THEORY OF COMPUTING, Volume 1 (2005), pp. 149–176
http://theoryofcomputing.org

A Non-linear Time Lower Bound for Boolean Branching Programs

Miklós Ajtai

Received: October 6, 2004; revised: May 5, 2005; published: October 5, 2005.

Abstract: We give an exponential lower bound for the size of any linear-time Boolean branching program computing an explicitly given function. More precisely, we prove that for all positive integers k and for all sufficiently small $\varepsilon > 0$, if n is sufficiently large then there is no Boolean (or 2-way) branching program of size less than $2^{\varepsilon n}$ which, for all inputs $X \subseteq \{0,1,\ldots,n-1\}$, computes in time kn the parity of the number of elements of the set of all pairs $\langle x,y\rangle$ with the property $x \in X$, $y \in X$, $x < y$, $x+y \in X$.

For the proof of this fact we show that if $A = (a_{i,j})_{i=0,j=0}^{n}$ is a random n by n matrix over the field with 2 elements with the condition that "A is constant on each minor diagonal," then with high probability the rank of each δn by δn submatrix of A is at least $c\delta|\log\delta|^{-2}n$, where $c > 0$ is an absolute constant and n is sufficiently large with respect to δ.

(A preliminary version of this paper has appeared in the Proceedings of the 40th IEEE Symposium on Foundations of Computer Science.)

ACM Classification: F.2.2, F.2.3

AMS Classification: 68Q17, 68Q15

Key words and phrases: complexity theory, lower bounds, space complexity, branching programs, Hankel matrices, matrix rigidity

Authors retain copyright to their work and grant Theory of Computing unlimited rights to publish the work electronically and in hard copy. Use of the work is permitted as long as the author(s) and the journal are properly acknowledged. For the detailed copyright statement, see http://theoryofcomputing.org/copyright.html.

© 2005 Miklós Ajtai

DOI: 10.4086/toc.2005.v001a008

1 Introduction

1.1 The statement of the main result

A Boolean (or 2-way) branching program is a finite directed acyclic graph (which may contain parallel edges) with a unique source node, so that each non-sink node is labeled by one of the input variables $x_0,\dots,x_{n-1}$, each non-sink node has outdegree two, each edge is labeled by an element of $\{0,1\}$ so that the two outgoing edges of a non-sink node always get different labels, and each sink-node is labeled by an element of $\{0,1\}$. If an input is given we start from the unique source node and go along a path according to the following rule. If we are at node v and the label of v is the variable x_i then we leave v on the unique outgoing edge whose label is the value of x_i. This path will end in a sink node; the label of the sink-node is the output of the program at the given input, the length of the path is the computational time at the given output, the maximal length of a path in the graph that we may get from an input this way is the length (or depth) of the branching program. The number of nodes in the graph is the size of the branching program.

This model describes a very general way of computing where the computational time measures the number of accesses to the individual bits of the input and the size measures the number of different states of the machine performing the computations. We do not measure the computational time needed to determine the next state of our machine (that is, the next node in the graph along the path). We may also think about this model as a random access machine whose input registers contain a single input bit, with a working memory containing $\log_2 M$ bits where M is the size of the branching program.

Our goal is to give an explicit function which cannot be computed with a Boolean branching program in linear time if the size of the branching program is $2^{\varepsilon n}$. The function has to assign to each $\{0,1\}$-sequence of length n a single $\{0,1\}$-value. We identify the set of all $\{0,1\}$-sequences of length n with the set of all subsets of $\{0,1,\dots,n-1\}$, that is, our function f will assign to each subset X of $\{0,1,\dots,n-1\}$ a $\{0,1\}$-value. For such a set X let $f(X)$ be the parity of the number of elements of the set of all pairs $\langle x,y\rangle$ with the property $x\in X$, $y\in Y$, $x<y$, and $x+y\in X$. We will say that $f(X)$ is the *parity of interior sums* for the set X, where the expression "interior" refers to the fact that both the terms in the sum and the sum itself must be in X.

Our main result is that the parity of interior sums for a set of n elements cannot be computed with a Boolean branching program in linear time if the size of the branching program is $2^{\varepsilon n}$ (see Theorem 3.4).

1.2 The history of the problem and related results

1.2.1 Boolean and R-way branching programs

One of the main goals of complexity theory is to describe explicitly given functions which cannot be computed in certain computational models with specified amount of resources. Branching programs form one of the most general models of computation, e.g. random access machines with memory M and running time t can be described by branching programs of size 2^M and time (depth) t. Because of the generality and simplicity of the model and its mentioned close connection to the practical random access models of computation, finding lower bounds for explicitly given functions in the branching program model in general, and with the given values of parameters in particular (linear time and sublinear memory) was always considered a question of great importance.

A generalization of the Boolean branching programs is the computational model of R-way branching programs. Since most of the results that we use in our proofs were established for R-way branching programs, we describe briefly its definition. The computation is done in the same way, along a path of a directed graph as in the Boolean case, with the following differences. A set Γ with R elements is given, the elements of Γ are the possible values of the input variables. Each non-sink node has outdegree R, and the outgoing edges are labeled by the elements Γ so that the R outgoing edges have R different labels. Each sink node is labeled by an element of Γ. If an input is given, that is, we assign an element of Γ to each of the input variables $x_1,\ldots,x_n$, then we follow a path of the graph in the following way. We start at the unique source node. If we are at node v and the label of node v is the variable x_i then we leave v on the unique outgoing edge whose label is the element of Γ assigned to the variable x_i. This path will end in a sink node. The element of Γ which is the label of this sink node is the output of the R-way branching program. The value of R, in the results most interesting for us, is n^c where c is a constant. (This corresponds to the random access machines where each register can contain $c\log_2 n$ bits.)

1.2.2 Branching programs with many output bits, and the time segmentation method

The computational model of R-way branching programs was introduced by Borodin and Cook [8], who proved a time-space trade-off for sorting n integers. This work also introduced a method for proving lower bounds about R-way branching programs in the special case where the number of output bits is relatively large compared to the time allowed for the computation. Several other lower bounds and time-space trade-offs of similar nature were given, see e. g. Abrahamson [1, 2], Beame [5], Karchmer [11], Reisch and Schnitger [12], and Yesha [14]. These lower bound proofs have a common high-level structure, namely the time is cut into short intervals and we use the fact that during such an interval any information that we can use about the past must be contained in the limited memory at the beginning of the interval. In particular if many output bits are provided in a single time interval, then these may depend only on those input values which are accessed during this time interval and the the content of the memory at the the beginning of the interval.

1.2.3 Lower bounds for explicit functions and decision problems

Using the same high level proof structure, and other new ideas, Beame, Saks and Jayram[1] [6] gave a lower bound on the computational time for an explicitly given function with a Boolean branching program of size $2^{o(n)}$. Namely they proved that there is an $\varepsilon > 0$ so that the question whether the quadratic form $\sigma^T Q\sigma$ is zero, (where σ is the input, a $\{0,1\}$-vector of length n, and Q is the $n\times n$ Sylvester matrix over the field with three elements) cannot be decided with a branching program of length $(1+\varepsilon)n$ and of size $2^{o(n)}$. (The proof shows that the theorem holds for $\varepsilon = .0178$.) This is the best previously known lower bound in the direction of our main result. In the same paper they gave a nonlinear lower bound on the length of an R-way branching program computing an explicitly defined function, (similar to the function they used in the Boolean case.) More precisely they prove that for all k there is an r_k so that for all sufficiently large n there is an (explicitly given) 0-1 valued function $g(x_1,\ldots,x_n)$ of n variables such that: (a) each variable is takes its values from a set of size r_k and (b) there is no r_k-way and size n^c branching program which computes $g(x_1,\ldots,x_n)$ in depth kn.

[1] T.S. Jayram, formerly Jayram S. Thathachar

The author of the present paper proved [4] that the element distinctness problem (where each "element" is the value of a variable) cannot be decided with an R-way branching program, for $R = c\log_2 n$, in length linear in n if the size of the program is at most $2^{\varepsilon n}$, provided that $c \geq 2$. (If the problem is to find two elements whose Hamming distance is smaller than $\frac{1}{4}c\log_2 n$ then for a similar lower bound on the length the necessary restriction on the size is only $2^{\varepsilon n \log_2 n}$.) These proofs are based on the analysis of certain combinatorial properties of the input, which are very similar to the combinatorial properties used in [6]. The high level structure of the proof still follows the time segmentation idea described earlier. Since the present proof uses some of the technical lemmata of [4], we will give later a more detailed description of its techniques. At this point we sketch only some of the basic ideas of the proof in [4]. Since these are also related to the proof methods of [6], this will show the additional ideas that are needed to make the time segmentation method work when the number of output bits is small.

1.2.4 Lower bounds for binary functions and relations

It is shown in [4] that if a function f can be computed in linear time with the given restrictions on the size then there are two large disjoint subsets, W_1, W_2, of the set of the input variables and an input χ with the following properties. For each $i = 1,2$ we may change the input χ in many different ways by changing the values of the variables in W_i only, so that the output does not change. Moreover, for $i = 1,2$ we can select a large set of changes Y_i so that even if we perform a change from Y_1 (on the values of the variables in W_1) and another one from Y_2 (on the values of the variables in W_2) simultaneously, the output remains unchanged.

In the case of the element distinctness problem we are able to chose an input χ which meets these requirements with the additional property that χ (which is a list of n integers) consists of pairwise distinct integers. Therefore, if our branching program solves the element distinctness program, its output is, say, 1. However we can prove the for a fixed $i = 1,2$ the inputs that we get from χ through the changes in Y_i, take more than $\frac{n}{2}$ different values on the set of variables W_i. Therefore there will be an integer x so that for both $i = 1,2$, x will be a value of a variable from W_i if we perform suitable change on χ from Y_i. Consequently performing both suitable changes simultaneously we get an input which contains x twice and still the unchanged output is 1. This contradicts the assumption that the program solves the element distinctness problem.

Similar ideas are used for the other relations, or functions in the mentioned lower bounds. In each case we need a function $F(x,y)$ or a relation $R(x,y)$ with two variables so that if we can separately change x and y in many different ways then among these changes there will be two so that performing them simultaneously we are able to change the value of $F(x,y)$ or $R(x,y)$. If R is the equality predicate then, as we have seen, this can be guaranteed if both sets of changes produces at least $\frac{n}{2}$ different input values. In the case of the Hamming distance problem described above the situation is even better since $\frac{n}{2}$ can be replaced by $n^{1-\varepsilon'}$ for some small constant $\varepsilon' > 0$ (see [4]). (This is the reason that the proof of the lower bound for the Hamming distance problem is much simpler and gives a stronger result than the proof for the element distinctness problem.)

1.2.5 Quadratic forms and rigidity

The binary function $F(x,y)$ can be also defined by a quadratic form $x^T By$. Assume that, as before, x and y independently run over large sets and now we want to guarantee that the $x^T By$ is not constant. Motivated by similar considerations, quadratic forms were studied by Borodin, Razborov, and Smolensky [9], Jayram [13], and Beame, Saks, and Thatachar [6]. The result in this direction that we will use in our paper is the following. (This was proved in more general forms in [9], [13], and [6], and also follows from the results of [10].) Suppose that the rank of the matrix B is r and x resp. y are taking values independently from m_1 resp. m_2 dimensional subspaces of an m dimensional vectorspace. If $m_1 + m_2 + r > 2m$ then the quadratic form $x^T By$ is not constant.

As we described earlier, in the lower bound proofs we can usually guarantee only that x and y can take values independently only in some limited sense, namely we can apply independent changes to two disjoint sets of variables. In [6] the mentioned property of the quadratic forms is applied in the following way. The lower bound is proved for a function of the form $x^T Ay$ where A is a suitably chosen, explicitly given, matrix over a finite field. This matrix has the property that each $\delta n \times \delta n$ submatrix which does not contain elements from the main diagonal is of rank at least $\delta^2 n$ where $\delta > 0$ is a small constant. (This may be considered as a rigidity property of the matrix A.) The input variables of the branching program take values from the field F. We pick two large sets of independent changes on two disjoint sets of variables of at least δn elements. The submatrix of A formed from the corresponding rows and columns will be the matrix B in the mentioned property of quadratic forms. This way it is possible to guarantee that, roughly speaking, under independent changes on these sets of variables, the quadratic form cannot remain constant, which makes the lower bound proof possible.

In the present paper we will follow the same strategy for our proof with the following improvements. We use two-way branching programs with a single output bit which creates three new problems. (1) It is more difficult to prove the existence of the two disjoint sets of variables which admit many independent changes that leave the output of the branching program unchanged. For this we use the machinery worked out in [4]. (2) The explicitly given matrix of [6] is a Sylvester matrix over the field F and so the size of the field must be at least n. We need something similar for F_2, the field with two elements. We do not give an explicit construction for such a matrix, but a random construction which depends only on a linear number of random bits which can be included in the input. (3) it is not enough for us if the rank of the submatrices of sizes $\delta n \times \delta n$ are $\delta^2 n$, we need much larger ranks; what we prove will be $\delta |\log \delta|^{-2} n$.

1.2.6 Summary of the history of the problem

Summarizing the historical developments about the lower bound techniques for branching programs, we can say that there were two parallel developments. The first is the time segmentation method which later was supplemented by the technique of considering changes of the values of variables on two disjoint sets: [8],[6],[4]. The second is the development of the algebraic techniques about quadratic forms based on matrices with rigidity properties, providing explicitly defined functions which were suitable for the lower bound proof techniques mentioned in the first direction of developments: [9],[13],[6]. The present paper uses the techniques of both of these directions.

1.3 Subsequent developments

A preliminary version of this paper was published in [3] containing all of the essential elements of the proofs presented here. Since then, the main result of this paper was further improved by Beame, Saks, Sun, and Vee in [7] by making the time/space lower bounds sharper and generalizing the theorem for the case of probabilistic branching programs. Their proofs use the results and techniques of the present paper (together with methods of different nature).

2 Overview of the proof

2.1 A lower bound for a nonexplicit function

Our proof in the present paper uses a technical lemma of the element distinctness result. As we have mentioned already in the introduction, it is shown in [4] that if a function f can be computed in linear time with the given restrictions on the size then there are two large disjoint subsets, W_1, W_2, of the set of the input variables and an input χ so that for each $i = 1, 2$ we may change the input χ in many different ways by changing only the values of the variables in W_i so that the output does not change; moreover these changes can be performed simultaneously on W_1 and W_2 so that the output still does not change. The ratio between the sizes of the sets W_i and the logarithm of the number of changes has a crucial importance in the proofs of the present paper. (A precise statement of this result is given in Lemma 3.5 below.)

We use this result to show that a quadratic form (which is *not* given explicitly) cannot be computed in linear time. The algebraic part of this proof (Lemma 3.11) is a theorem proved by Borodin, Razborov, and Smolensky [9] (and in more general forms by Jayram [13] and Beame, Saks, and Jayram [6]). We reduce the problem of giving a quadratic form with the required properties to a question about the ranks of the submatrices (or minors) of the matrix generating the quadratic form in a similar way as is done in [6]. In both cases the goal is to get a matrix A so that each $\lfloor \delta n \rfloor$ by $\lfloor \delta n \rfloor$ submatrix of the matrix A has rank at least $\psi(\delta)n$, for each $\delta > 0$, provided that n is sufficiently large with respect to δ, where the function ψ should be as large as possible. The Sylvester matrices used in [6] are explicitly given examples of such matrices with $\psi(\delta) = \delta^2$, provided that we consider only submatrices that do not contain any elements of the main diagonal. (This restriction does not affect the applicability of the matrix to the lower bound proof.)

2.2 Decreasing the randomness needed

Definition 2.1. We will call an $n \times n$ matrix $A = (a_{i,j})$ a *Hankel* matrix if $\forall i, j, k, l \in \{0, 1, \ldots, n-1\}, i + j = k + l$ implies $a_{i,j} = a_{k,l}$. In other words A is a Hankel matrix iff it is constant across minor diagonals.

Remark 2.2. A Hankel matrix is determined by only $2n - 1$ suitably chosen entries, e.g. by entries of the first row and last column. If a matrix is constant along all diagonals it is called a *Toeplitz* matrix. Reversing the ordering of the rows creates a one-to-one correspondence between Toeplitz matrices and Hankel matrices. Therefore all of our results concerning the ranks of Hankel matrices remain valid for Toeplitz matrices as well.

We show that if A is a random n by n Hankel matrix over the field with 2 elements, with uniform distribution on the set of all such matrices, then with high probability the described property about the ranks of the submatrices holds with $\psi(\delta) = c\delta|\log\delta|^{-2}$ for an absolute constant $c > 0$. As a consequence, using also the mentioned lemma from [4], we are able to show that if $\tilde{A}$ is the matrix that we get from A by replacing each entry in the main diagonal and above by 0, then the quadratic form $\langle\tilde{A}x,x\rangle$, where x is the input vector, cannot be computed with a branching program of linear length and size at most $2^{\varepsilon n}$.

2.3 From a non-explicit function to an explicit function

Of course this is not an explicitly given function; we only know that the lower bound holds for almost all matrices. However, we got the matrix by randomizing only $2n-1$ bits. Therefore if we include these bits in the input, then we get an explicitly given problem (with $3n-1$ input variables, where the described tradeoff holds between the length and size of any branching program computing the quadratic form). In other words, if $A(y)$, $y = \langle y_0,\ldots,y_{2n-2}\rangle$ denotes the Hankel matrix with $a_{i,j} = y_{i+j}$, then $\langle\tilde{A}(y)x,x\rangle$ cannot be computed in the given length and size from the input $\langle x,y\rangle$.

2.4 Obtaining a lower bound for the parity of interior sums problem

Assume now that $A = (a_{i,j})$ is a fixed Hankel matrix so that $\langle\tilde{A}x,x\rangle$ cannot be computed with a branching program with the given restrictions. Suppose that $x = \langle x_0,\ldots,x_{n-1}\rangle$ and $X = \{i \mid x_i = 1\}$, and

$$D = \{i+j \mid a_{i,j} = 1, i,j \in \{0,\ldots,n-1\}\} \ .$$

It is easy to see that $\langle\tilde{A}x,x\rangle$ is the parity of the number of all pairs $\langle i,j\rangle$, $i \in X$, $j \in X$ with the property $i < j$ and $i+j \in D$.

This will already imply that if two subsets, X,Y, of the set $\{1,2,...,2n\}$ are given, then the problem of computing the parity of the number of elements of the set of all pairs $\langle i,j\rangle$ with the property $i \in X$, $j \in X$, $i < j$, $i+j \in Y$ cannot be solved by a branching program of linear length and of size at most $2^{\varepsilon n}$. (The set D defined above will play the role of Y.) It will not be difficult to make a single set from the two sets X,Y, by taking into account the sizes of their elements, and so we will get that the task of computing, given as input an $X \subseteq \{1,2...,n\}$, the parity of the number of elements of the set of all pairs $\langle i,j\rangle$ with the property $i \in X$, $j \in X$, $i < j$, $i+j \in X$ cannot be accomplished by a branching program of linear length and of size at most $2^{\varepsilon n}$.

Finally we note that our results about random Hankel matrices remain true over any field with appropriate modifications. (See the remarks after Lemma 4.5, Lemma 4.7 and Lemma 4.9. See also the comment about the applicability of these modified versions to generalizations of Theorem 3.4 in the proof of Lemma 3.11.)

3 The reduction of the lower bound to a problem about Hankel matrices

3.1 The statement of Theorem 3.4

In this section we reduce the problem of giving a lower bound for the time needed to solve the problem described in the introduction to the existence of a matrix A which can be constructed from n bits with the property that each large submatrix of A has also relatively large rank.

Definition 3.1. If X, Y are sets then $\mathrm{Func}(X,Y)$ will denote the set of all functions, defined on X, taking values in Y.

A branching program as we will define below will be what is usually called a (deterministic) Boolean or 2-way branching program indicating that the input variables take their values from a set of size 2.

Definition 3.2. A *branching program* $\mathcal{B}$ with n input variables $x_0,\ldots,x_{n-1}$ is a five tuple

$$\langle \mathcal{G}, \mathtt{start}, \mathtt{sink}, \mathtt{var}, \mathtt{val} \rangle \ ,$$

with the following properties

(a). $\mathcal{G}$ is a finite directed acyclic graph, which may contain parallel edges

(b). `start` is the unique source node of $\mathcal{G}$,

(c). `var` is a function defined on the non-sink nodes of $\mathcal{G}$ with values in the set $\{x_0,\ldots,x_{n-1}\}$ of variables,

(d). `out` is a function defined on the set of sink nodes of $\mathcal{G}$ with values in $\{0,1\}$,

(e). `val` is a function defined on the set of edges with values in $\{0,1\}$,

(f). each non-sink node has out-degree 2, and the function `val` takes different values on the two outgoing edges.

An input for the branching program $\mathcal{B}$ is a $\{0,1\}$-assignment of the variables x_i. (Instead of such an assignment we usually will think about an input as a $\{0,1\}$-valued function η defined on $\{0,1,\ldots,n-1\}$ where $\eta(i)$ is the value of x_i.) If an input is given, then starting from `start` we go along a path in the graph in the following way. When we are at a non-sink node v then we look at the value of the variable $\mathtt{var}(v)$ and leave the node along the edge e where the value of $\mathtt{val}(e)$ is the same as the value of this variable. Since the graph is acyclic and finite, this way we will reach a sink-node w. $\mathtt{out}(w)$ will be the output of the branching program at the given input. The number of edges along the path determined this way by the input is the computational time of the branching program at the given input. The maximal computational time for the set of all inputs (that is, the maximal length of all paths arising from an input in the given way) is the *length* of the branching program. The *size* of the branching program is the number of nodes of $\mathcal{G}$.

Definition 3.3. Assume that X is a subset of $\{0,\ldots,n-1\}$. $N_+(X)$ will denote the number of all pairs $x,y \in X$, $x<y$ so that $x+y \in X$.

The following theorem is the main result of the present paper. It states that the parity of the interior sums of a subset of $0,1,\ldots,n-1$ cannot be determined by a branching program of size $2^{\varepsilon n}$ in linear time.

Theorem 3.4. *For all positive integers k, if $\varepsilon > 0$ is sufficiently small and n is sufficiently large then there is no branching program $\mathcal{B}$ with n inputs, of length at most kn and of size at most $2^{\varepsilon n}$, which for all inputs η computes the parity of $N_+(X_\eta)$ where $X_\eta = \{i \in \{0,1,\ldots,n-1\} \mid \eta(i) = 1\}$*

3.2 Results from earlier works

In the proof we will use the following lemma, Lemma 3.5, which is a consequence of Lemma A1 proved in [4] (called Lemma 9 in that paper). The proof of Lemma 3.5 from Lemma A1 is almost identical to the proof of Theorem 4 from Lemma 9 in [4] and does not require any new ideas. We describe this proof (of Lemma 3.5 from Lemma A1) in the last section.

The remaining part of the paper, starting with Lemma 3.7, is self-contained. We begin with the definitions needed to understand the statement of Lemma 3.5.

Definitions.

1. An *input* (of a branching problem with n input variables) is a function χ defined on $\{0,1,\ldots,n-1\}$ with values in $\{0,1\}$. A *partial input* is a function η defined on a subset of $\{0,1,\ldots,n-1\}$ with values in $\{0,1\}$.
2. Assume that χ is an input and η is a partial input. Then $\chi \wr \eta$ will denote the input which is identical to η on $\mathtt{domain}(\eta)$ and identical to χ on $\mathtt{domain}(\chi) \backslash \mathtt{domain}(\eta)$.
3. If $\delta \in \{0,1\}$ and $\mathcal{B}$ is a branching program, then $\mathcal{B}^{-1}(\delta)$ will denote the set of all inputs η so that the output of $\mathcal{B}$ at input η is δ.

Lemma 3.5. *For all positive integers k, if $\sigma_1 > 0$ is sufficiently small with respect to k, $\sigma_2 > 0$ is sufficiently small with respect to σ_1, $\varepsilon > 0$ is sufficiently small with respect to σ_2, n is sufficiently large with respect to ε, $\mathcal{B}$ is a branching program with n inputs of length at most kn and of size at most $2^{\varepsilon n}$, and $\delta \in \{0,1\}$ so that $|\mathcal{B}^{-1}(\delta)| \geq 2^{n-1}$, then there exist a $\chi \in \mathcal{B}^{-1}(\delta)$, $\lambda \in (\sigma_2, \sigma_1)$, $\mu \in (\sigma_2, \sigma_1)$, $W_i \subseteq \{0,1,\ldots,n-1\}$, $i = 1,2$, and sets of partial inputs Y_i, $i = 1,2$ defined on W_i satisfying the following conditions:*

(1). for all $i \in W_1$ and $j \in W_2$ we have $i < j$,

(2). $|W_1| = |W_2| = \mu n$,

(3). $|Y_1|, |Y_2| \geq 2^{\mu n - \lambda n}$,

(4). $\mu^{1+\frac{1}{100k}} \geq 2\lambda$, and

(5). for all $\eta_1 \in Y_1$, $\eta_2 \in Y_2$, we have $(\chi \wr \eta_1) \wr \eta_2 \in \mathcal{B}^{-1}(\delta)$.

3.3 Branching programs and matrix rigidity

Definition 3.6. Assume that A is an n by n matrix over the field F and f is a real-valued function defined on $(0,1]$. We say that the matrix A is *f-rigid* if for each $q = 1,\ldots,n$ and for each q by q submatrix B of A we have that the rank of B is at least $f(\frac{q}{n})n$.

The proof of Theorem 3.4 is based on Lemma 3.7 and Lemma 3.9 described below.

Lemma 3.7. *There is a $\delta > 0$ so that, for all $\gamma > 0$, if the function $g(x)$ is defined by $g(x) = \delta x|\log x|^{-2}$ if $x \in (\gamma, \frac{1}{2})$ and $g(x) = 0$ otherwise, then for each sufficiently large positive integer n there is an n by n Hankel matrix A over F_2, so that A is g-rigid.*

Remark 3.8. It would be much easier to prove the lemma with $g(x) = \delta^2 x$, though this is not enough for the present application.

We will prove Lemma 3.7 in the next section; more precisely, we will prove (Theorem 4.2) that a random matrix A taken with uniform distribution on the set of all Hankel matrices meets the requirements of the Lemma with high probability.

Definitions.

1. Assume that η is a function with values in $\{0,1\}$ defined on $\{0,1,\ldots,n-1\}$. u_η will denote the n-dimensional vector $\langle \eta(0),\ldots,\eta(n-1)\rangle$
2. The inner product of the n-dimensional vectors u, v will be denoted by $u \cdot v$.
3. Assume that $A = \{a_{i,j}\}_{i=0,j=0}^{n-1}$ is an n by n matrix. $\tilde{A}$ will denote the n by n matrix that we get from A by keeping every entry of A below the main diagonal and replacing all other entries by 0. In other words $\tilde{A} = \{b_{i,j}\}_{i=0,j=0}^{n-1}$, where $b_{i,j} = a_{i,j}$ for all $i > j$ and $b_{i,j} = 0$ for all $i \leq j$, $i = 1,\ldots,n$, $j = 1,\ldots,n$.

Lemma 3.9. *For all positive integers k, if $\sigma_1 > 0$ is sufficiently small with respect to k, $\sigma_2 > 0$ is sufficiently small with respect to σ_1, $\varepsilon > 0$ is sufficiently small with respect to σ_2, and n is sufficiently large with respect to ε, then the following holds. Assume that the function f is defined on $(0,1]$ by $f(x) = x^{1+\frac{1}{100k}}$ if $x \in (\sigma_1, \sigma_2)$ and $f(x) = 0$ otherwise. If A is an f-rigid n by n matrix A over F_2 then there is no branching program $\mathcal{B}$ with n inputs of length at most kn and of size at most $2^{\varepsilon n}$ which, for all inputs η, computes $\tilde{A}u_\eta \cdot u_\eta$.*

Remark 3.10. We use the matrix $\tilde{A}$ instead of A in the expression $\tilde{A}u_\eta \cdot u_\eta$ at the conclusion of the lemma, since over a field of characteristic 2 and for a symmetric matrix A, almost all of the terms of $Au_\eta \cdot u_\eta$ will have 0 coefficients.

Proof of Lemma 3.9. Assume that, contrary to our statement, there is a branching program $\mathcal{B}$ with the given properties which computes $\tilde{A}u_\eta \cdot u_\eta$. We apply Lemma 3.5 with the given values of $k, \sigma_1.\sigma_2, \varepsilon, n$ and with the given $\mathcal{B}$. According to Lemma 3.5 there exist $\delta \in \{0,1\}$, $\chi \in \mathcal{B}^{-1}(\delta)$, $\lambda, \mu \in (\sigma_2, \sigma_1)$, and W_i, Y_i, $i = 1,2$ with the properties listed in Lemma 3.5. Let $v = \langle v_0,\ldots,v_{n-1}\rangle$ be an n dimensional vector over F_2 defined in the following way. For all $i \notin W_1 \cup W_2$ let $v_i = \chi(i)$ and for all $i \in W_1 \cup W_2$

let $v_i = 0$. Recall that for $i = 1,2$, Y_i is a set of functions from W_i to $\{0,1\}$. We define a vector $w^{(\xi)} = \langle w_0^{(\xi)}, \ldots, w_{n-1}^{(\xi)} \rangle$ for all ξ in $Y_1 \cup Y_2$. If $i \in \mathtt{domain}(\xi)$ then $w_i^{(\xi)} = \xi(i)$, if $i \notin \mathtt{domain}(\xi)$ then $w_i^{(\xi)} = 0$. Let g_i be the following function defined on Y_i: for all $\xi \in Y_i$, $g_i(\xi) = \tilde{A}(v + w^{(\xi)}) \cdot (v + w^{(\xi)})$. Since the functions g_i take at most two different values there are $Y_i' \subseteq Y_i$ so that $|Y_i'| \geq \frac{1}{2}|Y_i|$ and g_i is constant on Y_i' for $i = 1,2$. Assume now that $\xi_1 \in Y_1'$, $\xi_2 \in Y_2'$ and let $\eta = (\chi \wr \xi_1) \wr \xi_2$. By Lemma 3.5, $\eta \in H$ and therefore

$$\begin{aligned} \tilde{A}u_\chi \cdot u_\chi = \tilde{A}u_\eta \cdot u_\eta &= \tilde{A}(v + w^{(\xi_1)} + w^{(\xi_2)}) \cdot (v + w^{(\xi_1)} + w^{(\xi_2)}) \\ &= -\tilde{A}v \cdot v + g_1(\xi_1) + g_2(\xi_2) + \tilde{A}w^{(\xi_1)} \cdot w^{(\xi_2)} + \tilde{A}w^{(\xi_2)} \cdot w^{(\xi_1)} \quad . \end{aligned}$$

$\tilde{A}u_\chi \cdot u_\chi$ and $\tilde{A}v \cdot v$ do not depend on the choices of ξ_1, ξ_2. By the definition of Y_1' and Y_2', $g_1(\xi_1) + g_2(\xi_2)$ is constant on $Y_1' \times Y_2'$. These facts imply that $\tilde{A}w^{(\xi_1)} \cdot w^{(\xi_2)} + \tilde{A}w^{(\xi_2)} \cdot w^{(\xi_1)}$, as a function of ξ_1, ξ_2, is also constant on $Y_1' \times Y_2'$. Condition (1) and the definition of $\tilde{A}$ implies that $\tilde{A}w^{(\xi_2)} \cdot w^{(\xi_1)}$ is identically 0 on $Y_1' \times Y_2'$; therefore $\tilde{A}w^{(\xi_1)} \cdot w^{(\xi_2)}$ is constant on $Y_1' \times Y_2'$. Let V_0 be the vectorspace all F_2-valued functions defined on $\{0, \ldots, n-1\}$, and let V_i, $i = 1,2$, be the subspace of functions that vanish outside W_i. The dimension of V_i is μn. We may assume that $Y_i, Y_i' \subseteq V_i$. Let ι_1 be the natural embedding of V_1 into V_0 and let π_2 be the orthogonal projection of V_0 onto V_2. B will be the linear map of V_1 into V_2 defined by $Bx = \pi_2 \tilde{A} \iota_1 x$. For all $\xi_1 \in V_1$, $\xi_2 \in V_2$ we have $\tilde{A}w^{(\xi_1)} \cdot w^{(\xi_2)} = B\xi_1 \cdot \xi_2$. If we fix the bases in both V_1 and V_2 which consist of those functions which take the value 1 at exactly one point and 0 everywhere else, then the matrix of B is a submatrix of $\tilde{A}$ consisting of those entries whose column numbers are in W_1 and row numbers are in W_2. By Condition (1) this submatrix of $\tilde{A}$ is identical to the corresponding submatrix of A. Therefore by the f-rigidity of A, the rank of B is at least $\mu^{1+\frac{1}{100k}} n$. We apply Lemma 3.11 (below) with $V_1, V_2, m \to \mu n, X \to Y_1', Y \to Y_2'$ and B. Condition (3) implies that

$$|Y_i'| \geq \frac{1}{2}|Y_i| \geq 2^{\mu n - \lambda n - 1} \quad .$$

Therefore, according to Lemma 3.11, the fact that $Bx \cdot y$ is constant on $\gamma_1(Y_1) \times \gamma_2(Y_2)$ implies that $2(\mu n - \lambda n) + \mu^{1+\frac{1}{100k}} n \leq 2\mu n$. This is however impossible since, by Condition (4), $\mu^{1+\frac{1}{100k}} > 2\lambda$. □

The following lemma, in more general forms, is proved in [9], [13], [6], and also follows from the results of [10]. To make the paper more self contained we provide a proof.

Lemma 3.11. *Assume that V_1, V_2 are m-dimensional vectorspaces over the field F_2, $X \subseteq V_1, Y \subseteq V_2$, $|X| \geq 2^{m_1}$, $|Y| \geq 2^{m_2}$ and B is a linear map of V_1 into V_2 so that the rank of B is at least r. If $m_1 + m_2 + r > 2m$ then the function $Bx \cdot y$, $x \in X$, $y \in Y$ is not constant on $X \times Y$.*

Proof of Lemma 3.11. Let x_0 be an arbitrary but fixed element of X and let $X' = \{x - x_0 \mid x \in X\}$. Clearly $|X| = |X'|$ and if $Bx \cdot y$ is constant on $X \times Y$ then $Bx \cdot y$ is identically 0 on $X' \times Y$. Therefore it is enough to prove that the assumptions of the lemma imply that $Bx \cdot y$ is not identically 0 on $X \times Y$. Assume that, contrary to our assertion, it is identically 0. Let H be the subspace in V_1 generated by X and G be the subspace in V_2 generated by Y. We have $BH \cdot G = 0$, that is, the subspaces BH and G are orthogonal. Therefore $\dim(BH) + \dim(G) \leq m$, where $\dim(W)$ denotes the dimension of the subspace W. Since the rank of B is at least r we have that $\dim(BH) \geq \dim(H) - (m - r)$. We have

$\dim(H)-(m-r)+\dim(G)\le m$. The lower bound on the sizes of the sets X,Y imply the following lower bound on the dimensions of the subspaces generated by them: $\dim(H)\ge m_1$, $\dim(G)\ge m_2$. This simply follows from the fact that a d-dimensional subspace has 2^d elements. We have $m_1-(m-r)+m_2\le m$, that is, $m_1+m_2-r\le 2m$ in contradiction to our assumption. □

Remark 3.12. The lower bounds on $\dim(H)$ and $\dim(G)$ remain true even if the field has characteristic different from 2, but we assume that the elements of X and Y have only $\{0,1\}$ coefficients in a suitably chosen basis of V_1 and V_2. See Lemma 7 of [13]. This is important for the generalization of Theorem 3.4 for fields with other characteristics.

3.4 The proof of the main result

Proof of Theorem 3.4. Assume that, contrary to our assertion, there is a branching program $\mathcal{B}$ with the given parameters which computes the parity of $N_+(X)$. Let $m=\lfloor\frac{n}{10}\rfloor$. We apply Lemma 3.9 with $n\to m$, $k\to ck$, where c is a sufficiently large absolute constant and $\varepsilon\to\frac{\varepsilon}{2}$. Assume that σ_1, σ_2 are picked with the properties described in the lemma.

Let g be the function defined in Lemma 3.7. Applying Lemma 3.7 with $n\to m$, $\gamma\to\sigma_2$ we get that there is an m by m g-rigid matrix $A=(a_{i,j})$ over F_2. If σ_1 is sufficiently small with respect to δ, then A will be f-rigid as well. Therefore by Lemma 3.9 there is no branching program of size at most $2^{\frac{\varepsilon}{2}n}$ which computes $u_\zeta\tilde{A}\cdot u_\zeta$ in time ckn for all ζ, where ζ is an F_2-valued function defined on $\{0,1,\dots,m-1\}$. Let $D=\{i+j\mid\ a_{i,j}=1\}$ and

$$X_\zeta=\{i\in\{0,1,\dots,m-1\}\mid\zeta(i)=1\}\ .$$

For any pair of sets of integers X,Z let $N_+(X,Z)$ be the number of pairs x,y, $x<y$ so that $x\in X$, $y\in X$ and $x+y\in Z$. The statement of Lemma 3.9 in our case is that the parity of $N_+(X_\zeta,D)$ cannot be decided by a branching program with the given restrictions on its parameters. We show that this problem can be reduced to the problem of determining the parity of $N_+(X_\eta)$ for a suitably chosen $\eta\in\mathrm{Func}(n,2)$ in a way which can be implemented by a linear-time branching program. Therefore our indirect hypothesis will contradict to Lemma 3.9. η is defined in the following way.

We define first two sets U_1,U_2. $U_1=2m+X_\zeta$, $U_2=4m+D$. Let η be the unique element of $\mathrm{Func}(\{0,1,\dots,n-1\},\{0,1\})$ so that $X_\eta=U_1\cup U_2$. Clearly, if $x,y\in X_\zeta$, $x<y$, and $x+y\in D$ then $2m+x\in X_\eta, 2m+y\in X_\eta$, $2m+x<2m+y$, and $(2m+x)+(2m+y)\in X_\eta$. Conversely, assume that $z,w\in X_\eta$, $z<w$ and $z+w\in X_\eta$. It is easy to see that this implies $z,w\in\{2m,\dots,3m-1\}$ and therefore $z-2m,w-2m\in X_\zeta$, $z-2m<w-2m$, and $(z-2m)+w(-2m)\in X_\zeta$. Therefore $N_+(X_\zeta,D)=N_+(X_\eta)$. We claim that each value of η can be computed in constant time by a branching program, and to do this the size of our program must be increased only by a factor of two since the extra memory needed for this step is only one bit. Indeed, assume that we want to determine the value of $\eta(i)$ for some $i\in\{0,1,\dots,n-1\}$. First the program decides whether $i\in U_1$ by checking whether $\zeta(i-2m)=1$. If not, then $\eta(i)=0$. If $\zeta(i-2m)=1$, then it has to decide whether $i\in U_2$. Since D is part of the input this can be decided by checking whether $i-4m\in D$. If the anser is no then $\eta(i)=0$, if the answer is yes then $\eta(i)=1$. Therefore we have reduced the problem of determining the parity of $N_+(X_\zeta,D)$ to the problem of determining the parity of $N_+(X_\eta)$. □

4 Random Hankel matrices

4.1 The statement of the result

In this section we show that with a positive probability all large submatrices of a random Hankel matrix have relatively large ranks.

Definition 4.1. The field with q elements will be denoted by F_q.

Theorem 4.2. *There exists a $c_1 > 0$ so that, for all $c_2 > 0$, if n is sufficiently large then the following holds: Assume that $A = \{a_{i,j}\}$, $i = 0,\ldots,n-1$, $j = 0,\ldots,n-1$ is a random n by n Hankel matrix over F_2, taken with uniform distribution on the set of all such matrices. Then with a probability greater than $\frac{1}{2}$, A has the following property:*

(6)**.** Suppose $S = \{s_0,\ldots,s_{q-1}\}$, $T = \{t_0,\ldots,t_{q-1}\}$ are subsets of $\{0,\ldots,n-1\}$ with q elements, where $c_2 n < q < \frac{n}{2}$, and $B_{S,T} = (a_{s_i,t_j})$, $i = 0,\ldots,q-1$, $j = 0,\ldots,q-1$ is the submatrix of A consisting of those entries whose row numbers are in S and column numbers are in T. Then the rank of $B_{S,T}$ is at least $c_1|\log(\frac{q}{n})|^{-2}q$.

4.2 Sketch of the proof

4.2.1 A natural but unsuccessful attempt

The most natural way to prove the statement of the theorem would be the following. Assume that the sets S,T are fixed. We give an upper bound M on the probability $p_{S,T}$ of the event that, for the randomization of A, the matrix $B_{S,T}$ defined for the fixed sets S and T has rank smaller than $c_1|\log(\frac{q}{n})|^{-2}$. If M multiplied by the number of choices for the pair $\langle S,T\rangle$ is smaller than $\frac{1}{2}$ then the assertion of the theorem clearly holds.

Unfortunately a proof of this type cannot work. Indeed if $q = cn$ then the number of pairs $\langle S,T\rangle$ is about $2^{2c(\log\frac{1}{c})n}$. On the other hand for a fixed pair S,T, in the worst case, the number of minor diagonals of A intersected by $S \times T$ can be as small as $2cn-1$. Each of the choices of 0s and 1s in A on these diagonals are equally probable so the probability that we get rank smaller than $c_1|\log(\frac{q}{n})|^{-2}$ is at least 2^{-2cn+1}. (It is not 0 since, e.g. the 0 matrix has such a small rank.) Since the absolute value of the exponent in the number of pairs is greater by a factor of $|\log\frac{1}{c}|$ than in the upper bound M, the product cannot be smaller than $\frac{1}{2}$ if c is a small constant.

4.2.2 Reducing the number of relevant submatrices

The main problem with the argument described above was that the number of pairs $\langle S,T\rangle$ is too large compared to the number of relevant minor diagonals. Let $\mathcal{S}$ be the set of these pairs, that is, the set of all pairs $\langle S,T\rangle$ so that $S,T \subseteq \{0,1,\ldots,n-1\}$ and $|S| = |T| = q$. We will be able to avoid the mentioned difficulty in the following way. Instead of working with the elements of $\mathcal{S}$, we will consider a smaller set $\mathcal{S}'$ consisting of pairs $\langle S',T'\rangle$ so that $|S'| \le q$, $|T'| \le q$ and with the property that for all $\langle S,T\rangle \in \mathcal{S}$ there is a $\langle S',T'\rangle \in \mathcal{S}'$ so that $S' \subseteq S$ and $T' \subseteq T$. (We will refer to this property by saying that $\mathcal{S}'$ is *dense* in $\mathcal{S}$.)

It is enough to show that for all $\langle S', T' \rangle \in \mathcal{S}'$ the rank of $B_{S',T'}$ is at least $c_1 |\log(\frac{q}{n})|^{-2} q$. Indeed, since $\mathcal{S}'$ is dense in $\mathcal{S}$, each matrix $B_{S,T}$ with $\langle S, T \rangle \in \mathcal{S}$ has a submatrix $B_{S',T'}$ with $\langle S', T' \rangle \in \mathcal{S}$ and so $\operatorname{rank}(B_{S,T}) \geq \operatorname{rank}(B_{S',T'}) \geq c_1 |\log(\frac{q}{n})|^{-2} q$.

We will define the set $\mathcal{S}'$ by constructing a function $\mathcal{F}$ defined on $\mathcal{S}$ so that for each $\langle S, T \rangle \in \mathcal{S}$, $\mathcal{F}(\langle S, T \rangle) = \langle S', T' \rangle$ with $S' \subseteq S$, $T' \subseteq T$. Clearly if such an $\mathcal{F}$ is given and $\mathcal{S}' = \{\mathcal{F}(\langle S, T \rangle) \mid \langle S, T \rangle \in \mathcal{S}\}$, then $\mathcal{S}'$ is dense in $\mathcal{S}$. We also have to make sure that $|\mathcal{S}'|$ is small and that we are able to give a good upper bound, for each fixed $\langle S', T' \rangle \in \mathcal{S}'$, on the probability that the rank of the matrix $B_{S',T'}$ is smaller than $c_1 |\log(\frac{q}{n})|^{-2}$.

4.2.3 The rank of an enlarged submatrix

First we describe our method of estimating the probability that the rank of a submatrix $B_{S,T}$ of A is small for a fixed pair $\langle S, T \rangle$. This will be based on the following observation. Assume that a pair $\langle S, T \rangle$, $S, T \subseteq \{0, 1, \ldots, n-1\}$, is fixed and $s > \max S = \max_{x \in S} x$, $t > \max T$, $S_1 = S \cup \{s\}$, and $T_1 = T \cup \{t\}$. Then with probability at least $\frac{1}{2}$ for the randomization of A we have that the rank of B_{S_1,T_1} is strictly greater than the rank of $B_{S,T}$. We will prove this statement in the following way. For all $k = 0, 1, \ldots, n-1$, let D_k be the minor diagonal of A containing the entries $a_{i,j}$ with $i + j = k$. We show that if the values of the entries of A are fixed on all minor diagonals D_k with $k < s + t$, then out of the two possible definitions of A on the minor diagonal D_{s+t}, at least one will yield a matrix B_{S_1,T_1} with the property that $\operatorname{rank}(B_{S_1,T_1}) > \operatorname{rank}(B_{S,T})$. The proof of this fact is a simple argument in linear algebra as described in the proof of Lemma 4.5.

Lemma 4.5 itself is a slight generalization of this assertion, stating that if we add not a single new element to S and another single element to T, but a set of new elements $\tilde{S}$ to S so that $\max S < \min \tilde{S}$, and a set of new elements $\tilde{T}$ to T so that $\max T < \min \tilde{T}$ then the resulting enlarged sets $S_1 = S \cup \tilde{S}$, $T_1 = T \cup \tilde{T}$ have the following property. If the values of the entries of A are fixed on all minor diagonals D_k with $k \leq \max S + \max T$, then for the randomization of A on the remaining minor diagonals we have that, with probability at least $1 - 2^{-|\tilde{S}+\tilde{T}|}$, $\operatorname{rank}(B_{S_1,T_1}) > \operatorname{rank}(B_{S,T})$. Indeed, there are $|\tilde{S} + \tilde{T}|$ minor diagonals which contain an entry $a_{s,t}$ of A with $s \in \tilde{S}$ and $t \in \tilde{T}$. According to the already described special case the randomization of the values of the entries on each of these diagonals will lead to the required increase of the rank with a probability of at least $\frac{1}{2}$. Since these randomizations are independent we get that the rank increases with probability at least $1 - 2^{-|\tilde{S}+\tilde{T}|}$.

4.2.4 Partitioning the rows and columns

What we have done so far is only good for estimating the probability of $\operatorname{rank}(B_{S_1,T_1}) > \operatorname{rank}(B_{S,T})$ for some pairs $\langle S, T \rangle$, $\langle S_1, T_1 \rangle$ where $S \subseteq S_1$, $T \subseteq T_1$. To get a lower bound on the probability of $\operatorname{rank}(B_{S,T}) > R$ for some integer R, we will partition S into subsets $S_1, \ldots, S_l$ and T into subsets $T_1, \ldots, T_l$ so that $\max S_i < \min S_{i+1}$ and $\max T_i < \min T_{i+1}$ for all $i = 1, \ldots, l-1$. If $\operatorname{rank}(B_{S,T}) \leq R = l - r$ then there must be at least r distinct elements i of $\{1, \ldots, l\}$ with the property $\operatorname{rank}(B_{\Gamma_i,\Lambda_i}) = \operatorname{rank}(B_{\Gamma_{i+1},\Lambda_{i+1}})$, where $\Gamma_j = S_1 \cup \ldots \cup S_j$ and $\Lambda_j = T_1 \cup \ldots \cup T_j$ for $j = 1, \ldots, l$. We will denote by E the set of all integers $i \in \{1, \ldots, l\}$ with this property. Then

(7). the probability of the event that we get equalities for every element of this set is at most

$$2^{-\sum_{i\in E}(|S_i+T_i|)} .$$

If we add these upper bounds for all of the possible choices for E, that is for all subsets of $\{1,\ldots,l\}$ with r elements, then we get an upper bound on the probability of $\mathrm{rank}(B_{S,T}) \leq l-r$. (This upper bound is formulated in a slightly more general form in Lemma 4.7.) We will use this estimate for each fixed choice for $\langle S,T\rangle \in \mathcal{S}'$ with suitable choices of the partitions $S_1,\ldots,S_l$, $T_1,\ldots,T_l$.

4.2.5 The choice of the partitions and the submatrices

Our remaining task is to define the function $\mathcal{F}$ so that

$$\mathcal{S}' = \{\mathcal{F}(\langle X_1,X_2\rangle) \mid \langle X_1,X_2\rangle \in \mathcal{S}\}$$

is dense in $\mathcal{S}$, select a pair of partitions for each $\langle S,T\rangle \in \mathcal{S}'$, and then add the corresponding upper bounds (with $R = c_1 q|\log(\frac{q}{n})|^{-2}$) for each $\langle S,T\rangle \in \mathcal{S}'$. The upper bounds will not depend on the choice of $\langle S,T\rangle \in \mathcal{S}'$, so we will have to prove that the common upper bound multiplied by $|\mathcal{S}'|$ is at most $\frac{1}{2}$.

When we define $\mathcal{F}(\langle X_1,X_2\rangle)$ for some $\langle X_1,X_2\rangle \in \mathcal{S}$, we will have already in mind the task of choosing suitable partitions of S and T, where $\langle S,T\rangle = \mathcal{F}(X_1,X_2)$. We give here a somewhat simplified definition of $\mathcal{F}$, the final definition will be provided in the proof of Lemma 4.9. Let t be a positive integer which is a large constant. We assume now, for the sake of simplicity, that $t^2|q$. For $j=1,2$ we partition X_j into $\frac{q}{t^2}$ subsets $K_1^{(j)},\ldots,K_{q/t^2}^{(j)}$ each containing exactly t^2 elements so that $\max K_i^{(j)} < \min K_{i+1}^{(j)}$ for $i=1,\ldots,\frac{q}{t^2}-1$. Clearly these properties uniquely determine both partitions.

For each fixed $i=1,\ldots,q/t^2$ we pick sets $J_i^{(j)} \subseteq K_i^{(j)}$, $j=1,2$, with exactly t elements so that $|J_i^{(1)}+J_i^{(2)}|$ is maximal. Since the sets $J_i^{(j)}$ have t elements this maximum is at most t^2. We will show (Lemma 4.3) that, since we pick the sets $J_i^{(j)}$ from sets of size t^2, this upper bound can be can be attained, and so $|J_i^{(1)}+J_i^{(2)}| = t^2$ for all $i=1,2,\ldots,\frac{q}{t^2}$. Let

$$Z_j = \bigcup_{i=1}^{q/t^2} J_i^{(j)}$$

for $j=1,2$. Now we define $\mathcal{F}$ by $\mathcal{F}(\{X_1,X_2\}) = \langle Z_1,Z_2\rangle$. For $j=1,2$ we will use the partition $J_1^{(j)},\ldots,J_{q/t^2}^{(j)}$ of the set Z_j when estimating $\mathtt{prob}(\mathrm{rank}(B_{Z_1,Z_2}) \leq c_1|\log(\frac{q}{n})|^{-2}q)$. The inequality of Condition (7) gives a good upper bound which can be easily evaluated (as a function of q,t and n) since in the exponents the value of the expression $-(|J_i^{(1)}+J_i^{(2)}|)$ is t^2, and the number of exceptional sets E can be also estimated without any problems. Finally the number of possible pairs $\langle Z_1,Z_2\rangle$ is at most $\binom{n}{q/t^2}^2$. These estimates lead to the conclusion of the theorem.

4.2.6 Why did it work?

From the description of the necessary estimates at the end of the last paragraph it is not clear what made it possible to get a good enough upper bound on the probabilities $\mathtt{prob}(\mathrm{rank}(B_{Z_1,Z_2}) \leq R)$, where

$R = c_1|\log(\frac{q}{n})|^{-2}q$, compared to the number of pairs $\langle Z_1, Z_2\rangle$. It is true that the number of pairs became smaller, since the sizes of the sets Z_j are smaller by a factor of t then the sizes of the sets X_j, but for smaller sets the upper bounds on the probabilities can be larger. Why is it that we gained more on the number of sets then lost on the upper bounds on the probabilities?

The answer is that the upper bound did not depend on the sizes of the sets Z_i but depended only on the common size of the sets $J_i^{(1)} + J_i^{(2)}$ which was t^2. The corresponding quantity for the pair $\langle X_1, X_2\rangle$ is $|K_i^{(1)} + K_i^{(2)}|$. Since we do not have any assumption about the sets $K_i^{(1)}, K_i^{(2)}$ other than that their sizes are t^2, in the worst case $|K_i^{(1)} + K_i^{(2)}|$ can be as small as $2t^2$. Therefore, although the sizes of the sets went down by a factor of t, the critical quantity in the upper bounds remained essentially unchanged. This guaranteed that we won more on decreasing the number of pairs then we lost on increasing the upper bound of the probabilities. To formulate the same phenomenon in the language of minor diagonals we may say that: although the ratio of the sizes of the sets X_j and Z_j is t, if we consider the number of minor diagonals intersecting the subsets $K_i^{(1)} \times K_i^{(2)}$ resp. $J_i^{(1)} \times J_i^{(2)}$ the ratio, at least in certain cases, is at most 2.

This completes the sketch of the proof of the theorem. In the remaining part of the section we gave a detailed proof of the mentioned lemmata and the theorem.

4.3 The proof of Theorem 4.2

Lemma 4.3. *Assume that t is a positive integer and U, V are sets of integers with $|U| = |V| = t^2$. Then there are $U' \subseteq U$, $V' \subseteq V$, so that $|U'| = |V'| = t$ and $|U' + V'| = t^2$.*

Remark 4.4. As the proof will show, the lemma remains true if we replace $|U| = |V| = t^2$ by the weaker assumption $|U| = |V| = t^2 - t + 1$.

Proof of Lemma 4.3. We have to select the subsets U', V' of U and V so that each has exactly t elements and all of the t^2 sums $u + v$, $u \in U'$, $v \in V'$ are different. Suppose that this does not hold for some selection of U', V', that is $u + v = \bar{u} + \bar{v}$ for some suitably chosen $u, \bar{u} \in U'$, $v, \bar{v} \in V'$. This would imply that $u - \bar{u} = \bar{v} - v$ and so the sets $(U' - U')_+$ and $(V' - V')_+$ are not disjoint, where for a set of integers X, $(X)_+ = \{x \in X \mid x > 0\}$. Therefore it is sufficient (and also necessary) to prove that there exist $U' \subseteq U$, $V' \subseteq V$ so that $|U'| = |V'| = t$ and $(U' - U')_+ \cap (V' - V')_+ = \emptyset$.

Let $U = \{u_0, \ldots, u_{t^2-1}\}$, $V = \{v_0, \ldots, v_{t^2-1}\}$ so that $u_0 < \ldots < u_{t^2-1}$ and $v_0 < \ldots < v_{t^2-1}$. We define the integers m_u, m_v by

$$m_u = \min\{u_{i+t-1} - u_i \mid i = 0, 1, .., t^2 - t\} \qquad \text{and} \qquad m_v = \min\{v_{i+t-1} - v_i \mid i = 0, 1, .., t^2 - t\} \ .$$

Suppose that, e.g. $m_u \leq m_v$, and let s be an integer with $m_u = u_{s+t-1} - u_s$. We claim that the choice $U' = \{u_s, u_{s+1}, \ldots, u_{s+t-1}\}$, $V' = \{v_{jt} \mid j = 0, 1, \ldots, t-1\}$ meets our requirements. Indeed, if $v_{jt}, v_{kt} \in V'$ and $j < k$ then $v_{kt} - v_{jt} \geq v_{(j+1)t} - v_{jt} > v_{jt+t-1} - v_{jt} \geq m_v \geq m_u$, and therefore each element of $(V' - V')_+$ is strictly greater than m_u. On the other hand $U' \subseteq [u_s, u_{s+t-1}] = [u_s, u_s + m_u]$; therefore $(U' - U')_+$ contains only integers not greater than m_u. Consequently $(U' - U')_+ \cap (V' - V')_+ = \emptyset$. □

Definitions.

1. func$(n,2)$ will denote the set of all functions defined on $\{0,\ldots,n-1\}$ with values in F_2. Similarly, func$([l,n),2)$ will denote the set of all functions defined on the interval $[l,n)=\{l,\ldots,n-1\}$ with values in F_2.
2. Assume that n_1,n_2 are positive integers and $f\in \text{func}(n_1+n_2-1,2)$. Then diag$(f,n_1,n_2)$ will be the n_1 by n_2 matrix $(d_{i,j})$, $i=0,\ldots,n_1-1$, $j=0,1,\ldots,n_2-1$, where $d_{i,j}=f(i+j)$.
3. Assume that n_1,n_2,k_1,k_2, are positive integers, $n_1>k_1$, $n_2>k_2$, $f\in \text{func}(k_1+k_2-1,2)$, and g is taken with uniform distribution from the set func$([k_1+k_2,n_1+n_2-1),2)$. $\Phi(n_1,n_2,f)$ will be a random variable whose value is diag$(f\cup g,n_1,n_2)$ (where $f\cup g$ is the unique common extension of f and g to $[0,n_1+n_2-1)$). $\Phi(n_1,n_2)$ will denote the random variable whose value is diag(h,n_1,n_2) where h is taken with uniform distribution from the set func$(n_1+n_2-1,2)$.
4. Suppose $A=(a_{i,j})$, $i=0,\ldots,n_1-1$, $j=0,1,\ldots,n_2-1$, is an n_1 by n_2 matrix and $S\subseteq\{0,1,\ldots,n_1-1\}$, $T\subseteq\{0,1,\ldots,n_2-1\}$. Then sub$(A,S,T)$ will denote the $|S|$ by $|T|$ matrix consisting of those entries of A which have row numbers in S and in column numbers in T.

Lemma 4.5. *Assume that n_1, n_2, k_1, and k_2 are positive integers, $k_1<n_1$, $k_2<n_2$, f is a function on $\{0,1,\ldots,k_1+k_2-1\}$ with values in F_2, $S\subseteq\{0,1,\ldots,n_1-1\}$, $T\subseteq\{0,1,\ldots,n_2-1\}$, and*

$$|(S\cap\{k_1,\ldots,n_1-1\})+(T\cap\{k_2,\ldots,n_2-1\})|\geq m \ .$$

Then with probability at least $1-2^{-m}$ the following holds: the rank of the matrix sub$(\Phi(n_1,n_2,f),S,T)$ *is greater than the rank of the matrix*

$$\text{sub}(\Phi(n_1,n_2,f),S\cap\{0,1,\ldots,k_1-1\},T\cap\{0,1,\ldots,k_2-1\}) \ .$$

Remark 4.6. If we define random Hankel matrices over an arbitrary field F so that the random entries of the Hankel matrices are picked from a finite subset D of F with uniform distribution, then our Lemma remains true if we substitute $1-|D|^{-m}$ for the probability $1-2^{-m}$. (Naturally we also have to modify the definition on $\Phi(n_1,n_2,f)$, since in this case f is a function whose values are in the set D.)

Proof of Lemma 4.5. Let $\Phi(n_1,n_2,f)=(\varphi_{i,j})$, $i=0,\ldots,n_1-1$, $j=0,\ldots,n_2-1$. For each $\ell=0,1,2,\ldots$ let $S_\ell=S\cap\{0,1,\ldots,\ell\}$, $T_j=T\cap\{0,1,\ldots,\ell\}$. For each $i\in S$, $j\in T$, $w_{i,j}$ will be a function defined on T_j by $w_{i,j}(x)=\varphi_{i,x}$ for all $x\in T_j$. Let r be rank of the matrix sub$(\Phi(n_1,n_2,f),S_{k_1-1},T_{k_2-1})$. r is the dimension of the vectorspace generated by the functions w_{i,k_2-1}, $i\in S_{k_1-1}$. Suppose that $\bar{S}\subseteq S_{k_1-1}$, $|\bar{S}|=r$ so that the set of functions $W=\{w_{i,k_2-1}\mid i\in\bar{S}\}$ are linearly independent.

According to the definition of $\Phi(n_1,n_2,f)$, we have to randomize a function g with values in F_2 which is defined on the interval $[k_1+k_2,n_1+n_2-1)$. We randomize the values of g sequentially for each $x\in[k_1+k_2,n_1+n_2-1)\cap(S+T)$. Assume that $x\in[k_1+k_2,n_1+n_2-1)]$ and $g(y)$ has been randomized already for all $y<x$. Suppose that for a suitably chosen $i\in S\cap\{k_1,\ldots,n_1-1\}$ and $j\in T\cap\{k_2,\ldots,n_2-1\}$ we have $i+j=x$. By the assumption of the lemma this will happen for at least m different values of x. Therefore, it is enough to show that for such an x the following holds with a probability at least $\frac{1}{2}$: the function $w_{i,j}$ is linearly independent from the set of functions $H=\{w_{l,j}|l\in\bar{S}\}$. (Such an independence obviously implies that the rank of the matrix sub$(\Phi(n_1,n_2,f),S,T)$ is greater than $|\bar{S}|=r$.) Before the randomization of $g(x)$ the function $w_{i,j}$ is known in every point of T_j with the exception of j. Since there

are two possibilities for the value of $w_{i,j}$ at j, we have two functions u, v; hence for the randomization of $g(x)$ we have $P(w_{i,j} = u) = P(w_{i,j} = v) = \frac{1}{2}$. Consequently, it is enough to show that at least one of the two vectors u, v is linearly independent from the set H. Indeed, if both are linearly dependent, then their difference is also linearly dependent on them, that is, $u - v = \sum_{s \in \bar{S}} \gamma_s w_{s,j}$ where $\gamma_s \neq 0$ for at least one $s \in \bar{S}$. We show that this is impossible. Indeed, $u - v$ is a function on T_j which is zero everywhere but at j and $(u-v)(j) = 1$. Consequently $j \geq k_2$ implies that the restriction of $u - v$ to T_{k_2-1} is 0. Therefore we get that $\sum_{s \in \bar{S}\gamma_s} \gamma_s w_{s,k_2} = 0$. The functions w_{s,k_2} are linearly independent so we have $\gamma_s = 0$ for all $s \in \bar{S}$, in contradiction to our assumption. □

Lemma 4.7. *Assume that for each $j = 1,2$, $I_1^{(j)}, \ldots, I_l^{(j)}$, is a partition of the interval $[0, \ldots, n)$ into pairwise disjoint subintervals, $S^{(1)}, S^{(2)} \subseteq \{0, 1, \ldots, n-1\}$, and $|(S^{(1)} \cap I_i^{(1)}) + (S^{(2)} \cap I_i^{(2)})| \geq m_i$ for all $i = 1, \ldots, l$. Then, for any positive integer r, the probability that the rank of* $\mathrm{sub}(\Phi(n,n), S^{(1)}, S^{(2)})$ *is not greater than $l - r$ is at most*

$$\sum_{1 \leq i_1 < \cdots < i_r \leq l} 2^{-m_{i_1} - \ldots - m_{i_r}} \ .$$

Remark 4.8. If we define the random Hankel matrix $\Phi(n,n)$ over an arbitrary field F in the way described in Remark 4.6 (just after Lemma 4.5), then our lemma remains true if we substitute $|D|^{-m_{i_1} - \ldots - m_{i_r}}$ for $2^{-m_{i_1} - \ldots - m_{i_r}}$ in the last expression of the lemma.

Proof of Lemma 4.7. Assume that for all $j = 1,2$, $x \in I_i^{(j)}$, $y \in I_{i+1}^{(j)}$ implies $x < y$, and assume further that for all $j = 1,2$, $i = 1, \ldots, l$, $I_i^{(j)} = [b_{j,i}, b_{j,i+1})$. Let

$$S_i^{(j)} = S^{(j)} \cap \bigcup_{k=1}^{i} I_k^{(1)} = S^{(j)} \cap [0, b_{j,i+1}) \ ,$$

and let $\Phi_i = \mathrm{sub}(\Phi(n,n), S_i^{(1)}, S_i^{(2)})$. If the rank of $X = \mathrm{sub}(\Phi(n,n), S^{(1)}, S^{(2)})$ is not greater than $l - r$ then there are r integers $1 \leq i_1 < \ldots < i_r \leq l$ so that the rank of Φ_{i_t} and Φ_{i_t+1} is the same for $t = 1, \ldots, r$. We show that for each fixed $i_1, \ldots, i_r$ the probability of this event is at most $2^{-m_{i_1} - \ldots - m_{i_r}}$ which clearly implies the statement of the lemma. Suppose that $i_1, \ldots, i_r$ are fixed. According to the definition of $\Phi(n,n)$ we randomize an $h \in \mathrm{func}(2n-1, 2)$. We pick the values of h on $[2n-1, 2)$ sequentially. Assume that for some $t \in \{1, \ldots, r\}$ the values of $h(0), \ldots, h(b_{i_t,1} + b_{i_t,2}) - 1$ have been already fixed. We define a function f on the set $\{0, \ldots, h(b_{i_t,1} + b_{i_t,2}) - 1\}$ by $f(y) = h(y)$ for all $y = 0, \ldots, h(b_{i_t,1} + b_{i_t,2}) - 1$. Now we randomize the values of $h(x)$ for all $x = b_{i_t,1} + b_{i_t,2}, \ldots, b_{i_{t+1},1} + b_{i_{t+1},2} - 1$. We apply Lemma 4.5 for this part of the randomization with $n_j \to b_{t+1,j}$, $k_j \to b_{i_t,j}$ for $j = 1,2$, $S \to S_t^{(1)}$, $T \to S_t^{(2)}$, $m \to m_{i_t}$, and for the function f defined above. We get that the probability of the event $\mathrm{rank}(\Phi_{i_{t-1}}) = \mathrm{rank}(\Phi_{i_t})$ is less than 2^{-m_t}. This implies that the probability that $\mathrm{rank}(\Phi_{i_{t-1}}) = \mathrm{rank}(\Phi_{i_t})$ for all $t = 1, \ldots, r$ is at most $2^{-m_1 - \ldots - m_r}$. □

The following lemma will be used to give an estimate on the probability that the rank of a matrix $\mathrm{sub}(A, S, T)$ is at least R where the sets $S, T \subseteq \{0, 1, \ldots, n-1\}$ are fixed and A is a random Hankel matrix over F_2. Since the statement of the lemma depends on many parameters we restate their roles as described in the "sketch of the proof" of Theorem 4.2. The sizes of the sets S, T are the same:

$|S| = |T| = q$. We may think of t as a large constant, although the lemma requires only $t^2 < q$. In the "sketch of the proof" (using the notation $S \to X_1, T \to X_2$) we made the simplifying assumption that $t^2|q$, and partitioned both S and T into $\frac{q}{t^2}$ subsets each of size t^2. Without the assumption $q^2|t$, we will take first subsets of both S and T with $t^2\lfloor\frac{q}{t^2}\rfloor$ elements and partition these subsets into classes each of size t^2. Therefore $Q = \lfloor\frac{q}{t^2}\rfloor$ of the lemma refers to the number of classes in these partitions. In the estimate given in the lemma the factor $2^{-(Q-R+1)t^2}$ is an upper bound on the probability that if we select $R-1$ pairs of corresponding classes from the two partitions, then the remaining ones do not increase the rank of $\mathrm{sub}(A,S,T)$ in the sense explained in the "sketch". The factor $\binom{Q}{Q-R+1}$ is the number of ways that we can select these $R-1$ pairs from the Q pairs of classes. The factor $\binom{n}{Qt}^2$ has the following meaning. In the proof we consider all of the pairs of subsets $S' \subseteq S$, $T' \subseteq S$ and estimate the probability that for at least one of them the rank of $\mathrm{sub}(A,S',T')$ will be R or greater. Then we multiply this estimate by the number of possible pairs of sets S', T'. Since we select S', T' with $|S'| = |T'| = Qt$ (they contain exactly t elements from each class) the number of possible selections is at most $\binom{n}{Qt}^2$.

Lemma 4.9. *Assume that n,q,R,t are positive integers, $t^2 < q < n$ and $R < \lfloor\frac{q}{t^2}\rfloor$. Suppose further that $A = \{a_{i,j}\}$, $i = 0,\ldots,n-1$, $j = 0,\ldots,n-1$ is a random n by n Hankel matrix over F_2, taken with uniform distribution on the set of all such matrices. Let p be the probability of the following event:*

(8). for all $S \subseteq [0,n), T \subseteq [0,n)$, $|S| = |T| = q$ the rank of the matrix $\mathrm{sub}(A,S,T)$ is at least R.

Then

$$p \geq 1 - \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-(Q-R+1)t^2}$$

where $Q = \lfloor\frac{q}{t^2}\rfloor$.

Remark 4.10. This lemma also remains true with some modifications over an arbitrary field if we randomize the Hankel matrix A according to the distribution described in the remark after Lemma 4.5. Namely we have to substitute $|D|^{-(Q-R+1)t^2}$ for $2^{-(Q-R+1)t^2}$ in the last expression of the lemma.

Proof of Lemma 4.9. We will define a function $\mathcal{F}$ on the set of all ordered pairs $\langle X_1, X_2\rangle$ with $X_j \subseteq \{0,\ldots,n-1\}$, for $j = 1,2$, $|X_1| = |X_2| = q$. Before getting into the details of this definition, recall from the "sketch of proof"of Theorem 4.2 that, roughly speaking, we get $\mathcal{F}(\langle X_1, X_2\rangle)$ in the following way. First we partition both X_1 and X_2 into classes of sizes t^2. Then we select subsets $J_i^{(1)}$, $J_i^{(2)}$ from each pairs of corresponding classes with the property that $|J_i^{(1)}| = |J_i^{(2)}| = t$, $|J_i^{(1)} + J_i^{(2)}| = t^2$. The pair $\langle \bigcup_i J_i^{(1)}, \bigcup_i J_i^{(2)}\rangle$ is the value of $\mathcal{F}$.

Each value of the function will be a pair $\langle Z_1, Z_2\rangle$ so that $Z_1, Z_2 \subseteq \{0,\ldots,n-1\}$ and $|Z_1| = |Z_2| = \lfloor\frac{q}{t^2}\rfloor t$. The definition is the following. Assume that the pair $\langle X_1, X_2\rangle$ is given with the described properties. For each $j = 1,2$ we pick pairwise disjoint subsets $K_1^{(j)},\ldots,K_Q^{(j)}$ of X_j, where $Q = \lfloor\frac{q}{t^2}\rfloor$, so that $|K_i^{(j)}| = t^2$ for all $j = 1,2$ $i = 1,\ldots,Q$ and $x \in K_i^{(j)}$, $y \in K_{i'}^{(j)}$ implies $x < y$ for all $j = 1,2$, $1 \leq i < i' \leq Q$. (By the definition of Q this is possible.)

Assume now that an $i = 1,\ldots,Q$ is fixed. We apply Lemma 4.3 with $U \to K_i^{(1)}$, $V \to K_i^{(2)}$. Let U', V' be the sets whose existence is stated in Lemma 4.3 and let $J_i^{(1)} = U'$, $J_i^{(2)} = V'$. Finally let $Z_j = \bigcup_{i=1}^{Q} J_i^{(j)}$

for $j = 1,2$ and let $\mathcal{F}(\langle X_1, X_2\rangle) = \langle Z_1, Z_2\rangle$. Clearly the pair $\langle Z_1, Z_2\rangle$ satisfies the conditions described above as well as the following additional properties:

(a). for all $j = 1,2$ $J_1^{(j)}, \ldots, J_Q^{(j)}$ is a partition of Z_j, $|J_i| = t$ for all $i = 1, \ldots, Q$,

(b). for all $j = 1,2$, $1 \leq i < i' \leq Q$ $x \in J_i^{(j)}$, $y \in J_{i'}^{(j)}$ implies $x < y$,

(c). for all $j = 1,2$ and $i = 1, \ldots, q$ we have $Z_j \subseteq X_j$ and $|J_i^{(1)}| + |J_i^{(2)}| = t^2$.

Assume now that $\langle Z_1, Z_2\rangle \in \mathtt{range}(\mathcal{F})$. We estimate the probability ρ_{Z_1,Z_2} of the event that the rank of the matrix $\mathrm{sub}(A, Z_1, Z_2)$ is smaller than R.

We apply Lemma 4.7 with $l \to Q$, $I_i^{(j)} \to J_i^{(j)}$, $S^{(1)} \to Z_1$, $S^{(2)} \to Z_2$, $m_i \to t^2$, $r \to Q - R + 1$. We get that ρ_{Z_1,Z_2} is at most $QQ - R + 12^{-(Q-R+1)t^2}$. Therefore, using that $|Z_j| = Qt$, we get that the probability that the rank of the matrix $\mathrm{sub}(A, Z_1, Z_2)$ is smaller than R for at least one $\langle Z_1, Z_2\rangle \in \mathtt{range}(\mathcal{F})$ is at most

$$|\mathtt{range}(\mathcal{F})| \binom{Q}{Q-R+1} 2^{-(Q-R+1)t^2} \leq \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-(Q-R+1)t^2} \quad .$$

For each pair S, T, with the properties given in the lemma, if $\mathcal{F}(\langle S, T\rangle) = \langle Z_1, Z_2\rangle$, then $Z_1 \subseteq S$, $Z_2 \subseteq T$ and this implies that $\mathrm{rank}(\mathrm{sub}(A, S, T)) \geq \mathrm{rank}(\mathrm{sub}(A, Z_1, Z_2))$ so we have the same upper bound on the probability that the rank of $\mathrm{sub}(A, S, T)$ is smaller than R. □

Proof of Theorem 4.2. Assume that $\theta > 0$ is sufficiently small, $c_1 > 0$ is sufficiently small with respect to θ, and $c_2 > 0$. Suppose further that n is sufficiently large and $c_2 n < q \leq n$. We apply Lemma 4.9 with n, q, $R = c_1 |\log(\frac{q}{n})|^{-1} q$, and $t = \lfloor \theta^{-1} |\log(\frac{q}{n})| \rfloor$. We get that the probability that rank of $\mathrm{sub}(A, S, T)$ is at least R is at least

$$p \geq 1 - \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-(Q-R+1)t^2}$$

where $Q = \lfloor \frac{q}{t^2} \rfloor$. We show that

$$\binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-(Q-R+1)t^2}$$

is at most $\frac{1}{2}$ by giving upper bounds in its factors. As we have remarked already at the end the of sketch of the proof, the crucial fact that leads to the desired result is that in the exponent of 2 we have the factor t^2 and not only t. We will see that in the actual estimates this t^2 makes it possible to get the strong upper bound we need.

We will use that if $0 < \alpha < \frac{1}{2}$, n is sufficiently large, and $x < \alpha n$ then

$$\binom{x}{\alpha n} \leq e^{2\alpha n \log \frac{1}{\alpha}} \quad .$$

Let $\gamma = \frac{q}{n}$, and $\lambda = \frac{Qt^2}{q}$. Clearly $c_2 < \gamma < 1$ and $\frac{1}{2} < \lambda \leq 1$. Hence

$$\binom{n}{Qt} = \binom{n}{\gamma\lambda t^{-1} n} \leq e^{2\gamma\lambda t^{-1} n \log(\gamma^{-1}\lambda^{-1}t)} = e^{2\gamma\lambda t^{-1} n(\log\gamma^{-1} + \log\lambda^{-1} + \log t)} \quad .$$

Using that $t^{-1}\log\gamma^{-1}=\theta$, $t^{-1}\log\lambda^{-1}\le t^{-1}\log 2\le t^{-1}\le\theta$, and $t^{-1}\log t\le t^{-\frac{1}{2}}\le\theta^{\frac{1}{2}}$ we get that

$$\binom{n}{Qt}^2\le e^{4\gamma\lambda(\theta+\theta+\theta^{\frac{1}{2}})n}\le 2^{\frac{1}{20}\gamma\lambda n}$$

and

$$\binom{Q}{Q-R+1}\le 2^Q=2^{\gamma\lambda t^{-2}n}\le 2^{\frac{1}{20}\gamma\lambda n}$$

if θ is sufficiently small. Moreover,

$$2^{-(Q-R+1)t^2}\le 2^{-\frac{1}{8}Qt^2}=2^{-\frac{1}{8}\gamma\lambda t^{-2}t^2n}=2^{-\frac{1}{8}\gamma\lambda n}\quad .$$

These inequalities imply that

$$\binom{n}{Qt}^2\binom{Q}{Q-R+1}2^{-(Q-R+1)t^2}\le 2^{\frac{1}{20}\gamma\lambda n+\frac{1}{20}\gamma\lambda n-\frac{1}{8}\gamma\lambda n}\le 2^{-(\frac{1}{8}-\frac{1}{10})\gamma\lambda n}<\frac{1}{2}$$

if n is sufficiently large. (Here we use that $c_2<\gamma$ and $\frac{1}{2}<\lambda$.) □

5 The proof of Lemma 3.5

5.1 A lemma about disjoint sets of variables

In this section we prove Lemma 3.5 using Lemma 9 of [4]. (This is the most important technical lemma of that paper with a long proof.) We reformulate below this result from [4] as Lemma A1 to make it consistent with the terminology of the present paper. In the proof we will also use other lemmata from [4]; we will formulate them as Lemma A2, Lemma A3, and Lemma A4 in the present paper. These latter three lemmata have short proofs (given in [4]) using only the definitions of the concepts contained in their statements. The following definitions are needed for the statement of Lemma A1.

Definitions.

1. Assume that $\mathcal{B}$ is a branching program with n input variables and η is an input for $\mathcal{B}$. (Recall that an input is a $\{0,1\}$-valued function defined on $\{0,1,\ldots,n-1\}$ with the meaning that the value $\eta(i)$ is assigned to the variable x_i.) At input η the branching program follows a path in the directed graph $\mathcal{G}$ as described in the definition of a branching program. We associate a time (a nonnegative integer) with each node of this path. If the path is $v_0,v_1,\ldots,v_i$, where v_0 is the source node and v_i is a sink node, then for all integers $t\in[0,1]$, we will say that the program is *at node v_t at time t* with respect to input η. We will use the notation $\mathtt{state}(t,\eta)=v_t$.
2. Assume that $\mathtt{state}(t,\eta)=v_t$, and $\mathtt{var}(v_t)=x_i$. In this case we will say that the program *accesses* the variable x_i at time t.
3. An input η is *visible* if each variable x_i, $i=0,1,\ldots,n-1$ is accessed at some time during the computation performed at input η.

4. Assume that $\mathcal{B}$ is a branching program so that the path associated with each input is of length l. In this case we will say that for each input the *length* of the program is l.

Additional assumptions about $\mathcal{B}$. Without loss of generality we will make the following two additional assumptions about the branching program $\mathcal{B}$ in the proof of Lemma 3.5.

(a) We assume that every input is visible. Indeed, we can modify the branching program $\mathcal{B}$ so that the new branching program $\mathcal{B}'$ first reads the value of each variable and then continues with the original computation of $\mathcal{B}$. The length and the size of the program are thus increased only by n. Moreover, if Lemma 3.5 holds for $\mathcal{B}'$ then clearly it also holds for $\mathcal{B}$ since, apart from the size and the depth of the program, Lemma 3.5 treats the program as a black box, it speaks only about the function defined by the branching program.

(b) We assume that, independently of the input, the length of the branching program is exactly kn, that is, for each input η the program reaches a sink node at time kn. This is not an essential restriction because there is another program $\mathcal{B}'$, whose size is larger than the size of $\mathcal{B}$ by only a factor of at most n^2, so that program $\mathcal{B}'$ works exactly the same way as program $\mathcal{B}$ but also counts the time and when $\mathcal{B}$ reaches a sink node v then it works further till time kn when it gives the same output as the output of $\mathcal{B}$ at node v. (We may assume that at each time t in this new additional time interval, the branching program $\mathcal{B}'$ accesses, e.g., the variable x_0.)

As a consequence of this second assumption, for each fixed input η, the function $\mathtt{state}(t,\eta)$ is defined for all $t = 0,1,\ldots,nk$ and the branching program accesses a variable at each time t for $t = 0,1,\ldots,kn-1$.

Definitions.

1. Suppose that σ is a real number with $\sigma \in (0,\frac{1}{2})$. We assume that a partition of the set $\{0,1,\ldots,kn-1\}$ into intervals is fixed so that the length of each interval is between σn and $2\sigma n$. $\mathcal{I}^{(\sigma)}$ will denote the set of these intervals. If the choice of σ is clear from the context, we will omit the superscript σ.
2. Assume that $T \subseteq \{0,1,\ldots,kn-1\}$ is a set of integers. The set of all integers $i \in \{0,1,\ldots,n-1\}$, so that the input variable x_i is accessed by the branching program at some $t \in T$, at input η, will be denoted by $\mathtt{register}(T,\eta)$. The set of all integers j in $\mathtt{register}(T,\eta)$ so that the value of variable x_j is not accessed at any time outside T at input η will be denoted $\mathtt{core}(T,\eta)$. Clearly $\mathtt{core}(T,\eta) \subseteq \mathtt{register}(T,\eta)$.

Remark 5.1. The notation $\mathtt{register}(\eta)$ was motivated by the fact that, in [4], instead of branching programs we work with random access machines, and so instead of reading the values of variables the machine reads the content of registers. To make the two papers more compatible we did not change this notation.

Definitions.

1. If a $\sigma > 0$ is given, $F \subseteq \mathcal{I}^{(\sigma)}$, and χ is an input, then $\mathtt{stem}(F,\chi)$ will denote the restriction of χ onto $\{0,1,\ldots,n-1\}\backslash\mathtt{core}(F,\chi)$.
2. Suppose that $T \subseteq \{0,\ldots,kn-1\}$. We say that x is at the *right border* of T if $x \notin T$ and $x-1 \in T$. The set of those integers which are at the right border of T will be denoted by $\mathtt{right}(T)$.

3. Suppose that $T \subseteq \{0,\ldots,kn-1\}$ and χ is an input. Let f be a function defined on the set $\mathtt{right}(T)$, so that for all $t \in \mathtt{right}(T)$ we have $f(t) = \mathtt{state}(t,\chi)$. We will call f the *right-state* function of the set T at input χ and will denote it by $\mathtt{rstate}_{T,\chi}$.

Remark 5.2. The significance of the set $\mathtt{core}(T,\chi)$, the right border of T, and the function $\mathtt{right}_{T,\chi}$ is the following. Assume that starting from the input χ we change the value of some of the variables in $\mathtt{core}(T,\chi)$ in a way that for the new input χ' we have $\mathtt{right}_{T,\chi} = \mathtt{right}_{T,\chi'}$. Then the output of the program remains unchanged.

Lemma A1. For all positive integer k, if $\sigma > 0$ is sufficiently small with respect to k, $\varepsilon > 0$ is sufficiently small with respect to σ, n is sufficiently large with respect to ε, $\mathcal{B}$ is a branching program with n input variables, $\mathcal{B}$ is of size at most $2^{\varepsilon n}$, for each input the length of the program is kn, and G is a set of visible inputs, then the following holds. There exist $\kappa > \sigma$, F_1, F_2, f_1, f_2, H satisfying the following conditions:

(9). $H \subseteq G$ and $|H| \geq 2^{-\kappa n}|G|$

(10). F_1, F_2 are disjoint subsets of $\mathcal{I}^{(\sigma)}$

(11). for all $i = 1,2$ and $j = 3-i$ if $\chi, \xi \in H$, and $\mathtt{stem}(F_i,\chi) = \mathtt{stem}(F_i,\xi)$, then $\mathtt{core}(F_j,\chi) = \mathtt{core}(F_j,\xi)$

(12). $|\mathtt{core}(F_i,\chi)| \geq \kappa^{\tau} n$ for all $\chi \in H$ and $i = 1,2$, where $\tau = 1 - \frac{1}{50k}$,

(13). $\mathtt{rstate}_{\chi,\bigcup F_i} = f_i$ for all $\chi \in H$, $i = 1,2$.

(14). $\kappa < 2^{-|\log\sigma|^{\frac{1}{4}}}$

Motivation. The intuitive meaning of Lemma A1 is the following. Suppose that a branching program works in linear time. Then, if we segment the time into intervals of length about σn, it is possible to select two disjoint sets of intervals F_1 and F_2 so that in each of them, for a large number of inputs χ, we access many variables (the ones in $\mathtt{core}(F_{3-i},\chi)$) that are not accessed anywhere else. Moreover the sets F_1 and F_2 can be selected with the additional property described below. If the state of the branching program is fixed at the right borders of F_1 and F_2 (Condition (13)), then $\mathtt{core}(F_1,\chi)$ and $\mathtt{core}(F_2,\chi)$ are independent from each other in the following sense. In order to know what is $\mathtt{core}(F_i,\chi)$, we do not have to know the values of the variables in $\mathtt{core}(F_{3-i},\chi)$ (Condition (11)). This last condition is the crucial part of the lemma, everything else in it can be proved by a simple counting argument.

Remarks.

1. Condition (14) was not included in the original statement of the lemma in [4] but its proof clearly implies it. The exact form of the upper bound on κ is not important for us, any upper bound of the type $\kappa < g(\sigma)$ where $\lim_{x\to\infty} g(x) = 0$ would be sufficient for the proof of Lemma 3.5.

2. We have changed the notation of the original lemma (by substituting κ for λ) to make it more compatible to the notation of Lemma 3.5.

3. The lemma in [4] was originally formulated for random access machines, however in the case when the possible contents of the input registers form a set with two elements, the notion of the random access machines used there is identical to the notion of (2-way) branching programs. Therefore Lemma A1 is a special case of Lemma 9 of [4].

4. There is a slight difference between the notation of the two papers: in [4] an input is a function defined on the set $\{1,\ldots,n\}$ while in the present paper it is defined on $\{0,1,\ldots,n-1\}$.

5. The proof of Lemma 3.5 from Lemma A1 is almost identical to the proof of Theorem 4 of [4] from Lemma 9 of the same paper.

5.2 Reducing Lemma 3.5 to Lemma A1

As we pick the values of the various parameters in Lemma 3.5 we will describe the values of the parameters of Lemma A1 when we use it in our proof.

Assume that k is given (we will apply Lemma A1 with the same value of k). Now we pick σ_1 and σ_2 so that σ_1 is sufficiently small with respect to k and σ_2 is sufficiently small with respect to σ_1. Let $\sigma = 3\sigma_2$. Let $\varepsilon > 0$ be sufficiently small with respect to σ_2, let n be sufficiently large with respect to ε, and let $\mathcal{B}$ be a branching program of length kn (for each input)and size at most $2^{\varepsilon n}$. (ε, $\mathcal{B}$ and n are the same in the two lemmata.) We pick $\delta \in \{0,1\}$ so that $|\mathcal{B}^{-1}(\delta)| \geq 2^{n-1}$. Let $G = \mathcal{B}^{-1}(\delta)$ in Lemma A1. (As we noted earlier we may assume that every input of $\mathcal{B}$ is visible, so G meets this requirement of Lemma A1.) Now we pick $\kappa, F_1, F_2, f_1, f_2, H$ with the properties listed in Lemma A1.

As a first step in the proof of Lemma 3.5 we prove that there is a $\bar{\chi} \in H$ so that for each $i = 1,2$ the following condition is satisfied:

(15)**.** assume that $s_i = |\mathtt{core}(F_i, \bar{\chi})|$ and $\bar{Y}_i$ is the set of all partial inputs η defined on $\mathtt{core}(F_i, \bar{\chi})$ so that $\bar{\chi} \wr \eta \in H$; then $|\bar{Y}_i| \geq \frac{1}{6} 2^{-\kappa n} 2^{s_i}$.

For the proof we use the following two lemmata from [4]. The first one, Lemma A2 is Lemma 10 in [4], the second one Lemma A2, is Proposition 3 in that paper.

Lemma A2. Suppose that $F \subseteq \mathcal{I}$, χ, ξ are inputs with $\mathtt{stem}(F,\chi) \neq \mathtt{stem}(F,\xi)$ and $\mathtt{rstate}_{\chi,\bigcup F} = \mathtt{rstate}_{\xi,\bigcup F}$. Then there is an $x \in \mathtt{domain}(\mathtt{stem}(F,\chi)) \cap \mathtt{domain}(\mathtt{stem}(F,\xi))$ so that $\chi(x) \neq \xi(x)$.

Lemma A3. Assume that $A \subseteq A'$ are finite sets, P is a partition of A, P' is a partition of A', each class of P is contained in a single class of P', and $d = |A||A'|^{-1}$. Then for all $\lambda > 0$, there are at most $\lambda|A|$ elements x of A so that if C, C' are the unique P, P' classes containing x then $|C||C'|^{-1} \leq \lambda d$.

As a first step in proving the existence of a $\bar{\chi} \in H$ so that for all $i = 1,2$ Condition (15) is satisfied, we fix an $i \in \{1,2\}$ and give a lower bound on the number of inputs $\bar{\chi} \in H$ with Condition (15) (with this fixed i). We define a partition $\mathcal{T}_i$ of H in the following way. $\forall \chi, \xi \in H$, ξ, χ belong to the same class iff $\mathtt{stem}(F_i,\chi) = \mathtt{stem}(F_i,\xi)$. It is a consequence of this definition that if χ and ξ do not belong to the same class of $\mathcal{T}_i$, then the functions $\mathtt{stem}(F_i,\chi)$ and $\mathtt{stem}(F_i,\xi)$ must be different (for the fixed value of i). Since the domains of these two functions, that is, $\{0,1,\ldots,n-1\}\backslash\mathtt{core}(F_i,\chi)$ and $\{0,1,\ldots,n-1\}\backslash\mathtt{core}(F_i,\xi)$ are not necessarily identical, in principle it would be possible for the

functions $\mathtt{stem}(F_i,\chi)$ and $=\mathtt{stem}(F_i,\xi)$ to be compatible, that is, to be identical on the intersection of their domains. However Condition (13) of Lemma A1 and Lemma A2 imply that this can never happen, that is,

(16). functions that belong to different classes of $\mathcal{T}_i$ are not compatible.

We will denote by H' the set of all inputs ζ so that there is a $\chi \in H$ with the property that ζ is an extension of $\mathtt{stem}(\chi,H)$. For each fixed $\chi \in H$ let W_χ be the set of all $\zeta \in H'$ so that ζ is an extension $\mathtt{stem}(F_i,\chi)$. Condition (16) implies that the sets W_χ, $\chi \in H$ (we take each of them only once) form a partition $\mathcal{T}'_i$ of H'. Clearly each class of $\mathcal{T}_i$ is contained in exactly one class of $\mathcal{T}'_i$.

We want to apply Lemma A3 with $A \to H$, $A' \to H'$, $P \to \mathcal{T}_i$, $P' \to \mathcal{T}'_i$, and $\lambda \to \frac{1}{3}$. Since, by the definition of G, we have $|G| \geq 2^{n-1}$, Condition (9) of Lemma A1 implies that $|H| \geq 2^{-\kappa n}2^{n-1}$. Obviously $|H'| \leq 2^n$ and so $d = |H||H'|^{-1} \geq \frac{1}{2}2^{-\kappa n}$. Therefore, according to Lemma A3, for at least $\frac{1}{3}|H|$ inputs χ from the set H, the following condition is satisfied: χ belongs to a class in $\mathcal{T}_i$ whose density in the unique class of $\mathcal{T}'_i$ containing it is at most $\frac{1}{3}d \geq \frac{1}{6}2^{-\kappa n}$. Let X_i be the set of all inputs $\chi \in H$ with this property. Since $|X_i| \leq \frac{1}{3}|H|$ for both $i=1$ and $i=2$ we have that $|H\setminus(X_1 \cup X_2)| \geq \frac{1}{3}|H|$. Let $\bar{\chi} \in H\setminus(X_1 \cup X_2)$. The definition of X_i implies that for all $i=1,2$, $\bar{\chi}$ belongs to a class of $\mathcal{T}_i$ whose density in the corresponding class of $\mathcal{T}'_i$ is greater than $\frac{1}{6}2^{-\kappa n}$. Since each class of $\mathcal{T}'_i$ contains exactly 2^{s_i} elements this implies that $\bar{\chi}$ meets the requirements of Condition (15).

Assume that $\bar{\chi}$ is fixed with Condition (15) and $\bar{Y}_i$, $i=1,2$ are the sets defined in the description of that property. We will prove the following:

(17). for all $\eta_i \in \bar{Y}_i$, $i=1,2$ we have $(\bar{\chi} \wr \eta_1) \wr \eta_2 \in \mathcal{B}^{-1}(\delta)$.

For the proof of this fact we use the following lemma which is Lemma 2 in [4]:

Lemma A4. Assume that χ is an input, η_1, η_2 are partial inputs, $T_1, T_2 \subseteq \{0,1,\ldots,nk-1\}$. If χ, η_1, η_2, T_1, and T_2 satisfy the following conditions, then $\mathcal{B}(\chi) = \mathcal{B}((\chi \wr \eta_1) \wr \eta_2)$.

(18). $\mathtt{domain}(\eta_1)$ and $\mathtt{domain}(\eta_2)$ are disjoint.

(19). T_1 and T_2 are disjoint.

(20). for all $i=1,2$ we have $\mathtt{domain}(\eta_i) \subseteq \mathtt{core}(T_i,\chi)$

(21). for all $i=1,2$ we have $\mathtt{rstate}_{T_i,\chi} = \mathtt{rstate}_{T_i,\chi\wr\eta_i}$

(22). for all $i,j \in \{1,2\}$, $i \neq j$ we have $\mathtt{domain}(\eta_i) \cap \mathtt{register}(T_j, \chi \wr \eta_j) = \emptyset$.

To prove that Condition (17) is satisfied by $\bar{\chi}$, we show that the assumptions of Lemma A4 hold with $\chi \to \bar{\chi}$, η_1, η_2, $T_1 \to \bigcup F_1$, and $T_2 \to \bigcup F_2$.

Condition (18). By the definitions of η_i and the function $\mathtt{core}$ we have $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i,\chi) \subseteq F_i$ for $i=1,2$. Condition (10) of Lemma A1 implies that $F_1 \cap F_2 = \emptyset$, so $\mathtt{domain}(\eta_1)$ and $\mathtt{domain}(\eta_2)$ are disjoint.

Condition (19). This is a consequence of Proposition (10) of Lemma A1.

Condition (20). By the definition of η_i this holds with equality.

Condition (21). This follows from $\chi, \bar{\chi} \wr \eta_i \in H$ and Condition (13) of Lemma A1.

Condition (22). Assume that $i,j \in \{1,2\}$, $i \neq j$ are fixed. We have $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i, \bar{\chi})$. Condition (11) of Lemma A1 and the fact $\bar{\chi} \wr \eta_j \in H$ together imply that $\mathtt{core}(F_i, \bar{\chi}) = \mathtt{core}(F_i, \bar{\chi} \wr \eta_j)$. Therefore $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i, \bar{\chi} \wr \eta_j)$. $F_1 \cap F_2 = \emptyset$ so at input $\bar{\chi} \wr \eta_j$ and at times belonging to the set $\bigcup F_j$ we can never access a variable in $\mathtt{core}(F_i, \bar{\chi} \wr \eta_j)$, and consequently

$$\mathtt{domain}(\eta_i) \cap \mathtt{register}(T_j, \bar{\chi} \wr \eta_j)) = \emptyset \ .$$

Since all of the assumptions of Lemma A4 hold, its conclusion must hold as well and so $\bar{\chi}$ satisfies Proposition (17).

We will need the following observation to conclude the proof. Let $\mathtt{core}(F_i, \bar{\chi}) = S_i$. For any $i = 1,2$ and for any $X \subseteq S_i$, there is an $\bar{Y}_i(X) \subseteq \bar{Y}_i$ so that $\eta(x) = \zeta(x)$ for all $\eta, \zeta \in \bar{Y}_i(X)$, $x \in S_i \backslash X$, and $|\bar{Y}_i(X)| \geq \frac{1}{6} 2^{-\kappa n} 2^{|X|}$. Indeed, we may partition the elements of $\bar{Y}_i$ into disjoint classes according to the values of its elements on the set $S_i \backslash X$. Since there are at most $2^{s_i - |X|}$ classes, at least one class must contain at least $2^{-s_i + |X|} |\bar{Y}_i|$ elements. $\bar{Y}_i(X)$ will be such a class. The stated lower bound on $|\bar{Y}_i(X)|$ and the lower bound on $|\bar{Y}_i|$ formulated in Condition (15) imply $|\bar{Y}_i(X)| \geq \frac{1}{6} 2^{-\kappa n} 2^{|X|}$.

By Condition (12) of Lemma A1 we have $|S_i| \geq \kappa^\tau n$ for $i = 1,2$. Let $[\frac{1}{2}\kappa^\tau n] = r$. Let z_i be the rth smallest element of S_i and assume that e.g. $z_1 \leq z_2$. Let W_1 be the set of the r smallest elements of S_1 and let W_2 be the set of the r largest elements of S_2. Let $Y_i = \bar{Y}_i(W_i)$ for $i = 1,2$. According to our previous observation we have

(23). for all $i = 1,2$, $|Y_i| \geq \frac{1}{6} 2^{|W_i| - \kappa n}$.

By the definitions of r, z_i, and W_i, Condition (1) is satisfied by W_1 and W_2. We claim that the other requirements of the lemma are also met by the following choice of the various parameters. We pick two partial inputs $\zeta_1 \in Y_1$, $\zeta_2 \in Y_2$ in an arbitrary way. Let $\chi = (\bar{\chi} \wr \zeta_1) \wr \zeta_2$, $\lambda = 2\kappa$, and $\mu = |W_1| n^{-1} = |W_2| n^{-1}$. ($W_i$, Y_i have already been defined.)

The definitions of $\sigma_1, \sigma_2, \varepsilon$, and δ at the beginning of the proof of Lemma 3.5 show that the requirements of the lemma, stated before the conditions $\lambda \in (\sigma_2, \sigma_1)$, $\mu \in (\sigma_2, \sigma_1)$, are met. $\lambda \in (\sigma_2, \sigma_1)$ is an immediate consequence of $\lambda = 2\kappa$, the inequalities $\sigma < \kappa$, $\kappa \leq 2^{-|\log \sigma|^{\frac{1}{4}}}$, and the fact that we choose $\sigma_2 = \frac{1}{3}\sigma$ so that it is sufficiently small with respect to σ_1.

The fact that $\mu \in (\sigma_2, \sigma_1)$ is a consequence of the following facts: $\tau = 1 - \frac{1}{50k}$ (cf. Condition (12) of Lemma A1), $\mu = |W_i| n^{-1}$, $|W_i| = [\frac{1}{2}\kappa^\tau n]$, $\sigma < \kappa \leq 2^{-|\log \sigma|^{\frac{1}{4}}}$, $\sigma = 3\sigma_2$, and σ_2 is sufficiently small with respect to σ_1. Indeed $\sigma = 3\sigma_2$ is sufficiently small with respect to σ_1 (for a fixed k), and so

$$\mu = \left[\frac{1}{2}\kappa^\tau n\right] n^{-1} \leq \frac{1}{2}\kappa^\tau \leq 2^{-|\log \sigma|^{\frac{1}{4}}(1 - \frac{1}{50k})} < \sigma_1 \ .$$

On the other hand

$$\mu = \left[\frac{1}{2}\kappa^\tau n\right] n^{-1} > \frac{1}{3}\kappa^\tau \geq \frac{1}{3}\sigma^{1 - \frac{1}{50k}} \geq \frac{1}{3}\sigma = \sigma_2$$

and so $\mu \in (\sigma_2, \sigma_1)$.

We have already seen that Condition (1) of Lemma 3.5 is satisfied.

Condition (2) of Lemma 3.5 is a consequence of the definition of μ.
Condition (3). Using the inequality of Condition (23) we get $|Y_i| \geq \frac{1}{6}2^{|W_i|-\kappa n} \geq 2^{|W_i|-\lambda n} = 2^{\mu n-\lambda n}$.
Condition (4). By the definition of $r = |W_i| = \mu n$ we have $\mu n = [\frac{1}{2}\kappa^\tau n]$ and so

$$\mu \geq \frac{1}{3}\kappa^\tau = \frac{1}{3}(\frac{\lambda}{2})^\tau = \frac{1}{3}(\frac{\lambda}{2})^{1-\frac{1}{50k}} \quad .$$

Therefore

$$\mu^{1+\frac{1}{100k}} \geq (\frac{1}{3})^{1+\frac{1}{100k}}(\frac{\lambda}{2})^{(1-\frac{1}{50k})(1+\frac{1}{100k})} \geq 2\lambda \quad .$$

(Here we used that by Condition (17), both κ and $\lambda > 0$ are sufficiently small with respect to k.)

Condition (5) of Lemma 3.5 is a consequence of Condition (17) and the definitions of χ and Y_i. These definitions imply that $(\chi \wr \eta_1) \wr \eta_2 = (\bar{\chi} \wr \eta_1') \wr \eta_2'$ where $\eta_i' = \eta_i \cup \zeta_i|_{S_i - W_i} \in \bar{Y}_i$.

References

[1] * K. ABRAHAMSON: Time-Space Tradeoffs for Branching Programs Contrasted With Those for Straight-Line Programs. In *27th IEEE FOCS*. IEEE Computer Society, 1986. 1.2.2

[2] * K. ABRAHAMSON: Time-Space Tradeoffs for Algebraic Problems on General Sequential Machines. *Journal of Computer and System Sciences*, 43:269–289, 1991. [JCSS:10.1016/0022-0000(91)90014-V]. 1.2.2

[3] * M. AJTAI: A non-linear time lower bound for boolean branching programs. In *40th IEEE FOCS*, pp. 60–70. IEEE Computer Society, 1999. [FOCS:10.1109/SFFCS.1999.814578]. 1.3

[4] * M. AJTAI: Determinism versus Non-Determinism for Linear Time RAMs with Memory Restrictions. *Journal of Computer and System Sciences*, 65:2–37, 2002. [JCSS:10.1006/jcss.2002.1821]. 1.2.3, 1.2.4, 1.2.5, 1.2.6, 2.1, 2.2, 3.2, 5.1, 5.1, 1, 3, 4, 5, 5.2, 5.2

[5] * P. BEAME: General Sequential Time-Space Tradeoff for Finding Unique Elements. *SIAM J. Computing*, 20(2):270–277, 1991. [SICOMP:20/0220017]. 1.2.2

[6] * P. BEAME, M. SAKS, AND T. S. JAYRAM: Time-space tradeoffs for branching programs. *Journal of Computer and System Sciences*, 63(4):542–572, 2001. [JCSS:10.1006/jcss.2001.1778]. 1.2.3, 1.2.5, 1.2.6, 2.1, 3.3

[7] * P. BEAME, M. SAKS, X. SUN, AND E. VEE: Time-space tradeoff lower bounds for randomized computation of decision problems. *Journal of the ACM*, 50(2):154–195, 2003. [JACM:10.1145/636865.636867]. 1.3

[8] * A. BORODIN AND S. A. COOK: A time-space tradeoff for sorting on a general sequential model of computation. *SIAM J. Computing*, 11:287–297, 1982. [SICOMP:11/0211022]. 1.2.2, 1.2.6

[9] * A. BORODIN, A. A. RAZBOROV, AND R. SMOLENSKY: On lower bounds for read-k-times branching programs. *Computational Complexity*, 3:1–18, 1993. 1.2.5, 1.2.6, 2.1, 3.3

[10] * B. CHOR AND O. GOLDREICH: Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. In *26th IEEE FOCS*, pp. 429–442. IEEE Computer Society, 1985. 1.2.5, 3.3

[11] * M. KARCHMER: Two Time-Space Tradeoffs for Element Distinctness. *Theoretical Computer Science*, 47:237–246, 1986. [TCS:10.1016/0304-3975(86)90150-7]. 1.2.2

[12] * S. REISCH AND G. SCHNITGER: Three applications of Kolmogorov-complexity. In *23rd IEEE FOCS*, pp. 45–52. IEEE Computer Society, 1982. 1.2.2

[13] * J. S. THATHACHAR: On separating the read-k-times branching program hierarchy. In *30th ACM STOC*, pp. 653–662. ACM, 1998. [STOC:10.1145/276698.276881]. 1.2.5, 1.2.6, 2.1, 3.3, 3.12

[14] * Y. YESHA: Time-Space Tradeoffs for Matrix Multiplication and the Discrete Fourier Transform of Any General Sequential Random-Access Computer. *Journal of Computer and System Sciences*, 29:183–197, 1984. [JCSS:10.1016/0022-0000(84)90029-1]. 1.2.2

AUTHOR

Miklós Ajtai
IBM Almaden Research Center
ajtai@almaden.ibm.com

ABOUT THE AUTHOR

MIKLÓS AJTAI received his Ph. D. from the Hungarian Academy of Sciences in 1975. His advisor was András Hajnal. He worked in the following areas: axiomatic set theory (independence proofs), lattice theory (posets with meet and join), combinatorics, the theory of random graphs, complexity theory, sorting networks, the theory of lattices (n-dimensional grids) and their applications to complexity theory and cryptography. He is a member of the Hungarian Academy of Sciences and was an invited speaker at ICM in 1998. He received the Knuth prize in 2003, and the IBM Corporate Award in 2000.