Small Pseudo-Random Sets Yield Hard Functions: New Tight Explicit Lower Bounds for Branching Programs

Alexander E. Andreev
University of Moscow
andreev@mntn.msk.su

Juri L. Baskakov
University of Moscow
baskakov@mntn.msk.su

Andrea E. F. Clementi
University of Rome *La Sapienza*
clementi@dsi.uniroma1.it

José D. P. Rolim*
University of Geneva
rolim@cui.unige.ch

Abstract

In several previous works the explicit construction of a computationally-hard function with respect to a certain class of algorithms or Boolean circuits has been used to derive small pseudo-random spaces. In this paper, we revert this connection by presenting two new direct relations between the efficient construction of pseudo-random (both two-sided and one-sided) sets for Boolean affine spaces and the explicit construction of Boolean functions having hard *branching program* complexity.

In the case of 1-read branching programs (1-$Br.Pr.$), we show that the construction of non trivial (i.e. of cardinality $2^{o(n)}$) *discrepancy sets* (i.e. two-sided pseudo-random sets) for Boolean affine spaces of dimension greater than $n/2$ yield a set of explicit Boolean functions having very hard 1-$Br.Pr.$ size. Then, by combining the previous constructions of ϵ-biased sample spaces for linear tests and a simple "Reduction" Lemma, we derive the required discrepancy set and obtain a Boolean function in P having 1-$Br.Pr.$ size not smaller than $2^{n-O(\log^2 n)}$ and a Boolean function in $\mathsf{DTIME}(2^{O(\log^2 n)})^{\mathsf{NP}}$ having 1-$Br.Pr.$ size not smaller than $2^{n-\log 4n}$. This bound is optimal and improves over the best previously known lower bound that was $2^{n-3n^{1/2}}$ [24].

As for the more general case of non deterministic, syntactic k-read branching programs (k-$Br.Pr.$), we introduce a new method to derive explicit, exponential lower bounds that involves the efficient construction of *hitting sets* (i.e. one-sided pseudo-random sets) for affine spaces of dimension $o(n/2)$. Using an appropriate "orthogonal" representation of small Boolean affine spaces, we provide these hitting sets thus obtaining an explicit Boolean function in P that has k-$Br.Pr.$ size not smaller than $2^{n^{1-o(1)}}$ for *any* $k = o\left(\frac{\log n}{\log\log n}\right)$. This improves the previous known lower bounds given in [8, 12, 21] for some range of k.

[*] Contact Author: Centre Universitaire Informatique, University of Geneva, 24 rue General Dufour, 1204 geneva (Switzerland), Fax: +41-22-705-7780

1 Introduction

Branching programs

Research in branching programs is extremly active from both theoretical and practical point of view [23, 13, 22, 26, 28]. Branching programs (*Br.Pr.*'s) have often been used as data structures to represent finite Boolean functions. Moreover, the problem of finding lower bounds for the size of *Br.Pr.*'s for explicit Boolean functions has been deeply studied in [8, 12, 22, 23]. Similarly to the case of circuit complexity, it turned out that the task of finding explicit lower bounds in the case of general *Br.Pr.*'s was not carried out very successfully. This intrinsic difficulty led to consider several restrictions, one of the most important ones is that of setting a fixed bound for the maximum number of times that an input variable can be read by a *Br.Pr.*

A *read-k-times Br.Pr.* is allowed to read each variable at most k times along any valid computation and, in a *syntactic read-k-times Br.Pr.*, this reading restriction holds for any path from the source to any sink. Notice that while a *read-1-time Br.Pr.* is always a syntactic *read-1-time Br.Pr.*, for $k \geq 2$ this does not hold. Clearly, the *Br.Pr.* model can also be made *non deterministic* by allowing the existence of more arcs labeled with the same value that leave the same node.

To our knowledge, the best known explicit lower bound for *read-1-time*-branching programs (1-*Br.Pr.*'s) is due to Savicky and Zak [23, 24]; they derived a Boolean function in P having, for any sufficiently large n, 1-*Br.Pr.* size not smaller than 2^{n-s}, where $s = O(n^{1/2})$. We also emphasize that for this model there are no known harder explicit functions even in classes larger than P, while a non explicit lower (and optimal) bound of size $\Theta(2^{n-\log n})$ is known.

Regarding non-deterministic syntactic read-k-times *Br.Pr.*'s (k-*Br.Pr.*'s) when $k \geq 2$, there are explicit lower bounds of exponential size when $k = O(\log n)$ [8, 12, 21]. Borodin *et al* [8] showed an explicit family of Boolean functions such that any k-*Br.Pr.* computing it must have a number of labeled edges not smaller than $\mathsf{exp}(\Omega(n)/(4^k * k^3)))$ when $k \leq c \log n$ for a fixed constant $0 < c < 1$. A lower bound of size $\mathsf{exp}(\Omega(\sqrt{n}/k^{2k}))$ has been independently proved by Okolnishnikova [21]. More recently, Jukna studied the Okolnishnikova's function and obtained an exponential gap between the size of any *Br.Pr.* required for such a function and that required for its complement. Finding exponential lower bounds when k grows faster than any logarithmic function or when the branching programs are non syntactic is still an open question.

Our results

In several previous works [2, 3, 6, 11, 19, 20] the explicit construction of a computationally-hard function with respect to a certain class $\mathcal{C}$ of algorithms (or Boolean circuits) has been used to derive small pseudo-random spaces able to de-randomize a somewhat probabilistic version of $\mathcal{C}$. Here we revert this connection by showing how to use efficient constructions of pseudo-random spaces for linear tests [18] to define a set of explicit functions having very hard complexity with respect to *read-1-time-branching programs* and *non deterministic syntactic read-k-times branching programs.*

Read-1-time-branching programs and discrepancy sets for large affine spaces

In the case of 1-*Br.Pr.*'s, Simon ans Szegedy [25] obtained a valuable combinatorial theorem (see Theorem 3.1) that provides a useful method to obtain lower bounds [23]. In short, our first result is the construction, for any $n, k > 0$ and for any $\alpha \in \{0,1\}^{kn}$, of a finite Boolean function $f^\alpha : \{0,1\}^n \rightarrow \{0,1\}$

for which Simon and Szegedy's Theorem implies the following property: if α is not a solution of a certain set $\mathcal{F}$ of linear systems having $\Theta(n)$ linear equations then f^{α} has hard 1-*Br.Pr.* size (the precise lower bound is a function of some parameters of the systems in $\mathcal{F}$). Therefore, in order to efficiently construct a family of hard functions, we need to derive an α that has the above property for any positive integers k and n.

Since the solution space of any linear system is an affine space, we would be able to efficiently carry out this task if we could construct a family of small *discrepancy sets* for affine spaces. More formally, let $\mathcal{AFF}(n,k,s)$ be the set of all n variables linear systems $Ax = b$ (where $A \in \{0,1\}^{s\times n}$, $x \in \{0,1\}^n$, and $b \in \{0,1\}^s$) of at most s linear functions in which at most k variables appear with non-zero coefficients (such variables are said *essential*). The systems in $\mathcal{AFF}(n,k,s)$ will be simply denoted as pairs (A,b).

Definition 1.1 *Let $\epsilon > 0$. A (multi)set $S \subseteq \{0,1\}^n$ is said to be ϵ-discrepant for $\mathcal{AFF}(n,k,s)$ if for any feasible system $(A,b) \in \mathcal{AFF}(n,k,s)$ with* $\text{rank}(A) = s$*, it holds $|\mathbf{Pr}_{x\in S}(Ax = b) - 2^{-s}| \le \epsilon$.*

Note that 2^{-s} equals the probability that x is a solution of (A,b) when x is chosen uniformly at random from $\{0,1\}^n$.

The case in which the linear system has only one equation, i.e. the classes $\mathcal{AFF}(n,k,1)$, has been studied extensively in the literature: in particular, Naor and Naor [18] introduced the following definition.

Definition 1.2 *A subset $S \subseteq \{0,1\}^n$ is* ϵ-discrepant *for* linear functions *if, for any Boolean linear function $f : \{0,1\}^n \to \{0,1\}$, $|\mathbf{Pr}_S(f(x) = 0) - \mathbf{Pr}_S(f(x) = 1)| \le \epsilon$. Further, S is said to be k-wise ϵ-discrepant if the "test" linear functions in the above definition can have at most k non-zero coefficients*

Note that, according to the definition in [18], a set is an ϵ-*biased* sample space w.r.t linear tests iff it is ϵ-discrepant for linear functions; furthermore, the restriction k on the number of essential variables in the definition of $\mathcal{AFF}(n,k,s)$ is the generalization of the definition of k-wise ϵ-discrepancy set for linear tests given by Naor and Naor.

One of the major reasons for constructing ϵ-discrepancy sets for linear functions relies on the fact that they can be used to de-randomize some probabilistic algorithms (see [18]).

The main result of [18] is the efficient construction of an ϵ-discrepancy set for linear functions of size $O((n/\epsilon)^{O(1)})$ and a k-wise ϵ-discrepancy set for linear functions of size $O((k\log n)/\epsilon)^{O(1)})$. Three simpler and better (in some parameter ranges) constructions of ϵ-discrepancy sets for linear functions have been introduced by Alon *et al* in [1]. All of them yield sample spaces of size $O((n/\epsilon)^2)$. Their third method is called the *powering construction* and will be extensively used in our paper (a formal description is given in Section 2.1). All such sample spaces can be constructed in time polynomial in n/ϵ.

However, the above efficient constructions of discrepancy sets for linear functions cannot be directly applied to our goal of constructing a Boolean function having hard 1-*Br.Pr.* since we need a discrepancy set for the more general case $\mathcal{AFF}(n,k,\Theta(n))$. The next result provides an elegant solution for this problem.

Lemma 1.1 (Reduction Lemma.) *Let $\epsilon > 0$. If $S \subseteq \{0,1\}^n$ is ϵ-discrepant for $\mathcal{AFF}(n,n,1)$ then it is 2ϵ-discrepant for $\mathcal{AFF}(n,n,n)$.*

This lemma can be proved by combining some properties of the Fourier coefficients of the characteristic functions of Boolean affine spaces and other technical results from [17]. In Section A.1, instead we provide a more direct and simpler proof of the Lemma that, as a result, does not require the use of Fourier analysis.

Thanks to this lemma, any previous construction of ϵ-discrepancy sets for linear functions turns out to be a 2ϵ-discrepancy set also for $\mathcal{AFF}(n,n,n)$.

Theorem 1.1 *Let $n > 0$. For any $k, s \leq n$ and for any $\epsilon > 0$ it is possible to efficiently construct an ϵ-discrepancy set $D(n,k,s)$ for $\mathcal{AFF}(n,k,s)$ of size $|D(n,k,s)| \leq (2k(\log n + 1))/\epsilon)^2$.*

The result concerning $\mathcal{AFF}(n,k,s)$ (i.e. when a non trivial bound on the number of essential variable is imposed) is obtained by providing a different version of the powering construction given by Alon *et al* [1].

Finally, by combining the construction of the parameterized family of Boolean functions f^{α}'s, the selection of the "correct" α by using Theorem 1.1 and Simon ans Szegedy 's Theorem, we achieve the following lower bounds which are exponentially larger than that obtained by Savicky and Zak [23].

Theorem 1.2 *It is possible to construct:*
a) a Boolean function in P having 1-Br.Pr. size not smaller than $2^{n-O(\log^2 n)}$;
b) a Boolean function in $\mathsf{DTIME}(2^{O(\log^2 n)})^{\mathsf{NP}} \cap \mathsf{P/poly}$ having 1-Br.Pr. size not smaller than $2^{n-\log 4n}$ (this bound is optimal);

Non deterministic syntactic read-k-times branching programs and hitting sets for small affine spaces

In the more general case of *k-Br.Pr.*'s, we introduce a new method to derive explicit, exponential lower bounds that uses concepts introduced by Borodin *et al* [8].

Our method relies on the following result. Given any Boolean function $f : \{0,1\}^n \to \{0,1\}$, let $N^1_f = \{x \in \{0,1\}^n : f(x) = 1\}$ and $L_{k-br}(f)$ be the size of smallest *k-Br.Pr.* that computes f.

Theorem 1.3 *Let $t \geq \log^2 n$, $k = o(\log n/\log t)$ and let $f(x_1,\ldots,x_n)$ be a boolean function such that $|N^1_f| \geq 2^{n-1}$ and $L_{k-br}(f) \leq 2^{n^{1-\epsilon}}$ for some constant $0 < \epsilon < 1$. Then, for sufficiently large n, the set N^1_f contains an affine space of dimension not smaller than $t/4$.*

The above theorem implies that the complement of the characteristic function of a family of *hitting sets* (i.e. one-sided random sets) [4, 5, 10, 15] for affine spaces is a good candidate to get hard *k-Br.Pr.* size. Thus, the next goal is to obtain an efficient construction of such hitting sets. In the general definition, given a class $\mathcal{F}$ of Boolean functions of n inputs, a subset $H \subseteq \{0,1\}^n$ is a *hitting set* for $\mathcal{F}$ if, for any non-zero function $f \in \mathcal{F}$, H contains at least one inputs on which f outputs 1. It is easy to verify that Theorem 1.1 guarantees a hitting set for $\mathcal{AFF}(n,k,s)$ if and only if $\epsilon < 1/2^s$. Indeed, in this case we have $|D(n,k,s)| \leq (2k(\log n+1))/\epsilon)^2 = 2k(\log n + 1)2^{2s}$ that gives a hitting set of non trivial size (i.e. of size $o(2^n)$) only when $s < n/2$, i.e., when the corresponding affine space is large. However, when the affine space is small there is a more efficient way to represent it by using its *orthogonal* space (which is large). Combining this idea with the Reduction Lemma, we derive an efficient construction of non trivial hitting sets in the case of small affine spaces.

Theorem 1.4 *For any $n > 0$ and $m \leq n$ it is possible to efficiently construct a hitting set $\mathcal{H}(n,m)$ for $\mathcal{AFF}(n,n,n-m)$ such that $|\mathcal{H}(n,m)| \leq 2^{n-m+m(O(1)\log n)/\sqrt{m}}$.*

The above construction thus provides a non trivial hitting set when $m \geq c\log^2 n$ for some constant $c > 0$. This is a strong improvement over the best previously known construction that yields a non trivial hitting set only when $n - m = o(n)$ [4, 17].

We then consider the complement function $F_n^4(x_1 \ldots x_n)$ of the characteristic function of the hitting set $\mathcal{H}(n,m)$ in Theorem 1.4 (for an appropriate choice of $\log n \leq m < n/2$). By definition of hitting sets, it should be clear that F_n^4 cannot contain any affine subspace of dimension not smaller than m. Hence, Theorem 1.3 implies the following lower bound that improves over those in [8] and [12] for some range of k.

Theorem 1.5 *Let $k = o(\frac{\log n}{\log\log n})$. There exists a Boolean function $F^4 = \{F_n^4 : \{0,1\}^n \rightarrow \{0,1\}, n > 0\}$ that belongs to* P *and, for sufficiently large n, $L_{k-br}(F_n^4) \geq 2^{n^{1-o(1)}}$.*

Organization of the paper

The rest of the paper consists of two main sections. Section 2 is devoted to the constructions of both discrepancy and hitting sets for affine spaces. Then, these constructions are used in Section 3 to derive the explicit lower bounds for branching programs.

2 Pseudo-random sets for affine spaces

Given a subset $W \subseteq \{0,1\}^n$, its size is denoted by $|W|$, and its probability with respect to the uniform distribution in $\{0,1\}^n$ is denoted by $\mathbf{Pr}(W)$; Instead, the notation $\mathbf{Pr}_S(W)$ refers to case in which the uniform distribution is defined on the sample space S.
A Boolean function $f : \{0,1\}^N \rightarrow \{0,1\}$ is said *linear* if a vector $a = (a_1, \ldots, a_N) \in \{0,1\}^N$ exists such that f can be written as $f(x_1, \ldots, x_N) = a_1x_1 \oplus \ldots \oplus a_Nx_N$. Given two binary vectors $a = (a_1, \ldots, a_N)$ and $b = (b_1, \ldots, b_N)$ from $\{0,1\}^N$, we define $c = a \oplus b$ as the vector whose the i-th component is given by the xor of the i-th components of a and b. The *inner* product is defined as $< a, b > = \ a_1b_1 \oplus \ldots \oplus a_Nb_N$. The finite field $GF(2^N)$ will be represented by the standard one-to-one mapping $\text{bin} : GF(2^N) \rightarrow \{0,1\}^N$ such that $\text{bin}(a+b) = \text{bin}(a) \oplus \text{bin}(b)$ and $\text{bin}(0) = (0, \ldots, 0)$.
Furthermore, given any element $a \in GF(2^N)$ and an integer $k > 0$, we will make use of the concatenation of powers $U(a,k) \in \{0,1\}^{kN}$ where $U(a,k) = \text{bin}(a^0)\text{bin}(a^1)\ldots\text{bin}(a^{k-1})$. For the sake of brevity, the notation a^j will replace the term $\text{bin}(a^j)$.

2.1 Discrepancy sets for large affine spaces

For any subset $S \subseteq \{0,1\}^N$ and for any $\alpha \in \{0,1\}^N$ we define

$$\varepsilon(S, \alpha) = |\ 1/2\ -\ (\ \sum_{x \in S} < \alpha, x > /|S|\ )\ |$$

and the "discrepancy" degree of S as $\varepsilon(S) = \max\{\varepsilon(S,\alpha)\ :\ \alpha \in \{0,1\}^N,\ \alpha \neq 0\}$. Note that $S \subseteq \{0,1\}^N$ is ϵ-discrepant for $\mathcal{AFF}(N,N,1)$ iff $\varepsilon(S) \leq \epsilon$. We now introduce the discrepancy degree w.r.t. systems of linear functions. For any choice of $\sigma \in \{0,1\}^s$ and $\alpha_i \in \{0,1\}^N$ for $i = 1, \ldots s$, we define

$$\pi(S, \alpha_1, \ldots, \alpha_s, \sigma_1, \ldots, \sigma_s) = \frac{|\{x \in S\ :\ < \alpha_i, x >= \sigma_i; i = 1, \ldots, s\}|}{|S|}$$

and we let $\varepsilon_s(S)$ be the maximum of the difference $|2^{-s} - \pi(S, \alpha_1, \ldots, \alpha_s, \sigma_1, \ldots, \sigma_s)|$, over all possible linearly independent s-ple of vectors α_i's and vectors $\sigma \in \{0,1\}^s$.

Our interest in the above definition relies on the fact that a set S is ϵ-discrepant for $\mathcal{AFF}(N, N, s)$ iff $\varepsilon_s(S) \leq \epsilon$. We can now present the Reduction Lemma according to the above definitions (for the proof see Section A.1).

Lemma 2.1 (The Reduction Lemma) *Let $S \subseteq \{0,1\}^N$. If $\alpha_1, \ldots, \alpha_s$ are linearly independent and $\sigma \in \{0,1\}^s$, then*

$$\left|2^{-s} - \pi(S, \alpha_1, \ldots, \alpha_s, \sigma_1, \ldots, \sigma_s)\right| \leq \left(2 - \frac{1}{2^{s-1}}\right) \varepsilon(S).$$

Hence if S is ϵ-discrepant w.r.t. $\mathcal{AFF}(N, N, 1)$ then it is 2ϵ-discrepant w.r.t. $\mathcal{AFF}(N, N, s)$, for any $s \leq N$.

We now describe the powering construction of the ϵ-discrepancy set $\mathcal{L}^*_{N,z} \subseteq \{0,1\}^N$ given by Alon *et al* in [1] and then we combine it with the Reduction Lemma in order to get a good discrepancy set for the general class $\mathcal{AFF}(N, N, N)$.

Definition 2.1 *[The Powering Sample Space] [1] The generic vector l in the sample space $\mathcal{L}^*_{N,z}$ is specified by two vectors $x, y \in \{0,1\}^z$. The i-th bit of l is the inner product of the i-th power of x and y. Clearly, we have that $|\mathcal{L}^*_{N,z}| = 2^{2z}$.*

Theorem 2.1 *[1] For any $N > 0$ and $z \leq N$, it holds $\varepsilon(\mathcal{L}^*_{N,z}) \leq \frac{N}{2^{z+1}}$. Hence $\mathcal{L}^*_{N,z}$ is $\frac{N}{2^{z+1}}$-discrepant for $\mathcal{AFF}(N, N, 1)$.*

Notice also that Alon *et al*'s ϵ-discrepancy set for linear functions is the set $\mathcal{L}^*_{N,z}$ for $z = \log(N/\epsilon)$. By applying the Reduction Lemma to Theorem 2.1 we can easily prove the following

Corollary 2.1 *If $\alpha_1, \ldots, \alpha_s \in \{0,1\}^N$ are linearly independent and $\sigma \in \{0,1\}^s$, then*

$$\left|2^{-s} - \pi(\mathcal{L}^*_{N,z}, \alpha_1, \ldots, \alpha_s, \sigma_1, \ldots, \sigma_s)\right| \leq N \cdot 2^{-z} .$$

Hence $\mathcal{L}^*_{N,z}$ is $(N2^{-z})$-discrepant for $\mathcal{AFF}(N, N, N)$.

The above combination of the Alon *et al*'s powering construction and the Reduction Lemma can be slightly modified in order to obtain a discrepancy set for $\mathcal{AFF}(N, k, s)$ that takes in consideration the number k of essential variables (this construction is described in Section A.2).

Theorem 2.2 (Discrepancy set for AFF(N,k,N)) *For any $N > 0$, and $0 \leq k, z \leq N$, it is possible to construct a set $\mathcal{D}_2(N, k, z)$ which is $(k\lceil \log N \rceil 2^{-z})$-discrepant for $\mathcal{AFF}(N, k, N)$.*

2.2 Hitting sets for small affine spaces

In this section we give an efficient construction of non trivial hitting sets for small affine spaces, i.e., for large systems of linear equations. The key idea of the construction is the fact that any affine subspace $\mathcal{S}$ can be represented by using its corresponding orthogonal subspace $\mathcal{S}^\perp$ in $\{0,1\}^n$, and if $\mathcal{S}$ has small dimension then $\mathcal{S}^\perp$ has a large dimension. The orthogonal subspace is thus described by a small linear system and so we can apply the powering construction of the previous section[1]. Our first main task will thus be the construction of a different representation of small affine spaces using their respective orthogonal subspaces.

If $A \subseteq \{0,1\}^n$ then rank(A) denotes the maximal number of linear independent vectors (A will be also considered as a Boolean matrix). Given $a \in \{0,1\}^n$, the term $[a]^j$ denotes the the first j bits of a. Furthermore, we will use the notation $[A]^j = \{[a]^j : a \in A\}$, and define $\text{rank}_j(A) = \text{rank}([A]^j)$. The small affine space that we want to hit is described by the following linear system

$$\{< a_i, x > = \ b_i \quad \text{where } a_i \in \{0,1\}^n,\ b_i \in \{0,1\} \ \ i = 1, 2, \ldots, n-m \ . \tag{1}$$

We let $A = \{a_1, \ldots, a_{n-m}\}$ and assume that rank$(A) = n - m$. As stated before we are interested in the case in which m is small, i.e., $0 \leq m < n/2$. Our first step is to partition the matrix A into $s \geq 16$ smaller submatrices having almost equal rank. Let $\mathcal{N}(n,s)$ be the set of vectors $N = (n_1, \ldots, n_s)$ with positive integer components such that $n_1 + \ldots + n_s = n$. Let $k = m/s$ and *wlog* assume that $k \geq 2$. Then, for sufficiently large n, given any $A = \{a_1, \ldots, a_{n-m}\}$, for any fixed $s \geq 16$, we can always choose a vector $N = (n_1, n_2, \ldots, n_s) \in \mathcal{N}(n,s)$ such that

$$\text{rank}_{n_1+\ldots+n_i}(A) - \text{rank}_{n_1+\ldots+n_{i-1}}(A) = n_i - m_i \text{ and } \ k \leq m_i \leq k+1 \ , \quad \text{for any } i = 1, \ldots, s. \tag{2}$$

Note that $m_1 + m_2 + \ldots + m_s = m$. Let $x = (x_1, x_2, \ldots, x_s)$, where the length of x_i is n_i. Then, by using a standart diagonalization method for linear systems, we can represent the solution space of the system in Eq.(1) with a set of smaller linear systems:

$$\{< b_1, x_1 > \ = \ u_1^{b_1},\ \forall\, b_1 \in B_1 \quad \ldots \ \ldots \quad \{< b_s, x_s > \ = \ u_s^{b_s},\ \forall\, b_s \in B_s$$

for some $B_i \subseteq \{0,1\}^{n_i}$ such that $|B_i| = n_i - m_i$ and rank$(B_i) = n_i - m_i$ (i=1,...,s) and for some $u_i^{b_i} \in \{0,1\}$, $i = 1 \ldots s$ (clearly, both B_i's and $u_i^{b_i}$'s depend on the system in Eq. (1)).

Let $t = \lceil \log s \rceil + 1$ and consider any subset $D = \{d_1, \ldots, d_{k-t}\}$ of vectors from $\{0,1\}^n$. Then the orthogonal space (w.r.t. $[D]^j$) $\mathcal{S}(D,j)$ is the set of solutions of system $< [d_i]^j, (x_1, x_2, \ldots, x_j) >= 0$ for $i = 1, \ldots, k-t$. For any $N \in \mathcal{N}(n,s)$, we then consider the sets of prefix combinations from $\mathcal{S}(D, n_j)$'s, i.e., $\mathcal{S}(D,N) = \mathcal{S}(D,n_1) \times \ldots \mathcal{S}(D,n_s)$. The next result provides the orthogonal condition that a set of vectors D must satisfy in order to let $\mathcal{S}(D,N)$ contain at least one solution of the linear system in Eq. (1) (for a formal proof see Section A.3.1).

Lemma 2.2 *Let $D = \{d_1, \ldots, d_{k-t}\} \subseteq \{0,1\}^n$ and $N \in \mathcal{N}(n,s)$. If for any $i = 1, 2, \ldots, s$ rank$(B_i \cup [D]^{n_i}) = n_i - m_i + k - t$, then $\mathcal{S}(D,N)$ contains at least one solution of the system in Eq. (1).*

Lemma 2.2 suggests us the way to construct a hitting set for the system in Eq. (1): it suffices to find a vector set D that satisfies the "orthogonal" condition in Lemma 2.2 for any possible choices of B_i's, $i = 1, \ldots, s$. It is not hard to see that a random subset D from $\{0,1\}^n$ of size $k-t$ satisfies, with high probability, the above condition. Our idea is to de-randomize the process by choosing our hitting set as the union of all subsets of size $k-t$ from the powering discrepancy set $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$; given any $16 \leq s \leq n$, for any subset $N \in \mathcal{N}(n,s)$ we let

$$\mathcal{H}(N) = \bigcup_{D \subseteq \mathcal{L}^*_{n,(\lceil \log n \rceil + t)} \ : \ |D| = k-t} \mathcal{S}(D,N) \ ,$$

[1] This new representation of small affine spaces does not allow to preserve the discrepancy property of our sample space but only their hitting property.

and we define our hitting set $\mathcal{H}(n,s)$ as the union of $\mathcal{H}(N)$ over all possible choices of N from $\mathcal{N}(n,s)$. The hitting property of $\mathcal{H}(n,s)$ is a consequence of the fact that, from Corollary 2.1, $\mathcal{L}^*_{n,(\lceil \log n\rceil+t)}$ contains at least one solution of any linear systems in $\mathcal{AFF}(n,n,t-1)$. This implies the following result (for a proof see Section A.3.2).

Lemma 2.3 *Let $t = \lceil \log s\rceil + 1$, $N = (n_1,\ldots,n_s) \in \mathcal{N}(n,s)$. Let $B_1,\ldots,B_s$ be such that for any $i = 1,\ldots,s$ $B_i \subseteq \{0,1\}^{n_i}$ with $|B_i| = n_i - m_i$ where $k \le m_i \le k+1$, and $\mathrm{rank}(B_i) = n_i - m_i$. Then a vector subset $D \subseteq \mathcal{L}^*_{n,(\lceil \log n\rceil+t)}$ exists such that $|D| = k-t$ and $\mathrm{rank}(B_i \cup [D]^{n_i}) = n_i - m_i + k - t$, for any $i = 1,\ldots,s$.*

We have now at hand all the ingredients to derive a hitting set for small affine subspaces.

Theorem 2.3 *For any $1 \le m \le n$, for any $16 \le s \le n$, the set $\mathcal{H}(n,s) \subseteq \{0,1\}^n$ is a hitting set for $\mathcal{AFF}(n,n,n-m)$. Furthermore, If $s = \Theta(\sqrt{m})$ then*

$$|\mathcal{H}(n,s)| \le 2^{n-m+m\frac{O(1)\log n}{\sqrt{m}}} .$$

Proof. Consider the system in Eq. (1). As before mentioned, for any $s \ge 16$ (and in particular for $s = \lceil\sqrt{m}\rceil$), we can choose vector $N \in \mathcal{N}(n,s)$ that satisfies Condition (2). Lemma 2.3 then implies that a subset D from $\mathcal{L}^*_{n,(\lceil \log n\rceil+t)}$ exists that satisfies the orthogonal condition in Lemma 2.2. This lemma implies that $\mathcal{S}(D,N)$ contains at least one solution of the system in Eq. (1). The first claim of the theorem thus follows from the fact that $\mathcal{S}(D,N) \subseteq \mathcal{H}(n,s)$.
Concerning the size of the hitting set, assume that $s = \lceil\sqrt{m}\rceil$; then by definition of k and t we have (here we omit some standard computations)

$$|\mathcal{H}(N)| \;\le\; \left(2^{2(\lceil \log n\rceil+t)}\right)^s \prod_{i=1}^{s} 2^{n_i-(k-t)} \;\le\; 2^{n-m+m\frac{20\log n}{\sqrt{m}}} .$$

It follows that

$$|\mathcal{H}(n,s)| \le \begin{pmatrix} n \\ s-1 \end{pmatrix} 2^{n-m+m\frac{20\log n}{\sqrt{m}}} \le 2^{n-m+m\frac{22\log n}{\sqrt{m}}} .$$

□

Finally, we remark that the set $\mathcal{H}(n,\lceil\sqrt(m)\rceil)$ turns out to be a non trivial hitting set (i.e. of size $o(2^n)$) for $\mathcal{AFF}(n,n,n-m)$ for any $m \ge c\log^2 n$ for some constant $c > 0$.

3 New explicit lower bounds for branching programs

In what follows we adopt notations and terminology from [23]. A *Br.Pr.* is a directed acyclic graph where one of the nodes, called *source*, has fan-in 0 and some other nodes, called *terminals*, have fan-out 0. Each non terminal nodes is labelled by the index of an input Boolean variable and has fan-out 2. The two arcs leaving a non terminal node are respectively labelled 0 and 1. A *Br.Pr.* computes a Boolean function $f : \{0,1\}^n \to \{0,1\}$ on a fixed input as follows. Starting the computation from the source node, if a generic node is reached, the corresponding input variable is tested and the computation chooses the arc corresponding to the actual value of this input variable. The process terminates when a sink node is reached and its label represents the output of f. The *size* of a *Br.Pr.* is the number of its nodes.

3.1 Lower bound for 1-read branching programs

We adopt notations and terminolgy introduced in [23]. In particular, for a *partial* input we mean any element from $\{0,1,*\}$ where the positions containing 0 or 1 mean that the corresponding input bit is fixed, while the notation $*$ mean that the corresponding input bit is free. We say that a partial input v is defined on the set $I \subseteq \{0,1\}^n$ if $v_i \in \{0,1\}$ iff $i \in I$. Partial inputs define subfunctions of a given Boolean functions $f : \{0,1\}^n \to \{0,1\}$ in the following natural way. For any set $I \subseteq \{1,2,...,n\}$, let $\mathcal{B}(I)$ be the set of all partial inputs defined on I. Given any partial input $v \in \mathcal{B}(I)$, the subfunction $f|_v$ of $f(x_1,\ldots,x_n)$ is obtained by setting $x_i = v_i$ for any $i \in I$. The set of inputs on which $f|_v$ is defined consists of the Boolean rectangle $R(v)$ of dimension $n - |I|$, $R(v) = \{(a_1,\ldots,a_n) : \ a_i = v(i), \ i \in I\}$. Given any function $f : \{0,1\}^n \to \{0,1\}$ and any subset $I \subseteq \{1,2,...,n\}$, let

$$\nu(f,I) \ = \ \max_{v \in \mathcal{B}(I)} |\{u \in \mathcal{B}(I) \ : \ f|_u = f|_v\}| \ .$$

In the case of 1-*Br.Pr.*'s, Simon and Szegedy introduced a nice technique to derive explicit lower bounds that enjoys of the following theorem.

Theorem 3.1 *[25] Let f be a Boolean function of n variables and let $r \le n$. The size of any* 1-*Br.Pr. computing f is at least* $2^{n-r}/(\max\{\nu(f,I) : \ |I| = n-r\})$.

Our next goal is to give a suitable interpretation of this theorem for the family of n-inputs Boolean functions $\mathsf{powF}_n = \{f^{\alpha}_{n,k} : \alpha \in \{0,1\}^{kn}\}$ where $f^{\alpha}_{n,k}(x)$ is the inner product between α and the concatenation of the first k powers of x, i.e. $f^{\alpha}_{n,k}(x) =< \alpha, U(x,k) >$. Let $I \subseteq \{1,2,...,n\}$ such that $|I| = n-r$; for any $u,v \in \mathcal{B}(I)$ we consider the *xor* vector $df(u,v) \in \{0,1\}^n$ where $df(u,v)_i = 1$ iff $i \in I$ and $v_i \neq u_i$. Then the condition (implicitly required by Theorem 3.1)

$$f^{\alpha}_{n,k}|_v \ = \ f^{\alpha}_{n,k}|_u \ . \tag{3}$$

is equivalent to require that for all $x \in R(v)$, $f^{\alpha}_{n,k}(x) \oplus f^{\alpha}_{n,k}(x \oplus df(v,u)) = 0$. Therefore, by definition of $f^{\alpha}_{n,k}$, it is equivalent to require that $\alpha \in \{0,1\}^{kn}$ is a solution of the following Boolean system of $|R(v)| = 2^r$ linear equations

$$\{< \alpha, (U(x,k) \oplus U(x \oplus df(v,u),k)) >= 0 \ , \qquad x \in R(v) \ . \tag{4}$$

So, in order to obtain a lower bound as large as possible, we need to efficiently find an α that does not satisfy the above system for any choice of I and for any $u \neq v$ from $\mathcal{B}(I)$. The first step to this aim is to choose r and k in order to make the equations linearly independent. If indeed we set $r = \lceil \log n \rceil + 1$, and $k = 2^{r+1}$ then, by definitions of $U(x,k)$ and $df(v,u)$, vectors $U(x,k) \oplus U(x \oplus df(v,u),k)$ are linearly independent for any $x \in R(v)$ and for any different $v,u \in \mathcal{B}(I)$. This implies that for any fixed choice of u and v such that $u \neq v$, the number of α satisfying the linear system is $2^{kn}2^{-h}$ where $h = 2^r$.

By applying standard counting arguments, it would be possible to show that a vector α selected uniformly at random from $\{0,1\}^{kn}$ will not satisfy the linear system in Eq. (4) for any I such that $|I| = n-r$ and for any different $u,v \in \mathcal{B}(I)$ with high probability. This however will not give an explicit and efficient construction of a Boolean function having asymptotically maximal 1-*Br.Pr.* complexity. The key idea is to use the discrepancy set $\mathcal{L}^*_{nk,t}$ (given in Section 2.1) to efficiently de-randomize the probabilistic construction described above. Indeed, let $G(y_1,\ldots,y_{2t})$ (the correct choice of the "price"

t is given later) be the Boolean generator of the set $\mathcal{L}^*_{nk,t}$. By applying Corollary 2.1 with $N = kn$, $s = 2^r$, and $z = t$ we easily have that the system in Eq. (4) for $\alpha = G(\beta)$ is satisfied for not more than $2^{2t}(2^{-2^r} + 2^{r+1}n2^{-t})$ different β's. It follows that the number of β's for which the system is satisfied for some (at least one) I, with $|I| = n - r$, and some (at least one pair) different $v, u \in \mathcal{B}(I)$ is at most

$$2^{2t}(2^{-2^r} + 2^{r+1}n2^{-t}) \left(\begin{array}{c} n \\ r \end{array} \right) 2^{2r}.$$

Now, if we choose $t \geq c \log^2 n$ for a sufficiently large positive constant c, we have that the above value is bounded by $2^{2t}o(1)$. It follows that an element $\beta_0 \in \{0,1\}^{2t}$ exists such that $\alpha_0 = G(\beta_0)$ does not satisfy System (4) for any I with $|I| = n - r$, and for any $v \neq u$ from $\mathcal{B}(I)$. By applying the above deterministc construction and Theorem 3.1, we can prove the following results.

Theorem 3.2 a). *There exists a function* $F^1 = \{F^1_n : \{0,1\}^n \to \{0,1\}, n > 0\}$ *that is constructable in* $\mathsf{DTIME}(2^{O(\log^2 n)})^{\mathsf{NP}} \cap \mathsf{P/poly}$ *and such that, for almost every* $n > 0$, $L_{1-br}(F^1_n) \geq 2^{n-\log(4n)}$.
b). *There exists a function* $F^2 = \{F^2_n : \{0,1\}^n \to \{0,1\}, n > 0\}$ *that is constructable in* P *and such that, for almost every* $n > 0$, $L_{1-br}(F^2_n) \geq 2^{n-O(\log^2 n)}$.

Proof. a). As described before the theorem, for any $n > 0$, we let $r = \lceil \log n \rceil + 1$, $k = 2^{r+1}$ and $t \geq c \log^2 n$. We then choose a "good" vector $\beta \in \{0,1\}^{2t}$ such that condition (3) does not hold for any I with $|I| = n - r$ and for any $v \neq u$ from $\mathcal{B}(I)$. Finally, we define $F^1_n = f^{G(\beta)}_{n,k}$. It is easy to verify that the function can be computed by a circuit of polynomial size provided that the correct β is given and also that $F^1 \in \mathsf{DTIME}(2^{O(\log^2 n)})^{\mathsf{NP}}$. Moreover, by applying Theorem 3.1, we have that $L_{1-br}(F^1_n) \geq 2^{n-r} \geq 2^n/4n$.
b). The construction is similar to that of the first case. The only difference is that the new function F^2_n has now an auxilary input of $s = O(\log^2 n)$ bits that specifies the input for the Boolean generator G. More formally, $F^2_n : \{0,1\}^s \times \{0,1\}^n \to \{0,1\}$, is defined as

$$F^2_n(\beta, x) = f^{G(\beta)}_{n,k}(x) \ .$$

Since G works in time polynomial in its output size (this is a direct consequence of the powering construction given in [1] - see Definition 2.1), it is easy to show that F^2 can be constructed in time polynomial in n. Since, in the worst-case, F^2 cannot be easier than F^1, the lower bound for F^2 is a conseqence of the first case and of the fact that the size of the input of F^2 is $O(\log^2 n)$ bits larger than the input of F^1.
□

3.2 A lower bound for k-branching programs

Theorem 3.3 *Let* $t \geq \log^2 n$ *and* $k = o\left(\frac{\log n}{\log t}\right)$. *Let* $f : \{0,1\}^n \to \{0,1\}$ *be such that* $|N^1_f| \geq 2^{n-1}$ *and* $L_{k-br}(f) = M \leq 2^{n^{1-\epsilon}}$ *for some constant* $0 < \epsilon < 1$. *Then, for sufficiently large* n, *the subset* N^1_f *contains a (at least one) Boolean affine subspace in* $\{0,1\}^n$ *of dimension at least* $t/4$.

In order to prove the above Theorem we will make use of some concepts and results untroduced by Borodin *et al* in [8].

Definition 3.1 *[8] A Boolean function $g(x_1, \ldots, x_n)$ is a (k, a)-rectangle if g can be represented in the form*

$$g \;=\; \bigwedge_{i=1}^{ka} g_i(X_i)$$

where g_i is a Boolean function depending only on variables from $X_i \subseteq \{x_1, \ldots, x_n\}$, $|X_i| \leq \lceil n/a \rceil$ and each variable belongs to at most k of the sets $\{X_1, \ldots, X_{ka}\}$.

The next lemma provides an interesting relation between (1,T)-rectangles and Boolean affine subspaces (the proof is given in Section A.4). Let N_g^1 be the set of x such that $g(x) = 1$.

Lemma 3.1 *If $g(x_1, \ldots, x_n)$ is a $(1, T)$-rectangle then N_g^1 contains an affine space of dimension at least $(|N_g^1|)/((n/T) + 1)$.*

Let $L_{k-br}(f)$ be the size of the smallest (non deterministic syntactic) *k-Br.Pr.* that computes f. Borodin *et al* showed that the existence of an upper bound on the *k-Br.Pr.* size of a given Boolean function yields a possible "rectangular" representation of it.

Lemma 3.2 *[8] Let $f(x_1 \ldots, x_n)$ be a Boolean function such that $L_{k-br}(f) \leq M$ and let a and k be positive integers. Then f is an OR of at most $(2M)^{2ka}$ (k, a)-rectangles.*

The next lemma provides a method to represent any (k, a)-rectangle (with any $k > 1$) as an OR of a certain number of $(\mathbf{1}, a(k))$-rectangles where $a(k)$ is exponential in k (the proof is given in Section A.5).

Lemma 3.3 *Let $g(x_1, \ldots, x_n)$ be a (k, a)-rectangle and assume that for some $t \geq 2$ it holds*

$$(4t^{k+1})/n \;+\; (4t^{2k+1}k)/a \;<\; 1 \;.$$

Then g is an OR of at most $2^{n-n/(2t^{k-1})}$ $(1, 2 \cdot t^k)$-rectangles.

Proof of Theorem 3.3

If we choose $a = 16kt^{2k+1}$ then, for sufficiently large n, we obtain

$$(4t^{k+1})/n \;+\; (4t^{2k+1}k)/a \;\leq\; o(n) + 1/4 < 1 \;.$$

By combining Lemma 3.2 and Lemma 3.3, we have that f is an OR of at most r $(1, 2 \cdot t^k)$-rectangles, where $r \leq (2M)^{2ka} 2^{n-n/(2t^{k-1})} \leq 2^{n-n/(4t^{k-1})}$. Since $|N_f^1| \geq 2^{n-1}$, the above inequality implies that N_f^1 contains at least one $(1, 2 \cdot t^k)$-rectangle of size not smaller than $(2^{n-1})/(2^{n-n/(4t^{k-1})}) = 2^{n/(4t^{k-1})-1}$. From Lemma 3.1 this rectangle contains an affine space of dimension at least

$$\frac{\log\left(2^{n/(4t^{k-1})-1}\right)}{(n/(2t^k)) + 1} \;=\; \frac{n/(4t^{k-1}) - 1}{(n/(2t^k)) + 1} \;\geq\; \frac{1}{2}\frac{n/(4t^{k-1})}{(n/(2t^k))} \;=\; \frac{t}{4} \;.$$

□

For any $n > 0$, we define the Boolean function $F_n^4(x_1, \ldots, x_n)$ as the complement of the characteristic function of the hitting set $\mathcal{H}(n, \sqrt{m})$ given by Theorem 2.3 where we set $m = \lceil \log^4 n \rceil$. From the construction of $\mathcal{H}(n, s)$ shown in Section 2.2, we can prove that the family $F^4 = \{F_n^4,\; n > 0\}$ can be constructed in polynomial time (a formal proof is given in Section A.6). Furthermore, from Theorem 2.3 we have that $|\mathcal{H}(n, O(\log^2 n))| = o(2^n)$; so, by applying Theorem 3.3, we obtain the following lower bound.

Corollary 3.1 *If $k = o(\frac{\log n}{\log \log n})$ then, for sufficiently large n, $L_{k-br}(F_n^4) \geq 2^{n^{1-o(1)}}$.*

Acknowledgements. We would like to thank Oded Goldreich and Luca Trevisan for several comments on a preliminary version of this paper. We also thank Sasha Razborov and Martin Sauerhoff for helpful pointers. Finally, Andrea Clementi wish to thank Riccardo Silvestri for his crucial hints in proving Theorem 3.2.

References

[1] N. Alon, O. Goldreich, J. Hastad, and R. Peralta (1990), "Simple Constructions of Almost k-wise Independent Random Variables", *Proc. of IEEE-FOCS*, Vol. 2, pp. 544-553.

[2] A. Andreev, A. Clementi, and J. Rolim (1997), "A New General De-randozation Method", *Journal of the Association for Computing Machinery (J. of A.C.M.)*, 45(1):179-213, January 1998.

[3] A. Andreev, A. Clementi, and J. Rolim (1997), "Worst-case Hardness Suffices for Derandomization: a New method for Hardness-Randomness Trade-Offs", in Proc. of *ICALP*, LNCS, 1256, pp. 177-187.

[4] A. Andreev, A. Clementi, J. Rolim (1997), "Efficient Constructions of Hitting Sets for Systems of Linear Functions", in *ECCC* TR96-029, (Extended Abstract in Proc. of *STACS'97*, LNCS 1200, pp.387–398).

[5] R. Armoni, M. Saks, A. Wigderson, and S. Zhou (1996 "Discrepancy Sets and Pseudorandom Generators for Combinatorial Rectangles", Proc. of *IEEE-FOCS*, pp. 412-421.

[6] Blum M., and Micali S. (1984), "How to generate cryptographically strong sequences of pseudorandom bits", *SIAM J. of Computing*, 13(4), pp. 850-864.

[7] R. Ben-Natan (1990), "On independent random variables over small sample spaces", *M.Sc. Thesis*, Computer Science Dept., Hebrew University, Jerusalem, Israel, Feb.

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

[9] Bshouty N. H. (1989), "On the Extended Direct Sum Problem Conjecture", Proc. of *ACM-STOC*, pp. 177-185.

[10] G. Even, O. Goldreich, M. Luby, N. Nisan and B. Velickovic (1992), "Approximations of general independent distributions", Proc. of *ACM-STOC*, pp 10-16.

[11] R. Impagliazzo, and A. Wigderson (1997), "P= BPP if E requires exponential circuits: Derandomizing the XOR lemma" Proc. of *ACM STOC*, pp. 220-229.

[12] S. Jukna (1995), "A Note on Read-k-times Branching programs", *RAIRO Theoretical Informatics and Applications*, 29 (1), pp.75-83. (also in *ECCC*, TR94-027).

[13] S. Jukna, A. Razborov, P. Savicky, I. Wegener (1997), "On P versus $NP \cap co-NP$ for Decision Trees and Read-Once Branching Programs", *ECCC*, TR97-023.

[14] J. Justesen (1972), "A Class of Constructive Asymptoticaly Good Algebraic Codes", *IEEE Transactions on Information Theory,* 18, pp. 652–656.

[15] D. Karger and D. Koller (1994), "(De)randomized construction of small sample spaces in NC", Proc. of *IEEE-FOCS*, pp. 252-263.

[16] Karchmer M., Raz R., and Wigderson A. (1991), "On Proving Super-Logarithmic Depth Lower Bounds via the Direct Sum in Communication Complexity", Proc. of *IEEE Structure in Complexity Theory*, pp. 299-304.

[17] E. Kushilevitz, and Y. Mansour (1993), " Learning Decision Trees Using the Fourier Spectrum", SICOMP 22(6), pp. 1331-1348. Early version: STOC 91.

[18] J.Naor, and M.Naor (1990), "Small-bias probability spaces: efficient constructions and aplications", Proc. of *ACM-STOC*, pp. 213-223.

[19] N.Nisan (1990), "Pseudo-random generators for Space-Bounded Computation", Proc. of *ACM-STOC*, pp. 204-212.

[20] Nisan N., and Wigderson A. (1994), "Hardness vs Randomness", *J. Comput. System Sci.*, 49, pp. 149-167.

[21] E.A. Okolnishnikova (1993), "On lower bounds for branching programs", *SiberianAdvances in Mathematics*, 3(1), pp. 152-166.

[22] M. Sauerhoff (1997), "A Lower Bound for Randomized Read-k-Times Branching Programs", *ECCC*, TR97-019.

[23] P. Savicky and S. Zak (1996), "A large lower bound for 1-branching programs", *ECCC*, TR96-036.

[24] P. Savicky and S. Zak (1998), "A read-once lower bound and a (1,+k)-hierarchy for branching programs", submitted paper.

[25] J. Simon and M. Szegedy (1993), "A new lower bound theorem for read only once branching programs and its applications", *Advances in Computational Complexity Theory (J.Cai, editor)*, *AMS-DIMACS Series*, 13, pp.183-193.

[26] J. S. Thathachar (1998), "On Separating the Read-k-Times Branching Program Hierarchy", Proc. of *ACM-STOC*, to appear.

[27] U. Vazirani (1986), "Randomness, Adversaries and Computation", *PhD Thesis*, EECS, UC Berkeley.

[28] I. Wegener (1988), "On the Complexity of Branching Programs and Decision Trees for Clique Functions", *J.ACM*, 35, pp. 461-477.

A Proofs

A.1 Proof of Lemma 2.1

Fix $S \subseteq \{0,1\}^N$ and a set of s linearly independent vectors $\alpha_1, \ldots, \alpha_s$ from $\{0,1\}^N$. For any $\sigma = (\sigma_1, \ldots, \sigma_s)$ we consider the "discrepancy degree" of S w.r.t. σ:

$$d(\sigma) = \pi(S, \alpha_1, \ldots, \alpha_s, \sigma_1, \ldots, \sigma_s)$$

and define $q(\sigma) = d(\sigma) \cdot |S|$. For any $b = (b_1, \ldots, b_s) \in \{0,1\}^s$, we also consider the linear combination $\alpha(b) = b_1\alpha_1 \oplus \ldots \oplus b_s\alpha_s$. The sum

$$Q(b) \ = \ \sum_{l \in S} < l, \alpha(b) > \ ,$$

can be written as

$$Q(b) \ = \ \sum_{l \in S} < l, b_1\alpha_1 \oplus \ldots \oplus b_s\alpha_s > \ = \ \sum_{l \in S}(b_1 < l, \alpha_1 > \oplus \ldots \oplus b_s < l, \alpha_s >) \ =$$

$$= \ \sum_{\gamma=(\gamma_1,\ldots,\gamma_s)\in\{0,1\}^s} \left(\sum_{l \in S \ : \ <l,\alpha_1>=\gamma_1,\ldots,<l,\alpha_s>=\gamma_s} < b, \gamma > \right) ;$$

so,

$$Q(b) = \sum_{\gamma\in\{0,1\}^s} q(\gamma) < b, \gamma > \ = \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma) < b, \gamma > \ . \tag{5}$$

Two cases may arise depending on whether or not $\sigma = 0$.

1.) For $\sigma \neq 0$, Eq. (5) implies that

$$\sum_{b \ : \ <b,\sigma>=1} Q(b) \ = \ \sum_{b \ : \ <b,\sigma>=1} \ \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma) < b, \gamma > \ =$$

$$= \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma) \sum_{b \ : \ <b,\sigma>=1} < b, \gamma > \ = \ q(\sigma)2^{s-1} + \sum_{\gamma\in\{0,1\}^s\setminus\{0,\sigma\}} q(\gamma)2^{s-2} \tag{6}$$

where the last equality is due to the fact the the number of $b \neq 0$ such that $< b, \sigma >= 1$ and $< b, \gamma >= 1$ for non zero $\sigma \neq \gamma$ is equal to 2^{s-2}. Similarly we have

$$\sum_{b \ : \ <b,\sigma>=0 \ , \ b\neq 0} Q(b) = \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma) \sum_{b \ : \ <b,\sigma>=0 \ , \ b\neq 0} < b, \gamma >= \sum_{\gamma\in\{0,1\}^s\setminus\{0,\sigma\}} q(\gamma)2^{s-2}. \tag{7}$$

Combining Eq.s (6) and (7), we get

$$q(\sigma)2^{s-1} \ = \ \sum_{b \ : \ <b,\sigma>=1} Q(b) \ - \ \sum_{b \ : \ <b,\sigma>=0 \ , \ b\neq 0} Q(b) \tag{8}$$

We remind that

$$\varepsilon(S) = \max\{\varepsilon(S, \alpha) \ : \ \alpha \in \{0,1\}^N, \ \alpha \neq 0\} \ \text{ where } \ \varepsilon(S, \alpha) = | \ 1/2 \ - \ (\ \sum_{x \in S} < \alpha, x > /|S| \ ) \ | \ ;$$

For any $b \neq 0$, it is then easy to prove that

$$|S|\left(\frac{1}{2}-\varepsilon(S)\right) \leq Q(b) \leq |S|\left(\frac{1}{2}+\varepsilon(S)\right) \ . \tag{9}$$

Then from Eq.s (8) and (9), we have that

$$2^{s-1}|S|\left(\frac{1}{2}-\varepsilon(S)\right)-(2^{s-1}-1)|S|\left(\frac{1}{2}+\varepsilon(S)\right) \leq q(\sigma)2^{s-1} \ \leq$$

$$\leq \ 2^{s-1}|S|\left(\frac{1}{2}+\varepsilon(S)\right)-(2^{s-1}-1)|S|\left(\frac{1}{2}-\varepsilon(S)\right) \ .$$

It follows that

$$2^{-s}-(2-2^{1-s})\varepsilon(S) \leq \frac{q(\sigma)}{|S|} \leq 2^{-s}+(2-2^{1-s})\varepsilon(S) \ .$$

2.) For $\sigma = 0$, we can obtain the same bounds in the following way.

$$\sum_{b\ :\ b\neq 0} Q(b) = \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma)\sum_{b} < b,\gamma >= \sum_{\gamma\in\{0,1\}^s\setminus\{0\}} q(\gamma)2^{s-1} \ = 2^{s-1}(|S|-q(0)).$$

This implies that

$$\frac{q(0)}{|S|} \ = \ 1-\frac{1}{2^{s-1}|S|}\sum_{b\ :\ b\neq 0} Q(b) \ . \tag{10}$$

By using the same argument yielding Eq. (9), we get

$$(2^s-1)|S|\left(\frac{1}{2}-\varepsilon(S)\right) \ \leq \ \sum_{b\ :\ b\neq 0} Q(b) \ \leq \ (2^s-1)|S|\left(\frac{1}{2}+\varepsilon(S)\right). \tag{11}$$

¿From Eq. 10 and 11, we obtain

$$\frac{q(0)}{|S|} \leq 1-\frac{2^s-1}{2^{s-1}}\left(\frac{1}{2}-\varepsilon(S)\right) = 1-(2-2^{1-s})\left(\frac{1}{2}-\varepsilon(S)\right) = 2^{-s}+(2-2^{1-s})\varepsilon(S)$$

In the same way, we can derive the lower bound for $\frac{q(0)}{|S|}$:

$$\frac{q(0)}{|S|} \ \geq \ 2^{-s}-(2-2^{1-s})\varepsilon(S).$$

A.2 A discrepancy set for linear systems with a limited number of essential variables

The aim of this section is to modify the powering construction given in the previous section in order to obtain a discrepancy set for $\mathcal{AFF}(N, k, s)$ that takes in consideration the number k of essential variables.

Let $n = \log N$ and $1 \leq z \leq n$. Consider the N elements $a_1, \ldots, a_N$ of $GF(2^n)$. For any $a \in GF(2^n)$, we define the concatenation of powers $U(a, k) = a^0 a^1 \ldots a^{k-1}$ and, given a vector $l \in \mathcal{L}^*_{kn,z}$, we construct the following vector of N bits

$$(< l, U(a_1, k) >, < l, U(a_2, k) >, \ldots, < l, U(a_N, k) >)$$

The new powering discrepancy set is then defined as

$$\mathcal{D}_2(N, k, z) \;=\; \bigcup_{l \in \mathcal{L}^*_{kn,z}} \{(< l, U(a_1, k) >, < l, U(a_2, k) >, \ldots, < l, U(a_N, k) >)\} \;.$$

Clearly, we have $|\mathcal{D}_2(N, k, z)| = |\mathcal{L}^*_{kn,z}| = 2^{2z}$. Informally speaking, the discrepancy property of $D(N, k, z)$ is a consequence of the fact that given any k pairwise different a_i, $i = 1, \ldots, k$ from $GF(2^n)$, the corresponding k vectors $U(a_i, k)$ in $GF(2^{nk})$ are linearly independent

Theorem A.1 (Discrepancy set for AFF(N,k,N)) *For any $N > 0$, and $0 \leq k, z \leq N$, the set $\mathcal{D}_2(N, k, z)$ is $(k\lceil \log N \rceil 2^{-z})$-discrepant for $\mathcal{AFF}(N, k, N)$.*

Proof. Consider the set W of all solutions of a fixed linear system $< l_1, x >= \beta_1, \ldots, < l_s, x >= \beta_s$ where vectors $l_1, \ldots, l_s$ from $\{0, 1\}^N$ are linearly independent. Assume also that the number of essential variables is at most k, i.e., the number of 1's in the *or* vector $l_1 \vee \ldots \vee l_s$ is at most k. Consider a fixed $l \in \mathcal{L}^*_{kn,z}$ and its corresponding string in $\mathcal{D}_2(N, k, z)$:

$$v_l \;=\; (< l, U(a_1, k) >, < l(U(a_2, k) >, \ldots, < l, U(a_N, k) >)$$

For $l_i = (\lambda^i_1, \ldots, \lambda^i_N)$ $(i = 1, \ldots, s)$, we define vectors $R_i \in GF(2^{kn})$ as

$$R_i \;=\; \lambda^i_1 U(a_1, k) \oplus \lambda^i_2 U(a_2, k) \oplus \ldots \oplus \lambda^i_N U(a_N, k).$$

It is possible to prove that $R_1, \ldots, R_s$ are linearly independent. This indeed follows from three facts: 1) any k vectors of type $U(a_j, k)$ are linearly independent; 2) at most k of such vectors appear in each R_i. 3) The linear combinations of them remain linearly independent. Then, for any $i = 1, \ldots, s$, we have

$$< l_i, v_l > = < (\lambda^i_1, \ldots, \lambda^i_N), (< l, U(a_1, k) >, \ldots, < l, U(a_N, k) >) > =$$

$$= \lambda^i_1 < l, U(a_1, k) > \oplus \ldots \oplus \lambda^i_N < l, U(a_N, k) >) > =$$

$$= < l, \lambda^i_1 U(a_1, k) \oplus \ldots \oplus \lambda^i_N U(a_N, k) > = < l, R_i > \;.$$

Our final step is to reduce to conditions required by Corollary 2.1. Indeed, the equation $< l_i, v_l >= \beta_i$ is equivalent to $l(R_i) = \beta_i$. It follows that, for any l such that $v_l \in W \cap \mathcal{D}_2(N, k, z)$, we have

$$< l, R_1 >= \beta_1 \ldots, < l, R_s >= \beta_s \;.$$

where $R_1, \ldots, R_s$ are linearly independent. Corollary 2.1 thus implies the thesis. □

A.3 Proofs of lemmas 2.2 and 2.3

A.3.1 Proof of lemma 2.2

From the hypothesis on $\mathrm{rank}(B_i \cup [D]^{n_i})$, if we assume that $u_i^{b_i} = 0$ for any $b_i \in [D]^{n_i}$, it easily follows that the system

$$\{< b_i, x_i > = \ u_i(b_i) \ \ , \forall b_i \in B_i \cup [D]^{n_i}$$

admits a solution and $B_i \cap [D]^{n_i} = \emptyset$. This immediatly implies the lemma.

A.3.2 Proof of lemma 2.3

We show an iterative method to construct the set D by adding one new element to it in order to satisfy the orthogonal property specified by the lemma. Assume we already have the set

$$D_j \subseteq \mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$$

with $|D| = j$ $(0 \leq j < k - t)$ such that

$$\mathrm{rank}(B_i \cup [D_j]^{n_i}) = n_i - m_i + j \ , \qquad i = 1, \ldots, s \ .$$

We then consider the homogeneous system

$$\{< b_i, x_i > = \ 0 \ , \quad \text{for any } b_i \in B_i \cup [D_j]^{n_i} \ ,$$

whose solution space has dimension $m_i - j$. Let $\{v_1^i, \ldots, v^i_{m_i - j}\}$ be a vector basis of this space, and define $V_j^i \in \{0,1\}^n$ as the concatentation of v_j^i and dummy 0's. Observe that for any $a \in \{0,1\}^n$, we have

$$< V_j^i, a > = < v_j^i, [a]^{n_i} > \ .$$

For any fixed $i \in \{1, \ldots, s\}$, the vector $[a]^{n_i}$ cannot be represented as a linear combination of vectors from $(B_i \cup [D_j]^{n_i})$ iff the following system is not satisfied

$$\Big\{ < V_t^i, a >= 0 \ \ , \ t = 1, \ldots, m_i - j \ . \tag{12}$$

For any fixed i, Corollary 2.1 implies that the probability that a vector a from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ satisfies the system in Eq. (12) is at most

$$2^{j - m_i} + 2^{-(\lceil \log n \rceil + t)} n \leq 2^{(k-t-1)-k} + 2^{-t} = \frac{3}{2} 2^{-t} = \frac{3}{2} 2^{-(\lceil \log s \rceil + 1)} \leq \frac{3}{4}\frac{1}{s} \ .$$

It follows that the probability that a vector a chosen uniformly at random from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ satisfies Condition (12) is true for at least one i from $\{1, \ldots, s\}$ is at most

$$s \frac{3}{4}\frac{1}{s} \ = \ \frac{3}{4} \ ,$$

So, there exists a from $\mathcal{L}^*_{n,(\lceil \log n \rceil + t)}$ such that for all $i \in \{1, \ldots, s\}$ the system in Eq. (12) is not satisfied. We then define $D_{j+1} \ = \ D_j \cup \{a\}$; the construction terminates when $j = k - t$ and we set $D = D_{k-t}$.

A.4 Proof of lemma 3.1

Let

$$g = \bigwedge_{i=1}^{a} g_i(X_i) \ .$$

Since $k = 1$, the sets X_i's are pairwise disjoint. Further, we have that

$$|N_g^1| \ = \ \prod_{i=1}^{a} |N_{g_i}| \ .$$

Let r be the number of i's such that $|N_{g_i}^1| > 1$. Since

$$|N_g^1| \ \leq \ \left(2^{\lceil \frac{n}{T} \rceil}\right)^r \ ,$$

then

$$r \ \geq \ \frac{|N_g^1|}{(n/T)+1} \ .$$

Let A_i be the affine space of maximum dimension d_i such that $A_i \subseteq N_{g_i}^1$. Observe that if $|N_{g_i}^1| > 1$ then $d_i \geq 1$ since any two different points yield an affine space of dimension 1. We thus consider the affine space $A = A_1 \times \ldots \times A_a$. The dimension of A is at least r.

A.5 Proof of lemma 3.3

For any fixed $i \in \{1, \ldots, ka\}$, consider the subset $X_i \subseteq \{x_1, \ldots, x_n\}$ where $|X_i| \leq \lceil n/a \rceil$ and assume that each variable belongs to at most k subsets from $\{X_1, \ldots, X_{ka}\}$. Let $q(i)$ be the number of sets X_j such that $x_i \in X_j$.

Let us fix an integer $t \geq 2$ and then select randomly a function $\phi : \{1, \ldots, ka\} \to \{1, \ldots, t\}$ with uniform distribution. Let us consider the Boolean function $\xi_{s,i}(\phi)$ defined over all possible choices of ϕ such that $\xi_{s,i}(\phi) = 1$ if for all j for which $x_i \in X_j$ it holds $\phi(j) = s$, and $\xi_{s,i}(\phi) = 0$ otherwise. Then, for any fixed s, we let

$$\Xi_s(\phi) \ = \ \sum_{i=1}^{n} \xi_{s,i}(\phi) \ .$$

Our first goal is study the expected behaviour of the above random variable. In particular, if $\mathbf{E}(*)$ denotes the expected value of a given random variable, then using Chebichev's inequaltiy, it is possible to prove the following bound. For any fixed $s = \{1, \ldots, t\}$ and for any $0 \leq \epsilon \leq 1$

$$\mathbf{Pr}\left(|\Xi_s - \mathbf{E}(\Xi_s)| \geq \epsilon \mathbf{E}(\Xi_s)\right) \ \leq \ \frac{1}{\epsilon^2}\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) \ . \qquad (13)$$

Now, let $g(x_1, \ldots, x_n)$ be a (k,a)-rectangle and assume that for some $t \geq 2$

$$\frac{4t^{k+1}}{n} + \frac{4t^{2k+1}k}{a} \ < \ 1 \ .$$

From Eq. (13), for any fixed s, we get

$$\mathbf{Pr}\left(\Xi_s \leq \frac{1}{2}\frac{n}{t^k}\right) \leq \mathbf{Pr}\left(\Xi_s \leq \frac{1}{2}\mathbf{E}\left(\Xi_s\right)\right) \leq \mathbf{Pr}\left(\left|\Xi_s - \mathbf{E}\left(\Xi_s\right)\right| \geq \frac{1}{2}\mathbf{E}\left(\Xi_s\right)\right) \leq 4\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) .$$

Consequently

$$\mathbf{Pr}\left(\exists s \in \{1,\ldots,t\} \,:\, \Xi_s \leq \frac{1}{2}\frac{n}{t^k}\right) \leq t \cdot 4\left(\frac{t^k}{n} + \frac{t^{2k}k}{a}\right) \leq \frac{4t^{k+1}}{n} + \frac{4t^{2k+1}k}{a} < 1 .$$

Therefore, a function ϕ exists such that

$$\forall s \in \{1,\ldots,t\} : \ \Xi_s(\phi) > \frac{1}{2}\frac{n}{t^k}$$

For any s, we can thus choose X_s^* such that

$$|X_s^*| = \lceil \frac{n}{2t^k} \rceil , \qquad \text{and } X_s^* \subseteq \bigcap_{j\ ,\ \phi(j)=s} X_j .$$

We then define

$$X_0^* \ = \ \{x_1,\ldots,x_n\} \setminus \left(\bigcup_{s=0}^{t} X_s^*\right) .$$

Since g is a (k,a)-rectangle, we can write g as follows

$$g(x_1,\ldots,x_n) \ = \ \bigwedge_{s=1}^{t}\left(\bigwedge_{j\ :\ \phi(j)=s} g_j(X_j)\right) .$$

Let v be a partial input from $\mathcal{B}(X_0^*)$ then, for any $j = 1,\ldots,ka$, if $\phi(j) = s$ then $X_j \setminus X_0^* \ = \ X_s^*$. It follows that the function

$$\left(\bigwedge_{i\in X_0^*} (x_i \oplus v(x_i))\right)\left(\bigwedge_{j\ :\ \phi(j)=s} g_j(X_j)\right)$$

depends only on the variables from X_s^*, so there exist functions $g_s^v(X_s^*)$, $s = 1,\ldots,t$, such that

$$\left(\bigwedge_{i\in X_0^*} (x_i \oplus v(x_i))\right) g(x_1,\ldots,x_n) = \left(\bigwedge_{s=1}^{t} g_j^v(X_s^*)\right) ,$$

and g can be written in the following form

$$g(x_1,\ldots,x_n) \ = \bigvee_{v\in\mathcal{B}(X_0^*)}\left(\bigwedge_{i\in X_0^*} (x_i \oplus v(x_i))\right)\left(\bigwedge_{s=1}^{t} g_j^v(X_s^*)\right) , \tag{14}$$

Observe now that if $s \neq v$ then $X_s^* \cap X_v^* = \emptyset$ and, moreover, for any $s = 1,\ldots,t$, $|X_s^*| \leq \lceil n/(2t^k)\rceil$. Consequently the representation in Eq. (14) of g is an OR of $2^{|X_0^*|}$ of $(1, 2t^k)$-rectangles. The Lemma is proved by observing that

$$|X_0^*| \leq n - t\left(\frac{n}{2t^k}\right) \ = \ n - \frac{n}{2t^{k-1}} .$$

A.6 An efficient method to compute F^4

For any $n > 0$, we define the Boolean function $F_n^4(x_1, \ldots, x_n)$ as the complement of the characteristic function of the hitting set $\mathcal{H}(n, \sqrt{m})$ given by Theorem 2.3 where we set $m = \lceil \log^4 n \rceil$.

Lemma A.1 *Let* $F^4 = \{F_n^4,\ n > 0\}$, *then* $F^4 \in \mathsf{P}$.

Proof. The hitting set $\mathcal{H}(n, s)$ can be written in the following recursive form

$$\mathcal{H}(n, s) = \bigcup_{k=1}^{n-(s-1)} \mathcal{H}(k, 1) \ \times \ \mathcal{H}(n-k, s-1)\ . \tag{15}$$

For any $t = 0, \ldots, n-s+1$, let $f_{n,s}^{t+1}(x_{t+1}, \ldots, x_n)$ be the characteristic function of $\mathcal{H}(n-t, s)$. Consider now the operator

$$F_{n,s}(x_1, \ldots, x_n) \ = \ (f_{n,s}^1(x_1, \ldots, x_n), f_{n,s}^2(x_2, \ldots, x_n), \ldots, f_{n,s}^{n-s+1}(x_{n-s+1}, \ldots, x_n))$$

From Eq. (15), we have

$$f_{n,s}^{t+1}(x_{t+1}, \ldots, x_n) \ = \ \bigvee_{k=1}^{n-k-(s-1)} \left(f_{k,1}^1(x_{k+1}, \ldots, x_{k+t}) \wedge f_{n,s-1}^{k+t+1}(x_{k+t+1}, \ldots, x_n) \right)\ .$$

It thus follows that the circuit complexity (so, the time complexity) of $F_{n,s}$ satisfies the following bound

$$L(F_{n,s}) \leq L(F_{n,s-1}) + n^3$$

since $f_{k,1}^1$ is an AND of linear functions. We thus have that $L(F_{n,s}) \leq n^4$. The lemma is then proved by simply observing that $f_{n,s}^1(x_1, \ldots, x_n)$ is the characteristic function of $\mathcal{H}(n, s)$.

□