Lower Bounds for Boolean Circuits with Finite Depth and Arbitrary Gates

Dmitriy Yu. Cherukhin

Mech.-Math. Dept., MSU, Moscow, 119992, Russia
cherukhin@gmail.com

Abstract. We consider bounded depth circuits over an arbitrary field K. If the field K is finite, then we allow arbitrary gates $K^n \to K$. For instance, in the case of field $GF(2)$ we allow any Boolean gates. If the field K is infinite, then we allow only polinomials.

For every fixed depth d, we prove a lower bound $\Omega(n\lambda_{d-1}(n))$ for the size (i.e. the number of wires) of any circuit for computing the cyclic convolution over the field K. In particular, for $d = 2, 3, 4$, our bounds are $\Omega(n^{1.5})$, $\Omega(n \log n)$ and $\Omega(n \log\log n)$ respectively; for $d \geqslant 5$, the function $\lambda_{d-1}(n)$ is slowly growing. On the Boolean model, our bounds are the best known for all even d and for $d = 3$. For $d = 2, 3$, we prove these bounds in previous papers [11, 13].

Key words: Boolean function, circuit, complexity, depth, lower bound, cyclic convolution.

1 Introduction

This paper concerns the problem of proving high lower bounds of complexity for explicitly given functions. At the present time, we don't know any explicit function (or a multi-output function) which has superlinear complexity in the model of unrestricted Boolean circuits, i.e. we can't prove that a computation of a given function require superlinear number of steps. That's why we consider restricted models of computation.

In this paper, we consider circuits with bounded depth and unbounded fan-in of each gate. Also, we consider several functional systems. In the Boolean case, we allow all Boolean functions as gates. We classify such circuits as medium strength circuits (like formulas over a complete basis). Size of a circuit is defined as the number of edges (i.e. wires) in it.

For every depth d, there are explicit Boolean multi-output functions that require circuits of superlinear size. For depth 2, the first superlinear lower bound was obtained in the paper [2]. The best known lower bound before our series of papers was $\Omega(\frac{n \log^2 n}{\log\log n})$ [9]. In the paper [11] we prove a lower bound $\Omega(n^{1.5})$.

Recall that depth-2 circuits are interesting because of Valiant's reduction [1]. Namely, a lower bound $\omega(\frac{n^2}{\log\log n})$ for depth-2 model implies superlinear lower bound for log-depth model. Note that the upper bound for any n-output function of n inputs is n^2.

For depth 3, the first (and also the best known before our papers) superlinear lower bound $\Omega(n \log\log n)$ was obtained in the paper [6]. In the paper [13] we prove a lower bound $\Omega(n \log n)$.

For even $d \geqslant 4$, the first and the best known lower bound $\Omega(n\lambda_d(n))$ was proved in [3]. Here function $\lambda_d(n)$ is slowly growing. We improve this bound. Namely, we obtain a bound $\Omega(n\lambda_{d-1}(n))$ for any $d \geqslant 2$. Our bound generalizes our previous bounds because $\lambda_1(n) = \Theta(n^{1.5})$ and $\lambda_2(n) = \Theta(n \log n)$. Our bound is also the best known for any even $d \geqslant 4$. In particular, for $d = 4$ our bound is $\Omega(n \log\log n)$.

Note, that for odd $d \geqslant 5$, the first and still the best known lower bound $\Omega(n\lambda_d(n))$ was proved in [5]. Our result doesn't improve this bound because $\lambda_{d-1}(n) = \Theta(\lambda_d(n))$ for any odd $d \geqslant 5$. Other lower bounds were obtained in papers [4, 7, 10, 14]. They does not exceed the best known bounds, but they applies to simplier or more interesting functions such as shift function and matrix multiplication.

2 Proof methods

Our lower bound is valid for cyclic convolution over an arbitrary field K. One can split our proof into two parts (Lemmas 1 and 2 respectively). In the first part we use a complexity-theoretic technique for the depth-2 circuits. We have introduced this technique in the paper [11]. Then, slight improves were done in papers [13, 14].

Here is a "sketch" of Lemma 1. Let I be a subset of inputs and J be a subset of outputs. Suppose there are a lot of connections between variables from I and functions computed at J. Precisely, we substitute variables from I by constants and count entropy of the set of all such subfunctions computed at J. Let this entropy be high (for the cyclic convolution, an entropy is high for many pairs (I, J)). Let V be a subset of nodes such that any output from J is computed using only nodes V. Then, there must be either many nodes in V (hence, many paths between sets J and V), or many paths between sets I and V.

The entropy of a multi-output function was introduced in the paper [14]. We consider a similar notion of entropy, however, we use an axiomatic definition of entropy (as in our paper [12]). This approach allows us to deal with many models of computations: both finite functions and arithmetic functions over an infinite field. For arithmetic circuits over an infinite field, there are higher lower bounds (see [10]) than for the Boolean circuits. So, our bounds are not the best known.

In the second part of our proof we use the graph-theoretic lemma from [10] (papers [3, 5] contain a similar technique). This lemma allows us to reduce a circuit from depth d to depth 2. Precisely, we reduce the bottom part of circuit (which has depth $d - 1$) to circuit of depth 1. After this reduction we apply our Lemma 1.

Note, that previous authors, using the same graph-theoretic technique, have proved weaker bounds. This is because they only considered the functions computed at the nodes, not their subfunctions when counting the size of information transferred from I to J. This approach leads to the well-known superconcentration property of graph [1–3, 5, 6, 9].

3 Functional systems and circuits

Let K be a set (may be, with an algebraic structure) and let $\mathcal{F}$ consists of some functions of a form $K^n \to K$. We say that $\mathcal{F}$ is a *functional system* if $\mathcal{F}$ is closed with respect to superposition and variable substitution. In other words, for any functions f, g from $\mathcal{F}$ (where f has n variables and g has k variables) and any indexes $i_1, \dots, i_{n+k-1}$, the function

$$f(g(x_{i_1}, \dots, x_{i_k}), x_{i_{k+1}}, \dots, x_{i_{n+k-1}})$$

must be in $\mathcal{F}$.

We consider the following functional systems:

(I) K is a finite set, $|K| \geqslant 2$ and $\mathcal{F}$ consists of all functions of a form $K^n \to K$;

(II) K is a field and $\mathcal{F}$ consists of all linear functions $K^n \to K$;

(III) K is an infinite field and $\mathcal{F}$ consists of all multi-variable polinomials over K.

Note, that if field K is infinite and two polinomials are equal at each point, then coefficients of these polinomials are equal too. For a finite field, it is not true (for example, $x \equiv x^2$ over $GF(2)$). That's why we allow only infinite fields in the system (III).

Now we are going to define the notion of a *circuit* over a functional system $\mathcal{F}$. Let us consider a finite directed acyclic graph. A node is called an *input* if it has no ingoing edges; a node is called an output if it has no outgoing edges. Let our graph has n inputs, identifying with variables $x_1, \dots, x_n$, and m outputs, identifying with variables $z_1, \dots, z_m$. Let to each non-input node v be assigned a function $g_v \in \mathcal{F}$, and let the ingoing edges of the node v are identifying with the arguments of g_v. Then the object constructed above is called a circuit over $\mathcal{F}$.

The *size* of a circuit is the number of edges in it; the *depth* of a circuit is the maximal length of directed path in it. In this paper (except for Lemma 1) we assume that depth of a circuit is at most d. The set of nodes in a circuit can be partitioned into *levels*. We number level by $0, 1, \dots, d$. For instance, all inputs belong to level 0. Without loss of generality we may assume that all outputs belong to level d and every edge goes from a level i to the level $i+1$ for some i.

For each node v there is a function f_v: $K^n \to K$ *computed at the node* v; f_v depends on input variables $x_1, \dots, x_n$. One can simply define the function f_v by induction. Note, that f_v is a superposition of functions $g_{v'}$; thus, $f_v \in \mathcal{F}$. Consider the m-output function $F = (f_{z_1}, \dots, f_{z_m})$, where $z_1, \dots, z_m$ are outputs of the circuit. We say that the function F is *computed by the circuit.*

Suppose a field K. If K is finite, then let $\mathcal{F}$ be the functional system (I); if K is infinite, then let $\mathcal{F}$ be the system (III). We define, for each integer n, an n-output function H_n: $K^{2n} \to K^n$, named *cyclic convolution.* Let $H_n = (h_1, \dots, h_n)$ and variables of each h_j are called $x_1, \dots, x_n,\ y_1, \dots, y_n$. By definition, put

$$h_j(x_1, \dots, x_n,\ y_1, \dots, y_n) = x_1 y_j + x_2 y_{j+1} + \ldots + x_n y_{j-1}.$$

The main result of this paper (Theorem 1) is the following. For every $d \geqslant 2$ and every field K, any depth-d circuit for computing the cyclic convolution has at least $\Omega(n\lambda_{d-1}(n))$ edges, where the function $\lambda_d(n)$ is defined in section 6.

4 Expressibility and entropy

Assume that $f, f_1, \dots, f_m$ are functions from a functional system $\mathcal{F}$ and they depend on variables $\tilde{y} = (y_1, \dots, y_n)$. We say that f is *expressible* through $f_1, \dots, f_m$ if for some function $\Phi \in \mathcal{F}$ of m variables we have

$$f(\tilde{y}) \equiv \Phi(f_1(\tilde{y}), \dots, f_m(\tilde{y})).$$

Let $\mathcal{E}(\cdot)$ be a nonnegative-valued functional defined on each finite set of functions $\{f_1, \dots, f_m\} \subseteq \mathcal{F}$. The functional $\mathcal{E}(\cdot)$ is called an *entropy* if the following conditions hold:

(a) the entropy of any single function is at most 1;

(b) the entropy of a set consisting of k independent variables (from the set $y_1, \dots, y_n$) equals k;

(c) if any function from a set F is expressible through a set G, then $\mathcal{E}(F) \leqslant \mathcal{E}(G)$;

(d) *subadditivity*: for any sets F and G

$$\mathcal{E}(F) + \mathcal{E}(G) \geqslant \mathcal{E}(F \cup G).$$

Now we define the entropy for each of our functional systems.

System (II). The entropy is the rank of a set of linear functions. It is clear that conditions (a)–(d) hold.

System (III). We define the entropy as the rank of the set of linear parts of given polinomials. Then conditions (a), (b) and (d) follow from matching conditions for the system (II). Condition (c) holds because the field K is infinite. Indeed, a linear part of product of two polinomials is linearly expressible through linear parts of these polinomials. Hence, if a polinomial f is expressible through polinomials $f_1, \dots, f_m$ then the linear part of f is expressible (in the sense of system (II)) through linear parts of $f_1, \dots, f_m$. Thus, the condition (c) follows from the matching condition for the system (II).

System (I). For a given set of functions $\{f_1, \dots, f_m\}$, let us consider the following equivalence relation on the set K^n. Two points from K^n are equivalent iff any function f_i takes equal values at these points. Let N be the number of equivalence classes for this relation. By definition, put

$$\mathcal{E}(\{f_1, \dots, f_m\}) = \log_{|K|} N. \qquad (1)$$

This entropy functional was used in the paper [14].

Since every function $K^n \to K$ takes at most $|K|$ different values, it follows that condition (a) holds. Condition (b) holds because an ordered set of k different variables takes all $|K|^k$ possible values. If every function from F is expressible through G, then any equivalence class for F consists of some equivalence classes for G. Hence, condition (c) holds. Finally, if there are N equivalence classes for F and M equivalence classes for G, then there are at most NM equivalence classes for $F \cup G$. Thus, condition (d) holds.

5 The complexity-theoretic lemma

The following lemma is the main complexity-theoretic ingredient in our result. The same technique was introduced in the paper [11]. Then, some slight improvements were done in papers [13, 14]. In this paper we make once more slight improvement.

We assume that a functional system $\mathcal{F}$ contains constants 0 and 1. Consider a circuit over the system $\mathcal{F}$ computing a multi-output function F: $K^{2n} \to K^n$ of variables $x_1, \ldots, x_n,\ y_1, \ldots, y_n$. Let $F = (f_1, \ldots, f_n)$. Let f_j^i denote the function obtained by substituting 1 for x_i and zeroes for the remaining variables $x_1, \ldots, x_n$ in f_j. Let I be a subset of inputs $x_1, \ldots, x_n$ and let J be a subset of outputs $z_1, \ldots, z_n$. For any node v, denote by $I(v)$ the set of all inputs from I such that there is a directed path from this input to the node v.

Lemma 1. *Let V be a subset of nodes such that any directed path from any input to the set J passes through a node from V. Then for any entropy functional $\mathcal{E}(\cdot)$, we have*

$$\sum_{v \in V} (|I(v)| + 1) \geqslant \mathcal{E}(\{f_j^i \mid x_i \in I,\ z_j \in J\}).$$

Proof. Recall, that a function f_v is computed at a node v, and f_v depends on variables $x_1, \ldots, x_n,\ y_1, \ldots, y_n$. Let f_v^i denote the function obtained by substituting 1 for x_i and zeroes for the remaining variables $x_1, \ldots, x_n$ in f_v; let f_v^0 denote the function obtained by substituting zeroes for all variables $x_1, \ldots, x_n$ in f_v.

Since any directed path from any input to the set J passes through the set V, we have that any function $f_j,\ z_j \in J$, is expressible through the functions $f_v,\ v \in V$. Then for any i, the function f_j^i is expressible through the functions $f_v^i,\ v \in V$. Hence, it follows from the entropy condition (c) that

$$\mathcal{E}(\{f_j^i \mid x_i \in I,\ z_j \in J\}) \leqslant \mathcal{E}(\{f_v^i \mid x_i \in I,\ v \in V\}). \tag{2}$$

Conditions (d) and (a) yield

$$\mathcal{E}(\{f_v^i \mid x_i \in I,\ v \in V\}) \leqslant \sum_{v \in V} \mathcal{E}(\{f_v^i \mid x_i \in I\}) \leqslant \sum_{v \in V} |\{f_v^i \mid x_i \in I\}|. \tag{3}$$

Note that if $x_i \notin I(v)$, then a function f_v does not depend on a variable x_i, and hence $f_v^i = f_v^0$. Therefore

$$\{f_v^i \mid x_i \in I\} = \{f_v^i \mid x_i \in I(v)\} \cup \{f_v^0\}.$$

Hence

$$|\{f_v^i \mid x_i \in I\}| \leqslant |I(v)| + 1. \tag{4}$$

The lemma follows from (2)–(4). □

Now we derive the corollary for the cyclic convolution. Consider a field K, the corresponding functional system $\mathcal{F}$ and any circuit over $\mathcal{F}$ for computing the cyclic convolution $H_n = (h_1, \ldots, h_n)$. Let k, l are positive integers and $kl \leqslant n$.

Let I_p^k denote the following subset of inputs $x_1, \ldots, x_n$: it begins with x_p and consists of k inputs one after the other, i.e. $I_p^k = \{x_p, x_{p+1}, \ldots, x_{p+k-1}\}$. We assume that the order of inputs $x_1, \ldots, x_n$ is cyclic, i.e. x_n is followed by x_1. Let $J_q^{k,l}$ denote the following subset of outputs: it begins with z_q and consists of l outputs one after the other with the step k, i.e. $J_q^{k,l} = \{z_q, z_{q+k}, \ldots, z_{q+kl-k}\}$ (the order is cyclic too).

Recall that $h_j^i = y_{i+j-1}$. Note that $i + j - 1$ takes kl subsequent values as i ranges over $p, p+1, \ldots, p+k-1$ and j over $q, q+k, \ldots, q+kl-k$. Hence the set $\{h_j^i \mid x_i \in I_p^k,\ z_j \in J_q^{k,l}\}$ consists of kl independent variables. Thus, the entropy condition (b) implies

$$\mathcal{E}(\{h_j^i \mid x_i \in I_p^k,\ z_j \in J_q^{k,l}\}) = kl.$$

Combining this with Lemma 1 we get the following.

Corollary 1. *Let V be a subset of nodes such that any directed path from any input to the set $J_q^{k,l}$ passes through a node from V. Then*

$$\sum_{v \in V} (|I_p^k(v)| + 1) \geqslant kl.$$

6 Slowly growing functions and the graph technique

This section contains a material (including definitions and claims) taken from the paper [10]. The similar technique was used in papers [3, 5].

Let a function f takes each natural number to a nonnegative integer, and for any $n \geqslant 2$, $f(n) < n$. Let $f^{(k)}$ denote the k-th degree of f under the composition, i.e. $f^{(k)} = f \circ f \circ \ldots \circ f$, where f is repeated k times. We define a function f^* as follows:

$$f^*(n) = \min\{k \mid f^{(k)}(n) \leqslant 1\}.$$

Now we define functions $\lambda_d(n)$:

$$\lambda_1(n) = \lfloor \sqrt{n} \rfloor, \quad \lambda_2(n) = \lceil \log_2 n \rceil, \quad \lambda_d(n) = \lambda_{d-2}^*(n),\ d = 3, 4, \ldots$$

The following claim contains properties of functions $\lambda_d(n)$. It implies that our bounds $\Omega(n\lambda_{d-1}(n))$ are superlinear (item 1), the bound for depth 4 is $\Omega(n \log\log n)$ (item 2), and for even $d \geqslant 4$, our bound is better than the previous bound $\Omega(n\lambda_d(n))$ (item 3).

Claim. 1) For any d, $\lambda_d(n)$ is a monotone increasing function tending to infinity on $n \to \infty$;

2) $\lambda_3(n) = \Theta(\log\log n)$;

3) if d is even or $d = 3$, then $\lambda_{d-1}(n) = \Omega(\lambda_d(n))$;

4) if d is odd and $d \geqslant 5$, then $\lambda_{d-1}(n) = \Theta(\lambda_d(n))$.

Proof. Items 1, 2 and 4 were proved in ([10], Claim 2.4). Item 3 is obvious for $d = 2, 3$. Let d is even and $d \geqslant 4$. We claim that if f is increasing function tending to infinity and $f(n) = o(n)$, then $f^*(n) = o(f(n))$. Indeed, $f(n) = o(n)$ and $f(n) \to \infty$ implies $f^{(2)}(n) = o(f(n))$. Moreover, $f^*(n) \leqslant f^{(2)}(n)+1$ because each iteration of f decreases the number by at least 1. Thus, $f^*(n) = o(f(n))$. In particular, for $d \geqslant 6$, $\lambda_{d-1}(n) = \Theta(\lambda_{d-2}(n)) = \Omega(\lambda_d(n))$. For $d = 4$, the proof is similar but it uses relation $f^*(n) = o(f^{(2)}(n))$ because $\lambda_3(n) = \Theta(\lambda_2^{(2)}(n))$. □

The following lemma is a graph-theoretic ingredient of out result. It says that if a depth-d graph has less than $\Omega(n\lambda_d(n))$ edges, then one can remove small sets of it's inputs, outputs and intermediate nodes so that there remains a little number of paths between inputs and outputs. This lemma helps us to reduce a depth-d graph to a depth-1 graph: paths mentioned above become an edges in the depth-1 graph.

Lemma 2. ([10], Lemma 1.1) *If $0 < \varepsilon < 1/400$ and if a depth-d graph consists of more than n nodes and less than $\varepsilon n\lambda_d(n)$ edges, then there are subsets I, J, W in sets of inputs, outputs and all nodes of the graph (respectively) such that*

a) $|I| \leqslant 5\varepsilon dn$, $|J| \leqslant 5\varepsilon dn$, $\sqrt{n} \leqslant |W| = o(n)$;

b) the number of directed paths from inputs to outputs which does not pass through the set $I \cup J \cup W$ is at most $\varepsilon n^2/|W|$.

7 The main result

Theorem 1. *If $d \geqslant 2$, K is an arbitrary field and $\mathcal{F}$ is the corresponding functional system of the type (I) or (III), then any depth-d circuit over $\mathcal{F}$ for computing the cyclic convolution H_n has $\Omega(n\lambda_{d-1}(n))$ edges.*

Proof. Assume the converse. Let L be the number of edges in the circuit. Then there exists ε such that $0 < \varepsilon < 1/400$ and $L < \varepsilon n\lambda_{d-1}(n)$. Let G be the graph consisting of all edges which are not outgoing edges of inputs, i.e. edges of second, third etc levels. The graph G has depth $d-1$. Applying Lemma 2 to the graph G, we find sets I, J and W. Denote $l = 4|W|$ and $k = \lfloor n/l \rfloor$. Then $kl \leqslant n$, and the restriction $\sqrt{n} \leqslant |W| = o(n)$ yields $kl = n(1 - o(1))$.

Here is a "sketch" of the following proof. For particular p and q, we consider the set of inputs I_p^k and the set of outputs $J_q^{k,l}$ of the original circuit (see section 5 for their definition). For applying Lemma 1, we need to define the subset V of nodes such that any path from any input to the set $J_q^{k,l}$ passes through V. We shall say that any path from any input to the set $J_q^{k,l}$ is "bad". So, we need to cut off all bad paths. The set $I \cup J \cup W$ cuts off a lot of bad paths; we include this set to the set V (note that we only need to include the set $J \cap J_q^{k,l}$ instead of J). By Lemma 2, the number of the remaining paths in the graph G is small. So, we can cut the remaining bad paths by a small subset V_1' of the first-level nodes (which are the inputs for the graph G).

By Lemma 1, the sum of $|I_p^k(v)|+1$, where v ranges over V, must be at least kl i.e. $n(1-o(1))$. This contradicts the following estimate of the sum. Recall that the size of the set $I_p^k(v)$ is at most k (because $I_p^k(v)$ is a subset of I_p^k). Since the set J is of size $\varepsilon' n$ for some small ε', then for particular q, the set $J \cap J_q^{k,l}$ is of size $\varepsilon' l$. By definition of l, the set W is of size $l/4$. Hence, the sum of $|I_p^k(v)|+1$, where v ranges over $(J \cap J_q^{k,l}) \cup W$, is smaller than n. It remains to show that this sum, where v ranges over $I \cup V_1'$, is smaller than n too. For particular p, the sum of $|I_p^k(v)|$ over the set $I \cup V_1'$ is small because the number of edges on the first level is small (note that the set $I \cup V_1'$ consists of first-level nodes). And the sum of 1's over the set $I \cup V_1'$ (i.e. the size of this set) is small by Lemma 2.

Now we define p and q more precisely. Let p be an index such that the total number of outgoing edges for the set I_p^k is minimal. Since the total number of outgoing edges for the set of all inputs does not exceed L, it follows that (for the choosed p) the total number of outgoing edges for I_p^k is at most kL/n. Indeed, summing all these total numbers over all p we obtain at most kL (because we count each input k times). Thus, the average value is kL/n, and the minimal value not exceeds the average value.

Denote the set of all first-level nodes by V_1. Recall that for each $v \in V_1$, the set $I_p^k(v)$ consists of all inputs from I_p^k connected with the node v by a directed path. Since v is at the first level, any such path consists of one edge. Hence, summing sizes of sets $I_p^k(v)$ over $v \in V_1$, we obtain the number of outgoing edges for the set I_p^k. Thus,

$$\sum_{v \in V_1} |I_p^k(v)| \leqslant \frac{kL}{n} < \frac{k \cdot \varepsilon n \lambda_{d-1}(n)}{n} = \varepsilon k \lambda_{d-1}(n).$$

Note that $\lambda_{d-1}(n) \leqslant \sqrt{n}$ (for any d), and since $l/4 = |W| \geqslant \sqrt{n}$, it follows that $k \leqslant \frac{1}{4}\sqrt{n}(1+o(1))$. Therefore,

$$\sum_{v \in V_1} |I_p^k(v)| \leqslant \varepsilon\sqrt{n} \cdot \frac{1}{4}\sqrt{n}(1+o(1)) = 0.25\varepsilon n(1+o(1)). \tag{5}$$

Let q be an index such that both following conditions hold:

a') $|J_q^{k,l} \cap J| \leqslant \frac{2l}{n} \cdot 5\varepsilon d n$;

b') the number of directed paths from the set V_1 to the set $J_q^{k,l}$ which does not pass through the set $I \cup J \cup W$, is at most $\frac{2l}{n} \cdot \varepsilon n^2/|W|$.

Since $|J| \leqslant 5\varepsilon d n$, it follows that the part of sets $J_q^{k,l}$ not satisfying the condition a') is less than 1/2 (the proof is similar to the above one where we choose p). Since the number of paths from V_1 to the set of all outputs is at most $\varepsilon n^2/|W|$, it follows that the part of sets $J_q^{k,l}$ not satisfying the condition b') is less than 1/2. Hence, there are the set $J_q^{k,l}$ satisfying both conditions.

Let V_1' be the set of all first-level nodes connected with the set $J_q^{k,l}$ by directed paths which does not pass through the set $I \cup J \cup W$ (the set V_1' cuts off the remaining "bad" paths). Since the number of paths does not exceed the number of their starting points, we have

$$|V_1'| \leqslant \frac{2l}{n} \cdot \frac{\varepsilon n^2}{|W|} = \frac{2\varepsilon ln}{l/4} = 8\varepsilon n. \tag{6}$$

Finally, applying Lemma 1 to the set $V = V_1' \cup I \cup (J_q^{k,l} \cap J) \cup W$, we obtain

$$kl \leqslant \sum_{v \in V} (|I_p^k(v)|+1) = \sum_{v \in V_1' \cup I} |I_p^k(v)| + |V_1'| + |I| + \sum_{v \in (J \cap J_q^{k,l}) \cup W} (|I_p^k(v)|+1). \tag{7}$$

Let us estimate each of the four summands at the right hand of (7). Since $V_1' \cap I \subseteq V_1$, the first summand is majorized by (5). Estimates for the second and third summands are (6) and item a) of Lemma 2 respectively. Let us estimate the fourth summand. Using inequality $|I_p^k(v)| \leqslant k$ and the condition a'), we have

$$\sum_{v \in (J \cap J_q^{k,l}) \cup W} (|I_p^k(v)| + 1) \leqslant (k+1)(|J \cap J_q^{k,l}| + |W|) \leqslant$$

$$2k\Big(2\frac{l}{n} \cdot 5\varepsilon dn + \frac{l}{4}\Big) = (0.5 + 20\varepsilon d)kl.$$

Thus, (7) implies

$$kl \leqslant 0.25\varepsilon n(1 + o(1)) + 8\varepsilon n + 5\varepsilon dn + (0.5 + 20\varepsilon d)kl.$$

Since ε is small, we have a contradiction. □

8 Conclusion

Note 1. Since our bound is uniform for all depths (specifically, $\Omega(n\lambda_{d-1}(n))$), then one can raise the following question. Is it the limit of our capacities or can our bound be improved at least for a particular d? We suppose that if it is possible to derive a bound $\Omega(n^{1.5+\varepsilon})$ for depth-2 circuits, then one can obtain a superlinear bound for log-depth circuits using a similar technique.

Here is the explanation. Let us consider a depth-2 circuit. Denote the number of edges in the first level by L_1 and the number of edges in the second level by L_2. We have an observation (see Corollary 2 in [13]) that the complexity measure $\sqrt{L_1 L_2}$ is more representative (for depth-2 circuits) than $L_1 + L_2$. But, for Valiant's circuits of depth 2, we have $L_1 = O(\frac{n^2}{\log\log n})$ and $L_2 = o(n^{1+\varepsilon})$, hence $\sqrt{L_1 L_2} = o(n^{1.5+\varepsilon})$. So, a lower bound $\Omega(n^{1.5+\varepsilon})$ seems to be interesting.

Note 2. For the functional system (I), one can use another entropy functional, namely, the Shannon entropy. Recall the equivalence relation on the set K^n (see the paragraph before (1)). Let $K_1, \ldots, K_N$ are the equivalence classes of this relation, and $p_i = |K_i|/|K|^n$. By definition, put

$$\mathcal{E}'(\{f_1, \ldots, f_m\}) = \sum_{i=1}^{N} p_i \log_{|K|} \frac{1}{p_i}.$$

We use the entropy $\mathcal{E}'(\cdot)$ instead of $\mathcal{E}(\cdot)$ in the paper [12] because the first holds the strong subadditivity: $\mathcal{E}'(F)+\mathcal{E}'(G) \geqslant \mathcal{E}'(F\cup G)+\mathcal{E}'(F\cap G)$. For the entropy $\mathcal{E}'(\cdot)$, there are inequalities which does not follow from the strong subadditivity; an example of such inequality was given in the paper [8]. So, one can try to improve our bound using the strong subadditivity or these stronger inequalities.

Acknowledgements. The author is grateful to Stasys Jukna for useful discussions. The work is partially supported by President RF grant for leading scientific school support NSh–4470.2008.1.

References

1. Valiant L. G.: Graph-theoretic arguments in low-level complexity. LNCS **53** (1977) 162–176
2. Pippenger N.: Superconcentrators of depth 2. J. of Computer and System Sciences **24** (1982) 82–90
3. Dolev D., Dwork C., Pippenger N., Wigderson A.: Superconcentrators, generalizers and generalized connectors with limited depth. Proc. 15th ACM STOC (1983) 42–51
4. Pudlák P., Savický P.: On shifting networks. Theoretical Comput. Sci. **116** (1993) 415–419
5. Pudlák P.: Communication in Bounded Depth Circuits. Combinatorica **14:2** (1994) 203–216
6. Alon N., Pudlák P.: Superconcentrators of depth 2 and 3; odd levels help (rarely). J. of Computer and System Sciences **48** (1994) 194–202
7. Pudlák P., Rödl V., Sgall J.: Boolean circuits, tensor ranks and communication complexity. SIAM J. on Computing **26:3** (1997) 605–633
8. Zhang Z., Yeung R. W.: On characterization of entropy function via information inequalities. IEEE Transactions on Information Theory. **44** (1998) 1440–1450
9. Radhakrishnan J., Ta-Shma A.: Bounds for dispersers, extractors and depth-two superconcentrators. SIAM J. of Discrete Mathematics **13:1** (2000) 2–24
10. Raz R., Shpilka A.: Lower Bounds for Matrix Product, in Bounded Depth Circuits with Arbitrary Gates. SIAM J. Comput. **32:2** (2003) 488–513
11. Cherukhin D. Yu.: The lower estimate of complexity in the class of schemes of depth 2 without restrictions on a basis. Moscow Univ. Math. Bull. **60:4** (2005) 42–44
12. Cherukhin D. Yu.: On complexity of informational networks. Manuscript, available at http://cherukhin.narod.ru/research/research.htm (2007), in Russian
13. Cherukhin D. Yu.: Lower Bounds for Depth-2 and Depth-3 Boolean Circuits with Arbitrary Gates. Proc. 3rd Int. Comput Sci Symposium in Russia (CSR-2008), to appear.
14. Jukna S.: Entropy of operators or why matrix multiplication is hard for small depth circuits. Electronic Colloquium on Computational Complexity, Report Nr. TR08-019, 2008