Electronic Colloquium on Computational Complexity, Report No. 19 (2008)

Note
Entropy of operators or why matrix multiplication is hard for small depth circuits

Stasys Jukna *†‡§

February 27, 2008

Abstract

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

We prove that every depth-2 circuit for f requires at least entropy(f) wires. This generalizes and substantially simplifies the argument used by Cherukhin in 2005 to derive the highest known lower bound $\Omega(n^{3/2})$ for the operator of cyclic convolutions. We then show that the multiplication of two $n^{1/2}$ by $n^{1/2}$ matrices over any finite field has entropy $\Omega(n^{3/2})$.

1 Introduction

One of the challenges in circuit complexity is to prove a nonlinear lower bound for log-depth circuits computing explicitly given boolean operator $f : \{0,1\}^n \to \{0,1\}^n$. This corresponds to simultaneous computation of the sequence $f = (f_1, \ldots, f_n)$ of boolean functions $f_j : \{0,1\}^n \to \{0,1\}$, where $f_j(x)$ is the j-th coordinate of the vector $f(x)$. An important result of Valiant [19] reduces this problem to proving a lower bound $\Omega(n^{1+\epsilon})$ on the number of wires in a depth-2 circuit computing a linear transformation $y = Ax$ over GF_2, 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.

A *depth-2 circuit* for $f : \{0,1\}^n \to \{0,1\}^m$ is a directed acyclic graph with n input nodes $x_1, \ldots, x_n$, and m output nodes $y_1, \ldots, y_m$. 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.

Let $s_2(f)$ denote the minimum size of a depth-2 circuit computing f.

*Institute of Mathematics and Computer Science, Vilnius, Lithuania

†Mailing address: University of Frankfurt, Institute of Informatics, Robert-Mayer-Str. 11-15, D-60054 Frankfurt, Germany.

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

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

ISSN 1433-8092

Superlinear lower bounds for $s_2(f)$ were obtained using graph-theoretic arguments by analyzing some superconcentration properties of the circuit as a graph [7, 8, 11, 10, 14]. Unfortunately, the approach based on superconcentrators cannot lead to lower bounds for depth-2 circuits larger than $\Omega(n \log^2 n)$, since there are depth-2 superconcentrators with $O(n \log^2 n)$ [9], and even with $O(n \log^2 n / \log\log n)$ [13] edges.

The (numerical) limitation of these results 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 a direct approach has led Cherukhin [5] to the highest known lower bound $s_2(f) = \Omega(n^{3/2})$ for an explicit boolean operator f—cyclic convolution computing n special *bilinear* forms $x^\top Ay$ over GF_2. (Recall that such a bound for a *linear* operator Ax would imply nonlinear lower bound for log-depth circuits.)

In this note we take a look at Cherukhin's argument from a more general and more simple perspective. This leads to a general lower bound $s_2(A) \geq \text{entropy}(f)$ for depth-2 circuits in terms of the entropy of the computed operators. The bound is very easy to prove and easy to apply. More importantly, it gives a simple explanation of *why* some operators require many wires. Since we allow *arbitrary* gates, the reason (quite naturally) turns out to be of information-theoretic nature: large number of wires is forced by the high entropy of operators, where the entropy on an arbitrary mapping $g : A \to B$ is defined as the maximum of $\log_2 |S|$ over all subsets $S \subseteq A$ on which g is injective.

Amazing simplicity of the proof itself indicates that this (high entropy) is a fundamental reason causing complexity in depth-2 circuits. Since entropy(f) is easy to compute, this gives us a handy tool to prove large lower bounds for a whole string of operators. We demonstrate this by a few-lines proof that the operator $f(X, Y) = X \cdot Y$ computing the product of two $\sqrt{n} \times \sqrt{n}$ matrices over an arbitrary finite field has entropy $\Omega(n^{3/2})$.

2 Entropy of function sets

Let D be a finite set with $d = |D|$ elements, and H some set of functions $h : D^n \to D$. Say that H *separates* a set of vectors $\Omega \subseteq D^n$ if each pair of vectors in Ω is separated by at least one function in H, that is, if for every pair $a \neq b \in \Omega$ there exists $h \in H$ such that $h(a) \neq h(b)$. In other words, a set of functions H separates Ω if the corresponding operator is injective on Ω. The maximum bit size $\log_d |\Omega|$ of a set Ω separated by H is the *entropy* of H, and is denoted by entropy(H). Note that we always have entropy$(H) \leq n$.

Our argument is based on the following two obvious facts.

Proposition 1. *If H contains r single variables, then* entropy$(H) \geq r$.

Proof. If H contains $x_1, \ldots, x_r$, then any set $\Omega \subseteq D^n$ of $|D|^r$ vectors, having the same values on all remaining $n - r$ variables, is separated by H. □

Say that a function h can be computed from a set of functions G if there exists a function $\varphi : D^k \to D$ such that $h = \varphi(g_1, \ldots, g_k)$ for some functions $g_1, \ldots, g_k$ in G.

Proposition 2. *For every set H of functions $h : D^n \to D$, we have* entropy$(H) \leq \min\{n, |H|\}$. *If all functions in H can be computed from the functions in G, then* $|G| \geq$ entropy(H).

Proof. To prove the first claim, let $h_1, \ldots, h_t$ be the functions in H, and assign to each vector $a \in D^n$ its code $H(a) = (h_1(a), \ldots, h_t(a))$ in D^t, $t = |H|$. If a set of vectors $\Omega \subseteq D^n$ is separated by H, then each vector a in Ω must receive its *own* code $H(a)$, implying that $|\Omega| \leq d^t$, and hence, $|H| = t \geq \log_d |\Omega|$.

For the second claim, just observe that $G(a) = G(b)$ implies $H(a) = H(b)$. Hence, any set of vectors separated by H must be also separated by G. □

3 Entropy and depth-2 circuits

Let D be an arbitrary finite set, e.g., some fixed finite field. We only require that D contains at least two elements, say, 0 and 1 (any other two distinct elements would work.)

Let $f = (f_1, \ldots, f_m)$ be a sequence of functions over the set D, all on the same set of variables. Fix some subset of variables $X = \{x_1, \ldots, x_n\}$, and let Y be the set of the remaining variables (these are *free* variables). The lower bound below holds for *any* choice of X.

With each subset of inputs $I \subseteq [n] = \{1, \ldots, n\}$ and each subset of outputs $J \subseteq [m]$ we associate the set of subfunctions

$$f[I, J] = \{f_j(\mathbf{e}_i, Y) \colon i \in I, j \in J\},$$

where $\mathbf{e}_i \in \{0,1\}^X$ is the vector $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with precisely one 1 in the i-th coordinate. Hence, $f[I, J]$ is the set of all (at most $|I \times J|$) functions $h : D^Y \to D$ in variables Y such that $h(Y)$ can be obtained from some function f_j with $j \in J$ by setting precisely one of the variables x_i with $i \in I_t$ to 1 and the rest to 0. Recall that $\text{entropy}(f[I, J]) \geq r$ if we can obtain r different single-variable functions $h(Y) = y_k$ in this way. Define the *entropy* of the operator f as

$$\text{entropy}(f) = \max \sum_{t=1}^{p} \text{entropy}(f[I_t, J_t]),$$

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 $[m]$.

Theorem 3. $s_2(f) \geq \text{entropy}(f)$.

Proof. Since the total number of wires in a depth-2 circuit is just the number of wires incident to its input or output nodes, it is enough to prove the following Lemma.

For any set I of inputs and any set J of outputs in a depth-2 circuit, let W_I be the set of all wires leaving I, and W_J be the set of all wires entering J.

Lemma 4. $|W_I| + |W_J| \geq \text{entropy}(f[I, J])$.

To prove the lemma, let U be the set of all nodes on the middle layer, and let $g_u(X, Y)$ denote the function computed at a node $u \in U$. For each node $i \in I \cup J$, let $U_i \subseteq U$ be the set its neighbors in U. With each input $i \in I$ and output $j \in J$ we associate the following sets of subfunctions

$$G_i = \{g_u(\mathbf{e}_i, Y) \colon u \in U_i\} \quad \text{and} \quad H_j = \{g_u(\mathbf{0}, Y) \colon u \in Uj\}.$$

Let $G = \bigcup_{i \in I} G_i$ and $H = \bigcup_{j \in J} H_j$. Note that $|G_i| \leq |U_i|$ (resp., $|H_j| \leq |U_j|$) is at most the number of wires leaving i (resp., entering j). Hence, $|G| \leq |W_I|$ and $|H| \leq |W_J|$. By

Proposition 2, it remains to show that each function $f_j(\mathbf{e}_i, Y)$ can be computed from functions in $G_i \cup H_j$.

Inputs of the j-th output gate are precisely the nodes in U_j. Hence, the function f_j computed at the j-th output gate must be computable from the functions g_u with $u \in U_j$. This means that also the subfunction $f_j(\mathbf{e}_i, Y)$ can be computed from the subfunctions $g_u(\mathbf{e}_i, Y)$ with $u \in U_j$. If $u \in U_j \cap U_i$, then the function $g_u(\mathbf{e}_i, Y)$ belongs to G_i by the definition of G_i. If $u \in U_j \setminus U_i$, then there is no wire between i and u, meaning that the value of g_u does not depend on the i-th variable x_i. In this case we have $g_u(\mathbf{e}_i, Y) = g_u(\mathbf{0}, Y)$, implying that $g_u(\mathbf{e}_i, Y)$ belongs to H_j. Hence, $f_j(\mathbf{e}_i, Y)$ can be computed from the functions in $G_i \cup H_j$.

This completes the proof of Lemma 4, and thus, the proof of Theorem 3. □

Theorem 3 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 variables in 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 to 0. (We say in this case that f *isolates* the variable y.) By Proposition 2, we then have entropy$(F_t) \geq |Y|$, implying that $s_2(f) \geq p|Y|$.

One of the most natural functions isolating *all* single variables is a scalar product function $f(x, y) = x_1y_1 + x_2y_2 + \cdots + x_ry_r$; then $f(\mathbf{e}_i, y) = y_i$ 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 n-convolution considered in [5], fall in this general (scalar product) frame. We illustrate this with one important example—matrix product.

Example 5 (Entropy of matrix product)**.** Given two $r \times r$ boolean matrices $X = (x_{i,j})$ and $Y = (y_{i,j})$ over a finite field D, our goal is to compute their product $Z = X \cdot Y$ over D. The corresponding operator $f = mult_n(X, Y)$ 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 [18], improved in [2], gives a circuit of size $O(n^{6/5})$. The only know lower bound in the unrestricted case, however, is the lower bound $2.5 \cdot n$ [4]. A lower bound $s_2(mult_n) = \Omega(n \log n)$ for depth-2, as well as nonlinear lower bounds for any constant depth, were proved in [14] using superconcentrators. For depth-2, entropy arguments yield much higher lower bound.

Lemma 6. entropy$(mult_n) = \Omega(n^{3/2})$.

Proof. Let $f = mult_n$, and let $\mathbf{e}_{i,k}$ be the boolean $r \times r$ matrix with precisely one 1 in the position (i, k). Since $f_{i,j} = \sum_{k=1}^{r} x_{i,k}y_{k,j}$, we have that $f_{i,j}(\mathbf{e}_{i,k}, Y) = y_{k,j}$ for all $j = 1, \ldots, r$. This implies that the i-th row $f_{i,1}(\mathbf{e}_{i,k}, Y), \ldots, f_{i,r}(\mathbf{e}_{i,k}, Y)$ of the product matrix $\mathbf{e}_{i,k} \cdot Y$ is just the k-th row $y_{k,1}, \ldots, y_{k,r}$ of Y. Hence, if we take $I_t = J_t = \{(t, 1), \ldots, (t, r)\}$ (the t-th row), then the set $f[I_t, J_t] = \{f_b(\mathbf{e}_a, Y) \colon a \in I_t, b \in J_t\}$ contains all $r^2 = n$ variables of Y. By Proposition 1, we have entropy$(f[I_t, J_t]) \geq n$ for each $t = 1, \ldots, r$, implying that entropy$(f) \geq rn = \Omega(n^{3/2})$. □

Remark 7 (Limitations)**.** How large can entropy of operators be? Recall that in the definition of entropy(f) of an (n,m)-operator $f : D^n \to D^m$, we first split the inputs into p blocks $I_1,\ldots,I_p$ of some sizes $a_1 \le a_2 \le \ldots \le a_p$, 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 $b_1 \ge b_2 \ge \ldots \ge b_p$. 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 in each set $f[I_i,J_i]$ we can have at most $|I_i \times J_i| = a_i b_i$ functions, the entropy of this set cannot exceed $a_i b_i$. If the partition is balanced, then Chebyshev's inequality yields

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

On the other hand, we have a trivial upper bound entropy(f) $\le pn$. Substituting $p \ge$ entropy(f)$/n$ in the previous inequality, we obtain that entropy(f) $\le n\sqrt{m}$. 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 Concluding remarks and open problems

A natural question is to extend the entropic approach to circuits of depth $d \ge 3$. It is clear that the entropy of the sets of functions computed at each level can only increase when going from outputs to inputs. The problem is to relate the entropy with the number of wires between these layers, like we have done this for depth two. At this point note that the proof of Lemma 4 also holds for circuits of *any* depth: it is enough to replace the setW_J by the set P_J of *paths* (not just wires) starting in the first (next to the inputs) layer and entering nodes in J. This yields

$$|W_I| + |P_J| \ge \text{entropy}(f[I,J]).$$

For depth-3 circuits ($d = 3$) this version of Lemma 4 can be used to derive lower bounds of the form $\Omega(n \log n)$. Such a lower bound for cyclic convolution is already proved in a forthcoming paper [6]. So, we only sketch how the same lower bound can be derived for the matrix multiplication using the entropy.

Let W_I is the number of wires between the nodes in the output and the first layer. The number of the remaining wires is $\sum_{i=1}^{m} d_i$, where $d_1 \ge d_2 \ge \ldots \ge d_m$ are the degrees of the the nodes on the third (next to the outputs) layer. The squares of these numbers give us a trivial upper bound $|P_J| \le \sum_{i=1}^{m} d_i^2$ on the number of paths between the first and the output layer. Knowing that this sum of squares must be large, at least entropy$(f[I,J]) - |W_I|$, it remains then to show that the sum $\sum_{i=1}^{m} d_i$ of the numbers themselves must be large. This can be done by using the following consequence of an interesting technical lemma from [10].

Lemma 8. *Let $a_1 \ge \ldots \ge a_m$ be a sequence of real numbers in some interval $[0,R]$ summing up to A. Then $\sum_{i=1}^{m} \sqrt{a_i} \ge \epsilon\sqrt{A} \cdot \ln(A/R)$, where $\epsilon > 0$ is an absolute constant.*

Proof. Let p be the maximal number such that the sum $a_{p+1} + \cdots + a_m$ of all but the first p numbers is smaller than $A/(p+1)$. Lemma 4 of [10] implies that then $\sum_{i=1}^{p} \sqrt{a_i} \ge \epsilon\sqrt{A} \cdot \ln p$. Since $A/2 \le A - A/(p+1) \le \sum_{i=1}^{p} a_i \le pR$ implies $p \ge A/2R$, we are done. □

For the operator $f = mult_n$ in $2n$ variables computing the product of two $r \times r$ matrices ($n = r^2$) this yields a lower bound of the form $r \cdot \Omega(\sqrt{n} \ln r) = \Omega(n \log n)$. Can this be improved to $\Omega(n^{1+\epsilon})$?

Actually, even the power of depth-2 circuits is far from being understood. As mentioned in the introduction, a lower bound $\Omega(n^{1+\epsilon})$ on the number of wires in a depth-2 circuit, computing an explicit linear transformation Ax over GF_2, would yield a nonlinear lower bound for log-depth circuits. To approach this problem, it is natural to first prove such a bound for *linear* depth-2 circuits, where we only allow linear functions (sums mod 2) as gates. For circuits over the real field a lower bound $\Omega(n^{3/2})$ was proved in [16]. 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 $\Omega(n \log^{3/2} n)$ [1, 10, 12]. In would be therefore interesting to extend the entropic approach to depth-2 circuits computing *linear* operators.

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 two circuit is a depth two circuit, where the gates on the middle layer compute ORs 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.

For a boolean $n \times n$ matrix $A = (a_{ij})$, let $f_A = (f_1, \ldots, f_n)$ be a sequence of boolean functions with $f_i(x) = \bigvee_{j=1}^{n} a_{ij} x_j$. That is, $f_A(x)$ computes a $(\wedge, \vee)$-boolean matrix-vector product Ax. Let $\mathrm{sym}_2(A)$ be the minimum number of nodes on the middle layer in a symmetric depth-2 circuit computing f_A. That is, now we count nodes, not wires.

Simple counting shows that matrices with $\mathrm{sym}_2(A) = \Omega(n)$ exist. The problem, due to Yao [20], is to exhibit an *explicit* boolean matrix A with large $\mathrm{sym}_2(A)$. In terms of set intersection representations of matrices, this problem was re-stated by Pudlák and Rödl in [12] (see Problem 10). To see the equivalence between $\mathrm{sym}_2(A)$ and their measure, 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 $a_{ij} = f_i(\mathbf{e}_j) = 1$ iff $|U_i \cap V_j| \in S$.

What we need is an explicit boolean $n \times n$ matrix A with $\mathrm{sym}_2(A) = \exp((\log \log n)^{\omega(1)})$. Together with the results of Yao [20], and Beigel and Tarui [3], this would yield a super-polynomial lower bound for *ACC circuits*. These are constant depth unbounded fanin circuits over a basis consisting of AND, OR and a finite number of modulo-counting functions: each such function gives the value 1 iff the number of 1's in the input is not divisible by p. When p is a prime, exponential lower bounds were proved by Razborov [15] and Smolensky [17]. However, the case of composite moduli p (even when one moduli $p = 6$ is allowed) remains widely open.

References

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

[2] D. Coppersmith, S. Winograd, Matrix multiplications via arithmetic progressions, J. Symb. Comp. 9 (1990) 251-280.

[3] R, Beigel, J. Tarui On ACC, Computational Complexity 4 (1994) 350–366.

[4] N. H. Bshouty, A lower bound for matrix multiplication, SIAM J. Comput. 18 (1982) 759-765.

[5] D. Yu. Cherukhin, The lower estimate of complexity in the class of schemes of depth 2 without restrictions on a basis, Vestnik Moscow University, Ser. 1, Mathematika 60(4) (2005) 54-56 (in Russian).

[6] D. Yu. Cherukhin, Lower bounds of complexity for depth-2 and depth-3 boolean circuits with arbitrary gates, in: Proc. 3rd Int. Comput. Sci. Symposium in Russia (CSR 2008) (to appear).

[7] D. Dolev, C. Dwork, N. Pippenger, A. Wigderson, Superconcentrators, generalizer and generalized connectors with limited depth, in: Proc. 15th STOC (1983) 42-51.

[8] N. Pippenger, Superconcentrators, SIAM J. Comput. 6 (1977) 298-304.

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

[10] P. Pudlák, Communication in bounded depth circuits, Combinatorica 14 (2) (1994) 203-216.

[11] P. Pudlák, P. Savický, On shifting networks, Theoretical Comput. Sci. 116 (1993) 415-419.

[12] P. Pudlák, V. Rödl, Some combinatorial-algebraic problems from complexity theory, Discrete Math. 136 (1994) 253-279.

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

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

[15] A. A. Razborov, Bounded-depth formulae over the basis $\{\&, \oplus\}$ and some combinatorial problem, in: S.I. Adian (ed.), Problems of Cybernetics, Complexity Theory and Applied Mathematical Logic (1988) 149–166 (in Russian).

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

[17] R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity, in: Proc. 19th STOC (1987) 77–82.

[18] V. Strassen, Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpoliationskoefizienten, Numer. Math. 20 (1973) 238-251.

[19] L. Valiant, Graph-theoretic methods in low-level complexity, in: Proc. 6th MFCS, Springer Lect. Notes in Comput. Sci. 53 (1977) 162-176.

[20] A. C. Yao, On ACC and threshold circuits, in: Proc. 31th FOCS (1990) 619–627.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/