions.

Since *entropy*(f) is relatively easy to compute, it gives us a handy tool to prove large lower bounds for a whole string of explicit operators. We demonstrate this by a tight estimate $\Theta(n^3)$ of the smallest number of wires in any depth-2 circuit computing the product of two n by n matrices over any finite field. This improves the highest previously known lower bound $s_2(f) = \Omega(n^2 \log n)$ for this operator derived in [22].

2. Results

In this section we first introduce the notion of entropy of operators, and state some its basic properties. Then we prove our main result—a general lower bound on the number of wires in depth-2 circuits in terms of the entropy (Lemma 2 and Theorem 1).

2.1. **Entropy of function sets.** In general, a (uniform) entropy of a mapping $f : U \to V$ is the smallest number of binary bits that are necessary to specify each single value $f(x)$ of f. That is, the logarithm $\log_2 |S|$ of the largest number of elements in a set $S \subseteq U$ on which f is injective. By looking at sets of functions as corresponding mappings, we arrive to the following definition.

Let $F = \{f_1, \ldots, f_m\}$ be a set of functions $f_j : \{0,1\}^n \to \{0,1\}$ on the same set of variables $x_1, \ldots, x_n$. Say that a set of vectors $S \subseteq \{0,1\}^n$ is *separated* by F, if for every pair of vectors $a \neq b \in S$ there is a function $f \in F$ with $f(a) \neq f(b)$, that is, if the corresponding operator $F : \{0,1\}^n \to \{0,1\}^m$ in injective on S. Define

$$\textit{entropy}(F) = \max\{\log_2 |S| \colon S \subseteq \{0,1\}^n \text{ and } F \text{ separates } S\}.$$

Example 1. If $F(\vec{x}) = A\vec{x}$ is a linear operator over GF_2 for some boolean $m \times n$ matrix A with $m \leq n$, then *entropy*(F) is just the rank r of A over GF_2, since the operator $F(a)$ takes the same value on 2^{n-r} out of all 2^n input vectors a.

Say that a boolean function f can be computed from a set of boolean functions G if there exists a boolean function φ such that $f = \varphi(g_1, \ldots, g_k)$ for some functions $g_1, \ldots, g_k$ in G. Note that, in any circuit, every function is computed from the set of functions computed at its inputs. In particular, every set of functions F on variables $x_1, \ldots, x_n$ is computable from $G = \{x_1, \ldots, x_n\}$.

Proposition 1. *Let F and G be some finite sets of boolean functions in n variables.*

(i) Upper bound: $\mathit{entropy}(F) \leq \min\{n, |F|\}$.
(ii) Lower bound: *if F contains r single variables, then $\mathit{entropy}(F) \geq r$.*
(iii) Main connection: *if every function in F can be computed from the functions in G, then $\mathit{entropy}(F) \leq \mathit{entropy}(G) \leq |G|$.*

Proof. (i) The set $F = \{f_1, \ldots, f_m\}$ defines a natural encoding of vectors $a \in \{0,1\}^n$ by vectors $F(a) = (f_1(a), \ldots, f_m(a))$ in $\{0,1\}^m$. If a set $S \subseteq \{0,1\}^n$ is separated by F, then each vector in S must receive its own code, implying that $|S| \leq 2^m = 2^{|F|}$, and hence, $\log_2 |S| \leq |F|$.

(ii) Suppose that F contains r single variables $x_1, \ldots, x_r$. Let $S \subseteq \{0,1\}^n$ be an arbitrary set of $|S| = 2^r$ vectors having the same values on all remaining $n - r$ variables. Since any pair of vectors $a \neq b \in S$ must differ in at least one of the first r coordinates, each such pair is separated by at least one of the variables $x_1, \ldots, x_r$.

(iii) Just observe that then $G(a) = G(b)$ implies $F(a) = F(b)$. Hence, any set separated by F must be also separated by G, implying that $\mathit{entropy}(F) \leq \mathit{entropy}(G) \leq |G|$, where the last inequality follows from (i). □

2.2. **Entropy of subfunctions and the number of wires.** Let F and G be two sets of boolean functions. We can think of F as a set of functions computed by some circuit at its output nodes, and G as a set of functions computed at some intermediate nodes. Fix some set $\vec{x} = (x_1, \ldots, x_n)$ of variables, and call them *main variables*. Let $\vec{y} = (y_1, \ldots, y_r)$ be the set of the remaining *auxiliary* variables.

We say that a main variable x_i is *critical* for a function $g(\vec{x}, \vec{y})$ if $g(\vec{e}_i, \vec{y}) \neq g(\vec{0}, \vec{y})$, where $\vec{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the vector of length n with precisely one 1 in the i-th coordinate.

Given a subset $X \subseteq \{x_1, \ldots, x_n\}$ of main variables, let $X(g)$ denote the set of all variables $x_i \in X$ which are critical for g. The number $|X(g)|$ of variables in $X(g)$ is the *weight* of g with respect to the set of variables X. The *weight* of a set G of functions, denoted by $\mathit{weight}_X(G)$, is the sum

$$\mathit{weight}_X(G) = \sum_{g \in G} |X(g)|$$

of weights of all its functions. We will see soon (Lemma 2) that, in depth-2 circuits, this number lower bounds to the number of wires leaving the inputs in X.

If every function in F can be computed from the functions in G, then Proposition 1(iii) implies $|G| \geq \mathit{entropy}(F)$. To get a similar (entropic) lower bound on $\mathit{weight}_X(G)$, we consider the following set F_X of subfunctions of the functions in F.

We define the set F_X of subfunctions of F with respect to X to be the set of all boolean functions $h(\vec{y})$ that can be obtained from some function $f \in F$ by setting some variable $x_i \in X$ to 1 and all the remaining main variables to 0. That is,

$$F_X = \{f(\vec{e}_i, \vec{y}) \colon f \in F, x_i \in X\}.$$

Note that F_X may contain up to $|X| \cdot |F|$ different functions.

Lemma 1 (Entropy and weight). *If every function in F can be computed from the functions in G, then*

$$\mathit{weight}_X(G) + |G| \geq \mathit{entropy}(F_X).$$

Proof. Since the functions in F can be computed from the functions in G, the subfunctions in F_X can be computed from the subfunctions in G_X, as well. By Proposition 1(iii), we have $\mathit{entropy}(F_X) \leq \mathit{entropy}(G_X) \leq |G_X|$. It remains, therefore, to show that $|G_X| \leq \mathit{weight}_X(G) + |G|$.

To show this, recall that G_X consists of all boolean functions $g(\vec{e}_i, \vec{y})$ obtained from some function $g \in G$ by setting some variable $x_i \in X$ to 1 and the remaining main variables to 0. If $x_i \notin X(g)$, then $g(\vec{e}_i, \vec{y}) = g(\vec{0}, \vec{y})$. Hence, for each $g \in G$, the set $\{g(\vec{e}_i, \vec{y}) \colon x_i \in X\}$ consist of at most $|X(g)|$ functions $g(\vec{e}_i, \vec{y})$ with $x_i \in X(g)$, and just one additional function $g(\vec{0}, \vec{y})$. Summing over all $g \in G$, we obtain that $|G_X| \leq |G| + \sum_{g \in G} |X(g)| = |G| + \mathit{weight}_X(G)$. □

Let now $f = (f_1, \ldots, f_m)$ be an operator and $F \subseteq \{f_1, \ldots, f_m\}$. Let also $X \subseteq \{x_1, \ldots, x_n\}$ be a subset of main variables. Lemma 1 yields the following basic relation between the entropy and the number of wires.

Lemma 2 (Entropy and the number of wires)**.** *In any depth-2 circuit computing f, the number of wires leaving the inputs in X or entering the outputs in F must be at least $\mathit{entropy}(F_X)$. Moreover, the outputs in F must have at least $\mathit{entropy}(F)$ neighbors on the middle layer.*

Proof. Let M be the set of all nodes on the middle layer joined by a wire with at least one output in F. Then F must be computable from the set $G = \{g_v \colon v \in M\}$ of boolean functions computed at the nodes $v \in M$. Proposition 1(iii) implies that $|M| \geq |G| \geq \mathit{entropy}(F)$, proving the second claim.

To prove the first claim, observe that we must have at least $|M| \geq |G|$ wires entering the outputs in F. Hence, by Lemma 1, it remains to show that at least $\mathit{weight}_X(G)$ wires must leave the inputs in X.

Each node $v \in M$ must be connected by a wire with each input $x_i \in X$ of which the function g_v depends. Hence, at least $|X(g_v)|$ wires must go from X to the node v. Since no wire can go to more than one node, the total number of wires from X to M must be at least $\sum_{v \in M} |X(g_v)| = \mathit{weight}_X(G)$. □

Remark 1. Note that the same argument also works for arbitrary (not just depth-2) circuits: If V is a set of nodes such that each path from X to F goes through at least one node in V, then $|V|$ plus the number of paths from X to V must be at least $\mathit{entropy}(F_X)$.

Define the *entropy* of an (n, m)-operator $f = (f_1, \ldots, f_m)$ as

$$\mathit{entropy}(f) = \max \sum_{t=1}^{p} \mathit{entropy}(\{f_j(\vec{e}_i, \vec{y}) \colon i \in I_t, j \in J_t\}), \tag{1}$$

where the maximum is over all partitions $I_1, \ldots, I_p$ of inputs $[n]$ and all partitions $J_1, \ldots, J_p$ of outputs $[m]$. Since the total number of wires in a depth-2 circuit is just the number of wires incident to its input or output nodes, Lemma 2 directly yields the following

Theorem 1 (Entropy Criterium)**.** *For every operator f, we have $s_2(f) \geq \mathit{entropy}(f)$.*

Remark 2. Theorem 1 can be readily extended to sequences of functions $f : D^n \to D$ for any *finite* set D. For this, it is enough to take the logarithm to the basis $|D|$ in the definition of the entropy. The rest is the same.

Remark 3. Taking *partitions* of inputs and outputs in the definition of $\mathit{entropy}(f)$ is not crucial. For each natural number k, we can define $\mathit{entropy}_k(f)$ as the maximum (1) over all subsets $I_1, \ldots, I_p$ of inputs and all subsets $J_1, \ldots, J_p$ of outputs such that no element belongs to more than k of these sets. Hence, taking partitions corresponds to $k = 1$. Now, if $d(i)$ is the number of wires leaving the input i, then the sum

$$\sum_{t=1}^{p} \sum_{i \in I_t} d(i) = \sum_{i=1}^{n} \sum_{t: i \in I_t} d(i) \leq k \sum_{i=1}^{n} d(i)$$

is at most k times larger than the total number $\sum_{i=1}^{n} d(i)$ of wires leaving the inputs. Since the same also holds for the number of wires entering the output nodes, Lemma 1 implies

$$s_2(f) \geq \max_{k \geq 1} \frac{1}{k} \cdot \mathit{entropy}_k(f).$$

Remark 4. Note that $\mathit{entropy}(f)$ is not only a lower bound on the size of any depth-2 circuit computing the operator $f(\vec{x}, \vec{y})$ on *all* inputs $(\vec{x}, \vec{y})$, but also on the size of any depth-2 circuit correctly computing f on special inputs of the form $(\vec{e}_i, \vec{y})$; on inputs $(\vec{x}, \vec{y})$, where $\vec{x}$ has more than one 1, the circuit may output arbitrary values.

3. Application: matrix multiplication

Theorem 1 allows one to show that $s_2(f)$ must be super-linear for many operators $f = (f_1, \ldots, f_m)$ on two sets of variables X and Y. For this, it is enough that we can split the set $F = \{f_1, \ldots, f_m\}$ of functions computed by this operator into some number p of disjoint sets $F_1, \ldots, F_p$ such that, for some partition $X_1, \ldots, X_p$ of the set of variables X, and for each $t = 1, \ldots, p$, we can obtain each single variable $y \in Y$ by taking some function $f \in F_t$ and fixing one its variable $x \in X_t$ to 1 and the rest of X to 0. (We say in this case that f *isolates* the variable y.) Then, by Proposition 1(ii), the set of subfunctions in each F_t with respect to the corresponding set of variables X_t must have entropy at least $|Y|$. By Theorem 1, we then have $s_2(f) \geq p|Y|$.

One of the most natural functions isolating *all* its single variables is a scalar product function $f(\vec{x}, \vec{y}) = x_1 y_1 + x_2 y_2 + \cdots + x_r y_r$; then $f(\vec{e}_i, \vec{y}) = y_i$ for all $i = 1, \ldots, r$. Hence, natural examples of operators of large entropy are sequences of particular scalar products. Many operators computing sequences of bilinear functions, including that of cyclic convolution considered in [6], fall in this general (scalar product) frame. We illustrate this with one important example—matrix product.

Given two $r \times r$ boolean matrices $X = (x_{i,j})$ and $Y = (y_{i,j})$, our goal is to compute their product $Z = X \cdot Y$ over GF_2. The corresponding operator $f = \mathrm{mult}_n(X, Y)$ has $n = 2r^2$ input variables, arranged in two matrices, and consists of $n = r^2$ scalar products $f_{i,j} = \sum_{k=1}^{r} x_{i,k} y_{k,j}$ corresponding to the entries of the product matrix $Z = (z_{i,j})$. (This time indexes of variables as well as of computed functions are *pairs* of numbers.)

Since mult_n is just a sequence of r^2 scalar products on $2r$ variables, $(2r)r^2 = 2n^{3/2}$ is a trivial upper bound, even in depth-1. If we put no restrictions on the depth, then Strassen's algorithm [24], improved in [7], gives a circuit of size $O(n^{1.2})$. The only know lower bound in the unrestricted case, however, is the lower bound $2.5 \cdot n$ proved in [4]. A lower bound $s_2(\mathrm{mult}_n) = \Omega(n \log n)$ for depth-2, as well as nonlinear lower bounds for any constant depth, were proved in [22] using superconcentrators. For depth-2, entropy arguments yield a tight estimate $s_2(\mathrm{mult}_n) = \Theta(n^{1.5})$.

Lemma 3. $\mathit{entropy}(\mathrm{mult}_n) \geq n^{3/2}$.

Proof. Let $f = \mathrm{mult}_n$, and let $\vec{e}_{i,k}$ be the boolean $r \times r$ matrix with precisely one 1 in the position (i,k). Since

$$f_{i,j}(\vec{x},\vec{y}) = x_{i,1} \cdot y_{1,j} + \cdots + x_{i,k} \cdot y_{k,j} + \cdots + x_{i,r} \cdot y_{r,j},$$

we have that

$$f_{i,j}(\vec{e}_{i,k},\vec{y}) = 0 \cdot y_{1,j} + \cdots + 1 \cdot y_{k,j} + \cdots + 0 \cdot y_{r,j} = y_{k,j},$$

for all $j = 1,\ldots,r$. That is, the i-th row $f_{i,1}(\vec{e}_{i,k},Y),\ldots,f_{i,r}(\vec{e}_{i,k},Y)$ of the product matrix $\vec{e}_{i,k} \cdot Y$ is just the k-th row $y_{k,1},\ldots,y_{k,r}$ of Y.

Hence, if we take $X_i = \{x_{i,1},\ldots,x_{i,r}\}$ (the i-th row of X) and $F_i = \{f_{i,1},\ldots,f_{i,r}\}$ (the i-th row of the product matrix), then the corresponding set of subfunctions of F_i with respect to the variables in X_i,

$$\left\{\begin{array}{cccc} f_{i,1}(\vec{e}_{i,1},Y) & f_{i,2}(\vec{e}_{i,1},Y) & \cdots & f_{i,r}(\vec{e}_{i,1},Y) \\ f_{i,1}(\vec{e}_{i,2},Y) & f_{i,2}(\vec{e}_{i,2},Y) & \cdots & f_{i,r}(\vec{e}_{i,2},Y) \\ \vdots & \vdots & & \vdots \\ f_{i,1}(\vec{e}_{i,r},Y) & f_{i,2}(\vec{e}_{i,r},Y) & \cdots & f_{i,r}(\vec{e}_{i,r},Y) \end{array}\right\} = \left\{\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,r} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,r} \\ \vdots & \vdots & & \vdots \\ y_{r,1} & y_{r,2} & \cdots & y_{r,r} \end{array}\right\}$$

contains all $r^2 = n$ variables of Y. Together with Proposition 1(ii), this implies that, for each $i = 1,\ldots,r$, the entropy of F_i with respect to X_i is at least n. Hence, $\mathit{entropy}(f) \geq rn = n^{3/2}$. □

Remark 5 (Limitations). How large can entropy of operators be? Recall that in the definition of $\mathit{entropy}(f)$ of an (n,m)-operator f, we first split the inputs into p blocks $I_1,\ldots,I_p$ of some sizes $a_1 \leq a_2 \leq \ldots \leq a_p$, and the outputs into p blocks $J_1,\ldots,J_p$ of some sizes $b_1,\ldots b_p$. Then we just take the sum of the entropies of the corresponding (to these blocks) sets of subfunctions. Say that a partition is *balanced* if $b_1 \geq b_2 \geq \ldots \geq b_p$. Note that the partition (into the rows) which we used for the matrix product is balanced—there all b_i's were even equal.

Since each of the sets $\{f_j(\vec{e}_i,\vec{y})\colon i \in I_t, j \in J_t\}$ can have at most $|I_i \times J_i| = a_i b_i$ functions, Proposition 1(i) implies that the entropy of this set cannot exceed $a_i b_i$. If the partition is balanced, then Chebyshev's inequality yields

$$\mathit{entropy}(f) \leq \sum_{i=1}^{p} a_i b_i \leq \frac{1}{p}\Big(\sum_{i=1}^{p} a_i\Big)\Big(\sum_{i=1}^{p} b_i\Big) \leq \frac{nm}{p}.$$

On the other hand, by Proposition 1(i), we have a trivial upper bound $\mathit{entropy}(f) \leq pn$. Substituting $p \geq \mathit{entropy}(f)/n$ in the previous inequality, we obtain that $\mathit{entropy}(f) \leq n\sqrt{m}$. Thus, at least with respect to balanced partitions, the entropy of any (n,m)-operator does not exceed $n\sqrt{m}$. In particular, for such partitions, matrix multiplication has the largest possible entropy $\Theta(n^{3/2})$ among all (n,n)-operators.

4. Open problems

In view of Valiant's result [25], reducing log-depth circuits to depth-2 circuits with unbounded fanin gates, it is important to first be able to prove a lower bound $s_2(f_A) = \Omega(n^{1+\epsilon})$ on the number of wires in a depth-2 circuit, computing an explicit *linear* operator $f_A(\vec{x}) = A\vec{x}$ over GF_2. No such bound is known even for *linear* depth-2 circuits, where we only allow linear functions (sums mod 2) as gates. In this case, matrices A requiring $\Omega(n^2/\log n)$ exist. So, again, the problem is to construct such a matrix.

Superlinear lower bounds $\Omega(n \log n)$ and even of the form $\Omega(n \log^{3/2} n)$ for linear depth-2 circuits were obtained by analyzing the determinant or the rigidity of the underlying matrix [11, 12, 1, 9, 16, 17, 5]. For such circuits over the real field a lower bound $\Omega(n^{3/2})$ was proved in [23]. However in their result it is essential that they use large integers in the matrix. It remains an open problem to prove such a bound for 0-1 matrices. For GF_2 the largest bound is $\Omega(n \log^{3/2} n)$ [19, 2, 20]. It remains unclear to what extent entropy can help when dealing with linear operators.

Actually, in the case of *arbitrary* gates, even extremal values of $s_2(f)$ are not well understood. We already know that some natural (n, n)-operators $f : \{0,1\}^n \to \{0,1\}^n$ require $s_2(f) = \Omega(n^{3/2})$ wires. On the other hand, $s_2(f) \leq n^2$ is a trivial upper bound for every such operator.

Problem 1. *What is the maximum of* $s_2(f)$ *over all operators* f*?*

The case of *linear* operators $f_A(\vec{x}) = A\vec{x}$ is particularly interesting. Note that the number 2^{n^2} of such operators is much smaller than the number 2^{n2^n} of all operators.

Problem 2. *What is the maximum of* $s_2(f_A)$ *over all linear operators* f_A*?*

It is well known that in the class of linear depth-2 circuits this maximum is $\Theta(n^2/\log n)$ (see, e.g., [19]). In the class of depth-2 circuits with *arbitrary* gates a lower bound $s_2(f_A) = \Omega(n \log n)$ holds already for triangular matrix A [16]. But the gap is still very large.

A less famous problem about depth-2 circuits, related to another old problem in circuit complexity (proving lower bounds for ACC circuits), is the following one.

A *symmetric* depth-2 circuit is a depth-2 circuit, where the gates on the middle layer compute parities of their inputs, and each output gate computes the same symmetric function of its inputs. That is, each output gate gives the value 1 iff the number of 1's in its input belongs to some specified for the whole circuit subset S of natural numbers. We also assume that there are no direct wires from an input to an output node.

Say that a circuit computing an operator $f = (f_1, \ldots, f_n)$ *represents* a given boolean $n \times n$ matrix $A = (a_{ij})$ if $f_i(\vec{e}_j) = a_{ij}$ for every i and j. That is, the circuit is only required to be correct on inputs with precisely one 1. Let $\mathrm{sym}_2(A)$ be the minimum number of nodes on the middle layer in a symmetric depth-2 circuit representing A. That is, now we count nodes, not wires.

This measure has an equivalent definition in terms of set intersections. Namely, $\mathrm{sym}_2(A)$ is the smallest number m such that it is possible to associate (not necessary different) vectors from $\{0,1\}^m$ to the rows and columns so that $a_{ij} = 1$ iff the scalar product over the reals of the corresponding vectors belongs to some fixed set $S \subseteq \{0, 1, \ldots, m\}$. To see the equivalence, just associate with each output node i and each input node j the sets U_i and V_j of all their neighbors on the middle layer. Then $a_{ij} = f_i(\vec{e}_j) = 1$ iff $|U_i \cap V_j| \in S$.

Using counting argument, the existence of matrices A with $\mathrm{sym}_2(A) = \Omega(n)$ is quite easy to show: there are 2^{n^2} different matrices, but only 2^{m+1} possibilities to chose S, and only $(2^m)^{2n}$ possibilities to assign the sets U_i and V_j.

Problem 3 (Yao)**.** *Prove* $\mathrm{sym}_2(A) = 2^{(\log\log n)^{\omega(1)}}$ *for an explicit matrix* A.

Together with the results of [26, 3], this would yield a super-polynomial lower bound for ACC-circuits. Unfortunately, no lower bound larger than a trivial one $\mathrm{sym}_2(A) \geq \log_2 n$ is known. If we would not require output gates be symmetric functions, then $O(\log n)$ nodes on the middle layer would be already enough to represent every matrix A. Since matrices with $\mathrm{sym}_2(A) = \Omega(n)$ exist, this means that the symmetry of gates is a severe restriction, and should be captured in the lower bounds argument. The difficulty here, however, is that the circuit can use *arbitrary* symmetric functions. High lower bounds of the form $\mathrm{sym}_2(A) = n^{\Omega(1)}$ are only known for circuits using arbitrary threshold or modular gates, or their negations [10]. An interesting question is to extend this to *interval* gates accepting a given input vector iff the number of 1's in it lies in a given interval S of consecutive numbers.

Problem 4. *Prove* $\mathrm{sym}_2(A) \geq (\log_2 n)^{1+\epsilon}$ *in the case of arbitrary interval gates.*

The next difficulty with Problem 3 is that the circuit is allowed to output arbitrary values on inputs $\vec{x}$ with more than one 1. Would we require that the circuit must correctly compute the whole transformation $f_A(\vec{x}) = A\vec{x}$ over GF_2, the problem would be trivial. Since *entropy*(f_A) is at least the rank $\mathrm{rk}(A)$ of A over GF_2 (see Example 1), Proposition 1(iii) implies that any depth-2 circuit (not just a symmetric one) *computing* f_A must have at least $\mathrm{rk}(A)$ nodes on the middle layer. In particular, if I is the $n \times n$ identity matrix, then n nodes are necessary. On the other hand, it can be shown that $m = 2\log_2 n$ nodes on this layer are already enough to *represent* the matrix I, i.e. that $\mathrm{sym}_2(I) \leq 2\log_2 n$. Sketch: put m nodes on the middle layer, associate with each input variable x_i its *own* $m/2$-element subset S_i of these nodes, and join the i-th output node with all middle nodes outside this set S_i. Then there is a path from the i-th input to the j-th output iff $i \neq j$.

References

[1] N. Alon, M. Karchmer, and A. Wigderson. Linear circuits over GF(2). *SIAM. J. Comput.* 19(6): 1064–1067, 1990.

[2] N. Alon and P. Pudlák. Superconcentrators of depth 2 and 3; odd levels help (rarely). *J. Comp. Sys. Sci.* 48: 194–202, 1994.

[3] R, Beigel and J. Tarui. On ACC. *Comput. Complexity* 4: 350–366, 1994.

[4] N. H. Bshouty. A lower bound for matrix multiplication. *SIAM J. Comput.* 18: 759–765, 1982.

[5] P. Bürgisser and M. Lotz. Lower bounds on the bounded coefficient complexity of bilinear maps. *J. ACM*, 51(3): 464–482, 2004.

[6] D. Y. Cherukhin. The lower estimate of complexity in the class of schemes of depth 2 without restrictions on a basis. *Moscow University Mathematics Bulletin* 60(4): 42–44, 2005.

[7] D. Coppersmith and S. Winograd. Matrix multiplications via arithmetic progressions. *J. Symb. Comp.* 9: 251–280, 1990.

[8] D. Dolev, C. Dwork, N. Pippenger, and A. Wigderson. Superconcentrators, generalizer and generalized connectors with limited depth. In *Proc. 15th STOC*, pages 42–51, 1983.

[9] J. Friedman. A note on matrix rigidity. *Combinatorica* 13: 235–239, 1993.

[10] S. Jukna. On set intersection representations of graphs. *J. Graph Theory* (submitted).

[11] J. Morgenstern. Note on a lower bound on the linear complexity of fast Fourier transform. *J. ACM* 20(2): 305–306, 1973.

[12] J. Morgenstern. The linear complexity of computation. *J. ACM* 22(2): 184–194, 1975.

[13] E. I. Nechiporuk. On a Boolean function. *Soviet Math. Doklady* 7(4): 999–1000, 1966.

[14] N. Pippenger. Superconcentrators. *SIAM J. Comput.* 6: 298–304, 1977.

[15] N. Pippenger. Superconcentrators of depth 2. *J. Comput. Syst. Sci.* 24: 82–90, 1982.

[16] P. Pudlák. Communication in bounded depth circuits. *Combinatorica* 14(2): 203–216, 1994.

[17] P. Pudlák. A note on the use of determinat for proving lower bounds on the size of linear circuits. *Inf. Process. Letters* 74: 197–201, 2000.

[18] P. Pudlák and P. Savický. On shifting networks. *Theoretical Comput. Sci.* 116: 415–419, 1993.

[19] P. Pudlák and V. Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Math.* 136: 253–279, 1994.

[20] P. Pudlák, V. Rödl and J. Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. Comput.* 26(3): 605–633, 1997.

[21] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.* 13(1): 2–24, 2000.

[22] R. Raz and A. Shpilka. Lower bounds for matrix product in bounded depth circuits with arbitrary gates. *SIAM J. Comput.* 32(2): 488–513, 2003.

[23] V. Shoup and R. Smolensky. Lower bounds for polynomial evaluation and interpolation problems. *Comput. Complexity* 6(4): 301–311, 1997.

[24] V. Strassen. Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpoliationskoefizienten. *Numer. Math.* 20: 238–251, 1973.

[25] L. Valiant. Graph-theoretic methods in low-level complexity. In *Proc. 6th MFCS*, Springer Lect. Notes in Comput. Sci. 53, pages 162–176, 1977.

[26] A. C. Yao. On ACC and threshold circuits. In *Proc. 31th FOCS*, pages 619–627, 1990.