Better Bounds for Threshold Formulas

Jaikumar Radhakrishnan
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

Abstract

We consider the computation of threshold functions using formulas over the basis {AND, OR, NOT}.

We show that every monotone formula that computes the threshold function T_k^n, $2 \le k \le \frac{n}{2}$, has size $\Omega(nk\log(\frac{n}{k-1}))$. The same lower bound is shown to hold even in the stronger monotone contact networks model.

We also show nearly optimal bounds on the size of $\Sigma\Pi\Sigma$ formulas computing T_k^n for small k.

1 Introduction

The threshold function T_k^n is the Boolean function on n inputs that takes the value 1 exactly when at least k of the input variables have value 1. Threshold functions play a central role in the investigation of the computational complexity of Boolean functions (see Boppana and Sipser [4], Wegener [22]). Their complexity has been studied in various circuit models. In this paper, we consider the computation of threshold functions using formulas and contact networks.

Our main result is the following. We show that every *monotone* formula computing T_k^n, $2 \le k \le \frac{n}{2}$, has size $\Omega(kn\log(\frac{n}{k-1}))$. The same lower bound is shown to hold for of the stronger monotone contact networks model. This result improves the earlier $\Omega(n\log n)$ lower bound due to Krichevskii [11] and Hansel [8], and the $\Omega(kn)$ lower bound due to Khrapchenko [10].

The complexity of computing T_k^n, for large thresholds, using constant depth circuits is well understood [4]. However, for small values of k these results do not imply superlinear lower bounds. Indeed, it has been shown by Newman, Ragde, and Wigderson [14] that for small values of k (bounded by a function of the form $(\log n)^r$ for some constant r), there do exist linear size constant depth circuits computing T_k^n. In contrast, formulas even with unbounded depth need size $\Omega(n\log n)$ to compute T_2^n. To better understand the computation of T_k^n by constant depth formulas, we consider $\Sigma\Pi\Sigma$ formulas computing T_k^n. A $\Sigma\Pi\Sigma$ formula has the form $\bigvee_{i=1}^{p}\bigwedge_{j=1}^{t_i}\bigvee_{q\in S_{ij}} q$, where each S_{ij} is a subset of variables and their negations. Our results are the following.

1. Suppose $k \le \frac{n}{2}$. Then, every $\Sigma\Pi\Sigma$ formula computing T_k^n has size $\exp(\Omega(\sqrt{k}))n$.
2. Suppose k and n are large and $k \le \sqrt{n}$. Then, every $\Sigma\Pi\Sigma$ formula computing T_k^n has size at least
$$\exp(\delta(k))n\log n,$$
where $\delta(k) = \frac{1}{50}\sqrt{\frac{k}{\ln k}}$. This is an improvement of the $\Omega(kn\log(\frac{n}{k-1}))$ lower bound due to Newman, Ragde, and Wigderson [14].
3. Suppose k and n are large and $k \le \sqrt{n}$. Then, there exist $\Sigma\Pi\Sigma$ formulas computing T_k^n with size at most
$$\exp(2\sqrt{k}\ln k)n\log n.$$

Note that the second result is stronger than the first only for small values of k, when the $\log n$ factor becomes significant in comparison to the $\exp(\sqrt{k})$ factor. In this paper we shall discuss the results 2 and 3 only; a proof for result 1 is presented in [18].

1.1 Relation to previous work

Over the basis {AND, OR, NOT}, Paterson, Pippenger, and Zwick [15] show that T_k^n can be computed using formulas of size $O(n^{4.85})$. The existence of polynomial size *monotone* formulas for computing T_k^n is implied by the $O(\log n)$ depth sorting network due to Ajtai, Komlós, and Szemerédi [1]. The existence of more efficient monotone threshold formulas was shown by Valiant [21] and Boppana [2]. Valiant showed that the majority function ($T_{\lfloor n/2\rfloor}^n$) can be computed using monotone formulas of size $O(n^{5.3})$. Boppana generalized Valiant's result and showed that T_k^n can be

CH3062-7/91/0000/0314$01.00 © 1991 IEEE

computed by monotone formulas of size $O(k^{4.3}n\log n)$. This is the best upper bound known for computing T_k^n using monotone formulas.

For the basis {AND, OR, NOT}, Khrapchenko [10] showed that any formula computing T_k^n has size $k(n-k+1)$. Hansel [8] and Krichevskii [11] (see also [14, 16]) showed that any formula computing T_2^n has size $\Omega(n\log n)$. For the complete binary basis, Pudlák [17] showed that any formula computing T_k^n, $2 \leq k \leq n-1$, has size $\Omega(n\log\log n)$; Fischer, Meyer, and Paterson [7] showed that the majority function needs formula size $\Omega(n\log n)$.

For monotone formulas, the only lower bounds known were those implied by the results of Hansel, Krichevskii, and Khrapchenko, stated above, which hold even if the formula contains negations. Snir [20] showed an $\Omega(kn\log n)$ lower bound for a problem arising in the context of hypergraph covering. This result and its connection to $\Sigma\Pi\Sigma$ threshold formulas are discussed below. However, it is not clear how Snir's result may be used to derive our results for monotone threshold formulas.

Related to the monotone formula model is the model of the *monotone contact networks*. This model is similar to the model of monotone nondeterministic branching programs. This model was considered by Lupanov [12], who showed that any monotone contact network computing T_2^n has size $\Omega(\log n/\log\log n)$. Krichevskii [11] improved this lower bound to $\Omega(n\log n)$. It follows from the results of Moore and Shannon [13] that any monotone contact network (even if constant 1 is allowed) that computes T_k^n must have size $k(n-k+1)$. Recently Razborov [19] showed that any contact network (where negations are allowed) that computes the majority function must have size $\Omega(n\log\log\log^* n)$.

A different approach to improving the known lower bounds for formulas computing T_k^n was chosen by Newman, Ragde, and Wigderson [14]. They considered following the method used by Hansel and Krichevskii for T_2^n. Hansel and Krichevskii obtained their lower bounds in two stages. First, they showed that any formula computing T_2^n may be converted to a $\Sigma\Pi\Sigma$ formula without increasing the size. Next, they showed that any $\Sigma\Pi\Sigma$ formula computing T_2^n has size $\Omega(n\log n)$. Since there exist formulas computing T_2^n of size $O(n\log n)$, this lower bound is the best possible.

To generalize the result of Hansel and Krichevskii, Newman, Ragde, and Wigderson studied $\Sigma\Pi\Sigma$ formulas computing T_k^n for bigger values of k. They showed, under the assumption that the fanin for the AND's is restricted to k, that any $\Sigma\Pi\Sigma$ formula computing T_k^n, for $2 \leq k \leq n$, has size at least $n^{\frac{\log n - \log(k-1)}{\log k - \log(k-1)}}$.

Under their assumption, the problem is equivalent to the problem of covering the complete k-uniform hypergraph using k-partite hypergraphs. In this setting, the problem was studied earlier by Snir [20] who obtained identical lower bounds. Using a random family of k-partite hypergraphs one may obtain $\Sigma\Pi\Sigma$ formulas of size $O(\sqrt{k}\exp(k)n\log n)$. It was shown by Radhakrishnan [18] that the results of Snir can be improved using the technique of Körner [9], and Fredman and Komlós [6]. This implies that any $\Sigma\Pi\Sigma$ formula computing T_k^n, with the restriction that the fanin for the AND's be k, has size $\Omega(\frac{\exp(k)}{k\sqrt{k}}n\log n)$. Thus there exist essentially tight bounds on the size of restricted $\Sigma\Pi\Sigma$ formulas computing T_k^n.

1.2 Overview

In this paper, we use an information theoretic inequality due to Fredman and Komlós to improve the known lower bounds for formulas computing T_k^n. In section 4, we present our results for monotone threshold formulas. In section 5, we present the results on $\Sigma\Pi\Sigma$ formulas. For our lower bound, we do not impose any restrictions on the $\Sigma\Pi\Sigma$ formulas, that is, the fanin of the AND's need not be k and the formula may contain negations. We also show that $\Sigma\Pi\Sigma$ threshold formulas exist whose size is close to the lower bound. These formulas do not use negations.

2 Notation

In the following $[n]$ will denote the set $\{1, 2, \ldots, n\}$; for a set S, $\binom{S}{k}$ will denote the set of all k size subsets of S.

For a function f on n variables, we shall assume that the variables are $x_1, x_2, \ldots, x_n$. We say that f *accepts* $T \subseteq [n]$ if f evaluates to 1 when all the variables $x_i, i \in T$ are given the value 1 and the remaining variables are given the value 0. We often identify a set of variables with the set of indices of those variables. A function f is said to be *l-immune* if it accepts no $T \subseteq [n]$ with $|T| \leq l$. Thus, the threshold function T_k^n is $(k-1)$-immune.

We shall use the standard notation from graph theory. The graphs we consider will usually be undirected and simple. The size of the largest independent set in a graph G will be denoted by $\alpha(G)$; *size*(G) will denote the number of non-isolated vertices in G. A function f with domain $V(G)$ will be called a *coloring* of G if $f(i) \neq f(j)$ whenever $(i,j) \in E(G)$. For a graph G, G^N will denote the subgraph induced by the non-isolated vertices of G.

We shall think of a monotone formula as a rooted binary tree where each internal vertex is either an AND or an OR, and each leaf is a variable.

Definition 2.1 A *monotone contact network* is a directed graph $G = (V, E)$ that has two distinguished vertices s (source) and t (sink). (Multiple edges are permitted to exist in the graph.) The non-distinguished vertices will be called *internal* vertices. Each edge in the network has a label from the set $\{x_1, x_2, \ldots, x_n\}$. A monotone contact network N computes a function f_N from $\{0,1\}^n$ to $\{0,1\}$ as follows. Given an input y in $\{0,1\}^n$ each edge of N is set to 0 or 1 according to the value of its label. Then, $f_N(y) = 1$ if there is a path from s to t using only edges with value 1, and $f_N(y) = 0$ otherwise. The *size* of a contact network is the total number of edges in it.

Definition 2.2 A *depth two contact network* is a contact network where each edge is incident on s or t. Further, s has no edge coming in and t has no edge going out.

We shall say that a contact network *accepts* a set A if the function it computes accepts A. We shall extend this terminology to apply to the vertices of a contact network. For example, we shall say that a vertex p of the contact network N accepts a set A if there is a path of all 1's from p to t on input A. Thus, the contact network accepts precisely those inputs that are accepted by the distinguished vertex s. A vertex is said to be *r-immune* if it does not accept any input of size less than or equal to r. For example, if a contact network computes T_k^n, then the distinguished vertex s is $(k-1)$-immune.

A $\Sigma\Pi\Sigma$ formula has the form $\bigvee_{i=1}^{p} \bigwedge_{j=1}^{t_i} \bigvee_{q \in S_{ij}} q$, where each S_{ij} is a subset of variables and their negations. The size of a formula f, denoted by size(f), is the number of occurrences of variables in it. Thus if f is the $\Sigma\Pi\Sigma$ formula shown above, then $\text{size}(f) = \sum_{i=1}^{p} \sum_{j=1}^{t_i} |S_{ij}|$.

A $\Pi\Sigma$ formula has the form $\bigwedge_{j=1}^{t} \bigvee_{q \in S_j} q$. Here S_j is a subset of variables and their negations.

Let F be a formula on n variables and let $T \subseteq [n]$. We use $F|_T$ to denote the formula obtained from F by fixing the variables appearing in T at 1. We shall, however, continue to think of $F|_T$ as a formula on n variables (only some of the variables do not appear explicitly in its representation).

3 Information theoretic inequalities

For a random variable X, let *support*(X) denote the set of values that X assumes. We shall need the following standard definitions from information theory.

Given a random variable X with finite support, its entropy is given by

$$H(X) = -\sum_{x \in \text{support}(X)} \Pr[X = x] \log \Pr[X = x].$$

If X and Y are random variables then (X, Y) will be the random variable taking values in support$(X) \times$ support(Y) according to the joint distribution of X and Y. For notational convenience we denote the entropy of this random variable by $H(X, Y)$ instead of $H((X, Y))$. The entropy of a function f will be the entropy of the random variable $f(X)$, where X assumes values in the domain of f with uniform distribution.

If X and Y are random variables, then the conditional entropy of Y given X, denoted by $H(Y|X)$, is given by

$$H(Y|X) = H(X, Y) - H(X).$$

For random variables X and Y with some joint distribution and $x \in$ support(X), let Y_x be a random variable such that

$$\Pr[Y_x = y] = \Pr[Y = y | X = x].$$

Then, it can be shown that

$$H(Y|X) = \mathrm{E}(H(Y_X)). \quad (1)$$

The following information theoretic measure on graphs was introduced by Fredman and Komlós [6]. This notion of graph entropy is different from the more sophisticated notion of graph entropy due to Körner [9].

Definition 3.1 (Coloring Entropy)
Let $G = (V, E)$ be a graph. Let f be the coloring of the graph G^N with minimum entropy. The coloring entropy of G is given by

$$H(G) = \frac{\text{size}(G)}{|V(G)|} H(f).$$

(If $E(G)$ is empty, then $H(G) = 0$.)

In the rest of this paper, when we say entropy of a graph G we shall mean the coloring entropy of G. The following lemma is due to Fredman and Komlós [6]. It is a special case of the subadditivity property of Körner's entropy [9]. The proof given here is a generalization of the one due to Pippenger [16].

Lemma 3.2 Let $G, G_1, \ldots, G_l$ be graphs with vertex set $[n]$ such that $G = G_1 \cup G_2 \cup \ldots \cup G_l$. Then

$$\sum_{i=1}^{l} H(G_i) \geq \log(\frac{n}{\alpha(G)}).$$

Proof: Let X be a random variable taking values in $V(G)$ with uniform distribution. Then $H(X) = \log n$.

For $i = 1, \ldots, l$, let f_i be a coloring of the subgraph G_i^N, with minimum entropy, as in the definition of $H(G_i)$. For $i = 1, \ldots, l$, let Y_i be the random variable taking values in the range of f_i, defined by

$$\Pr[Y_i = j] = \frac{|f^{-1}(j)|}{\text{size}(G_i)}.$$

The random variables $Y_1, Y_2, \ldots, Y_l$ are independent of each other and of X.

For $i = 1, \ldots, l$, let the random variable X_i be defined by

$$X_i = \begin{cases} f(X) & \text{if } X \text{ is non-isolated in } G_i; \\ Y_i & \text{otherwise.} \end{cases}$$

Note that $H(Y_i) = H(X_i) = H(f_i)$. Further, observe that for each value of X the random variables $X_1, X_2, \ldots, X_l$ are independent. Now,

$$\begin{aligned} &H(X_1, X_2, \ldots, X_l) - H((X_1, \ldots, X_l)|X) \\ &\quad = H(X) - H(X|(X_1, X_2, \ldots, X_l)). \end{aligned} \tag{2}$$

Since $G = \cup_{i=1}^{l} G_i$, for any value of $(X_1, X_2, \ldots, X_l)$, the random variable X can take at most $\alpha(G)$ different values. It follows that

$$H(X|(X_1, X_2, \ldots, X_l)) \leq \log \alpha(G).$$

Since $X_1, X_2, \ldots, X_l$ are independent for each value of X, we have from (1) that

$$H((X_1, X_2, \ldots, X_l)|X) = \sum_{i=1}^{l} H(X_i|X).$$

Noting that $H(X_1, X_2, \ldots, X_l) \leq \sum_{i=1}^{l} H(X_i)$, we get from (2)

$$\sum_{i=1}^{l}(H(X_i) - H(X_i|X)) \geq \log(\frac{n}{\alpha(G)}).$$

From (1), we have $H(X_i|X) = (1 - \frac{\text{size}(G_i)}{\alpha(G_i)})H(X_i)$. It follows that

$$\sum_{i=1}^{l} H(G_i) \geq \log(\frac{n}{\alpha(G)}).$$

This completes the proof of the lemma. □

Since a bipartite graph has a coloring with entropy at most 1, we get the following corollary to lemma 3.2.

Corollary 3.3 Let $G_1, G_2, \ldots, G_l$ be bipartite graphs with vertex set $[n]$. If $G = G_1 \cup G_2 \cup \ldots \cup G_l$, then

$$\sum_{i=1}^{l} \text{size}(G_i) \geq n \log(\frac{n}{\alpha(G)}). \quad □$$

We shall need the following two results. Let $L(x) = \log((x+1)e)$. The following lemma is from Csiszár and Körner [5].

Lemma 3.4 For a random variable X concentrated on non-negative integers

$$H(X) \leq \text{L}(\text{E}(X)). \quad □$$

The following lemma is due to Boppana [3]

Lemma 3.5 Every graph G with n vertices and m edges has a coloring with entropy at most $L(\frac{m}{n})$. □

4 Monotone formulas

In this section we shall extend the results of Krichevskii and Khrapchenko and show improved lower bounds on the size of monotone formulas computing T_k^n. Let $L_M(T_k^n)$ be the minimum size of a monotone formula computing T_k^n. It is easy to see that $L_M(T_k^n) = L_M(T_{n-k+1})$. The case $k = 1$ is trivial. We shall consider the case when $2 \leq k \leq \frac{n}{2}$.

Note that a monotone formula can be converted to a monotone contact network of the same size by representing the OR's in parallel and the AND's in series. Hence, to show lower bounds on the size of monotone formulas computing T_k^n, it suffices to show lower bounds on the size of contact networks computing T_k^n.

The following lemma is implicit in the work of Krichevskii [11].

Lemma 4.1 *Every 1-immune monotone contact network N can be converted to a depth two contact network $\hat{N}$ such that*

1. *$\hat{N}$ is 1-immune.*
2. *Size of $\hat{N}$ is at most the size of N.*
3. *Every input accepted by N is accepted by $\hat{N}$.*

Proof: We first convert N to a contact network N' that accepts exactly the same set of inputs as N and has size at most the size of N. Then we convert N' to a depth two contact network $\hat{N}$.

Let V_1 be the set of vertices in N that are reachable from s using only those edges that have label x_1.

Delete all edges incident on vertices in V_1 with label x_1. Add new edges connecting s to each vertex in V_1. Label the new edges with label x_1.

We repeat this procedure for the other labels $x_2, \ldots, x_n$. The final network thus obtained is N'. At each stage the new network accepts exactly the same set of inputs as the old network. In particular, N' is 1-immune. Moreover, the number of new edges added is at most the number of edges deleted.

The network N' has the following property. If a vertex v has an edge from s with label x_i then no edge leaving v has label x_i. Let v be an internal vertex of N'. Let A_v be the set of labels on the edges (s, v) and B_v be the labels on the edges (v, w) leaving v. Then A_v and B_v are disjoint sets. Further, the size of N' is at least $\sum_v |A_v| + |B_v|$.

The network $\hat{N}$ is constructed as follows. The set of vertices for $\hat{N}$ is the same as the set of vertices for N'. For each internal vertex v add $|A_v|$ edges of the form (s, v), one for each label in A_v. Similarly, add $|B_v|$ edges of the form (v, t) one for each label in B_v.

Clearly, the size of $\hat{N}$ is at most the size of N'. Since A_v is disjoint from B_v, $\hat{N}$ is 1-immune. It only remains to verify that $\hat{N}$ accepts all the inputs that N' accepts. Let y be accepted by N'. Thus on input y we obtain a path of all 1's from s to t. Since N' is 1-immune, this path must have length at least 2. Let v_1 be the second vertex on this path. Then, the edge (s, v_1) and an edge going out of v_1 are set to 1 on input y. Thus, there is a path generated from s to t via v_1 in $\hat{N}$. Hence, $\hat{N}$ accepts y. □

Lemma 4.2 *Let N be a 1-immune monotone contact network that accepts all the sets of size k. Then the size of N is at least $n\log(\frac{n}{k-1})$.*

Proof: By lemma 4.1 we may assume that N is a depth two contact network. For each internal vertex v of N, let A_v be the set of labels that appear on the edges of the form (s, v) and B_v be the set of labels that appear on the edges of the form (v, t). Let G_v be the undirected bipartite graph with vertex set $\{x_1, \ldots, x_n\}$ and edge set $E(G_v) = A_v \times B_v$. Note that size of N is at least $\sum_v \text{size}(G_v)$. Let $G_N = \bigcup_v G_v$.

Since N accepts all the sets of size k we have that G_N has no independent set of size k. By lemma 3.3 we have that $\sum_v \text{size}(G_v) \geq n\log(\frac{n}{k-1})$. The lemma follows from this. □

Lemma 4.3 Let $k \geq 2$ and let N be a monotone contact network computing T_k^n. Then

$$\text{size}(N) \geq \left\lfloor \frac{k}{2} \right\rfloor n\log(\frac{n}{k-1}).$$

Proof: We shall use induction on d to show that the following assertion holds for all positive integers d.

If N is a $(2d-1)$-immune monotone contact network that accepts all sets of size k, then size$(N) \geq dn\log(\frac{n}{k-1})$.

The basis case, when $d = 1$, is lemma 4.2 above. Assume that the assertion is true with $d = r$, for some positive integer r. We shall show that the assertion holds for $d = r + 1$.

Suppose that N is a $(2(r+1)-1)$-immune contact network that accepts all sets of size k. Let V' be the set of vertices that accept some input of size at most two. Note that t is in V' and s is not in V'. Let V_2 be those vertices in V' that are 1-immune. Let $V_1 = V' - V_2$.

Let M_1 be the network obtained from N by deleting all the vertices outside V' and the edges incident on them. The source of M_1 is obtained by collapsing all the vertices in V_2 to a single node. The sink of M_1 will be t. Let N_1 be the network obtained from N by collapsing all the vertices in V' to form the sink. The source of N_1 will be s. Note that size$(N) \geq$ size$(M_1)+$ size(N_1).

Clearly, M_1 is 1-immune. Since N is $(2(r+1)-1)$-immune and all vertices in V' accept some set of size at most 2, N_1 is $(2r-1)$-immune. Also, N_1 accepts all sets of size k. We may thus apply the assertion with $d = r$ to N_1 and obtain

$$\text{size}(N_1) \geq rn\log(\frac{n}{k-1}). \tag{3}$$

Next we show that M_1 accepts all the inputs that N accepts. Let y be an input accepted by N. Then, there is a path in N from s to t with all labels set to 1 by y. Let v be the last vertex on this path that is 1-immune. (Note that s is 1-immune, hence there is at least one such node.) All the vertices after v are not 1-immune, hence those vertices are all in V_1. We claim that $v \in V_2$. Since the successor of v on the path is in V_1, v accepts a set of size at most two. Thus v is a 1-immune vertex in V'. It follows that v is in V_2. Thus y creates an all 1's path from a vertex in V_2 to t all of whose edges are contained in M_1. Thus M_1 accepts all the sets that N accepts. It follows that M_1 accepts all sets of size k. By lemma 4.2 we have

$$\text{size}(M_1) \geq n\log(\frac{n}{k-1}). \tag{4}$$

Combining (3) and (4) we get

$$\text{size}(N) \geq \text{size}(M_1) + \text{size}(N_1) \geq (r+1)n\log(\frac{n}{k-1}).$$

This completes the induction step.

We may now complete the proof the lemma by taking $d = \lfloor \frac{k}{2} \rfloor$ in the above assertion. □

Corollary 4.4 Every monotone formula computing T_k^n, for $2 \le k \le \frac{n}{2}$, has size $\Omega(kn\log(\frac{n}{k-1}))$. □

5 ΣΠΣ formulas

In this section we present the results on ΣΠΣ formulas computing T_k^n. As explained in the introduction, our lower bound results are interesting only for small values of k. For the lower bound results we shall assume that n and k are large numbers and $k \le (\log\log n)^2$.

5.1 Preliminaries

Definition 5.1 (Fredman-Komlós graph) Let f be a formula on n variables. For $k \ge 2$, the graph $G(f,k)$ is defined by

$$\begin{aligned} V(G(f,k)) &= \{(C,x) : C \in \binom{[n]}{k-2} \\ &\qquad \textit{and } x \in [n] - C\}; \\ E(G(f,k)) &= \{((C,x),(D,y)) : C = D \\ &\qquad \textit{and } f \textit{ accepts } C \cup \{x,y\}\}. \end{aligned}$$

In the special case of $k = 2$, we may think of $G(f,k)$ as a graph with vertex set $[n]$ where (i,j) is an edge if and only if $\{i,j\}$ is accepted by f. In our discussion, the parameter k, in the above definition, will often be clear from the context. For notational convenience, we write $G(f)$ instead of $G(f,k)$, in that case.

Let F_0 be a formula computing T_k^n. Then $G(F_0)$ consists of $\binom{n}{k-2}$ components, where each component is a complete graph on $n-k+2$ vertices. We shall denote this graph by $C(k)$. Let G be a subgraph $C(k)$ with $V(G) = V(C(k))$. The subgraph of G induced by the vertices (C,x), with the same value for C will be called a *block* of $G(f)$. We shall denote this block by G_C. Thus, there are $\binom{n}{k-2}$ blocks, one for each $C \in \binom{[n]}{k-2}$.

Let $\hat{H}(G)$ be the sum of the entropies of all the blocks of G. For example, if $G = C(k)$, then each block is a complete graph with $n-k+2$ vertices. Thus, each block has entropy $\log(n-k+2)$. Therefore, $\hat{H}(C(k)) = \binom{n}{k-2}\log(n-k+2)$. Since $L(x)$ is a concave function we get the following corollary to lemma 3.5.

Corollary 5.2 If G is a subgraph of $C(k)$ with $V(G) = V(C(k))$, then

$$\hat{H}(G) \le \binom{n}{k-2} L(\frac{|E(G)|}{|V(G)|}). \quad □$$

We call a ΠΣ formula f *k-optimal* if f is $(k-1)$-immune and no $(k-1)$-immune ΠΣ formula g satisfies size$(g) <$ size(f) and $G(f,k)$ is a subgraph of $G(g,k)$.

Lemma 5.3 Let $k \ge 2$ and let $f = \bigwedge_{j=1}^{t} \bigvee_{q \in S_j} q$ be a k-optimal formula. Then, no S_j has more than $k-1$ negated variables.

Proof: Suppose the lemma does not hold. Say S_{j_0} has at least k negated variables. Consider the formula g obtained by omitting S_{j_0}. Clearly $G(f,k)$ is a subgraph of $G(g,k)$. Also size$(g) <$ size(f). Further if g accepts $T \subseteq [n]$, $|T| < k$ then so would f, for $\bigvee_{q \in S_{j_0}} q$ evaluates to 1 on any such T. Since f is k-optimal, f is $(k-1)$-immune and hence does not accept any such T. Hence g does not accept such a T and is $(k-1)$-immune. This contradicts the optimality of f and the proof is complete. □

Lemma 5.4 Let $f = \bigwedge_{j=1}^{t} \bigvee_{q \in S_j} q$ be a 1-optimal formula. Then no two S_j have the same negated variable.

Proof: By the optimality of f we may assume that no S_j has both a variable and its negation. Assume that the lemma does not hold. Let S_i and S_j have the same negated variable, say $\overline{x}_1$. By the previous lemma they have no other negated variable. Let g be the formula obtained from f by omitting S_i. As before $G(f,2)$ is a subgraph of $G(g,2)$ and size$(g) <$ size(f). We claim that g is 1-immune. For, if g accepts T and $1 \notin T$ then f accepts T. Also, g does not accept $\{1\}$ for $\bigvee_{q \in S_j} q$ evaluates to 0 on $\{1\}$. Thus if g accepts T and $|T| < 2$ then so does f. Since f is 1-immune, g is 1-immune. This contradicts the optimality of f and we have established the lemma. □

5.2 ΠΣ formulas

Consider the ΣΠΣ formula $F = \bigvee_{i=1}^{p} \bigwedge_{j=1}^{t_i} \bigvee_{q \in S_{ij}} q$. Suppose that F computes T_k^n. Then $G(F)$ consists of $\binom{n}{k-2}$ disjoint complete graphs of size $n-k+2$. Let $A_i = \bigwedge_{j=1}^{t_i} \bigvee_{q \in S_{ij}} q$. Note that $G(F)$ is the union of the graphs $G(A_i)$, $i = 1,\ldots,p$. Roughly speaking, we shall show that if the size of A_i is small, then $\hat{H}(G(A_i))$ is also small. Notice that since the union of the $G(A_i)$ is $C(k)$, we have that $\sum_{i=1}^{p} \hat{H}(G(A_i)) \ge \binom{n}{k-2}\log(n-k+2)$.

To relate size(A_i) to $\hat{H}(G(A_i))$, we need two results. The combinatorial result, lemma 5.6, states, roughly speaking, that if size(A_i) is small then only a small number of blocks are nonempty in $G(A_i)$. Due to technical difficulties, introduced by the presence of negated variables, we actually show that if some edges are deleted from $G(A_i)$ then most of the blocks are empty. The edges deleted from the different $G(A_i)$ put together are so few that they do not contribute significantly to the entropy of the final graph. The proof of this result is long and technical. We shall not include it in this paper. The detailed proof may be found in [18].

However, this result, in itself, is not sufficient to complete the proof of the lower bound. Even if the number of nonempty blocks is small, each such block may be very dense and could still have entropy which, for our purposes, is not sufficiently small. To bound the entropy of a block, we observe that the edges contained in a block correspond to the edges accepted by a certain 1-immune $\Pi\Sigma$ formula. For example, the edges contained in the block corresponding to $D \in \binom{[n]}{k-2}$ are in direct correspondence with the edges of $G(A_i|_D, 2)$. Note that $G(A_i|_D, 2)$ has only one block and hence $H(G(A_i|_D, 2)) = \hat{H}(G(A_i|_D, 2))$. The second result (lemma 5.5) relates the size of a 1-immune $\Pi\Sigma$ formula with the entropy of its graph.

We now present the proof of the lower bound in detail.

Lemma 5.5 Let $A = \bigwedge_{j=1}^{t} \bigvee_{q\in S_j} q$ be a 1-immune $\Pi\Sigma$ formula on n variables. Then

$$H(G(A,2)) \leq 2L(\frac{\text{size}(A)}{n}).$$

Proof: We may assume that A is 1-optimal. By lemma 5.3 an S_j may have at most one negated variable. Since A does not accept the empty set not all S_j have a negated variable. By lemma 5.4 no two S_j have the same negated variable. Let $S_1, S_2, \ldots, S_{t'}$ not have any negated variable. Further, let the negated variable in $S_{t'+j}$ be x_j for $1 \leq j \leq t-t'$.

Let G_1 be the subgraph of $G(A,2)$ with vertex set $[n]$ and consisting of all edges that have at least one end in $\{1, 2, \ldots, t-t'\}$. Let G_2 be the graph with vertex set $[n]$ consisting of the remaining edges of $G(A,2)$.

Now, for $1 \leq j \leq t-t'$ if $(i,j) \in E(G(A,2))$ then $x_i \in S_{t'+j}$. It follows that $|E(G_1)| \leq \sum_{j=1}^{t-t'}(|S_{t'+j}| - 1) \leq \text{size}(A)$. By lemma 3.5, we have that there is a coloring of G_1 with entropy at most $L(\frac{\text{size}(A)}{n})$.

Next we consider G_2. Let $\chi : [n] \to [t']$ be defined as follows.

$$\begin{aligned}\chi(j) &= 1 \;\; \textit{for } 1 \leq j \leq t-t'; \\ \chi(j) &= \min\{r : S_r \textit{ does not contain } x_j\} \\ &\qquad \textit{for } t-t' < j \leq n.\end{aligned}$$

Since A is 1-immune, every variable x_j not appearing in the negated form in A satisfies $x_j \notin S_r$ for some r, $1 \leq r \leq t'$. Thus, χ is well defined. We claim that χ is a coloring of G_2. For let $(i_1, i_2) \in E(G_2)$. Since vertices $1, \ldots, t-t'$ are isolated in G_2, $t-t' < i_1, i_2 \leq n$. Suppose χ colors both i_1 and i_2 by the same color, say r. Then $\bigvee_{q\in S_r} q$ evaluates to 0 on $\{i_1, i_2\}$ and hence (i_1, i_2) is not an edge of $G(A,2)$ and therefore not an edge in G_2. This contradicts our assumption. Hence χ is a coloring of G_2. By our definition $\sum_{j=1}^{n}(\chi(j) - 1) \leq \text{size}(A)$. Thus if the random variable X takes values in $[n]$ with uniform distribution then $E(\chi(X) - 1) \leq \frac{\text{size}(A)}{n}$. It follows from lemma 3.4 that

$$H(G_2) \leq H(\chi(X)) \leq H(\chi(X) - 1) \leq L(\frac{\text{size}(A)}{n}).$$

Since both G_1 and G_2 have colorings with entropy at most $L(\frac{\text{size}(A)}{n})$, we have that their union has a coloring with entropy at most $2L(\frac{\text{size}(A)}{n})$. □

The following lemma is the combinatorial result discussed earlier.

Lemma 5.6 Let $A = \bigwedge_{j=1}^{t} \bigvee_{q\in S_j}$ be a $(k-1)$-immune $\Pi\Sigma$ formula. Let $\Gamma = \{\gamma \in \binom{[n]}{k} : A \textit{ accepts } \gamma\}$. Let $\alpha(k) = \frac{1}{6e^2}\sqrt{\frac{k}{\ln k}}$.

(a) Suppose size$(A) \leq \frac{n}{2}$. Let

$$\Psi = \{\overline{a} \in \binom{[n]}{k-2} : \exists\, x, y \; \overline{a} \cup \{x, y\} \in \Gamma\}.$$

Then $|\Psi| \leq \left(\frac{\text{size}(A)}{n}\right) e^{-\alpha(k)} \binom{n}{k-2}$.

(b) Suppose size$(A) > \frac{n}{2}$. Then there exists a set $\Delta \subseteq \Gamma$, $|\Delta| \leq n^{-\frac{1}{3}}\binom{n}{k}$ such that if

$$\Psi = \{\overline{a} : \exists\, x, y \; \overline{a} \cup \{x, y\} \in \Gamma - \Delta\},$$

then $|\Psi| \leq \left(\frac{\text{size}(A)}{n}\right) e^{-\alpha(k)} \binom{n}{k-2}$. □

This result states that if we remove $\binom{k}{2} n^{-\frac{1}{3}} \binom{n}{k}$ edges from $G(A, k)$, then the number of nonempty blocks is small.

Lemma 5.7 Let A be a $(k-1)$-immune $\Pi\Sigma$ formula on n variables. Let G' be a subgraph of $G(A, k)$ with at most $(\frac{\text{size}(A)}{n}) e^{-\alpha(k)} \binom{n}{k-2}$ nonempty blocks.

(a) If $\text{size}(A) \leq ne^{\alpha(k)}$, then $\hat{H}(G') \leq \binom{n}{k-2}\frac{\text{size}(A)}{n}e^{-\alpha(k)}2L(\frac{\text{size}(A)}{n-k+2})$.

(b) If $\text{size}(A) > ne^{\alpha(k)}$, then $\hat{H}(G') \leq \binom{n}{k-2}2L(\frac{\text{size}(A)}{n-k+2})$.

Proof: We think of G' as consisting of $\binom{n}{k-2}$ disjoint blocks G'_D, one for each $D \in \binom{[n]}{k-2}$. Observe that G'_D is a subgraph of $G(A|_D, 2)$. (Here we think of $A|_D$ as a formula on $n-k+2$ variables). Since A is $(k-1)$-immune $A|_D$ is 1-immune. By lemma 5.5,

$$H(G'_D) \leq 2L(\frac{\text{size}(A|_D)}{n-k+2}) \leq 2L(\frac{\text{size}(A)}{n-k+2}).$$

This gives us an upper bound on the entropy of any nonempty block. Since the number of nonempty blocks in G' is at most $\frac{\text{size}(A)}{n}e^{-\alpha(k)}\binom{n}{k-2}$, we can conclude

$$\hat{H}(G') \leq \binom{n}{k-2}\frac{\text{size}(A)}{n}e^{-\alpha(k)}2L(\frac{\text{size}(A)}{n-k+2}).$$

Since the number of nonempty blocks is at most $\binom{n}{k-2}$ we always have that

$$\hat{H}(G') \leq \binom{n}{k-2}2L(\frac{\text{size}(A)}{n-k+2}). \quad \square$$

5.3 The lower bound

We are now ready to prove the main result of this section.

Theorem 5.8 If $F = \bigvee_{i=1}^{p}\bigwedge_{j=1}^{t_i}\bigvee_{q\in S_{ij}} q$ is a $\Sigma\Pi\Sigma$ formula computing T_k^n then

$$\text{size}(F) \geq \exp(\delta(k))n\log n,$$

where $\delta(k) = \frac{1}{50}\sqrt{\frac{k}{\ln k}}$.

Proof: Let $A_i = \bigwedge_{j=1}^{t_i}\bigvee_{q\in S_{ij}} q$, let $\alpha(k)$ be as in lemma 5.6. Let

$$\begin{aligned} I_1 &= \{i : \text{size}(A_i) \leq \frac{n}{2}\};\\ I_2 &= \{i : \frac{n}{2} < \text{size}(A_i) \leq ne^{\alpha(k)}\};\\ I_3 &= \{i : \text{size}(A_i) > ne^{\alpha(k)}\}.\end{aligned}$$

For $i \in I_2$ we may, by lemma 5.6, write $G(A_i) = G_1(A_i,k) \cup G_2(A_i,k)$, where $G_1(A_i,k)$ has at most $\frac{\text{size}(A)}{n}e^{-\alpha(k)}\binom{n}{k-2}$ nonempty blocks and $G_2(A_i,k)$ has at most $\binom{k}{2}n^{-\frac{1}{3}}\binom{n}{k}$ edges. Now,

$$\begin{aligned} G(F) = \bigcup_i G(A_i) &= \bigcup_{i\in I_1} G(A_i) \cup \bigcup_{i\in I_2} G_1(A_i,k) \cup \\ &\quad \bigcup_{i\in I_2} G_2(A_i,k) \cup \bigcup_{i\in I_3} G(A_i).\end{aligned}$$

Let $G' = \bigcup_{i\in I_2} G_2(A_i,k)$. Then,

$$|E(G')| \leq |I_2|\binom{k}{2}n^{-\frac{1}{3}}\binom{n}{k}.$$

Since all the blocks of $C(k)$ are complete graphs with $n-k+2$ vertices, we get, using lemma 3.2, that

$$\sum_{i\in I_1}\hat{H}(G(A_i)) + \sum_{i\in I_2}\hat{H}(G_1(A_i,k)) + \hat{H}(G') + \sum_{i\in I_3}\hat{H}(G(A_i)) \geq \binom{n}{k-2}\log(n-k+2).$$

By corollary 5.2, lemma 5.7 and lemma 5.6 we have that

$$\sum_{i\in I_1\cup I_2}\frac{\text{size}(A_i)}{n}e^{-\alpha(k)}2L(\frac{\text{size}(A_i)}{n-k+2}) + L(\frac{|E(G')|}{|V(G')|}) + \sum_{i\in I_3}2L(\frac{\text{size}(A_i)}{n-k+2}) \geq \log(n-k+2).$$

The lemma follows from the last inequality. We omit the details. $\square$

5.4 The upper bound

In this section we show that there exist $\Sigma\Pi\Sigma$ formulas for computing T_k^n, for n and k large enough and $k \leq \sqrt{n}$, of size at most

$$e^{2\sqrt{k}\ln k}n\log n.$$

For simplicity, assume that $k^{\frac{1}{2}}$ is an integer. We construct the formulas in two stages. In the first stage we construct $\Pi\Sigma$ formulas. These formulas are $(k-1)$-immune and they accept a large proportion of all inputs that a formula computing T_k^n must accept. In the next stage, we take the disjunction of random copies of this formula and obtain a $\Sigma\Pi\Sigma$ formula computing T_k^n.

Lemma 5.9 There exists a $\Pi\Sigma$ formula computing $T_l^{l^2}$ of size at most

$$\binom{l^2}{l-1}(l^2-l+1).$$

Proof: Let

$$F = \bigwedge_{S \in \binom{[l^2]}{l^2-l+1}} \bigvee_{j \in S} x_j.$$

It is easy to verify that F computes $T_l^{l^2}$. Also,

$$\text{size}(F) = \binom{l^2}{l-1}(l^2 - l + 1). \quad \square$$

Lemma 5.10 There exists a $(k-1)$-immune $\Pi\Sigma$ formula G such that $\text{size}(G) \le \binom{k}{\sqrt{k}} n$ and G accepts at least $\left(\frac{1}{e^2\sqrt{k}}\right)^{\sqrt{k}} \binom{n}{k}$ sets of size k.

Proof: Let $l = \sqrt{k}$. Let $D_1, D_2, \ldots, D_l$ be a partition of $[n]$ into l equal parts. For each $i = 1, \ldots, l$, let $D_i^1, D_i^2, \ldots, D_i^{l^2}$ be a partition of D_i into l^2 equal parts. Thus $|D_i^j| = \frac{n}{l^3}$.

Let F_i be the formula obtained from the formula F in lemma 5.9 by replacing the variable x_j by $\bigvee_{q \in D_i^j} x_q$. That is,

$$F_i = \bigwedge_{S \in \binom{[l^2]}{l^2-l+1}} \bigvee_{j \in S} \bigvee_{q \in D_i^j} x_q$$

Let $G = \bigwedge_{i=1}^{l} F_i$. Note that G is a $\Pi\Sigma$ formula and it is $(k-1)$-immune. We have

$$\begin{aligned} \text{size}(F_i) &= \binom{l^2}{l-1}(l^2 - l + 1)\frac{n}{l^3}; \\ \text{size}(G) &= \frac{l^2 - l + 1}{l^2}\binom{l^2}{l-1} n \\ &\le \binom{l^2}{l} n. \end{aligned}$$

The number of sets of size k accepted by G is given by

$$\prod_{i=1}^{l} (\textit{the number of sets of size } l \textit{ accepted by } F_i)$$

$$= \left[\binom{l^2}{l}\left(\frac{n}{l^3}\right)^l\right]^l \ge \left(\frac{1}{e^2 l}\right)^l \binom{n}{l^2}.$$

Since $l^2 = k$, the lemma follows from this. $\square$

Theorem 5.11 There exists a $\Sigma\Pi\Sigma$ formula of size at most $e^{2\sqrt{k}\ln k} n \log n$ computing T_k^n.

Proof: Let r be a parameter to be chosen later. We take r independent copies of the formula G described in lemma 5.10 by randomly permuting the variable set. Let these random copies be $G_1, G_2, \ldots, G_r$. For any fixed set T of size k,

$$\Pr[G_i \textit{ does not accept } T] \le \left(1 - \left(\frac{1}{e^2\sqrt{k}}\right)^{\sqrt{k}}\right).$$

Since the G_i are independently chosen, the probability that some k-set is accepted by none of $G_1, \ldots, G_r$ is at most

$$\binom{n}{k}\left(1 - \left(\frac{1}{e^2\sqrt{k}}\right)^{\sqrt{k}}\right)^r.$$

For $r = k(e^2\sqrt{k})^{\sqrt{k}} \log n$, we get that this probability is less than 1. Hence there must be some r copies $\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_r$, such that every set of size k is accepted by at least one of them. Let our $\Sigma\Pi\Sigma$ formula for T_k^n be $\hat{F} = \bigvee_{i=1}^{r} \hat{G}_i$.

Clearly, $\hat{F}$ is $(k-1)$-immune. It accepts every set of size k and by monotonicity every set of size at least k. Further,

$$\text{size}(\hat{F}) \le \binom{k}{\sqrt{k}}(e^2\sqrt{k})^{\sqrt{k}} kn \log n \le e^{2\sqrt{k}\ln k} n \log n.$$

$\square$

6 Conclusions

We have shown a lower bound of $\Omega(kn\log(\frac{n}{k-1}))$ on the size of monotone formulas computing T_k^n, for $2 \le k \le \frac{n}{2}$. We also showed that the same bound holds in the monotone contact networks model. The following questions remain open.

1. Our bound is still far from the best upper bound known. Can this gap be reduced?
2. Does the $\Omega(kn\log(\frac{n}{k-1}))$ lower bound hold on the size of formulas computing T_k^n, for $2 \le k \le \frac{n}{2}$, even when negations are allowed? Note that there exist contact networks with size $O(kn)$, if negations are allowed.
3. Our result does not improve Khrapchenko's lower bound for the majority function. Can a lower bound better than $\Omega(n^2)$ be shown for the size of monotone formulas computing the majority function?
4. In the monotone formulas model, the complexities of computing T_2^n and T_{n-1}^n are the same. However, this is not obvious in the monotone contact networks model. Is there a lower bound of $\Omega(n\log n)$ on the size of monotone contact networks computing T_{n-1}^n ?

In this paper, we also considered $\Sigma\Pi\Sigma$ formulas computing T_k^n, for small k, and obtained essentially tight bounds on the size of such formulas. Can these results be generalized to depths greater than three?

Acknowledgments

I wish to thank Endre Szemerédi for his generous contribution to this work. I thank Ravi Boppana for his advice and encouragement; his suggestions contributed greatly to the presentation in this paper. I am grateful to Shiva Chaudhuri, Rafi Heiman, Zoli Király, Ilan Newman, Gábor Tardos, and Avi Wigderson for their comments on the earlier versions of this paper.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. *Sorting in* $c \log n$ *parallel steps.* Combinatorica, 3, 1983, pp. 1–19.

[2] R. B. Boppana. *Amplification of Probabilistic Boolean Formulas.* Advances in Computing Research, Vol. 5, 1989, pp. 27–45.

[3] R. B. Boppana. *Optimal Separations Between Concurrent-Write Parallel Machines.* Proceedings of the twenty first ACM STOC, 1989, pp. 320–326.

[4] R. B. Boppana and M. Sipser. *The Complexity of Finite Functions.* Chapter 14, The Handbook of Theoretical Computer Science, (J. van Leeuwen, ed.), Elsevier Science Publishers B. B., 1990, pp. 759–804.

[5] I. Csiszár and J. Körner. *Information Theory, Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1982.

[6] M. Fredman and J. Komlós. *On the size of Separating Systems and Perfect Hash Functions.* Siam J. Alg. Disc. Meth., vol 5 1984, pp. 61–68.

[7] M. J. Fischer, A. R. Meyer, and M. S. Paterson. $\Omega(n \log n)$ *Lower bounds on the length of Boolean formulas.* SIAM J. Comp., Vol. 11, No. 3, 1982, pp. 416–427.

[8] G. Hansel. *Nombre minimal de contacts de fermature nécessaires pour réaliser une fonction booléenne symétrique de n variables.* C. R. Acad. Sci. Paris 258 (1964), pp. 6037–6040.

[9] J. Körner. *Fredman-Komlós Bound and Information Theory.* Siam J. Alg. Disc. Meth., vol. 7, 1986, pp. 560–570.

[10] V. M. Khrapchenko. *A method of obtaining lower bounds for the complexity* π*-schemes.* Math. Notes Acad. Sci. USSR 11 (1972), pp. 474–479.

[11] R. E. Krichevskii. *Complexity of contact circuits realizing a function of logical algebra.* Sov. Phys. Dokl. 8 (1964) pp. 770–772.

[12] O. B. Lupanov. *On comparing the complexity of realization of monotone contact networks containing only closing contacts and by arbitrary contact networks.* Sov. Phys. Dokl. 7 (1962) pp. 486–489.

[13] E. F. Moore and C. E. Shannon. *Reliable circuits using reliable relays.* J. Franklin Inst. 262, 1956, pp. 191–208 and pp. 281–297.

[14] I. Newman, P. Ragde, and A. Wigderson. *Perfect Hashing, Graph Entropy and Circuit Complexity* Proceedings of the fifth annual conference on Structure in Complexity Theory, 1990, pp. 91–90.

[15] M. S. Paterson, N. Pippenger, and U. Zwick. *Faster circuits and shorter formulae for multiple addition, multiplication and symmetric Boolean functions.* Proceedings of the thirty first IEEE FOCS, 1990, pp. 642–650.

[16] N. Pippenger. *An Information-Theoretic Method in Combinatorial Theory.* Journal of Combinatorial Theory (A) 23, 1977, pp. 99-104.

[17] P. Pudlák. *Bounds for Hodes-Speckor theorem.* Logic and Machines: Decision Problems and Complexity, Lecture Notes in Computer Science, Vol 171, (Springer, Berlin, 1984), pp. 421–445.

[18] J. Radhakrishnan. *Graph Entropy and Formula Complexity.* Manuscript, 1991.

[19] A. A. Razborov. *Lower Bounds for Computing Symmetric Boolean Functions by Switching and Rectifier Networks.* Manuscript.

[20] M. Snir. *The Covering Problem of Complete Uniform Hypergraphs.* A note, Discrete Math. 27, 1979, pp. 103–105.

[21] L. G. Valiant. *Short monotone formulae for the majority function.* Journal of Algorithms 5, 1984, pp. 363–366.

[22] I. Wegener. *The Complexity of Boolean Functions.* Wiley-Teubner Series in Computer Science, 1987.