Electronic Colloquium on Computational Complexity, Report No. 26 (1999)

A Non-linear Time Lower Bound for Boolean Branching Programs

(Preliminary version)

Miklós Ajtai
IBM Almaden Research Center

Abstract. We prove that for all positive integer k and for all sufficiently small $\epsilon > 0$ if n is sufficiently large then there is no Boolean (or 2-way) branching program of size less than $2^{\epsilon n}$ which for all inputs $X \subseteq \{0, 1, ..., n-1\}$ 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 "$\forall i, j, k, l \in \{0, 1, ..., n-1\}$, $i + j = k + l$ implies $a_{i,j} = a_{k,l}$" then with a 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.

Introduction. A Boolean (or 2-way) branching program is a finite directed acyclic grap with a unique source node, so that each non-sink node is labeled by one of the input variables $x_0, ..., x_{n-1}$, 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 on nodes in the graph is the size of the branching program.

This model describes a very general way of computation where the computational time measures of the number of accesses to the individual bits of the input, 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 containg $\log_2 M$ bits where M is the size of the branching program.

ISSN 1433-8092

We give an explicit function, described in the abstract, which cannot be computed with a Boolean branching program in linear time if the size of the branching program is $2^{\epsilon n}$. The best previously known result in this direction is a lower bound given by Beame, Saks and Thathachar for the computation of an explicitly given function, namely they show that there is an $\epsilon > 0$ 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 elemts) cannot be decided with a branching program of length $(1+\epsilon)n$ and of size $2^{O(n)}$. (The proof shows that the theorem holds for $\epsilon = .0178$.) For multi-output functions severaltime space tradoffs were known already for a long time, see e.g [BC], [Bea].

A generalization of the Boolean branching program is the R-way branching program where each input variable takes its values from a set Γ of size R and each node in the graph corresponding to the branching program has R outgoing edges each by a different element of Γ. The output, and length of the program defined in the natural way. In [BST] a nonlinear lower bound is given, on the length of an R-way branching program computing an explicitly defined function, (similar to the function 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 $g(x_1, ..., x_n)$ of n variables such that: (a) each variable is taking 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 $g(x_1, ..., x_n)$ in depth kn.

If $R = c \log n$ then the R-way branching program corresponds to the random access machine whose input registers contain $c \log_2 n$ bits. It has been proved in [A] that the element distinctness problem (where each "element" is the content of a register), 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 $2^{\epsilon n}$, 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 $2^{\epsilon n \log_2 n}$.) These proofs are based on the ananlysis of certain combinatorial properties of the input, which are very similar to the combinatorial properties used in [BTS].

Our proof in the present paper uses the technical lemmata of the element distinctness results. Namely, it is shown in [A] 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 the values of the variables in W_i only, 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 1 below.)

We use this result to show that a qaudratic form (which is NOT given explicitly) cannot be computed in linear time. The algebraic part of this proof (Lemma 5) is a theorem proved in [BRS] (and in more general forms in [Tha] and [BST]). We reduce the problem of giving a quadratic form with the required properties to the question about the ranks of the submatrices (or minors) of the matrix generating the quadratic form, in a similar way as it is done in [BST]. In both cases the goal is to get a matrix A so that each $[\delta n]$ by $[\delta n]$ 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 [BST] 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.) We will call an $n \times n$ matrix $A = (a_{i,j})$ a Hankel matrix if "$\forall i, j, k, l \in \{0, 1, ..., n-1\}$, $i+j = k+l$ implies $a_{i,j} = a_{k,l}$". 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 a 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 [A], we are able to show that if $\tilde{A}$ is the marix that we get from A by replacing each entries 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 $2^{\epsilon n}$. 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 qaudratic form.) In other words if $A(y)$, $y = \langle y_0, ..., y_{2n-2}\rangle$ denotes the Hankel matrix whith $a_{i,j} = y_{i+j}$, then $\langle \tilde{A}(y)x, x\rangle$ cannot be computed in the given length and size from the input $\langle x, y\rangle$. 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 $x = \langle x_0, ..., x_{n-1}\rangle$ and $X = \{i | x_i = 1\}$, and $D = \{i + j | a_{i,j} = 1, i, j \in \{0, ..., n-1\}\}$. 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 $i + j \in D$. By encoding the set D by a part of the input we will be able to show that the problem "compute 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 solved by a branching program of linear length and of size at most $2^{n\epsilon}$.

Finally we note that our results about random Hankel matrices remain true over any fields with appropriate modifications. (See the remarks after Lemma 7, Lemma 8 and Lemma 9. See also a comment, about the applicability of these modified version for generalizations of Theorem 1, in the proof of Lemma 5.)

1. 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.

Notation. If X, Y are sets then $\mathrm{Func}(X, Y)$ will denote the set of all functions defined on X, whose values are 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 are taking their values from a set of size 2.

Definition. A branching program $\mathcal{B}$ with n input variables $x_0, ..., x_{n-1}$ 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,

(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 of variables $\{x_0, ..., x_{n-1}\}$,

(d) `out` is a function defined the 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 have 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 $0, 1$-evaluation of the variables x_i. (Instead of such an evaluation we ususally will think about an input as a $0, 1$-valued function η defined on $\{0, 1, ..., n-1\}$ 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. Assume that X is a subset of $\{0,...,n-1\}$. $N_+(X)$ will denote the number of all pairs $x, y \in X$, $x < y$ so that $x + y \in X$.

Theorem 1. *For all positive integer k, if $\epsilon > 0$ 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 $2^{\epsilon n}$, which for all inputs η computes the parity of $N_+(X_\eta)$ where $X_\eta = \{i \in \{0,1,...,n-1\} | \eta(i) = 1\}\}$*

In the proof we will use the following lemma, Lemma 1, which is a consequence of Lemma A of [A] (called Lemma 9 in that paper) as we will explain below. The reader who is not familiar with the details of the proofs in [A] may accept the statement of Lemma 1 without a proof. (In this case Lemma A is not needed.) The remaining part of the paper, strating with Lemma 2 is self contained. We give first the definitions needed to understand the statement of Lemma 1.

Definition. 1. An input (of a branching problem with n input variables) is a function χ defined on $\{0,1,...,n-1\}$ with values in $\{0,1\}$. A partial input is a function η defined on a subset of $\{0,1,...,n-1\}$ 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 $\mathcal{H}(\mathcal{B}, \delta)$ will denote the set of all inputs η so that the output of $\mathcal{B}$ at input η is δ.

Lemma 1. *For all positive integer k if $\sigma_1 > 0$ is sufficiently small with respect to k, $\sigma_2 > 0$ is sufficiently small with respect to σ_1, $\epsilon > 0$ is sufficienlty 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 $2^{\epsilon n}$, $\delta \in \{0,1\}$ so that $|\mathcal{H}(\mathcal{B}, \delta)| \geq 2^{n-1}$, then there exist a $\chi \in \mathcal{H}(\mathcal{B}, \delta)$, $\lambda \in (\sigma_2, \sigma_1)$, $\mu \in (\sigma_2, \sigma_1)$, $W_i \subseteq \{0,1,...,n-1\}$, $i = 1,2$, and sets of partial inputs Y_i, $i = 1,2$ defined on W_i, so that*

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

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

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

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

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

Proof. We start with the following lemma proved in [A]. Originally it was formulated for random access machines, however in the case when the possible contents of the input registers form a set with two elements (that is, $\alpha = 2$ with the terminology

used there), then the notion of random access machines is identical to the notion of (2-way) branching programs. In the main theorem of [A] about the element distinctness problem, we assumed that $\alpha > n^2$. However this assumption was not used for all of the lemmata in its proof so we are able to use a lemma from this proof without any modification. (There is a slight difference between the notation of the two papers: in [A] an input is a function defined on the set $\{1, ..., n\}$ while in the present paper it is defined on $\{0, 1, ..., n-1\}$.)

Definition. Assume that $\mathcal{B}$ is a branching program with n input variables. An input η is visible if each variable x_i, $i = 0, 1, ..., n-1$ occurs as $\mathtt{var}(e)$ for some edge e on the path determined by the computation at η.

Remark. 1. For the proof of Theorem 1 we may assume that every input is visible, since it is easy to modify a branching program in a way that first the program reads the value of each variables and then it continues with the original computation. This way the length and the size of the program is increased only by n. Therefore we may assume throughout the proof of the lemma that every input is visible.

For the understanding of the statement of Lemma A below, the following concepts, defined in [A], are needed: $\mathcal{I}$ a partition of the time interval, $\mathtt{core}(F, \chi)$, $\mathtt{stem}(F, \chi)$, $\mathtt{rstate}_{\chi, F}$.

Lemma A. *For all positive integer k, if $\sigma > 0$ is sufficiently small with repsect to k and $\epsilon > 0$ 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, of length at most kn, and of size at most $2^{\epsilon n}$, 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 with the following properties:*

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

(7) F_1, F_2 *are disjoint subsets of* $\mathcal{I}$

(8) *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* $\mathtt{core}(F_j, \chi) = \mathtt{core}(F_j, \xi)$

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

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

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

Remarks. 1. Property (11) was not included in the original statement of the lemma in [A] but its proof clearly implies it. The exact form of the upper bound on κ is not important for us, we will use only that $\kappa < \sigma_1$ for some $\sigma_1 > 0$ which is sufficiently small with respect to k.

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 1.

We may prove Lemma 1 from Lemma A in a similar way as Theorem 4 of [A] was proved by using Lemma A. Some parts of the proofs are almost identical, so in this case we will refer only to the proof in [A]. As we pick the values of the various parameters in Lemma 1 we will tell what will be the values of the parameters of Lemma A when we use it to prove Lemma 1.

Assume that k is given (we will apply Lemma A 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 $\epsilon > 0$ be sufficiently small with respect to σ_2 and n be sufficiently large with respect to ϵ and let $\mathcal{B}$ be a branching program of length at most kn and of size at most $2^{\epsilon n}$. (ϵ, $\mathcal{B}$ and n are the same in the two lemmata.) We pick $\delta \in \{0,1\}$ so that $|\mathcal{H}(\mathcal{B},\delta)| \geq 2^{n-1}$. (We will write $\mathcal{H}$ for $\mathcal{H}(\mathcal{B},\delta)$.) Let $G = \mathcal{H}$ in Lemma A. Now we pick κ, F_1, F_2, f_1, f_2, H according to Lemma A.

With the same argument as in the proof of Theorem 4 in [A] (the only change is that we replace c_0 by $\frac{1}{2}$) we get that there is a $\bar{\chi} \in H$ so that

(12) *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* $|\bar{Y}_i| \geq \frac{1}{6} 2^{-\kappa n} 2^{s_i}$.

(This is the analogue of (28) in [A].) The same way as it is done in [A] (using Lemma 2 of [A]) at the end of the proof of Theorem 4, we may also prove that

(13) *for all* $\eta_i \in \bar{Y}_i$, $i = 1,2$ *we have* $(\bar{\chi} \wr \eta_1) \wr \eta_2 \in \mathcal{H}$.

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.

By (9) 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 e.g. that $z_1 \leq z_2$. 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 $|Y_i| \geq \frac{1}{6} 2^{|W_i| - \kappa n}$. 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 partitions $\zeta_1 \in Y_1$, $\zeta_2 \in Y_2$ in an arbitrary way. Let $\chi = (\bar{\chi} \wr \zeta_1) \wr \zeta_2$, $\lambda = 2\kappa$, $\mu = |W_1| n^{-1} = |W_2| n^{-1}$, ($W_i$, Y_i have been already defined).

We have already seen that (1) is satisfied.

(2) is a consequence of the definition of μ and the following facts: $|W_i| = [\frac{1}{2}\kappa^\tau]$, $\kappa \leq 2^{-|\log \sigma|^{\frac{1}{4}}}$, $\sigma = 3\sigma_2$ and σ_2 is sufficienlty small with respect to σ_1.

(3). $|Y_i| \geq \frac{1}{6} 2^{|W_i| - \kappa n} \geq 2^{|W_i| - \lambda n} = 2^{\mu n - \lambda n}$.

(4). By the definition of $r = |W_i| = \mu n$ we have $\mu n = [\frac{1}{2}\kappa^\tau n]$ and so $\mu \geq \frac{1}{3}\kappa^\tau = \frac{1}{3}(\frac{\lambda}{2})^\tau = \frac{1}{3}(\frac{\lambda}{2})^{1-\frac{1}{50k}}$. Therefore $\mu^{1+\frac{1}{100k}} \geq (\frac{1}{3})^{1+\frac{1}{100k}}(\frac{\lambda}{2})^{(1-\frac{1}{50k})(1+\frac{1}{100k})} \geq 2\lambda$. (Here we used that by 13, both κ and $\lambda > 0$ are sufficiently small with respect to k.)

(5) is a consequence of (13) and the defintions 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 $\eta_i' = \eta_i \cup \zeta_i|_{S_i - W_i} \in \bar{Y}_i$. *Q.E.D.*(Lemma 1).

Definitions. 1. 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 $q = 1, ..., n$ and for each q by q submatrix B of A we have that the rank of B is at least $f(\frac{q}{n})n$.

2. Suppose that $A = (a_{i,j})_{i=0,j=0}^{n-1}$ is an n by n matrix over the field F. We say that A is a Hankel matrix if for all $i, j, k, l \in \{0, ..., n-1\}$, $i + j = k + l$ implies $a_{i,j} = a_{k,l}$.

The proof of Theorem **1** is based on the following two lemmata.

Lemma 2. *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, $\epsilon > 0$ is sufficiently small with respect to σ_2 and n is sufficiently large with repsect 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 $x \in (\sigma_2, \sigma_1)$ and $f(x) = 0$ otherwise. If there is an n by n Hankel matrix A over F_2 so that A is f-rigid over F_2, then there is no branching program $\mathcal{B}$ with n inputs, of length at most kn and of size at most $2^{\epsilon n}$, which for all inputs η decides the parity of $N_+(X_\eta)$ where $X_\eta = \{i \in \{0, 1, ..., n-1\} \mid \eta(i) = 1\}$*

Lemma 3. *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.*

We will prove Lemma **3** in the next section, more precisely we will prove (Theorem 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 $\{0, 1, ..., n-1\}$. u_η will denote the n-dimensional vector $\langle \eta(0), ..., \eta(n-1) \rangle$

2. The inner products of the n-dimensional vectors u, v will be denoted by $\langle u, v \rangle$.

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 entries of A below the main diagonal and replacing all of the 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 $i \leq j$, $i = 1, ..., n$, $j = 1, ..., n$.

Lemma 4. *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, $\epsilon > 0$ is sufficiently small with respect to σ_2 and n is sufficiently large with repsect 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 $2^{\epsilon n}$, which for all inputs η computes $\tilde{A}\langle u_\eta, u_\eta\rangle$.*

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

Proof. Assume that contrary to our statement there is a branching program $\mathcal{B}$ with the given properties which computes $\langle \tilde{A}u_\eta, u_\eta\rangle$. We apply Lemma 1 with the given values of $k, \sigma_1.\sigma_2, \epsilon, n$ and with the given $\mathcal{B}$. According to Lemma 1 there exist $\delta \in \{0,1\}$, $\chi \in \mathcal{H}(\mathcal{B}, \delta)$, $\lambda, \mu \in (\sigma_2, \sigma_1)$, $W_i, Y_i, i = 1,2$ with the properties listed in Lemma 1. Let $v = \langle v_0, ..., v_{n-1}\rangle$ 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 $v_i = 0$. We define a vector $w^{(\xi)} = \langle w_0^{(\xi)}, ..., w_{n-1}^{(\xi)}\rangle$ 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 $\xi \in Y_i$, $g_i(\xi) = \langle \tilde{A}(v + w^{(\xi)}), v + w^{(\xi)}\rangle$. 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 1, $\eta \in H$ and therefore $\langle \tilde{A}u_\chi, u_\chi\rangle = \langle \tilde{A}u_\eta, u_\eta\rangle = \langle \tilde{A}(v + w^{(\xi_1)} + w^{(\xi_2)}, v + w^{(\xi_1)} + w^{(\xi_2)}\rangle = -\langle \tilde{A}v, v\rangle + g_1(\xi_1) + g_2(\xi_2) + \langle \tilde{A}w^{(\xi_1)}, w^{(\xi_2)}\rangle + \langle \tilde{A}w^{(\xi_2)}, w^{(\xi_1)}\rangle$. $\langle \tilde{A}u_\chi, u_\chi\rangle$ and $\langle \tilde{A}v, v\rangle$ 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 $\langle \tilde{A}w^{(\xi_1)}, w^{(\xi_2)}\rangle + \langle \tilde{A}w^{(\xi_2)}, w^{(\xi_1)}\rangle$ as a function of ξ_1, ξ_2 is also constant on $Y_1' \times Y_2'$. (1) and the definition of $\tilde{A}$ implies that $\langle \tilde{A}w^{(\xi_2)}, w^{(\xi_1)}\rangle$ is identically 0 on $Y_1' \times Y_2'$, therefore $\langle \tilde{A}w^{(\xi_1)}, w^{(\xi_2)}\rangle$ is constant on $Y_1' \times Y_2'$. Let V_0 be the vectorspace all F_2 valued functions defined on $\{0, ..., n-1\}$, and let V_i, $i = 1,2$ 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 $\langle \tilde{A}w^{(\xi_1)}, w^{(\xi_2)}\rangle = \langle B\xi_1, \xi_2\rangle$. 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 (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 **5** (below) with V_1, V_2, $m \to \mu n$, $X \to Y_1'$, $Y \to Y_2'$ and B. (3) implies that

$|Y_i'| \geq \frac{1}{2}|Y_i| \geq 2^{\mu n - \lambda n - 1}$. Therefore, according to Lemma 5, the fact that $\langle Bx, y\rangle$ 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 (4), $\mu^{1+\frac{1}{100k}} > 2\lambda$. *Q.E.D.*(Lemma 4)

The following lemma in more general forms are proved in [BRS], [Tha], [BST]. To make the paper more self contained we provide here a proof.

Lemma 5. *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 $\langle Bx, y\rangle$, $x \in X$, $y \in Y$ is not constant on $X \times Y$.*

Let x_0 be an arbitrary but fixed element of X and let $X' = \{x - x_0 | x \in X\}$. Clearly $|X| = |X'|$ and if $\langle Bx, y\rangle$ is constant on $X \times Y$ then $\langle Bx, y\rangle$ is identically 0 on $X' \times Y$. Therefore it is enough to prove that the assumptions of the lemma imply that $\langle Bx, y\rangle$ 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 $\langle BH, G\rangle = 0$, 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 $\dim(H) - (m - r) + \dim(G) \leq m$. 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: $\dim(H) \geq m_1$, $\dim(G) \geq m_2$. This simply follows from the fact that a d-dimensional subspace has 2^d elements. (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 suitably chosen basis of V_1 and V_2. See [Tha] Lemma 7. This is important for the generalization of Theorem 1 for fields with other characteristics.) We have $m_1 - (m - r) + m_2 \leq m$, that is, $m_1 + m_2 - r \leq 2m$ in contradiction to our assumption. *Q.E.D.*(Lemma 5)

Proof of Theorem 1. 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 $m = [\frac{n}{10}]$. We apply Lemma 4 with $n \to m$, $k \to ck$ where c is a sufficiently large absolute constant and $\epsilon \to \frac{\epsilon}{2}$. Assume that σ_1, σ_2 are picked with the properties described in the lemma.

Let g be the function defined in Lemma 3. Applying Lemma 3 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 4 there is no branching program of size at most $2^{\frac{\epsilon}{2}n}$ which computes $\langle u_\zeta \tilde{A}, u_\zeta\rangle$ in time ckn, for all ζ, where ζ is an F_2 valued function defined on $\{0, 1, ..., m-1\}$. Let $D = \{i + j |\ a_{i,j} = 1\}$, $X_\zeta = \{i \in \{0, 1, ..., m-1\} | \zeta(i) = 1\}$. For any pair of

sets of integers X, Z let $N_+(X, Z)$ the number of pairs x, y, $x < y$ so that $x \in X$, $y \in X$ and $x + y \in Z$. The satement of Lemma 4 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 4. 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 $\mathrm{Func}(\{0, 1, ..., n-1\}, \{0, 1\})$ so that $X_\eta = U_1 \cup U_2$. Clearly "$x, y \in X_\zeta$, $x < y$ and $x + y \in D$" implies that "$2m + x \in X_\eta, 2m + y \in X_\eta$, $2m + x < 2m + y$ 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 $z, w \in \{2m, ..., 3m - 1\}$ and therefore $z - 2m, w - 2m \in X_\zeta$, $z - 2m < w - 2m$ and $(z - 2m) + w(-2m) \in X_\zeta$. therefore $N_+(X_\zeta, D) = N_+(X_\eta)$. Morever each value of η can be computed by in constant time by a branching program in (the size of the program must be increased only by a factor of two, since the extra memory needed for this step is only one bit). *Q.E.D.*(Theorem 1)

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

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

Theorem 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 $A = \{a_{i,j}\}$, $i = 0, ..., n-1$, $j = 0, ..., n-1$ 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:*

(14) Suppose $S = \{s_0, ..., s_{q-1}\}$, $T = \{t_0, ..., t_{q-1}\}$ are subsets of $\{0, ..., n-1\}$ with q elements, where $c_2 n < q$, and $B = (a_{s_i, t_j})$, $i = 0, ..., q-1$, $j = 0, ..., q-1$ 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 is at least $c_1 |\log(\frac{q}{n})|^{-2} q$.

Lemma 6. *Assume that $t > 10$ 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 $|U' + V'| \geq \frac{1}{4} t^2$.*

Proof. We will pick pairs u_i, v_i, $u_i \in U$, $v_i \in V$ sequentially and we will have $U' = \{u_1, ..., u_t\}$, $V' = \{v_1, ..., v_t\}$. Assume that $U_i = \{u_1, ..., u_i\}$, $V_i = \{v_1, ..., v_i\}$. We will pick u_i, v_i so that the number $|U_i + V_i| - |U_{i-1} + V_{i-1}|$ is maximal for $i = 2, ..., t$. (u_1, v_1 are arbitrary.) We show that this definition implies that for all $i > \lceil \frac{t}{2} \rceil + 1$ we have that either $|U_{i-1} + V_{i-1}| \geq \frac{1}{4}t^2$ or $|U_i + V_i| - |U_{i-1} + V_{i-1}| > \frac{3t}{8}$ which clearly implies the lemma. Indeed, assume that $i > \lceil \frac{t}{2} \rceil + 1$ but $|U_{i-1} + V_{i-1}| < \frac{1}{4}t^2$. This latter inequality implies that for each fixed u_j, $j = 1, ..., i-1$ there are at least $\frac{3}{4}t^2$ elements v of V so that $u_j + v \notin U_{i-1} + V_{i-1}$. Let W_j be the set of all such elements v. Clearly $W_j \subseteq V \backslash V_{i-1}$. We have that $\sum_{j=1}^{i-1} |W_j| \geq (i-1)\frac{3}{4}t^2 \geq \frac{t}{2}\frac{3}{4}t^2 \geq \frac{3}{8}t^3$. Therefore there is a $v \in V \backslash V_{i-1}$ which is contained in at least $\frac{3}{8}t^2$ sets W_j, and so $|(v + U_{i-1}) \backslash (U_{i-1} + V_{i-1})| \geq \frac{3}{8}t$. Let $v_i = v$ and let u_i be an arbitrary element of $U \backslash U_{i-1}$. (Choosing u_i in a similar way as w_i we may improve the constant $\frac{1}{4}$ in the conclusion of the lemma.) We have that $|U_i + V_i| - |U_{i-1} + V_{i-1}| \geq |(v_i + U_{i-1}) \backslash (U_{i-1} + V_{i-1})| \geq \frac{3}{8}t$. *Q.E.D.*(Lemma 6)

Definitions. 1. func$(n, 2)$ will be the set of all functions defined on $\{0, ..., n-1\}$ with values in F_2. func$([l, n), 2)$ will denote the set of all functions defined on the interval $[l, n) = \{l, ..., n-1\}$ with values in F_2.

2. Assume that n_1, n_2 are positive integers, $f \in$ func$(n_1 + n_2 - 1, 2)$. Then diag(f, n_1, n_2) will be the n_1 by n_2 matrix $(d_{i,j})$, $i = 0, ..., n_1 - 1$, $j = 0, 1, ..., n_2 - 1$, where $d_{i,j} = f(i + j)$.

3. Assume that n_1, n_2, k_1, k_2, are positive integers, $n_1 > k_1$, $n_2 > k_2$ $f \in$ 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 $A = (a_{i,j})$, $i = 0, ..., n_1 - 1$, $j = 0, 1, ..., n_2 - 1$, is an n_1 by n_2 matrix and $S \subseteq \{0, 1, ..., n_1 - 1\}$, $T \subseteq \{0, 1, ..., n_2 - 1\}$. 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 7. *Assume that n_1, n_2, k_1, k_2 are positive integers, $k_1 < n_1$, $k_2 < n_2$, f is a function on $\{0, 1, ..., k_1 + k_2 - 1\}$ with values in F_2, $S \subseteq \{0, 1, ..., n_1 - 1\}$, $T \subseteq \{0, 1, ..., n_2 - 1\}$ and $|(S \cap \{k_1, ..., n_1 - 1\}) + (T \cap \{k_2, ..., n_2 - 1\})| \geq m$. Then with a probability of 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* sub$(\Phi(n_1, n_2, f), S \cap \{0, 1, ..., k_1 - 1\}, T \cap \{0, 1, ..., k_2 - 1\})$.

Remark. 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. Let $\Phi(n_1,n_2,f)=(\varphi_{i,j})$, $i=0,...,n_1-1$, $j=0,...,n_2-1$. For each $j=0,1,2,...$ let $S_j=S\cap\{0,1,...,j\}$, $T_j=T\cap\{0,1,...,j\}$. 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 $\mathrm{sub}(\Phi(n_1,n_2,f),S_{k_1-1},T_{k_2-1})$. 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 $W=\{w_{i,k_2-1}|i\in\bar{S}\}$ 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 $i\in S\cap\{k_1,...,n_1-1\}$ and $j\in T\cap\{k_2,...,n_2-1\}$ 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 of 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 $\mathrm{sub}(\Phi(n_1,n_2,f),S,T)$ is greater than $|\bar{S}|=r$.) Before the randomization of $g(x)$ the funcion $w_{i,j}$ 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 so the for the randomization of $g(x)$ we have that $P(w_{i,j}=u)=P(w_{i,j}=v)=\frac{1}{2}$. 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. *Q.E.D.*(Lemma 7)

Lemma 8. *Assume that for each $j=1,2$, $I_1^{(j)},...,I_l^{(j)}$, is a partition of the interval $[0,...,n)$ into pairwise disjoint subintervals, $S^{(1)},S^{(2)}\subseteq\{0,1,...,n-1\}$ and $|(S^{(1)}\cap I_i^{(1)})+(S^{(2)}\cap I_i^{(2)})|\geq m_i$ for all $i=1,...,l$. 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* $\sum\{2^{-m_{i_1}-...-m_{i_r}}|1\leq i_1<...<i_r\leq l\}$

Remark. If we define the random Hankel matrix $\Phi(n,n)$ over an arbitrary field F in the way described in the Remark after Lemma 7, then our Lemma remains true if we substitute $|D|^{-m_1-\ldots-m_r}$ for $2^{-m_{i_1}-\ldots-m_{i_r}}$ in the last expression of the Lemma.

Proof. 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 $1 \le i_1 < \ldots < i_r \le l$ 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. Asume that for some $t \in \{1,\ldots,r\}$ the values of $h(0),\ldots,h(b_{i_t,1}+b_{i_t,2})-1$ has 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 7 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 of " $\mathrm{rank}(\Phi_{i_t-1}) = \mathrm{rank}(\Phi_{i_t})$ for all $t=1,\ldots,r$" is at most $2^{-m_1-\ldots-m_r}$. *Q.E.D.*(Lemma 8)

Lemma 9. *Assume that n,q,R,t are positive integers, $t^2<q<n$ and $R<[\frac{q}{t^2}]$. 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:*

(15) 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

$$p \ge 1 - \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-\frac{1}{4}(Q-R+1)t^2}$$

where $Q=[\frac{q}{t^2}]$.

Remark. This lemma remain also 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 7. Namely we have to substitute $|D|^{-\frac{1}{4}(Q-R+1)t^2}$ for $2^{-\frac{1}{4}(Q-R+1)t^2}$ in the last expression of the lemma.

Proof of Lemma 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 $j=1,2$, $|X_1|=|X_2|=q$. Each value of the

function will be a pair $\langle Z_1, Z_2 \rangle$ so that $Z_1, Z_2 \subseteq \{0, ..., n-1\}$ and $|Z_1| = |Z_2| = [\frac{q}{t^2}]t$. The defintion 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 $K_1^{(j)}, ..., K_Q^{(j)}$ of X_j, where $Q = [\frac{q}{t^2}]$, so that $|K_i^{(j)}| = t^2$ for all $j = 1, 2$ $i = 1, ..., Q$ 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 $i = 1, ..., Q$ is fixed. We apply Lemma 6 with $U \to K_i^{(1)}$, $V \to K_i^{(2)}$. Let U', V' be the sets whose exitence is stated in Lemma 6 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 Z_1, Z_2 meets the requirements described before the actual definition of $\mathcal{F}$. They also have the following additional properties:

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

(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 $i = 1, ..., q$ we have $J_i^{(1)} + J_i^{(2)} \geq \frac{1}{4}t^2$.

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

We apply lemma 8 with $l \to Q$, $I_i^{(j)} \to J_i^{(j)}$, $S^{(1)} \to Z_1$, $S^{(2)} \to Z_2$, $m_i \to \frac{1}{4}t^2$, $r \to Q - R + 1$. We get that ρ_{Z_1, Z_2} is at most $\binom{Q}{Q-R+1} 2^{-(Q-R+1)\frac{1}{4}t^2}$. 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 $|\mathtt{range}(\mathcal{F})| \binom{Q}{Q-R+1} 2^{-(Q-R+1)\frac{1}{4}t^2} \leq \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-(Q-R+1)\frac{1}{4}t^2}$. 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. *Q.E.D.*(Lemma 9).

Proof of Lemma 2. Assume that $\theta > 0$ is sufficiently small and $c_1 > 0$ is sufficiently small with respect to θ, and $c_2 > 0$. Suppose further that n is sufficiently sufficiently large and $c_2 n < q \leq n$. We apply Lemma 9 with n, q, $R = c_1 |\log(\frac{q}{n})|^{-1} q$, $t = [\theta^{-1} |\log(\frac{q}{n})|]$. We get that the probability that rank of $\mathrm{sub}(A, S, T)$ is at least R is at at least $p \geq 1 - \binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-\frac{1}{4}(Q-R+1)t^2}$ where $Q = [\frac{q}{t^2}]$. We show that $\binom{n}{Qt}^2 \binom{Q}{Q-R+1} 2^{-\frac{1}{4}(Q-R+1)t^2}$ is at most $\frac{1}{2}$ by giving upper bounds in its factors. We will use that if $0 < \alpha < \frac{1}{2}$ and n is sufficiently large, and $x < \alpha n$ then $\binom{x}{\alpha n} \leq e^{2\alpha n \log \frac{1}{\alpha}}$. Let $\gamma = \frac{q}{n}$, and $\lambda = \frac{Qt^2}{q}$. Clearly $c_2 < \gamma < 1$ and $\frac{1}{2} < \lambda \leq 1$.

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

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

$\binom{n}{Qt}^2 \leq e^{4\gamma\lambda(\theta+\theta+\theta^{\frac{1}{2}})n} \leq 2^{\frac{1}{20}\gamma\lambda n}$ if θ is sufficiently small.
$\binom{Q}{Q-R-1} \leq 2^Q = 2^{\gamma\lambda t^{-2}n} \leq 2^{\frac{1}{20}\gamma\lambda n}$ it θ is sufficiently small.
$2^{-\frac{1}{4}(Q-R+1)t^2} \leq 2^{-\frac{1}{8}Qt^2} = 2^{-\frac{1}{8}\gamma\lambda t^{-2}t^2 n} = 2^{-\frac{1}{8}\gamma\lambda n}$.

These inequalities imply that $\binom{n}{Qt}^2\binom{Q}{Q-R+1}2^{-\frac{1}{4}(Q-R+1)t^2} \leq 2^{\frac{1}{20}\gamma\lambda n+\frac{1}{20}\gamma\lambda n-\frac{1}{8}\gamma\lambda n} \leq 2^{-(\frac{1}{8}-\frac{1}{10})\gamma\lambda n} < \frac{1}{2}$ if n is sufficiently large, (here we use that $c_2 < \gamma$ and $\frac{1}{2} < \lambda$). *Q.E.D.*(Theorem 2)

References

[A] M. Ajtai, Determinism versus Non-Determinism for Linear Time RAMs with Memory Restrictions, ECCC http://www.eccc.uni-trier.de/eccc/ (revised version)

[BC] A. Borodin and S. Cook, A time-space tradeoff for sorting on a general sequential model of computation, SIAM J. Comput., 11, (1982), pp. 287-297

[Bea] P. Beame, A General Sequential Time-Space Tradeoff for Finding Unique Elements, SIAM J. Comput., 20, (1991) No. 2,pp. 270-277.

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

[BST] P.W. Beame, M. Saks and J. S. Thathachar. Time-space Tradeoffs for Branching programs. 39th Annual Symposium on Foundations of Computer Science, 1998. pp. 254-263, or ECCC http://www.eccc.uni-trier.de/eccc/

[Tha] J. S. Thathachar, On separating the read-k-times branching program hierarchy, Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, 1998, pp. 653-662.

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