Finite Limits and Lower Bounds for Circuits Size*

Stasys Jukna[†‡]

Abstract

The lower bounds problem in circuit complexity theory may be looked as the problem about the possibility to diagonalize over *finite* sets of computations. In this paper we show that Sipser's notion of "finite limit" is the right diagonal for different models of small-depth circuits. This is partly a survey paper, but it also contains various new results as well as new proofs of know ones.

Key words: Computational complexity, small depth circuits, AC^0, ACC, branching programs, lower bounds problem, diagonalization, cube colorings, discrepancy.

1 Introduction

Important property of computations is their "locality": at each step the decision about the next one is made by looking at only small (in most models, constant) peace of information conducted so far. It is therefore natural to use this locality in lower bound proofs. Suppose we are given some "hard" function f and want to prove that it is really hard, i.e. that f requires long computations. Assume the opposite and, starting from some string in $f^{-1}(0)$, try to construct a computation for it which is a "limit" of accepting computations, meaning that on each *local part* it coincides with at least one accepting computation. By locality of computations, the machine cannot distinguish this computation from the set of all accepting computations, and hence, makes an error (accepts a string from $f^{-1}(0)$).

There are two general ideas how to construct such a "limit computation": the "topological approach" of Sipser [23] and the "method of approximations" of Razborov [20]. Roughly, these ideas tell us how, given a $0,1$-matrix

*Tech. Rep. # 94-06, Uni Trier, 1994.

[†]Department of Computer Science, University of Trier, D-54286 Trier, Germany. E-mail: jukna@@uni-trier.de

[‡]Research supported by DFG-project Me 1077/5-2.

M (rows are computations) and a family $\mathcal{F}$ of subsets of its columns (sets in $\mathcal{F}$ corresponds to gates), to construct a "limit row", i.e. a row which is not in M but coincides on each set of columns from $\mathcal{F}$ with at least one row of M. The idea (unlike its realization !) is very simple: try to construct the desired row either as a limit point in an appropriate topology on the set of strings or by applying appropriate boolean functions to the columns of M. The idea was materialized for some circuit models: countable nondeterministic circuits [23, 4], monotone circuits [19], switching-and-rectifier networks [21] and parity branching programs [16] (see [25] for a survey). All these methods actually do the same: they diagonalize computations by constructing a "limit computation". There are also other situations (not covered by [25]) were the notion of "finite limit" works. In this paper we show how do limits appear in lower bounds proofs for small depth circuits and small depth branching programs. Our main goal was to look at these methods in a "clean form" in order to see what diagonalization tricks they actually use and what do we need to get better lower bounds. We do this in the following order.

In the next section we outline the way in which "limit computations" appear as diagonals in lower bound proofs and sketch two general diagonalization ideas mentioned above. In Section 3 we define the notions of k-limit, k-diagonal and k-closed set. Diagonal is a limit which lies outside the set; a set is closed if it contains no diagonal, i.e. no limit outside the set. We then state some basic facts about closed sets and introduce one combinatorial characteristic of (boolean) functions. This characteristic is denoted by $\Delta_k(f)$ and is defined as the minimal possible number of blocks in a partition of the domain of f such that each block is k-closed and f is constant on each of them. In the remaining sections we investigate this characteristic for various values of k. We start this investigation in Section 4 by proving that, for all values of k, the number $\Delta_k(f)$ is large for almost all functions. In sections 5 and 6 we show that $\Delta_k(f)$ with $k \ll n$ (n = the number of variables of f) is the lower bound for the size of AC^0 circuits. In Section 7 we show that $\Delta_k(f)$ with $k = n - 1$ is the lower bound for the size of ACC circuits, i.e. unbounded fanin $\{\wedge, \vee, \neg, \mathrm{mod}_m\}$ circuits. In the last section we show that some modification of $\Delta_k(f)$ is connected with time-space tradeoffs for Turing machines and gives lower bounds for read-r-times branching programs.

2 Intuition

To be more specific, fix some basis, i.e. some set H of boolean functions $h : \{0,1\}^k \to \{0,1\}$. (One may take, e.g. the standard fanin 2 basis $H = \{\wedge, \vee, \neg\}$.) A circuit over the basis H is a stright line program over H, that is a sequence $P = (g_1, g_2, \ldots, g_t)$ with $g_i = x_i$ for $1 \leq i \leq n$, and for every $i > n$, $g_i = h(g_{i_1}, \ldots, g_{i_k})$ for some $i_1, \ldots, i_k < i$ and $h \in H$. Those g_i's are "gates" of P; their number t is the "size" of P. The computation of P on a string $a \in \{0,1\}^n$ is the string

$$P(a) = ((g_1(a), g_2(a), \ldots, g_t(a)) \in \{0,1\}^t; \qquad (*)$$

it is an accepting computation if its last coordinate $g_t(a) = 1$. (Clearly, not every vector in $\{0,1\}^t$ is an accepting computation or even a computation of P on some input). The circuit P computes the boolean function $f(x_1, \ldots, x_n)$ in a usual way: $f(a) = 1$ iff $P(a)$ is an accepting computation. Let $U = f^{-1}(1) \subseteq \{0,1\}^n$ denote the set of "ones" of f. If $C \subseteq \{0,1\}^t$ is the set of all accepting computations of the circuit P, then $U = C|_{\{1,\ldots,n\}}$, i.e. U is the set of projections of vectors in C on the first n coordinates. The set C itself is very special ("locally defined") subset of $\{0,1\}^t$, namely, for each string $z \in C$ and every $i > n$, the i-th bit z_i of z depends only on some collection of k previous bits $z_{i_1}, \ldots, z_{i_k}$ of z. Thus, the lower bounds problem is the following question: given a set $U \subseteq \{0,1\}^n$ and a number $t \geq 0$, prove (or disprove) that $U \neq C|_{\{1,\ldots,n\}}$ for any "locally defined" set $C \subseteq \{0,1\}^t$. In other words, we want to prove that the set U cannot be separated from its complement $\{0,1\}^n \setminus U$ by making only small number of "local tests".

For a subset $S = \{i_1, \ldots, i_{k+1}\} \subset \mathbb{N} = \{1, 2, \ldots\}$, a *test* on S is a function $\phi : \{0,1\}^{\mathbb{N}} \to \{0,1\}$ for which there is some $h \in H$ such that for any (infinite) string $z \in \{0,1\}^{\mathbb{N}}$, $\phi(z) = 1 \iff h(z_{i_1}, \ldots, z_{i_k}) = z_{i_{k+1}}$; subset S is the *support* of this test. Observe that each gate in P makes the test: "is the output bit consistent with input bits?". We can therefore look the circuit P as the *set* Φ of tests made by its gates; the size is the number of tests in Φ. The set C of accepting computations of the circuit Φ is then simply the set of those and only those strings in $\{0,1\}^t$ that pass all the tests in Φ, i.e. $C = \left\{ z \in \{0,1\}^t : \bigwedge_{\phi \in \Phi} \phi(z) = 1 \right\}$. The lower bounds problem is to prove that if t is too small then $U \neq C|_{\{1,\ldots,n\}}$.

One can obtain such a proof by constructing a string $b \in \{0,1\}^t$ such that:

(i) the initial part $(b_1, \ldots, b_n)$ of b is not in U, and

(ii) b is a "limit" for the set C is the following sense: on each subset $S \in \text{support}(\Phi)$, the string b coincides with at least one string in C.

By (ii), the string b passes all tests, which means that $b \in C$. But by (i), the projection of b is not in U, and hence, $U \neq C|_{\{1,\ldots,n\}}$.

Thus, what we need is to *diagonalize* over the set C in order to find a limit string outside C. This can be done by combining (or "fusing", a term proposed in [25]) strings in C into a new string. There are two general ideas how to construct such a diagonal: the "topological approach" and the "method of approximations". (Wigderson in [25] proposed to look them as one "fusion method".) In this paper we will follow the first, "topological" idea but let us briefly sketch both of them.

The Topological Approach. Sipser [23] suggested approaching the NP vs. coNP question by studying its infinite analog: the classical result of Lebesque that the class of analytic sets is not closed under complements. Sipser provides a new topological proof of this result, which can hopefully be "finitized". The idea is to choose an appropriate notion of "large set" and construct the diagonal using some kind of "backward induction" preserving largeness of sets. A sequence of strings $a^1, a^2, \ldots$ in $f^{-1}(1)$ converging to a string $a^* \in f^{-1}(0)$ is constructed so that the associated sequence of accepting computations $z^1, z^2, \ldots$ converges to an accepting computation for a^*. Convergence here is in terms of an appropriate topology on the set of all strings. In [23] natural discrete topology (open set being the set of extension of a finite string) is shown to work in infinite case. The main role of "largeness" is to ensure (under the assumption that the number of tests in Φ is small) the possibility to continue the process until a^* is reached.

The Method of Approximations. The second idea, the "method of approximations", was suggested by Razborov [20] and developed by Karchmer and Wigderson in [15, 16, 17, 25]. The idea is to look the set of accepting computations $C \subseteq \{0,1\}^t$ as the $m \times t$ matrix with $m = |C|$, and try to obtain the desired limit string b by combining ("fusing", as suggested by Wigderson [25]) the rows into a new one. This is done by applying some special ("fusing") boolean functions $F : \{0,1\}^m \to \{0,1\}$ to the columns of C. Each such F produces the string $b = (F(c^1), \ldots, F(c^t)) \in \{0,1\}^t$ where $c^1, \ldots, c^t$ are columns of C. If we are lucky, our fusing function F preserves (in a natural sense) all the local tests in Φ and satisfies $f(F(c^1), \ldots, F(c^n)) = 0$; then b is the desired diagonal. Razborov in [20] has proved that if f has no circuit over $\{\wedge, \vee, \neg\}$ of size t then, potentially, the diagonal can be obtained by fusing with only *monotone* functions. There are, however, too much such "candi-

dates for a limit" (double exponent in $m = |U|$) and it is hard to choose the right one. Fortunately, Razborov was able to do this in [21] where he proves a super-linear lower bound on the size of switching-and-rectifier networks for the Majority function. Karchmer and Wigderson have then proved similar lower bound for the Majority in the class of parity branching programs. This was done using *linear* functions F. The third interesting class of fusing functions is that of monotone *self-dual* functions, i.e monotone boolean functions satisfying [1] $F(a) \vee \neg F(\neg a) = 1$. Karchmer in [15] shows how self-dual fusing functions can be used to re-prove Razborov's lower bound for monotone circuits. Recently, Ben-David, Karchmer and Kushilevitz [4] have demonstrated that the limit in Sipser's proof about analytic sets can be also obtained by fusing computations via self-dual functions.

3 Limits and Closed Sets

Consider the hypercube $\mathtt{E}_q^n$ where $\mathtt{E}_q = \{0, 1, \ldots, q-1\}$. Elements of $\mathtt{E}_q^n$ are vectors $a = (a_1, \ldots, a_n)$ with $a_i \in \mathtt{E}_q$. The *projection* of a onto a set of coordinates $S = \{i_1, \ldots, i_k\} \subseteq [n] \rightleftharpoons \{1, \ldots, n\}$ is the vector $a|_S \rightleftharpoons (a_{i_1}, \ldots, a_{i_k})$.

Let $\mathcal{F}$ be a family of subsets of $[n]$. A vector $b \in \mathtt{E}_q^n$ is an *$\mathcal{F}$-limit* of a subset $A \subseteq \mathtt{E}_q^n$ if for every set $S \in \mathcal{F}$ there exists a vector $a \in A$ such that $a \neq b$ but $a|_S = b|_S$. A *k-limit* is an $\mathcal{F}$-limit with $\mathcal{F} = [n]^k \rightleftharpoons \{S \subseteq [n] : |S| = k\}$. Let $\lim_{\mathcal{F}}(A)$ ($\mathsf{lim}_k(A)$) denote the set of all $\mathcal{F}$-limits (k-limits) for A. Then

$$\lim_1(A) \supseteq \lim_2(A) \supseteq \cdots \supseteq \lim_n(A) = \emptyset.$$

Each limit of A (if any) can lie inside as well as outside A. In this last case, the limit is a "diagonal" for A: it does not belong to A but this fact cannot be captured by looking at only small peaces of coordinates. Formally, we say that a vector $b \in \mathtt{E}_q^n$ is an *$\mathcal{F}$-diagonal* for a subset $A \subseteq \mathtt{E}_q^n$ if $b \notin A$ but b is an $\mathcal{F}$-limit of A, i.e. if $b \in \lim_{\mathcal{F}} \setminus A$. If $\mathcal{F} = [n]^k$ we say that b is a *k-diagonal* for A.

The main goal of this paper is to show that various lower bound methods in circuit complexity theory are actually different solutions of the following (purely combinatorial) "diagonalization problem":

> *exhibit an explicit set $U \subseteq \mathtt{E}_q^n$ such that in every partition of U into "small" number of blocks, at least one block has a k-limit in $\mathtt{E}_q^n \setminus U$.*

[1] For a vector $a \in \{0,1\}^m$, $\neg a = (\neg a_1, \ldots, \neg a_m)$.

In order to have a diagonal, the set must have at least one limit. To make this limit a diagonal, we have to force it be *outside* the set. And this is hard. Much easer is to prove that some limit lies *inside* the set. Namely, every sufficiently large set contains at least one of its limits. Although such limit is not a diagonal, let us state this fact (pointed out to the author by M. Sipser [22]) explicitly.

Proposition 1 *Let* $A \subseteq \mathtt{E}_q^n$. *If* $|A| > q^k\binom{n}{k}$ *then* $\mathsf{lim}_k(A) \cap A \neq \emptyset$, *i.e.* A *contains at least one of its* k*-limits.*

Proof. Suppose that $\mathsf{lim}_k(A) \cap A = \emptyset$. This, in particular, means that for every vector $a \in A$ there exists a subset $S(a) \subseteq [n]$, $|S(a)| = k$, such that $a|_{S(a)} \neq b|_{S(a)}$ for all $b \in A - \{a\}$. In other words, each vector $a \in A$ has its own "pattern", i.e. a set of k coordinates in which this vector differs from all other vectors in A. Hence, for any two different vectors a and b in A we have that either $S(a) \neq S(b)$ or $S(a) = S(b)$ but $a_i \neq b_i$ for some $i \in S(a)$, i.e. one can assign to each vector $a \in A$ its "pattern $\langle S, a|_S \rangle$ so that different vectors have different patterns. There are $\binom{n}{k}$ possible subsets of k coordinates, on each of which vectors can have no more than q^k possible values. Thus, there are at most $\binom{q}{k}q^k$ possible patterns $\langle$subset,value$\rangle$. Since each vector in A must have its own pattern, we conclude that $|A| \leq \binom{n}{k}q^k$. ∎

Definition. A set A is $\mathcal{F}$*-closed* if $\mathsf{lim}_{\mathcal{F}}(A) \subseteq A$, and $\mathcal{F}$*-open* otherwise. A set is k-closed if it is $\mathcal{F}$-closed with $\mathcal{F} = [n]^k$.

Remark: For $A \subseteq \mathtt{E}_q^n$, let $\mathrm{dist}(A) \rightleftharpoons \min\{d(a,b) : a,b \in A,\ a \neq b\}$ where $d(a,b)$ is the usual Hamming distance between the vectors a and b, i.e. $d(a,b) \rightleftharpoons |\{i : a_i \neq b_i\}|$. If $\mathrm{dist}(A) > n - k + 1$ then A is k-closed because then $\lim_k(A) = \emptyset$. Thus, sets with large distance are closed. On the other hand, by Proposition 1, such closed sets cannot be very large: if $\mathrm{dist}(A) \geq n - k + 1$ then no two vectors in A can coincide on k coordinates; this, in particular, means that $\mathsf{lim}_k(A) \cap A = \emptyset$, and hence, $|A| \leq q^k\binom{n}{k}$.

When dealing with closed sets it is sometimes useful to keep the following alternative definition in mind. Say that a function $\phi : \mathtt{E}_q^n \to \{0,1\}$ is k*-variable function* if ϕ depends on at most k variables. A function $f : \mathtt{E}_q^n \to \{0,1\}$ is a *generalized* k*-CNF* if it is a product (AND) of k-variable functions.

Proposition 2 *A set* $A \subseteq \mathtt{E}_q^n$ *is* k*-closed if and only if* $A = \Phi^{-1}(1)$ *for some generalized* k*-CNF* Φ.

Proof. Given $A \subseteq \mathbb{E}_q^n$, associate with each $S \subseteq [n]$ the k-variable function $\phi_S : \mathbb{E}_q^S \to \{0,1\}$ defined by: $\phi_S(y) = 1 \iff y = x|_S$ for some $x \in A$. One may easily check that the product $\Phi_A = \prod \phi_S$ over all $S \subseteq [n]$ with $|S| = k$, is the desired generalized k-CNF Φ. ∎

Corollary 3 *If A and B are k-closed then so is $A \cap B$.*

Corollary 4 *The hypercube $\mathbb{E}_q^n$ has at most $2^{q^k\binom{n}{k}}$ k-closed subsets.*

Proof. For each $S \subseteq [n]$ with $|S| = k$, there are 2^{q^k} functions $\phi_S : \mathbb{E}_q^S \to \{0,1\}$. ∎

Definition. For a function $f : \mathbb{E}_q^n \to \{0,1\}$, let $\Delta_k(f)$ denote the minimal possible number of blocks in a partition of $\mathbb{E}_q^n$ such that

- each block is k-closed, and
- f is constant on each block.

Note that $\Delta_1(f) \geq \Delta_2(f) \geq \ldots \geq \Delta_n(f)$ where $\Delta_n(f) = 1$ if f is a constant function, and $\Delta_n(f) = 2$ otherwise (since both blocks of the partition $\mathbb{E}_q^n = f^{-1}(0) \cup f^{-1}(1)$ are n-closed).

Various lower bound arguments in circuit complexity theory consist of two steps: the "reduction" step and then the "diagonalization" step :

1. Prove that if f has small circuit then $\Delta_k(f)$ is also small. The parameter k as well as the real meaning of "small" depend on a circuit model only.
2. Prove that $\Delta_k(f)$ is large.

In what follows we will concentrate on the second, "diagonalization" step. The first, "reduction" step will be only sketched.

4 Limits for Random Functions

For $k = 1$, k-closed sets are, by Proposition 2, exactly the subcubes of $\mathbb{E}_q^n$, i.e. sets of the form

$$A = \left\{ x \in \mathbb{E}_q^n \,:\, x_{i_1} = a_{i_1}, \ldots, x_{i_m} = a_{i_m} \right\}$$

where $a_{i_j} \in \mathtt{E}_q$, $1 \leq j \leq m \leq n$. The number $n - m$ is the *dimension* of this subcube. Thus, it is easy to exhibit a function $f : \mathtt{E}_q^n \to \{0,1\}$ with large $\Delta_1(f)$. Take for example, the Parity function $\mathsf{Par}_n = x_1 \oplus \cdots \oplus x_n$ over $\mathtt{E}_2^n$. Then $\Delta_1(\mathsf{Par}_n) = 2^n$, since Par_n is non-constant on each subcube of dimension $\neq 0$.

For larger values of k the problem becomes harder. One may ask if at all $\Delta_k(f)$ can be large if k is large? Not surprisingly, the answer is - yes, almost all functions f have this property.

For a function $f : \mathtt{E}_q^n \to \{-1,1\}$, its *discrepancy*, $\mathrm{disc}_A(f)$, on a set $A \subseteq \mathtt{E}_q^n$ is the absolute value of the sum $\sum_{x \in A} f(x)$. If our function f ranges over $\{0,1\}$ instead of $\{-1,1\}$ then its discrepancy is defined to be that of $(-1)^f$. Define

$$\mathrm{disc}_k(f) = \max\{\mathrm{disc}_A(f) \ : \ A \text{ is } k\text{-closed}\}$$

Note that

$$\Delta_k(f) \geq q^n/\mathrm{disc}_k(f) \tag{1}$$

This is because $f(A) \equiv const \iff \mathrm{disc}_A(f) = |A|$, and hence, f is non-constant on all k-closed sets containing more than $\mathrm{disc}_k(f)$ vectors.

Theorem 5 *For almost all functions $f : \mathtt{E}_q^n \to \{0,1\}$ we have that*

$$\Delta_k(f) \geq \sqrt{\frac{q^{n-k}}{2\binom{n}{k}}} \tag{2}$$

Theorem 5 follows directly from (1) and the following simple lemma.

Lemma 6 *Consider the random function $\mathsf{f} : \mathtt{E}_q^n \to \{-1,1\}$ taking the values -1 and $+1$ independently with probability $1/2$. Then*

$$\mathrm{Prob}[\mathrm{disc}_k(\mathsf{f}) > N] < 2\exp\left(-\alpha q^k \binom{n}{k}\right)$$

where $\alpha = 1 - \ln 2 = 0.31...$ and

$$N = \sqrt{2q^{n+k}\binom{n}{k}}.$$

Proof. By the Chernoff's bound [2] we have for any subset $A \subseteq \mathbb{E}_q^n$ that

$$\begin{aligned} \text{Prob}[\text{disc}_A(\mathsf{f}) > N] &< 2e^{-N^2/2\#(A)} \\ &\leq 2e^{-q^k\binom{n}{k}}. \end{aligned}$$

By Corollary 4, there are at most $\exp\left((1-\alpha)q^k\binom{n}{k}\right)$ closed sets, and the desired bound follows. ∎

5 Limits for $k = 1$

Even for $k = 1$ the number $\Delta_k(f)$ gives lower bounds for AC^0 circuits, i.e. bounded depth unbounded fanin AND/OR circuits. Recall that by Proposition 2, a set $A \subseteq \mathbb{E}_2^n$ is 1-closed iff A is a subcube of $\mathbb{E}_2^n$.

The "bottom-up" method of random restrictions of Furst, Saxe and Sipser [9]; Yao [26] and Håstad [12] have led to the following lower bounds criterion for AC^0 circuits.

Theorem 7 **([5])** *Suppose that $f : \mathbb{E}_2^n \to \{0,1\}$ has an AC^0 circuit of depth d and size L. Then f is constant on at least one subcube of dimension $m \rightleftharpoons n/3(10\log L)^{d-2} - \log L$, and hence, $\Delta_1(f) \leq 2^{n-m}$.*

The Parity function $\mathsf{Par}_n = x_1 \oplus \cdots \oplus x_n$ has $\Delta_1(\mathsf{Par}_n) = 2^n$. Hence $n/3(10\log L)^{d-1} - \log L \leq 0$ which gives the lower bound $L \geq \exp\left(\Omega(n^{1/(d-1)})\right)$ for this function.

6 Limits for $1 \ll k \ll n$

Theorem 8 **([13])** *Suppose that $f : \mathbb{E}_2^n \to \{0,1\}$ has an AC^0 circuit of depth 3 and size L. Then there is a subfunction f' of f on $\Omega(n)$ variables such that*

$$L \geq \Delta_k(f') \qquad \textit{where} \qquad k = \log L.$$

Remark: What is new in this bound after we have Theorem 7? The largest bound Theorem 7 can give in depth 3 is $\exp(\Omega(\sqrt{n}))$ achieved by the Parity function. On the other hand, it is known ([18]) that almost all functions

[2] If $X_1, \ldots, X_m$ are mutually independent random variables with $\text{Prob}[X_i = +1] = \text{Prob}[X_i = -1] = 1/2$ then for every $a > 0$, $\text{Prob}[|X_1 + \cdots X_m| > a] < 2e^{-a^2/2m}$.

require depth 3 AC^0 circuits of size $2^n/\log n$. Thus, our seemingly good understanding of AC^0 circuits is not complete: we know why the Parity function is AC^0-hard but we do not know why other (in fact, most) functions are harder. Theorem 8 allows (at least potentially) to achieve this bound. For random function $\mathsf{f}: \mathtt{E}_2^n \to \{0,1\}$, Theorem 5 gives

$$\Delta_k(\mathsf{f}) \geq \sqrt{\frac{2^{n-k}}{2\binom{n}{k}}}$$

which by Theorem 8 gives the lower bound $\exp\left(n^{1-o(1)}\right)$ for f.

Open Problem 9 *Exhibit an explicit sequence of functions* $f_n : \mathtt{E}_2^n \to \{0,1\}$ *with* $\Delta_k(f_n) \geq \exp(n^\alpha)$ *for* $k = n^\alpha$ *with* $\alpha > 1/2$.

In order to estimate $\Delta_k(f)$ for $k > 1$, we have to look at the behaviour of f on the class of k-closed sets which is much richer than that of subcubes. It was shown in [13] that appropriate modification of well-known Erdós-Rado lemma about "sunflowers" [8] can help to construct limits for $k \gg 1$. It appears, however, that the same result can be achieved using Sipser's topological argument.

Let us, for example, prove $\Delta_k(T_s^n) \geq \binom{n}{s} \cdot k^{-s}$ where $T_s^n(a) = 1$ iff $|a| \geq s$; $|a|$ is the number of ones in a. Take any partition of $\mathtt{E}_2^n$ into k-closed blocks, and suppose that the number of blocks is $r < \binom{n}{s} \cdot k^{-s}$. What we need is to prove that then $T_s^n \neq const$ on at least one the blocks. At least one of the blocks B contains at least $\binom{n}{s}/r > k^s$ vectors from the s-th slice The function accepts all these vectors, hence it is forced to accept *all* the vectors in the block B (since T_s^n is constant on B). At this step we need a "diagonalization" lemma (Lemma 10 bellow) stating that any set A of more than k^s vectors from the s-th slice has a k-limit b with less than s ones. Then $T_s^n(b) = 0$. But $b \in B$ since B is k-closed, and hence, T_s^n must accept this vector b, a contradiction.

Lemma 10 *Let* $A \subseteq \{0,1\}^n$. *Suppose that* A *has more than* k^s *vectors with exactly* s *ones and no vectors with less than* s *ones. Then* A *is not* k*-closed.*

This lemma was proved in [13] by induction similar to that used in the Sunflower Lemma. Let us give another proof in order to stress the similarity[3] with Sipser's topological argument [23] (see also Section 2).

[3] This is of no means the "finitized" version of the argument in [23] one may expect since we do not have non-deterministic variables here. The presence of nondeterministic bits makes our knowledge about the set A less definite.

Proof. Assume to the contrary that A is k-closed. Then $A = \Phi^{-1}(1)$ for some generalized k-CNF Φ. Consider the set $A_0 \rightleftharpoons \{a \in A : |a| = s\}$. For $B \subseteq A_0$ and $a \in \{0,1\}^n$ we say that B is *large at* a if $b \geq a$ for all $b \in B$, and $|B| > k^{s-|a|}$. Note that A_0 itself is large at $\vec{0} \rightleftharpoons (0,\ldots,0)$. Our goal is to construct a sequence of sets $A_0 \supseteq A_1 \supseteq \cdots$ and a sequence of vectors $\vec{0} = a^0, a^1, \ldots$ such that $|a^i| = i$ and A_i is large at a^i. As in Sipser's proof we perform a construction in stages.

Stage i ($i < s-1$): Since $|a^i| = i < s$, $a^i \notin A$. Hence,the vector a^i does not pass some test $\phi(x_{j_1},\ldots,x_{j_k})$ from Φ. Since all vectors in A_i pass this test, every vector in A_i must differ from a^i is at least one of the coordinates $S = \{j_1,\ldots,j_k\}$. By largeness of A_i, $b \geq a^i$ for all $b \in A_i$. Hence, for each $b \in A_i$ there is a $j \in S$ such that the j-th coordinate of b is 1 whereas that of a^i is 0. There must be, therefore, one coordinate $j \in S$ which corresponds to at least $1/k$ fraction of vectors in A_i. Let A_{i+1} be this fraction, and a^{i+1} be a^i with the j-th coordinate replaced by 1. The set A_{i+1} is large at a^{i+1} since $|A_{i+1}| \geq |A_i|/k > k^{s-|a^i|-1} = k^{s-|a^{i+1}|}$. Go to Stage $i+1$.

Upon completion of all $s-2$ stages we obtain a vector a^{s-1} with $s-1$ ones. We claim that this vector passes *all* the tests in Φ, which gives the desired contradiction. The argument is the same as above. Assume the opposite that $\phi(a^{s-1}) = 0$ for some ϕ from Φ. Let S be the support of ϕ. Since A_{s-1} is large at a^{s-1}, $|A_{s-1}| > k^{s-|a^{s-1}|} = k$. Each vector in A_{s-1} has its own position were it has 1 and a^{s-1} has 0; in all other positions it coincides with a^{s-1}. But we have more than $k = |S|$ vectors in A_{s-1}. Hence, at least one $b \in A_{s-1}$ must coincide with a^{s-1} on S. But then $\phi(b) = \phi(a^{s-1}) = 0$ whereas, by the definition, all the vectors in A_{s-1} (and hence, the vector b) pass all the tests in Φ, a contradiction. ∎

7 Limits for $k \approx n$

Given a function $f : \mathtt{E}_q^n \to \{0,1\}$ with $q \geq 2$, define its *boolean version* $\hat{f} : \mathtt{E}_2^{mn} \to \{0,1\}$ with $m = \lceil \log q \rceil + 1$, as follows. Fix the standard encoding of integers in $\mathtt{E}_q = \{0,1,\ldots,q-1\}$ by 0,1-strings of length m (zeros allowed on the left). The function $\hat{f}$ has mn boolean inputs; given a boolean vector in $\{0,1\}^{mn}$, the function $\hat{f}$ first makes the partition of its variables into n subsequent blocks of length m each, looks each block as representing an integer in $\mathtt{E}_q$ and computes the value of f on these integers.

Theorem 11 *Let $f : \mathbb{E}_q^n \to \{0,1\}$. Suppose that the boolean version of f can be represented as*

$$\hat{f} = SYM(h_1, \ldots, h_L) \tag{3}$$

where SYM is any symmetric boolean function and each h_i is an arbitrary boolean function on at most $n-1$ variables. Then

$$L \geq (\Delta_{n-1}(f))^{1/n} - 1.$$

Proof. Partition the set of mn (boolean) variables of $\hat{f}$ into n equal-size blocks and let H_i be the set of those functions h_j which have *no* variables in the i-th block. (This is possible since each h_j has only $n-1$ variables and there are n blocks). Since SYM is symmetric function, the value of $\hat{f}$ is uniquely determined by the value of the sum $F = F_1 + ... + F_n$ of n functions $F_i = \sum_{h \in H_i} h$. Important here is that for each $i \in [n]$, the function F_i *does not depend* on coordinates in the i-th block. Look at these sums F_i as functions on $\mathbb{E}_q^n$. That is, each F_i has n variables (ranging over $\mathbb{E}_q^n$), and does not depend on the i-th one.

Now assign to each vector $a \in \mathbb{E}_q^n$ the color $\langle F_1(a), \ldots, F_n(a) \rangle$. Since each sum F_i can take at most $L+1$ different values, we need at most $(L+1)^n$ colors to color all the vectors in $\mathbb{E}_q^n$. Clearly, each color class (i.e. each set of vectors in $\mathbb{E}_q^n$ with the same color) is $(n-1)$-closed, and on each of them the function f is constant. Therefore, $\Delta_{n-1}(f) \leq (L+1)^n$ which gives the desired lower bound on L. ∎

Let us say here several words on why Theorem 11 is interesting. Beigel and Tarui [3] have shown (based on earlier results of Toda [24], Allender [1] and Yao [27]) that any boolean function $\hat{f}$ on N boolean variables in ACC (the class of all functions which have bounded depth polynomial size circuits with AND, OR, NOT and MOD_r gates) can be represented [4] in the form (3) with $L = \exp\left((\log N)^{O(1)}\right)$ and $n = (\log N)^{O(1)}$.

Open Problem 12 *Exhibit an explicit sequence of functions $f_n : \mathbb{E}_q^n \to \{0,1\}$ with $\Delta_{n-1}(f) \geq q^\alpha$ for $n = (\log\log q)^{\omega(1)}$ and $\alpha \geq n/\log q$.*

By Theorem 11, for any such function, the corresponding sequence $\hat{f}_n$ of boolean versions is *not* in ACC.

[4] Green, Köbler and Torán [11] have then showed that the results remains true with one concrete symmetric function, namely, with MidBit function $f(x_1, \ldots, x_n)$ which computes 1 iff the middle bit in the binary representation of $\sum x_i$ is 1.

Theorem 5 says that for random f, $\Delta_{n-1}(\mathsf{f}) \geq \sqrt{q/2n}$, which is more than enough. The problem is, of course, to prove similar (or even much weaker) lower bound on $\Delta_{n-1}(f)$ for an *explicit* function f. We have seen above how this can be done for moderate values of k. For large k's the situation is harder, and so far we have only two non-trivial lower bounds for particular explicit functions $f : \mathtt{E}_q^n \to \{0,1\}$:

$$\Delta_{n-1}(f) \to \infty \text{ for each fixed } n \text{ as } q \to \infty \qquad \text{(CFL bound)}$$

proved by *Chandra, Furst* and *Lipton* [7], and

$$\Delta_{n-1}(f) \geq q^{\Omega(1/4^n)} \qquad \text{(BNS bound)}$$

proved by *Babai, Nissan* and *Szegedy* in [2]. The first proof is an application of Gallai's theorem from Ramsey theory. The second derives an upper bound on the discrepancy $\mathrm{disc}_{n-1}(f)$ of f on $(n-1)$-closed sets; the inequality (1) then gives the desired lower bound on $\Delta_{n-1}(f)$. Let us look at these techniques in more details.

7.1 The CFL Bound: Coloring Spheres

Let us first observe that, when dealing with $\Delta_{n-1}(f)$, we are actually dealing with special colorings of spheres in $\mathtt{E}_q^n$.

A *sphere around* $a \in \mathtt{E}_q^n$ is a set S of n vectors

$$S = \{(a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n) \,:\, b_i \neq a_i, i = 1, \ldots, n\}.$$

The sphere[5] S is specified by two vectors: the *center* a, and the *character* $b = (b_1, \ldots, b_n)$; they must differ in all coordinates. Thus, there are exactly $(q-1)^n$ spheres around each vector, and each vector belongs to at most $n(q-1)$ spheres.

Say that a coloring $\chi : \mathtt{E}_q^n \to [r]$ *respects* a function $f : \mathtt{E}_q^n \to \{0,1\}$ if it uses different colors for vectors in $f^{-1}(0)$ and $f^{-1}(1)$ i.e if $f(a) \neq f(b)$ implies $\chi(a) \neq \chi(b)$.

Proposition 13 *$\Delta_{n-1}(f)$ is the minimal number r for which there exists an r-coloring $\chi : \mathtt{E}_q^n \to [r]$ such that*

(1) χ respects f, and

(2) if some sphere is monochromatic then its center has the same color.

[5] A sphere should not be mixed with *Hamming sphere of radius one* which is uusally defined as a set of all $(q-1)^n$ vectors $(a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n)$ with $1 \leq i \leq n$ and $b_i \in \mathtt{E}_q^n - \{a_i\}$.

Proof. A set $A \subseteq \mathbb{E}_q^n$ is $(n-1)$-closed iff A contains no sphere with center outside A. ∎

EXAMPLE 1. Take $n = 2$. The hypercube $\mathbb{E}_q^2$ is a complete graph on q vertices. Define $f : \mathbb{E}_q^2 \to \{0,1\}$ by: $f(x,y) = 1$ iff $x = y$. Then $\Delta_1(f) = q$. To see this take any coloring $\chi : \mathbb{E}_q^2 \to [q-1]$. By Pigeonhole principle, at least one color class $\chi^{-1}(i)$ contains two edges (x,x) and (y,y) with $x \neq y$. These two vectors form a sphere around the vector (x,y). By (2), $\chi(x,y) = \chi(x,x)$ which means that χ does not respect f.

EXAMPLE 2. Fix a vector $a \in \mathbb{E}_q^n$, and let $\mathcal{S}$ be the set of all spheres around a. We have many such spheres, namely, $|\mathcal{S}| = (q-1)^n$. How many colors we need in order to leave no of these spheres monochromatic? The answer is: two colors are enough. Define the coloring $\chi : \mathbb{E}_q^n \to [2]$ by $\chi(x) = 0$ if x differs from a exactly in the first coordinate, and $\chi(x) = 1$ otherwise.

EXAMPLE 3. Consider the hyperplane $P_{q,n} \rightleftharpoons \{(x_1, \ldots, x_n) : x_1 + \ldots + x_n = q\}$ in $\mathbb{E}_q^n$, and let $f_{q,n}$ be the characteristic function of this hyperplane. It is proved in [7] that

$$\Delta_{n-1}(f_{q,n}) \to \infty \text{ for each fixed } n \text{ as } q \to \infty \tag{4}$$

Proof. Let us say that a sphere $S \subseteq \mathbb{E}_q^n$ around a vector a is *nice* if its character is $a + \lambda = (a_1 + \lambda, \cdots, a_n + \lambda)$ for some $\lambda \neq 0$. A *ball* is a sphere with its center. Gallai's theorem (see, e.g. [10]) says that:

> *For every r and n there exists a $p = p(n,r)$ such that every r-coloring of $\mathbb{E}_p^n$ leaves at least one nice ball monochromatic.*

With this theorem at hand, (4) can be derived as follows. Fix an arbitrary $n \in \mathbb{N}$ and assume to the contrary that there exists an r such that $\Delta_{n-1}(f_{q,n}) \leq r$ for all $q \in \mathbb{N}$. By Proposition 13, for every q there exists a coloring $\chi : P_{q,n} \to [r]$ which leaves *no* sphere in the plane $P_{q,n}$ monochromatic (since centers of spheres in $P_{q,n}$ lie outside $P_{q,n}$). Take p large enough and $q \rightleftharpoons np$. Then there is a bijection $\nu : \mathbb{E}_p^{n-1} \to P_{q,n}$ which sends a vector $a = (a_1, \ldots, a_{n-1}) \in \mathbb{E}_p^{n-1}$ to the vector $(a_1, \ldots, a_{n-1}, q - \sum a_i) \in P_{q,n}$. This bijection and the coloring $\chi : P_{q,n} \to [r]$ induce the r-coloring χ' of $\mathbb{E}_p^{n-1}$ by $\chi'(a) \rightleftharpoons \chi(\nu(a))$.

By Gallai's theorem, there exists an integer $p = p(n,r)$ such that every r-coloring $\chi' : \mathbb{E}_p^{n-1} \to [r]$ leaves at least one nice ball $B \subseteq \mathbb{E}_p^{n-1}$ monochromatic. Simple but crucial observation is that every nice ball in the hypercube $\mathbb{E}_p^{n-1}$ corresponds to a sphere in the hyperplane $P_{q,n}$. Namely, if $a \in \mathbb{E}_p^{n-1}$ is the center and $a + \lambda$ the character of the ball B, then the corresponding

sphere S around the vector $a = (a_1, \ldots, a_{n-1}, q - \sum a_i - \lambda)$ with the character $a + \lambda$ lies in $P_{q,n}$. By our assumption, the coloring χ leaves no sphere in $P_{q,n}$ monochromatic. That is $\chi(S) \not\equiv const$. But then also $\chi'(B) \not\equiv const$, a contradiction with Gallai's theorem. ∎

7.2 The BNS Bound: Bounding the Discrepancy

Recall that the discrepancy, $\mathrm{disc}_A(f)$, of a function $f : \mathtt{E}_q^n \to \{-1, 1\}$ on a set $A \subseteq \mathtt{E}_q^n$ is the absolute value of the sum $\sum_{x \in A} f(x)$, and that $\Delta_k(f) \geq q^n / \mathrm{disc}_k(f)$ where $\mathrm{disc}_k(f)$ is the maximum of $\mathrm{disc}_A(f)$ over all k-closed subsets $A \subseteq \mathtt{E}_q^n$ (see Section 4).

For an index $i \in [n]$ and a constant $u \in \mathtt{E}_q$, let $f_{i \to u} \rightleftharpoons f(x_1, \ldots, x_{i-1}, u, x_{i+1}, \ldots, x_n)$ denote the corresponding subfunction of f. If f depends on n variables then $f_{i \to u}$ depends on at most $n - 1$ variables. The simplest way to estimate the discrepancy is to do this by induction on the number of variables n. The basic trick used in [2] is actually the following recursion.

Lemma 14 *For every function $f : \mathtt{E}_q^n \to \{-1, 1\}$ and every index $i \in [n]$,*

$$\mathrm{disc}_{n-1}(f) \leq \left[q^{n-1} \sum_{u,v \in \mathtt{E}_q} \mathrm{disc}_{n-2}(f_{i \to u} \cdot f_{i \to v}) \right]^{1/2}. \tag{5}$$

Proof. Take a $(n-1)$-closed set $A \subseteq \mathtt{E}_q^n$ for which $\mathrm{disc}_{n-1}(f) = \mathrm{disc}_A(f)$, and let $\Phi = \prod_{j=1}^{n} \phi_j$ be the corresponding generalized $(n-1)$-CNF. The only we know about these functions is that ϕ_j ($j \in [n]$) *does not* depend on the j-th variable and that $A = \Phi^{-1}(1)$. Then

$$\mathrm{disc}_{n-1}(f) = \mathrm{disc}_A(f) = \left| \sum_{u \in \mathtt{E}_q} \sum_{\vec{x}} f_{i \to u}(\vec{x}) \cdot \Phi_{i \to u}(\vec{x}) \right| = \sum_{\vec{x}} h(x)$$

where for each $\vec{x} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathtt{E}_q^{n-1}$,

$$h(\vec{x}) = \phi_i(\vec{x}) \left| \sum_u f_{i \to u}(\vec{x}) \cdot \Phi'_{i \to u}(\vec{x}) \right| \leq \left| \sum_u f_{i \to u}(\vec{x}) \cdot \Phi'_{i \to u}(\vec{x}) \right|$$

and $\Phi' = \prod_{j \neq i} \phi_j$.

By Couchy-Schwarz inequality

$$[\mathrm{disc}_{n-1}(f)]^2 = \left[\sum_{\vec{x}} h(\vec{x}) \right]^2 \leq q^{n-1} \sum_{\vec{x}} h(\vec{x})^2 \leq q^{n-1} \sum_{u,v} \left| \sum_{\vec{x}} f_{i \to u} \cdot f_{i \to v} \cdot \Phi'_{i \to u} \cdot \Phi'_{i \to v} \right|.$$

where, for each pair $u, v \in \mathtt{E}_q$, the inner sum is $\leq \mathrm{disc}_{n-2}(f_{i\to u} \cdot f_{i\to v})$ since the function $\Phi'_{i\to u} \cdot \Phi'_{i\to v}$ is a generalized $(n-2)$-CNF, and hence, defines an $(n-2)$-closed subset of $\mathtt{E}_q^{n-1}$. ∎

Let, as before, $m = \lceil \log q \rceil + 1$. For an integer $x \in \mathtt{E}_q$, let $\hat{x} \subseteq [m]$ denote the set of those positions on which the standard binary code of x has 1. Keeping this bijection $\mathtt{E}_q \ni x \longleftrightarrow \hat{x} \in [m]$ in mind, the function "generalized inner product" $f_{m,n} : \mathtt{E}_q^n \to \{-1, 1\}$ is defined by: $f_{m,n}(x_1, \ldots, x_n) = 1 \iff |\hat{x}_1 \cap \cdots \cap \hat{x}_n| \equiv 1 (\mathrm{mod}\ 2)$. That is, $f_{m,n}$ computes 1 iff the number of positions where all the (binary codes of) inputs have 1, is odd. For $f \rightleftharpoons f_{m,n}$ and any two constants $u, v \in \mathtt{E}_q$, the product $f_{i\to u} \cdot f_{i\to v}$ coincides[6] with $f_{m_{u,v},n-1}$ where $m_{u,v} = |\hat{u} \setminus \hat{v}| + |\hat{v} \setminus \hat{u}|$. Moreover, for each $i \in [m]$, there are exactly $\binom{m}{i}$ pairs u, v for which $m_{u,v} = i$. Using this property and the recursion (5), one obtains the recursion

$$\mathrm{disc}_{n-1}(f_{m,n}) \leq \left[q^{n-1} \sum_{i=0}^{m} \binom{m}{i} \mathrm{disc}_{n-2}(f_{i,n-1}) \right]^{1/2}. \tag{6}$$

This gives the estimate $\mathrm{disc}_{n-1}(f_{m,n}) \leq (\mu_n)^m$ where μ_n is given by the recursion: $\mu_1 \rightleftharpoons 0$ (since $f_{m,1}(x)$ is simply the parity of x) and $\mu_n \rightleftharpoons \sqrt{2^{n-1}(1+\mu_{n-1})}$. This is because by (6),

$$\mathrm{disc}_{n-1}(f_{m,n}) \leq \left[q^{n-1} \sum_{i=0}^{m} \binom{m}{i} (\mu_{n-1})^i \right]^{1/2} = \left[2^{m(n-1)} (1+\mu_{n-1})^m \right]^{1/2} = (\mu_n)^m .$$

It is easy to verify by induction on n that $\mu_n \leq 2^n - 2^{-n}$. Thus,

$$\begin{aligned} \Delta_{n-1}(f_{m,n}) &\geq q^n / \mathrm{disc}_{n-1}(f_{m,n}) \geq 2^{mn} / (\mu_n)^m \geq \left(1 - 2^{-2n}\right)^{-m} \\ &\geq \exp\left(\Omega(m/4^n)\right) = q^{\Omega(1/4^n)}. \end{aligned}$$

8 Limits for Small Families $\mathcal{F}$

Recall that k-limits are $\mathcal{F}$-limits with $\mathcal{F}$ being the whole family of k-subsets of $[n]$. Let us now look what happens if we take only small families.

Definition. For a function $f : \mathtt{E}_q^n \to \{0, 1\}$, let $\Delta_{k,\ell}(f)$ denote the minimal possible number of blocks in a partition of $\mathtt{E}_q^n$ such that f is constant on each block and

[6] This is the main and the only (!) property of the Generalized Inner Product used. Thus, similar lower bounds of $\Delta_{n-1}(f)$ can be proved in the same way for any function f which has products $f_{i\to u} \cdot f_{i\to v}$ among its subfunctions.

$(*)$ each block is $\mathcal{F}$-closed for some $\mathcal{F} \subseteq [n]^k$ with $|\mathcal{F}| = \ell$.

The *degree* of a family $\mathcal{F} \subseteq [n]^k$ is the maximal number of sets in $\mathcal{F}$ with non-empty intersection. Let $\Delta_k^r(f)$ denote the corresponding measure with $(*)$ replaced by

$(**)$ each block is $\mathcal{F}$-closed for some $\mathcal{F} \subseteq [n]^k$ with $\deg(\mathcal{F}) \leq r$.

Following the proof of Theorem 1 in [6], one can derive the following reduction for branching programs. The *depth* of a branching program is the length of a longest s-t path. A program is *read-r-times* if each variable in each s-t path appears no more than r times.

Theorem 15 *Suppose that $f : \mathtt{E}_2^n \to \{0,1\}$ has a non-deterministic branching program of size L and depth T. Then, for any k,*

$$L \geq \left(\Delta_{k,\ell}(f)\right)^{1/\ell}$$

where $\ell = T/k$. Moreover, if the program is read-r-times then the same holds with $\Delta_{k,\ell}(f)$ replaced by $\Delta_k^r(f)$.

Proof. (Sketch). Let P be a non-deterministic branching program which computes f and has size L and depth T. With each s-t path p in P associate the sequence $w_0, \ldots, w_\ell$ ($\ell = T/k$) of its nodes (the "trace" of p): $w_0 = s$, $w_\ell = t$ and for each $1 \leq i < \ell$, the node w_i is selected as follows. Take w_i to be the first node u of an edge (u,v) in p after w_{i-1}, for which the following holds: the function $\phi_{w_{i-1},u}$ computed by a subprogram of P with w_{i-1} as start and u as final node, depends on $< k$ variables and $\phi_{w_{i-1},v}$ depends on $\geq k$ variables. In such a way, each trace defines a generalized k-CNF $\phi_1 \wedge \cdots \wedge \phi_\ell$. There are at most L^ℓ traces. Therefore, f is an OR of at most L^ℓ generalized k-CNF's, and thus, $\Delta_{k,\ell}(f) \leq L^\ell$ which gives the desired lower bound on L. ∎

One of (many) long-standing problems in complexity theory is to exhibit an explicit function f, any non-deterministic log-space machine computing which should require super-linear time. By Theorem 15, this could be resolved by proving good lower bounds on $\Delta_{k,\ell}(f)$. If f has a non-deterministic log-space machine then f has a non-deterministic branching program of size $L \leq n^{O(1)}$. The time T is the depth of this program. By Theorem 15, $T \geq k \cdot \ell$ for all k, ℓ satisfying $\ell = \Omega\left(\frac{\log \Delta_{k,\ell}(f)}{\log n}\right)$.

Open Problem 16 *Exhibit an explicit sequence of functions* $f_n : \mathbb{E}_2^n \to \{0,1\}$ *such that* $\ell = \Omega\left(\frac{\log \Delta_{k,\ell}(f)}{\log n}\right)$ *for some* k, ℓ *satisfying* $k \cdot \ell = \omega(n)$. By Theorem 15, any non-deterministic log-space machine computing such f_n requires time $T = \omega(n)$.

Thus, here both Δ and k must be large. The only relaxation is that we have to consider only small subfamilies $\mathcal{F}$ of $[n]^k$. We need therefore arguments for the following generalized "diagonalization problem":[7]

> *Given integers* k *and* ℓ, *exhibit an explicit set* $U \subseteq \mathbb{E}_q^n$ *such that in every partition of* U *into "small" number of blocks, at least one block is* $\mathcal{F}$*-closed for* no *family* $\mathcal{F} \subseteq [n]^k$ *with* $|\mathcal{F}| = \ell$.

It is not clear how to diagonalize using only the knowledge that $\mathcal{F}$ has few sets. This can be done, however, if we know more. Namely, if these sets do not overlap much, i.e. if $\mathcal{F}$ has sufficiently small degree.

For $A \subseteq \mathbb{E}_2^n$, let $\mathrm{dist}(A)$ be the minimal Hamming distance between two different vectors in A. For any $d \in \mathbb{N}$ there are explicit sets $U \subseteq \{0,1\}^n$ with $\mathrm{dist}(U) = 2d+1$ and $|U| \geq 2^n/(n+1)^d$ (BCH codes, for example). The lemma bellow says that *no* large subset of U is closed w.r.t. families of small degree, and hence, $\Delta_k^r(U)$ must be large[8]. Similar "diagonalization lemma" for small degree families was used in [14] to prove that some linear code functions require exponential the size non-deterministic read-r-times branching programs.

Lemma 17 *Let* $A \subseteq \mathbb{E}_2^n$, $\mathcal{F} \subseteq [n]^k$ *and* $r = \deg(\mathcal{F}) \leq n$. *Suppose that*

$$|A| > \frac{2^n}{\binom{\alpha n}{d}\binom{\beta n}{d}} \tag{7}$$

with $\alpha = \left(\frac{k}{en}\right)^r$, $\beta = 1 - \frac{kr}{n}$ *and* $d = (\mathrm{dist}(A)-1)/2$. *Then* A *is not* $\mathcal{F}$*-closed.*

Proof. Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ and assume to the contrary that A is $\mathcal{F}$-closed. Note that $m \leq n \cdot \deg(\mathcal{F})/k = nr/k$ because each element of $[n]$ belongs to at most $\deg(\mathcal{F})$ sets of $\mathcal{F}$. The proof consists of two stages. We first prove that A must be $\mathcal{H}$-closed for some family $\mathcal{H}$ containing only two

[7] Of course, this makes sense only if, for some $\mathcal{F} \subseteq [n]^k$ with $|\mathcal{F}| \ll \binom{n}{k}$, we can prove something better as for $\mathcal{F} = [n]^k$, the case considered in previous sections.

[8] Namely, $\Delta_k(U) \geq \left(\frac{\alpha\beta n}{d^2}\right)^d$ with parameters α and β defined in Lemma 17

"sufficiently disjoint" sets, and prove that it is not possible if $|A|$ is large enough to satisfy (7).

Our first goal is to show that A is $\mathcal{H}$-closed for some family $\mathcal{H} = \{S, T\}$ containing only two sets $S, T \subseteq [n]$ such that $|S \setminus T| \geq \alpha n$ and $|T \setminus S| \geq \beta n$. Take uniformly at random a subset $I \subseteq \{1, \ldots, m\}$ with $|I| = r = \deg(\mathcal{F})$, and consider two (random) sets $S = \cup_{i \in I} S_i$ and $T = \cup_{i \notin I} S_i$. Since each point $x \in [n]$ can belong to at most r of the sets $S_1, \ldots, S_m$, this point belongs to $S \setminus T$ with probability at least $\binom{m}{r}^{-1} > \alpha$. This implies that the mean of $|S \setminus T|$ is at least αn. Fix a set I for which $|S \setminus T| \geq \alpha n$. Since $|S| \leq kr$, we conclude that $|T \setminus S| = n - |S| \geq \beta n$. Moreover, A is $\{S, T\}$-closed.

Split now the set A into $p \leq 2^{|S \cap T|}$ blocks $A_1, \ldots, A_p$ such that all the vectors in one block coincide on $S \cap T$. Let $\Gamma_d(n)$ be the number of vectors in a Hamming ball of radius d in $\mathbb{E}_2^n$, i.e. $\Gamma_d(n) = 1 + \sum_{i=1}^{d} \binom{n}{i} > \binom{n}{d}$. Any two vectors of A differ in at least $2d+1$ positions. Hence, if we fix the values of some t coordinates then no more than $2^{n-t}/\Gamma_d(n-t)$ of all 2^{n-t} possible extensions can be in A. Since the whole set A was $\{S, T\}$-closed, each block A_i is $\{S \setminus T, T \setminus S\}$-closed. Thus, for every i,

$$|A_i| \leq \left|A_i|_S\right| \cdot \left|A_i|_T\right| \leq \frac{2^{|T \setminus S|}}{\Gamma_d\left(|T \setminus S|\right)} \cdot \frac{2^{|S \setminus T|}}{\Gamma_d\left(|S \setminus T|\right)},$$

and hence, $|A| = \sum_{j=1}^{p} |A_i| \leq 2^n / \Gamma_d(\alpha n)\Gamma_d(\beta n)$, a contradiction with (7). □

Acknowledgment. I am grateful to Michael Sipser for telling me the notion of finite limit and Johan Håstad and Pavel Pudlák for many fruitful discussions.

References

[1] E. Allender. A note on the power of threshold circuits. in: *Proc. 30th IEEE Symp. on Foundations of Computer Science* 1989, pp. 580-584.

[2] L. Babai, N. Nisan, M. Szegedy. Multiparty protocols, pseudorandom generators for Logspace, and time-space trade-offs. *Journal of Comput. Syst. Sci.*, vol. 45 (1992), pp. 204-232.

[3] R. Beigel, J. Tarui. On ACC. in: *Proc. 32nd IEEE Symp. on Foundations of Computer Science* 1991, pp. 783-792.

[4] S. Ben-David, M. Karchmer, E. Kushilevitz. On ultrafilters and NP. In *Proc. of the 26th ACM Symposium on Theory of Computing*, (1994) (to appear)

[5] R. Boppana, M. Sipser. The complexity of finite functions. In: *The Handbook of Theoretical Computer Science*, (J. van Leeuwen, ed.), Elsvier Science Publischers B.V., 1990, pp. 759-804.

[6] A. Borodin, A. Razborov and R. Smolensky, On lower bounds for read-k times branching programs, *Computational Complexity*, **3** (1993), 1-18.

[7] A. K. Chandra, M. L. Furst, R. J. Lipton. Multi-party protocols. In *Proc. of the 15th ACM Symposium on Theory of Computing*, (1983), pp. 94-99.

[8] P. Erdös and R. Rado. Intersection theorems for systems of sets, *J. London Math. Soc.* **35** (1960), pp 85-90.

[9] M. Furst, J. Saxe, M. Sipser. Parity, circuits and the polynomial time hierarchy, *Math. Systems Theory* **17** (1984), pp. 13-27.

[10] R. Graham, B. Rothschild, J. Spencer. *Ramsey Theory*. 2nd ed., Wiley, New-York, 1990.

[11] F. Green, J. Köbler, J. Torán. The power of Middle Bit. In: *Proceedings of the 7th Annual Symposium on Structure in Complexity Theory,* 1992, pp. 111-117.

[12] J. Håstad. *Almost Optimal Lower Bounds for Small Depth Circuits,* Advances in Computing Research, (1987), Vol 5, pp 143-170.

[13] J. Håstad, S. Jukna and P. Púdlak, Top-down lower bounds for depth 3 circuits, in: *Proc. 34th IEEE Symp. on Foundations of Computer Science* 1993, pp. 124-129.

[14] S. Jukna. *A note on read-k times branching programs.* Technical Report 448, Universität Dortmund, 1992. (To appear in: *RAIRO Theoretical Informatics and Applications*).

[15] M. Karchmer. On proving lower bounds for circuit size. In: *Proceedings of the 8th Annual Symposium on Structure in Complexity Theory,* (1993), pp. 112-118.

[16] M. Karchmer, A. Wigderson. On span programs. In: *Proceedings of the 8th Annual Symposium on Structure in Complexity Theory,* (1993).

[17] M. Karchmer, A. Wigderson. Characterizing non-deterministic circuit size. In *Proc. of the 25th ACM Symposium on Theory of Computing*, (1993)

[18] O. B. Lupanov. Implementing the algebra of logic functions in terms of bounded depth formulas in the basis $\{\wedge, \vee, \neg\}$. *Soviet Phys. Doklady,* 6 (1961), pp. 750-752.

[19] A. Razborov. Lower bounds for the monotone complexity of some Boolean functions. *Soviet Math. Dokl.*, 31 (1985), pp. 354-357.

[20] A. Razborov. On the method of approximations. In *Proc. of the 21st ACM Symposium on Theory of Computing*, (1989), pp. 167-176.

[21] A. Razborov. Lower bounds on the size of switching-and-rectifier networks for symmetric Boolean functions (in Russian). *Math. Notes of the Academy of Sciences of the USSR,* 48(6) (1990), pp. 79-91.

[22] M. Sipser. *Private communication* (August, 1991)

[23] M. Sipser. A topological view of some problems in complexity theory. In *Colloquia Mathematica Societatis János Bolyai* **44** (1985), pp 387-391.

[24] S. Toda. On the computational power of PP and $\oplus$P. in: *Proc. 30th IEEE Symp. on Foundations of Computer Science* 1989, pp. 514-519.

[25] A. Wigderson. The fusion method for lower bounds in circuit complexity. *Combinatorics, Paul Erdös is Eighty* (Vol 1), Miklós, Sós ands Szönyi (Eds.), Bolyay Math. Society, (1993), pp. 453-468.

[26] A.C. Yao. Separating the polynomial time hierarchy by oracles. In: *Proc. 26th Ann. IEEE Symp. on Foundations of Computer Science* (1985), pp. 1-10.

[27] A.C. Yao. On ACC and threshold circuits. In: *Proc. 31st Ann. IEEE Symp. on Foundations of Computer Science* 1990, pp. 619-627.