Electronic Colloquium on Computational Complexity, Report No. 58 (2000)

Approximation of Boolean Functions by Combinatorial Rectangles[*]

Martin Sauerhoff[†]
FB Informatik, LS 2, Universität Dortmund, 44221 Dortmund, Germany
e-Mail: sauerhof@ls2.cs.uni-dortmund.de

Abstract

This paper deals with the number of monochromatic combinatorial rectangles required to approximate a Boolean function on a constant fraction of all inputs, where each rectangle may be defined with respect to *its own* partition of the input variables. The main result of the paper is that the number of rectangles required for the approximation of Boolean functions in this model is very sensitive to the allowed error: There is an explicitly defined sequence of functions $f_n\colon \{0,1\}^n \to \{0,1\}$ such that f_n has rectangle approximations with a *constant* number of rectangles and one-sided error $1/3+o(1)$ or two-sided error $1/4+2^{-\Omega(n)}$, but, on the other hand, f_n requires *exponentially* many rectangles if the error bounds are decreased by an arbitrarily small constant.

Rectangle partitions and rectangle approximations with the same partition of the input variables for all rectangles have been thoroughly investigated in communication complexity theory. The complexity measures where each rectangle may have its own partition are used as tools for proving lower bounds in branching program theory. As applications of the main result, two separation results for *read-once branching programs* are presented.

First, the relationship between nondeterminism and randomness for read-once branching programs is investigated. It is shown that the analogs of the complexity classes NP and BPP defined in terms of read-once branching program size are incomparable if the error for the randomized model is bounded by a constant smaller than $1/3$. The second result is that unambiguous nondeterministic read-once branching programs, i. e., programs with at most one accepting computation path for each input, for the function f_n from the main result have exponential size. Together with a linear upper bound on the size for unrestricted nondeterminism, this implies that the analogs of the classes UP and NP for read-once branching programs are different.

Keywords: Branching programs, communication complexity, lower bounds, approximation, nondeterminism, randomness.

[*] A part of the results in this paper has been presented in *Proc. of STACS '98, LNCS 1373*, Springer 1998.

[†] This work has been supported by DFG grant We 1066/9-1.

ISSN 1433-8092

1 Introduction

In this section, we introduce rectangle approximations and the respective complexity measures studied in the paper. After this, we present the main result and discuss its applications.

1.1 Rectangle Approximations

We start with definitions of the main concepts of the paper.

Definition 1.1: Let X be a set of n variables, and let $\Pi = (X_1, X_2)$ be a *balanced partition* of X, i. e., $X = X_1 \cup X_2$, $X_1 \cap X_2 = \emptyset$ and $\big||X_1| - |X_2|\big| \le 1$. A function $r\colon \{0,1\}^n \to \{0,1\}$ defined on the variable set X is called a *(combinatorial) rectangle defined with respect to* Π (Π-*rectangle* for short), if there are functions $r_1, r_2\colon \{0,1\}^n \to \{0,1\}$ defined on X such that $r = r_1 \wedge r_2$ and r_i does not essentially depend on X_i, for $i = 1, 2$.

As usual, we identify Boolean vectors and variable assignments. If there is only one (fixed) partition of the variables, this can also be abstracted away, as usually done in communication complexity theory. The partition is then implicitly given by working with functions written as $f\colon X \times Y \to \{0,1\}$, where X and Y are finite sets. A combinatorial rectangle in $X \times Y$ is a set $R = A \times B$, where $A \subseteq X$ and $B \subseteq Y$. Obviously, this coincides with the above definition if X and Y are explicitly encoded by Boolean assignments to variables.

Here, we usually work with several partitions of the variables, though.

Definition 1.2: Let $f\colon \{0,1\}^n \to \{0,1\}$ be defined on the variable set X. A *rectangle partition representing* f is a collection of rectangles $r_1, \ldots, r_k$, where r_i is defined with respect to a balanced partition Π_i of X, such that

(i) for $i = 1, \ldots, k$, either $r_i^{-1}(1) \subseteq f^{-1}(0)$ or $r_i^{-1}(1) \subseteq f^{-1}(1)$;

(ii) $r_1^{-1}(1) \cup \cdots \cup r_k^{-1}(1) = \{0,1\}^n$ and $r_i^{-1}(1) \cap r_j^{-1}(1) = \emptyset$ for different i, j.

Define $C(f)$, the *(deterministic) rectangle complexity of* f, as the minimal number of rectangles in a rectangle partition representing f. The *(deterministic) single-partition rectangle complexity*, $C^{\mathrm{s}}(f)$, is the minimal number of rectangles in a rectangle partition for f where all rectangles are defined with respect to the *same* balanced partition of the input variables.

Rectangle partitions have been studied extensively in communication complexity theory as a combinatorial tool for proving lower bounds on the complexity of two-party protocols (see the monographs [25, 33] for definitions and a thorough introduction).

It is well-known that the measure $C^{\mathrm{s}}(f)$ (rectangle complexity in the single-partition case) is closely related to the complexity $D(f)$ of *deterministic two-party communication protocols* for f, we have:

Proposition 1.3 (Yao [59] / Halstenberg and Reischuk [24]):

$$\log C^{\mathrm{s}}(f) \le D(f), \quad \textit{and} \quad D(f) = O\Big(\big(\log C^{\mathrm{s}}(f)\big)^2\Big).$$

In this paper, we deal with "imperfect" representations of functions by rectangle partitions as described in the next definition.

Definition 1.4: Let $\mu\colon \{0,1\}^n \to [0,1]$ be an arbitrary probability distribution. A *rectangle approximation for* $f\colon \{0,1\}^n \to \{0,1\}$ *with (two-sided) error* ε *with respect to* μ is a rectangle partition representing a function $g\colon \{0,1\}^n \to \{0,1\}$ with

$$\mu\left(\{x \mid f(x) \neq g(x)\}\right) \leq \varepsilon,$$

where $0 \leq \varepsilon < 1/2$. The rectangle approximation has *one-sided error* ε if

$$\begin{aligned}&\mu\left(\{x \mid f(x) = 0 \wedge g(x) = 1\}\right) \leq \varepsilon \cdot \mu\big(f^{-1}(1)\big), \quad \text{and}\\ &\mu\left(\{x \mid f(x) = 1 \wedge g(x) = 0\}\right) = 0.\end{aligned}$$

In this case, we allow that $0 \leq \varepsilon < 1$.

For $0 \leq \varepsilon < 1/2$, define $C_{\varepsilon}^{\mathrm{A},\,\mu}(f)$, the *complexity of rectangle approximations for* f *with respect to* μ, as the minimum of $C(g)$ taken over all functions g which fulfill the above error bound for two-sided error. Define $C_{1,\,\varepsilon}^{\mathrm{A},\,\mu}(f)$ analogously for one-sided error and all $0 \leq \varepsilon < 1$. Let $C_{\varepsilon}^{\mathrm{A},\,\mu,\,\mathrm{s}}(f)$ and $C_{1,\,\varepsilon}^{\mathrm{A},\,\mu,\,\mathrm{s}}(f)$ denote the respective measures for a single balanced partition of the input variables.

We use the upper index "uniform" instead of μ for the uniform distribution over an arbitrary input space.

Observe that for all f and μ, one of the two constant functions is always a trivial approximation with two-sided error $1/2$, and the constant 0 is a trivial approximation with one-sided error 1.

The measure $C_{\varepsilon}^{\mathrm{A},\,\mu,\,\mathrm{s}}(f)$ has been analyzed in the context of so-called distributional communication complexity. The (μ,ε)*-distributional communication complexity of* f is the minimum complexity of a deterministic two-party communication protocol which correctly computes f on at least a $(1-\varepsilon)$-fraction of all inputs with respect to μ. By Proposition 1.3, it follows immediately that the logarithm of $C_{\varepsilon}^{\mathrm{A},\,\mu,\,\mathrm{s}}(f)$ is a lower bound on the (μ,ε)-distributional complexity of f. Such bounds have been proven, e. g., in [9, 17, 45, 60]. Furthermore, lower bounds on the distributional complexity directly yield lower bounds for randomized public-coin communication complexity (for details, see again the monographs [25, 33]).

Rectangle complexity with multiple partitions of the input variables has first been used explicitly by Borodin, Razborov, and Smolensky [16] for proving exponential lower bounds on the size of nondeterministic read-once branching programs (the next subsection will give definitions of this and other types of branching programs). They have considered the *nondeterministic rectangle complexity*, which is defined as the minimum number of rectangles required to *cover* the 1-inputs of the given function (i. e., rectangles may overlap). Implicitly, already the papers of Jukna [28] and Krause, Meinel, and Waack [32] contain lower bounds on this measure.

Furthermore, Borodin, Razborov, and Smolensky have introduced a generalized notion of rectangles (baptized "(k,a)-rectangles" in [30]) for proving lower bounds on nondeterministic (syntactic) read-k-times branching programs. Additional results of this kind have been obtained by Okol'nishnikova [41] and Jukna [30].

Finally, we remark that also the most recent lower bounds for linear-length branching programs [2, 3, 12, 13] employ representations of Boolean functions by appropriately defined generalizations of combinatorial rectangles.

Lower bounds on the complexity of rectangle approximations with multiple partitions of the input variables have first been proven in the conference version of this work [47]. By extending the technique of Borodin, Razborov, and Smolensky, these results have been applied to prove exponential lower bounds on the size of randomized read-once branching programs. Exponential lower bounds for randomized (syntactic) read-k-times branching programs have been obtained in the same way by using generalized rectangles instead of the usual ones.

Thathachar [51] has improved these results in order to separate the so-called (syntactic) read-k-times hierarchy for branching programs. More precisely, he has shown that there are functions for which deterministic read-$(k+1)$-times branching programs have polynomial size, while nondeterministic or randomized read-k-times branching programs have exponential size. In the context of this paper and the notation used here, Thathachar's paper implies the following:

Theorem (Thathachar [51]):

There is a sequence of explicitly defined functions $f_n\colon \{0,1\}^n \to \{0,1\}$ *such that*
(1) $C^{\mathrm{A,\,uniform}}_{1/9+\delta_n}(f_n) = O(1)$*, where* $\delta_n = 2^{-\Omega(n)}$*;*
(2) $C^{\mathrm{A,\,uniform}}_{\varepsilon}(f_n) = 2^{\Omega(\sqrt{n})}$*, for all* $\varepsilon \le (1/3)\cdot 2^{-25}$.

In the appendix of this paper, it is shown that the gap between the error bounds in this theorem can be closed: For the same function, an exponential lower bound even holds for two-sided error $1/9-\gamma_n$, for all $\gamma_n > 0$ with $\gamma_n = \Omega(1/\mathrm{Poly}(n))$. Furthermore, a similar gap between the complexity for one-sided error $1/4+o(1)$ and $1/4-\gamma'_n$, for all $\gamma'_n > 0$ with $\gamma'_n = \Omega(1/\mathrm{Poly}(n))$ is shown.

1.2 The Main Result

It is a well-known fact that the error probability of a randomized communication protocol with bounded error can be decreased below an arbitrary constant by repeating the protocol a constant number of times with independent assignments to the random bits ("probability amplification"). Thus, error probability is not a really important parameter here.

Contrary to this observation, we also learn from the known results that the error bound has a decisive influence on the complexity of rectangle approximations in the single-partition model. Razborov [45] has proven for the *disjointness function* DISJ_n (which decides whether two subsets of the set $\{1,\ldots,n\}$ are disjoint) that there is a distribution μ over the input space of DISJ_n such that $C^{\mathrm{A},\,\mu,\,\mathrm{s}}_{\varepsilon}(\mathrm{DISJ}_n) = 2^{\Omega(n)}$ for all constants $\varepsilon < 1/180$. On the other hand, $\mu\left(\mathrm{DISJ}_n^{-1}(1)\right) = 3/4$, and thus the function is trivially approximated by the constant 1 with error $1/4$ with respect to μ. Hence, we have an unbounded increase of the complexity if the error is decreased by some small positive constant.

For the *inner product function over* $\mathbb{Z}_p$, p a prime, one obtains a similar increase of complexity, but for an arbitrarily small constant decrease of the error bound. This function checks whether the standard inner product of two n-bit vectors is different from 0 in $\mathbb{Z}_p$. Babai, Kimmel, and Hayes [10] have proven that exponentially many rectangles are required to approximate the inner product function over $\mathbb{Z}_p$ in the single partition model and with respect to the uniform distribution if the error is bounded by a constant smaller than $1/p$, whereas the function is trivially approximated by the constant 1 with error bounded by $1/p + 2^{-\Omega(n)}$.

We show here that the described sensitivity to the error bound also occurs in the general model where the partition of the input variables may be chosen differently for different rectangles, even for the uniform distribution over the input space, and even when the error is decreased only by an arbitrary small positive constant. We consider the following function, where $[P]$ is used to denote the Boolean function which is equal to 1 if the predicate P is true, and 0 otherwise.

Definition 1.5: Define the function $\mathrm{MS}_n\colon \{0,1\}^{n^2} \to \{0,1\}$ ("ModSum") on the $n \times n$ matrix $X = (x_{ij})_{1\le i,j\le n}$ of Boolean variables by

$$\mathrm{MS}_n(X) := \mathrm{RT}_n(X) \vee \mathrm{CT}_n(X),$$

where $\mathrm{RT}_n\colon \{0,1\}^{n^2} \to \{0,1\}$ ("RowTest") is defined by

$$\mathrm{RT}_n(X) := \sum_{i=1}^{n} \bigl[x_{i,1} + \cdots + x_{i,n} \equiv 0 \bmod 3\bigr] \bmod 2$$

and $\mathrm{CT}_n(X) := \mathrm{RT}_n\bigl(X^\top\bigr)$ ("ColumnTest").

We prove the following upper and lower bounds on the complexity of rectangle approximations for MS_n with respect to the uniform distribution.

Theorem 1.6: *Let* $N = n^2$ *(the input size of* MS_n*).*

(1) $C^{\mathrm{A,\,uniform}}_{1,\,1/3+\delta_N}(\mathrm{MS}_n) = O(1)$ *and* $C^{\mathrm{A,\,uniform}}_{1/4+\delta'_N}(\mathrm{MS}_n) = 1$, *where* $\delta_N = o(1)$ *and* $\delta'_N = 2^{-\Omega(\sqrt{N})}$;

(2) $C^{\mathrm{A,\,uniform}}_{1,\,1/3-\gamma_N}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$ *and* $C^{\mathrm{A,\,uniform}}_{1/4-\gamma'_N}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$, *for all* $\gamma_N, \gamma'_N > 0$ *with* $\gamma_N, \gamma'_N = \Omega(1/\mathrm{Poly}(N))$.

We present another variant of the above main theorem where we allow ourselves to choose a *nonuniform* distribution over the input space instead of the uniform one. By adjusting the distribution, we obtain larger bounds on the error for which we still get exponential lower bounds on the complexity of rectangle approximations.

Theorem 1.7: *Let* $N = n^2$. *There is a probability distribution* $\mu\colon \{0,1\}^N \to [0,1]$ *such that*

(1) $C^{\mathrm{A},\,\mu}_{1,\,1/2+\delta_N}(\mathrm{MS}_n) = O(1)$ *and* $C^{\mathrm{A},\,\mu}_{1/3+\delta'_N}(\mathrm{MS}_n) = 1$, *where* $\delta_N = o(1)$ *and* $\delta'_N = 2^{-\Omega(\sqrt{N})}$;

(2) $C^{\mathrm{A},\,\mu}_{1,\,1/2-\gamma}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$ *and* $C^{\mathrm{A},\,\mu}_{1/3-\gamma'}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$, *for all constants* $\gamma, \gamma' > 0$.

1.3 Applications for Branching Programs

One of the most important and, seemingly, also most difficult unresolved tasks in complexity theory is to find out how the computational power of randomized algorithms relates to that of deterministic and nondeterministic algorithms. If one believes that the concepts determinism, randomness, and nondeterminism differ, one has to prove separation results for complexity classes such as P, NP, and BPP, which still seem to be out of reach today due to the lack of appropriate techniques for proving lower bounds.

One approach in this situation is to resort to relativization or to results based on unproven (but plausible) assumptions. On the other hand, one may also study alternative models of computation which promise to be easier to handle by combinatorial tools than the somewhat "unwieldy" classical Turing machine.

The latter approach has been quite successful for some nonuniform models of computation, such as circuits, communication protocols, and branching programs. Branching programs allow to describe sequential computations in an especially handy way. Furthermore, complexity classes defined in terms of other well-known nonuniform models of computation may be equivalently characterized in terms of branching programs.

Definition 1.8: A *(deterministic) branching program (BP)* on the variable set $\{x_1, \dots, x_n\}$ is a directed acyclic graph with one source and two sinks labeled by the constants 0 and 1, resp. Each non-sink node is labeled by a variable x_i and has exactly two outgoing edges carrying labels 0 and 1, resp. This graph represents a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ in the following way. To compute $f(a)$ for some input $a \in \{0,1\}^n$, start at the source node. For a non-sink node labeled by x_i, check the value of this variable and follow the edge which is labeled by this value (this is called a "test of variable x_i"). Iterate this until a sink node is reached. The value of f on input a is the value of the reached sink. For a fixed input a, the sequence of nodes visited in this way is uniquely determined and is called the *computation path for a*. The *size* of a branching program G is the number of its nodes and is denoted by $|G|$.

Usually, we consider sequences of BPs representing sequences $(f_n)_{n\in\mathbb{N}}$ of Boolean functions, where $f_n\colon \{0,1\}^n \to \{0,1\}$. In order to simplify notation, we will frequently talk of functions where we really mean "sequences of functions." As a further convention, we omit indices indicating the input size wherever no confusion arises.

Nondeterministic and randomized BPs may be defined by introducing additional "guessing nodes" or "probabilistic nodes." For the moment, we may imagine that, when reaching such nodes during a computation, the successor on the computation path is nondeterministically guessed or determined by flipping a fair coin, resp. More precise definitions are given later on.

It is a well-known fact that sequences of functions which can be computed by BPs of polynomial size can also be computed within logarithmic space on the nonuniform variant of Turing machines and vice versa [18, 42]. Hence, it is an important problem to prove superpolynomial lower bounds on the size of BPs for explicitly defined sequences of functions.

So far, not even superlinear lower bounds for deterministic BPs are known. Nevertheless, superpolynomial and exponential lower bounds could be proven for several restricted types of BPs.

One major goal in branching program theory is to extend the available techniques to more and more general models. The last years have brought some astonishing progress along this line. We give only a brief overview here (see, e. g., [56] for more details).

A *read-once branching program* is a (deterministic) BP where each variable may appear at most once on each path from the source to a sink. This restricted type of BPs has been the first one for which exponential lower bounds on the size could be proven [55, 61], and it has been extensively studied. A general technique presented by Simon and Szegedy [50] summarizes what is behind most of the known proofs of lower bounds for deterministic read-once BPs.

More recently, Gál [19] and Bollig and Wegener [15] have obtained exponential lower bounds on the size of deterministic read-once BPs even for "very simple" functions, i. e., functions with polynomial size CNF and DNF, resp. Exponential lower bounds for nondeterministic read-once BPs have been proven by Jukna [28], Krause, Meinel, and Waack [32], Borodin, Razborov, and Smolensky [16]. The randomized case has been first handled in the conference version of this work. Further results for the nondeterministic and randomized variant of read-once BPs are discussed below.

Read-once BPs are the special case $k = 1$ of *(syntactic) read-k-times branching programs* where each variable may appear at most k times on each path from the source to a sink. An even more general model are *semantic read-k-times branching programs* where the restriction on the read access is only required to hold for all computation paths (instead of *all* paths). Results for these variants of BPs have been proven in [13, 16, 30, 41, 47, 51].

The latest records in the competition for lower bounds on the size of less and less restricted variants of BPs have been achieved for *linear-length branching programs*, which are BPs where the length of the longest path from the source to one of the sinks is bounded by a linear function in the input size. Beame, Saks, and Thathachar [13], Ajtai [2, 3] and Beame, Saks, Sun, and Vee [12] have proven exponential lower bounds for this model.

In this paper, we focus on another topic of branching program theory. Given the polynomial relationship between the size of BPs and the logarithm of the space complexity for nonuniform Turing machines, BPs are also a suitable model for comparing the power of determinism, nondeterminism, and randomness in the nonuniform, space-restricted setting. Complexity classes for BPs analogous to P, NP, BPP etc. are defined by replacing polynomial time complexity with polynomial branching program size in the respective model. Due to the difficulty of proving separation results, it is not too surprising that already for the more restricted variants of BPs there are several open problems concerning the relationship between the different modes of computation. Here we deal with such problems for read-once BPs. First, we summarize known results.

The relationship between determinism and nondeterminism has been studied already in the early papers on read-once BPs. Jukna [28] and Krause, Meinel, and Waack [32] have shown that the *permutation matrix function* requires exponential size for nondeterministic read-once BPs, whereas one easily proves that its complement has linear size in the same model [32]. More recently, Jukna, Razborov, Savický, and Wegener [27] have improved upon this by presenting sequences of functions with exponential size for deterministic read-once BPs which are even contained in the analog of the complexity class $\mathrm{NP} \cap \mathrm{coNP}$ for read-once BPs.

Initial results on the relationship between determinism and randomness have been obtained by Ablayev and Karpinski [1], who have considered a randomized variant of *OBDDs* (*ordered binary decision diagrams*). An OBDD is a restricted read-once BPs where the sequence in which the variables appear along each path from the source to a sink has to be consistent with a fixed order. Ablayev and Karpinski have presented a sequence of functions $f_n\colon \Sigma^n \to \{0, 1\}$, where $|\Sigma| = 4$, which can be represented by randomized OBDDs with small, one-sided error and polynomial size, but which require exponential size for deterministic (unrestricted) read-once BPs. (Their upper bound also works for a Boolean encoding of the functions f_n.)

The results contributed in this paper are described in the following.

NP versus BPP for Read-Once BPs. In the classical setting, it seems to be unlikely that NP $\subseteq$ BPP, since Ko [31] has shown that this would imply NP $=$ RP as well as a collapse of the polynomial time hierarchy to BPP. On the other hand, BPP may well be contained in NP, we may even have P $=$ BPP. Some support for the conjecture that P $=$ BPP is provided by recent derandomization results for BPP-algorithms (see, e. g., [7, 26, 57]).

Analogous questions have been studied for space-bounded complexity. Already Gill [20] has shown that NL $=$ RL, but it is nevertheless unknown whether the classes NL and BPL are different. (RL and BPL are the classes of languages which can be decided by probabilistic Turing machines with bounded one-sided and two-sided error, resp., using at most logarithmic space.) The situation for the nonuniform setting is the same in this respect, we have NL/Poly $=$ RL/Poly $\subseteq$ BPL/Poly, but it is open whether this inclusion is proper.

Here we deal with the relationship between nondeterminism and randomness for read-once BPs. We first prove that the permutation matrix function has randomized OBDDs with small, one-sided error of polynomial size. Together with exponential lower bound for nondeterministic read-once BPs from [28, 32], we obtain that P $\subsetneqq$ RP and BPP $\not\subseteq$ NP for read-once BPs. With respect to the relationship between determinism and randomness, this has been strengthened in [49] by an exponential gap between the size of deterministic read-once BPs and randomized read-once BPs with *zero error*, implying that P $\subsetneqq$ ZPP for read-once BPs.

On the other hand, we prove that the function considered in the main result of the paper requires exponential size for randomized read-once BPs with two-sided error bounded by a constant smaller than $1/3$ or one-sided error bounded by a constant smaller than $1/2$, while it can be represented in linear size by nondeterministic read-once BPs. Hence, the analog of the class NP is not contained in BPP and RP $\subsetneqq$ NP if the error allowed for the randomized models is not too large.

Randomized read-once BPs also show the high sensitivity on the error bound already observed for rectangle approximations. Our results imply that decreasing the bound imposed on the error probability of a randomized read-once BP by an arbitrarily small constant may result in an exponential blowup of the size. This is contrary to the situation for probabilistic Turing machines and for randomized general BPs, where the error probability may be decreased below an arbitrary small constant while maintaining polynomial size by “probability amplification.”

UP versus NP for Read-Once BPs. Valiant [52] has introduced the subclass UP of NP which contains the languages decidable by nondeterministic Turing machines with *at most one* accepting computation for each input. Obviously, $\mathrm{P} \subseteq \mathrm{UP} \subseteq \mathrm{NP}$, but it is not known whether any of these inclusions are proper. The results of Valiant and Vazirani [53] provide some evidence for the hypothesis $\mathrm{P} \subsetneqq \mathrm{UP} = \mathrm{NP}$. On the other hand, there is an oracle A with $\mathrm{P}^A = \mathrm{UP}^A \subsetneqq \mathrm{NP}^A$ [14]. Finally, we note that separating P from UP would also have consequences for the area of cryptography: it is known that there are polynomial time one-way functions if and only if $\mathrm{P} \subsetneqq \mathrm{UP}$ [23].

For the setting of nonuniform, logarithmically space-bounded computations, Allender and Reinhardt [6] have shown that unambiguous nondeterminism is indeed as powerful as the unrestricted version: the nondeterministic analogs of the classes UL and NL coincide, i. e., we have UL/Poly $=$ NL/Poly.

A contrary result holds for two-party communication protocols. Yannakakis [58] has proven that deterministic communication complexity is at most quadratically larger than unambiguous communication complexity. Furthermore, the results of Mehlhorn and Schmidt [36] can be exploited to obtain a function of input size n which has nondeterministic communication complexity $O(\log n)$ and unambiguous communication complexity $\Omega(n)$ (unpublished observation due to M. Dietzfelbinger).

The latter result also implies that the concepts of unambiguous and unrestricted nondeterminism no longer coincide for a very restricted type of Turing machines, namely for nonuniform Turing machines with logarithmic space-bound and *one-way access* to their input tape. More precisely, 1-UL/Poly $\subsetneqq$ 1-NL/Poly, where the prefix “1-” to the complexity classes indicates one-way access to the input tape. (This unpublished result is attributed to M. Dietzfelbinger by Allender *et al.* [5]. The article also contains an improved version of the result.)

Furthermore, the above fact can also be formulated in terms of restricted BPs. Analogous to the proof for one-way Turing machines, one obtains that there is a sequence of functions which has nondeterministic OBDDs of polynomial size, but for which unambiguous OBDDs require exponential size.

We complement these results here by showing that the function from the main result on rectangle approximations requires exponential size for unambiguous read-once BPs. Together with a linear upper bound on the size for (unrestricted) nondeterministic read-once BPs, this especially implies that the analogs of the classes UP and NP are different also for read-once BPs.

Overview. The rest of the paper is organized as follows. In the next section, we describe the technique used for proving lower bounds on the complexity of rectangle approximations. Section 3 and Section 4 show how this technique is applied for the proof of the main result and the improved version for a nonuniform distribution over the input space, resp. Then we deal with the implications of these results for read-once BPs (Section 5). In Sections 6, 7 and 8 we fill in some technical details of the proof of the main result left out at the beginning. We first give a short introduction into some algebraic concepts and techniques (Section 6), and then apply these to prove two central lemmas (Section 7 and 8). Finally, in the appendix we improve Thathachar’s result for rectangle approximations with respect to the error bounds.

2 Proof Technique

In this section, we describe the technique used for proving lower bounds on the complexity of rectangle approximations. This is an extension of Yao's technique [60] for proving lower bounds on the distributional communication complexity. We consider approximations with respect to an arbitrary probability distribution μ over the input space here.

For proving large lower bounds on the complexity of rectangle approximations, we look for functions f with the following two properties:

- The fraction of 0-inputs of f (with respect to the given distribution μ on the input space) does not tend to 0 or 1 with increasing input size, i. e., there are constants $\alpha > 0$ and $\beta < 1$ such that $\alpha < \mu\bigl(f^{-1}(0)\bigr) < \beta$.
- For each rectangle r, $r^{-1}(1)$ only contains "few" 0-inputs compared to the overall size of the rectangle. More precisely, if $\mu\bigl(r^{-1}(1)\bigr)$ is not exponentially small with respect to the input size of f, then the ratio $\mu\bigl(r^{-1}(1) \cap f^{-1}(0)\bigr) / \mu\bigl(r^{-1}(1)\bigr)$ is bounded by a constant smaller than 1.

The role of 0- and 1-inputs may be swapped. We concentrate on 0-inputs in this description since we will apply the technique in this way later on.

The first property is obviously necessary; otherwise, we could approximate f by one of the constant functions with small error. We first give a formal definition of the second property and then show that both properties together ensure that rectangle approximations for f have large complexity.

Definition 2.1: Let $f : \{0,1\}^n \to \{0,1\}$ be defined on the variable set X, $|X| = n$. Let μ be an arbitrary probability distribution over $\{0,1\}^n$. Suppose that there are a constant α, $0 \le \alpha \le 1$, and a sequence of real numbers $(\delta_n)_{n\in\mathbb{N}}$, such that for each rectangle r defined with respect to an arbitrary balanced partition of X,

$$\mu\bigl(r^{-1}(1) \cap f^{-1}(0)\bigr) \le \alpha \cdot \mu\bigl(r^{-1}(1)\bigr) + \delta_n. \tag{LD}$$

Then we say that f has the *low* 0*-density property with respect to rectangles and to the distribution* μ *with parameters* α *and* δ_n.

In the applications of the proof technique, the values of δ_n in this definition will be exponentially small in n. The following theorem summarizes the proof technique.

Theorem 2.2: *Let* $f : \{0,1\}^n \to \{0,1\}$ *be defined on the variable set* X, $|X| = n$. *Let* μ *be an arbitrary probability distribution over* $\{0,1\}^n$. *Suppose that* f *has the low* 0*-density property with parameters* α *and* δ_n. *Then*

(1) $C^{\mathrm{A}}_{1,\varepsilon}(f) \ge \delta_n^{-1} \cdot \Bigl((1-\alpha) \cdot \mu\bigl(f^{-1}(0)\bigr) - \alpha \cdot \varepsilon \cdot \mu\bigl(f^{-1}(1)\bigr)\Bigr)$, *for all* $\varepsilon < 1$*; and*

(2) $C^{\mathrm{A}}_{\varepsilon}(f) \ge \delta_n^{-1} \cdot \Bigl((1-\alpha) \cdot \mu\bigl(f^{-1}(0)\bigr) - \max(1-\alpha, \alpha) \cdot \varepsilon\Bigr)$, *for all* $\varepsilon < 1/2$.

Proof: For technical reasons, it is easier to start with the second part.

Part (2): Let P be a rectangle partition representing $g\colon \{0,1\}^n \to \{0,1\}$, where the rectangles of the partition are defined with respect to balanced partitions of the variable set X. Suppose that P is a rectangle approximation for f with two-sided error ε with respect to μ. For $c \in \{0,1\}$, let $r_1^c, \ldots, r_{k_c}^c$ be the rectangles on which g computes the result c. We only work with the sets of input assignments belonging to these rectangles, defined by $R_i^c := \bigl(r_i^c\bigr)^{-1}(1)$ for all c and i.

In the following, we derive a lower bound on k_0. First, observe that, since the sets R_i^0, R_i^1 form a partition of $\{0,1\}^n$,

$$\mu\bigl(f^{-1}(0)\bigr) \;=\; \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(0)\bigr) + \sum_{i=1}^{k_1} \mu\bigl(R_i^1 \cap f^{-1}(0)\bigr). \tag{1}$$

Define

$$e_0 \;:=\; \sum_{i=1}^{k_1} \mu\bigl(R_i^1 \cap f^{-1}(0)\bigr), \quad \text{and} \quad e_1 \;:=\; \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(1)\bigr).$$

Then we have $e_0 + e_1 \le \varepsilon$, since P approximates f with error ε with respect to μ.

Summing up inequality (LD) from the low 0-density property for all rectangles R_i^0, $i = 1, \ldots, k_0$, yields

$$k_0 \cdot \delta_n \;\ge\; \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(0)\bigr) - \alpha \cdot \sum_{i=1}^{k_0} \mu\bigl(R_i^0\bigr).$$

Using that $\mu\bigl(R_i^0\bigr) = \mu\bigl(R_i^0 \cap f^{-1}(0)\bigr) + \mu\bigl(R_i^0 \cap f^{-1}(1)\bigr)$ for all i, we may rewrite this as

$$k_0 \cdot \delta_n \;\ge\; (1-\alpha) \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(0)\bigr) - \alpha \cdot \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(1)\bigr).$$

By Equation (1),

$$k_0 \cdot \delta_n \;\ge\; (1-\alpha) \left[\mu\bigl(f^{-1}(0)\bigr) - \sum_{i=1}^{k_1} \mu\bigl(R_i^1 \cap f^{-1}(0)\bigr) \right] - \alpha \cdot \sum_{i=1}^{k_0} \mu\bigl(R_i^0 \cap f^{-1}(1)\bigr),$$

and thus, using the definitions of e_0 and e_1,

$$k_0 \cdot \delta_n \;\ge\; (1-\alpha) \cdot \mu\bigl(f^{-1}(0)\bigr) - (1-\alpha) \cdot e_0 - \alpha \cdot e_1. \tag{2}$$

We still have to take into account that $e_0 + e_1 \le \varepsilon$. The right hand side of Inequality (2) is minimized by maximizing $(1-\alpha) \cdot e_0 + \alpha \cdot e_1$ subject to the constraint $e_0 + e_1 \le \varepsilon$. It follows that

$$k_0 \cdot \delta_n \;\ge\; (1-\alpha) \cdot \mu\bigl(f^{-1}(0)\bigr) - \max(1-\alpha, \alpha) \cdot \varepsilon.$$

This yields the claimed lower bound.

Part (1): We can simply re-use the above proof by exploiting that, in the case of one-sided error, we have $e_0 = 0$ and $e_1 \leq \varepsilon \cdot \mu\big(f^{-1}(1)\big)$. Inequality (2) turns into

$$k_0 \cdot \delta_n \;\geq\; (1-\alpha) \cdot \mu\big(f^{-1}(0)\big) - \alpha \cdot \varepsilon \cdot \mu\big(f^{-1}(1)\big),$$

which gives the desired result. □

3 Proof of the Main Result

First, we present three combinatorial lemmas which will be used later on. Then we prove Theorem 1.6. For the convenience of the reader, we repeat the definition of the function for which we are going to prove the main result.

Definition 3.1: Define the function $\mathrm{MS}_n \colon \{0,1\}^{n^2} \to \{0,1\}$ on the $n \times n$ matrix $X = (x_{ij})_{1 \leq i,j \leq n}$ of Boolean variables by

$$\mathrm{MS}_n(X) := \mathrm{RT}_n(X) \vee \mathrm{CT}_n(X),$$

where $\mathrm{RT}_n \colon \{0,1\}^{n^2} \to \{0,1\}$ is defined by

$$\mathrm{RT}_n(X) := \sum_{i=1}^{n} \big[x_{i,1} + \cdots + x_{i,n} \equiv 0 \bmod 3\big] \bmod 2$$

and $\mathrm{CT}_n(X) := \mathrm{RT}_n\big(X^\top\big)$.

The first lemma below deals with properties of balanced partitions of the input variables of MS_n.

Lemma 3.2: *Let $\Pi = (X_1, X_2)$ be a partition of the variables in the $n \times n$ matrix $X = (x_{ij})_{1 \leq i,j \leq n}$ with $\big||X_1| - |X_2|\big| \leq 1$. Call a row or a column of X* mixed *if X_1 contains at least 2 and at most $n-2$ variables of it.*

Let $\beta < 1/\sqrt{2}$ be a constant. Then for n large enough, there are either at least $\lfloor \beta n \rfloor$ mixed rows or columns with respect to Π.

Proof: Call a row *dense* if it contains at least $n-1$ variables from X_1, and *sparse* if it contains at most one variable from X_1. Observe that a row cannot be both dense and sparse, and that a row is mixed exactly if it is neither dense nor sparse. Let r_d, r_s be the number of rows which are dense and sparse, resp.

Case 1: $r_d, r_s \geq 2$. Let i_1, i_2 be two different dense rows. Then at most two variables in these rows are not from X_1, and thus there are at least $n-2$ columns j with $x_{i_1,j} \in X_1$ and $x_{i_2,j} \in X_1$. Analogously, there are two sparse rows i_3, i_4 and at least $n-2$ columns j with $x_{i_3,j} \notin X_1$ and $x_{i_4,j} \notin X_1$. It follows that there are at least $n-4$ mixed columns.

Case 2: $r_d \leq 1$ or $r_s \leq 1$. W. l. o. g., assume that the latter occurs (otherwise, swap the roles of X_1 and X_2 in the whole proof). If $r_d \leq (1-\beta)n - 1$, $n - (r_d + r_s) \geq \beta n$ rows are mixed and we are finished. Hence, assume that $r_d > (1-\beta)n - 1$ for the following.

Suppose that n is large enough such that $(1-\beta)n - 1 \geq 1$. Then we have $r_d \geq 2$. Let $I \subseteq \{1, \ldots, n\}$ be the set of indices of dense rows, $|I| = r_d$. Define

$$J := \{j \mid \text{there are } i_1, i_2 \in I, i_1 \neq i_2 \text{ such that } x_{i_1,j} \in X_1 \text{ and } x_{i_2,j} \in X_1\}.$$

Since the rows with index in I are dense, the number of X_2-variables in these rows is bounded from above by $|I| = r_d$. On the other hand, the total number of X_2-variables in all columns whose index is not in J is bounded from below by $(r_d - 1) \cdot (n - |J|)$. Putting these two bounds together, we get

$$|J| \geq n - \frac{r_d}{r_d - 1}.$$

Since $r_d \geq 2$, it follows that $|J| \geq n - 2$.

Now each column with index $j \in J$ is mixed if less than $n - r_d - 1$ variables x_{ij} with $i \notin I$ are contained in X_1. Let c be the number of columns in J which are not mixed and thus contain at least $n - r_d - 1$ additional X_1-variables in rows outside of I.

Putting the above results together, we have obtained the following lower bound on the total number of variables in X_1:

$$|X_1| \geq r_d \cdot (n-1) + c \cdot (n - r_d - 1).$$

On the other hand, $|X_1| \leq n^2/2 + 1/2$. Hence,

$$\frac{1}{2}\left(n^2 + 1\right) \geq r_d \cdot (n-1) + c \cdot (n - r_d - 1).$$

Solving for c, we obtain

$$c \leq \frac{\left(n^2+1\right)/2 - (n-1)r_d}{n - 1 - r_d}. \qquad (*)$$

(Observe that $r_d < n - 1$, since $r_d \cdot (n-1) \leq |X_1| \leq \left(n^2+1\right)/2$.)

It is easy to verify that the right hand side of $(*)$ is decreasing in r_d if n is large enough. Using that $r_d > (1-\beta)n - 1$, we obtain

$$\begin{aligned} c &< \frac{\left(n^2+1\right)/2 - \big((1-\beta)n - 1\big)(n-1)}{\beta n} \\ &= \left(1 - \frac{1}{2\beta}\right)n + \frac{2}{\beta} - 1 - \frac{1}{2\beta n}. \end{aligned}$$

By the above definitions, we have at least $|J| - c \geq n - 2 - c$ mixed columns. We have shown that

$$n - 2 - c > \frac{1}{2\beta} \cdot n - \frac{2}{\beta} - 1 + \frac{1}{2\beta n} = \frac{1}{2\beta} \cdot n - O(1).$$

Since β is a constant with $\beta < 1/\sqrt{2}$, it follows that $n - 2 - c \geq \beta n$ for n large enough. □

The next two lemmas constitute the core part of the proof of the main result and are required to apply the technique presented in the last section. We choose the uniform distribution on the input space here.

As already remarked, we require that the fraction of 0-inputs of the function under consideration does not tend to 0 or 1 for increasing input size. The following statement implies an asymptotically tight bound on the fraction of 0-inputs for MS_n (remember that $\mathrm{MS}_n = \mathrm{RT}_n \vee \mathrm{CT}_n$):

Lemma 3.3: *Let* $\xi, \eta \in \{0, 1\}$. *Then*

$$\left|\mathrm{RT}_n^{-1}(\xi) \cap \mathrm{CT}_n^{-1}(\eta)\right| \cdot 2^{-n^2} = 1/4 \pm 2^{-\Omega(n)}.$$

The proof of this lemma is technically involved and therefore deferred to its own section, Section 7.

By Lemma 3.2, we know that for an arbitrary balanced partition of the input matrix of MS_n, there are either many mixed rows or columns. We claim that in the first case, the function RT_n (RowTest) is hard to approximate with respect to the given partition, whereas in the second case CT_n (ColumnTest) is hard to approximate. For this, we prove an upper bound on the *discrepancy* of the respective function.

Definition 3.4: Let $f : \{0, 1\}^n \to \{0, 1\}$ be an arbitrary function, and let Π be an arbitrary balanced partition of the input variables of f. For an arbitrary Π-rectangle r define the *discrepancy of f with respect to r*, $\mathrm{Disc}(f, r)$, by

$$\mathrm{Disc}(f, r) := \Big|\left|f^{-1}(1) \cap r^{-1}(1)\right| - \left|f^{-1}(0) \cap r^{-1}(1)\right|\Big| \cdot 2^{-n}.$$

Let $\mathrm{Disc}(f, \Pi)$ denote the maximum of $\mathrm{Disc}(f, r)$ taken over all Π-rectangles r.

The following technical lemma provides a bound on the discrepancy of a suitable class of subfunctions of RT_n and CT_n.

Lemma 3.5: *Let* $c = (c_0, c_1, \ldots, c_m)$, *where* $c_0 \in \mathbb{Z}_2$ *and* $c_1, \ldots, c_m \in \mathbb{Z}_3$. *Define the function* $\mathrm{RT}_c^* \colon \{0, 1\}^{2m} \times \{0, 1\}^{2m} \to \{0, 1\}$ *on vectors* $u^1, u^2, v^1, v^2 \in \{0, 1\}^m$ *by*

$$\mathrm{RT}_c^*\big((u^1, u^2), (v^1, v^2)\big) := \left[\sum_{i=1}^{m} \left[u_i^1 + u_i^2 + v_i^1 + v_i^2 \equiv c_i \bmod 3\right] \equiv c_0 \bmod 2\right].$$

Let $\Pi = (U, V)$, *where* $U = \left\{u_i^1, u_i^2 \mid i = 1, \ldots, m\right\}$ *and* $V = \left\{v_i^1, v_i^2 \mid i = 1, \ldots, m\right\}$. *Then*

$$\mathrm{Disc}(\mathrm{RT}_c^*, \Pi) \le 2^{-m} + 3^{-m}.$$

A proof of this fact is given later on in Section 8. Intuitively, the functions RT_c^* are "very similar" to the well-known *inner product function* (the standard inner product in $\mathbb{Z}_2$), for which discrepancy bounds have been proven in communication complexity theory (see, e. g., [17]).

We can now easily apply this fact to get the desired upper bounds on the discrepancy of RT_n or CT_n.

Lemma 3.6 (Discrepancy Lemma): *Let $\Pi = (X_1, X_2)$ be a partition of the variables in the matrix $X = (x_{ij})_{1 \le i,j \le n}$ with $\big||X_1| - |X_2|\big| \le 1$. Suppose that m rows of X are mixed with respect to Π. Then*

$$\mathrm{Disc}(\mathrm{RT}_n, \Pi) \le 2^{-m} + 3^{-m}.$$

An analogous statement holds for mixed columns instead of rows and CT_n *instead of* RT_n.

Proof: We only prove the claim for RT_n. For the ease of notation, we omit subscripts indicating the input size in the following. For a function f and an assignment a to some variables of f, we use f_a to denote the subfunction of f obtained by setting variables to constants according to a.

Let r be an arbitrary Π-rectangle. Our goal is to prove that $\mathrm{Disc}(\mathrm{RT}, r) \le 2^{-m} + 3^{-m}$.

In each of the m mixed rows of the matrix X with respect to Π, choose two different variables from X_1 and two different variables from X_2. Let $X_1' \subseteq X_1$ and $X_2' \subseteq X_2$ be the sets of the chosen variables. Observe that $\left|X_1' \cup X_2'\right| = 4m$. Let $\Pi' := \left(X_1', X_2'\right)$.

We set all variables in $X \backslash \left(X_1' \cup X_2'\right)$ to constants according to an arbitrary assignment a. Then all variables except those in $X_1' \cup X_2'$ are fixed. We consider the subfunction RT_a of RT defined on the remaining variables in $X_1' \cup X_2'$, and the subfunction r_a of the rectangle r, which is a rectangle with respect to the balanced partition Π'.

It is easy to see that there are constants $c_0 \in \mathbb{Z}_2$ and $c_1, \ldots, c_m \in \mathbb{Z}_3$ (depending on the assignment a) such that

$$\mathrm{RT}_a\big(X_a\big) = \left[\sum_{i=1}^{m} \left[u_i^1 + u_i^2 + v_i^1 + v_i^2 \equiv c_i \bmod 3\right] \equiv c_0 \bmod 2\right],$$

where u_i^1, u_i^2 and v_i^1, v_i^2 are used to denote the unfixed X_1- and X_2-variables, resp., in the ith mixed row, for $i = 1, \ldots, m$. Hence, the subfunction RT_a is of the type described in Lemma 3.5, and we have

$$\mathrm{Disc}\big(\mathrm{RT}_a, \Pi'\big) \le 2^{-m} + 3^{-m}.$$

Since r_a is a Π'-rectangle, we obtain

$$\Big|\left|r_a^{-1}(1) \cap \mathrm{RT}_a^{-1}(0)\right| - \left|r_a^{-1}(1) \cap \mathrm{RT}_a^{-1}(1)\right|\Big| \cdot 2^{-4m} \le 2^{-m} + 3^{-m}. \qquad (*)$$

This statement holds for all assignments a to $X \backslash \big(X_1' \cup X_2'\big)$. Due to the law of total probability,

$$\sum_{\substack{\text{ass. } a \text{ to} \\ X \backslash (X_1' \cup X_2')}} \left|r_a^{-1}(1) \cap \mathrm{RT}_a^{-1}(c)\right| \cdot 2^{-4m} \cdot 2^{-(n^2 - 4m)} = \left|r^{-1}(1) \cap \mathrm{RT}^{-1}(c)\right| \cdot 2^{-n^2},$$

for $c \in \{0, 1\}$. Applying this to $(*)$ gives the desired result,

$$\mathrm{Disc}(\mathrm{RT}, r) = \Big|\left|r^{-1}(1) \cap \mathrm{RT}^{-1}(0)\right| - \left|r^{-1}(1) \cap \mathrm{RT}^{-1}(1)\right|\Big| \cdot 2^{-n^2} \le 2^{-m} + 3^{-m}.$$

□

Now we are prepared to prove the main theorem, which we cite below for the convenience of the reader.

Theorem 1.6: *Let $N = n^2$ (the input size of* MS_n*).*

(1) $C^{\mathrm{A,\,uniform}}_{1,\,1/3+\delta_N}(\mathrm{MS}_n) = O(1)$ *and* $C^{\mathrm{A,\,uniform}}_{1/4+\delta'_N}(\mathrm{MS}_n) = 1$*, where* $\delta_N = o(1)$ *and* $\delta'_N = 2^{-\Omega(\sqrt{N})}$*;*

(2) $C^{\mathrm{A,\,uniform}}_{1,\,1/3-\gamma_N}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$ *and* $C^{\mathrm{A,\,uniform}}_{1/4-\gamma'_N}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$*, for all* $\gamma_N, \gamma'_N > 0$ *with* $\gamma_N, \gamma'_N = \Omega(1/\mathrm{Poly}(N))$.

Proof of Theorem 1.6: *Part (1):* The bound for two-sided error follows directly from Lemma 3.3: The constant function 1, which is itself a rectangle, approximates MS_n with two-sided error at most $1/4 + 2^{-\Omega(n)} = 1/4 + 2^{-\Omega(\sqrt{N})}$.

It remains to handle the case of one-sided error. Let $\Pi_{\mathrm{rows}} = (X_1, X_2)$ be a balanced partition of X where both parts X_1, X_2 only contain complete rows of X, except possibly for one row which is divided "as equally as possible." It is easy to see that RT_n can be computed by a deterministic two-party communication protocol with respect to Π_{rows} using at most 3 bits of communication. By Proposition 1.3, this yields a rectangle partition P representing RT_n with at most 8 rectangles. The 1-inputs for MS_n in $\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)$ are the only inputs mapped to the wrong value by P. Using Lemma 3.3, we can thus bound the relative error of P on the 1-inputs of MS_n by

$$\frac{\left|\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\right|}{\left|\mathrm{MS}_n^{-1}(1)\right|} \le \frac{1/4 + \varepsilon_n}{3/4 - \varepsilon'_n}$$

where $\varepsilon_n,\ \varepsilon'_n = 2^{-\Omega(n)}$. This is of order $1/3 + o(1)$, as claimed.

Part (2): We are going to show that MS_n has the low 0-density property with respect to the uniform distribution and appropriate parameters. Let r be an arbitrary rectangle defined with respect to a balanced partition $\Pi = (X_1, X_2)$ of the input variables of MS_n. We claim that

$$\left|r^{-1}(1) \cap \mathrm{MS}_n^{-1}(0)\right| \cdot 2^{-n^2} \le \alpha \cdot \left|r^{-1}(1)\right| \cdot 2^{-n^2} + \delta_n,$$

for $\alpha = 1/2$ and exponentially small δ_n defined below.

We first apply Lemma 3.2. W. l. o. g., assume that there are $m := \lfloor \beta n \rfloor$ mixed rows of the input matrix X with respect to Π, for a constant β chosen such that $\beta < 1/\sqrt{2}$. By Lemma 3.6, we get

$$\Big|\left|r^{-1}(1) \cap \mathrm{RT}_n^{-1}(0)\right| - \left|r^{-1}(1) \cap \mathrm{RT}_n^{-1}(1)\right|\Big| \cdot 2^{-n^2} \le 2^{-m} + 3^{-m}.$$

Hence, especially,

$$\left|r^{-1}(1) \cap \mathrm{RT}_n^{-1}(0)\right| \cdot 2^{-n^2} \le (1/2) \cdot \left|r^{-1}(1)\right| \cdot 2^{-n^2} + (1/2) \cdot \left(2^{-m} + 3^{-m}\right),$$

and furthermore, since $\mathrm{MS}_n^{-1}(0) \subseteq \mathrm{RT}_n^{-1}(0)$,

$$\left|r^{-1}(1) \cap \mathrm{MS}_n^{-1}(0)\right| \cdot 2^{-n^2} \le (1/2) \cdot \left|r^{-1}(1)\right| \cdot 2^{-n^2} + (1/2) \cdot \left(2^{-m} + 3^{-m}\right).$$

We have thus shown that MS_n has the low-0-density property with parameters $\alpha = 1/2$ and $\delta_n := (1/2) \cdot \left(2^{-m} + 3^{-m}\right)$, where $m = \lfloor \beta n \rfloor$.

It only remains to apply Theorem 2.2 from the last section. We conclude that

$$\begin{aligned} C^{\mathrm{A}}_{1,\varepsilon}(f) &\ge \delta_n^{-1} \cdot \Big((1/2) \cdot \left|\mathrm{MS}_n^{-1}(0)\right| \cdot 2^{-n^2} - (1/2) \cdot \varepsilon \cdot \left|\mathrm{MS}_n^{-1}(1)\right| \cdot 2^{-n^2}\Big), \quad \text{and} \\ C^{\mathrm{A}}_{\varepsilon}(f) &\ge \delta_n^{-1} \cdot \Big((1/2) \cdot \left|\mathrm{MS}_n^{-1}(0)\right| \cdot 2^{-n^2} - (1/2) \cdot \varepsilon\Big), \end{aligned}$$

for all appropriate ε. We have

$$\left|\mathrm{MS}_n^{-1}(0)\right| \cdot 2^{-n^2} = 1/4 \pm 2^{-\Omega(n)}, \quad \text{and} \quad \left|\mathrm{MS}_n^{-1}(1)\right| \cdot 2^{-n^2} = 3/4 \pm 2^{-\Omega(n)}.$$

Thus, the above lower bounds are still of order $2^{\Omega(n)} = 2^{\Omega(\sqrt{N})}$ if the error bounds are $\varepsilon = 1/3 - \gamma_N$ for one-sided error and $\varepsilon = 1/4 - \gamma'_N$ for two-sided error, where $\gamma_N, \gamma'_N > 0$ are chosen such that $\gamma_N, \gamma'_N = \Omega(1/\mathrm{Poly}(N))$. □

4 Improving the Error Bounds

In this section, we prove Theorem 1.7, the alternative version of the main result with a nonuniform distribution over the input space. Our aim is to adjust the distribution such that we get the best possible error bounds.

In Section 3, we have proven a combinatorial lemma saying that, for each partition of the input matrix of MS_n, there are either many rows or many columns which are "mixed," i. e., contain at least two variables of either side of the partition. Here we extend this lemma as follows.

Suppose that we have many mixed rows with respect to the given partition. We are going to show that we can choose a large subset of mixed rows and two pairs of variables in each of these rows, one pair from each side of the partition, such that no column contains two variables from both sides of the partition. This will ensure that, while the subfunction of RT_n (RowTest) defined on the chosen variables is "difficult," the subfunction of CT_n (ColumnTest) is "easy."

We present the desired combinatorial lemma in the following abstract form.

Lemma 4.1: *Let $X = (x_{ij})_{1 \le i,j \le n}$ be a matrix with entries from $\{0, 1, *\}$.*

For a set $I \subseteq \{1, \ldots, n\}$ of row indices, call a column j split with respect to *I if there are $i_0, i_1 \in I$ such that $x_{i_0,j} = 0$ and $x_{i_1,j} = 1$. Suppose that there is a set $I \subseteq \{1, \ldots, n\}$, $|I| = m$, of row indices such that*

(i) each row i with $i \in I$ contains exactly two 0- and exactly two 1-entries, and the remaining entries in these rows are $*$-entries;

(ii) all rows i with $i \notin I$ contain only $*$-entries.

Then there is a set $I^* \subseteq I$ with $|I^*| \ge m/16$ such that no column is split with respect to I^*.

Proof: We assign colors from $\{0, 1\}$ independently and uniformly at random to the columns of X. Let $\chi(j)$ be the random variable describing the color of column j, where $j = 1, \ldots, n$. Define I_χ as the set of rows which is obtained by starting with the complete set I and removing all rows which have an entry with the "wrong" color $\overline{\chi(j)}$ in column j, for $j = 1, \ldots, n$.

We show that the expected number of remaining rows, $E[I_\chi]$, is still $|I|/16 = m/16$. For $i \in I$, define $S_\chi(i)$ as the random variable which is equal to 1 if row i is contained in I_χ, and equal to 0 otherwise. Since each row $i \in I$ has exactly two entries of color 0 and 1, resp., we have $E[S_\chi(i)] = 1/16$. Hence,

$$E[I_\chi] = \sum_{i \in I} E[S_\chi(i)] = |I|/16 = m/16.$$

This implies that there is a fixed coloring χ^* with $|I_{\chi^*}| = m/16$. It only remains to choose $I^* := I_{\chi^*}$. □

Analogously to Section 3, we use a discrepancy bound to establish the new lower bound on the complexity of rectangle approximations for MS_n. We prove the following extended version of the "discrepancy lemma" (Lemma 3.6) from Section 3.

Lemma 4.2 (Extended Discrepancy Lemma): *Let $\Pi = (X_1, X_2)$ be a partition of the variables in the matrix $X = (x_{ij})_{1 \le i,j \le n}$ with with $\big||X_1| - |X_2|\big| \le 1$. Suppose that m rows of X are mixed with respect to Π.*

Let r be an arbitrary Π-rectangle. Then, for $c \in \{0, 1\}$,

$$\Big|\big|r^{-1}(1) \cap \mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(c)\big| - \big|r^{-1}(1) \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(c)\big|\Big| \cdot 2^{-n^2} \;=\; 2^{-\Omega(m)}.$$

An analogous statement holds for mixed columns instead of rows and exchanged roles of RT_n and CT_n.

Proof: The proof is essentially along the same lines as the proof of Lemma 3.6 in Section 3. Again, we omit subscripts indicating the input size for the ease of notation, and we use f_a for the subfunction of f belonging to an assignment a.

We only prove the statement for mixed rows. Let $\Pi = (X_1, X_2)$ be the given partition with m mixed rows. First, we apply Lemma 4.1. For this, we choose two variables from X_1 and two from X_2 in each of the mixed rows and identify them with 0- and 1-entries, resp. The lemma yields a subset I of the indices of mixed rows with $|I| \ge m/16$ such that for each column j of X, all variables x_{ij} with $i \in I$ are either contained in X_1 or in X_2, but not both. Let $m' := |I|$.

Let $X_1' \subseteq X_1$ and $X_2' \subseteq X_2$ be the sets of the variables chosen in the rows with index in I, and let $\Pi' := \big(X_1', X_2'\big)$. Furthermore, let a be an arbitrary assignment to the variables in $X \backslash \big(X_1' \cup X_2'\big)$. We already know that RT_a has small discrepancy with respect to the partition Π' from Section 3. Substituting m for m' in Lemma 3.5, we obtain:

$$\mathrm{Disc}\big(\mathrm{RT}_a, \Pi'\big) \le 2^{-m'} + 3^{-m'}. \tag{$*$}$$

Now we consider the function CT_a. We have ensured that no column of X is split with respect to the rows with index in I, which means that all variables of a column (which are not set to constants) either belong to X_1' or to X_2', but not both. Hence, it is easy to compute CT_a by a deterministic communication protocol with respect to the partition $\Pi' = \left(X_1', X_2'\right)$: obviously, 1 bit of communication is sufficient.

By Proposition 1.3 from the introduction, we conclude that, for $c \in \{0, 1\}$, there are Π'-rectangles $r_{c,1}$ and $r_{c,2}$ such that $r_{c,1}^{-1}(1) \cap r_{c,2}^{-1}(1) = \emptyset$ and

$$\mathrm{CT}_a^{-1}(c) \ = \ r_{c,1}^{-1}(1) \cup r_{c,2}^{-1}(1).$$

Since r_a and $r_{c,1}, r_{c,2}$ are all Π'-rectangles, $r_a \wedge r_{c,1}$ and $r_a \wedge r_{c,2}$ are also Π'-rectangles. Hence, by $(*)$,

$$\left|\left|\left(r_a^{-1}(1) \cap r_{c,i}^{-1}(1)\right) \cap \mathrm{RT}_a^{-1}(0)\right| - \left|\left(r_a^{-1}(1) \cap r_{c,i}^{-1}(1)\right) \cap \mathrm{RT}_a^{-1}(1)\right|\right| \cdot 2^{-4m'} \leq 2^{-m'} + 3^{-m'},$$

for $i = 1, 2$ and $c \in \{0, 1\}$. Using that $r_{c,1}^{-1}(1) \cap r_{c,2}^{-1}(1) = \emptyset$ and summing for $i = 1, 2$, we get

$$\begin{aligned}\left|\left|r_a^{-1}(1) \cap \mathrm{CT}_a^{-1}(c) \cap \mathrm{RT}_a^{-1}(0)\right| - \left|r_a^{-1}(1) \cap \mathrm{CT}_a^{-1}(c) \cap \mathrm{RT}_a^{-1}(1)\right|\right| \cdot 2^{-4m'} \\ \leq \ 2\left(2^{-m'} + 3^{-m'}\right).\end{aligned}$$

By summing over all assignments a to $X \setminus \left(X_1' \cup X_2'\right)$ (applying the law of total probability), we obtain the desired result:

$$\begin{aligned}\left|\left|r^{-1}(1) \cap \mathrm{CT}^{-1}(c) \cap \mathrm{RT}^{-1}(0)\right| - \left|r^{-1}(1) \cap \mathrm{CT}^{-1}(c) \cap \mathrm{RT}^{-1}(1)\right|\right| \cdot 2^{-n^2} \\ \leq \ 2\left(2^{-m'} + 3^{-m'}\right) = 2^{-\Omega(m)}.\end{aligned}$$

□

Finally, we apply the above lemma to prove the new result on the complexity of rectangle approximations for MS_n. We use the distribution over the input space of MS_n which assigns the measure 0 to the "easy" inputs in the set $\mathrm{RT}_n^{-1}(1) \cup \mathrm{CT}_n^{-1}(1)$. More precisely, let

$$A := \left(\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(0)\right) \cup \left(\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\right) \cup \left(\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(0)\right).$$

Define $\mu\colon \{0, 1\}^{n^2} \to [0, 1]$ for $x \in \{0, 1\}^{n^2}$ by $\mu(x) := |A|^{-1}$ if $x \in A$, and $\mu(x) := 0$ otherwise. For this distribution μ, we get the result announced in the introduction:

Theorem 1.7: *Let $N = n^2$.*
(1) $C_{1,\,1/2+\delta_N}^{\mathrm{A},\,\mu}(\mathrm{MS}_n) = O(1)$ and $C_{1/3+\delta_N'}^{\mathrm{A},\,\mu}(\mathrm{MS}_n) = 1$, where $\delta_N = o(1)$ and $\delta_N' = 2^{-\Omega(\sqrt{N})}$;
(2) $C_{1,\,1/2-\gamma}^{\mathrm{A},\,\mu}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$ and $C_{1/3-\gamma'}^{\mathrm{A},\,\mu}(\mathrm{MS}_n) = 2^{\Omega(\sqrt{N})}$, for all constants $\gamma, \gamma' > 0$.

Proof: We only describe the proof of the lower bounds, the upper bounds are obtained in the same way as for Theorem 1.6. We use the technique from Section 2. It follows from Lemma 3.3 that

$$\mu\bigl(\mathrm{MS}_n^{-1}(0)\bigr) = (1/3)\cdot(1+\varepsilon_n), \quad \text{and} \quad \mu\bigl(\mathrm{MS}_n^{-1}(1)\bigr) = (2/3)\cdot(1+\varepsilon'_n),$$

where $|\varepsilon_n|, |\varepsilon'_n| \to 0$ for $n \to \infty$. In the remainder of the proof, we show that MS_n has the low-0-density property with respect to μ and parameters $\alpha := 1/2$ and δ_n with $\delta_n = 2^{-\Omega(n)}$.

Let r be an arbitrary rectangle defined with respect to a balanced partition of the input variables of MS_n. Suppose that at least $m = \lfloor \beta n \rfloor$ rows of X are mixed with respect to the partition of r, where $\beta < 1/\sqrt{2}$ is a constant. By Lemma 4.2,

$$\Bigl|\bigl|r^{-1}(1)\cap \mathrm{RT}_n^{-1}(0)\cap \mathrm{CT}_n^{-1}(c)\bigr| - \bigl|r^{-1}(1)\cap \mathrm{RT}_n^{-1}(1)\cap \mathrm{CT}_n^{-1}(c)\bigr|\Bigr|\cdot 2^{-n^2} \;\le\; \delta_n,$$

for $c \in \{0,1\}$ and some δ_n with $\delta_n = 2^{-\Omega(m)} = 2^{-\Omega(n)}$. For $c = 0$, this especially implies

$$\bigl|r^{-1}(1)\cap \mathrm{RT}_n^{-1}(0)\cap \mathrm{CT}_n^{-1}(0)\bigr|\cdot 2^{-n^2} \;\le\; \bigl|r^{-1}(1)\cap \mathrm{RT}_n^{-1}(1)\cap \mathrm{CT}_n(0)\bigr|\cdot 2^{-n^2} + \delta_n.$$

Furthermore, since

$$\mathrm{RT}_n^{-1}(0)\cap \mathrm{CT}_n^{-1}(0) = \mathrm{MS}_n^{-1}(0)\cap A, \quad \text{and} \quad \mathrm{RT}_n^{-1}(1)\cap \mathrm{CT}_n^{-1}(0) \subseteq \mathrm{MS}_n^{-1}(1)\cap A,$$

we get

$$|A|\cdot 2^{-n^2}\cdot \mu\bigl(r^{-1}(1)\cap \mathrm{MS}_n^{-1}(0)\bigr) \;\le\; |A|\cdot 2^{-n^2}\cdot \mu\bigl(r^{-1}(1)\cap \mathrm{MS}_n^{-1}(1)\bigr) + \delta_n,$$

or, equivalently,

$$\mu\bigl(r^{-1}(1)\cap \mathrm{MS}_n^{-1}(0)\bigr) \;\le\; (1/2)\cdot \mu\bigl(r^{-1}(1)\bigr) + \rho\cdot\delta_n,$$

where $\rho := 2^{n^2}/(2|A|)$. Again by Lemma 3.3, $\rho = O(1)$. This is the desired low-0-density property for MS_n with respect to μ. It only remains to substitute the above facts into Theorem 2.2. □

5 Applications for Read-Once Branching Programs

In this section, we prove the complexity theoretical results for randomized read-once BPs announced in the introduction. We first give a formal definition of randomized branching programs and present some elementary facts concerning this model. In the second subsection, it is described how lower bounds on the complexity of rectangle approximations can be used to prove lower bounds for randomized read-once branching programs. We apply this technique to get the desired results in the final two subsections.

5.1 Definitions and Basic Facts

Definition 5.1: A *randomized branching program* is a branching program defined on two disjoint sets of variables $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_r\}$ which has the additional property that on each path from the source to a sink, each variable from Y occurs at most once. The variables in Y are called *probabilistic variables*, and nodes labeled by these variables are called *probabilistic nodes*. The other variables and nodes are called *non-probabilistic*.

Let $g\colon \{0,1\}^n \times \{0,1\}^r \to \{0,1\}$ be the function represented by a given randomized branching program G according to the deterministic semantics of branching programs (Definition 1.8). Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function defined on the variables in X. For each assignment $x \in \{0,1\}^n$ to the variables in X, define the *error probability of G on x with respect to f* by

$$\mathrm{err}_{G,f}(x) := \mathrm{Pr}_y\{g(x,y) \neq f(x)\},$$

where the assignments y to the Y-variables are chosen according to the uniform distribution over $\{0,1\}^r$.

We call G a *randomized branching program for f* with

(1) *unbounded two-sided error,* if for all $x \in \{0,1\}^n$, $\mathrm{err}_{G,f}(x) < 1/2$;

(2) *unbounded one-sided error,* if for all $x \in \{0,1\}^n$,

$$\begin{aligned} \mathrm{err}_{G,f}(x) &= 0, \quad \text{if } f(x) = 0; \\ \mathrm{err}_{G,f}(x) &< 1, \quad \text{if } f(x) = 1; \end{aligned}$$

(3) *two-sided error ε,* for constants ε with $0 \le \varepsilon < 1/2$, if for all $x \in \{0,1\}^n$, $\mathrm{err}_{G,f}(x) \le \varepsilon$;

(4) *one-sided error ε,* for constants ε with $0 \le \varepsilon < 1$, if for all $x \in \{0,1\}^n$,

$$\begin{aligned} \mathrm{err}_{G,f}(x) &= 0, \quad \text{if } f(x) = 0; \\ \mathrm{err}_{G,f}(x) &\le \varepsilon, \quad \text{if } f(x) = 1. \end{aligned}$$

We subsume the first two types of error under the label “unbounded error,” while “bounded error” means one of the last two types.

For randomized branching programs with unbounded one-sided error, we use the more common name *nondeterministic branching program* in the following. The definition of nondeterministic branching programs given here coincides with the standard definitions (see, e. g., Meinel [37, 38] and Razborov [44]), which requires that a path consistent with an assignment $x \in \{0,1\}^n$ from the source to the 1-sink (an *accepting path*) exists if and only $x \in f^{-1}(1)$. In the nondeterministic case, the variables in Y are called *nondeterministic variables*, and nodes labeled by these variables are called *nondeterministic nodes*.

It will be convenient to define complexity classes for randomized BPs analogously to the standard classes for Turing machines.

Definition 5.2: Let P-BP, NP-BP, and PP-BP denote the classes of sequences of functions representable by deterministic, nondeterministic, and randomized BPs with unbounded two-sided error, resp. Let RP-BP and BPP-BP denote the classes of sequences of functions representable by randomized BPs with one-sided error bounded by a constant $\varepsilon < 1$ and two-sided error bounded by a constant $\varepsilon < 1/2$, resp.

The following inclusions are obvious from the definitions.

Proposition 5.3: $\quad$ P $\subseteq$ RP $\subseteq$ BPP $\subseteq$ PP, $\quad$ RP $\subseteq$ NP $\subseteq$ PP.

Several simple facts for randomized BPs may be proven essentially in the same way as for probabilistic Turing machines. For example, we can adapt the well-known technique of iterating probabilistic computations to decrease the error probability of randomized BPs:

Proposition 5.4 (Probability amplification):

(1) Let G be a randomized BP representing $f\colon \{0,1\}^n \to \{0,1\}$ with one-sided error ε, $0 \le \varepsilon < 1$. Then there is a randomized BP G' for f with one-sided error ε^m and size $|G'| = O(m|G|)$.

(2) Let G be a randomized BP representing $f\colon \{0,1\}^n \to \{0,1\}$ with two-sided error ε, $0 \le \varepsilon < 1/2$. Let $0 \le \varepsilon' \le \varepsilon$. Then there is a randomized BP G' for f with two-sided error ε' and size $|G'| = O\big(m^2|G|\big)$, where $m = O\big(\log\big((\varepsilon')^{-1}\big)(1/2-\varepsilon)^{-2}\big)$.

Proof: *Part (1):* We use copies $G_1, \dots, G_m$ of G with disjoint sets of probabilistic variables and identify the 1-sink of G_i with the source of G_{i+1}, for $i = 1, \dots, m-1$. The resulting randomized BP obviously has the claimed properties.

Part (2): We start with a deterministic read-once BP G_0 of size $O\big(m^2\big)$ representing the *threshold function* which computes 1 if the number of ones in the input of length m is at least $\lceil m/2 \rceil$, and 0 otherwise (see [54] for the easy construction of such a read-once BP). We replace each node of G_0 by a copy of G, identifying the c-sink of G with the c-successor of the node, for $c \in \{0,1\}$. Each copy uses its own set of probabilistic variables. The resulting randomized BP G' has size $O\big(m^2|G|\big)$. Using Chernoff bounds, we can prove that G' represents f with two-sided error at most $2 \cdot \exp\big(-(1/2-\varepsilon)^2 m\big)$. This is bounded from above by ε' if we choose $m := \big\lceil \ln\big(2/\varepsilon'\big)\,(1/2-\varepsilon)^{-2} \big\rceil$. □

The derandomization technique of Ajtai and Ben-Or [4] for probabilistic circuits is also applicable for randomized BPs. As a consequence, the complexity classes for bounded error (one- and two-sided) turn out to coincide with P-BP.

Proposition 5.5: $\quad$ RP-BP = BPP-BP = P-BP.

We will see that the situation becomes different for restricted BPs.

Proof: By Proposition 5.4, we can decrease the error probability of a given randomized BP for an n-variable function with two-sided error $\varepsilon < 1/2$ to less than 2^{-n} while maintaining polynomial size. As for probabilistic circuits, the resulting randomized BP can be made deterministic by setting the probabilistic variables to constants in an appropriate way. □

Randomized BPs are defined in a such a way that they may be simulated by probabilistic nonuniform Turing machines and vice versa analogously to the well-known result for the deterministic case [18, 42]. Especially, this gives us the following results (where PL/Poly is the class of all sequences of functions computable by a probabilistic nonuniform Turing machine with unbounded two-sided error using at most logarithmic space).

Proposition 5.6: $\text{NP-BP} = \text{NL}/\text{Poly}, \quad \text{PP-BP} = \text{PL}/\text{Poly}.$

This is proven by a straightforward modification of the well-known simulations for the deterministic case. For details, see [37] and [48], resp. (In [48], it is also shown how the class BPL/Poly can be characterized in terms of randomized BPs.)

For the simulations used here, it is crucial that the probabilistic variables of a randomized BP may only be read once. Results of Babai, Nisan, and Szegedy [11] and Nisan [40] lead to the conjecture that, for the scenario of uniform space-bounded computation, the model where random bits may be accessed more than once (without explicitly storing them) is more powerful than the usual read-once model. The following facts indicate that dropping the read-once restriction is also likely to lead to a more powerful model for randomized BPs with unbounded error. Let NP*-BP and PP*-BP be the analogs of the classes NP-BP and PP-BP, resp., for the model where probabilistic variables may be read arbitrarily often.

Proposition 5.7: $\text{NP}^*\text{-BP} = \text{NP}/\text{Poly}, \quad \text{PP}^*\text{-BP} = \text{PP}/\text{Poly}.$

For proofs, see again [37] and [48], resp. This justifies the read-once restriction imposed on the probabilistic variables for randomized BPs.

In the remainder of this subsection, we discuss randomized variants of restricted BPs, which are obtained analogously to Definition 5.1 by requiring that the non-probabilistic variables fulfill the respective restriction. Thus, a *randomized read-once branching program* is a randomized BP where each non-probabilistic variable may appear at most once on each path from the source to a sink.

Complexity classes for randomized read-once BPs are introduced analogously to Definition 5.2, and are denoted by P-BP1, NP-BP1, BPP-BP1, and so on. Additionally, we consider the following classes. For an arbitrary sequence of real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ with $0 \le \varepsilon_n < 1$, let

$$\begin{aligned}\text{RP}_{\varepsilon_n}\text{-BP1} := \{(f_n)_{n\in\mathbb{N}} \mid \exists\, (G_n)_{n\in\mathbb{N}}\colon\ & G_n \text{ is a rand. read-once BP repr. } f_n \text{ with}\\ & \text{one-sided error } \varepsilon_n \text{ and } |G_n| = \text{Poly}(n)\ \}.\end{aligned}$$

Define BPP$_{\varepsilon_n}$-BP1 analogously for sequences $(\varepsilon_n)_{n\in\mathbb{N}}$ with $0 \le \varepsilon_n < 1/2$ and two-sided instead of one-sided error.

As for general BPs, we have some trivial inclusion relations between the basic complexity classes.

Proposition 5.8: $\text{P-BP1} \subseteq \text{BPP-BP1} \subseteq \text{PP-BP1}, \quad \text{RP-BP1} \subseteq \text{NP-BP1} \subseteq \text{PP-BP1}.$

For randomized read-once BPs, it is not as easy as for randomized general BPs to mimic known proofs for probabilistic Turing machines, because it is no longer obvious how computations may be iterated. Especially, the proof of Proposition 5.4 (probability amplification) does not work anymore. In fact, we are going to prove in this section that it *cannot* work: An analog of Proposition 5.4 for read-once BPs instead of unrestricted BPs does not exist.

Hence, it is not even obvious that RP-BP1 $\subseteq$ BPP-BP1. Nevertheless, we can prove this without probability amplification using the idea described in the following.

Lemma 5.9: *Let G be a randomized read-once BP which represents $f : \{0, 1\}^n \to \{0, 1\}$ with one-sided error $\varepsilon < 1$. Let $r \geq 1$. Then there is a randomized read-once BP G' with size at most $O(|G| + r)$ which represents f with two-sided error at most $\varepsilon/(1 + \varepsilon) + 2^{-r}$.*

Proof: It is easy to see that for each $\delta \in \{i \cdot 2^{-r} \mid 0 < i < 2^r\}$, there is a randomized BP $G_{r,\delta}$ which consists only of at most r probabilistic nodes (i. e. no non-probabilistic nodes except for the sinks), and which has the property that the 1-sink is reached with probability δ for a random assignment to the probabilistic variables.

For the construction of G', we identify the 1-sink of such a randomized BP $G_{r,\delta}$ with the 1-sink of G, and the 0-sink with the source of G. The error probability of G' with respect to f is bounded by $\max\{1 - \delta, \delta + (1 - \delta)(1 - \varepsilon)\}$. This is minimized by choosing δ as close as possible to $\delta_{\text{opt}} := \varepsilon/(1+\varepsilon)$. Since δ may be chosen from the set $\{i \cdot 2^{-r} \mid 0 < i < 2^r\}$, we can ensure that $|\delta - \delta_{\text{opt}}| < 2^{-r}$. The resulting randomized BP G' for this value of δ represents f with two-sided error at most $\varepsilon/(1 + \varepsilon) + 2^{-r}$ and size $O(|G| + r)$. □

Corollary 5.10: RP-BP1 $\subseteq$ BPP-BP1.

5.2 Proof Technique for Randomized Read-Once BPs

In this subsection, we describe how lower bounds on the complexity of rectangle approximations may be used to derive lower bounds on the size of randomized read-once BPs.

First, we show that lower bounds on the complexity of rectangle partitions yield lower bounds on the size of deterministic read-once BPs:

Theorem 5.11: *Let G be a deterministic read-once BP representing $f : \{0, 1\}^n \to \{0, 1\}$. Then*

$$C(f) \leq 2n|G|.$$

The proof of this theorem is along the same lines as the proof of the corresponding fact for rectangle *covers* and *nondeterministic* read-once BPs due to Borodin, Razborov, and Smolensky [16]. The proof given here also uses ideas of Okol'nishnikova [41].

Proof: First, we simplify the structure of the given read-once BP G. A read-once BP is called *regular* if for each node v the same set of variables is tested on all paths from the source to v. Furthermore, it is required that on each path from the source to the sinks all variables are tested. It is easy to see that an arbitrary read-once BP G with n variables can be converted into a regular read-once BP G' of size $|G'| \leq 2n|G|$ by inserting dummy tests.

Let G' be a regular read-once BP obtained from the given read-once BP G for f. Let $X = \{x_1, \dots, x_n\}$ be the variable set of G and f. We introduce some more notation.

- For two nodes v, w of G', let $X(v, w) \subseteq X$ be the set of all variables tested on paths from v to w, including the variable at v and excluding the variable at w.
- For an arbitrary assignment $a \in \{0, 1\}^n$ to the variables of G', define $f_{v,w}(a) = 1$ iff there is a path from v to w which is consistent with the assignment a (i. e., for each node u on the path labeled by a variable x_i, the path runs through the a_i-edge starting at u).

Notice that $f_{v,w}$ does not essentially depend on variables from $X \backslash X(v, w)$.

Define a subset C of the nodes of G' as follows. For each path from the source to a sink, include the node reached after testing $\lfloor n/2 \rfloor$ variables. Let s be the source of G', and let t_0 and t_1 be the 0- and the 1-sink, resp. For $v \in C$, define $r_v^0 := f_{s,v} \wedge f_{v,t_0}$ and $r_v^1 := f_{s,v} \wedge f_{v,t_1}$. These functions are combinatorial rectangles due to the definition of C and the regularity of G'. Since each computation path in G' runs through exactly one node of C, we have

$$f^{-1}(0) = \bigcup_{v \in C} \left(r_v^0\right)^{-1}(1), \quad \text{and } r_v^0 \wedge r_w^0 = 0, \text{ for } v \neq w;$$
$$f^{-1}(1) = \bigcup_{v \in C} \left(r_v^1\right)^{-1}(1), \quad \text{and } r_v^1 \wedge r_w^1 = 0, \text{ for } v \neq w.$$

Hence, the rectangles $r_v^0, r_v^1, v \in C$, form a rectangle partition representing f. Their number is bounded by $|C| \leq |G'| \leq 2n|G|$. □

As an easy corollary of the above theorem, we obtain that the complexity of rectangle *approximations* may be used to lower bound the size of deterministic read-once BPs which *approximate* a given function. We give a definition for approximating (unrestricted) BPs below, an extension to the various restricted variants of BPs (especially, to approximating read-once BPs) is obvious.

Definition 5.12: Let $\mu\colon \{0, 1\}^n \to [0, 1]$ be an arbitrary probability distribution. A deterministic BP G is an *approximating BP for* $f\colon \{0, 1\}^n \to \{0, 1\}$ *with two-sided error* ε *with respect to* μ, where $0 \leq \varepsilon < 1/2$, if G represents a function $g\colon \{0, 1\}^n \to \{0, 1\}$ with

$$\mu\left(\{x \mid f(x) \neq g(x)\}\right) \leq \varepsilon;$$

and it is an *approximating BP with one-sided error* ε, $0 \leq \varepsilon < 1$, if

$$\mu\left(\{x \mid f(x) = 0 \wedge g(x) = 1\}\right) \leq \varepsilon \cdot \mu\big(f^{-1}(1)\big), \quad \text{and}$$
$$\mu\left(\{x \mid f(x) = 1 \wedge g(x) = 0\}\right) = 0.$$

Corollary 5.13: *Let* $f\colon \{0, 1\}^n \to \{0, 1\}$ *be an arbitrary function, and let* μ *be an arbitrary probability distribution on* $\{0, 1\}^n$.

(1) Let G be an approximating read-once BP for f with one-sided error ε, $0 \leq \varepsilon < 1$, with respect to μ. Then $C_{1,\varepsilon}^{\mathrm{A},\mu}(f) \leq 2n|G|$.

(2) Let G be an approximating read-once BP for f with two-sided error ε, $0 \leq \varepsilon < 1/2$, with respect to μ. Then $C_{\varepsilon}^{\mathrm{A},\mu}(f) \leq 2n|G|$.

Proof: Follows directly from the definitions and Theorem 5.11. □

Finally, we apply the above insights to randomized read-once BPs. This is done using a simple counting argument originally due to Yao [60].

Lemma 5.14 (Yao's trick): *Let G be a randomized read-once BP representing the function $f\colon \{0,1\}^n \to \{0,1\}$ with two-sided error ε, $0 \le \varepsilon < 1/2$, with respect to a probability distribution μ on $\{0,1\}^n$. Then there is an approximating read-once BP G' for f with two-sided error ε with respect to μ and size at most $|G|$. An analogous statement holds in the case of one-sided error.*

Proof: We only prove the statement for two-sided error, the case of one-sided error can be handled in the same way. Let G be a randomized read-once BP representing f with two-sided error ε. Let $Y = \{y_1, \ldots, y_r\}$ be the set of probabilistic variables of G. Let $g\colon \{0,1\}^n \times \{0,1\}^r \to \{0,1\}$ denote the function computed by G according to the deterministic semantics of BPs.

We know that, for all assignments $a \in \{0,1\}^n$ to the non-probabilistic variables,

$$\sum_{b \in \{0,1\}^r} 2^{-r} \cdot [g(a,b) \ne f(a)] \le \varepsilon.$$

Hence, also

$$\sum_{a \in \{0,1\}^n} \mu(a) \sum_{b \in \{0,1\}^r} 2^{-r} \cdot [g(a,b) \ne f(a)] \le \varepsilon.$$

By changing the order of summation, we get

$$\sum_{b \in \{0,1\}^r} 2^{-r} \sum_{a \in \{0,1\}^n} \mu(a) \cdot [g(a,b) \ne f(a)] \le \varepsilon.$$

It follows that there is at least one assignment $b_0 \in \{0,1\}^r$ to the probabilistic variables with

$$\mu\{a \mid a \in \{0,1\}^n,\ g(a,b_0) \ne f(a)\} \le \varepsilon.$$

Define G' as the read-once BP obtained from G by setting the variables $y_1, \ldots, y_r$ to constants according to b_0. This is done by redirecting all edges leading to a y_i-node to its b_i-successors and deleting the y_i-node afterwards, for $i = 1, \ldots, r$. We obviously obtain an approximating read-once BP for f with two-sided error ε and size at most $|G|$. □

The following theorem summarizes the proof technique for randomized read-once BPs, which we call "rectangle technique" for easier reference. It follows by simply putting together Corollary 5.13 and Lemma 5.14.

Theorem 5.15 (Rectangle Technique for Randomized Read-Once BPs):
Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be an arbitrary function, and let μ be an arbitrary probability distribution on $\{0, 1\}^n$.

(1) Let G be a randomized read-once BP representing f with one-sided error ε with respect to μ, where $0 \leq \varepsilon < 1$. Then $C_{1,\varepsilon}^{\mathrm{A},\mu}(f) \leq 2n|G|$.

(2) Let G be a randomized read-once BP representing f with two-sided error ε, with respect to μ, where $0 \leq \varepsilon < 1/2$. Then $C_{\varepsilon}^{\mathrm{A},\mu}(f) \leq 2n|G|$.

5.3 NP versus BPP for Read-Once Branching Programs

In this subsection, we compare the power of randomized read-once BPs with that of nondeterministic read-once BPs. First, we show that randomized read-once BPs with two-sided error may be exponentially smaller than nondeterministic ones. For this, we consider the following well-known function.

Definition 5.16: The *permutation matrix function* PERM_n is defined on an $n \times n$ matrix $X = (x_{ij})_{1 \leq i,j \leq n}$ of Boolean variables by $\mathrm{PERM}_n(X) = 1$ iff X is a permutation matrix, i. e., if each row and each column contains exactly one entry equal to 1.

Jukna [29] and Krause, Meinel, and Waack [32] have independently shown that nondeterministic read-once BPs for PERM have exponential size. Krause, Meinel, and Waack have also shown that PERM $\in$ coNP-BP1. We complement this by the insight that PERM_n can be represented in polynomial size even by randomized OBDDs, i. e., by randomized read-once BPs where the non-probabilistic variables appear in the same order on each path from the source to a sink.

Theorem 5.17:

(1) For all ε_n with $0 \leq \varepsilon_n < 1$ and $\varepsilon_n = \Omega(1/\mathrm{Poly}(n))$, $\neg\mathrm{PERM}_n$ can be represented by randomized OBDDs with one-sided error ε_n and polynomial size;

(2) each nondeterministic read-once BP representing PERM_n has size $2^{\Omega(n)}$.

Proof: It only remains to show Part (1). The construction uses the well-known *fingerprinting technique* due to Freivalds (see, e. g., the monograph [39] for details on the history). Ablayev and Karpinski [1] have first applied this technique to the construction of randomized OBDDs.

We use the following representation of PERM_n. Let $x_i = (x_{i,1}, \ldots, x_{i,n})$ be the ith row of X. Let $|x|$ be the value of x interpreted as a binary representation. Then $\mathrm{PERM}_n(X) = 1$ if and only if

$$\sum_{i=1}^{n} |x_i| = 2^n - 1 \quad \wedge \quad \text{all } x_i \text{ contain exactly one entry equal to 1.}$$

We apply the fingerprinting technique to check probabilistically whether the binary representation of the sum of the values $|x_i|$ is equal to the vector $(1, \ldots, 1) \in \{0, 1\}^n$.

The randomized OBDD G for PERM_n starts with a tree of probabilistic nodes at the top by which we randomly choose a prime number from the set P_m of the m smallest primes (m is fixed below). For each prime $p \in P_m$, we append a deterministic OBDD BP G_p at the respective leaf of the tree. In G_p, the variables of the input matrix are read in a rowwise order. This allows to simultaneously compute the sum of all $|x_i|$ and to check whether each x_i contains exactly one entry equal to 1. If at least one row with zero or more than one entry equal to 1 is found, or if the sum of the $|x_i|$ modulo p is not equal to $2^n - 1 \bmod p$, the 0-sink is reached. Otherwise, the 1-sink is reached. It easy to see how G_p can be constructed by standard techniques such that $|G_p| = \Theta\left(p \cdot n^2\right)$.

If $\mathrm{PERM}_n(X) = 1$, the 1-sink is reached for all p. The randomized OBDD errs if $\mathrm{PERM}_n(X) = 0$, the matrix X has exactly one entry equal to 1 in each row, and the sum of all $|x_i|$ is equal to $2^n - 1$ modulo the randomly chosen prime p. Since

$$\left| \sum_{i=1}^{n} |x_i| - \left(2^n - 1\right) \right| \le n \cdot 2^{n-1},$$

there are fewer than $n - 1 + \lceil \log n \rceil$ primes for which the sum of the $|x_i|$ is equal to $2^n - 1$ modulo p. Hence, the error probability can be bounded from above by $2n/m$. For $m := \left\lceil \varepsilon_n^{-1} \cdot 2n \right\rceil$, this bound is small enough. By the prime number theorem, $|P_m| = \Theta(m \log m)$. Thus, the overall size of G for the above choice of m is $O\left(n^4 \log n \cdot \varepsilon_n^{-1}\right)$. □

Corollary 5.18:

(1) P-BP1 $\subsetneqq$ RP-BP1;

(2) RP-BP1 $\neq$ coRP-BP1;

(3) BPP-BP1 $\not\subseteq$ NP-BP1 $\cup$ coNP-BP1.

Proof: The first two parts follow immediately from the above facts on PERM_n. For the third part, consider the function $2\mathrm{PERM}_n\colon \{0,1\}^{2n^2} \to \{0,1\}$, defined on two Boolean $n \times n$ matrices X and Y by $2\mathrm{PERM}_n(X, Y) := \mathrm{PERM}_n(X) \wedge \neg\mathrm{PERM}_n(Y)$. By Theorem 5.17, this function is contained in the class BPP-BP1. From the exponential lower bound on the size of nondeterministic read-once BPs for PERM_n it follows that 2PERM is neither contained in NP-BP1 nor in coNP-BP1. □

It is much harder to show that nondeterminism can be more powerful than randomness for read-once BPs. We require a function which is “easy” enough to be computable by nondeterministic read-once BPs of small size, but for which we nevertheless can apply the proof technique from the last subsection. The function MS_n from the main result on rectangle approximations has the desired properties. More precisely, we obtain the following results.

Theorem 5.19: *Let* $N = n^2$ *(the input size of* MS_n*).*

(1) The function MS_n *can be represented in size* $O(N)$ *by*

(a) randomized read-once BPs with one-sided error $1/2$*; and*

(b) randomized read-once BPs with two-sided error $1/3 + \delta_N$*,*
for arbitrary $\delta_N > 0$ *with* $\log(1/\delta_N) = \mathrm{Poly}(N)$.

(2) Let $\gamma, \gamma' > 0$ *be arbitrary constants. Then*

(a) each randomized read-once BPs for MS_n *with one-sided error* $1/2 - \gamma$*, and*

(b) each randomized read-once BPs for MS_n *with two-sided error* $1/3 - \gamma'$

requires size $2^{\Omega(\sqrt{N})}$.

Proof: *Part (1):* We first describe two deterministic sub-BPs G_r and G_c. In G_r, we read the variables of the input matrix rowwise and evaluate RT_n (it easy to see how this can be done in a read-once BP using standard techniques). Likewise, we evaluate CT_n in G_c reading the variables columnwise. A randomized read-once BP for MS_n is now obtained by adding a single probabilistic node which allows to choose randomly between G_r and G_c. This BP has obviously linear size and one-sided error $1/2$. The result for two-sided error follows by Lemma 5.9 (choose $r = \lceil \log(1/\delta_N) \rceil$).

Part (2): Both lower bounds follow by applying the proof technique from Theorem 5.15 to the results from Theorem 1.7. □

We remark that the additional positive term in the error bound for Part (1b) is only required to account for the "rounding error" incurred by representing the constant probability $1/3$ in binary with polynomial length. This term disappears if we allow to assign outcomes of biased coinflips (probabilities $1/3$ and $2/3$) to the probabilistic variables of a randomized BP instead of fair ones as in the standard model.

Theorem 5.19 immediately yields the following results on complexity classes:

Corollary 5.20:

(1) $\mathrm{NP\text{-}BP1} \not\subseteq \mathrm{BPP\text{-}BP1}_{1/3-\gamma}$*, for all constants* $\gamma > 0$*;*

(2) $\mathrm{RP\text{-}BP1}_{1/2-\gamma} \subsetneqq \mathrm{NP\text{-}BP1}$*, for all constants* $\gamma > 0$*;*

(3) $\mathrm{RP\text{-}BP1}_{1/2-\gamma} \subsetneqq \mathrm{RP\text{-}BP1}_{1/2}$ *and* $\mathrm{BPP\text{-}BP1}_{1/3-\gamma'} \subsetneqq \mathrm{BPP\text{-}BP1}_{1/3+\delta_n}$*,*
for all $\delta_n > 0$ *with* $\log(1/\delta_n) = \mathrm{Poly}(n)$ *and all constants* $\gamma, \gamma' > 0$.

Part (3) of this theorem shows that there is no "probability amplification" technique for randomized read-once BPs similar to Proposition 5.4 for general BPs. Decreasing the error probability by an arbitrarily small constant may lead to an exponential blowup of the size for randomized read-once BPs.

5.4 Unrestricted versus Unambiguous Nondeterminism for Read-Once Branching Programs

We now deal with the power of nondeterminism for read-once BPs. We consider the following restricted nondeterministic model.

Definition 5.21: A nondeterministic read-once BP is called *unambiguous read-once BP* if for each input there is at most one accepting computation path. Let UP-BP1 denote the class of sequences of functions with unambiguous read-once BPs of polynomial size.

We are going to prove that multiple accepting paths for the same input have to be allowed to exploit the full power of nondeterministic read-once BPs. We already know that the function MS_n can be represented in linear size by nondeterministic read-once BPs according to Theorem 5.19 (Part (1a)). On the other hand, every unambiguous read-once BP for this function requires exponential size:

Theorem 5.22: *Each unambiguous read-once BP for* MS_n *has size* $2^{\Omega(n)}$.

Corollary 5.23: UP-BP1 $\subsetneq$ NP-BP1.

In order to prove Theorem 5.22, we use the following variant of Theorem 5.11 from Subsection 5.2.

Theorem 5.24: *Let G be an unambiguous read-once BP for the function $f\colon \{0,1\}^n \to \{0,1\}$ defined on the variable set X, $|X| = n$. Then there are combinatorial rectangles $r_1, \dots, r_k$ (each with its own partition of the input variables) such that*

(i) $k \le 2n|G|$;

(ii) $r_1^{-1}(1) \cup \cdots \cup r_k^{-1}(1) = f^{-1}(1)$ *and* $r_i^{-1}(1) \cap r_j^{-1}(1)$ *for* $i \neq j$.

Proof: This is very much similar to the proof of Theorem 5.11, as well as to the proof of Borodin, Razborov, and Smolensky establishing an analogous fact for nondeterministic read-once BPs and covers of the 1-inputs by rectangles.

We use the notation from the proof of Theorem 5.11, and we assume that the given unambiguous read-once BP G is regular with respect to variables in X (which increases the size by a factor of at most $2n$). For each 1-input of f, there is exactly one accepting computation path from the source s of G to the 1-sink t_1. Hence,

$$f^{-1}(1) = \bigcup_{v \in C} \left(r_v^1\right)^{-1}(1), \quad \text{and} \quad r_v^1 \wedge r_w^1 = 0, \quad \text{for } v \neq w,$$

where $r_v^1 = f_{s,v} \wedge f_{v,t_1}$, for $v \in C$, are the combinatorial rectangles from the proof of Theorem 5.11. Obviously, these rectangles have the required properties. □

Proof of Theorem 5.22: Let G be an unambiguous read-once BP representing MS_n. By Theorem 5.24, there is a partition of $\mathrm{MS}_n^{-1}(1)$ into rectangles $r_1, \ldots, r_k$, where $k \le 2n|G|$ and the rectangles are defined with respect to balanced partitions of the input matrix X of MS_n. Let Π_i be the partition of the inputs used by rectangle r_i, for $i = 1, \ldots, k$.

We start with a sketch of the essence of the proof. First, observe that

$$\mathrm{MS}_n^{-1}(1) = \big(\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\big) \cup \big(\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(0)\big) \cup \big(\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)\big).$$

Consider an arbitrary rectangle r_i from the given partition of $\mathrm{MS}_n^{-1}(1)$. We claim that the following happens: Either $r_i^{-1}(1)$ is exponentially small, or around half of the inputs in $r_i^{-1}(1)$ are from the set $\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)$ (or both). Then all rectangles which are not exponentially small and which are used to cover the sets $\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)$ or $\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)$ "overlap" into the set $\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)$. As a consequence, the set $\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)$ is either covered twice by rectangles (which is not allowed), or the used rectangles have to be exponentially small.

We now prove this in detail. Fix a constant $\beta < 1/\sqrt{2}$. By Lemma 3.2, either at least $m := \lfloor \beta n \rfloor$ rows or columns of X are mixed for each partition Π_i (or both). Define

$$I := \{i \mid \text{there are at least } m \text{ mixed rows with respect to } \Pi_i\},$$

and $J := \{1, \ldots, k\} \backslash I$. Notice that for each $i \in J$, there are at least m mixed columns with respect to Π_i by Lemma 3.2. To simplify notation, we define $R_i := r_i^{-1}(1)$ for $i = 1, \ldots, k$.

Let μ be the uniform distribution on the assignments to the input variables of MS_n. By the "extended discrepancy lemma," Lemma 4.2, we have

$$\sum_{i \in I} \mu\big(R_i \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)\big) \ge \sum_{i \in I} \mu\big(R_i \cap \mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\big) - |I| \cdot \delta_n, \qquad (1)$$

$$\sum_{j \in J} \mu\big(R_j \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)\big) \ge \sum_{j \in J} \mu\big(R_j \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(0)\big) - |J| \cdot \delta_n, \qquad (2)$$

where $\delta_n = 2^{-\Omega(m)} = 2^{-\Omega(n)}$.

Furthermore, since $R_i \cap \mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(0) = \emptyset$ for all $i = 1, \ldots, k$,

$$\sum_{i \in I} \mu\big(R_i \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(0)\big) \le |I| \cdot \delta_n, \quad \text{and} \qquad (3)$$

$$\sum_{j \in J} \mu\big(R_j \cap \mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\big) \le |J| \cdot \delta_n. \qquad (4)$$

The sets R_i form a partition of the 1-inputs of MS, thus we can combine $(1) + (4)$ and $(2) + (3)$ to obtain

$$\sum_{i \in I} \mu\big(R_i \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)\big) \ge \mu\big(\mathrm{RT}_n^{-1}(0) \cap \mathrm{CT}_n^{-1}(1)\big) - (|I| + |J|) \cdot \delta_n, \qquad (5)$$

$$\sum_{j \in J} \mu\big(R_j \cap \mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(1)\big) \ge \mu\big(\mathrm{RT}_n^{-1}(1) \cap \mathrm{CT}_n^{-1}(0)\big) - (|I| + |J|) \cdot \delta_n. \qquad (6)$$

Finally, adding (5) and (6) yields

$$\mu\big(\mathrm{RT}_n^{-1}(1)\cap\mathrm{CT}_n^{-1}(1)\big)\geq \\ \mu\big(\mathrm{RT}_n^{-1}(1)\cap\mathrm{CT}_n^{-1}(0)\big)+\mu\big(\mathrm{RT}_n^{-1}(0)\cap\mathrm{CT}_n^{-1}(1)\big)-2(|I|+|J|)\cdot\delta_n.$$

By Lemma 3.3,

$$\mu\big(\mathrm{RT}_n^{-1}(1)\cap\mathrm{CT}_n^{-1}(0)\big)+\mu\big(\mathrm{RT}_n^{-1}(0)\cap\mathrm{CT}_n^{-1}(1)\big)-\mu\big(\mathrm{RT}_n^{-1}(1)\cap\mathrm{CT}_n^{-1}(1)\big)\geq 1/4-\varepsilon_n,$$

where $\varepsilon_n = 2^{-\Omega(n)}$. Hence,

$$k=|I|+|J|\geq\frac{1}{2}\cdot\delta_n^{-1}\,(1/4-\varepsilon_n)=2^{\Omega(n)},$$

which yields the desired bound on the size of G, since $|G|\geq k/(2n)$. □

6 Algebraic Tools

Counting the number of solutions of equations is a fundamental combinatorial problem. It is also one of the foundations for the proofs of the results in this paper. In this section, we present algebraic tools which allow to count the numbers of solutions of equations over finite fields in many cases. The presentation heavily relies upon Babai's lecture notes [8]. For the proofs not given here and further background information we refer to his paper, or to standard textbooks like [35].

6.1 A Brief Introduction to Characters over Finite Abelian Groups

For the following, let $(G,+)$ be a finite, abelian group. Let $|G|$ denote the order of G.

A *character of* G is a homomorphism from G to the complex unit circle, i. e., a homomorphism $\chi : G\to\mathbb{C}$ where $|\chi(a)|=1$ for all $a\in G$. We use $\widehat{G}$ to denote the set of all characters of G. The special character χ with $\chi(a)=1$ for all $a\in G$ is called the *trivial character of* G and is denoted by χ_0.

From the definition, we can immediately conclude that, for all characters χ, $\chi(0)=1$ (since $\chi(0)=\chi(0+0)=\chi(0)\cdot\chi(0)$). Furthermore, for all $a\in G$, $\chi(-a)=\chi(a)^{-1}=\overline{\chi(a)}$ (where the bar denotes complex conjugation). This is because $\chi(a)\cdot\chi(-a)=\chi(a-a)=\chi(0)=1$. Finally, it also follows that $(\chi(a))^{|G|}=\chi\,(|G|\cdot a)=\chi(0)=1$ for all $a\in G$, i. e., the values of χ are $|G|$th roots of unity in $\mathbb{C}$. (By $n\cdot a=a\cdot n$, where $n\in\mathbb{Z}$, we mean the nth power of a in additive notation, i. e., $a+\cdots+a$, n times.)

For $\chi,\psi\in\widehat{G}$, the *product character* $\chi\cdot\psi\in\widehat{G}$ is defined by

$$(\chi\cdot\psi)(g):=\chi(g)\cdot\psi(g),\quad \text{for all } g\in G.$$

It is easy to verify that $\widehat{G}$ becomes a group under the multiplication of characters defined in this way, which is called the *character group of G*

We collect some important facts on the structure of character groups.

Proposition 6.1: *Let $G = H_1 \times H_2$, i. e., G is the direct product of the groups H_1 and H_2, and let $\chi \in \widehat{H_1}$, $\psi \in \widehat{H_2}$. Then $\varphi := \chi \times \psi$ defined by*

$$\varphi(g, h) := \chi(g) \cdot \psi(h), \quad \textit{for all } g \in H_1, h \in H_2,$$

is a character of G. Moreover, all characters of G are of this form, and

$$\widehat{G} \cong \widehat{H_1} \times \widehat{H_2}.$$

This can also be verified in an elementary way. Together with the structure theorem for finite abelian groups (see, e. g., [34]), the above proposition implies:

Theorem 6.2: $\quad G \cong \widehat{G}$.

Especially, we have $\left|\widehat{G}\right| = |G|$. An important tool for handling characters is presented in the following.

Theorem 6.3 (Orthogonality relations):

(1) For $\chi, \psi \in \widehat{G}$,

$$\frac{1}{|G|} \sum_{g \in G} \chi(g)\overline{\psi(g)} = \begin{cases} 1, & \textit{if } \chi = \psi; \\ 0, & \textit{otherwise.} \end{cases}$$

(2) For $g, h \in G$,

$$\frac{1}{|G|} \sum_{\chi \in \widehat{G}} \chi(g)\overline{\chi(h)} = \begin{cases} 1, & \textit{if } g = h; \\ 0, & \textit{otherwise.} \end{cases}$$

Characters may be used to generalize the well-known discrete Fourier transform. As shown in the next subsection, such generalized Fourier transforms are a useful tool for computing the number of solutions of equations.

Definition 6.4: Let $f : G \to \mathbb{C}$ be an arbitrary function. The *Fourier transform* of f is the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by

$$\widehat{f}(\chi) := \sum_{a \in G} \chi(a)\, f(a), \quad \text{for all } \chi \in \widehat{G}.$$

Proposition 6.5: *The mapping $\mathcal{F}$ of functions to their Fourier transform has an inverse $\mathcal{F}^{-1}$. This maps a function $F : \widehat{G} \to \mathbb{C}$ to the function $f : G \to \mathbb{C}$ given by*

$$f(a) = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \overline{\chi(a)} F(\chi), \quad \textit{for } a \in G.$$

Proof: We only have to check that the above formula, applied to the Fourier transform $\widehat{f}$ of a function f, indeed recovers the original function. For $a \in G$,

$$\sum_{\chi \in \widehat{G}} \overline{\chi(a)}\,\widehat{f}(\chi) = \sum_{\chi \in \widehat{G}} \overline{\chi(a)} \sum_{b \in G} \chi(b)\, f(b) = \sum_{b \in G} f(b) \sum_{\chi \in \widehat{G}} \chi(b-a).$$

By Theorem 6.3, the last sum is equal to $|G|$, if $b = a$, and equal to 0, if $b \neq a$. Hence,

$$\sum_{\chi \in \widehat{G}} \overline{\chi(a)}\,\widehat{f}(\chi) = |G| \cdot f(a),$$

as desired. □

At the end of this subsection, we consider some concrete examples of character groups which will pop up in the following.

Proposition 6.6:

(1) *Let $G = \mathbb{Z}_p \times \mathbb{Z}_q$, where p and q are different primes. For each $u \in \mathbb{Z}_p$ and $v \in \mathbb{Z}_q$, define the function $\chi_{u,v} \colon G \to \mathbb{C}$ by*

$$\chi_{u,v}(w) := e^{2\pi i u w/p} \cdot e^{2\pi i v w/q}, \quad \textit{for all } w \in G,$$

where G, $\mathbb{Z}_p$ and $\mathbb{Z}_q$ are represented by subsets of $\mathbb{Z}$ for the computation of the exponents (notice that $\mathbb{Z}_p \times \mathbb{Z}_q \cong \mathbb{Z}_{pq}$). Then $\chi_{u,v}$ is a character of G, and all characters of G are obtained in this way.

(2) *Let $G = (\mathbb{Z}_p^m, +)$, where p is a prime, $m \geq 1$ an integer, and "$+$" the addition of vectors in $\mathbb{Z}_p^m$. Define the standard inner product in $\mathbb{Z}_p^m$ by $\langle u, v\rangle := \sum_{i=1}^m u_i v_i$ for vectors $u, v \in \mathbb{Z}_p^m$. For each $u \in \mathbb{Z}_p^m$, the function $\chi_u \colon \mathbb{Z}_p^m \to \mathbb{C}$ defined by*

$$\chi_u(v) := e^{2\pi i \langle u,v\rangle/p}, \quad \textit{for } v \in \mathbb{Z}_p^m,$$

is a character of $\mathbb{Z}_p^m$, and all characters are obtained in this way.

Proof: Since we always have $\left|\widehat{G}\right| = |G|$, it only remains to verify that, in both cases, the given functions are different characters of their respective groups. This is done by elementary calculations. □

6.2 On the Number of Solutions of Equations over Finite Abelian Groups

Let $(G, +)$ be an arbitrary finite abelian group. Let $A_1, \ldots, A_n \subseteq G$ and $b \in G$. We consider the equation

$$x_1 + \cdots + x_n = b, \quad \text{where } x_1 \in A_1, \ldots, x_n \in A_n. \tag{EQ}$$

We are interested in the number of solutions of (EQ) and define

$$S(A_1, \ldots, A_n; b) := \left|\{x \mid x = (x_1, \ldots, x_n) \in A_1 \times \cdots \times A_n \wedge x_1 + \cdots + x_n = b\}\right|.$$

These numbers can be computed by the following formula:

Theorem 6.7: *For arbitrary $A_1, \dots, A_n \subseteq G$ and $b \in G$,*

$$S(A_1, \dots, A_n; b) \;=\; \frac{2^n}{|G|} + \frac{1}{|G|} \cdot \sum_{\chi \in \widehat{G}, \chi \neq \chi_0} \overline{\chi(b)} \prod_{k=1}^{n} \sum_{a \in A_k} \chi(a).$$

Proof: We apply generating functions in the same way as this is usually done in combinatorics. Instead of generating functions of the type $\sum_{n \geq 1} c_n z^n$, we use generalized Fourier series. Define the function $F\colon \widehat{G} \to \mathbb{C}$ by

$$F(\chi) := \prod_{k=1}^{n} \sum_{a \in A_k} \chi(a)$$

for all $\chi \in \widehat{G}$. The idea is that the function F captures all information on the different possible ways to sum elements from $A_1, \dots, A_n$. A term $\sum_{a \in A_k} \chi(a)$ represents the $|A_k|$ possible ways to choose the kth summand in (EQ) from the set A_k: the factor $\chi(a)$ in the overall product belongs to the decision to include a in the sum.

We have

$$F(\chi) \;=\; \prod_{k=1}^{n} \sum_{a \in A_k} \chi(a) \;=\; \sum_{a_1 \in A_1, \dots, a_n \in A_n} \chi(a_1 + \cdots + a_n) \;=\; \sum_{b \in G} \chi(b)\, S(A_1, \dots, A_n; b).$$

Thus, F is the Fourier transform of the function $f\colon G \to \mathbb{C}$ defined by $f(b) := S(A_1, \dots, A_n; b)$. The formula for the inverse Fourier transform yields

$$f(b) \;=\; \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \overline{\chi(b)}\, F(\chi) \;=\; \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \overline{\chi(b)} \prod_{k=1}^{n} \sum_{a \in A} \chi(a).$$

Pulling out the term for $\chi = \chi_0$ gives the claimed formula. □

At the end of this section, we discuss an important special version of the above formula which will be used later on.

Let A be an arbitrary $m \times n$ matrix over $\mathbb{Z}_p$, where p is a prime, and let $b \in \mathbb{Z}_p^m$ be an arbitrary vector. It is easy to count the number of solutions $x \in \mathbb{Z}_p^n$ of the linear system of equations $A \cdot x \equiv b \bmod p$: there are exactly $p^{n - \mathrm{rank}(A)}$ such solutions, where $\mathrm{rank}(A)$ is the rank of the matrix A over $\mathbb{Z}_p$.

This simple formula no longer holds if we restrict ourselves to *Boolean* solutions. Using the tools developed in this section, we can nevertheless come up with an exact formula also for this case.

Theorem 6.8: *Let p be a prime. Let A be an arbitrary $m \times n$ matrix over $\mathbb{Z}_p$ with column vectors $a_1, \ldots, a_n \in \mathbb{Z}_p^m$, and let $b \in \mathbb{Z}_p^m$. Furthermore, define $\omega := e^{2\pi i/p}$, and let $\langle \cdot, \cdot \rangle$ denote the standard inner product in $\mathbb{Z}_p^m$, i. e., $\langle u, v \rangle = \sum_{k=1}^m u_k v_k \bmod p$ for $u, v \in \mathbb{Z}_p^m$. Then the number of $x \in \{0, 1\}^n$ with $A \cdot x \equiv b \bmod p$ is exactly*

$$p^{-m} \cdot 2^n + p^{-m} \sum_{\substack{v \in \mathbb{Z}_p^m \\ v \neq 0}} \omega^{-\langle b, v \rangle} \prod_{k=1}^{n} \left(1 + \omega^{\langle a_k, v \rangle}\right).$$

Variants of this formula appear in different disguises in the literature (see, e. g., [21] for the restricted case $m = 1$ and [43] for a closely related result for $p = 2$).

Proof: By application of Theorem 6.7. We choose $G = (\mathbb{Z}_p^m, +)$, and $A_k := \{0, a_k\}$, for $k = 1, \ldots, n$. The characters of $\mathbb{Z}_p^m$ have already been described in Proposition 6.6. □

7 Proof of Lemma 3.3

It is easy to prove that approximately one half of all Boolean $n \times n$ matrices X satisfy one of the equations $\mathrm{RT}_n(X) = \xi$ or $\mathrm{CT}_n(X) = \eta$, resp., for arbitrary $\xi, \eta \in \{0, 1\}$. Lemma 3.3 states that, in spite of the fact that both of these equations are defined on the same set of variables, they behave as if they were independent of each other: approximately $1/4$ of all Boolean matrices X satisfy $\mathrm{RT}_n(X) = \xi$ *and* $\mathrm{CT}_n(X) = \eta$. In this section, we prove this astonishing fact.

Proof of Lemma 3.3: Let $X = (x_{ij})_{1 \le i, j \le n}$. We have $\mathrm{RT}_n(X) = \xi$ and $\mathrm{CT}_n(X) = \eta$ if and only if there are $r_1, \ldots, r_n \in \mathbb{Z}_3$ and $c_1, \ldots, c_n \in \mathbb{Z}_3$ such that

$$\begin{array}{rclrcl} x_{1,1} + x_{1,2} + \cdots + x_{1,n} & \equiv & r_1 & x_{1,1} + x_{2,1} + \cdots + x_{n,1} & \equiv & c_1 \\ x_{2,1} + x_{2,2} + \cdots + x_{2,n} & \equiv & r_2 & x_{1,2} + x_{2,2} + \cdots + x_{n,2} & \equiv & c_2 \\ & \vdots & & & \vdots & \\ x_{n,1} + x_{n,2} + \cdots + x_{n,n} & \equiv & r_n & x_{1,n} + x_{2,n} + \cdots + x_{n,n} & \equiv & c_n \end{array} \tag{1}$$

and, additionally,

$$\sum_{i=1}^{n} [r_i \equiv 0 \bmod 3] \equiv \xi \bmod 2 \quad \wedge \quad \sum_{i=1}^{n} [c_i \equiv 0 \bmod 3] \equiv \eta \bmod 2. \tag{2}$$

Let $x_i := (x_{i,1}, \ldots, x_{i,n})$, for $i = 1, \ldots, n$, and $x := (x_1, \ldots, x_n)^\top$. Furthermore, let $b := (r_1, \ldots, r_n, c_1, \ldots, c_n)^\top \in \mathbb{Z}_3^{2n}$. Then (1) is equivalent to

$$A \cdot x \equiv b \bmod 3, \tag{3}$$

where A is the $2n \times n^2$ matrix defined as follows (empty spaces indicate 0-entries):

$$A := \begin{array}{c}\\ \left[\begin{array}{c|c|c|c} 1\ \ 1\ \ \cdots\ \ 1 & & & \\ & 1\ \ 1\ \ \cdots\ \ 1 & & \\ & & \ddots & \\ & & & 1\ \ 1\ \ \cdots\ \ 1 \\ \hline \begin{matrix}1&&&\\&1&&\\&&\ddots&\\&&&1\end{matrix} & \begin{matrix}1&&&\\&1&&\\&&\ddots&\\&&&1\end{matrix} & \cdots & \begin{matrix}1&&&\\&1&&\\&&\ddots&\\&&&1\end{matrix} \end{array}\right]\end{array} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\1\\ \ddots\\1\end{matrix}}\right\} n \\ \left.\vphantom{\begin{matrix}1\\1\\ \ddots\\1\end{matrix}}\right\} n \end{array}$$

$$\text{(each of the column blocks: } \overbrace{\qquad\qquad}^{n}\ \overbrace{\qquad\qquad}^{n}\ \cdots\ \overbrace{\qquad\qquad}^{n}\text{)}$$

Our first aim is to estimate the number of Boolean solutions of the linear system (3) for fixed r_i and c_i. Later on, we deal with the number of possible choices for the r_i and c_i.

Claim: *The number of solutions $x \in \{0,1\}^{n^2}$ of system (3) is $3^{-(2n-1)} \cdot 2^{n^2} \cdot \left(1 \pm 2^{-\Omega(n)}\right)$, if*

$$r_1 + \cdots + r_n \equiv c_1 + \cdots + c_n \bmod 3, \qquad (*)$$

and 0 *otherwise.*

Proof of the Claim: By elementary transformations of the system (3), it follows that $(*)$ is necessary for the existence of solutions.

Now suppose that $(*)$ is fulfilled. Define $m := 2n$. Let $a_0, \ldots, a_{n^2-1} \in \mathbb{Z}_3^m$ denote the column vectors of the coefficient matrix A. Let $e_0, \ldots, e_{m-1}$ be the standard basis of $\mathbb{Z}_3^m$, i. e., e_k is the vector which has a 1 at position k and zeros everywhere else. Then we have $a_k = e_{\lfloor k/n \rfloor} + e_{n+k \bmod n}$, for $k = 0, \ldots, n^2 - 1$. Let N be the number of Boolean solutions of (3). By Theorem 6.8,

$$N = 3^{-m} \cdot 2^{n^2} + 3^{-m} \sum_{\substack{v \in \mathbb{Z}_3^m \\ v \neq 0}} \omega^{-\langle b,\, v \rangle} \prod_{k=0}^{n^2-1} \left(1 + \omega^{\langle a_k,\, v \rangle}\right), \quad \text{where } \omega = e^{2\pi i/3}. \qquad (\#)$$

Let us first compute the value of the sum which is obtained by substituting the special vector

$$u = (u_0, \ldots, u_{m-1}) = (\underbrace{1, \ \ldots\ , 1}_{n}, \underbrace{-1, \ldots, -1}_{n})$$

for v.

Since $u_{\lfloor k/n \rfloor} = 1$ and $u_{n+k \bmod n} = -1$ for all k, we get

$$\prod_{k=0}^{n^2-1} \left(1 + \omega^{\langle a_k,\, u \rangle}\right) = \prod_{k=0}^{n^2-1} \left(1 + \omega^{u_{\lfloor k/n \rfloor} + u_{n+k \bmod n}}\right) = 2^{n^2}.$$

Furthermore, we have $\omega^{-\langle b,\, u\rangle} = 1$ because of $(*)$. The same arguments work for $-u$ instead of u. Hence, we can rewrite (#) as

$$N = 3 \cdot 3^{-m} \cdot 2^{n^2} + 3^{-m} \sum_{v \notin \{0,u,-u\}} \omega^{-\langle b,\, v\rangle} \prod_{k=0}^{n^2-1} \left(1 + \omega^{\langle a_k,\, v\rangle}\right),$$

where the summation is over all vectors $v \in \mathbb{Z}_3^m$ not contained in the set $\{0, u, -u\}$. By subtracting $3^{-(m-1)} \cdot 2^{n^2}$ on both sides and taking absolute values, we obtain

$$\begin{aligned} \varepsilon_n := \left| N - 3^{-(m-1)} \cdot 2^{n^2} \right| \;&\leq\; 3^{-m} \sum_{v \notin \{0,u,-u\}} \prod_{k=0}^{n^2-1} \left| 1 + \omega^{\langle a_k,\, v\rangle} \right| \\ &=\; 3^{-m} \sum_{\substack{v \notin \{0,u,-u\} \\ v=(v_0,\ldots,v_{m-1})}} \prod_{k=0}^{n^2-1} \left| 1 + \omega^{v_{\lfloor k/n\rfloor} + v_{n+k \bmod n}} \right|. \end{aligned}$$

We want to show that

$$\varepsilon_n = 3^{-(m-1)} \cdot 2^{n^2} \cdot 2^{-\Omega(n)}, \quad \text{or, equivalently,} \quad 3^m \cdot \varepsilon_n = 3 \cdot 2^{n^2 - \Omega(n)} = 2^{n^2 - \Omega(n)}.$$

To do this, we have to get a relatively precise bound on the above sum of product terms. One may verify easily that, for arbitrary integers ℓ,

$$\left| 1 + \omega^{\ell} \right| = \sqrt{2}(1 + \cos(2\pi\ell/3))^{1/2} = \begin{cases} 2, & \text{if } \ell \equiv 0 \bmod 3; \\ 1, & \text{if } \ell \equiv 1, 2 \bmod 3. \end{cases}$$

Thus, we require that sufficiently "few" of the factors in the above products are equal to 2. (Observe that, if only for one vector $v \notin \{0, u, -u\}$ all factors would be equal to 2, then we would already have "lost.")

For $v \in \mathbb{Z}_3^m$, define

$$Z(v) := \left| \left\{ j \;\middle|\; 0 \leq j \leq n^2 - 1 \,\wedge\, v_{\lfloor j/n\rfloor} + v_{n+j \bmod n} \equiv 0 \bmod 3 \right\} \right|.$$

With this notation, we have established above that $3^m \cdot \varepsilon_n \leq \sum_{v \notin \{0,u,-u\}} 2^{Z(v)}$. For the following, fix a vector $v \in \mathbb{Z}_3^m$. Identify $\mathbb{Z}_3$ with $\{0, 1, 2\}$. Our goal is to characterize $Z(v)$ in terms of the numbers of 0-, 1-, and 2-entries in $v = (v_0, \ldots, v_{m-1})$. For $c \in \{0, 1, 2\}$, define

$$\begin{aligned} \alpha_c(v) &:= \left| \{ j \mid 0 \leq j \leq n-1 \,\wedge\, v_j \equiv c \bmod 3 \} \right|, \quad \text{and} \\ \beta_c(v) &:= \left| \{ j \mid n \leq j \leq 2n-1 \,\wedge\, v_j \equiv c \bmod 3 \} \right|. \end{aligned}$$

We have $\alpha_0(v) + \alpha_1(v) + \alpha_2(v) = n$ and $\beta_0(v) + \beta_1(v) + \beta_2(v) = n$. In order to compute $Z(v)$, we have to count the numbers $j = 0, \ldots, n^2 - 1$ such that $v_k + v_\ell \equiv 0 \bmod 3$, where $k = \lfloor j/n \rfloor$ and $\ell = n + j \bmod n$. The same number is obtained if we consider all possible values for ℓ, i. e., $\ell = n, \ldots, 2n-1$, and count all $k \in \{0, \ldots, n-1\}$ such that $v_k + v_\ell \equiv 0 \bmod 3$. Thus,

$$Z(v) = \sum_{\ell=n}^{2n-1} Z_\ell(v), \;\text{ where } Z_\ell(v) := \left| \{ k \mid 0 \leq k \leq n-1 \,\wedge\, v_k + v_\ell \equiv 0 \bmod 3 \} \right|.$$

Since $v_k + v_\ell \equiv 0 \Leftrightarrow v_k \equiv -v_\ell$, we have $Z_\ell(v) = \alpha_{-v_\ell}(v)$ (computing the subscript of α in $\mathbb{Z}_3$). Thus,

$$Z(v) = \sum_{\ell=n}^{2n-1} Z_\ell(v) = \sum_{c \in \{0,1,2\}} \alpha_{-c}(v)\beta_c(v) = \alpha_0(v)\beta_0(v) + \alpha_1(v)\beta_2(v) + \alpha_2(v)\beta_1(v).$$

Putting the above insights together, we obtain:

$$\begin{aligned}
3^m \cdot \varepsilon_n &\leq \sum_{v \notin \{0,u,-u\}} 2^{Z(v)} \\
&= \sum_{v \notin \{0,u,-u\}} 2^{\alpha_0(v)\beta_0(v) + \alpha_1(v)\beta_2(v) + \alpha_2(v)\beta_1(v)} \\
&= \sum_{\substack{0 \leq \alpha_0,\alpha_1,\alpha_2,\, \beta_0,\beta_1,\beta_2 \leq n \\ \alpha_0+\alpha_1+\alpha_2 = \beta_0+\beta_1+\beta_2 = n \\ \alpha_0\beta_0,\, \alpha_1\beta_2,\, \alpha_2\beta_1 \neq n^2}} \binom{n}{\alpha_0, \alpha_1, \alpha_2}\binom{n}{\beta_0, \beta_1, \beta_2} 2^{\alpha_0\beta_0 + \alpha_1\beta_2 + \alpha_2\beta_1}.
\end{aligned}$$

Observe that the terms for indices with $\alpha_0\beta_0 = n^2$, $\alpha_1\beta_2 = n^2$ or $\alpha_2\beta_1 = n^2$ which are excluded in the above sum are all equal to 2^{n^2}. Thus,

$$3^m \cdot \varepsilon_n + 3 \cdot 2^{n^2} \leq \sum_{\substack{0 \leq \alpha_0,\alpha_1,\alpha_2,\, \beta_0,\beta_1,\beta_2 \leq n \\ \alpha_0+\alpha_1+\alpha_2,\, \beta_0+\beta_1+\beta_2 = n}} \binom{n}{\alpha_0, \alpha_1, \alpha_2}\binom{n}{\beta_0, \beta_1, \beta_2} 2^{\alpha_0\beta_0 + \alpha_1\beta_2 + \alpha_2\beta_1} =: R.$$

By the multinomial theorem,

$$\begin{aligned}
R &= \sum_{\substack{0 \leq \alpha_0,\alpha_1,\alpha_2 \leq n \\ \alpha_0+\alpha_1+\alpha_2 = n}} \binom{n}{\alpha_0, \alpha_1, \alpha_2} \left(2^{\alpha_0} + 2^{\alpha_1} + 2^{\alpha_2}\right)^n \\
&= \sum_{k=0}^{n} \sum_{\ell=0}^{n-k} \binom{n}{k}\binom{n-k}{\ell} \left(2^k + 2^\ell + 2^{n-k-\ell}\right)^n.
\end{aligned}$$

Remember that we want to show that $3^m \cdot \varepsilon_n = 2^{n^2 - \Omega(n)}$. Hence, it remains to prove an upper bound of order $3 \cdot 2^{n^2} + 2^{n^2-\Omega(n)}$ on the above sum. This is provided in Lemma 7.1 at the end of the section. □

Now we know the number of solutions of System (3) for fixed row and column sums r_i and c_i. It remains to count the number of different choices for the r_i and c_i. For this, we have to take into account Equation (2) and the necessary condition for solutions from the above claim.

Before we estimate the number of r_i and c_i fulfilling these equations, we present another technical tool. For $c \in \mathbb{Z}_2$ and $d \in \mathbb{Z}_3$, define

$$N_{n,c,d} := \left|\left\{(x_1,\ldots,x_n) \in \mathbb{Z}_3^n \;\middle|\; \sum_{i=1}^{n} [x_i \equiv 0 \bmod 3] \equiv c \bmod 2 \wedge x_1 + \cdots + x_n \equiv d \bmod 3\right\}\right|.$$

Claim: *For arbitrary $c \in \mathbb{Z}_2$ and $d \in \mathbb{Z}_3$, $N_{n,c,d} = (1/6) \cdot 3^n \cdot \left(1 \pm 2^{-\Omega(n)}\right)$.*

Proof of the claim: We apply Theorem 6.7 from the last section. We consider the group $G = \mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$. Define $A := \{(1,0),(0,1),(0,-1)\} \subseteq G$, for $k = 1,\ldots,n$, and $b := (c,d) \in G$. Let $y_k = (y_{k,1}, y_{k,2}) \in A$, for $k = 1,\ldots,n$. We have $y_1 + \cdots + y_n = b$ in G if and only if

$$y_{1,1} + \cdots + y_{n,1} \equiv c \bmod 2 \quad \wedge \quad y_{1,2} + \cdots + y_{n,2} \equiv d \bmod 3.$$

Furthermore, $y_{k,1} = [y_{k,2} \equiv 0 \bmod 3]$ for all k due to the definition of A. Hence, the number of solutions of $y_1 + \cdots + y_n = b$ in G with $y_k \in A$ for all k is equal to $N_{n,c,d}$. The formula from Theorem 6.7 yields

$$N_{n,c,d} = \frac{1}{6} \cdot 3^n + \frac{1}{6} \cdot \sum_{u \in \mathbb{Z}_6,\, u \neq 0} \overline{\chi_u(b)} \prod_{k=1}^{n} \sum_{a \in A} \chi_u(a),$$

where χ_u, $u \in \mathbb{Z}_6$, is defined by $\chi_u(v) := e^{2\pi i u v/2} \cdot e^{2\pi i u v/3} = (-1)^{uv} \cdot e^{2\pi i u v/3}$ for $v \in \mathbb{Z}_6$ (see Proposition 6.6).

We have $(1,0) = 3$, $(0,1) = 4$, and $(0,-1) = 2$ in $\mathbb{Z}_6$. Thus

$$\begin{aligned}
\sum_{a \in A} \chi_u(a) &= (-1)^{2u} \cdot e^{4\pi i u/3} + (-1)^{3u} \cdot e^{6\pi i u/3} + (-1)^{4u} \cdot e^{8\pi i u/3} \\
&= e^{2\pi i u/3} + e^{-2\pi i u/3} + (-1)^u = 2 \cdot \operatorname{Re}\left(e^{2\pi i u/3}\right) + (-1)^u \\
&= \begin{cases} 3, & \text{if } u \equiv 0 \quad \bmod 6; \\ -2, & \text{if } u \equiv 1,5 \bmod 6; \\ 1, & \text{if } u \equiv 3 \quad \bmod 6; \\ 0, & \text{if } u \equiv 2,4 \bmod 6. \end{cases}
\end{aligned}$$

We substitute this into the above formula and obtain the estimate

$$\left| N_{n,c,d} - \frac{1}{6} \cdot 3^n \right| \leq \frac{1}{6} \cdot \sum_{u \neq 0} \left| \sum_{a \in A} \chi_u(a) \right|^n = \frac{1}{6} \cdot \left(2^{n+1} + 1\right).$$

This proves the claim. □

At this point, we can put all results together. For each $d \in \mathbb{Z}_3$, there are exactly $N_{\xi,d} \cdot N_{\eta,d}$ choices for $r_1,\ldots,r_n \in \mathbb{Z}_3$ and $c_1,\ldots,c_n \in \mathbb{Z}_3$ such that Equation (2) is fulfilled and

$$r_1 + \cdots + r_n \equiv c_1 + \cdots + c_n \equiv d \bmod 3.$$

Altogether, we have

$$\sum_{d\in\mathbb{Z}_3} N_{\xi,d}\cdot N_{\eta,d} = \frac{1}{12}\cdot 3^{2n}\cdot\left(1\pm 2^{-\Omega(n)}\right)$$

choices for the r_i and c_i. For each of these choices, we obtain $3^{-2n+1}\cdot 2^{n^2}\cdot\left(1\pm 2^{-\Omega(n)}\right)$ inputs in $\mathrm{RT}_n^{-1}(\xi)\cap\mathrm{CT}_n^{-1}(\eta)$ by our first claim. Thus, the total number of inputs is

$$\frac{1}{12}\cdot 3^{2n}\cdot\left(1\pm 2^{-\Omega(n)}\right)\cdot 3^{-2n+1}\cdot 2^{n^2}\cdot\left(1\pm 2^{-\Omega(n)}\right) = \frac{1}{4}\cdot 2^{n^2}\cdot\left(1\pm 2^{-\Omega(n)}\right).$$

□

To complete the proof, it only remains to verify the following fact already used above.

Lemma 7.1:

$$\sum_{k=0}^{n}\sum_{\ell=0}^{n-k}\binom{n}{k}\binom{n-k}{\ell}\left(2^k+2^\ell+2^{n-k-\ell}\right)^n = 3\cdot 2^{n^2}+2^{n^2-\Omega(n)}.$$

Proof: Our plan is to split the summation intervals of the inner and outer sum and to handle the resulting partial sums separately. We first remove the term for $k=n$, this is equal to $(2^n+2)^n$. Let S be the remaining double sum, where $k\le n-1$.

We start with a decomposition of the outer sum. Let α be a constant with $0<\alpha<1/2$ (fixed later on). Define S_1, S_2, and S_3 as the partial sums which are obtained by restricting the outer index k to the intervals $0\le k\le\lfloor\alpha n\rfloor$, $\lceil\alpha n\rceil\le k\le\lfloor(1-\alpha)n\rfloor$, and $\lfloor(1-\alpha)n\rfloor\le k\le n-1$, resp. Then $S\le S_1+S_2+S_3$.

Sum S_2: Let us first consider S_2. We have $\lceil\alpha n\rceil\le k\le\lfloor(1-\alpha)n\rfloor$, and for integers k, this is equivalent to $\alpha n\le k\le(1-\alpha)n$. We observe that, for fixed k or ℓ, the function $2^k+2^\ell+2^{n-k-\ell}$ attains its maximal values at the borders of the summation interval of the remaining variable. Using the bounds for k and ℓ, we thus get

$$2^k+2^\ell+2^{n-k-\ell}\le 2^k+1+2^{n-k}\le 2^{(1-\alpha)n+1}+1.$$

Hence,

$$S_2 \;\le\; \left(2^{(1-\alpha)n+1}+1\right)^n\sum_{k=\lceil\alpha n\rceil}^{\lfloor(1-\alpha)n\rfloor}\binom{n}{k}\sum_{\ell=0}^{n-k}\binom{n-k}{\ell} \;\le\; \left(2^{(1-\alpha)n+1}+1\right)^n 3^n,$$

where we have generously estimated the sum of the binomial coefficients by using the binomial theorem. Furthermore,

$$\left(2^{(1-\alpha)n+1}+1\right)^n 3^n \;=\; 2^{(1-\alpha)n^2+n+(\log_2 3)n}\left(1+2^{-(1-\alpha)n-1}\right)^n.$$

Since α is a positive constant, this is upper bounded by $2^{\gamma n^2}$ for some constant $\gamma < 1$. Thus, the sum S_2 turns out to be "very small."

Sums S_1 and S_3: We further split the sums S_1 and S_3 into three partial sums each, depending on the value of ℓ. Let β be a constant with $0 < \beta < 1/2$. For $i = 1, 3$, define $S_{i,1}$, $S_{i,2}$, and $S_{i,3}$ as the partial sums which are obtained from S_i by restricting ℓ to the intervals $0 \le \ell \le \lfloor \beta(n-k) \rfloor$, $\lceil \beta(n-k) \rceil \le \ell \le \lfloor (1-\beta)(n-k) \rfloor$, and $\lfloor (1-\beta)(n-k) \rfloor \le \ell \le n-k$, resp. We observe that

$$\sum_{\ell = \lfloor (1-\beta)(n-k) \rfloor}^{n-k} \binom{n-k}{\ell} \left(2^k + 2^\ell + 2^{n-k-\ell}\right)^n = \sum_{\ell=0}^{\lfloor \beta(n-k) \rfloor} \binom{n-k}{\ell} \left(2^k + 2^\ell + 2^{n-k-\ell}\right)^n .$$

Hence, $S_{i,3} = S_{i,1}$ for $i = 1, 3$, and $S_1 \le 2 \cdot S_{1,1} + S_{1,2}$ as well as $S_3 \le 2 \cdot S_{3,1} + S_{3,2}$. It is therefore sufficient to derive estimates for the sums $S_{1,1}$, $S_{1,2}$ and $S_{3,1}$, $S_{3,2}$.

Sum $S_{1,1}$: Let us look at $S_{1,1}$ first, where $0 \le k \le \lfloor \alpha n \rfloor$ and $0 \le \ell \le \lfloor \beta(n-k) \rfloor$. We pull out the term for $k = \ell = 0$, which is equal to $(2^n + 2)^n$, and estimate the sum of the remaining terms. Let $S'_{1,1}$ denote this sum. For indices k and ℓ where $k \ge 1$, we get

$$2^k + 2^\ell + 2^{n-k-\ell} \le \max \left\{ 2^{n-1} + 3,\ 2^{\lfloor \alpha n \rfloor} + 1 + 2^{\lfloor (1-\alpha) n \rfloor} \right\} .$$

Since α is a constant with $0 < \alpha < 1$, the maximum is equal to $2^{n-1} + 3$ for n large enough. The same upper bound is obtained for indices k and ℓ where $\ell \ge 1$. Thus,

$$S'_{1,1} \ \le \ \left(2^{n-1} + 3\right)^n \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \sum_{\ell=0}^{\lfloor \beta(n-k) \rfloor} \binom{n-k}{\ell}.$$

For the partial sums of the first binomial coefficients, we use the well-known asymptotically optimal estimate (see, e. g., [22]):

$$\sum_{\ell=0}^{\lfloor \beta(n-k) \rfloor} \binom{n-k}{\ell} \ \le \ 2^{H(\beta)(n-k) - (1/2)\log_2(n-k) + c} \ \le \ 2^{H(\beta)(n-k) + c'},$$

where c and c' are constants and $H(x) = -\big(x \log_2 x + (1-x)\log_2(1-x)\big)$ is the entropy function.

Hence,

$$\begin{aligned} S'_{1,1} \ &\le \ \left(2^{n-1} + 3\right)^n \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} 2^{H(\beta)(n-k) + c'} \ \le \ \left(2^{n-1} + 3\right)^n 2^{H(\beta) n + c'} \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \\ &\le \ \left(2^{n-1} + 3\right)^n 2^{(H(\alpha) + H(\beta)) n + c''}, \quad \text{for some constant } c'' \\ &= \ 2^{n^2 - (1 - H(\alpha) - H(\beta)) n + c''} \left(1 + 3 \cdot 2^{-(n-1)}\right)^n . \end{aligned}$$

By choosing α and β small enough such that $H(\alpha) + H(\beta) < 1$, we obtain a bound of order $2^{n^2-\Omega(n)}$. Altogether, $S_{1,1} = 2^{n^2} + 2^{n^2-\Omega(n)}$.

Sum $S_{3,1}$: Here we have $\lfloor (1-\alpha)n \rfloor \le k \le n-1$ and $0 \le \ell \le \lfloor \beta(n-k) \rfloor$. Due to these bounds, we again get $2^k + 2^\ell + 2^{n-k-\ell} \le 2^{n-1} + 3$, and

$$S_{3,1} \;\le\; \left(2^{n-1}+3\right)^n \sum_{k=\lfloor (1-\alpha)n \rfloor}^{n-1} \binom{n}{k} \sum_{\ell=0}^{\lfloor \beta(n-k) \rfloor} \binom{n-k}{\ell}.$$

These partial sums of binomial coefficients may be estimated in the same way as for $S'_{1,1}$, which yields $S_{3,1} = 2^{n^2-\Omega(n)}$.

Sum $S_{1,2}$: For the sum $S_{1,2}$, where $0 \le k \le \lfloor \alpha n \rfloor$ and $\lceil \beta(n-k) \rceil \le \ell \le \lfloor (1-\beta)(n-k) \rfloor$, we can apply the same ideas as for the sum S_2 at the beginning. We have

$$2^k + 2^\ell + 2^{n-k-\ell} \le 2^k + 2^{(1-\beta)(n-k)+1} + 1 \le \max\left\{2^{(1-\beta)n+1} + 2,\; 2^{\alpha n} + 2^{(1-\alpha)(1-\beta)n+1}\right\}.$$

Since α and β are positive constants, the maximum is bounded from above by $2^{\gamma' n}$ for some constant $\gamma' < 1$ if n is large enough, and we obtain

$$S_{1,2} \;\le\; 2^{\gamma' n^2} \sum_{k=0}^{\lfloor \alpha n \rfloor} \binom{n}{k} \sum_{\ell=\lceil \beta(n-k) \rceil}^{\lfloor (1-\beta)(n-k) \rfloor} \binom{n-k}{\ell} \;\le\; 2^{\gamma' n^2} \cdot 3^n \;\le\; 2^{\gamma'' n^2}$$

for some constant $\gamma'' < 1$.

Sum $S_{3,2}$: For the last remaining sum, we have $\lfloor (1-\alpha)n \rfloor \le k \le n-1$ and $\lceil \beta(n-k) \rceil \le \ell \le \lfloor (1-\beta)(n-k) \rfloor$. Due to these bounds,

$$2^k + 2^\ell + 2^{n-k-\ell} \le 2^k + 2^{(1-\beta)(n-k)+1} \le \max\left\{2^{n-1} + 2^{2-\beta},\; 2^{(1-\alpha)n} + 2^{\alpha(1-\beta)n+1}\right\}.$$

For n large enough, the maximum is equal to $2^{n-1} + 2^{2-\beta}$. This yields

$$\begin{aligned}
S_{3,2} \;&\le\; \left(2^{n-1} + 2^{2-\beta}\right)^n \sum_{k=\lfloor (1-\alpha)n \rfloor}^{n-1} \binom{n}{k} \sum_{\ell=\lceil \beta(n-k) \rceil}^{\lfloor (1-\beta)(n-k) \rfloor} \binom{n-k}{\ell} \\
&\le\; \left(2^{n-1} + 2^{2-\beta}\right)^n \sum_{k=\lfloor (1-\alpha)n \rfloor}^{n-1} \binom{n}{k} 2^{n-k} \\
&\le\; \left(2^{n-1} + 2^{2-\beta}\right)^n \, 2^{\lfloor \alpha n \rfloor} \sum_{k=\lfloor (1-\alpha)n \rfloor}^{n-1} \binom{n}{k} \\
&\le\; \left(2^{n-1} + 2^{2-\beta}\right)^n \, 2^{\lfloor \alpha n \rfloor} 2^{H(\alpha)n+c},
\end{aligned}$$

for some constant c. If we choose α small enough such that $\alpha + H(\alpha) < 1$, this bound is of order $2^{n^2-\Omega(n)}$.

We collect the conditions imposed on α and β. We require that $0 < \alpha < 1/2$, $0 < \beta < 1/2$ and additionally $H(\alpha) + H(\beta) < 1$ and $\alpha + H(\alpha) < 1$. Obviously, α and β can be chosen in this way. Altogether, we have proven:

$$\begin{aligned} S_1 &\leq 2 \cdot S_{1,1} + S_{1,2} = 2 \cdot 2^{n^2} + 2^{n^2-\Omega(n)}, \\ S_2 &\leq 2^{\gamma n^2}, \quad \text{for a constant } \gamma < 1, \\ S_3 &\leq 2 \cdot S_{3,1} + S_{3,2} = 2^{n^2-\Omega(n)}. \end{aligned}$$

Thus, $S = 2 \cdot 2^{n^2} + 2^{n^2-\Omega(n)}$, and together with the term for $k = n$, $(2^n + 2)^n = 2^{n^2} + 2^{n^2-\Omega(n)}$, this is of the claimed size. □

8 Proof of Lemma 3.5

In this section, we always work with a fixed partition of the input variables, and we write combinatorial rectangles as Cartesian products of sets of input assignments. We are going to prove the following bound on the discrepancy of subfunctions of RT_n already used in Section 1.6.

Lemma 3.5: *Let $c = (c_0, c_1, \ldots, c_m)$, where $c_0 \in \mathbb{Z}_2$ and $c_1, \ldots, c_m \in \mathbb{Z}_3$. Define the function* $\mathrm{RT}^*_c \colon \{0,1\}^{2m} \times \{0,1\}^{2m} \to \{0,1\}$ *on vectors $x^1, x^2, y^1, y^2 \in \{0,1\}^m$ by*

$$\mathrm{RT}^*_c\big((x^1, x^2), (y^1, y^2)\big) := \left[\sum_{i=1}^{m} \left[x_i^1 + x_i^2 + y_i^1 + y_i^2 \equiv c_i \bmod 3\right] \equiv c_0 \bmod 2\right].$$

Let $R = A \times B$, where $A, B \subseteq \{0,1\}^{2m}$. Then

$$\mathrm{Disc}(\mathrm{RT}^*_c, R) \leq 2^{-m} + 3^{-m}.$$

The proof of this lemma is based on an adaptation of a technique due to Babai, Hayes, and Kimmel [10]. The key notion of this technique is a measure called “multicolor discrepancy,” which generalizes the discrepancy measure used in communication complexity theory (see Definition 3.4) to arbitrary finite, abelian groups instead of $\mathbb{Z}_2$. This can be applied to RT^*_c by “decomposing” the function into suitable functions over $\mathbb{Z}_3$ as “building blocks.”

We first describe a slightly extended version of the technique from [10].

8.1 Multicolor Discrepancy

For the whole section, let X_1, X_2 be fixed finite sets, and let $X := X_1 \times X_2$. Let G be a finite abelian group.

Definition 8.1: Let $f\colon X_1 \times X_2 \to G$ be an arbitrary function and R be a combinatorial rectangle in $X = X_1 \times X_2$, i. e., $R = A \times B$, where $A \subseteq X_1$, $B \subseteq X_2$.

For every $Y \subseteq G$, define the *strong Y-discrepancy of f with respect to R*, $\Gamma_Y(f, R)$, by

$$\Gamma_Y(f, R) := |\varepsilon_Y(f, R)|, \quad \text{where} \quad \varepsilon_Y(f, R) := \frac{1}{|X|}\left(\left|f^{-1}(Y) \cap R\right| - |R| \cdot \frac{|Y|}{|G|}\right).$$

The expression $(1/|X|) \cdot |f^{-1}(Y) \cap R|$ measures the portion of the inputs in R which is mapped to "colors" in the set Y by the function f. Intuitively, the strong Y-discrepancy of f is close to zero for *all* rectangles R iff the Y-colored inputs are "randomly" distributed in the input space. More precisely, this means that every rectangle R gets approximately the same number of Y-colored inputs as if we would label the inputs in R by values chosen randomly from G according to the uniform distribution, which would give an expected number of $|R| \cdot |Y|/|G|$ inputs with color from Y.

Babai, Hayes, and Kimmel consider strong Y-discrepancy for one-element sets Y in [10]. Observe that $\varepsilon_Y(f, R) = \sum_{y \in Y} \varepsilon_{\{y\}}(f, R)$.

The goal is to derive small upper bounds on the strong discrepancy $\Gamma_Y(f, R)$ for a given function f and arbitrary rectangles R. Babai, Hayes, and Kimmel observed that there is a way to obtain such bounds by using the technique of character sums (or generalized Fourier transforms). For the following, we use the definitions and basic facts already introduced in Section 6. Additionally, we use δ_A for the characteristic function of an arbitrary set $A \subseteq G$, i. e., $\delta_A(x) := 1$ if $x \in A$, and $\delta_A(x) := 0$, otherwise.

To derive bounds on strong discrepancy, Babai, Hayes and Kimmel consider the following alternative measure, which may appear to be rather unrelated to strong discrepancy at the first glance.

Definition 8.2: Let $f\colon X \to G$ be an arbitrary function and $R = A \times B$, where $A \subseteq X_1$, $B \subseteq X_2$. Furthermore, let $\chi \in \widehat{G}$ be a character of G. Define the *weak χ-discrepancy of f with respect to R* by

$$\Gamma_\chi^{\text{weak}}(f, R) := \frac{1}{|X|}\left|\sum_{x \in R} \chi(f(x))\right|.$$

The following fact proven in [10] (Prop. 2.9) provides a basic relation between weak discrepancy and strong Y-discrepancy for one-element sets $Y = \{y\}$.

Proposition 8.3: *For all $\chi \in \widehat{G}$, $\chi \neq \chi_0$,*

$$\Gamma_\chi^{\text{weak}}(f, R) = \left|\sum_{y \in G} \chi(y)\, \varepsilon_{\{y\}}(f, R)\right|.$$

By this proposition,

$$\Gamma_\chi^{\text{weak}}(f, R) \le |G| \cdot \max_{y \in G} |\varepsilon_{\{y\}}(f, R)| = |G| \cdot \max_{y \in G} \Gamma_{\{y\}}(f, R).$$

This may serve as a justification for the term “weak discrepancy.”

Babai, Hayes, and Kimmel discuss the relationship between strong and weak discrepancy in more detail in [10]. The decisive point for the applicability of the whole approach of “multicolor discrepancy” is that it is also possible to bound strong discrepancy in terms of weak discrepancy. Below, we present the central lemma which establishes such a relation. We consider strong Y-discrepancy for arbitrary sets Y. The same has already been proven for the case $|Y| = 1$ in [10] (Lemma 2.7).

Lemma 8.4:

$$\Gamma_Y(f, R) \leq \frac{1}{|G|} \sum_{\substack{\chi \in \widehat{G} \\ \chi \neq \chi_0}} \left|\widehat{\delta}_Y(\chi)\right| \cdot \Gamma^{\mathrm{weak}}_{\overline{\chi}}(f, R).$$

Proof: By a straightforward adaptation of the proof of Lemma 2.7 from [10]. First, we observe that, by the formula for the inverse Fourier transform,

$$\delta_Y(y) = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \overline{\chi(y)}\, \widehat{\delta}_Y(\chi), \quad \text{for all } y \in G.$$

Substituting this into the definition of $\varepsilon_Y(f, R)$, we get

$$\begin{aligned} \varepsilon_Y(f, R) &= \sum_{y \in G} \varepsilon_{\{y\}}(f, R)\, \delta_Y(y) = \sum_{y \in G} \varepsilon_{\{y\}}(f, R) \cdot \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \overline{\chi(y)}\, \widehat{\delta}_Y(\chi) \\ &= \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{\delta}_Y(\chi) \sum_{y \in G} \overline{\chi(y)}\, \varepsilon_{\{y\}}(f, R). \end{aligned}$$

For $\chi = \chi_0$, the inner sum is

$$\sum_{y \in G} \varepsilon_{\{y\}}(f, R) = \frac{1}{|X|} \sum_{y \in G} \left(\left| f^{-1}(y) \cap R \right| - |R| \cdot \frac{1}{|G|} \right) = 0.$$

Hence,

$$\Gamma_Y(f, R) = |\varepsilon_Y(f, R)| \leq \frac{1}{|G|} \sum_{\substack{\chi \in \widehat{G} \\ \chi \neq \chi_0}} \widehat{\delta}_Y(\chi) \left| \sum_{y \in G} \overline{\chi(y)}\, \varepsilon_{\{y\}}(f, R) \right|.$$

By Proposition 8.3, this is the claimed estimate. □

8.2 Some Facts on Fourier Transforms

We present two lemmas which will be used later on. The aim is to describe the size of sets in terms of the Fourier transforms of their characteristic function. For the required facts on characters of finite groups, we refer to the appendix.

Lemma 8.5: *Let G and G' be finite abelian groups, and let $\varphi\colon G \to G'$ be a group homomorphism. Then the set $\left\{\chi \circ \varphi \mid \chi \in \widehat{G'}\right\}$ is a subgroup of $\widehat{G}$, and there is subgroup H of G such that $\widehat{H} = \left\{\chi \circ \varphi \mid \chi \in \widehat{G'}\right\}$. Furthermore, if φ is onto ($\varphi(G) = G'$), then $\left|\widehat{H}\right| = |H| = |G'|$.*

Proof: It is easy to verify that $\left\{\chi \circ \varphi \mid \chi \in \widehat{G'}\right\}$ is a subgroup of $\widehat{G}$ by elementary calculations. Since $G \cong \widehat{G}$ (Theorem 6.2), it follows that there is a subgroup H of G with $\widehat{H} = \left\{\chi \circ \varphi \mid \chi \in \widehat{G'}\right\}$. The last claimed fact is again easy to verify. □

Lemma 8.6: *Let G be a finite abelian group, and let H be a subgroup of G.*

(1) For $A \subseteq G$, $\displaystyle\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|^2 = |A|.$

(2) For $A, B \subseteq G$, $\displaystyle\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|\left|\widehat{\delta}_B(\chi)\right| \leq \sqrt{|A||B|}.$

Proof: *Part (1):* By the definition of the Fourier transform, we have

$$\begin{aligned}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|^2 &= \sum_{\chi\in\widehat{H}}\overline{\left(\sum_{a\in G}\chi(a)\,\delta_A(a)\right)}\sum_{b\in G}\chi(b)\,\delta_A(b)\\ &= \sum_{\chi\in\widehat{H}}\overline{\left(\sum_{a\in A}\chi(a)\right)}\sum_{b\in A}\chi(b)\\ &= \sum_{\chi\in\widehat{H}}\sum_{a,b\in A}\overline{\chi(a)}\chi(b)\\ &= \sum_{a,b\in A}\sum_{\chi\in\widehat{H}}\chi(a-b) \;=\; |H|\cdot|A|.\end{aligned}$$

The last line follows from the orthogonality relations for the characters of H (Theorem 6.3).

Part (2): By the Cauchy-Schwarz inequality in $\mathbb{R}^{|H|}$,

$$\begin{aligned}\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|\left|\widehat{\delta}_B(\chi)\right| &\leq \frac{1}{|H|}\left(\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|^2\right)^{1/2}\left(\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_B(\chi)\right|^2\right)^{1/2}\\ &= \left(\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_A(\chi)\right|^2\right)^{1/2}\left(\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\left|\widehat{\delta}_B(\chi)\right|^2\right)^{1/2} \;=\; \sqrt{|A||B|}.\end{aligned}$$

For the final step, we have applied the first part of the lemma. □

8.3 Application to RowTest

We describe the function RT^*_c in the following way. Let $c = (c_0, c_1, \dots, c_m)$, where $c_0 \in \mathbb{Z}_2$ and $c_1, \dots, c_m \in \mathbb{Z}_3$.

Define $\varphi \colon \mathbb{Z}_3^m \times \mathbb{Z}_3^m \to \mathbb{Z}_3^m$ by $\varphi(u, v) := u + v$, where $u, v \in \mathbb{Z}_3^m$. Furthermore, define $f : \mathbb{Z}_3^{2m} \times \mathbb{Z}_3^{2m} \to \mathbb{Z}_3^m$ by $f(x, y) := \varphi(x) + \varphi(y)$, where $x, y \in \mathbb{Z}_3^{2m}$. Finally, let

$$Y_c := \left\{ x \in \mathbb{Z}_3^m \;\middle|\; \sum_{i=1}^{m} [x_i \equiv c_i \bmod 3] \equiv c_0 \bmod 2 \right\}.$$

Then, for all $x, y \in \{0, 1\}^{2m} \subseteq \mathbb{Z}_3^{2m}$, $\mathrm{RT}^*_c(x, y) = 1$ iff $f(x, y) \in Y_c$.

Notice that, by these definitions, we have extended the input space $\{0, 1\}^{2m} \times \{0, 1\}^{2m}$ of RT^*_c to the larger input space $\mathbb{Z}_3^{2m} \times \mathbb{Z}_3^{2m}$ of f. This is crucial for the smooth application of the algebraic concepts used for the technique of multicolor discrepancy. For the remainder of this section, we work within the groups $G := \left(\mathbb{Z}_3^{2m}, +\right)$ or $G' := \left(\mathbb{Z}_3^{m}, +\right)$ (where $+$ denotes the usual vector addition).

Lemma 8.7: *For all $c = (c_0, c_1, \dots, c_m)$, where $c_0 \in \mathbb{Z}_2$ and $c_1, \dots, c_m \in \mathbb{Z}_3$,*

$$\max_{\chi \in \widehat{G'},\, \chi \neq \chi_0} \left| \widehat{\delta}_{Y_c}(\chi) \right| \leq 2^{m-1}.$$

Proof: We first consider the case $c_1 = \cdots = c_m = 0$. Define $A_m := Y_{0,0,\dots,0}$ and $B_m := Y_{1,0,\dots,0}$, i. e., A_m contains the vectors in $\mathbb{Z}_3^m$ with an even number of zero entries, whereas B_m contains the vectors with an odd number of zero entries. Define $\omega := e^{2\pi i/3}$. Then all characters of $G' = \left(\mathbb{Z}_3^m, +\right)$ are obtained by defining $\chi_u(v) := \omega^{\langle u,v \rangle}$ for $u, v \in \mathbb{Z}_3^m$, where $\langle u, v \rangle := \sum_{i=1}^{3} u_i v_i$ is the standard inner product in $\mathbb{Z}_3^m$ (Proposition 6.6). Finally, for $u \in \mathbb{Z}_3^m$ let

$$S_m(u) := \sum_{v \in A_m} \omega^{\langle u,v \rangle}, \quad \text{and} \quad T_m(u) := \sum_{v \in B_m} \omega^{\langle u,v \rangle}.$$

By these definitions, $\widehat{\delta}_{Y_c}(\chi_u) = S_m(u)$ or $\widehat{\delta}_{Y_c}(\chi_u) = T_m(u)$ (depending on the value of c_0).

First, we consider the case $u = 0$. Then

$$S_m(u) = |A_m| = \sum_{k=0}^{m} \binom{m}{k} \frac{1}{2} (1 + (-1))^k \, 2^{m-k} = \frac{1}{2} \left(3^m - 1\right),$$
$$T_m(u) = |B_m| = \frac{1}{2} \left(3^m + 1\right).$$

Now let $u = (u_1, \dots, u_m) \neq 0$. There is at least one i such that $u_i \neq 0$. Define $u' := (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_m)$, and for an arbitrary vector $v = (v_1, \dots, v_m) \in \mathbb{Z}_3^m$, let $v' := (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_m)$.

We have

$$\begin{aligned}
S_m(u) = \sum_{v\,\in\,A_m} \omega^{\langle u,v\rangle} &= \sum_{\substack{v\,\in\,A_m\\ v_i\,=\,0}} \omega^{\left\langle u',v'\right\rangle} + \sum_{\substack{v\,\in\,A_m\\ v_i\,=\,1}} \omega^{\left\langle u',v'\right\rangle+u_i} + \sum_{\substack{v\,\in\,A_m\\ v_i\,=\,-1}} \omega^{\left\langle u',v'\right\rangle-u_i}\\
&= \sum_{v'\,\in\,B_{m-1}} \omega^{\left\langle u',v'\right\rangle} + \omega^{u_i} \sum_{v'\,\in\,A_{m-1}} \omega^{\left\langle u',v'\right\rangle} + \overline{\omega^{u_i}} \sum_{v'\,\in\,A_{m-1}} \omega^{\left\langle u',v'\right\rangle}\\
&= T_{m-1}\left(u'\right) + \left(\omega^{u_i} + \overline{\omega^{u_i}}\right) S_{m-1}\left(u'\right)\\
&= T_{m-1}\left(u'\right) - S_{m-1}\left(u'\right)
\end{aligned}$$

Analogously, $T_m(u) = S_{m-1}\left(u'\right) - T_{m-1}\left(u'\right)$. Hence, if $u' \neq 0$,

$$\begin{aligned}
S_m(u) &= -2 \cdot S_{m-1}\left(u'\right), \quad \text{and}\\
T_m(u) &= -2 \cdot T_{m-1}\left(u'\right).
\end{aligned}$$

Let u have k nonzero entries, and let u' be a vector obtained from u by deleting $k-1$ nonzero entries (and decreasing the size of the vector accordingly). Then, by induction,

$$S_m(u) = (-2)^{k-1} \cdot S_{m-k+1}\left(u'\right) = (-2)^{k-1}\left(T_{m-k}(0) - S_{m-k}(0)\right) = (-2)^{k-1}.$$

Analogously,

$$T_m(u) = (-2)^{k-1} \cdot T_{m-k+1}\left(u'\right) = (-2)^{k-1}\left(S_{m-k}(0) - T_{m-k}(0)\right) = (-1)^k\, 2^{k-1}.$$

Especially, we obtain that $|S_m(u)|, |T_m(u)| \leq 2^{m-1}$ for all $u \neq 0$.

It finally remains to consider the general case where $c = (c_0, c_1, \ldots, c_m)$ is arbitrarily chosen. Let $c' := (c_1, \ldots, c_m)^\top \in \mathbb{Z}_3^m$. Obviously, we have $Y_c = A_m - c'$ or $Y_c = B_m - c'$.

The claim follows from the fact that the absolute value of a Fourier coefficient of a function is invariant under translations of the inputs. To see this, define $\tau_a \colon G' \to G'$ by $\tau_a(x) := x + a$, where $a \in G'$ is fixed. Let $g \colon G' \to \mathbb{C}$ be an arbitrary function. Then

$$|\widehat{g \circ \tau_a}(\chi)| = \left|\sum_{u \in G'} \chi(u) g(u+a)\right| = \left|\sum_{u \in G'} \chi(u+a) g(u)\right| = \left|\chi(a) \sum_{u \in G'} \chi(u) g(u)\right| = |\widehat{g}(\chi)|.$$

□

Lemma 8.8: *Let $c = (c_0, c_1, \ldots, c_m)$, where $c_0 \in \mathbb{Z}_2$, $c_1, \ldots, c_m \in \mathbb{Z}_3$. Let $G = \left(\mathbb{Z}_3^{2m}, +\right)$ and $G' = \left(\mathbb{Z}_3^m, +\right)$. Let $f : G \times G \to G'$ be defined by $f(x,y) := \varphi(x) + \varphi(y)$ for $x, y \in G$, where $\varphi \colon G \to G'$ with $\varphi(u,v) := u + v$ for $(u,v) \in G = \mathbb{Z}_3^{2m}$. Then*

$$\Gamma_{Y_c}(f, R) \leq 3^{-4m} \cdot 2^{m-1} \cdot \sqrt{|R|}$$

for all rectangles $R = A \times B$, where $A, B \subseteq G$.

Proof: We have

$$\begin{aligned}
|\varepsilon_{Y_c}(f,R)| &\le 3^{-m} \sum_{\substack{\chi\in\widehat{G'}\\ \chi\neq\chi_0}} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot \Gamma^{\text{weak}}_{\chi}(f,R) \qquad \text{(by Lemma 8.4)}\\
&= 3^{-m} \sum_{\substack{\chi\in\widehat{G'}\\ \chi\neq\chi_0}} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot 3^{-4m} \left|\sum_{(x,y)\in R} \chi(\varphi(x)+\varphi(y))\right| \qquad \text{(by Def. 8.2)}\\
&= 3^{-m} \sum_{\substack{\chi\in\widehat{G'}\\ \chi\neq\chi_0}} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot 3^{-4m} \left|\sum_{x\in A} \chi(\varphi(x))\right| \left|\sum_{y\in B} \chi(\varphi(y))\right| \qquad \text{(using } R = A\times B\text{)}
\end{aligned}$$

The mapping φ is obviously a group homomorphism (a linear transform) from $G = \mathbb{Z}_3^{2m}$ to $G' = \mathbb{Z}_3^m$, and it is onto. By Lemma 8.5, there is a subgroup H of G with $\widehat{H} := \left\{\chi\circ\varphi \mid \chi\in\widehat{G'}\right\}$, and $|H| = 3^m$. By the definition of the Fourier transform, we have

$$\left|\sum_{x\in A} \chi(\varphi(x))\right| = \left|\widehat{\delta}_A(\chi\circ\varphi)\right|$$

for all $\chi \in \widehat{G'}$. Lemma 8.6 from Section 8.2 yields

$$3^{-m}\sum_{\chi\in\widehat{G'}} \left|\sum_{x\in A} \chi(\varphi(x))\right| \left|\sum_{y\in B} \chi(\varphi(y))\right| = 3^{-m}\sum_{\psi\in\widehat{H}} \left|\widehat{\delta}_A(\psi)\right| \left|\widehat{\delta}_B(\psi)\right| \le \sqrt{|A||B|}. \qquad (*)$$

Now we are ready to complete the estimate for $\varepsilon_{Y_c}(f,R)$. We have

$$\begin{aligned}
|\varepsilon_{Y_c}(f,R)| &\le 3^{-m} \sum_{\substack{\chi\in\widehat{G'}\\ \chi\neq\chi_0}} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot 3^{-4m} \left|\sum_{x\in A} \chi(\varphi(x))\right| \left|\sum_{y\in B} \chi(\varphi(y))\right|\\
&\le 3^{-4m} \cdot \max_{\chi\in\widehat{G'},\,\chi\neq\chi_0} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot 3^{-m} \sum_{\substack{\chi\in\widehat{G'}\\ \chi\neq\chi_0}} \left|\sum_{x\in A} \chi(\varphi(x))\right| \left|\sum_{y\in B} \chi(\varphi(y))\right|\\
&\le 3^{-4m} \cdot \max_{\chi\in\widehat{G'},\,\chi\neq\chi_0} \left|\widehat{\delta}_{Y_c}(\overline{\chi})\right| \cdot \sqrt{|A||B|} \qquad \text{(using } (*)\text{)}\\
&\le 3^{-4m} \cdot 2^{n-1} \cdot \sqrt{|A||B|} \qquad \text{(by Lemma 8.7).}
\end{aligned}$$

□

Finally, we use Lemma 8.8 to prove the desired upper bound on the discrepancy of RT_c^*.

Proof of Lemma 3.6: Let $S := \{0,1\}^{2m} \times \{0,1\}^{2m}$. We use the trivial embedding of S into $X = \mathbb{Z}_3^{2m} \times \mathbb{Z}_3^{2m}$. Let $R \subseteq S$. Then we have

$$\begin{aligned}(\mathrm{RT}_c^*)^{-1}(1) \cap R &= f^{-1}(Y_c) \cap R, \quad \text{and} \\ (\mathrm{RT}_c^*)^{-1}(0) \cap R &= f^{-1}(\overline{Y_c}) \cap R,\end{aligned}$$

where f is the function from Lemma 8.8. Furthermore, $\overline{Y_{(c_0,c_1,\ldots,c_m)}} = Y_{(\overline{c_0},c_1,\ldots,c_m)}$, and

$$|Y_{(0,c_1,\ldots,c_m)}| = \frac{1}{2}\left(3^m - 1\right), \quad \text{and} \quad |Y_{(1,c_1,\ldots,c_m)}| = \frac{1}{2}\left(3^m + 1\right).$$

This follows in the same way as for the case $c_1 = \cdots = c_m = 0$ considered in the proof of Lemma 8.7.

Let $\varepsilon := \max\{\Gamma_{Y_c}(f,R), \Gamma_{\overline{Y_c}}(f,R)\}$, where R is an arbitrary rectangle in $S = \{0,1\}^{2m} \times \{0,1\}^{2m}$. By the definition of strong discrepancy,

$$\begin{aligned}\frac{1}{|X|}\left|\left|f^{-1}(Y_c) \cap R\right| - |R| \cdot |Y_c|/|G'|\right| &\le \varepsilon, \quad \text{and} \\ \frac{1}{|X|}\left|\left|f^{-1}(\overline{Y_c}) \cap R\right| - |R| \cdot |\overline{Y_c}|/|G'|\right| &\le \varepsilon.\end{aligned}$$

Thus,

$$\frac{1}{|S|}\left|\left|f^{-1}(Y_c) \cap R\right| - \left|f^{-1}(\overline{Y_c}) \cap R\right|\right| \le 2\varepsilon \cdot \frac{|X|}{|S|} + \frac{|R|}{|S||G'|}\underbrace{\left||Y_c| - |\overline{Y_c}|\right|}_{=1} \le 2\varepsilon \cdot \frac{|X|}{|S|} + \frac{1}{|G'|}.$$

Rewriting this for the function RT_c^* instead of f, we obtain

$$\frac{1}{|S|}\left|\left|(\mathrm{RT}_c^*)^{-1}(1) \cap R\right| - \left|(\mathrm{RT}_c^*)^{-1}(0) \cap R\right|\right| \le 2\varepsilon \cdot \frac{|X|}{|S|} + \frac{1}{|G'|}.$$

It only remains to substitute the upper bound

$$\varepsilon \le 3^{-4m} \cdot 2^{m-1} \cdot \sqrt{|R|}$$

from Lemma 8.8, which yields

$$\frac{1}{|S|}\left|\left|(\mathrm{RT}_c^*)^{-1}(1) \cap R\right| - \left|(\mathrm{RT}_c^*)^{-1}(0) \cap R\right|\right| \le 2^{-3m} \cdot \sqrt{|R|} + 3^{-m} \le 2^{-m} + 3^{-m}.$$

The last inequality follows using the trivial bound $|R| \le 2^{4m}$. □

Acknowledgments

I would like to thank Ingo Wegener and Martin Dietzfelbinger for supporting this work by proof reading, discussions and helpful hints. Thanks to Thomas Hofmeister for improving Lemma 4.1 and simplifying the proof. Further thanks go to Jayram Thathachar for the important hint to look at the multicolor discrepancy technique of Babai *et al.*, and to Eric Allender for the pointer to his papers on the “UL versus NL” question.

Appendix: Improving Thathachar's Result

Thathachar has proven that a function which is closely related to the function from the main result of this paper has deterministic read-$(k+1)$-times BPs of polynomial size, but requires exponential size for randomized and nondeterministic read-k-times BPs (for not too large k). In this part of the appendix, we improve the lower bound for the randomized case with respect to the error bound.

First, we sketch the technique for proving lower bounds on the size of randomized read-k-times BPs (for arbitrary k) which is behind Thathachar's original result as well as the improved one. This proof technique has first been used in the conference version of this work [47] and is based on ideas of Borodin, Razborov, and Smolensky [16] for the nondeterministic case. The key notion of the technique are generalized rectangles defined as follows.

Definition A.1: Let X be a set of variables, $n := |X|$. Let k, a be integers, where $k \geq 1$ and $2 \leq a \leq n$. Let sets $X_1, \dots, X_{ka} \subseteq X$ be given with

(i) $X_1 \cup \dots \cup X_{ka} = X$ and $|X_i| \leq \lceil n/a \rceil$, for $i = 1, \dots, ka$;

(ii) each variable from X appears in at most k of the sets X_i.

A function $r\colon \{0,1\}^n \to \{0,1\}$ is called (k,a)*-rectangle in* $\{0,1\}^n$ *with respect to* $X_1, \dots, X_{ka}$ (or *generalized rectangle*, if we do not care about the parameters) if there are functions $r_1, \dots, r_{ka}\colon \{0,1\}^n \to \{0,1\}$ such that

(i) r_i does not essentially depend on the variables from X_i, for $i = 1, \dots, ka$;

(ii) $r = r_1 \wedge \dots \wedge r_{ka}$.

By setting $k = 1$ and $a = 2$ in this definition, we obtain combinatorial rectangles (as in Definition 1.1) as a special case. The following theorem shows that deterministic read-k-times BPs yield partitions of the input space into (k,a)-rectangles analogous to the partitions into $(1,2)$-rectangles studied in the main part of the paper.

Theorem A.2: *Let G be a deterministic read-k-times BP representing the function $f\colon \{0,1\}^n \to \{0,1\}$ defined on variables from the set X, $|X| = n$. Let $a \geq 2$ be an integer. Then there are (k,a)-rectangles $r_1, \dots, r_t$ (each with its own sets $X_1, \dots, X_{ka}$ according to Definition A.1) such that*

(i) $t \leq (2|G|)^{ka}$*;*

(ii) $r_i^{-1}(1) \subseteq f^{-1}(0)$ *or* $r_i^{-1}(1) \subseteq f^{-1}(1)$*, for all* $i = 1, \dots, t$*;*

(ii) $r_1^{-1}(1) \cup \dots \cup r_t^{-1}(1) = \{0,1\}^n$ *and* $r_i^{-1}(1) \cap r_j^{-1}(1)$ *for* $i \neq j$*.*

The proof is essentially along the same lines as the proof of the analogous result for nondeterministic read-k-times BPs in the paper [16] of Borodin, Razborov, and Smolensky. The above theorem allows to exploit lower bounds on approximations of functions by *generalized* rectangles to prove lower bounds on randomized read-k-times BPs. This works exactly in the same way as described in Section 5.2, we simply have to substitute generalized rectangles where $(1,2)$-rectangles have been used before. A detailed description can be found in [46] and [48].

Next, we define Thathachar's function. Let $q \neq 2$ be a prime and $k \geq 2$. We consider k-dimensional matrices as inputs, the indices of matrix entries are from the hypercube $\{1, \ldots, n\}^k$. For $d \in \{1, \ldots, k\}$ and $i \in \{1, \ldots, n\}$, we define the index set

$$I_i^d := \{(i_1, \ldots, i_{n^k}) \in \{1, \ldots, n\}^k \mid i_d = i\},$$

this is "the ith hyperplane in the dth direction" (e. g., for $k = 2$, the sets I_i^1, I_i^2 contain the indices of rows and columns, resp.). Notice that $\left|I_i^d\right| = n^{k-1}$ for all i and d. For the whole section, let X be a k-dimensional Boolean matrix of variables and let X_i^d be the set of variables in X corresponding to the index set I_i^d.

Definition A.3: Define $\mathrm{CHSP}_n^k\colon \{0,1\}^{n^k} \to \{0,1\}$ ("Conjunctive Hyperplanar Sum-of-Products") on the k-dimensional matrix X of Boolean variables by

$$\mathrm{CHSP}_n^k(X) := \bigwedge_{1 \leq d \leq k} \mathrm{PT}_{n,d}^k(X),$$

where $\mathrm{PT}_{n,d}^k\colon \{0,1\}^{n^k} \to \{0,1\}$ is defined for $d \in \{1, \ldots, k\}$ by

$$\mathrm{PT}_{n,d}^k(X) := \left[\sum_{i=1}^{n} \bigoplus_{x \in X_i^d} x \equiv 0 \bmod q\right].$$

(As usual, "$\oplus$" denotes the addition in $\mathbb{Z}_2$, $a \oplus b := (a+b) \bmod 2$ for $a, b \in \{0,1\}$.)

For the following, we allow that k is a function of n, but we assume that q is a constant with respect to n. We remark that in Thathachar's original paper, the function CHSP_n^k is defined for input matrices over $\{-1, 1\}$. It is easy to verify that Thathachar's results also hold for the usual Boolean input space due to the one-to-one and onto correspondence between the encodings.

Thathachar has proven the following.

Theorem A.4 (Thathachar [51]): *Let* $N = n^{k+1}$ (*the input size of* CHSP_n^{k+1}).

(1) The complement of CHSP_n^{k+1}, $\neg\mathrm{CHSP}_n^{k+1}$, *can be represented by nondeterministic read-once BPs of size* $O\big((k+1)N\big)$;

(2) each nondeterministic read-k-times BP and each randomized read-k-times BP with two-sided error $(1/3) \cdot 2^{-(2q-1)^{k+1}}$ *for* CHSP_n^{k+1} *has size* $\exp\big(\Omega\big(N^{1/(k+1)} \cdot k^{-3} \cdot 2^{-2k}\big)\big)$.

The first part of this theorem is easy to see. In a nondeterministic read-once BP for $\neg\mathrm{CHSP}_n^{k+1}$, we guess a single direction $d \in \{1, \ldots, k+1\}$ and then evaluate $\neg\mathrm{PT}_{n,d}^{k+1}$ by a deterministic read-once BP of size $O\big(n^{k+1}\big)$. For a 1-input of $\neg\mathrm{CHSP}_n^{k+1}$, at least one of the $k+1$ functions $\neg\mathrm{PT}_{n,d}^{k+1}$ yields the output 1. This nondeterministic read-once BP can also be seen as a randomized read-once BP with one-sided error $1 - 1/(k+1)$, hence we even have $\mathrm{CHSP}_n^{k+1} \in \mathrm{coRP}_{1-1/(k+1)}$-BP1.

The lower bounds in Part (2) are based on the technique of Borodin, Razborov, and Smolensky for the nondeterministic case and the variant of this technique described above for the randomized case, resp. We improve the result for the randomized case as follows.

Theorem A.5: *Let $N = n^{k+1}$ and $k = O(\log n)$. Let $\gamma_N, \gamma'_N > 0$ be arbitrarily chosen such that $\gamma_N, \gamma'_N = \Omega(1/\mathrm{Poly}(N))$. Then each randomized read-k-times BP for* CHSP_n^{k+1} *with*
(1) two-sided error $q^{-(k+1)} - \gamma_N$; or
(2) one-sided error $(q-1)/\big(q^{k+1}-1\big) - \gamma'_N$
has size $\exp\big(\Omega\big(N^{1/(k+1)} \cdot k^{-3} \cdot 2^{-2k}\big)\big)$.

The most difficult part of the proof of Theorem A.5 has already been done by Thathachar. He has shown that the function CHSP_n^{k+1} has low 1-density with respect to (k, a)-rectangles and the uniform distribution, where a is suitably chosen.

Lemma A.6 (Thathachar [51]): *Let $a := 144 \cdot k \cdot 2^k$, and let r be an arbitrary (k, a)-rectangle in $\{0,1\}^N$, $N = n^{k+1}$. Then*

$$\left|r^{-1}(1) \cap \left(\mathrm{CHSP}_n^{k+1}\right)^{-1}(1)\right| \cdot 2^{-N} \;\le\; \alpha \cdot \left|r^{-1}(1)\right| \cdot 2^{-N} + \delta_N,$$

where $\alpha := 1/q$ and $\delta_N := 2\gamma^{\left(6(k+1)2^{k+1}\right)^{-1} N^{1/(k+1)}}$, $\gamma := \cos(\pi/q)^{1/80} < 1$.

The key to the improvement with respect to the error bounds is the following asymptotically precise estimate of the number of 1-inputs of CHSP_n^k:

Lemma A.7: *Let $N = n^k$ and $k = 2^{o(n)}$. Then*

$$\left|\left(\mathrm{CHSP}_n^k\right)^{-1}(1)\right| \cdot 2^{-N} = q^{-k} \cdot \left(1 \pm 2^{-\Omega\left(N^{1/k}\right)}\right).$$

We prove this later on. First, we put the lemmas together to obtain the desired result.

Proof of Theorem A.5: *Part (1):* We apply the "rectangle technique" for (k, a)-rectangles, where we choose $a := 144 \cdot k \cdot 2^k$. Let G be a randomized read-k-times BP representing CHSP_n^{k+1} with two-sided error ε. Then G yields an approximation of CHSP_n^{k+1} with respect to the uniform distribution over $\{0,1\}^N$ which also has two-sided error ε and uses at most $(2|G|)^{ka}$ (k, a)-rectangles. The version of Theorem 2.2 for (k, a)-rectangles and exchanged roles of 0- and 1-inputs yields the lower bound

$$\delta_N^{-1} \cdot \left((1-\alpha) \cdot \left|\left(\mathrm{CHSP}_n^{k+1}\right)^{-1}(1)\right| - \max(1-\alpha, \alpha) \cdot \varepsilon\right)$$

on the number of (k, a)-rectangles in such an approximation. Hence,

$$|G| \;\ge\; \frac{1}{2} \cdot \left[\delta_N^{-1} \cdot \left((1-\alpha) \cdot \left|\left(\mathrm{CHSP}_n^{k+1}\right)^{-1}(1)\right| - \max(1-\alpha, \alpha) \cdot \varepsilon\right)\right]^{1/(ka)}.$$

Plugging in the results from Lemma A.6 and A.7 yields

$$|G| \;\ge\; \frac{1}{2} \cdot \left[\delta_N^{-1} \cdot \left(\left(1 - \frac{1}{q}\right) \cdot \left(q^{-(k+1)} - \vartheta_N - \varepsilon\right)\right)\right]^{1/(ka)},$$

where $\vartheta_N = 2^{-\Omega\left(N^{1/(k+1)}\right)} = 2^{-\Omega(n)}$.

Substituting $\varepsilon = q^{-(k+1)} - \gamma_N$, we get

$$|G| \;\geq\; \frac{1}{2} \cdot \delta_N^{-1/(ka)} \cdot \left(\left(1 - \frac{1}{q}\right) \cdot (\gamma_N - \vartheta_N) \right)^{1/(ka)} .$$

By assumption, we have $\gamma_N \geq 1/p(N)$ for some polynomial p and N large enough. Since $k = O(\log n)$, it follows that $N = n^{k+1} = n^{O(\log n)} = 2^{O\left(\log^2 n\right)}$, and thus $\gamma_N = 2^{-O\left(\log^2 n\right)}$. On the other hand, $\vartheta_N = 2^{-\Omega(n)}$. Hence, $\gamma_N - \vartheta_N \geq c \cdot \gamma_N$ for some constant $c > 0$ and N large enough, and

$$(\gamma_N - \vartheta_N)^{1/(ka)} \;=\; 2^{-O\left(\log^2 n \cdot k^{-2} \cdot 2^{-k}\right)}.$$

Since

$$\delta_N^{-1/(ka)} \;=\; 2^{\Omega\left(n \cdot k^{-3} \cdot 2^{-2k}\right)},$$

the lower bound for $|G|$ is of the required size.

Part (2): Let G be a randomized read-k-times BP representing CHSP_n^{k+1} with one-sided error ε. Analogously to the first part, but with the lower bound for one-sided error from Theorem 2.2, we obtain

$$\begin{aligned} |G| \;&\geq\; \frac{1}{2} \cdot \left[\delta_N^{-1} \cdot \left((1-\alpha) \cdot \left| \left(\mathrm{CHSP}_n^{k+1}\right)^{-1}(1) \right| - \alpha \cdot \varepsilon \cdot \left| \left(\mathrm{CHSP}_n^{k+1}\right)^{-1}(0) \right| \right) \right]^{1/(ka)} \\ &\geq\; \frac{1}{2} \cdot \delta_N^{-1/(ka)} \left[\left(1 - \frac{1}{q}\right) \left(q^{-(k+1)} - \vartheta_N\right) - \frac{1}{q} \cdot \varepsilon \cdot \left(1 - q^{-(k+1)} - \vartheta'_N\right) \right]^{1/(ka)} , \end{aligned}$$

where $\vartheta_N, \vartheta'_N = 2^{-\Omega\left(N^{1/(k+1)}\right)} = 2^{-\Omega(n)}$. Now we substitute $\varepsilon = (q-1)/\left(q^{k+1} - 1\right) - \gamma'_N$ and estimate the term within the brackets. We have

$$\begin{aligned} &\left(1 - \frac{1}{q}\right) \cdot q^{-(k+1)} - \frac{1}{q} \cdot \varepsilon \cdot \left(1 - q^{-(k+1)}\right) \\ &\quad=\; \left(1 - \frac{1}{q}\right) \cdot q^{-(k+1)} - \left(\left(1 - \frac{1}{q}\right) \left(q^{k+1} - 1\right)^{-1} - \frac{\gamma'_N}{q} \right) \cdot \left(1 - q^{-(k+1)}\right) \\ &\quad=\; \frac{\gamma'_N}{q} \cdot \left(1 - q^{-(k+1)}\right) \;\geq\; \frac{\gamma'_N}{q} \cdot \left(1 - q^{-2}\right) \;=\; c \cdot \gamma'_N, \end{aligned}$$

for some constant $c > 0$. Furthermore,

$$\begin{aligned} &-\left(1 - \frac{1}{q}\right) \cdot \vartheta_N + \frac{1}{q} \cdot \varepsilon \cdot \vartheta'_N \\ &\quad=\; -\left(1 - \frac{1}{q}\right) \cdot \vartheta_N + \left(\left(1 - \frac{1}{q}\right) \cdot \left(q^{k+1} - 1\right)^{-1} - \frac{\gamma'_N}{q} \right) \cdot \vartheta'_N \\ &\quad=\; 2^{-\Omega(n)}. \end{aligned}$$

Hence, the term within the brackets above is of order $c \cdot \gamma'_N - 2^{-\Omega(n)}$. Since $\gamma'_N = \Omega(1/\mathrm{Poly}(N))$, we obtain that the lower bound for $|G|$ is of the desired size analogously to the first part. □

In the remainder of the section, we prove Lemma A.7. As for the function MS_n in Section 7, we will have to count the number of solutions of equations over finite fields. We prove the following tool in advance.

Lemma A.8: *Let n be an arbitrary positive integer, q an odd prime, and $c \in \mathbb{Z}_{2q}$. Define*

$$S(n,q,c) \;:=\; \left|\left\{x \in \mathbb{Z}_2^n \;\middle|\; x_1 + \cdots + x_n \equiv c \bmod (2q)\right\}\right|.$$

Then

$$\left|S(n,q,c) - 2^n/(2q)\right| \;\le\; (1 - 1/(2q)) \cdot 2^{n/2} \cdot (1 + \cos(\pi/q))^{n/2}.$$

If q is a constant with respect to n, then $S(n,q,c) = 2^n/(2q) \cdot \left(1 \pm 2^{-\Omega(n)}\right)$.

Proof: Our aim is to apply Theorem 6.7 from the first part of the appendix. We choose $G = (\mathbb{Z}_{2q}, +)$. The character group of G consists of the functions χ_u, $u \in \mathbb{Z}_{2q}$, defined by

$$\chi_u(v) = (-1)^{uv} \cdot \omega^{uv}, \quad \text{for all } v \in \mathbb{Z}_{2q};$$

where $\omega = e^{2\pi i/q}$ and the computation of the exponents is done in $\mathbb{Z}$. Using Theorem 6.7, we obtain the estimate

$$\left|S(n,q,c) - 2^n/(2q)\right| \;\le\; 1/(2q) \cdot \sum_{u \in \mathbb{Z}_{2q},\, u \neq 0} |1 + \chi_u(1)|^n \;\le\; 1/(2q) \cdot \sum_{u \in \mathbb{Z}_{2q},\, u \neq 0} \left|1 + (-1)^u \cdot \omega^u\right|^n .$$

We have $|1 + (-1)^u \cdot \omega^u|^2 = 2\big(1 + \cos(\pi u(1 + 2/q))\big)$. The function $\cos(\pi u(1+2/q))$ attains its maximal value 1 if $u \equiv 0 \bmod (2q)$, and its minimal value -1 if $u \equiv q \bmod (2q)$. For $u \not\equiv 0 \bmod (2q)$, the maximal value is obtained by choosing u such that $u(1 + 2/q)$ is as close to an even integer as possible. Since the distance is at least $1/q$, we have $\cos(\pi u(1 + 2/q)) \le \cos(\pi/q)$ for $u \not\equiv 0 \bmod (2q)$. Substituting this into the above estimate gives the first part of the claim. For the second part, we use that $\cos(\pi/q) = 1 - (\pi/q)^2/2 + O\big((\pi/q)^4\big)$ by Taylor series expansion. □

Proof of Lemma A.7: Let X be the k-dimensional input matrix of CHSP_n^k. We start by "guessing" the results of the parity checks for all kn hyperplanes, let these be the constants $p_i^d \in \mathbb{Z}_2$, for $d \in \{1, \ldots, k\}$ and $i \in \{1, \ldots, n\}$. We have $\mathrm{CHSP}_n^k(X) = 1$ iff

$$\sum_{x \in X_i^d} x \equiv p_i^d \bmod 2, \quad \text{for all } d \in \{1, \ldots, k\} \text{ and } i \in \{1, \ldots, n\}; \tag{1}$$

and

$$\sum_{i=1}^{n} p_i^d \equiv 0 \bmod q, \quad \text{for all } d \in \{1, \ldots, k\}. \tag{2}$$

Equation (1) can also be seen as a linear system of equations for the n^k variables of X in $\mathbb{Z}_2$. We recursively define the $kn \times n^k$ coefficient matrix of this system.

First, let M_1 be the $n \times n$ identity matrix. For $k > 1$, define the $kn \times n^k$ matrix M_k as follows (empty spaces indicate zero-entries):

$$M_k := \begin{array}{c} \begin{array}{cccc} \overbrace{\hphantom{1\ 1\ \cdots\ 1}}^{n^{k-1}} & \overbrace{\hphantom{1\ 1\ \cdots\ 1}}^{n^{k-1}} & \hphantom{\ddots} & \overbrace{\hphantom{1\ 1\ \cdots\ 1}}^{n^{k-1}} \end{array} \\ \left[\begin{array}{c|c|c|c} 1\ 1\ \cdots\ 1 & & & \\ & 1\ 1\ \cdots\ 1 & \ddots & \\ & & & 1\ 1\ \cdots\ 1 \\ \hline M_{k-1} & M_{k-1} & \cdots & M_{k-1} \end{array}\right] \end{array} \begin{array}{l} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n \\ \left.\vphantom{\begin{array}{c}M\\M\end{array}}\right\} (k-1)n \end{array}$$

The matrix M_k consists of n lines containing n^{k-1} consecutive ones each in the upper part and of n copies of the $(k-1)n \times n^{k-1}$-dimensional matrix M_{k-1} in the lower part.

Let $x = (x_1, \ldots, x_{n^k})$ and $b := (p_1^1, \ldots, p_n^1, \ldots, p_1^k, \ldots, p_n^k) \in \mathbb{Z}_2^{kn}$. By these definitions, Equation (1) becomes

$$M_k \cdot x \equiv b \bmod 2. \tag{3}$$

We count the number of solutions of this system for fixed p_i^d. As the second step, we will count the number of possible choices for the p_i^d.

We prove by induction that M_k has rank $kn - (k-1)$. For $k = 1$, the claim is obviously true. Now consider the matrix M_k, $k > 1$. We assume that M_{k-1} has rank $(k-1)n - (k-2)$. For $i = 1, \ldots, n$ call the columns $(i-1)n^{k-1} + 1, \ldots, in^{k-1}$ in M_k the *ith block*. Apply the following column-transformations on M_k: Add the first block to all $n-1$ other blocks, which cancels out all copies of M_{k-1} in the lower part except in the first block and changes all zeros to ones in the first row of the blocks $2, \ldots, n$. It is easy to see that the set of column vectors in the blocks $2, \ldots, n$ obtained in this way has rank $n-1$. Furthermore, no column vector from the first block is a linear combination of columns in the blocks $2, \ldots, n$ and vice versa. Finally, the column vectors of the first block have rank $(k-1)n - (k-2)$ by assumption. Hence, M_k has rank $(k-1)n - (k-2) + (n-1) = kn - (k-1)$ altogether.

Now we apply the following row transformations to M_k in order to simplify the System (3). For $d = 1, \ldots, k-1$, add the rows $(d-1)n + 2, \ldots, (d-1)n + n$ as well as the rows $dn + 1, \ldots, dn + n$ to row $(d-1)n + 1$. In each modified row $(d-1)n + 1$, this cancels out all entries in the coefficient matrix, and on the right hand side of the equation we obtain the new constant

$$\sum_{i=1}^{n} p_i^d + \sum_{i=1}^{n} p_i^{d+1}.$$

Let $\widetilde{M}_k$ the matrix obtained from M_k by removing the rows $(d-1)n+1$, $d = 1, \ldots, k-1$. Let $\widetilde{b}$ be the right hand side obtained from b in the same way. Then we can replace system (3) by

$$\widetilde{M}_k \cdot x \equiv \widetilde{b} \bmod 2 \quad \wedge \tag{4}$$

$$\sum_{i=1}^{n} p_i^d + \sum_{i=1}^{n} p_i^{d+1} \equiv 0 \bmod 2, \quad \text{for } d = 1, \ldots, k-1. \tag{5}$$

We have proven above that $\widetilde{M}_k$ has full rank. Hence, System (3) has exactly $2^{n^k - kn + k - 1}$ solutions if (5) is fulfilled, and no solution otherwise.

It remains to count the number of the p_i^d fulfilling (2) and (5). We first notice that (5) is equivalent to

$$\sum_{i=1}^{n} p_i^1 \equiv \sum_{i=1}^{n} p_i^2 \equiv \cdots \equiv \sum_{i=1}^{n} p_i^k \bmod 2.$$

For $c \in \mathbb{Z}_2$ define

$$N_c := \left|\left\{(x_1, \ldots, x_n) \in \mathbb{Z}_2^n \;\middle|\; x_1 + \cdots + x_n \equiv c \bmod 2 \;\wedge\; \sum_{i=1}^{n} x_i \equiv 0 \bmod q\right\}\right|,$$

to count the number of possible choices for a set of constants $p_1^d, \ldots, p_n^d$ for fixed parity c. Since $\mathbb{Z}_{2q} \cong \mathbb{Z}_2 \times \mathbb{Z}_q$, we have $N_0 = S(n, 2q, 0)$ and $N_1 = S(n, 2q, q)$. By Lemma A.8,

$$N_0 = \frac{2^n}{2q} \cdot (1 + \gamma_n), \quad N_1 = \frac{2^n}{2q} \cdot \left(1 + \gamma_n'\right),$$

where $|\gamma_n|, |\gamma_n'| = 2^{-\Omega(n)}$. The total number of choices for the p_i^d fulfilling (2) and (5) is

$$N_0^k + N_1^k = \left(\frac{2^n}{2q}\right)^k \cdot \left((1 + \gamma_n)^k + (1 + \gamma_n')^k\right).$$

Since $k = 2^{o(n)}$, this is of order $2 \cdot (2^n/(2q))^k \cdot \left(1 \pm 2^{-\Omega(n)}\right)$. For each of these choices we obtain $2^{n^k - kn + k - 1}$ 1-inputs for CHSP_n^k. Hence, the total number of 1-inputs is

$$2^{n^k - kn + k - 1}\left(N_0^k + N_1^k\right) = 2^{n^k} \cdot 2^{-kn + k - 1} \cdot 2 \left(\frac{2^n}{2q}\right)^k \cdot \left(1 \pm 2^{-\Omega(n)}\right) = 2^{n^k} \cdot q^{-k} \cdot (1 + o(1)).$$

□

References

[1] F. Ablayev and M. Karpinski. On the power of randomized branching programs. In *Proc. of the 23rd Int. Coll. on Automata, Languages, and Programming (ICALP), LNCS 1099*, 348–356. Springer, 1996.

[2] M. Ajtai. Determinism versus non-determinism for linear time RAMs with memory restrictions. In *Proc. of the 31st Ann. ACM Symp. on Theory of Computing (STOC)*, 632–641, 1999.

[3] M. Ajtai. A non-linear time lower bound for Boolean branching programs. In *Proc. of the 40th IEEE Symp. on Foundations of Computer Science (FOCS)*, 60–70, 1999.

[4] M. Ajtai and M. Ben-Or. A theorem on probabilistic constant depth computations. In *Proc. of the 16th Ann. ACM Symp. on Theory of Computing (STOC)*, 471–474, 1984.

[5] E. Allender, J. Jiao, M. Mahajan, and V. Vinay. Non-commutative arithmetic circuits: depth reduction and size lower bounds. *Theoretical Computer Science*, 209:47–86, 1998.

[6] E. Allender and K. Reinhardt. Making nondeterminism unambiguous. *SIAM J. Comp.*, 29(4):1118–1131, 2000. Earlier version in *Proc. of 38th FOCS*, 244–253, 1997.

[7] A. E. Andreev, A. E. F. Clementi, J. D. P. Rolim, and L. Trevisan. Weak random sources, hitting sets, and BPP simulations. *SIAM J. Comp.*, 28(6):2103–2116, 1999.

[8] L. Babai. The Fourier transform and equations over finite abelian groups. Lecture Notes, version 1.2, Dec. 1989.

[9] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proc. of the 27th IEEE Symp. on Foundations of Computer Science (FOCS)*, 337–347, 1986.

[10] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. In *Proc. of the 30th Ann. ACM Symp. on Theory of Computing (STOC)*, 673–682, 1998.

[11] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace and time-space trade-offs. *Journal of Computer and System Sciences*, 45:204–232, 1992.

[12] P. Beame, M. Saks, X. Sun, and E. Vee. Super-linear time-space tradeoff lower bounds for randomized computation. Technical Report 25, Electr. Coll. on Comp. Compl., 2000.

[13] P. Beame, M. Saks, and J. S. Thathachar. Time-space tradeoffs for branching programs. In *Proc. of the 39th IEEE Symp. on Foundations of Computer Science (FOCS)*, 254–263, 1998.

[14] R. Beigel, H. Buhrman, and L. Fortnow. NP might not be as easy as detecting unique solutions. In *Proc. of the 30th Ann. ACM Symp. on Theory of Computing (STOC)*, 203–208, 1998.

[15] B. Bollig and I. Wegener. A very simple function that requires exponential size read-once branching programs. *Information Processing Letters*, 66:53–57, 1998.

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

[17] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comp.*, 17(2):230–261, 1988.

[18] A. Cobham. The recognition problem for the set of perfect squares. In *Proc. of the 7th Symposium on Switching an Automata Theory (SWAT)*, 78–87, 1966.

[19] A. Gál. A simple function that requires exponential size read-once branching programs. *Information Processing Letters*, 62:13–16, 1997.

[20] J. Gill. Computational complexity of probabilistic Turing machines. *SIAM J. Comp.*, 6(4):675–695, Dec. 1977.

[21] M. Goldmann. A note on the power of majority gates and modular gates. *Information Processing Letters*, 53:321–327, 1995.

[22] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley Publishing Company, Reading, MA, 1994.

[23] J. Grollman and A. L. Selman. Complexity measures for public-key cryptosystems. *SIAM J. Comp.*, 17:309–335, 1988.

[24] B. Halstenberg and R. Reischuk. On different modes of communication. *SIAM J. Comp.*, 22(5):913–934, 1993.

[25] J. Hromkovič. *Communication Complexity and Parallel Computing*. EATCS Texts in Theoretical Computer Science. Springer, Berlin, 1997.

[26] R. Impagliazzo and A. Wigderson. P $=$ BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proc. of the 29th Ann. ACM Symp. on Theory of Computing (STOC)*, 220–228, 1997.

[27] S. Jukna, A. Razborov, P. Savický, and I. Wegener. On P versus NP $\cap$ co-NP for decision trees and read-once branching programs. In *Proc. of the 22nd Int. Symp. on Mathematical Foundations of Computer Science (MFCS), LNCS 1295*, 319–326. Springer, 1997. To appear in *Computational Complexity*.

[28] S. P. Jukna. Lower bounds on communication complexity. *Mathematical Logic and Its Applications*, 5:22–30, 1987.

[29] S. P. Jukna. The effect of null-chains on the complexity of contact schemes. In *Proc. of Fundamentals of Computation Theory (FCT), LNCS 380*, 246–256. Springer, 1989.

[30] S. P. Jukna. A note on read-k times branching programs. *Theoretical Informatics and Applications*, 29(1):75–83, 1995.

[31] K.-I. Ko. Some observations on the probabilistic algorithms and NP-hard problems. *Information Processing Letters*, 14(1):39–43, Mar. 1982.

[32] M. Krause, C. Meinel, and S. Waack. Separating the eraser Turing machine classes L_e, NL_e, co-NL_e and P_e. In *Proc. of the 13th Int. Symp. on Mathematical Foundations of Computer Science (MFCS), LNCS 324*, 405–413. Springer, 1988.

[33] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 1997.

[34] S. Lang. *Algebra*. Addison-Wesley, Redwood City, 2nd edition, 1984.

[35] R. Lidl and H. Niederreiter. *Introduction to finite fields and their applications*. Cambridge University Press, Cambridge, 2nd edition, 1994.

[36] K. Mehlhorn and E. Schmidt. Las-Vegas is better than determinism in VLSI and distributed computing. In *Proc. of the 14th Ann. ACM Symp. on Theory of Computing (STOC)*, 330–337, 1982.

[37] C. Meinel. *Modified Branching Programs and Their Computational Power*. Habilitationsschrift, Humboldt-Universität Berlin, 1988. Published as *LNCS 370*, Springer.

[38] C. Meinel. Polynomial size Ω-branching programs and their computational power. *Information and Computation*, 85:163–182, 1990.

[39] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge, 1995.

[40] N. Nisan. On read-once vs. multiple access to randomness in logspace. *Theoretical Computer Science*, 107:135–144, 1993.

[41] E. A. Okol'nishnikova. On lower bounds for branching programs. *Siberian Advances in Mathematics*, 3(1):152–166, 1993.

[42] P. Pudlák and S. Zák. Space complexity of computations. Technical report, Univ. Prague, 1983.

[43] A. A. Razborov. Bounded-depth formulas over the basis $\{\&, \oplus\}$ and some combinatorial problems. In S. I. Adian, editor, *Problems of Cybernetics, Complexity Theory and Applied Mathematical Logic*, 149–166. VINITI, Moscow, 1988. In Russian.

[44] A. A. Razborov. Lower bounds for deterministic and nondeterministic branching programs. In *Proc. of Fundamentals of Computation Theory (FCT), LNCS 529*, 47–60. Springer, 1991.

[45] A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

[46] M. Sauerhoff. A lower bound for randomized read-k-times branching programs. Technical Report 19, Electr. Coll. on Comp. Compl., 1997.

[47] M. Sauerhoff. Lower bounds for randomized read-k-times branching programs. In *Proc. of the 15th Ann. Symp. on Theoretical Aspects of Computer Science (STACS), LNCS 1373*, 105–115. Springer, 1998.

[48] M. Sauerhoff. *Complexity Theoretical Results for Randomized Branching Programs*. PhD thesis, Univ. of Dortmund. Shaker, Aachen, 1999.

[49] M. Sauerhoff. On the size of randomized OBDDs and read-once branching programs for k-stable functions. In *Proc. of the 16th Ann. Symp. on Theoretical Aspects of Computer Science (STACS), LNCS 1563*, 488–499. Springer, 1999.

[50] J. Simon and M. Szegedy. A new lower bound theorem for read-only-once branching programs and its applications. In J.-J. Cai, editor, *Advances in Computational Complexity Theory, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 13*, 183–193. American Mathematical Society, 1993.

[51] J. Thathachar. On separating the read-k-times branching program hierarchy. In *Proc. of the 30th Ann. ACM Symp. on Theory of Computing (STOC)*, 653–662, 1998.

[52] L. G. Valiant. Relative complexity of checking and evaluating. *Information Processing Letters*, 5:20–23, 1976.

[53] L. G. Valiant and V. V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47:85–93, 1986.

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

[55] I. Wegener. On the complexity of branching programs and decision trees for clique functions. *Journal of the ACM*, 35(2):461–471, Apr. 1988.

[56] I. Wegener. *Branching Programs and Binary Decision Diagrams—Theory and Applications*. Monographs on Discrete and Applied Mathematics. SIAM, Philadelphia, PA, 2000.

[57] A. Wigderson. De-randomizing BPP: The state of the art. In *Proc. of the 14th IEEE Int. Conf. on Computational Complexity*, 1999.

[58] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.

[59] A. C. Yao. Some complexity questions related to distributive computing. In *Proc. of the 11th Ann. ACM Symp. on Theory of Computing (STOC)*, 209–213, 1979.

[60] A. C. Yao. Lower bounds by probabilistic arguments. In *Proc. of the 24th IEEE Symp. on Foundations of Computer Science (FOCS)*, 420–428, 1983.

[61] S. Žák. An exponential lower bound for one-time-only branching programs. In *Proc. of the 11th Int. Symp. on Mathematical Foundations of Computer Science (MFCS), LNCS 176*, 562–566. Springer, 1984.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'