Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:
ECCC
TR96-029
FTP: ftp.eccc.uni-trier.de:/pub/eccc/
WWW: http://www.eccc.uni-trier.de/eccc/
Email: ftpmail@ftp.eccc.uni-trier.de with subject 'help eccc'

Towards efficient constructions of hitting sets that derandomize BPP

Alexander E. Andreev*
Department of Mathematics
University of Moscow

Andrea E. F. Clementi
Dipartimento di Scienze dell'Informazione
University of Rome

José D. P. Rolim†
Centre Universitaire d'Informatique
University of Geneva

(Extended Abstract)

Abstract

A subset $\mathcal{H} \subseteq \{0,1\}^n$ is a *Hitting Set* for a class $\mathcal{R}$ of boolean functions with n inputs if, for any function $f \in \mathcal{R}$ such that $\Pr(f = 1) \geq \delta$ (where $\delta \in (0,1)$ is some fixed value), there exists an element $\vec{h} \in \mathcal{H}$ such that $f(\vec{h}) = 1$.

The efficient construction of Hitting Sets for non trivial classes of boolean functions is a fundamental problem in the theory of derandomization. Our paper presents a new method to efficiently construct Hitting Sets for the class of *systems of boolean linear functions*.

Systems of boolean linear functions can be also considered as the algebraic generalization of boolean *combinatorial rectangular* functions studied by Linial *et al* in [11]. In the restricted case of boolean rectangular functions, our method (even though completely different) achieves equivalent results to those obtained in [11].

Our method gives also an interesting upper bound on the circuit complexity of the solutions of any system of *linear equations* defined over a finite field.

Furthermore, as preliminary result, we show a new upper bound on the circuit complexity of integer *monotone* functions that generalizes the upper bound previously obtained by Lupanov in [12].

*Partially supported by Grant N 95-01-00707 of Russian Foundation for Fundamental Researches,

†Contact author: Centre Universitaire d'Informatique, University of Geneva, 24 rue General Dufour, CH 1204 Geneva. E-mail: rolim@cui.unige.ch Fax ++41-22-705-7780

1 Introduction

This work is motivated by a recent result established by [5] in the theory of derandomization. Informally speaking, this result states that *quick Hitting Set Generators* can replace *quick Pseudorandom Generators* [15] in derandomizing BPP-algorithms. More precisely, in [5] it is proved that an efficient construction of a Hitting Set for the class of boolean functions having linear circuit-size complexity is sufficient to prove $P = BPP$. Consequently, a major, challenging goal in this area is now the development of algorithmic techniques to efficiently construct Hitting Sets for more and more general classes of boolean functions which can eventually culminate in the final efficient construction of Hitting Sets for boolean functions having linear circuit-size complexity. There is also a more practical goal in doing this research. Indeed, in the last decade, efficient constructions of *Hitting Sets* for particular families of finite functions have played an important role to reduce randomness in some probabilistic algorithms [3, 8, 9, 11, 16].

Our paper presents a new method to efficiently construct Hitting Sets for an important natural class of boolean functions which is the algebraic generalization of the class of *combinatorial rectangle boolean functions* studied in [11]. Our results give a new positive step in achieving the above mentioned goal of this area.

We study the class (denoted as $\mathcal{L}(n,k)$) of boolean functions that can be expressed as *systems* (i.e. logical conjunctions) of boolean *linear* functions, i.e.

$$f(x_1,\ldots,x_n) = \bigwedge_{j=1}^{k} \left(a_1^j x_1 \oplus a_2^j x_2 \oplus \ldots \oplus a_n^j x_n \oplus b^j\right) \quad a_i^j, b^j \in \{0,1\},\ k \leq n\ . \tag{1}$$

The complexity and the properties of boolean linear functions have been the subject of several studies over the past few years [1, 6, 7, 10]. Informally speaking, the main interest in linear functions is motivated by the fact that they have "small" (i.e. polynomial) circuit-size complexity [13] but they have a rather rich behavior recently used in [6] to approximate general boolean functions within a good trade-off between circuit-size complexity and the corresponding degree of approximation. More precisely, in [6], it is proved that any general boolean function $f : \{0,1\}^n \to \{0,1\}$ always admits a suitable combination of linear boolean functions $S : \{0,1\}^n \to \{0,1\}$ with equivalent (i.e. polynomially related) circuit-size complexity, and which agrees with f on a large fraction of inputs. Even though the problem to efficiently transform a Hitting Set for a function class into a Hitting Set for another class approximated by the former is still an open difficult question, the positive result in the approximation of general boolean functions using linear functions [5] should give the idea of the small complexity "gap" between linear and general boolean functions. This small gap provides a significant motivation in finding Hitting Sets for the class of systems of linear functions.

- *Previous Results on Hitting Sets.* In which follows, we consider the standard definition of circuit-size complexity of finite functions; moreover, given any boolean sequence $\vec{x} \in \{0,1\}^n$, we will use the term *complexity of* $\vec{x}$ to refer to the circuit-size complexity of the corresponding boolean function $x : \{0,1\}^{\lceil \log(n+1) \rceil} \to \{0,1\}$ where $x(i)$ is the i-th bit of $\vec{x}$. The circuit-size complexity of a finite function f (a finite sequence $\vec{x}$) will be denoted as $L(f)$ ($L(\vec{x})$).

A subset $\mathcal{H} \subseteq \{0,1\}^n$ is a *Hitting Set* for a class $\mathcal{R}$ of boolean functions with n inputs if, for any function $f \in \mathcal{R}$ such that $\Pr(f = 1) \geq \delta$ (where $\delta \in (0,1)$ is some fixed value), there exists an element $h \in \mathcal{H}$ such that $f(h) = 1$. A natural, well studied question concerning Hitting Sets is the *witness finding problem*: given a positive integer $n > 0$, and a positive number δ with $0 < \delta < 1$, find a subset $\mathcal{H} \subseteq \{0,1\}^n$ such that, for any *witness set* $\mathcal{W} \subseteq \{0,1\}^n$ with $|\mathcal{W}|/2^n \geq \delta$, we have $|\mathcal{W} \cap \mathcal{H}| > 0$. It is

immediate to verify that the witness finding problem consists of finding a Hitting Set for the class of all n-input boolean functions f such that $\Pr(f = 1) \geq \delta$.

Karp, Pippenger, and Sipser [9], and Sipser [16] introduced a randomized method to solve the witness finding problem that uses $O(n)$ random bits. Chor and Goldreich [8] derived a simpler algorithm that uses n random bits. This algorithm can be considered the best result in solving the witness finding problem for general witness sets. More recently, research has turned the attention also on non trivial restrictions of the problem where some combinatorial properties on the witness sets are imposed. An interesting positive result in this direction is that introduced by Linial, Luby, Saks, and Zuckermann [11]. They gave a deterministic method to construct an efficient solution for the witness finding problem when the witness sets are *combinatorial rectangles* (this result is described later).

Our results for systems of linear functions. In this paper, we study the witness finding problem for the class $\mathcal{L}(n,k)$ $(k \leq n)$ of boolean functions that can be represented in the form defined in Eq. 1. We also denote as $\mathcal{L}(n,k,q)$ $(q \leq n)$ the class of boolean functions in $\mathcal{L}(n,k)$ having at most q non-zero columns in the matrix $A = [a_i^j]$ defined in Eq. (1) (q is commonly called the number of *essential* variables). Observe that since for any *non null* function $f \in \mathcal{L}(n,k)$ we have $\Pr(f = 1) \geq 2^{-k}$, then the role of parameter δ in the definition of the witness finding problem is now replaced by the term 2^{-k}. The main results of this paper can be stated in the following way.

Theorem 1 *Let ϵ be any positive constant such that $0 < \epsilon < 1/3$.*

- *If $k \geq n^{2/3+\epsilon}$, then it is possible to construct a Hitting Set $\mathcal{H} \subseteq \{0,1\}^n$ for $\mathcal{L}(n,k)$ such that $|\mathcal{H}| \leq 2^{O(k)}$.*
- *If $\log\log n \leq k \leq n^{2/3}$ and $k \geq q^{2/3+\epsilon}$, then it is possible to construct a Hitting Set $\mathcal{H} \subseteq \{0,1\}^n$ for $\mathcal{L}(n,k,q)$ such that $|\mathcal{H}| \leq 2^{O(k)}$.*
- *If $k \leq \log\log n$ and $q \geq q^{2/3+\epsilon}$, then it is possible to construct a Hitting Set $\mathcal{H} \subseteq \{0,1\}^n$ for $\mathcal{L}(n,k,q)$ such that $|\mathcal{H}| \leq 2^{O(k)} \log^2 n$.*
- *If $k \leq \min\{n^{2/3}, q^{2/3}\}$, then it is possible to construct a Hitting Set $\mathcal{H} \subseteq \{0,1\}^n$ for $\mathcal{L}(n,k,q)$ such that $|\mathcal{H}| \leq 2^{O(k)\log n}$*

In all cases, the time required by the construction is polynomially bounded in the size of the output sequence.

Observe also that, for the first three cases, the size of the Hitting Sets is almost optimal since a simple lower bound for the size of Hitting Sets for $\mathcal{L}(n,k)$ is 2^k. Furthermore, a Hitting Set for the class $\mathcal{L}(n,k)$ corresponds to a subset containing at least one solution for *any* feasible linear system of the form $A\vec{x} = \vec{b}$ where A is a $k \times n$ boolean matrix, $\vec{x} \in \{0,1\}^n$ and $\vec{b} \in \{0,1\}^k$ (we will consider only feasible systems and thus, in the following, we will omit the term feasible). It follows that our results provides also interesting upper bounds for the size of the minimal space which contains at least one solution for any of such linear systems. Under this point of view, since our method is based only on algebraic properties of linear algebra in finite fields, we can derive equivalent upper bounds for systems of linear equations on finite (non boolean) fields. In order to obtain the upper bounds for the case in which the $k \times n$ linear system is defined over the field $GF(Q)$ of cardinality Q, it is sufficient to replace the basis 2 with value Q in the formulas listed in the above theorem. However, we will not mention this possible generalization further in this extended abstract.

- *Connection between combinatorial rectangles and systems of linear equations*

As mentioned above, Linial *et al* [11] studied the witness finding problem in the case of combinatorial rectangles. A *combinatorial rectangle* is any subset of the form $\mathcal{R} = R_1 \times R_2 \ldots \times R_n$ where $R_i \subseteq \{0, .., m-1\}$. The goal here is to generate a subset $\mathcal{H} \subseteq \{0, .., m-1\}^n$ that has non empty intersection with every combinatorial rectangle $\mathcal{R}$ whose size (also denoted as *volume*) is at least δm^n. Linial *et al*'s algorithm generates a hitting set $\mathcal{H}$ whose size is polynomial in $m \log n(1/\delta)$, and the running time is polynomial in $mn(1/\delta)$.

Observe that when $m = 2$ (i.e. the boolean case), then the characteristic function of a generic combinatorial rectangle $\mathcal{R} \subseteq \{0,1\}^n$ can be expressed as a system of *boolean linear functions*:

$$f(x_1, \ldots, x_n) = \bigwedge_{j=1}^{k} (x_{i(j)} \oplus b^j) \tag{2}$$

Indeed suppose, for instance, that $n = 3$ and the rectangle is $\mathcal{R} = \{0,1\} \times \{0\} \times \{1\}$; the corresponding characteristic function can be written as

$$f^{\mathcal{R}}(x_1, x_2, x_3) = (0 \cdot x_1 \oplus 1) \wedge (1 \cdot x_2 \oplus 1) \ \wedge \ (1 \cdot x_3 \oplus 0) = (x_2 \oplus 1) \wedge (x_3) \ .$$

Observe also that, given any boolean *rectangular* function $f^{\mathcal{R}}$ represented by Eq. 2, the size of the corresponding rectangle $\mathcal{R} \subseteq \{0,1\}^n$ easily verify the following equation $|\mathcal{R}| = 2^{-k} \cdot 2^n$. Thus, the role of the parameter δ in the definition of combinatorial rectangles is now replaced by the term 2^{-k}. More formally, the class of boolean rectangular functions having volume parameter $\delta = 2^{-k}$ is strictly contained in the class $\mathcal{L}(n, k, k)$ which always satisfies one of the first three cases of Theorem 1.

Boolean rectangle functions represent thus the intersection between Linial *et al*'s work and our work: while Linial *et al* provide an efficient construction of Hitting Set for general (i.e. non boolean) rectangular functions, our work gives an efficient solution of the same problem but for the class of boolean functions in which the "rectangular" condition is relaxed into the much more general condition expressed by Eq. 1. In the case of boolean rectangular functions, we thus give an another (completely different) method to construct Hitting Sets which has equivalent performances to those obtained by Linial *et al.*

Adopted techniques and further results. It is easy to see that the witness finding problem for the class $\mathcal{L}(n, k)$ corresponds to find a subset $\mathcal{H} \subseteq \{0,1\}^n$ which contains at least one solution of any system of $k \leq n$ linear *equations* in n variables, i.e.:

$$A \times \vec{x} \ = \ \vec{b} \ , \ \text{ where } \ A \in \{0,1\}^{k \times n}, \ \vec{x} \in \{0,1\}^n, \ \text{ and } \ \vec{b} \in \{0,1\}^k. \tag{3}$$

We first consider "large" linear systems (i.e. when $k > n^{2/3}$). In this case, we show a suitable matrix decomposition method which permits us to prove the following interesting characterization of the space of solutions of a linear system.

Lemma 1 *Let ϵ be constant, $0 < \epsilon < 1/3$. If $k \geq n^{2/3+\epsilon}$ then any linear system of type (3) has at least one solution (i.e. an n-bit sequence) with complexity at most $O\left(\frac{k}{\log k}\right)$.*

From this Lemma, we then derive the Hitting Set construction stated in the first item of Theorem 1. Then, we show how it is possible to reduce the other cases (i.e. when $k \leq n^{2/3}$) to the case of Lemma 1 by using some new properties of *boolean linear operators* which are stated in Lemmas 9 and 10.

The method adopted for proving Lemma 1 requires the use of integer *monotone* functions. An integer monotone function is any function $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, s\}$, such that $f(i) \leq f(j)$ for any $i < j$.

The circuit-size complexity of these functions has been studied by Lupanov [12] in the restricted case $n = s$. However, our construction requires an upper bound which holds also for the case $s < n$, and Lupanov's method cannot be applied in this case. Another contribution of our paper is the generalization of Lupanov's result.

Theorem 2 *If* $f : \{1, \ldots, n\} \to \{1, \ldots, s\}$ *with* $s \leq n$, *then* $L(f) \leq (1 + o(1)) \frac{\log \binom{n}{s}}{\log s} + (\log n)^{O(1)}$.

2 Hitting Sets for "large" linear systems

2.1 Complexity of monotone functions

The construction of the Hitting Set for the class $\mathcal{L}(n, k)$ requires the use of integer *monotone* functions, i.e., $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, s\}$, such that $f(i) \leq f(j)$ for any $i < j$. The complexity of these functions has been studied by Lupanov [12] in the restricted case $n = s$. However, in our construction we need an upper bound for the general case $s < n$.

Consider the following special coding of monotone sequences. Let $U = u_1, \ldots, u_t$ where $1 \leq u_1 < \ldots < u_t < n$, define the function $NUM_{n,t}(U)$ as follows

$$NUM_{n,t}(U) \;=\; \sum_{i=1}^{t} \binom{u_i}{i} \;.$$

By $DENUM_{n,t}(m)$ we denote the "inverse" operator, i.e.:

$$\text{if } NUM_{n,t}((u_1, \ldots, u_t)) = m \text{ then } DENUM_{n,t}(m) = (u_1, \ldots, u_t).$$

Lemma 2 *For any* $n > 0$ *and* $t \leq n$, *we have*

$$L(NUM_{n,t}) \leq O((t \log n)^4) \;, \qquad L(DENUM_{n,t}) \leq O((t \log n)^4) \;.$$

Let $1 \leq u_1 < \ldots < u_{ts} \leq n$ and consider the operator

$$F_U(i) \;=\; (u_{t(i-1)+1}, u_{t(i-1)+2}, \ldots, u_{t(i-1)+t}) \;.$$

By using Lemma 2, we can prove the following result.

Lemma 3 *If* $s \leq n$ *then*

$$L(F_U) \;\leq\; \frac{\log \binom{n}{ts}}{\log s} + O((t \log n)^4) \;.$$

We can now prove the generalization of Lupanov's result [12].

Theorem 3 *If* $f : \{1, \ldots, n\} \to \{1, \ldots, s\}$ *with* $s \leq n$, *then*

$$L(f) \leq (1 + o(1)) \frac{\log \binom{n}{s}}{\log s} + (\log n)^{O(1)}.$$

Sketch of the proof. Let $m \leq st$ and let $v_1, \ldots, v_{st}$ be a monotone sequence such that $f(i) = j$ if $v_{i-1} < j \leq v_i$. If we define $u_i = v_i + i - 1$, we have $1 \leq u_1 < \ldots < u_{st}$. Consider the sequence $U = (u_1, u_2, ..., u_{st})$ and the corresponding operator F_U. It is not hard to prove that $L(f) \leq L(F_U) + (t \log n)^2$. We can choose s and t such that

$$\log st \;=\; (1 + o(1)) \log m \;, \qquad \log t = o(\log s) \;, \qquad t \geq (log n)^2 \;,$$

The theorem then follows by applying Lemma 3. □

2.2 The complexity of solutions of "large" linear systems

Any non null function $f \in \mathcal{L}(n, k)$ represented by Eq. 1 verifies the following equation: $\Pr(\mathrm{f} = 1) = 2^{-\mathrm{r(A)}} \geq 2^{-\mathrm{k}}$. Furthermore, a subset $\mathcal{H} \subseteq \{0,1\}^n$ is a Hitting Set for $\mathcal{L}(n, k)$ iff $\mathcal{H}$ contains at least one solution for any feasible system of type (3). This equivalence result will be strongly used in deriving our Hitting Sets.

Given a boolean (k, n)-matrix A, consider the following column partition. Let $n = n_1 + \ldots + n_s$ and let A_i be a boolean (k, n_i)-matrix, such that

$$A \;=\; (A_1, A_2, ..., A_s) \;. \tag{4}$$

Define $r_s = \mathrm{r}(A_s)$ where $\mathrm{r}(A)$ denotes the rank of A, and

$$r_i \;=\; \mathrm{r}((A_i, A_{i+1}, \ldots, A_n)) - \mathrm{r}((A_{i+1}, A_{i+2}, \ldots, A_n)) \;, \qquad i = 1, \ldots, s-1 \;.$$

Then consider the linear system

$$A\vec{x} \;=\; \vec{b} \;, \tag{5}$$

where $\vec{x} \in \{0,1\}^n$ and $\vec{b} \in \{0,1\}^k$. Note that, without loss of generality, we can always assume that A has maximum rank, i.e., $\mathrm{r}(A) = k$. Using the above matrix representation, it is possible to show an interesting relation between the solutions of System (5) and the Hitting Sets for the classes $\mathcal{L}(n_i, r_i)$'s.

Lemma 4 *For any $i = 1, \ldots, s$, let $\mathcal{H}_i$ be a Hitting Set for $\mathcal{L}(n_i, r_i)$. Then, for any $\vec{b} \in \{0,1\}^k$, there exists a solution of System (5) which belongs to the set*

$$\mathcal{H}_1 \times \mathcal{H}_2 \times \ldots \times \mathcal{H}_s.$$

The following lemma states that the above matrix representation actually exists.

Lemma 5 *Let $r, m \leq n$ and $r \leq m$. Then, given any (k, n)-matrix A, it is possible to construct Representation (4) of A which satisfies the following conditions.*

$$r_i \leq r \;,\; n_i \leq m \;, \qquad i = 1, 2, ..., s \;, \tag{6}$$

and

$$s \leq \frac{k}{r} + \frac{n}{m} \;. \tag{7}$$

Let $[\vec{a}]^j$ be the prefix of length j of sequence $\vec{a}$ and, for any set S of boolean sequences, define $[S]^j \;=\; \{[\vec{a}]^j \;:\; \vec{a} \in S\}$. Hitting Sets for systems of linear functions satisfy the following *monotone* property.

Lemma 6 *If $\mathcal{H}$ is a Hitting Set for $\mathcal{L}(n,k)$ then, for any $n' \leq n$ and $k' \leq k$, the set $[\mathcal{H}]^{n'}$ is a Hitting Set for $\mathcal{L}(n',k')$.*

The above Lemmas imply the following result.

Lemma 7 *Let $r \leq m$ and $\mathcal{H}$ be a Hitting Set for $\mathcal{L}(m,r)$. Assume that Condition (6) is verified. Then there exists a solution of System (5) which belongs to the set*

$$[\mathcal{H}]^{n_1} \times [\mathcal{H}]^{n_2} \times \ldots \times [\mathcal{H}]^{n_s} \ .$$

Given any class $\mathcal{R}$ of boolean functions, the function $\lambda(\mathcal{R})$ denotes the minimum size of a Hitting Set for $\mathcal{R}$. Using the probabilistic method [2], it is possible to prove the following result.

Lemma 8 *For any $n > 0$ and $k \leq n$, we have*

$$\lambda(\mathcal{L}(n,k)) \ \leq \ 2^k(n+1)k.$$

We can now prove the main result of this section.

Theorem 4 *Let ϵ be a positive constant such that $0 < \epsilon < 1/3$. If $k \geq n^{2/3+\epsilon}$ then any system of type (5) has at least one solution with complexity at most $O\left(\frac{k}{\log k}\right)$.*

Sketch of the proof. From Lemma 5, we can construct the matrix representation in Eq. (4) which satisfies Conditions (6) and (7) (the choice of parameters r and m are given later). Then, Lemma 7 implies that there exists a solution of System (5) belonging to the set

$$[\mathcal{H}]^{n_1} \times [\mathcal{H}]^{n_2} \times \ldots \times [\mathcal{H}]^{n_s} \ ,$$

where $\mathcal{H}$ is a Hitting Set for $\mathcal{L}(m,r)$. We now show how to compute a sequence $\vec{a} = \vec{a}_1 \ldots \vec{a}_s$ where $\vec{a}_i \in [\mathcal{H}]^{n_i}$. We assume that $\mathcal{H} = \{\vec{h}_1, \ldots \vec{h}_{|\mathcal{H}|}\}$. For any $i = 1, \ldots, s$, $Q(i)$ denotes the index for which $\vec{a}_i = [\vec{h}_{Q(i)}]^{n_i}$. Define $NUM(u)$ as the function which gives, for any $u = 1, \ldots, n$, the index of the submatrix of A which contains column u. In other terms, $NUM(u)$ is uniquely determined by the following condition

$$\sum_{i=1}^{NUM(u)-1} n_i \ < \ u \ \leq \ \sum_{i=1}^{NUM(u)} n_i \ .$$

Define also $LEN(i) = \sum_{t=1}^{i-1} n_t$, and $SF(v) = \vec{h}_v$. Finally, let $SEL(i, \alpha_1, \ldots, \alpha_m) = \alpha_i$ if $1 \leq i \leq m$, and 0 otherwise. We can derive the u-th bit of $\vec{a}$ using the following sequence of computations

$$i = NUM(u) \ ; \qquad l = LEN(i) \ ; \qquad j = u - l \ ; \qquad p = Q(i) \ ;$$

$$\vec{\alpha} = SF(p) \ ; \qquad \vec{a}(u) \ = \ SEL(j, \vec{\alpha}) \ .$$

The ranges of the parameters used in the above computations are the following

$$u, l \in \{1, 2, .., n\}, \ j \in \{1, 2, .., m\}, \ i \in \{1, 2, ..., s\}, \ p \in \{1, 2, .., |\mathcal{H}|\}, \ \vec{\alpha} \in \{0,1\}^m.$$

It is then easy to prove the following bound for the complexity of $\vec{a}$:

$$L(\vec{a}) \;\leq\; L(NUM) + L(LEN) + O(\log n) + L(Q) + L(SF) + L(SEL)\ . \tag{8}$$

In which follows we give upper bounds for every element of the above sum. From Theorem 3 we have

$$L(NUM) \leq\; O\left(\frac{s}{\log s}\log\frac{n}{s} + (\log n)^{O(1)}\right)\ . \tag{9}$$

Since $LEN : \{1, 2, .., s\} \rightarrow \{1, 2, .., n\}$ then its output consists of $\log n$ bits; hence, by using Lupanov's result [12], we obtain

$$L(LEN) \;\leq\; (1 + o(1))\frac{s}{\log s}\log n\ . \tag{10}$$

An equivalent argument holds for functions $Q : \{1, .., s\} \rightarrow \{1, .., |\mathcal{H}|\}$, and $SF : \{1, .., |S|\} \rightarrow \{0, 1\}^m$:

$$L(Q) \;\leq\; (1 + o(1))\frac{s}{\log s}\log|\mathcal{H}| \;\text{ and }\; L(SF) \;\leq\; (1 + o(1))\frac{|\mathcal{H}|}{\log|\mathcal{H}|}m\ . \tag{11}$$

The function SEL can be easily constructed within the following circuit complexity

$$L(SEL) \;\leq\; O(m)\ . \tag{12}$$

By replacing Eq.s (9-12) in Eq. 8, we get

$$L(\vec{a}) \;\leq\; O\left(\frac{s}{\log s}(\log|\mathcal{H}| + \log n) + \frac{|\mathcal{H}|}{\log|\mathcal{H}|}m\right)\ . \tag{13}$$

By Lemma 5, we have that $s \leq \frac{k}{r} + \frac{n}{m}$. If we choose $r = \epsilon \log n$ and $m = \lceil\frac{nr}{k}\log n\rceil$, then we have $s \leq (1 + o(1))\frac{k}{r}$.

From Lemma 8, we have that $|\mathcal{H}| = 2^r(m + 1)r$ and, consequently,

$$\frac{s}{\log s}(\log|S| + \log n) \leq O\left(\frac{\frac{k}{r}}{\log\frac{k}{r}}(\log m + r + \log n)\right) \leq\; O\left(\frac{k}{\log k}\right)\ . \tag{14}$$

Furthermore,

$$\frac{|\mathcal{H}|}{\log|\mathcal{H}|}m \;\leq\; O\left(\frac{2^r(m+1)r}{r + \log m}m\right) \;\leq\; O\left(\frac{1}{\log n}2^r m^2\right) \leq$$

$$\leq O\left(\frac{1}{\log n}2^r\left(\frac{nr}{k}\log n\right)^2\right) \leq O\left((\log n)^3 2^r\left(\frac{n}{k}\right)^2\right) \;\leq O((\log n)^3 n^\epsilon n^{2/3-2\epsilon}) \leq$$

$$\leq\; O((\log n)^3 n^{2/3-\epsilon}) \;\leq\; O\left(\frac{k}{\log k}\right)\ . \tag{15}$$

From Eq.s (13), (14), and (15) we get $L(\vec{a}) \leq O(k/\log k)$. □

2.3 Construction of Hitting Sets for "large" linear systems

Let $\mathcal{F}(n,l)$ be the set of all sequences $\vec{a} \in \{0,1\}^n$ such that $L(\vec{a}) \leq l$. The following corollary is a consequence of Theorem 4 and of the Lupanov's bound [12] on the size of $\mathcal{F}(n,l)$.

Corollary 1 *Let ϵ be a positive constant such that $0 < \epsilon < 1/3$. If $k \geq n^{2/3+\epsilon}$ then there exists a constant c (which can be efficiently derived from the proof of Theorem 4) such that $\mathcal{F}(n, c\frac{k}{\log k})$ is a Hitting Set for $\mathcal{L}(n,k)$. Furthermore, the size of this Hitting Set is such that $|\mathcal{F}(n, c\frac{k}{\log k})| \leq 2^{O(k)}$.*

We can now use a standard procedure to generate all boolean sequences in $\mathcal{F}(n,l)$, where l is a fixed, known upper bound. In particular, we can construct a *Hitting Set Generator* (in short HSG, see [5]) $H : \{0,1\}^{h(n)} \rightarrow \{0,1\}^n$ where $h(n) = O(l \log(\log n + l))$. The HSG H is an algorithm which considers any sequence of $h(n)$ bits as the description of a boolean circuit having complexity l: if the description is correct then it generates the corresponding table of the circuit outputs, otherwise it returns the string $\vec{0}$. It is not hard to verify that the subset of strings generated by applying H to any possible inputs (i.e. the set $H(\{0,1\}^{h(n)})$) contains $\mathcal{F}(n,l)$. We thus have the following result.

Corollary 2 *Let ϵ be a positive constant such that $0 < \epsilon < 1/3$. If $k \geq n^{2/3+\epsilon}$ then it is possible to generate the Hitting Set $\mathcal{F}(n, c\frac{k}{\log k})$ for the class $\mathcal{L}(n,k)$ in polynomial time in $2^{O(k)}$.*

Note. It is easy to see that 2^k is a lower bound on the size of Hitting Sets for $\mathcal{L}(n,k)$. It follows that the logarithm of the size of our Hitting Set is optimal. In terms of rectangular boolean functions, since the volume parameter δ is equal to 2^{-k}, we thus generate a Hitting Set $\mathcal{H}$, for the class of boolean rectangular functions with n inputs and volume parameter δ, which has size polynomially bounded in $1/\delta$. Furthermore, the time to construct $\mathcal{H}$ is polynomially bounded in $(1/\delta)n$.

3 Hitting Sets for "small" linear systems

In this section, we describe a reduction technique whose goal is to extend the previous construction to the case of small (i.e. $k \leq n^{2/3}$) linear systems. This method works for the class $\mathcal{L}(n,k)$ with no restrictions on k but when a particular condition on the number of non-zero columns in the system matrix A is assumed. Let us now introduce the class of systems of linear functions determined by this new condition and its relation with boolean rectangular functions.

A function $f(x_1,..,x_n) \in \mathcal{L}(n,k)$ belongs to the subclass $\mathcal{L}(n,k,q)$ if it can be represented by Eq. (1) where matrix $A = [a_i^j]$ has at most q non-zero vertical columns. In which follows, we will consider the case in which k and q satisfy the following inequality: $k \geq q^{2/3+\epsilon}$ (intuitively, the number of linear functions in the system must be relative large with respect to the number q of essential variables). Since rectangular boolean functions are linear systems in which there is exactly one variable in every linear function (see Eq. 2) then it is easy to verify that the class of rectangular boolean functions with n variables and with volume $2^{-k}2^n$ is contained in the class $\mathcal{L}(n,k,k=q)$ which always satisfies the condition $k \geq q^{2/3+\epsilon}$.

3.1 Some properties of linear operators

In this section, we show some useful properties of *linear operators* which will be used in the above mentioned reduction. The set of all linear functions l's with m variables, such that $l(0,..,0) = 0$, is denoted as Lin_m. Moreover, a vectorial function $\vec{l} = (l_1,..,l_s) \in \mathsf{Lin}_m^s$ $(s \geq 1)$ is called *linear operator*. We will use the following result obtained by Nechiporuk [13].

Theorem 5 [13] *For any linear operator* $\vec{l} = (l_1, ..., l_s) \in \mathsf{Lin}_m^s$ *($s \geq 1$), we have*

$$L(\vec{l}) = O\left(\frac{ms}{\log m}\right) + O(m).$$

For any $\vec{a}, \vec{d} \in \{0,1\}^m$, consider now the "agreement" function $\xi_{\vec{a},\vec{d}} : \mathsf{Lin}_m^s \to \{0,1\}$ defined as $\xi_{\vec{a},\vec{d}}(\vec{l}) = \Pi_{i=1}^s (l_i(\vec{a}) \oplus l_i(\vec{d}) \oplus 1)$, where $\vec{l} = (l_1, .., l_s)$. Then, we define

$$\Xi_{\mathcal{A},\vec{d}}(\vec{l}) = \sum_{\vec{a} \in \mathcal{A}\ ,\vec{a} \neq \vec{d}} \xi_{\vec{a},\vec{d}}(\vec{l})\ .$$

If we consider $\vec{l}$ randomly selected from the space Lin_m^s with uniform probability, then we have the following upper bound on the expected value of $\Xi_{\mathcal{A},\vec{d}}$.

Lemma 9

$$\mathbf{E}\left(\Xi_{\mathcal{A},\vec{d}}\right) \leq |\mathcal{A}| 2^{-s}.$$

The above lemma can be used to derive the existence of an integer function which has the following particular "injectivity" property. This function will be one of the key ingredients in the reduction shown in the next section.

Lemma 10 *Let $\mathcal{A} \subseteq \mathcal{B} \subseteq \{1, 2, ..., n\}$. For any $s \geq 1$, there exists a function $f : \{1, .., n\} \to \{1, .., q\}$, with $q = |\mathcal{A}| + 2^s$, such that for any $a \in \mathcal{A}$ and $b \in \mathcal{B}$ $(a \neq b)$ we have $f(a) \neq f(b)$, and*

$$L(f) \leq O(|\mathcal{A}||\mathcal{B}| 2^{-s}(s + \log n) + s \log n)\ .$$

3.2 Hitting Sets for case $k \geq \max\{\log^2 n, q^{2/3+\epsilon}\}$

The proof of the following theorem gives the main reduction which allows us to extend the results for large linear systems to the case of small linear systems.

Theorem 6 *Let ϵ be a constant such that $0 < \epsilon < 1/3$. If $k \geq \max\{\log^2 n, q^{2/3+\epsilon}\}$ then any system of type (5) has at least one solution with complexity at most $O\left(\frac{k}{\log k}\right)$.*

Sketch of the proof. Consider a system $A\vec{x} = \vec{b}$ where A is a boolean $k \times n$-matrix with $\mathrm{r}(A) = k$, $\vec{x} \in \{0,1\}^n$ and $\vec{b} \in \{0,1\}^k$. Assume also that A satisfies the conditions of the theorem. Let $\mathcal{A} \subseteq \{1, .., n\}$ be the subset of indexes which describes a subset of A-columns of size and rank k. If $\mathcal{B}$ denotes the set of all indexes corresponding to non-zero columns of A, then we easily have that $\mathcal{A} \subseteq \mathcal{B}$. Let $s = \lfloor ((2/3 + \epsilon)/(2/3 + \epsilon/2)) \log q \rfloor$. From Lemma 10, there exists a function $f : \{1, .., n\} \to \{1, .., n'\}$ where $n' = 2^s + k$ such that for any $a \in \mathcal{A}$ and $b \in \mathcal{B}$, with $a \neq b$, we have $f(a) \neq f(b)$. Furthermore,

$$L(f) \leq O(|\mathcal{A}||\mathcal{B}| 2^{-s}(s + \log n) + s \log n).$$

Since

$$n' \leq q^{\frac{2/3+\epsilon}{2/3+\epsilon/2}} + k \leq k^{\frac{1}{2/3+\epsilon/2}} + k \leq 2 * k^{\frac{1}{2/3+\epsilon/2}}\ ,$$

then it is not hard to prove that, for a.e. n, $k \geq (n')^{2/3+\epsilon/3}$.

We now define a linear transformation for system $A\vec{x} = \vec{b}$ which leads us to the case of large systems described in the Section 2.3. The linear transformation is defined by the following equations

$$x_i \;=\; y_{f(i)} \;, \qquad i = 1, ..., n \;.$$

The properties of function f in Lemma 10 implies that the new obtained system has still rank k. If $\vec{a} = (a_1, a_2, ..., a_{n'})$ is a solution of the new system, then $\vec{\alpha} = (a_{f(1)}..a_{f(n)})$ is a solution for system $A\vec{x} = \vec{b}$. Furthermore, we have $L(\vec{\alpha}) \leq L(f) + L(\vec{a})$. We can now apply Theorem 4 thus proving that there exists a solution $\vec{a}$ of the new obtained system such that $L(\vec{a}) \;\leq\; O\left(\frac{k}{\log k}\right)$. From Lemma 10 and from the fact that $k = \Omega(\log^2 n)$, there exists a positive constant c for which

$$L(f) \;\leq\; O(k * q * 2^{-s}(s + \log n) + s \log n) \;\leq$$

$$O(k * 2^{-c\epsilon k}(\log k + \log n) + \log k \log n) \;= O\left(\frac{k}{\log k}\right),$$

Consequently, we have $L(\vec{\alpha}) = O\left(\frac{k}{\log k}\right)$.

□

The above theorem implies that the set $\mathcal{F}(n, c(k/\log k)$, for some constant $c > 0$, is a Hitting Set for the class $\mathcal{L}(n, k, q)$ when $k \geq \max\{\log^2 n, q^{2/3+\epsilon}\}$. We can thus repeat the same Hitting Set construction sketched in Section 2.3 and obtain equivalent results to those in Corollaries 1 and 2.

3.3 Hitting Sets for case $k < (\log n)^2$, $k \geq q^{2/3+\epsilon}$

Given a linear system $A\vec{x} = \vec{b}$, where A is a boolean $k \times n$-matrix with $\mathrm{r}(A) = k$, $\vec{x} \in \{0,1\}^n$ and $\vec{b} \in \{0,1\}^k$, let $\mathcal{B}$ be the set of all indexes corresponding to non-zero columns of A. We consider some finite field $GF(Q)$ and the function $f_u : \{1, .., n\} \to GF(Q)$ with $u \in GF(Q)$ defined as follows. Let $m = \lceil \log(n+1) \rceil$ and let $\vec{a} = (a_1, a_2, ..., a_m)$ be the standard binary representation of an integer $i \in \{1, .., n\}$. Then

$$f_u(\vec{a}) \;=\; \sum_{i=1}^{m} a_i * u^{i-1} \;.$$

where $+$ and $*$ are the operations defined in $GF(Q)$. Let $\vec{a}, \vec{b} \in \mathcal{B}$ such that $\vec{a} \neq \vec{b}$ (here $\mathcal{B}$ is considered as a set of boolean sequences of length $m = \lceil \log(n+1) \rceil$). Then the Equation

$$f_u(\vec{a}) \;=\; f_u(\vec{b}) \;. \tag{16}$$

is equivalent to the following

$$\sum_{i=1}^{m} (a_i - b_i) * u^{i-1} \;.$$

The above equation can be true for at most $m - 1$ different u's. It follows that if $Q > m * q^2$ then there exists at least one element $u \in GF(Q)$ for which Eq. (16) is false for any pair $\vec{a}, \vec{b} \in \mathcal{B}$ such that $\vec{a} \neq \vec{b}$. Thus, We have the same property of Lemma 10 and, hence, we can apply the same method of the previous case (i.e. $k = \Omega(\log^2 n)$) in order to construct a solution for system $A\vec{x} = \vec{b}$. If $\mathcal{H}$ is a Hitting Set for the class $\mathcal{L}(Q, k, q)$, then the set of sequences

$$(a_{f_u(1)}, a_{f_u(2)}, ..., a_{f_u(n)}) \;, \quad u \in GF(Q) \;, \; \vec{a} \in \mathcal{H} \;, \tag{17}$$

will be a Hitting Set for $\mathcal{L}(n, k, q)$. Its size is at most $Q|\mathcal{H}|$.

Theorem 7 *Let ϵ be a constant such that $0 < \epsilon < 1/3$. If $k \geq \max\{(\log\log n)^3, q^{2/3+\epsilon}\}$, then we can construct a Hitting Set for $\mathcal{L}(n,k,q)$ whose size is bounded by $2^{O(k)}$. The time to construct the Hitting Set is $2^{O(k)}n$.*

Sketch of the proof. We can choose Q such that $Q = O(q^2 \log n)$. In this case we have $k > (\log Q)^2$ and, from Theorem 6, we have that the obtained Hitting Set is such that $|\mathcal{H}| \leq 2^{O(k)}$. It follows that the size of the Hitting Set defined in Eq (17) is bounded by

$$O(q^2 \log n 2^{O(k)}) \;=\; O(2^{O(k)}) \;.$$

□

Theorem 8 *Let ϵ be a constant, $0 < \epsilon < 1/3$. If $k \geq q^{2/3+\epsilon}$ then we can construct a Hitting Set for $\mathcal{L}(n,k,q)$ whose size is bounded by $2^{O(k)}(\log n)^2$. The time to construct the Hitting Set is $2^{O(k)}(\log n)^2 n$.*

Sketch of the proof. If $k \geq \log\log n$ then we choose $Q = O(q^2 \log n$. Since $k > (\log\log Q)^3$ we can apply Theorem 7 and obtain a Hitting Set for the class $\mathcal{L}(Q,k,q)$. By considering the construction defined in Eq. 17, we derive a Hitting Set for the class $\mathcal{L}(n,k,q)$ whose size is bounded by

$$O(q^2 \log n 2^{O(k)}) \;=\; O(2^{O(k)} \log n) \;.$$

If $k \leq \log\log n$ then we can apply the construction of case $k = \lceil \log\log n \rceil$ and, consequently, the size of the obtained Hitting Set is bounded by

$$O(q^2 \log n \; 2^{O(k)}) \;=\; O(2^{O(k)} \log n) \;+\; O(\log^2 n).$$

□

3.4 Hitting Sets for case $k \leq \min\{n^{2/3}, q^{2/3}\}$

Let $\{0,1\}^n_k$ be the set of all sequences in $\{0,1\}^n$ with at most k units. Unfortunately, when $k \leq q^{2/3}$ we are not able to construct Hitting Sets having the same (almost optimal) size of the previous cases.

Lemma 11 *The set $\{0,1\}^n_k$ is a Hitting Set for $\mathcal{L}(n,k)$.*

Lupanov [12] proved that the complexity of any sequence in $\{0,1\}^n_k$ is at most

$$(1+o(1))\frac{k \log \frac{n}{k}}{\log k} + O(\log n) \;.$$

The above lemma permits us to repeat the same construction of the Set $\mathcal{F}(n,l)$ shown in Section 2.3, thus proving the following result.

Theorem 9 *If $k \leq n^{2/3}$ (and no restriction for q), then the set $\mathcal{H} = \mathcal{F}(n, O(\frac{k \log \frac{n}{k}}{\log k} + \log n))$ is a Hitting Set for the class $\mathcal{L}(n,k,q)$. Furthermore, we have that $|\mathcal{H}| = O(2^{k \log n})$ and the time to construct $\mathcal{H}$ is polynomial in its size.*

References

[1] Allender E, Beals R, and Ogihara M. (1996), "The complexity of matrix rank and feasible systems of linear equations", in Proc. of *28-th ACM STOC*, to appear. Also available by ftp/www in ECCC (Tech. Rep. 1996).

[2] Alon N. and Spencer J.H. (1992), *The Probabilistic Method*, Wiley-Interscience Publication.

[3] Andreev, A.E.(1994), "Almost optimal hitting sets", *Dokl. Akad. Nauk RUSSIA* to appear.

[4] Andreev A. (1995), "The complexity of nondeterministic functions", *Information and Computation*, to appear.

[5] Andreev A.E., Clementi A.E.F. and Rolim J.D.P. (1996), "Hitting Sets derandomize BPP", in Proc. of *23-th ICALP* LNCS, Springer-Verlag, to appear. Also available by ftp/www in ECCC (Tech. Rep. 1996).

[6] Andreev A.E., Clementi A.E.F. and Rolim J.D.P. (1996), "Optimal bounds for the approximation of boolean functions and some applications", in Proc. of *13-th STACS*, LNCS, Springer-Verlag (1996). Also available by ftp/www in ECCC (Tech. Rep. 1995).

[7] Andreev A.E., Clementi A.E.F. and Rolim J.D.P. (1996), "On the parallel computation of boolean functions on unrelated inputs", in Proc. of *4-th Israeli Symposium on Theory of Computing and Systems (ISTCS'96)*, to appear.

[8] Chor B., and O. Goldreich (1989), "On the Power of Two-Point Based Sampling", *J. Complexity*, 5, 96-106.

[9] Karp R., Pippenger N., and Sipser M. (1982) "Time-Randomness, Tradeoff", presented at *AMS Conference on Probabilistic Computational Complexity.*

[10] Karpinski, M., and Luby, M. (1993), "Approximating the number of solutions to a GF(2) Formula", *J. Algorithms*, 14, pp.280-287.

[11] Linial N., Luby M., Saks M., and Zuckerman D. (1993), "Efficient construction of a small hitting set for combinatorial rectangles in high dimension", in *Proc. 25th ACM STOC*, 258-267.

[12] Lupanov, O.B. (1965), "About a method circuits design - local coding principle", *Problemy Kibernet.* 14, pp.31-110. (in Russian). *Systems Theory Res.* v.14, 1966 (in English).

[13] Nechiporuk, E.I. (1965), About the complexity of gating circuits for the partial boolean matrix, Dokl. Akad. Nauk SSSR 163, pp.40-42. (In Russian). English translation in Soviet Math. Docl.

[14] Nisan N. (1990), *Using Hard Problems to Create Pseudorandom Generators*, *ACM Distinguished Dissertation*, MIT Press.

[15] Nisan N., and Wigderson A. (1994), "Hardness vs Randomness", *J. Comput. System Sci.* 49, 149-167 (also presented at the *29th IEEE FOCS*, 1988).

[16] Sipser M. (1986), "Expanders, Randomness or Time vs Space", in *Proc. of 1st Conference on Structures in Complexity Theory*, LNCS 223, 325-329.