Electronic Colloquium on Computational Complexity, Report No. 77 (1998)

Determinism versus Non-Determinism for Linear Time RAMs with Memory Restrictions

(Preliminary version)

Miklós Ajtai
IBM Almaden Research Center

Abstract. Our computational model is a random access machine with n read only input registers each containing $c \log n$ bits of information and a read and write memory. We measure the time by the number of accesses to the input registers. We show that for all k there is an $\epsilon > 0$ so that if n is sufficiently large then the elements distinctness problem cannot be solved in time kn with ϵn bits of read and write memory, that is, there is no machine with this values of the parameters which decides whether there are two different input registers whose contents are identical. We also show that there is a simple decision problem that can be solved in constant time (actually in two steps) using non-deterministic computation, while there is no deterministic linear time algorithm with $\epsilon n \log n$ bits read and write memory which solves the problem. More precisely if we allow kn time for some fixed constant k, then there is an $\epsilon > 0$ so that the problem cannot be solved with $\epsilon n \log n$ bits of read and write memory if n is sufficiently large. The decision problem is the following: "Find two different input registers, so that the Hamming distance of their contents is at most $\frac{1}{4} c \log n$". $\frac{1}{4}$ can be replaced by any fixed $0 < \gamma < \frac{1}{2}$ if c is sufficiently large with respect to γ. We actually show that the promise problem : "decide whether all occurring Hamming distances are greater than $(\frac{1}{2} - \gamma) c \log n$ or there is at least one which is smaller than $\gamma c \log n$" where $\gamma > 0$ is an arbitrarily small constant, cannot be solved by a nonlinear algorithm with the described limitations even if we know that we only get inputs where one of these conditions hold. (In this case ϵ may depend on γ too).

Introduction. One of the main goals of complexity theory is the separation of non-deterministic and deterministic computation. We solve the problem for random access machines with certain restrictions on the size of their working memory. Although the restrictions are strong, the working memory must be smaller than the input, still, under certain circumstances this computational model is realistic as we will explain later. We also use realistic search problems for the sepration results. The first problem is the element distinctness problem, that is, we have to decide whether there are different input registers with identical contents. This problem is of great practical and theoretical interest, it has been studied in great detail in various computational

ISSN 1433-8092

models, particularly in the comparison model (see [BFKLT], [BFMUW], [K], [Y]). A time space tradoff $TS = \Omega(n^2)$ for the elements distinctness problem on comparison-based branching programs was conjectured by Borodin et al in [BFKLT]. A. Yao [Y] proved a tradeoff $TS = \Omega(n^{2-\epsilon(n)})$, where $\epsilon(n) = O(1/(\log n)^{\frac{1}{2}})$, which is very close to optimal since $TS = O(n^2)$ is achievable even for sorting in the range $c_1 \log n \leq S \leq c_2 n/\log n$ (See [PR]). The best upper bounds for the element distinctness problem are given in the RAM model. We can solve the element distinctness problem with bucket sorting in our RAM in linear time with $c'n \log n$ bits of read and write memory, where c' is a suitably choosen constant (see [AHU]). This is a determinstic (non-probabilistic) algorithm. Our lower bounds are also about non-probabilistic algorithms. For the element distinctness problem we give a probabilistic algorithm which solves it in time kn with ϵn bits of read and write memory provided that $k > 0$ is sufficiently large with respect to ϵ and n is sufficiently large with respect to k. Moreover our algorithm can be implemented in a random access machine defined in the usual sense, that is where the memory is organized into registers and we allow only arithmetic operations etc. (For the exact statement of this result see Theorem 5 in the last section.) This makes it very unlikely that our lower bound for the element distinctness problem can be improved since the ratio between the determenistic lower bound and the probabilistic upper bound is only a large constant.

The other problem that we study is the Hamming distance problem, that is, find two different input registers so that the Hamming distance of their contents is smaller than $\gamma c \log n$ where $\gamma < \frac{1}{2}$ and the input registers contain $c \log n$ bits. This problem and in particular its promise problem version mentioned in the abstract is also of great theoretical and practical importance although much less is known about it than about the element distinctness problem. Searching for two items with small Hamming distance is a typical task e.g. in recognition problems, reservation systems, data compression problems and generally in any search problem where nearly identical objects have to be located. Here our result is stronger since we allow more read and write memory than in the element distinctness problem, still our proof is much easier. The reason is probably that the elements distinctness result is sharper, as we explained our lower bound is very likely close to the best possible, while for the Hamming distance problem we do not have any reason to assume that the lower bound cannot be significantly improved.

The main restriction in our computational model is the memory restriction, that is, the fact that the read and write memory is smaller (although only by a constant factor in the Hamming distance problem and a factor of constant times $\log n$ for the element distinctness problem), than the size of the input. However this is a common situation in practice. In certain search problems we may search all of our available memory (e.g hard disk). In this case the available free memory for computation (random acces or

not) can be much smaller then the input. Modifying the input may be impossible or unwise because of safety reasons. In a similar way we may have read only access to a huge database where we cannot alter the data. Again our input is much larger then our workspace. In our examples the input is usually not located in a random access memory, so our computational model does not describe well the expense of reading from such a memory, still our lower bounds have meaningful consequences.

The most elegant way to describe our computational model in an abstract form is the R-way branching programs. Since we do not have any limitation on the computation done between the accesses to the input, we may simply assume that the possible states of the R/W memory are the nodes of a directed graph and each node is associated with a register which will be accessed in the given state and there are outgoing edges from the nodes each associated with the possible contents of the accessed register. The machine follows a path on this graph starting from a distinguished node and from each node the path continues along the edge which is labeled by the content of the register associated with the node. The number of outgoing edges is R, in our case $R = n^c$. This computational model was introduced by Borodin and Cook (see [BC]) and they gave a time space tradeoff for sorting n integers. They introduced a technique for proving lower bounds for R-way branching problems where the number of output bits is relatively large compared to the time allowed. This method was successfully used for proving several lower bounds and space time tradeoffs for problems of similar nature. (See e.g. Beame [Be]) Our problem does not belong to this category since the output is a single bit. Still the same very high level idea, cutting the time into short intervals and knowing that all we can use about the past of the computation in an interval must be contained in the limited memory at the beginning of the interval, is applicable.

The strongest known separation theorem between deterministic and non-deterministic computation is the theorem of Paul, Pippinger, Szemerédi and Trotter (see [PPST]) stating that non-deterministic linear time is more powerful than deterministic linear time for multitape Turing machines. The proof of this theorem is also using a segmentation of the time. In this case, although the overall size of the memory can be larger than the input, still the geometry of the machine acts as a local memory limitation.

We formulated our results in the random access model since this motivates the choice of the parameters in our computational models. Although the notion of 1-way branching is simpler and perhaps more natural, n^c-way branching problems are motivated by the usual register sizes in actual random access machine. Since an address of a register is usally stored in one (or a few) registers, the $c \log n$ register size is a natural choice. Our results has no consequences for the Boolean (that is 1-way) case, since we lose a factor of $c \log n$ during the translation.

The number of lower bound results for computing various explicitly given functions with a branching program when there are certain limitation on the parameters (size, depth, width, number of aceesses of the input bits, "R", etc.) is very large so we do not try to give a review of them. The results that are closest to our present problem, are lower bounds on the computation of explicit functions given by Beame, Saks and Thathachar ([BST]). There is also a strong similarity between the proof techniques of [BST] and the present paper. They gave a nonlinear lower bound on the depth of an R-way branching program computing an explicitly defined function, where the size is n^c and R depends only on k and not on n. More precisely they proved that for all k there is a r_k so that for all sufficiently large n there is an (explicitly given) function $g(x_1, ..., x_n)$ of n variables with $0, 1$ values so that: (a) each variable is taking it 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. (In our RAM model this means that the number of bits in the input register is $\log r_k$ the working memory is $c \log n$ bits and the time is kn). For the 1-way case they have a similar lower bound but k can be only a fixed constant larger than 1. (The paper also gives an overview of our present state of knowledge about branching programs.)

We will first prove the Hamming distance result. The proof of the element distinctness lower bound will be built on top of this proof in the sense that the basic idea of the proof is the same but for the actual realization we need a much more complicated and argument. For this latter proof we use some of the lemmata proved for the Hamming distance lower bound.

We prove our result not directly for Hamming distances but for a much larger class of search problems. Assume that a binary relation Q is given and our problem is whether there are two distinct registers so that for their contents x, y we have $Q(x, y)$. We give a condition for the relation Q so that if this condition is satisfied then the problem cannot be decided in linear time with the described restrictions on the memory.

Definition. Assume that Q is a binary relation defined on the finite set A and $\lambda > 0$ is a positive real number. We say that the relation Q is λ-full on A iff the following holds:

for all $B \subseteq A$, $C \subseteq A$ if $|B| > \lambda|A|$, $|C| > \lambda|A|$ then there exist $x \in B$, $y \in C$ so that $Q(x, y)$.

Examples. 1. The relation equality is $\frac{1}{2}$-full on any finite set $|A|$. Indeed if both B and C has more than $\frac{1}{2}|A|$ elements, then they must have a common element.

2. (a) Assume that there is a graph G whose set of vertices is A, and we denote by $N(X)$ the neighborhood of each set $X \subseteq A$, that is the set of points which are either in X or connected by an edge to an element of X. Suppose further that $1 \geq \lambda > 0$ and the graph has the following expansion property:

for each $Y \subseteq A$ with more than $\lambda|A|$ elements we have $|N(Y)| > \frac{1}{2}|A|$.

We claim that the binary relation Q defined on A by "$Q(x,y)$ iff the distance of x and y in the graph G is at most 2" is λ-full on A. Indeed if $|B| > \lambda|A|$, $|C| > \lambda|A|$, then $N(B) > \frac{1}{2}|A|$, and $N(C) > \frac{1}{2}|A|$, so $N(B) \cap N(C) \neq \emptyset$, which implies our claim.

(b) If G is a graph on A, and $X \subseteq A$, then let $N_d(X)$ be the d neighborhood of N, that is, the set of points in A whose distance from X is at most d. If G has the expansion property that for all $X \subseteq A$, $|X| > \lambda|A|$ we have $|N_d(X)| > \frac{1}{2}|A|$, then the relation "the distance of x and y is at most $2d$" is λ-full.

The following example is a special case of (b) and it gives a way to construct a non-trivial λ-full relation for a fixed λ and for an arbirarily large universe.

(c) Assume that $1 > \lambda > 0$ is fixed, and that $a > 0$, $c > 0$ so that for each integer n there is a c-regular graph G_n on the vertex set A_n with n points so that for each $X \subseteq A_n$, $|X| < \frac{n}{2}$ we have $|N(X)| > (1+a)|X|$. Then there is a d depending only on λ, a, c so that the relation "the distance of x and y is less than $2d$" is λ-full on A_n

Definition. Let $\mathcal{R}(x,y)$ be a binary relation on the finite set A. Assume that $\mu > 0$ is a real-number and m is a positve integer. We say that $\mathcal{R}$ is (μ, m)-sparse if the following holds: the number of sequences $a_1, ..., a_m$ formed from the elements of A so that $\mathcal{R}(a_i, a_j)$ for some $1 \leq i < j \leq m$ is at most $\mu|A|^m$.

Our general condition which guarantees that the decision problem associated with Q cannot be solved in linear time will be essentially the following:

(a) Q is $n^{-\delta}$-full for some constant δ where $0 < \delta < c$. (The number of bits in the registers is $c \log n$) and

(b) Q is $(\frac{1}{2}, n)$-sparse.

In the reduction of the Hamming distance theorem to the more general result we will use the following statement which connects the notions of Hamming distance and λ-full relations:

(*) *For all $0 < \gamma < \frac{1}{2}$ there is a $\tau < 1$ so that if m is sufficently large then on the set $A^{(m)}$ of all $0,1$ sequences of length m, the following relation Q is $2^{-\tau m}$ full: "the Hamming distance of x and y is at most γm"*

The Hamming distance problem. For the exact formulation of our results we need some definitions. First we give a detailed but informal description of a random access machine with n read-only input registers and $\log \beta$ bits of read an write memory. ($\log x$ will denote the logarithm of base 2.) We will consider the following random access machine $\mathcal{M}$. It has n input registers each may contain an element of the set $\{0, 1, ..., \alpha - 1\}$ where α is a positive integer. (We usually will assume that $\alpha \leq n^c$ for some constant c). The machine is only able to read the contents of the input registers,

but it is not able to change them. The machine has β different states, we may think of them as the various possible states of its read and write memory if this memory can store $\log_2 \beta$ bits of information. We usually will assume that β is about $2^{\epsilon n \log n}$, in other words the read and write memory may consist of ϵn registers each with $\log n$ bits. $\mathcal{S}$ will denote the set of the states of $\mathcal{M}$. There is a state of the machine called the initial state what we will denote by `init`. We assume that a function φ is fixed, defined on $\mathcal{S}$ with values in $\{1, ..., n\}$. (At state S the machine will have access to the content of the resgister $\varphi(S)$).

Another function $\mathcal{G}(x, y)$ is also given which is defined for all pairs x, y where $x \in \mathcal{S}$ and $y \in \{0, ..., \alpha - 1\}$, the values of $\mathcal{G}$ are in $\mathcal{S}$. (As we will see below $\mathcal{G}$ describes how are the states of $\mathcal{M}$ are changed.)

Finally "`out`" is a function defined on the set $\mathcal{S}$ with values in $0, 1$. (`out` will determine the output of the machine at a given state.)

$\mathcal{M}$ works in the following way. Assume that each input register contains a nonnegative integer less than α. We will denote the content of register i by $\eta(i)$ for $i = 1, ..., n$. At time 0 the $\mathcal{M}$'s state is `init`.

The machine $\mathcal{M}$ changes its states according to the following rule. Assume that S is the state of the machine at time t. Then at time $t + 1$ the state of the machine is $\mathcal{G}(S, \eta(\varphi(S)))$. The state of the machine at time t will be denoted by $\mathtt{state}(t, \eta)$.

We will think that $\mathcal{M}$ is working for a fixed amount of time t_0 with the input η. The output of the machine at time t_0 and at input η will be $\mathtt{out}(\mathtt{state}(t_0, \eta))$ which also will be denoted by $\mathtt{out}_{t_0}(\eta)$.

Now we give a more concise formal definition of $\mathcal{M}$.

Definitions. 1. The machine $\mathcal{M}$ is a sequence $\langle n, \alpha, \beta, t_0, \mathcal{S}, \mathtt{init}, \mathcal{G}, \mathtt{out}, \varphi \rangle$, where n, α, β, t_0 are positve integers, $\mathcal{S}$ is a finite set, $\mathtt{init} \in \mathcal{S}$, $\mathcal{G}(x, y)$ is a function of two variables defined for all pairs x, y where $x \in \mathcal{S}$ and $y \in \{0, ..., \alpha - 1\}$ with values in $\mathcal{S}$, `out` is a $0, 1$-valued function defined on $\mathcal{S}$, φ is a function defined on $\mathcal{S}$ with values in $\{1, ..., n\}$ and $\beta = |\mathcal{S}|$.

2. A function η defined on the set $\{1, ..., n\}$ with values in the set $0, ..., \alpha - 1$ will be called an input for $\mathcal{M}$.

3. We define a function $\mathtt{state}(t, \eta)$ for all nonnegative integer t and for all input η, by recursion on t:

$\mathtt{state}(0, \eta) = \mathtt{init}$

$\mathtt{state}(t + 1, \eta) = \mathcal{G}(\mathtt{state}(t, \eta), \varphi(\mathtt{state}(t, \eta)))$

4. We define the output of $\mathcal{M}$ at input η as $\mathtt{out}(\mathtt{state}(t_0, \eta))$. We will also use the abbreviation $\mathtt{out}_{t_0}(\eta) = \mathtt{out}(\mathtt{state}(t_0, \eta))$

Sometimes we will not write out the complete sequence of objects defining the machine, we may say that $\mathcal{M} = \langle n, \alpha, \beta, t_0 \rangle$ is a machine and assume that the missing elements of the sequence are denoted in the usual way.

Theorem 1 . *For all $0 < \gamma < \frac{1}{2}$ there exists a $c_1 > 0$ so that for all positive integer k there is an $\epsilon > 0$ so that if n is sufficiently large, $\alpha > n^{c_1}$, $\beta < 2^{\epsilon n \log n}$ then the following holds:*

there is no machine $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ so that for any input η we have $\mathtt{out}_{kn}(\eta) = 1$ iff there are $i, j \in \{1, ..., n\}$, $i \neq j$ so that the Hamming distance of $\eta(i)$ and $\eta(j)$ is less than $\gamma \log_2 \alpha$

Although we do not formulate here the "promise problem" promised in the abstract, but our proof gives this result too without any extra work. The proof of Theorem 1 will be based on the the following more general theorem.

Theorem 2 . *There exists a $c_1 > 0$, so that for all positive integer k, and for all real number $\delta \in (0, c_2)$, if $\epsilon > 0$ is sufficiently small and n is sufficiently large, $\alpha > n^{c_1}$, $\beta < 2^{\epsilon n \log n}$, and Q is an $n^{-\delta}$-full binary relation on $\{0, ..., \alpha - 1\}$, with the property:*

(1) Q is $(\frac{1}{2}, n)$-sparse

then there is no machine $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ so that for all inputs η we have

(2) $\mathtt{out}_{kn}(\eta) = 1$ iff there are $i, j \in \{1, ..., n\}$, $i \neq j$ so that $Q(\eta(i), \eta(j))$.

Sketch of the proof of Theorem 2. We will show that there is an input χ so that χ does not satisfy Q and there are two disjoint sets W_1, W_2, and sets of partial inputs A_i defined on W_i, $i = 1, 2$ so that $|A_i|$ is so large that $|\bigcup_{\eta \in A_i} \mathtt{range}(\eta_i)| > n^{-\delta}\alpha$ (see Lemma 3 for the necessary lower bound on $|A_i|$) We will also be able to select these object with the additional property that for each pair $\eta_1 \in A_1$, $\eta_2 \in A_2$ we have $\mathtt{out}_{kn}(\chi) = \mathtt{out}_{kn}((\chi \wr \eta_1) \wr \eta_2)$.

The $n^{-\delta}$-fullness of Q implies that there are x_1, x_2 $\eta_1 \in A_1, \eta_2 \in A_2$ with $x_1 \in \mathtt{range}(\eta_1)$, $x_2 \in \mathtt{range}(\eta_2)$ and $Q(x_1, x_2)$. Since W_1 and W_2 are disjoint we have that $(\chi \wr \eta_1) \wr \eta_2$ satisfy Q. Moreover because of the mentioned property of the sets A_1, A_2 we have that the output of the machine at input χ is the same as at input $\chi' = (\chi \wr \eta_1) \wr \eta_2$. So we got two inputs χ, χ' which provide the same output and one satisfies Q while the other is not. Therefore the machine cannot decide whether its input satisfies Q.

To carry out this program we first have to know whether for our input χ there are partial inputs η at all so that $\mathtt{out}_{kn}(\chi) = \mathtt{out}_{kn}(\chi\eta)$. Since we assumed that for at least half of the all possible inputs χ we have $\mathtt{out}(\chi) = 0$, by a simple counting argument we get that indeed there must be many partial inputs η with the required property. Let X be the set of all inputs χ with $\mathtt{out}(\chi) = 0$. We show in Lemma 7 that if $B \subseteq \{1, ..., n\}$ then for most of the inputs $\chi \in X$ the number of partial inputs η so that $\chi \wr \eta \in X$ is very large, it will be close to $|X|\alpha^{n-|B|}$. The Lemma gives

a more precise connection between the various parameters. In particular it follows (see Lemma 8) that X still have a large subset whose elements χ have the following property: for all large enough $Z \subseteq \{1, ..., n\}$ the number of partial inputs η defined on Z so that $\chi \wr \eta \in X$ is still large.

We have seen that there are many inputs χ that can be changed in many different ways without changing the output. This cannot help in itself since this statement remains true if we do not speak about outputs, but ask only whether the sequence satisfies Q. This is the reason why we need changes represented by partial inputs η_1, η_2 which take place simultaneously. We want to isolate η_1, η_2 from eachother so that if $\mathtt{out}(\chi \wr \eta_i) = \mathtt{out}(\chi)$, that is, applying them separately does not change the output, then this remains true if they applied together. We cut the total time $[0, kn]$ into small subintervals of length σn, where σ is sufficiently small with respect to k. We want to choose η_1, η_2 in a way that their domains W_1, W_2 are in different subintervals. On top of that we also want W_1, W_2 to be large enough so that we have many different choices for η_1, η_2 defined on them. To achieve this we partition the set of registers so that two are in the same class if they are seen by the machine in the same intervals. An average register is seen only k-times so we may throw away those registers that are seen more than $2k$ times and still we have at least $\frac{n}{2}$ registers. We consider the classes only containing these registers. Again we throw away the classes which are too small and at the end what remained is still a partition of at least $\frac{1}{4}n$ registers in each class at least $c(\sigma, k)n$ registers, and registers in a single class are seen exactly in the same intervals. Let Γ_χ be this partition. W_1 and W_2 will be classes of Γ_χ for a suitably chosen χ. With a counting argument we can show that there are classes so that the set of intervals where the registers in W_1 and W_2 are accessed, are not only distinct but disjoint. We will use such a pair W_1, W_2. The definition of Γ_χ is given right before Lemma 5, the lemma itself formulates the mentioned properties.

Assume now that a χ and the classes W_1, W_2 are fixed so that the set of intervals where their registers are seen are disjoint. We will take a partial input η_i on W_i so that not only $\mathtt{out}(\chi) = \mathtt{out}(\chi \wr \eta_i)$ but also the state of the machine, when leaving of each of the intervals, where W_i is accesed, is the same at input χ and at input $\chi \wr \eta_i$. If T_i is the set of times t which are contained in an interval where W_i is accessed then it means that the state of the machine at each point of the right border of T_i is the same at inputs χ and $\chi \wr \eta_i$. This neutralizes the changes made by η_i. At the end of the interval where the elements of W_i is accessed the state of the machine will remain the same. Since we have a limit on the number of states (the memory) the values for the states at the end of the intervals can be fixed in a way so that still they occur for many χ and so for a large number of them there will be many partial inputs η_i with the required properties. We may think at first that the isolation of η_1 from η_2 implies that $\mathtt{out}(\chi) = \mathtt{out}(\chi \wr \eta_1 \wr \eta_2)$. This is not necessarily true because it is possible that

e.g. in an interval where only the registers of W_1 are seen at input χ, at input $\chi \wr \eta_1$ some new registers, say registers from W_2 are also seen. However if we exclude this possiblity then even the two changes applied together does not change the output. (see Lemma 2)

This last condition can be satisfied if we consider only the set H of those inputs χ where W_1 and W_2 are classes of Γ_χ and we are looking for partial inputs η_i with the property $\chi \wr \eta_i \in H$. To satisfy the earlier requirements too we also want for all inputs $\chi \in H$ the state of the machine to be the same at the right border of T_i for $i = 1, 2$. (T_i is determined uniquely by W_i). We consider the possible pairs of sets W_1 and W_2 and the possible functions giving the states of the machines at the right borders and show that the number of choices for these object is so small compared to α^n, the total number of inputs, that for at least one of the choices the set of corresponding inputs is still large. (Lemma 6). Let H be this set. We also require that for each $\chi \in H$, $\mathtt{out}(\chi) = 0$. Now we may complete the proof easily. H is so large that by Lemma 8 there is a large set of inputs A_i in W_i, $i = 1, 2$ so that for any choices $\eta_i \in A_i$ we have $\chi \wr \eta_i \in H$. $|A_i|$ is so large that by Lemma 3 the partial inputs in it altogether take more then $n^{-\delta}\alpha$ values. Therefore by the $n^{-\delta}$-fullness of Q we have that Q is satisfied by $\chi \wr \eta_1 \wr \eta_2$ for some $\eta_i \in A_i$. We have that $0 = \mathtt{out}(\chi) = \mathtt{out}(\chi \wr \eta_1 \wr \eta_2)$=1, a contradiction.

Proof of Theorem 2. We assume that contrary to our statement there is a machine $\mathcal{M}$ with property (2). For the proof of the theorem assume that k, δ are fixed. We pick a positive real number σ so that it is sufficiently small with respect to k and assume that ϵ is sufficiently small with respect to σ, k, δ, and n is sufficiently large with respect to k, σ, δ and ϵ.

Definitions. 1. A partial input η will be a function defined on a subset of the set $\{1, ..., n\}$ with values in $\langle 0, 1, ..., \alpha - 1\rangle$.

2. If 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 at every other points.

3. If T is a set of integers we say that $\mathcal{M}$ accesses a register u in a set T if there is a $t \in T$ so that $\mathcal{M}$ accesses u at time t.

4. Suppose that $T \subseteq \{0, ..., kn - 1\}$. 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 $\mathtt{rigth}(T)$.

5. Suppose that $T \subseteq \{0, ..., kn - 1\}$ and χ is an input. Let f be a function defined on the set $\mathtt{rigth}(T)$, so that for all $t \in \mathtt{rigth}(T)$ 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 $\mathtt{rstate}_{T,\chi}$.

6. Assume that χ is an input we say that χ satisfies Q if there are $1 \leq i < j \leq n$ so that $Q(\chi(i), \chi(j))$.

Lemma 1. *Assume that* $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ *is a machine satisfying condition* (2), χ *is an input and* A_1, A_2 *are sets of partial inputs so that* $\mathtt{domain}(\eta_1) \cap \mathtt{domain}(\eta_2) = \emptyset$ *for any* $\eta_1 \in A_1$, $\eta_2 \in A_2$. *If* $|\bigcup_{\eta \in A_i} \mathtt{range}(\eta_i)| > n^{-\delta}\alpha$ *for* $i = 1, 2$, *then there exist* $\eta_1 \in A_1$, $\eta_2 \in A_2$ *with* $\mathtt{out}_{kn}((\chi \, l \, \eta_1) \, l \, \eta_2) = 1$.

Proof. Since Q is $n^{-\delta}$-full on $\{0, 1, ..., \alpha - 1\}$ there are $x_i \in \bigcup_{\eta \in A_i} \mathtt{range}(\eta_i)$ for $i = 1, 2$ so that $Q(x_1, x_2)$. Let $\eta_i \in A_i$ be the partial inputs taking the values x_i. (2) implies that conclusion of the lemma.

We will show that there is an input χ so that χ does not satisfy Q and there are two disjoint sets W_1, W_2, and sets of partial inputs A_i defined on W_i, $i = 1, 2$ so that $|A_i|$ is so large that $|\bigcup_{\eta \in A_i} \mathtt{range}(\eta_i)| > n^{-\delta}\alpha$ (see Lemma 3 for the necessary lower bound on $|A_i|$). We will also be able to select these objects with the additional property that for each pair $\eta_1 \in A_1$, $\eta_2 \in A_2$ we have $\mathtt{out}_{kn}(\chi) = \mathtt{out}_{kn}((\chi \, l \, \eta_1) \, l \, \eta_2)$.

The $n^{-\delta}$-fullness of Q implies that there are x_1, x_2 $\eta_1 \in A_1, \eta_2 \in A_2$ with $x_1 \in \mathtt{range}(\eta_1)$, $x_2 \in \mathtt{range}(\eta_2)$ and $Q(x_1, x_2)$. Since W_1 and W_2 are disjoint we have that $(\chi \, l \, \eta_1) \, l \, \eta_2$ satisfy Q. Moreover, because of the mentioned property of the sets A_1, A_2 we have that the output of the machine at input χ is the same as at input $\chi' = (\chi \, l \, \eta_1) \, l \, \eta_2$. So we got two inputs χ, χ' which provide the same output and one satisfies Q while the other is not. Therefore the machine cannot decide whether its input satisfies Q.

From this description it is still missing how is it possible to guarantee the required properties of A_1 and A_2. Lemma 2 and 4 below give necessary conditions for this.

Definitions. Assume that T is a set of integers. The set of all registers i so that i is accessed by the machine $\mathcal{M}$ at some $t \in T$, at input η will be denoted by $\mathtt{register}(T, \eta)$. The set of all registers in $\mathtt{register}(T, \eta)$ which are not accessed at any time outside T, at input η will be denoted $\mathtt{core}(T, \eta)$. Clearly $\mathtt{core}(T, \eta) \subseteq \mathtt{register}(T, \eta)$.

Lemma 2 *Assume that* χ *is an input,* η_1, η_2 *are partial inputs,* $T_1, T_2 \subseteq \{0, 1, ..., nk - 1\}$. *If* $\chi, \eta_1, \eta_2, T_1, T_2$ *satisfy the following conditions then* $\mathtt{out}_{kn}(\chi) = \mathtt{out}_{kn}((\chi \, l \, \eta_1) \, l \, \eta_2)$.

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

*(***4.2***)* T_1 *and* T_2 *are disjoint.*

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

*(***6.2***)* *for all* $i = 1, 2$ *we have* $\mathtt{rstate}_{T_i, \chi} = \mathtt{rstate}_{T_i, \chi l \eta_i}$

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

Proof. We will see how the computation is changed by the given changes of the input. The set T_i is the union of a set K_i of disjoint intervals for $i = 1,2$. We assume that $|K_i|$ is minimal. Let $K = K_1 \cup K_2$. (4.2) implies that the elements of K are still disjoint. Let $K = \{I_1, ..., I_r\}$ where every element of I_s is smaller than any elements of I_{s+1} for $s = 1, ..., r-1$. First we prove by induction according to s, that if h_s is the unique element of $\mathtt{rigth}(I_s)$, then $\mathtt{state}(\chi, h_s) = \mathtt{state}(\chi \wr \eta_1, h_s) = \mathtt{state}(\chi \wr \eta_2, h_s) = \mathtt{state}(\chi', h_s)$, where $\chi' = (\chi \wr \eta_1) \wr \eta_2$.

Let $\chi_i = \chi \wr \eta_i$. Assume $s = 1$. The initial segment of the computation before we enter the time interval I_1 is the same at inputs χ, χ', χ_1, χ_2, since at input χ by (5.2) we do not access any register in the domains of η_i, $i = 1,2$ so changing the content of these registers in any way does not influence the computation. Assume that e.g. $I_1 \in K_1$. In this case we claim that the computation at input χ' in I_1 will proceed in the same way as at input $\chi \wr \eta_1$. Indeed, according to (7.2) with $i = 2$, $j = 1$, during the computation at input $\chi \wr \eta_1$ we do not access any of the registers in the domain of η_2, so the change of their contents do not influence our computation. We get that $\mathtt{state}(\chi \wr \eta_1, h_1) = \mathtt{state}(\chi', h_1)$ and furthermore by (6.2) that this common value is also equal to $\mathtt{state}(\chi, h_1)$. We may use the same argument that we have used before I_1 to show that the computation at input χ_2 in this interval remains the same as at input χ so we also have $\mathtt{state}(\chi, h_1) = \mathtt{state}(\chi_2, h_1)$. The general inductive step, and the computation after h_r can be handled in a similar way, using the fact that by the inductive hypothesis we start the computation at time h_{s-1} in the same state at all of the four possible inputs.

Lemma 3. *Assume that χ is an input, $D \subseteq \{1, .., n\}$ and A is a set of partial inputs defined on D. If $s = |\bigcup_{\eta \in A} \mathtt{range}(\eta)|$ then $s^{|D|} \geq |A|$. As a consequence if $|A| > (n^{-\delta}\alpha)^{|D|}$ then $s > n^{-\delta}\alpha$.*

Proof. The number of functions defined on a set of size $|D|$ and taking at most s different values is at most $s^{|D|}$.

Lemma 4 *Assume that $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ is a machine, χ is an input, and A_1, A_2 are sets of partial inputs, $T_1, T_2 \subseteq \{0, 1, ..., nk-1\}$. If $\mathcal{M}, \chi, A_1, A_2, T_1, T_2$ satisfy the following three conditions then there are $\eta_1 \in A_1$, $\eta_2 \in A_2$ and $x_1 \in \mathtt{range}(\eta_1)$, $x_2 \in \mathtt{range}(\eta_2)$ so that for $\chi' = (\chi \wr \eta_1) \wr \eta_2$ we have $\mathtt{out}_{kn}(\chi) = \mathtt{out}_{kn}(\chi')$, $\exists 1 \leq i < j \leq n$, $x_1 = \chi'(i), x_2 = \chi'(j)$, and $Q(x_1, x_2)$.*

(8.4) *for all $i = 1,2$ there is a set W_i so that $\mathtt{domain}(\eta) = W_i$ for all $\eta \in A_i$.*

(9.4) *for all $\eta_1 \in A_1$, $\eta_2 \in A_2$ conditions (3.2), (4.2), (5.2), (6.2), (7.2) are satisfied by $\chi, \eta_1, \eta_2, T_1, T_2$.*

(10.4) *$|A_i| > (n^{-\delta}\alpha)^{|W_i|}$ for all $i = 1,2$.*

Proof. ((10.4)) and Lemma 3 with $D \to W_i$ imply that $|\bigcup_{\eta\in A_i} \mathtt{range}(\eta_i)| > n^{-\delta}\alpha$. Therefore by the definition of $n^{-\delta}$ fullness we have $\eta_i \in A_i$ $x_i \in \mathtt{range}(\eta_i)$, $i = 1, 2$ so that $Q(x_1, x_2)$. By (3.2) the domains of η_1, η_2 are disjoint so $(\chi \restriction \eta_1) \restriction \eta_2$ takes the values x_1, x_2 at distinct places. $\mathtt{out}(\chi) = \mathtt{out}(\chi')$ is a consequence of Lemma 2.

Definitions. 1. We partition the set $\{0, 1, ..., kn - 1\}$ into intervals so that the length of each interval is between σn and $2\sigma n$. Let $\mathcal{I}$ be the set of these intervals. Suppose that an arbitrary input χ is fixed. We define a partition $\mathcal{R}_\chi$ on the set of input registers $\{1, ..., n\}$. u and v will be in the same class of $\mathcal{R}_\chi$ iff for each $I \in \mathcal{I}$, $\mathcal{M}$, at input χ, accesses u in I if and only if it accesses v in I. (That is, if the machine looks at u and v exactly in the same time intervals $I \in \mathcal{I}$.)

Let $\mathcal{R}'_\chi$ be the set of those classes of $\mathcal{R}_\chi$ whose elements are accessed, at input χ, in at most $2k$ different time intervals $I \in \mathcal{I}$. Γ_χ will denote the set of all classes of $\mathcal{R}'_\chi$ which have more than $\frac{1}{4}|\mathcal{R}'_\chi|^{-1}n$ elements.

2. Assume that χ is an input and C is a class of Γ_χ. The set of those intervals I of $\mathcal{I}$ wich satisfy $C \subseteq \mathtt{register}(I, \chi)$ will be denoted by $\mathtt{set}(C, \chi)$. In other words $\mathtt{set}(C, \chi)$ consists of exactly those intervals of $\mathcal{I}$ where all of the elements of C are accessed at input χ. (Note that by the definition of $\mathcal{R}_\chi$ either all or none of the elements of a $C \in \Gamma_\chi \subseteq \mathcal{R}_\chi$ is accessed in an $I \in \mathcal{I}$).

Lemma 5. *If χ is an arbitrary input then*

(11.5) $|C| \geq \frac{1}{4}\sigma^{2k}k^{-2k}n$ *for all* $C \in \Gamma_\chi$

(12.5) $|\bigcup_{C\in\Gamma_\chi} C| \geq \frac{1}{4}n$

(13.5) *There are* $W_1, W_2 \in \Gamma_\chi$ *with* $\mathtt{set}(W_1, \chi) \cap \mathtt{set}(W_2, \chi) = \emptyset$

Proof. (11.5). By the definition of $\mathcal{I}$ we have $|\mathcal{I}| \leq \sigma^{-1}k$. Each class of $\mathcal{R}_\chi$ is uniquely determined by at most $2k$ elements of $\mathcal{I}$. Therefore $|\mathcal{R}'_\chi| \leq \sum_{j=0}^{2k} \binom{\sigma^{-1}k}{j} \leq 2\binom{\sigma^{-1}k}{2k} \leq (\sigma^{-1}k)^{2k}$. Therefore the definition of Γ implies (11.5).

(12.5). First we note that $|\bigcup_{C\in\mathcal{R}'_\chi} C| \geq \frac{n}{2}$. Indeed, if a register is not in $\bigcup_{C\in\mathcal{R}'_\chi} C$, then it was accessed more than $2k$ times. Since there are altogether kn steps in the computation we may have no more than $\frac{n}{2}$ such registers. The definition of Γ_χ implies that Γ_χ contains each class of $\mathcal{R}_{\chi'}$ whose size is more than the half of the average class size in $\mathcal{R}'_\chi$. Clearly these classes contain at least half of the elements of $\bigcup_{C\in P'_\chi} C$.

(13.5). Let W_1 be an arbitrary element of Γ_χ. Assume that

(*) $\mathtt{set}(W, \chi) \cap \mathtt{set}(W_1, \chi) \neq \emptyset$

for some $W \in \Gamma_\chi$. Then there is an $I \in \mathcal{I}$ contained in $\mathtt{set}(W, \chi) \cap \mathtt{set}(W_1, \chi)$. $\Gamma_\chi \subseteq \mathcal{R}_\chi$ implies that $W \subseteq \mathtt{register}(I, \chi)$. Therefore $W \subseteq \bigcup\{\mathtt{register}(I, \chi) | I \in \mathtt{set}(W_1, \chi)\}$. So the set of registers contained in all $W \in \Gamma_\chi$ satisfying (*) is covered by the set $\bigcup\{\mathtt{register}(I, \chi) | I \in \mathtt{set}(W_1, \chi)\}$ and consequently the number of these registers is at most $\sum\{|\mathtt{register}(I, \chi)| \mid I \in \mathtt{set}(W_1, \chi)\} \leq 2k2\sigma n$. Since σ is

sufficiently small with respect to k, this inequality and (12.5) implies that there must be at least one class of Γ_χ which does not satisfy (*).

Lemma 6. *Assume that k is a positive integer, σ is sufficiently small with respect to k, ϵ is sufficiently small with respect to σ, n is sufficiently large with respect to ϵ, $\beta \leq 2^{\epsilon n \log n}$ and the machine $\mathcal{M}$ satisfies the conditions (1) and (2) of Theorem 2. Then there is a set H of inputs, and there are $W_1, W_2 \subseteq \{1, ..., n\}$, $J_1, J_2 \subseteq \mathcal{I}$ and functions f_1, f_2 so that for all $\chi \in H$ we have:*

(14.6) $|H| \geq 2^{-5k\epsilon n \log n}\alpha^n$

(15.6) $W_1, W_2 \in \Gamma_\chi$

(16.6) $J_i = \mathtt{set}(W_i, \chi)$ *for* $i = 1, 2$

(17.6) *if* $T_i = \bigcup_{I \in J_i} I$ *then* $f_i = \mathtt{rstate}_{T_i,\chi}$ *for* $i = 1, 2$

(18.6) $J_1 \cap J_2 = \emptyset$ *and* $W_1 \cap W_2 = \emptyset$

(19.6) $\mathtt{out}_{kn}(\chi) = 0$

Proof. According to our assumptions about $\mathcal{M}$ there are at least $\frac{1}{2}\alpha^n$ inputs χ with (19.6). Assume that a χ is fixed with (19.6). By (13.5) of Lemma 5 there are $W_1, W_2 \in \Gamma_\chi$ so that $W_1 \cap W_2 = \emptyset$. Let $J_i = W_i$, $T_i = \bigcup_{I \in J_i} I$, $f_i = \mathtt{rstate}_{T_i,\chi}$, for $i = 1, 2$. Clearly (15.6), (16.6), (17.6), (18.6) are satisfied. We show that the number of possible choices for the sequence $\langle W_1, W_2, J_1, J_2, f_1, f_2\rangle$ is at most $2^{-1+5k\epsilon n \log n}$. Therefore for at least one choice of the sequence the number of corresponding inputs χ is at least $\frac{1}{2}\alpha^n 2^{1-5k\epsilon n \log n} = 2^{-5k\epsilon n \log n}\alpha^n$.

The number of choices for the pair $W_1, W_2 \subseteq \{1, ..., n\}$ is at most 2^{2n}. The number of choices for the pair J_1, J_2 is at most $|\mathcal{I}|^{4k}$. Since $\mathcal{I}$ has at most $\sigma^{-1}k$ elements this is at most $\sigma^{-4k}k^{4k}$.

The domains of the functions f_i contain at most $2k$ elements since $\mathtt{rigth}(T_i)$ has at most $2k$ elements. The range of each function is in $\mathcal{S}$ which has at most $\beta \leq 2^{\epsilon n \log n}$ elements. Therefore the number of possible pairs of functions f_1, f_2 is at most $2^{4k\epsilon n \log n}$.

The product of all of our upper bounds is at most $2^{-1+5kn \log n}$ if n is sufficiently large. *Q.E.D.*(Lemma 6)

Lemma 7. *Assume that X is a set of functions defined on the set A with values in $\{1, ..., \alpha\}$ and $|B| \subseteq A$. For each fixed $f \in X$ let $\nu_B(f)$ be the number of $g \in X$ so that f and g is identical on $A - B$. Then for each $\lambda > 0$ the number of functions f with $\nu_B(f) \leq \lambda |X| \alpha^{-|A-B|}$ is at most $\lambda |X|$.*

Proof. We partition X into at most $\alpha^{|A-B|}$ classes: f, g are in the same class if their restrictions to $A - B$ is identical. The average size of a class is therefore at

least $|X|\alpha^{-|A-B|}$. Therefore the number of functions counted in $\nu_B(f)$ are in classes whith sizes less than λ times the average. Clearly these classes can cover at most $\lambda|X|$ elements of $|X|$.

Lemma 8. *Assume that $1 > \kappa > 0$, $\rho > 0$ and τ is sufficiently small with respect to both κ and ρ, $\omega = n^{-\rho}$ and n is sufficiently large with respect to τ, and X is a set of inputs with $|X| \geq 2^{-\tau n \log n}\alpha^n$. Then there is a $\chi \in X$ so that for all $Z \subseteq \{1,...,n\}$, $|Z| \geq \kappa n$ there are at least $(\omega\alpha)^{|Z|}$ partial inputs η defined on Z so that $\chi \wr \eta \in X$.*

Proof. Let Z be a fixed set with $|Z| \geq \kappa n$. We apply Lemma 7 with $\lambda \to |X|^{-1}\alpha^{n-|Z|}(\omega\alpha)^{|Z|}$, $A \to \{1,...,n\}$, $B \to Z$. We get that the number of inputs χ with $\nu_Z(\chi) \leq |X|^{-1}\alpha^{n-|Z|}(\alpha\omega)^{|Z|}|X|\alpha^{-|n-|Z||} = (\alpha\omega)^Z$ is at most $\lambda|X| \leq 2^{\tau n \log n}\alpha^{-n}\alpha^{n-|Z|}(\omega\alpha)^{|Z|}|X| = 2^{\tau n \log n}\omega^{|Z|}|X| \leq 2^{\tau n \log n}\omega^{\kappa n}|X|$. This holds for all sets $Z \subseteq \{1,...,n\}$ with $\kappa n \leq |Z|$. There are at most 2^n such sets. Therefore we have that the number of inputs χ so that $\nu_Z(\chi) \leq (\alpha\omega)^Z$ for at least one Z is not larger than $2^n 2^{\tau n \log n}(\omega)^{\kappa n}|X| = 2^{(1+\tau \log n + \kappa \log_2 \rho)n}|X| = 2^{(1+\tau \log n - \kappa\rho \log n)n}|X|$. Since τ is sufficiently small with respect to κ and ρ, the exponent in the last expression is negative, that is, there is a $\chi \in X$ with $\nu_Z(\chi) > (\omega\alpha)^{|Z|}$ for all Z of the required size which implies the statement of the lemma. *Q.E.D.*(Lemma 8).

Now we may complete the proof of Theorem 2. Let H, W_i, J_i, T_i, f_i, $i = 1,2$ be the set and functions defined in Lemma 6. We apply Lemma 8 with $X \to H$, $\rho \to \delta$ $\tau \to 5k\epsilon$, $\kappa \to \frac{1}{4}\sigma^{2k}k^{-2k}$. Let χ be an element of H whose existence is stated in the lemma. Lemma 6 implies that $W_i \in \Gamma_\chi$ for $i = 1,2$. Therefore by (11.5) of Lemma 5 we have that the conclusion of Lemma 8 holds with $Z \to W_i$, $i = 1,2$. So if A_i is the set of all partial inputs defined on W_i so that $\chi \wr \eta \in H$ then $|A_i| \geq (n^{-\delta}\alpha)^{|W_i|}$.

We claim that χ, A_i, T_i satisfy the conditions of Lemma 4.

(8.4) follows from the definiton of A_i.

(10.4) as we have already seen is a consequence of Lemma 8.

Assume now that $\eta_1 \in A_1$ and η_2 in A_2 and check the individual conditions in (9.4).

(4.2) is a consequence of (18.6).

(5.2) follows from (16.6) and from the definition of T_i in (17.6).

(6.2) We know that $\chi \wr \eta_i \in H$ and so (17.6) implies this condition.

(7.2) $\chi \wr \eta_j \in H$ therefore by (15.6) $\mathtt{domain}(\eta_i) = W_i$ is a class of $\Gamma_{\chi \wr \eta_j}$. Consequently its registers are not accessed outside T_i. As we have already seen T_1 T_2 are disjoint that is none of the registers of W_i is accessed in T_j at input $\chi \wr \eta_j$.

Since all of the requirements of Lemma 4 are met we have that there exist η_1, η_2, x_1, x_2 with the properties described in the lemma. By (19.6) $\mathtt{out}_{kn}(\chi) = 0$ and so by Lemma 4 $\mathtt{out}_{kn}((\chi \wr \eta_1) \wr \eta_2) = 0$ while x_1 and x_2, which are taken by $(\chi \wr \eta_1) \wr \eta_2$ at

different points satisfy the relation Q. This is clearly in contradiction with property (2). *Q.E.D.*(Theorem 2)

Remark. There is a way to simplify slightly the proof of Theorem 2. Namely with a somewhat weaker the lower bound in (14.6) we may include in Lemma 6 the following condition

(20) for all $\eta, \xi \in H$ we have $\Gamma_\eta = \Gamma_\xi$.

Indeed the number of possible sets Γ_η is at most $c(\sigma,k)^n$, where $c(k,n)$ depends on only σ and k. The reason for this is that Γ_η is a set of disjoints subsets of $\{1,...,n\}$ and $|\Gamma_\eta|$ remains below a bound depending only on σ and k. Therefore we lose only a factor of $2^{\log c(\sigma,k)n}$ in the lower bound of (14.6). We may conclude the proof using the modified version of Lemma 6 with only Lemma 7, that is the more complicated Lemma 8 is not needed. The disadvantage of this proof is that it is less suitable for further improvements, for example for proving the theorem for λ-full relations with greater values of λ. The reason for this is that the the lower bound in (14.6) for the simplified proof will decrease faster (than the corresponding bound for the original proof) if λ gets larger. (The partition Γ_η will have more classes and so the total number of possible partitions Γ_η increases drastically.)

Proof of Theorem 1. In the proof we will use the following three well-known facts. $A^{(m)}$ will denote the set of all $0,1$ sequences of length m. The first is the following theorem of Harper. (For a proof see e.g. Bollobás, Combinatorics, §16. ([Bo])).

Theorem A (Harper). *Assume that for all $X \subseteq A^{(m)}$, $N(X)$ is the set of sequences whose Hamming distance is at most one from at least on point of X. Then for all $X \subseteq A^{(m)}$, $|X| \geq \sum_{i=0}^{r} \binom{m}{i}$ implies $N(X) \geq \sum_{i=0}^{r+1} \binom{m}{i}$.*

Proposition 1. *for all $0 < a < \frac{1}{2}$ there is a $0 < b < 1$ so that if m is suffciently large then $\sum_{0 \leq i < am} \binom{m}{i} < 2^{bm}$*

This can be proved by estimating the binomial coefficients using Stirling's formula.

Proposition 2. *for all $0 < \gamma < \frac{1}{2}$, there is a $\theta > 0$ so that if m is sufficiently large and x, y are random $0,1$ sequences taken independently and with uniform distribution from the set of all such sequences, then the probability that the Hamming distance of x and y is smaller than γ is less than $2^{-\theta m}$.*

This follows easily from Proposition 1.

In the proof of Theorem 1 we assume, for the sake of simplicity, that α is a power of 2 namely $\alpha = 2^m$. Let Q_ξ be the following relation on $0, 1, ..., \alpha - 1$: "the Hamming distance of x and y is less than ξ". First we prove that

(21) *For all $\gamma > 0$ there is a $0 < \delta' < 1$ so that if m is sufficiently large then $Q_{\gamma m}$ is $\alpha^{-\delta'}$ full.*

Proof. Using Proposition 1 with $a \to \frac{1}{2} - \frac{\gamma}{3}$ we pick a $0 < \delta' < 1$ so that $\sum_{0 \le i < (\frac{1}{2} - \frac{\gamma}{3})m} \binom{m}{i} < 2^{(1-\delta')m}$. Assume that $B_i \subseteq A^{(m)}$, $|B_i| > 2^{-\delta' m} 2^m$ for $i = 1, 2$. We claim that $N_{\frac{1}{2}\gamma m}(B_i)$, that is, the set of sequences whose Hamming distance from at least one point in B_i is at most $\frac{1}{2}\gamma m$, contains more than $\frac{1}{2}2^m$ elements for $i = 1, 2$. This will imply that $N_{\frac{1}{2}\gamma m}(B_1) \cap N_{\frac{1}{2}\gamma m}(B_2) \neq \emptyset$ and therefore B_1, B_2 have points whose distance is at most γm. We have $\sum_{0 \le i < (\frac{1}{2} - \frac{\gamma}{3})m} \binom{m}{i} < 2^{(1-\delta')m} \le |B_i|$. Therefore applying Theorem A repeatedly we get by induction on j that for all $j \ge 0$ $\sum_{0 \le i < j + (\frac{1}{2} - \frac{\gamma}{3})m} \binom{m}{i} \le |N_j(B_i)|$. Now let $j = [\frac{\gamma}{2}m]$. $j + (\frac{1}{2} - \frac{\gamma}{3})m \ge \frac{m}{2} + 1$. We have $2^{m-1} < \sum_{0 \le i \le \frac{n}{2}+1} \binom{m}{i} \le N_j(B_i) \le N_{\frac{\gamma}{2}m}(B_i)$, which completes the proof of (21).

We need the following to show that Q satisfies (1):

(22) *For all $0 < \gamma < \frac{1}{2}$ if c_1 is sufficiently large (where $\alpha = 2^m > n^{c_1}$), m is sufficiently large with respect to c_1, and η is a random input with uniform distribution on the set of all inputs, then with a probability of at least $\frac{1}{2}$ we have that for all $1 \le i < j \le n$ $\neg Q_{\gamma m}(\eta(i), \eta(j))$.*

Proof of (22). For any fixed $1 \le i < j \le n$ let $p_{i,j}$ the probability of the event $Q_{\gamma m}(\eta(i), \eta(j))$. By Proposition 2 we have that this probability is at most $2^{-\theta m}$ where θ depends only on γ. Let c_1 be sufficiently large with respect to θ. We have $p_{i,j} \le 2^{-\theta m} \le 2^{-\theta c_1 \log_2 n} \le 2^{-3 \log_2 n} \le n^{-3}$. Therefore the probability that $Q(\eta(i), \eta(j))$ holds for at least on pair i, j, $i \neq j$ is at most $\binom{n}{2} n^{-3} < \frac{1}{2}$ which completes the proof of (22).

We may now complete the proof of Theorem 1. We show that the requirements of Theorem 2 are met with $Q \to Q_{\gamma m}$ and $\delta \to \delta' c_1^{-1}$ where δ' satisfies (21). This choice for δ implies that Q is indeed $n^{-\delta} = \alpha^{-\delta'}$-full. (1) is a consequence of (22) and (2) implies the conclusion of Theorem 1. (The proof also shows that the "promise problem" version of Thoerem 1, mentioned in the introduction also holds.)

The element distinctness problem. In this section we prove the analogue of Theorem 1 for the equality relation.

Theorem 3. *For all positive integer k there is an $\epsilon > 0$ so that if n is sufficiently large, $\alpha \ge n^2$, $\beta < 2^{\epsilon n}$ then the following holds:*

there is no machine $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ so that for any input η we have $\mathtt{out}_{kn}(\eta) = 1$ iff there are $i, j \in \{1, ..., n\}$, $i \neq j$ so that $\eta(i) = \eta(j)$.

The theorem holds not only for the equality relation but for a larger class of λ-full relations as formulated below in Theorem 4.

Theorem 4. *For all $c_0 > 0$ and positive integer k there is an $\epsilon > 0$ so that if n is sufficiently large, $\alpha \geq n^2$, $\beta < 2^{\epsilon n}$ and $\mathcal{R}(x,y)$ is a $\frac{1}{2}$-full and (c_0, n) sparse relation then the following holds:*

there is no machine $\mathcal{M} = \langle n, \alpha, \beta, kn \rangle$ so that for any input η we have $\mathtt{out}_{kn}(\eta) = 1$ *iff there are $i, j \in \{1, ..., n\}$, $i \neq j$ so that $\mathcal{R}(\eta(i), \eta(j))$.*

Remarks. 1. The theorem remains true if we assume only that $\mathcal{R}$ is $(2^{-\epsilon' n}, n)$-sparse, where $\epsilon' > 0$ is sufficiently small with respect to k, and our proof actually gives this result without any substantial changes.

2. The theorem remains true if we assume only that the relation $\mathcal{R}$ is $1-\delta$-full where δ is an arbitrarily small constant. We need only minor modifications of the present proof to get this result. In this case $\epsilon > 0$ may depend on δ too. The mentioned two changes of parameters in the theorem can be executed simultaneously.

First we show that Theorem 4 implies Theorem 3. The relation equality is clearly $\frac{1}{2}$-full. We show that if $\alpha \geq n^2$, then it is (c_1, n)-sparse on $\{1, ..., \alpha\}$ for some absolute constant c_1. We pick $a_1, ..., a_n$ independently and with uniform distribution from $\{1, ..., \alpha\}$. If p_0 is the probability of the event "there is no $0 < i < j \leq n$ so that $\mathcal{R}(a_i, a_j)$" then $p_0 \geq \prod_{i=1}^{n-1}(1 - \frac{i}{n^2}) \geq \prod_{i=1}^{n-1}(1 - \frac{n}{n^2}) = \prod_{i=1}^{n-1}(1 - \frac{1}{n}) \sim \frac{1}{e}$. Therefore there is a constant $c_1 > 0$ so that for any sufficiently large n we have $p_0 > c_1$.

Proof of Theorem 4. Assume now that $c_0 > 0$, k is an arbitrary integer, and let $\epsilon > 0$ be sufficiently small with respect to k and n is sufficiently large with respect to k, ϵ. Suppose further that contrary to the assertion of the theorem there is a machine $\mathcal{M}$ with the described restrictions on its parameters, which decides for an arbitrary input η whether it has two identical values. Let H_1 be the set of all inputs χ with $\mathtt{out}_{kn}(\chi) = 0$. The assumed (c_0, n) sparsity of $\mathcal{R}$ implies that $|H_1| > c_0 \alpha^n$. Our plan is to to find a $\chi \in H_1$ and two partial inputs η_1, η_2 with disjoint domains so that $\mathtt{out}_{kn}(\chi \, l \, \eta_1 \, l \, \eta_2) = \mathtt{out}_{kn}(\chi) = 0$ and $\mathcal{R}(\eta_1(u), \eta_2(v))$ for some $u \in \mathtt{domain}(\eta_1)$ $v \in \mathtt{domain}(\eta_2)$. Since $\chi \, l \, \eta_1 \, l \, \eta_2$ takes these values (and at different places) we reach a contradiction.

In the proof we will assume that c_0, k, ϵ, n, and the machine $\mathcal{M}$ has been fixed with the mentioned properties. All of the notions which depend on a machine (e.g. the functions `rstate`, `register` etc.) will refer to this particular machine. Sometimes we will define numbers which depend e.g. only on k but not on n. In this case we state the quantification of the variables k, ϵ, n again, and, even if we do not mention the machine $\mathcal{M}$ itself, we assume that it is fixed with the given properties.

We have defined earlier the set of intervals $\mathcal{I}$ depending on a parameter σ. We will use the same defintion here as well. We have already defined, before Lemma 2, the functions $\mathtt{register}(T,\eta)$, and $\mathtt{core}(T,\eta)$ where T is a set of times, and η is an input. We will extend this notation, namely if $F \subseteq \mathcal{I}$ then by definition $\mathtt{register}(F,\eta) = \mathtt{register}(\bigcup F,\eta)$ $\mathtt{core}(F,\eta) = \mathtt{core}(\bigcup F,\eta)$.

Definitions. 1. If $F \subseteq \mathcal{I}$ and χ is an input then $\mathtt{stem}(F,\chi)$ will denote the restriction of χ onto $\{1,...,n\} - \mathtt{core}(F,\chi)$. $\mathtt{fan}(F,\chi)$ will be the set of all inputs η with $\mathtt{stem}(F,\chi) = \mathtt{stem}(F,\eta)$. If H is a set of inputs then $\mathtt{fan}(H,F,\chi) = H \cap \mathtt{fan}(F,\chi)$

2. An input χ is called visible if every elements of $\{1,...,n\}$ (that is, every registers) are accessed at input χ.

Remark. Without loss of generality we may assume that every input is visible. Indeed e.g. we may suppose that our machine starts to work by accessesing each register once. This assumption adds only n to the time needed to solve any problem. In the proof we will assume that every input is visible. As a consequence we have that $\mathtt{core}(F,\chi)$ consists of those registers which are not accessed outside F, (in contrast to the original defitnion where we assumed that they are accessed in F.)

Sketch of the proof of Theorem 4. As in the proof of Theorem 2 our plan is the following. We start with the set H_1 of all inputs where the output of the machine is 0. We have $|H_1| > c_0\alpha^n$. Our goal is to find a $\chi \in H_1$ and two disjoint sets of registers U_1, U_2 and on U_i a set of partial inputs Y_i, $i = 1,2$ so that

(a) $|Y_i|$ is so large that it guarantees that $|\bigcup_{\eta\in Y_i} \mathtt{range}(\eta)| > \frac{\alpha}{2}$ and

(b) for all $\eta_1 \in Y_1$, $\eta_2 \in Y_2$ we have $\mathtt{out}_{kn}(\chi \wr \eta_1 \wr \eta_2) = 0$.

This leads to a contradiction since (a) implies that there are $x_i \in \bigcup_{\eta\in Y_i} \mathtt{range}(\eta)$ so that $\mathcal{R}(x_1,x_2)$ and therefore we may pick $\eta_1 \in Y_1, \eta_2 \in Y_2$ so that $x_i \in \mathtt{range}(\eta_i)$ which implies that $\chi' = \chi \wr \eta_1 \wr \eta_2$ takes the values x_1, x_2 at different places.

We will pick the sets of registers U_i in the following way. As in the proof of Theorem 2 we define a partition $\mathcal{I}$ of the time interval $[0, k-1)$ into subintervals whose length is about σn where $\sigma > 0$ is sufficiently small with respect to k, but $\epsilon > 0$ (from the upper bound on the memory) is sufficiently small with respect to σ. We will pick at random two disjoint subsets F_1 and F_2 of $\mathcal{I}$. (The common size of F_i, $i = 1,2$ will be chosen carefully. We will return to this question in the remark after Lemma 12.) Suppose F_i has been selected. We will define U_i by $U_i = \mathtt{core}(F_i,\chi)$ for a suitably chosen input $\chi \in H_1$. We will show that there is a large set $H \subseteq H_1$ of inputs so that any $\chi \in H$ will be good in the definition of U_i and Y_i will be defined as the set of all partial inputs η defined on U_i with the property $\chi \wr \eta \in H$ for $i = 1,2$. We will give the definition of the sets F_i is three steps. In each step we reduce the requirements on F_i, H to another set of requirements, (which may not be simpler but can be more easily satisfied) and in the last step with a probabilistic construction we show that our requirements can be met. The three steps will be described by three lemmata

Lemma 9, Lemma 11 and Lemma 13 (for the understanding of Lemma 13, Lemma 12 and the definition of the function $\mathtt{acc}_l$ before it is also necessary). The reader who want first to get a complete picture about structure of the proof without going into the technical details may read the statements of the given lemmata and definition together with the remarks immediately before and after each lemma.

Lemma 9. *For all positive integer k if $\sigma > 0$ is sufficiently small then there is a $\lambda > 0$ so that if $\epsilon > 0$ is sufficiently small and n is sufficiently large and G is a set of visible inputs then the following holds. There exist F_1, F_2, f_1, f_2, H with the following properties:*
(23.9) *$H \subseteq G$ and $|H| \geq 2^{-\lambda n}|G|$*
(24.9) *F_1, F_2 are disjoint subsets of $\mathcal{I}$*
(25.9) *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 $\mathtt{core}(F_j, \chi) = \mathtt{core}(F_j, \xi)$*
(26.9) *$|\mathtt{core}(F_i, \chi)| \geq 2\lambda n$ for all $\chi \in H$ and $i = 1, 2$*
(27.9) *$\mathtt{rstate}_{\chi, \bigcup F_i} = f_i$ for all $\chi \in H$, $i = 1, 2$.*

Remarks. 1. As a motivation for this lemma we describe where the proof of Theorem 2 breaks down if we try to adapt it to the present problem. We also point out the changes which make after all the basic proof technique applicable. The statement of the present lemma contains all the necessary modifications. The lemma in itself implies the theorem. (The proof is given right after these remarks.) The remaining part of this section is the proof of the lemma.

In the proof of Theorem 2 we used the following observation. If X is a large set of inputs and B is a large set of registers, then for a random element χ of X we have that with high probability there are many different partial inputs η on B so that $\chi \wr \eta \in X$. (See Lemma 7). In applying this lemma one of the difficulties is that the set B is fixed in advance, it cannot depend on the input χ. In other words if we first pick a random χ then depending on it a set B_χ then there is no guarantee that a similar assertion will hold. (In the case of Theorem 2 something like that was still true, as formulated in Lemma 8. We may think that for each B the number of exceptional inputs χ, where B does not behave in the required way, was so small that even if we threw out the exceptional inputs for *all* B still enough inputs remained. This is not the case with the present choice of the parameters.) The need for picking first χ and then B comes from the fact that one of the conditions of Lemma 2 is $\mathtt{domain}(\eta_i) = \mathtt{core}(T_i, \chi)$. $\mathtt{domain}(\eta_i)$ will have the role of set B so we will know it only after χ has been selected.

However, in the special case $B = \mathtt{core}(F_1, \chi)$, an analogue of Lemma 15 holds in spite of the fact that B depends on χ. The reason for this is the following. In this

special case we may pick a random input $\chi \in H$ in the following way. First we want to randomize χ outside $\mathtt{core}(F_1, \chi)$. The problem is that we do not know yet what is the set $\mathtt{core}(F_1, \chi)$. However the set $\{1, ..., n\} - \mathtt{core}(F_1, \chi)$ can be randomized at the same time as we randomize the value of χ on it (with the condition $\chi \in H$). We can do it by performing the computation in each interval of $\mathcal{I}$ outside F_1 and picking a random value for the content of any registers (according to the distribution induced by H) that is accessed during this computation. Condition (27.9) guarantees that we can start performing the computation at the left border of any interval of $\mathcal{I} - F_1$ which comes right after an interval of F_1. (Lemma 10 formulated and proved below ensures that we get this way the right distribution.) Therefore we can decide what will be $\{1, ..., n\} - \mathtt{core}(F_1, \chi)$ and therefore $\mathtt{core}(F_1, \chi)$ without giving any information about the values of χ on $\mathtt{core}(F_1, \chi)$. Therefore if we now continue the selection of χ onto $\mathtt{core}(F_1, \chi)$ the situation is the same as if $\mathtt{core}(F_1, \chi)$ would be the fixed subset B of Lemma 7. Condition (26.9) of the lemma guarantees that this set will be sufficietly large.

There is another problem with the proof of Theorem 2 under the present circumstances. In that proof we guaranteed condition (7.2) of Lemma 2, that is the fact that in the time set T_j during the computation at input $\chi \wr \eta_j$ we will not look at any of the registers in $\mathtt{domain}(\eta_i)$, in a very strong way. Namely we were able to prove that the times were we look at $\mathtt{domain}(\eta_i)$ is the same at the inputs χ and $\chi \wr \eta_i$. This cannot be ensured now. The counting argument in the proof of this fact breaks down because of the changes in the values of the parameters. Therefore we guarantee condition (7.2) in a new way through condition (25.9). Since in our case $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i, \chi)$ this condition with the original notation just says that $\mathtt{core}(F_j, \chi) = \mathtt{core}(F_j, \chi \wr \eta_i)$. This, by the definition of the function $\mathtt{core}$, will easily imply the required property. Actually the implication is so easy that we may feel that we only reformulated the original requirement. Indeed the most difficult part of the proof of Theorem 4 is that F_1, F_2 can be selected with property (25.9) and the other properties in the lemma. The main idea behind this part of the proof is the following. In (25.9) we want that $\mathtt{stem}(F_i, \chi)$ uniquely determines $\mathtt{core}(F_j, \chi)$. As a first step we try to satisfy all of the other conditions of the lemma and a weakened version of (25.9) where $\mathtt{stem}(F_i, \chi)$ does not determine $\mathtt{core}(F_j, \chi)$ uniquely but leaves relatively few choices for it. (Lemma 11.) The fact that this is possible is not that surprising since if we change χ on $\mathtt{core}(F_i, \chi)$ then the contents of at most $2\sigma^{-1}|F_i|$ registers are changed. (Assuming that (27.9)) holds for the inputs involved.) If these registers would be randomly distributed with respect to $\mathtt{core}(F_j, \chi)$ then relatively few would be in it, so a change in χ on $\mathtt{core}(F_i, \chi)$ would change $\mathtt{core}(F_j, \chi)$ only a little and so that total number of sets $\mathtt{core}(F_j, \chi)$ would be small. (The assumed randomness will be guaranteed by the random choice of F_1 and F_2.) Form this weakened version

of (25.9) we get the original one by taking a subset of the input set H in a way that from the few choices for $\mathtt{core}(F_j, \chi)$ only one remains.

2. If we want to prove the modified version of Theorem 4 where only $1-\delta$-fullness is assumed about the relation $\mathcal{R}$, then we have to replace $2\lambda n$ by $\bar{c}\lambda n$ where $\bar{c}$ is sufficiently large with respect to δ. In the proof of Lemma 9 this additional requirement does not cause any difficulties.

Before we give the proof of Lemma 9 we show that it implies Theorem 4.

Proof of Theorem 4. We apply Lemma 9 with $G = H_1$. (As we have noticed already after the definition of visibility we may assume that every input is visible.) Let $\lambda, F_1, F_2, f_1, f_2, H$ be given with the properties described in the lemma.

We show that

(28) *the number of elements of $\chi \in H$ which satisfy the following condition is at least $\frac{1}{3}|H|$: $|\mathtt{fan}(H, F_i, \chi)| \geq c_0 \frac{1}{3} 2^{-\lambda n} \alpha^{s_i}$, where $s_i = |\mathtt{core}(F_i, \chi)|$, $i = 1, 2$.*

First we estimate the number of inputs χ so that (28) holds for only a single fixed value of i, say, $i = 1$. We define a partition $\mathcal{T}$ of H by "$\chi, \xi \in H$ are in the same class iff $\mathtt{stem}(F_1, \chi) = \mathtt{stem}(F_1, \xi)$". The following lemma says that if χ and ξ are not in the same class then the functions (partial inputs) $\mathtt{stem}(F_1, \chi)$, $\mathtt{stem}(F_1, \xi)$ are not only different but also incompatible, that is, there is a $x \in \mathtt{domain}(\mathtt{stem}(F_1, \chi)) \cap \mathtt{domain}(\mathtt{stem}(F_1, \xi))$ so that $\mathtt{stem}(F_1, \chi)(x) \neq \mathtt{stem}(F_1, \xi)(x)$. (This needs to be proved because two different functions may be compatible if they have different domains, but take identical values in points where both are defined.)

Lemma 10. *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 $x \in \mathtt{domain}(\chi) \cap \mathtt{domain}(\xi)$ so that $\chi(x) \neq \xi(x)$*

Proof. We show that if $\mathtt{stem}(F, \chi)$ and $\mathtt{stem}(F, \xi)$ are compatible then they are identical. The set $\bigcup_{I \in \mathcal{I} - F} I$ can be covered by intervals. Let $J_1, ..., J_r$ be such a covering where the number of intervals is minimal. Let us consider the computation in a time interval J_l for an arbitrary l, at both inputs χ and ξ. $\mathtt{rstate}_{\chi, \bigcup F} = \mathtt{rstate}_{\xi, \bigcup F}$ implies that the computation starts at the same state of the machine for the two inputs. The compatibility of $\mathtt{stem}(F, \chi)$ and $\mathtt{stem}(F, \xi)$ implies that the two computations will be exactly the same until a register is accessed which is not in the domain of one of these inputs. This, by the definition of $\mathtt{stem}$, cannot happen in the interval J_l which is disjoint from F. Since every register in the domain of either $\mathtt{stem}(F, \chi)$ or $\mathtt{stem}(F, \xi)$ is accessed at least once in $\bigcup J_j$ we get that their domains are equal. *Q.E.D.*(Lemma 10)

Let H' be the set of those inputs ζ which are the extensions of a $\mathtt{stem}(F_1, \chi)$ for some $\chi \in H$. We define a partition $\mathcal{T}'$ of H' in the following way. The inputs $\zeta_1, \zeta_2 \in H'$ belong to the same class of $\mathcal{T}'$ iff there is a $\chi \in H$ so that both ζ_1 and ζ_2 are extensions of $\mathtt{stem}(F_1, \chi)$. Clearly $H \subseteq H'$ and every class of $\mathcal{T}$ is contained in a class of $\mathcal{T}'$.

Proposition 3 . *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 $|C||C'|^{-1} \leq \lambda d$*

Proof. Let X be the set of classes C of P with $|C||C'|^{-1} \leq \lambda d$, where C' is the unique P' class containing C. The total number of elements x with the required property is at most $\sum_{C \in X} |C| \leq \sum_{C \in X} \lambda d|C'| \leq \lambda d \sum_{C \in P} |C'| = \lambda d|A'| = \lambda|A|$. *Q.E.D.*(Proposition 3)

Let d be the density of H in H' that is $d = |H||H'^{-1}|$. (23.9) and the defintion of c_0 implies that the density of H in the set of all inputs is at most $c_0 2^{-\lambda n}$. H' is a set of inputs so we have $d \geq c_0 2^{-\lambda n}$. We apply Proposition 3 with $A \to H$, $P \to \mathcal{T}$, $\lambda \to \frac{1}{3}$. We get that the number of elements χ of H which belong to a $\mathcal{T}$-class whose density in the corresponding $\mathcal{T}'$-class is at most $\frac{1}{3}d \geq \frac{1}{3}c_0 2^{-\lambda n}$ is at most $\frac{1}{3}|H|$. The same is true if we define the corresponding partitions for F_2. Therefore at least $\frac{1}{3}|H|$ elements belong to a class of $\mathcal{T}$ whose density in $\mathcal{T}'$ is at least $\frac{1}{3}c_0 2^{-\lambda} n$ for $i = 1, 2$. This imples (28). Therefore there is a $\chi \in H$ so that the following holds for $i = 1, 2$:

(29) *assume that $s_i = |\mathtt{core}(F_i, \chi)|$, and Y_i is the set of all partial inputs η defined on $\mathtt{core}(F_i, \chi)$ so that $\chi \wr \eta \in H$. Then $|Y_i| \geq c_0 \frac{1}{3} 2^{-\lambda n} \alpha^{s_i}$.*

Therefore Lemma 3 and (26.9) implies that $|\bigcup\{\mathtt{range}(\eta) | \eta \in Y_i\}| > \frac{\alpha}{2}$ for $i = 1, 2$. Consequently, by the $\frac{1}{2}$-fullness of the relation $\mathcal{R}$, there are $u_i \in \eta_i \in Y_i$ for $i = 1, 2$ so that $\mathcal{R}(u_1, u_2)$.

We claim that the conditions of Lemma 2 are satisfied by χ, η_1, η_2, $T_1 \to \bigcup F_1$, $T_2 \to \bigcup F_2$.

(3.2). By (24.9) F_1 and F_2 are disjoint, so according to the definition of the function $\mathtt{core}$, the sets $\mathtt{core}(F_i, \chi) = \mathtt{domain}(\eta_i)$, $i = 1, 2$ are also disjoint.

(4.2) follows from (24.9)

(5.2) This holds with equality.

(6.2) This is a consequence of $\chi \wr \eta_i \in H$ and (27.9)

(7.2) $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i, \chi)$. (25.9) and $\chi \wr \eta_j \in H$ implies that $\mathtt{core}(F_i, \chi) = \mathtt{core}(F_i, \chi \wr \eta_j)$. We got $\mathtt{domain}(\eta_i) = \mathtt{core}(F_i, \chi \wr \eta_j)$. Since F_1 and F_2 are disjoint none

of the registers of $\mathtt{core}(F_i, \chi \wr \eta_j)$ is accessed from $\bigcup F_j$ at input $\chi \wr \eta_j$. *Q.E.D.*(Lemma 9).

Remark. We will use the following lemma in the proof of Lemma 9. This Lemma is similar in content to Lemma 9, only its conditions are somewhat relaxed. E.g. (25.9) which states that $\mathtt{stem}(F_i, \chi)$ is uniquely determines $\mathtt{core}(F_j, \chi)$ is replaced replaced by (33.11) and (34.11) which require only that for most of the inputs χ if $\mathtt{stem}(F_i, \chi)$ is given then there are relatively few choices for $\mathtt{core}(F_j, \chi)$. Condition (27.9) is left out altogether, since our bound on the working memory implies that that there are relatively few choices for the functions $\mathtt{rstate}_{\chi, \bigcup F_i}$, so as in the proof of Theorem 2 we may guarantee the condition by taking a subset of H.

Lemma 11. *For all positive integer k there is a $0 < \tau < 1$ so that for all sufficiently small $\sigma > 0$, there is a $\kappa > 0$, so that for all sufficiently small $\epsilon > 0$ and sufficiently large n if G is a set of visible inputs, then there exist F_1, F_2, $\mathcal{L}_1, \mathcal{L}_2$, D, D_0, with the following properties:*

(30.11) *κ is sufficiently small with respect to τ*

(31.11) *$D_0 \subseteq D \subseteq G$ and $|D| \geq 2^{-\kappa n}|G|$ and $|D_0| \geq \frac{3}{4}|D|$.*

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

(33.11) *For all $i = 1, 2$, $\mathcal{L}_i(x, y)$ is a binary function which is defined for all pairs x, y, where x is a partial input and y is a subset of $\mathtt{domain}(x)$. Moreover for each fixed x_0, $\mathcal{L}_i(x_0, y)$ as a function of y is a one-to-one map of the set of subsets of $\mathtt{domain}(x_0)$ onto a set of positive integers.*

(34.11) *for all $\chi \in D_0$ and $i = 1, 2$, $j = 3 - i$ the number of elements ξ of the set $\mathtt{fan}(D, F_i, \chi)$ with $\mathcal{L}_i(\mathtt{stem}(F_i, \chi), \mathtt{core}(F_j, \xi)) \leq 2^{\kappa n}$ is at least $\frac{7}{8}|\mathtt{fan}(D, F_i, \chi)|$.*

(35.11) *$|\mathtt{core}(F_i, \chi)| \geq \kappa^{\tau} n$ for $i = 1, 2$ and $\chi \in D_0$*

We show first that Lemma 11 implies Lemma 9. Assume that k is given and we pick a τ according to Lemma 11. Suppose now that $\sigma > 0$ is sufficiently small. and we pick κ according to Lemma 11. Let $\lambda = 4\kappa$ and suppose that $\epsilon > 0$ is sufficiently small, n is sufficiently large and G is a set of visible inputs. Let $F_1, F_2, \mathcal{L}_1, \mathcal{L}_2, D$ be the elements whose existence is guaranteed by Lemma 11.

We define two functions $\mathcal{D}_i$, $i = 1, 2$ on D by
$\mathcal{D}_i(\chi) = \mathcal{L}_1(\mathtt{stem}(F_i, \chi), \mathtt{core}(F_{2-i}, \chi))$. Let D' be the subset of D_0 where the values of both $\mathcal{D}_1$ and $\mathcal{D}_2$ are at most $2^{\kappa n}$. If $D^{(i)}$ is the set of all elements of D where the value of $\mathcal{D}_i$ is at most $2^{\kappa n}$, then $D' = D_0 \cap D^{(1)} \cap D^{(2)}$. To get a lower bound on $|D'|$ first we give a lower bound on $D^{(i)}$. Assume $i \in \{1, 2\}$ is fixed. We define a partition P_i, of D so that on each class of P_i, $\mathtt{stem}(F_i, \chi)$ as a function of χ is constant for $i = 1, 2$, and P_i is maximal with this property. Let W be a class of P_i. By (34.11) we have that $|W - D^{(i)}| \leq \frac{1}{8}|W|$ and so $|D - D^{(i)}| \leq \frac{1}{8}|D|$. Since this is true for

$i = 1, 2$ we have $|D - (D^{(1)} \cap D^{(2)})| \leq \frac{1}{4}|D|$. By (31.11) $|D_0| \geq \frac{3}{4}|D|$, so we have $|D'| = |D_0 - (D^{(1)} \cap D^{(2)})| \geq \frac{1}{4}|D|$.

We partition D' according to the values of both $\mathcal{D}_1$ and $\mathcal{D}_2$, that is χ and ξ will be in the same class iff $\mathcal{D}_i(\chi) = \mathcal{D}_i(\xi)$, $i = 1, 2$. Let C be a class of this partition with a maximal number of elements. Clearly $|C| \geq 2^{-2\kappa n}|D'| \geq 2^{-2-2\kappa n}|D|$. Now we define a partition P of C by "$\chi, \xi \in C$ are in the same class iff $\mathtt{rstate}_{\chi, \bigcup F_i} = \mathtt{rstate}_{\xi, \bigcup F_i}$" for $i = 1, 2$. Since the number of possible functions $\mathtt{rstate}_{\chi, F_i}$ is at most $2^{\epsilon n 2\sigma^{-1}k}$ and ϵ is sufficiently small with respect to k, σ and κ, we have that there is a class H of this partition so that $|H| \geq 2^{-\epsilon 2\sigma^{-1}kn}|C| \geq 2^{-\epsilon 2\sigma^{-1}kn}2^{-2-2\kappa n}|D| \geq 2^{-3\kappa n}|D| \geq 2^{-4\kappa n}|G|$. The fact that H is a single class of P implies that there are functions f_1, f_2 so that for all $i = 1, 2$ and for all $\chi \in H$ we have $\mathtt{rstate}_{\chi, \bigcup F_i} = f_i$, that is, (27.9) holds.

(23.9) is a conseqence of $\lambda = 4k$ and the inequality $|H| \geq 2^{-4\kappa}|G|$ proved above.

(24.9) follows from (32.11)

(25.9). Suppose $i \in \{1, 2\}$, $j = 3 - i$. $H \subseteq C$ and the definition of C imply that if $\mathtt{stem}(F_i, \chi) = \mathtt{stem}(F_i, \xi)$ then break

$$\mathcal{L}_i(\mathtt{stem}(F_i, \chi), \mathtt{core}(F_j, \chi)) = \mathcal{L}_i(\mathtt{stem}(F_i, \xi), \mathtt{core}(F_j, \xi)).$$

According to (33.11) $\mathcal{L}_i$, as a function of its second variable is a one-to-one map (for a fixed value of the first variable), therefore $\mathtt{core}(F_j, \chi) = \mathtt{core}(F_j, \xi)$.

(26.9) is a consequence of (35.11), $\lambda = 4\kappa$ and $0 < \kappa < 1$, $0 < \tau < 1$ and the fact that κ is sufficiently small with respect to τ.

Definition. If χ is an input, and l is a nonnegative integer, then $\mathtt{acc}_l(\chi)$ will denote the set of all registers which, at input χ, are accessed from exactly l different elements of $\mathcal{I}$.

Remarks. 1 For the defintion of $\mathtt{acc}_l(\chi)$ it is irrelevant whether a register has been accessed only once or several times from an interval of $\mathcal{I}$.

2. The goal of the next lemma is to show that for all visible inputs χ there is an l so that $|\mathtt{acc}_l(\chi)|$ is large, both in an absolute sense and compared to the numbers $|\mathtt{acc}_i|$, $i = 1, ..., l - 1$.

Lemma 12. *Suppose that $k \geq 2$ is a positive integer, $\sigma > 0$ is sufficiently small, n is sufficiently large, and χ is a visible input. Then there is a positive integer l with $1 \leq l \leq 2k$ and a $\mu > 0$, so that $\frac{1}{\mu}$ is an integer and*

(36.12) $|\mathtt{acc}_l(\chi)| \geq \sigma^{\frac{1}{4}\mu}n$,

(37.12) *for all $i = 0, 1, ..., l - 1$ we have $|\mathtt{acc}_i(\chi)| \leq \sigma^{(l+1)\mu}n$*

(38.12) $|\log \sigma|^{-\frac{1}{2}} \log k \leq \mu \leq (10k)^{2k+1}|\log \sigma|^{-\frac{1}{2}} \log k$

Remarks. 1. We will use this lemma to select the common size of the sets F_1 and F_2 whose existence is stated in Lemma 11. F_1, F_2 will be a pair of disjoint

random subsets of $\mathcal{I}$ with t elements where $t = [\sigma^{-1+\mu}]$ and μ is given by Lemma 12. The importance of the gap, described in the lemma, in the sequence $|\mathtt{acc}_i|$ between $i \leq l-1$ and $i = l$ is the following. In the proof of Lemma 11 we will estimate the number of elements of the set $X_0 \cap \mathtt{core}(F_j, \chi)$, where X_0 is a given set of registers. To get an upper bound will be easier if we may disregard the registers in $\mathtt{acc}_i(\chi)$ for $i = 1, ..., l-1$. The inequalities in (36.12) and (37.12) will make this possible. (36.12) will be used again when we prove a lower bound on $|\mathtt{core}(F_j, \chi)|$.

2. The function $|\log \sigma|^{-\frac{1}{2}}$ in the upper and lower bounds of (38.12) can be replaced by any function $f(\sigma) \geq 0$ so that $\lim_{\sigma \to 0} f(\sigma) = 0$ and $\lim_{\sigma \to 0} |\log \sigma| f(\sigma) = \infty$. This change does not affect the application of the lemma in the proof of Lemma 11

Proof. We define a sequence μ_r, l_r, for $r = 0, 1, 2, ...$ by recursion on r until a pair $\mu = \mu_r, l = l_r$ satisfies the conditions of our lemma. The pair μ_r, l_r, $r = 0, ..., 2k$ will satisfy the following conditions:

(39) $|\mathtt{acc}_{l_r}(\chi)| \geq \sigma^{\frac{1}{4}\mu_r} n$,

(40) μ_r is an integer and $l_r \leq 2k$ is a positive integer for all $r \geq 0$, moreover $l_r < l_{r-1}$ for all $r > 0$,

(41) $|\log \sigma|^{-\frac{1}{2}} \log k \leq \mu_r \leq (10k)^{r+1} |\log \sigma|^{-\frac{1}{2}} \log k$

Assume $r = 0$. The number of registers which are accessed at most in $2k$ different intervals of $\mathcal{I}$ is at least $\frac{n}{2}$, otherwise the total number of accesses would be more than kn. Therefore there is an l', $1 \leq l' \leq 2k$ so that there are at least $\frac{1}{4k} n$ registers which are accessed in exactly l' different intervals of $\mathcal{I}$, and so $|\mathtt{acc}_{l'}(\chi)| \geq \frac{1}{4k} n$. Let $\mu' = |\log \sigma|^{-\frac{1}{2}} \log k$, $\mu_0 = [\frac{1}{\mu'}]^{-1}$. $|\mathtt{acc}_{l'}(\chi)| \geq \frac{1}{4k} n \geq k^{-\frac{1}{4}|\log \sigma|^{\frac{1}{2}}} n = \sigma^{\frac{1}{4}\mu'} n \geq \sigma^{\frac{1}{4}\mu_0} n$. Therefore if $l_0 = l'$ then (39) holds with $r = 0$. We have $|\log \sigma|^{-\frac{1}{2}} \log k = \mu' \leq \mu_0 \leq 10 |\log \sigma|^{-\frac{1}{2}} \log k$, therefore μ_0, l_0 meet the requirements of (41). By the definitions of μ_0 and l_0 ((40) also holds.

Assume now that μ_s, l_s has been already defined for $s = 0, .., r-1$ so that (39), (40), (41) hold. If $\mu = \mu_{r-1}, l = l_{r-1}$ satisfy (37.12) we do not define μ_r, l_r. Assume that they do not satisfy (37.12). Then there exists an $l_r < l_{r-1}$, $r \geq 1$ so that $|\mathtt{acc}_{l_r}(\chi)| > \sigma^{(l_{r-1}+1)\mu_{r-1}} n$. (Since χ is visible we may assume that $l_r \neq 0$). Let $\mu_r = 4(l_{r-1} + 1)\mu_{r-1}$. $\neg$(37.12) implies $|\mathtt{acc}_{l_r}(\chi)| \geq \sigma^{\frac{1}{4}\mu_r} n$, and by the inductive assumption we have $\mu_r = 4(l_{r-1}+1)\mu_{r-1} \leq 4(2k+1)\mu_{r-1} \leq 10k\mu_{r-1} \leq (10k)^{r+1} |\log \sigma|^{-\frac{1}{2}} \log k$.

Since the numbers l_r form a decreasing sequence of positive integers and $l_0 \leq 2k$, we have that for some $r \leq 2k-1$ the pair l_r, μ_r cannot be defined. This is only possible if the pair $l = l_{r-1}, \mu = \mu_{r-1}$ meets the requirements of the lemma. *Q.E.D.*(Lemma 12).

Lemma 13. *Assume that $\gamma > 0$, $k \geq 2$ is an integer, $\sigma > 0$ is sufficiently small, n is sufficiently large, G is a set of visible inputs, $\chi \in G$, l, μ are the numbers whose existence are guaranteed by Lemma* 12 *and $t = [\sigma^{-1+\mu}]$. Suppose further that F_1, F_2*

is a random pair of disjoint subsets of $\mathcal{I}$ each with t elements taken with uniform distribution from the set of all pairs with this property. Then with a probability of at least $1-\gamma$ we have:

(42.13) $|\mathtt{core}(F_i,\chi)| \geq \sigma^{(l+\frac{1}{2})\mu}n$ *for* $i=1,2$.

Moreover for all $i=1,2$ if we randomize F_1, F_2 as described above and then

(43.13) *if we pick a random partial input ζ defined on $\mathtt{core}(F_i,\chi)$ so that $\chi \wr \zeta \in G$, with uniform distribution on the set of all such partial inputs, then with a probability of at least $1-\gamma$, for the randomization of ζ we have:* $|\mathtt{register}(F_i,\chi \wr \zeta) \cap \mathtt{core}(F_{3-i},\chi)| \leq \sigma^{(l+\frac{3}{4})\mu}n$

Remarks. 1. The upper bound on $|\mathtt{register}(F_i,\chi \wr \zeta) \cap \mathtt{core}(F_{3-i},\chi)|$ in (43.13) is essentially smaller than the lower bound in (42.13). This will be used to show that the various sets $\mathtt{core}(\chi \wr \zeta)$ are relatively close to each other in the metric defined by the size of the symmetric difference. (Lemma 16 gives the connection between the estimate in (43.13) and distances between the sets $\mathtt{core}(\chi \wr \zeta)$.) This closeness implies that there are relatively few sets of the form $\mathtt{core}(\chi \wr \zeta)$. This will make it possible to define the functions $\mathcal{L}_i$ with the properties described in Lemma 11.

2. In the proof of this lemma (for both statements) we will have two steps. First we will estimate the expected value of the numbers $|\mathtt{core}(F_i,\chi)|$ resp. $|\mathtt{register}(F_i,\chi \wr \zeta) \cap \mathtt{core}(F_{3-i},\chi)|$, then, based on these estimate we get the bounds which hold with high probability. The second step will be easier for the upper bound since in this case Markov's inequality can be used. In the case of the lower bound our solution is more complicated. (See the remark before Lemma 14.) To get the estimates on the expected value we will use the inequalities connecting l,μ,σ and $|\mathtt{acc}_j(\chi)|$ stated in Lemma 12. The definition of the function `core` will make it possible to estimate the probability that a fixed register $x \in \mathtt{acc}_j(\chi)$ is in $\mathtt{core}(F_i,\chi)$, where F_i has the distribution described in the lemma.

Proof. Assume that $\gamma,k,\sigma,n,G,\chi,l,\mu,t$ are fixed with the properties listed in the lemma. We start with the proof of (43.13) and assume that e.g $i=1$. First we estimate the expected value of $|\mathtt{register}(F_1,\chi \wr \zeta) \cap \mathtt{core}(F_2,\chi)|$ provided that we randomize F_2 only. More precisely suppose that $F_1 \subseteq \mathcal{I}$, $|F_1| = t$ and a partial input ζ defined on $\mathtt{core}(F_1,\chi)$ so that $\chi \wr \zeta \in G$ are fixed with these porperties but otherwise in an arbitrary way. We now take a random $F_2 \subseteq \mathcal{I}$, $|F_2| = t$ so that F_1 and F_2 are disjoint with uniform distribution on the set of all such sets F_2. We estimate first the expected value of $|\mathtt{register}(F_1,\chi \wr \zeta) \cap \mathtt{core}(F_2,\chi)|$ with respect to this randomization. Let x be a fixed element of $\mathtt{register}(F_1,\chi \wr \zeta)$, we estimate the probability p_x of the event $x \in \mathtt{core}(F_2,\chi)$. (The expected value in question will be $\sum\{p_x | x \in \mathtt{register}(F_1,\chi \wr \zeta)\}$.)

First assume that $x \in \mathtt{acc}_i(\chi)$ for some $i < l$. In this case we use the trivial bound $p_x \leq 1$. Let Z_1 be the set of all registers x with this porperty.

Assume now that $x \in \mathtt{acc}_i(\chi)$ for some $i \geq l$. Let Z_2 be the set of all registers x with this property. If x is accessed in an interval of F_1, at input χ then $p_x = 0$, since F_1 and F_2 are disjoint and $\mathtt{core}(F_2, \chi)$ cannot contain a register which is accessed outside F_2 at input χ. If x is not accessed in any of the intervals of F_1 then let $X \subseteq \mathcal{I}$ be the set of intervals where x is accessed. We have $X \cap F_1 = \emptyset$. p_x is equal to the probability of $X \subseteq F_2$. We can compute this probability by sequentially deciding about each element of X whether it is in F_2. This gives $p_x = \frac{t}{|\mathcal{I}|-t} \times \frac{t-1}{|\mathcal{I}|-t-1} \times \ldots \times \frac{t-|X|+1}{|\mathcal{I}|-t-|X|+1}$. Since $|X| \geq l$ we have $p_x \geq \frac{t}{|\mathcal{I}|-t} \times \frac{t-1}{|\mathcal{I}|-t-1} \times \ldots \times \frac{t-l+1}{|\mathcal{I}|-t-l+1}$. $t = [\sigma^{-1+\mu}]$ implies that $\sigma^{-1+\mu}$ is an upper bound for all of the nominators. To get a lower bound on the denominators we use the inequalities $\frac{1}{2}\sigma^{-1} \leq |\mathcal{I}|$ and $t \leq \frac{1}{8}\sigma^{-1}$, $l \leq \frac{1}{8}\sigma^{-1}$. The first inequality follows from the defintion of $\mathcal{I}$, the third from the fact that $l \leq 2k$ and σ is sufficiently large with respect to k, and the second inequality is a consequence of the lower bound on μ given in (38.12), since this lower bound implies that $\sigma^{\mu} \leq k^{-|\log \sigma|^{\frac{1}{2}}}$. Using the three inequalities we get the lower bound $\frac{1}{4}\sigma^{-1}$ on the denominators. Therefore we have $p_x \leq \frac{\sigma^{l(-1+\mu)}}{4^{-l}\sigma^{-l}} = 4^l\sigma^{l\mu}$.

We have $E(|\mathtt{register}(F_1, \chi \restriction \zeta) \cap \mathtt{core}(F_2, \chi)|) \leq \sum\{p_x | x \in \mathtt{register}(F_1, \chi \restriction \zeta)\} \leq \sum_{x \in Z_1} p_x + \sum_{x \in Z_2} p_x \leq |Z_1| + |Z_2| 4^l \sigma^{l\mu}$.

By (37.12) we have that $|Z_1| \leq l\sigma^{(l+1)\mu}n$. $|Z_2| \leq |\mathtt{register}(F_1, \zeta)| \leq |\bigcup F_1| \leq t\sigma n \leq \sigma^{-1+\mu}\sigma n = \sigma^{\mu}n$. Therefore $E(|\mathtt{register}(F_1, \zeta) \cap \mathtt{core}(F_2, \chi)|) \leq l\sigma^{(l+1)\mu}n + 4^l\sigma^{l\mu}\sigma^{\mu}n \leq (4^l + l)\sigma^{(l+1)\mu}n$. Since this is true for any fixed F_1 and ζ, we have that if we first ranomize both F_1 and F_2 and then ζ as described in the statement of the lemma then for the randomization of all of the three elements we have $E(|\mathtt{register}(F_1, \zeta) \cap \mathtt{core}(F_2, \chi)|) \leq (4^l + l)\sigma^{(l+1)\mu}n$. The upper bound that we claim on $|\mathtt{register}(F_1, \zeta) \cap \mathtt{core}(F_2, \chi)|$ in (43.13) is $\sigma^{(l+\frac{3}{4})\mu}n$. The ratio of the upper bound and the expected value is $\lambda = \sigma^{-\frac{1}{4}\mu}(4^l + 1)^{-1}$. Therefore by Markov's inequality $|\mathtt{register}(F_i, \chi \restriction \zeta) \cap \mathtt{core}(F_{3-i}, \chi)| \leq \sigma^{(l+\frac{3}{4})\mu}n$ holds with a probability of at least $1 - \lambda^{-1}$, for the randomization of F_1, F_2 and ζ. Therefore using Markov's inequality again it is easy to see that with a probability of at least $1 - \lambda^{-\frac{1}{2}}$ for the randomization of F_1 and F_2 we get such a pair F_1, F_2 so that with a probability of at least $1 - \lambda^{-\frac{1}{2}}$ for the randomization of ζ this inequality holds. (Here we apply Markov's inequality for the random variable, depending on the choice of F_1 and F_2 only, whose value is the probability, for the randomization of ζ, that the inequality in (43.13) holds.) The lower bound on μ in (38.12) and $l \leq 2k$ imply that $\lambda^{-\frac{1}{2}}$ is sufficiently small with respect to γ if σ is sufficiently small with respect to k and γ, which implies (42.13).

For the proof of (42.13) assume e.g. that $i = 1$. We will show that even $|\mathtt{core}(F_1, \chi) \cap \mathtt{acc}_l(\chi)| \geq \sigma^{(l+\frac{1}{2})\mu}$ with a probability close to 1.

Remark. We may give a lower bound on the expected value of $|\mathtt{core}(F_1,\chi)\cap \mathtt{acc}_l(\chi)|$, using similar arguments as in the proof of (42.13), however in the case of lower bounds there is no analogue of Markov's inequality and so we do not get automatically a lower bound on $|\mathtt{core}(F_i,\chi)\cap\mathtt{acc}_l(\chi)|$, which holds with high probability. The following Lemma is a general result which makes possible in certain cases to get a lower bound on random variable, which holds with high probability, by using a lower bound on its expected value. We assume that the random variable is the sum of those values of a function of l variables where each varibale is restricted to the same random subset of a finite universe.

Definition. If A is a set and l is a positive integer, then $[A]^l$ will denote the set of those subsets of A which contain exactly l elements.

Lemma 14. *There is a function g defined on the set of positive integers with positive real values so that for all positive integer l and for all sufficiently small $\iota > 0$ there is a $\delta > 0$ so that if m is a sufficiently large positive integer, $s \geq m^{1-\iota}$ then the following holds. Assume that A is a set with m elements, w is a function on $[A]^l$ with nonnegative real values so that for all $a \in A$ we have*
(44.14) $\sum\{w(X) \mid X \in [A]^l,\ a \in X\} \leq m^{-1+\iota}\sum\{w(X) \mid X \in [A]^l\}$
and B is a random subset of A with uniform distribution on $[A]^s$ and $\Lambda = \sum\{w(X) \mid X \in [B]^l\}$. Then, the probability of the following event is at least $1-2^{-m^\delta}$:

$$\Lambda \geq g(l)E(\Lambda) = g(l)\binom{s}{l}\binom{m}{l}^{-1}\sum\{w(X) \mid X \in [A]^l\}$$

For the proof of Lemma 14 we need the following lemma.

Lemma 15. *For all sufficiently small $\iota > 0$ there is an $\iota' > 0$ so that for all sufficiently large positive integer m the following holds. Assume that D is a finite set with m elements, $r \geq m^{1-\iota}$ is an integer and f is a nonnegative real-valued function on D so that for any $X \subseteq D$ with $|X| \leq m^{2\iota}$ we have $\sum_{x\in X} f(x) \leq \frac{1}{10}\sum_{x\in D} f(x)$*

Suppose further that R is taken at random with uniform distribution from $[D]^r$. Then with a probability greater than $1-2^{-m^{\iota'}}$ we have that

$$\sum_{x\in R} f(x) \geq \frac{1}{2}(\frac{r}{m}\sum_{x\in D} f(x))$$

Proof. We define an odering $\leq_f$ of the set D with the property "for all $d, d' \in D$, $d \leq_f d'$ implies $f(d) \leq_f f(d')$". Let P be a partition of D into intervals (according ot this ordering) so that the lengths of each interval is between $\frac{1}{4}m^{2\iota}$ and $(\frac{1}{4}+\frac{1}{100})m^{2\iota}$. Let I be a fixed interval. The expected value of $|R \cap I|$ is $\frac{r}{m}|I|$. We claim that with a probability of at least $1-2^{-m^{\iota''}}$ we have that $|R \cap I| \geq \frac{7}{8}\frac{r}{m}|I|$, where $\iota'' > 0$ depends only on ι. This can be easily proved e.g. by expressing the probability of $|R \cap I| = k$ by binomial coefficients and then estimating the sum of the corresponding binomial coefficients using Stirling's formula. Since the number of different intervals I is at most m, we get that there is an $\iota' > 0$ so that with a probability $p \geq 1-2^{m^{\iota'}}$ for all intervals $I \in P$ we have $|R \cap I| \geq \frac{3}{4}\frac{r}{m}|I|$. Let $I_1, ..., I_q$ be the intervals of P in the order induced on them by the ordering $\leq_f$ on D. If we take two consecutive intervals I_i, I_{i+1} then all of the values of f on I_{i+1} are greater than all of its values on I_i. Let $S_i(R) = \sum_{d \in I_i \cap R} f(d)$.

These facts imply that with a probability p for all $i = 2, ..., q$ we have that $S_{i+1}(R) \geq \frac{3}{4}E(S_i(R))$. This implies that $\sum_{i=1}^{q} S_i(R) \geq \frac{3}{4}\sum_{i=1}^{q-1} E(S_i(R)) = \frac{3}{4}E(\sum_{i=1}^{q} S_i(R)) - E(S_q(R)) = \frac{3}{4}(\frac{r}{m}\sum_{x \in D} f(x)) - E(S_q(R))$. We give an upper bound on $E(S_q(R))$. $|I_q| \leq m^{2\iota}$, therefore by our assumption $\sum_{x \in I_q} f(x) \leq \frac{1}{10}\sum_{x \in D} f(x)$ and so $E(S_q(R)) \leq \frac{1}{10}\frac{r}{m}\sum_{x \in D} f(x)$. Consequently $\sum_{x \in R} f(x) = \sum_{i=1}^{q} S_i(R) \geq \frac{3}{4}\frac{r}{m}\sum_{x \in D} f(x) - \frac{3}{4}\frac{1}{10}\frac{r}{m}\sum_{x \in D} f(x) \geq \frac{1}{2}\frac{r}{m}\sum_{x \in D} f(x)$. *Q.E.D.*(Lemma 15)

Proof of Lemma 14. We will use the following notation. If $x_1, ..., x_l$ is an arbitrary sequence of length l from the elements of A then $w(x_1, ..., x_l) = w(\{x_1, ..., x_l\})$ if $|\{x_1, ..., x_l\}| = l$, otherwise $w(x_1, ..., x_l) = 0$. Let $r = [\frac{s}{l}]$. We randomize B in the following way first we pick a random sequence $B_1, ..., B_l$ so that $B_i \in [A]^r$ for $i = 1, ..., l$ with uniform distribution on the set of all sequences with this property. The definition of r implies that $\nu = |\bigcup_{i=1}^{r} B_i| \leq s$. We pick a random subset B' of $A - \bigcup_{i=1}^{l} B_i$ with ν elements with uniform distribution on the set of all sets with this property. Let $B = B' \cup \bigcup_{i=1}^{l} B_i$. Clearly B has uniform distribution on $[A]^s$. (Note that the sets B_i are not necessarily disjoint.)

We have $\sum\{w(X)|X \in [B]^l\} \geq (l!)^{-1}\sum\{w(x_1, ..., x_l)|x_i \in B_i, i = 1, ..., l\}$, since each term on the right (apart from the order of the varaiables x_i) on the left-hand-side too. We will give a lower bound on the right-hand-side which holds with high probability for a random B. We may compute the expected value E_1 of $\sum\{w(x_1, ..., x_l)|x_i \in B_i, i = 1, ..., l\}$ by adding the probabilities $p(a_1, ..., a_l)$ of the events that the terms $w(a_1, ..., a_l)$ occur in the sum, for all $a_1, ..., a_l \in A$, $\{a_1, ..., a_l\} \in [A]^l$. Let $a_1, ..., a_l$, $\{a_1, ..., a_l\} \in [A]^l$ be fixed. Since the events $a_i \in B_i$ are independent we have $p(a_1, ..., a_l) = (\frac{r}{m})^l$. Therefore $E_1 = l!(\frac{r}{m})^l\sum\{w(X)|X \in [A]^l\}$. This makes it

possible to replace $\sum\{w(X)|X \in [A]^l\}$ by $(l!)^{-1}(\frac{r}{m})^{-l}E_1$ in the conclusion of Lemma 14. We get that it is enough to prove the lemma if the conclusion is

$$\Lambda \geq g(l)(l!)^{-1}(\frac{r}{m})^{-l}\binom{s}{l}\binom{m}{l}^{-1}E_1$$

We show that the coefficent of E_1 in this expression remains below a bound depending only on l. If l is fixed and m, s tends to infinity with $s \geq m^{1-\iota}$ then $\lim(\frac{s}{m})^{-l}\binom{s}{l}\binom{m}{l}^{-1} = 1$. We also have that $(\frac{r}{m})^{-1} = (\frac{s}{m})^{-1}\frac{s}{r} = (\frac{s}{m})^{-1}s[\frac{s}{l}]^{-1} \leq (\frac{s}{m})^{-1}s(\frac{s}{2l})^{-1} = (\frac{s}{m})^{-1}2l$. Using these facts we get that if m is sufficiently large with respect to l, then $(l!)^{-1}(\frac{r}{m})^{-l}\binom{s}{l}\binom{m}{l}^{-1} \leq (l!)^{-1}2^l l^l(\frac{s}{m})^{-l}\binom{s}{l}\binom{m}{l}^{-1} \leq (l!)^{-1}2^l l^l\frac{1}{2}$. Therefore it is enough to show that the lemma is true if we replace the conclusion by $\Lambda \geq g(l)(l!)^{-1}2^l l^l\frac{1}{2}E_1 = \tilde{g}(l)E_1$ for a function $\tilde{g} > 0$. Since $\Lambda \geq (l!)^{-1}\sum\{w(x_1, ..., x_l)|x_i \in B_i, i = 1, ..., l\}$ it is sufficient to prove the lemma if the conclusion is $\sum\{w(x_1, ..., x_l)|x_i \in B_i, i = 1, ..., l\}\bar{g}(l)E_1$ for a function $\bar{g}$.

We define a random variable h_j whose value depends only on the choice of $B_1, ..., B_j$. Namely, let

$$h_j = \sum\{w(x_1, ..., x_l)|x_i \in B_i, i = 1, ..., j, x_i \in A, i = j+1, ..., l\}$$

We will prove by induction on j that with a probability of at least $1 - j2^{-m^{\iota'}}$ we have that $h_j > \frac{1}{2}E(h_j)$, where $\iota' > 0$ depends only on ι. Our inductive statement for $j = l$ implies the required inequality.

Assume now that the statement holds for $i = 1, ..., j-1$ for some $j = 1, ...l$. We may write h_j in the form of $\sum_{a \in B_j} f(a)$, where

$$f(a) = \sum\{w(x_1, ..., x_{j-1}, a, x_{j+1}, ..., x_l)|x_i \in B_i, i = 1, ..., j-1, x_i \in A, i = j+1, ..., l\}$$

Assume now that $B_1, ..., B_{j-1}$ has been already randomized. We want to apply Lemma 15 to the function f with $D \to A$, $r \to [\frac{s}{l}]$. Suppose that $X \subseteq A$, $|X| \leq m^{2\iota}$. If $\sum_{x \in X} f(x) \leq \frac{1}{10}\sum_{x \in A} f(x)$ does not hold then there is an $x_0 \in A$ so that $f(x_0) \geq \frac{1}{10}m^{-2\iota}\sum_{a \in A} f(a)$. By the definition of $f(a)$ we have $\sum_{\in A} f(a) = h_{j-1}$, therefore according to the inductive assumption with a probability of at least $1 - (j-1)2^{-m^{\iota'}}$ (for the randomization of $B_1, ..., B_{j-1}$) we have that $h_{j-1} \geq g_j(l)E(h_{j-1}) \geq (\frac{1}{2l})^l g_j(l)\sum\{w(Y)|Y \in [A]^l\}$. This implies that for such a sequence $B_0, ..., B_{j-1}$ we would get $f(x_0) > \bar{g}_j(l)m^{-2\iota}\sum\{w(Y)|Y \in [A]^l\}$ in contradiction to (44.14) if m is sufficiently large. Therefore with a probability of at least $1 - (j-1)2^{-m^{\iota'}}$ for the randomization of $B_1, ..., B_{j-1}$ Lemma 15 can be applied for the function f and we

get that for the randomization of B_j with a probability of at least $1-2^{-m^{\iota'}}$ we have $\sum_{x\in B_j} f(x) \geq \frac{1}{2}\frac{r}{m}\sum_{x\in A} f(x) \geq \frac{1}{2}E(\sum_{x\in B_j} f(x)) \geq \frac{1}{2}E(h_j)$ which implies the inductive statement for j. *Q.E.D.*(Lemma 15)

Now we may conclude the proof of Lemma 13. We want to apply Lemma 14. Let l be the integer from the proof of Lemma 13 and let $\delta = \gamma$, assume that $\iota > 0$ is the number whose existence is stated in Lemma 14. We apply the Lemma with $A \to \mathcal{I}$. If $X \subseteq \mathcal{I}$, $|X| = l$, then $w(X)$ will be the number of registers from $\mathtt{acc}_l(\mathcal{I})$ which are accessed from each elements of X. Let $m = |\mathcal{I}| \leq 2k\sigma^{-1}$, $m \geq k\sigma^{-1}$ and $s = t = [\sigma^{-1+\mu}]$. We have to show that $s \geq m^{1-\iota}$. The upper bound in (38.12) and the fact that ρ is sufficiently small with respect to k, γ implies that μ is sufficiently small with respect to k,l, γ, and ι. So we have $m^{1-\iota} \leq (2k\sigma^{-1})^{1-\iota} \leq [\sigma^{-1+\mu}] = s$. Now we check (44.14) the second requirement of Lemma 14. On the left-hand-side we have the number of registers which are accessed from the interval a (and also $l-1$ other elements of $\mathcal{I}$.) Since a, as an element of $\mathcal{I}$, contains at most σn elements, we have that the left-hand-side is at most σn. By (36.12) and $|\mathcal{I}| \geq \sigma^{-1}$ the right-hand-side is at least $|\mathcal{I}|^{-1+\iota}\sigma^{\frac{1}{4}\mu}n \geq \frac{1}{2}\sigma^{1-\iota+\frac{1}{4}\mu}n$. As we have seen earlier, the upper bound in (38.12) implies that μ is sufficiently small with respect to ι so we have that the right-hand-side is greater than σn.

The random set B is F_1, and so, according to the conclusion of the lemma we have that with a probability of at least $1-2^{-m^\gamma} > 1-\gamma$ we have $|\mathtt{core}(F_1,\chi)\cap\mathtt{acc}_l(\chi)| \geq g(l)\binom{s}{l}\binom{m}{l}^{-1}\sigma^{\frac{1}{4}\mu}n$. Using that both s and m are sufficiently large with respect to l we get $\binom{s}{l}\binom{m}{l}^{-1} = \frac{s(s-1)\ldots(s-l+1)}{m(m-1)\ldots(m-l+1)} \geq (\frac{1}{2})^l(\frac{s}{m})^l \geq (\frac{1}{2})^l(\frac{1}{2}\sigma^{-1+\mu})^l(2\sigma^{-1}k)^{-l} \geq (\frac{1}{4k})^l\sigma^{l\mu}$. Since σ is sufficiently small with repsect to l, this implies (42.13). *Q.E.D.*(Lemma 13)

Notation. If A and B are sets then $A\Delta B$ will denote their symmetric difference.

Lemma 16. *Assume that χ,η are inputs F_1, F_2 are disjoint subsets of $\mathcal{I}$, $\mathtt{stem}(F_1,\chi) = \mathtt{stem}(F_1,\eta)$ and $\mathtt{rstate}_{\chi,\bigcup F_1} = \mathtt{rstate}_{\eta,\bigcup F_1}$. Then*

(45.16) $\mathtt{core}(F_2,\chi) - \mathtt{core}(F_2,\eta) \subseteq \mathtt{register}(F_1,\eta)\cap\mathtt{core}(F_2,\chi)$.

(46.16) $|\mathtt{core}(F_2,\eta)| = |\mathtt{core}(F_2,\chi)|$ *implies*
$|\mathtt{core}(F_2,\eta)\Delta\mathtt{core}(F_2,\chi)| \leq 2|\mathtt{register}(F_1,\eta)\cap\mathtt{core}(F_2,\chi)|$.

Remark. The condition $|\mathtt{core}(F_2,\eta)| = |\mathtt{core}(F_2,\chi)|$ in (46.16) does not restrict essentially the applicability of this lemma, since the number of possible values for $|\mathtt{core}(F_2,\eta)|$ is at most n. Therefore from a set of inputs we may always take a subset with density at least $\frac{1}{n}$ so that $|\mathtt{core}(F_2,\xi)|$ is the same for any input ξ in the subset.

Proof. $\mathtt{stem}(F_1,\chi) = \mathtt{stem}(F_1,\eta)$ and $\mathtt{rstate}_{\chi,\bigcup F_1} = \mathtt{rstate}_{\eta,\bigcup F_1}$ implies that the computations at inputs η and χ are identical outside F_1, that is the states of the machine at the two inputs are the same at time t for each $t \in [0,kn] - F_1$. Assume now that $x \in \mathtt{core}(F_2,\chi) - \mathtt{core}(F_2,\eta)$. Register x is accessed in F_2 at input χ. Since $F_1 \cap F_2 = \emptyset$ our previous remark implies that that it is also accessed in F_2 at input η. Therefore by $x \notin \mathtt{core}(F_2,\eta)$, x must be accessed at input η at some time outside F_2. This cannot happen outside F_1 since there the two computations are identical and $x \in \mathtt{core}(F_2,\chi)$ implies that x is not accessed outside F_2 at input χ. Therefore x must be accessed in F_1 at input η, that is, $x \in \mathtt{register}(F_1,\eta)$ which completes the proof of (45.16).

$|\mathtt{core}(F_2,\eta)| = |\mathtt{core}(F_2,\chi)|$ implies that their symmetric difference consists of two disjoint subsets with identical cardinalities. We gave an upper bound on one of them in (45.16). *Q.E.D.*(Lemma 16)

Proof of Lemma 11. Suppose that a positive integer k is fixed. Let $\tau = 1 - \frac{1}{40k}$. Suppose that $\sigma > 0$ is sufficiently small with respect to k, $\epsilon > 0$ is sufficiently small respect to k,σ, n is sufficiently large with respect to k,σ,ϵ and G is a visible set of input. For each fixed χ in G Lemma 12 guarantees the existence of a pair of numbers l,μ with the properties listed in the lemma. l is an integer in the interval $[1,2k]$, μ^{-1} is a positive integer and by (38.12) we have $\mu^{-1} \leq |\log\sigma|^{\frac{1}{2}}$. Therefore there are at most $2k|\log\sigma|^{\frac{1}{2}} \leq \sigma^{-1}$ possibilities for the choice of this pair. Consequently there is a subset D_1 of G with at most $\sigma^{-1}|G|$ elements so that for each $\chi \in D_1$ the pair l,μ is the same. In the following l and μ will denote these common values for all $\chi \in D_1$.

We define a partition P' of D_1: $\chi,\xi \in D_1$ are in the same class iff $|\mathtt{core}(F_i,\chi)| = |\mathtt{core}(F_i,\xi)|$ for $i = 1,2$. This partition has at most n^2 classes therefore it has a class D so that $|D| \geq n^{-2}|D_1| \geq n^2\sigma^{-1}|G|$. We will use later that

(47) for all $\chi,\xi \in D$ and $i = 1,2$ we have $|\mathtt{core}(F_i,\chi)| = |\mathtt{core}(F_i,\xi)|$.

Let $\kappa = \sigma^{(l+\frac{5}{8})\mu}$ We pick the sets F_1 and F_2 the same way as in Lemma 13, that is, $t = [\sigma^{-1+\mu}]$ and F_1,F_2 is a random pair of disjoint subsets of $\mathcal{I}$ each with t elements taken with uniform distribution from the set of all pairs with this property. According to Lemma 13 (with $G \to D$), for any fixed $\chi \in D$ if we pick F_1 and F_2 at random then with a probability of at least $1-\gamma$ both (42.13) and (43.13) hold. We will use a consequence of this fact in case when we randomize χ as well. We pick χ at random with uniform distribution from D and independently F_1,F_2 with the distribution described above. For this randomization we have that the resulting elements χ,F_1,F_2 satisfy condition (42.13) and (43.13) with a probability of at least $1-\gamma$. Therefore we may pick a fixed value for F_1 and F_2 so that if we randomize only χ then (42.13) and (43.13) hold with a probability of at least $1-\gamma$. Let F_1, F_2 be these fixed values and let D_0 be the set of all χ satisfying (42.13) and (43.13). Clearly $|D_0| \geq (1-\gamma)|D|$.

Now we define the functions $\mathcal{L}_i$, $i = 1, 2$. Let x be a partial input. Assume first that there is a $\chi \in D_0$ so that $\mathtt{stem}(F_i, \chi) = x$. For each $y \subseteq \mathtt{domain}(x)$ let $g(y)$ be the number of $\eta \in \mathtt{fan}(D, F_i, \chi)$ so that the symmetric difference of $\mathtt{core}(F_{3-i}, \eta)$ and $\mathtt{core}(F_{3-i}, \chi)$ is at most $4\sigma^{(l+\frac{3}{4})\mu}n$. We define now an ordering "$\leq_x$" on all of the subsets of $\mathtt{domain}(x)$ so that $z \leq_x y$ implies $g(z) \geq g(y)$ (apart from this property the ordering can be arbitrary). For each $y \subseteq \mathtt{domain}(x)$, $\mathcal{L}_i(x, y)$ will be the rank of y according to the ordering $\leq_x$, (the rank of an element y is the number of elements which are not greater than y). If there is no $\chi \in D_0$ so that $\mathtt{stem}(F_i, \chi) = x$ then $\mathcal{L}_i(x, y)$ as a function of y will be an arbitrary one-to-one map of the set of subsets of $\mathtt{domain}(x_0)$ onto a set of positive integers.

We show now that F_1, F_2, $\mathcal{L}_1, \mathcal{L}_2$, D, D_0 and κ meet the requirements of Lemma 11.

(30.11). This is a consequence of the defintion of κ, the lower bound on μ in (38.12) and the fact that σ is sufficiently small with respect to τ,

(31.11). We have seen that $|D| \geq n^{-2}\sigma^{-1}|G|$ and $|D_0| \geq |(1-\gamma)|D||$. Lemma 13 allows us to choose $\gamma > 0$ as an arbitrarily small constant.

(32.11). F_1 and F_2 are disjoint by their definition.

(33.11). The definition of $\mathcal{L}_i$ had two cases according to x. In the first case, for a fixed x, the value of $\mathcal{L}_i(x, y)$ is the rank of y in an ordered set, so $\mathcal{L}_1(x, y)$ as a function of y is indeed a one-to-one map, and the values are clearly positive integers. In the second case we defined $\mathcal{L}_1$ with no other purposes than to meet these requirements.

(34.11). Assume that $\chi \in D_0$. By the definition of D_0 both (42.13) and (43.13) are satisfied. Assume that i is fixed and $j = 3-i$. Let $R = \mathtt{fan}(D, F_i, \chi)$, $x = \mathtt{stem}(F_i, \chi)$. For any $\theta > 0$ let Ψ_θ be the set of all subsets y of $\mathtt{domain}(x)$ so that $|\mathtt{core}(F_j, \chi)\Delta y| \leq \theta\sigma^{(l+\frac{3}{4})\mu}n$. R_θ will denote the set of all $\eta \in R$ so that $\mathtt{core}(F_j, \eta) \in \Psi_\theta$.

(43.13) implies that for at least $(1-\gamma)|R|$ elements $\eta \in R$ we have $|\mathtt{register}(F_i, \eta) \cap \mathtt{core}(F_j, \chi)| \leq \sigma^{(l+\frac{3}{4})\mu}n$. Therefore by (47) and Lemma 16 we have $|\mathtt{core}(F_j, \chi)\Delta\mathtt{core}(F_j, \eta)| \leq 2|\mathtt{register}(F_i, \eta) \cap \mathtt{core}(F_j, \chi)| \leq 2\sigma^{(l+\frac{3}{4})\mu}n$. We got that $|R_2| \geq (1-\gamma)|R|$.

This implies that if g is the function defined in the defitinion of $\mathcal{L}_i$ then for all $y \in \Psi_2$ we have that $g(y)) \geq (1-\gamma)|R|$. On the other hand if $y \notin \Psi_8$ then $g(y) \leq \gamma|R|$. Therefore in the ordering $\leq_x$ all of the elements of Ψ_2 are smaller than all of the elements outside Ψ_8. Therefore the rank of all of the elemnts of Ψ_2 is at most $|\Psi_8|$. This implies that for all $\eta \in R_1$ we have $\mathcal{L}_i(x, \mathtt{core}(F_j, \eta)) \leq |\Psi_8|$. Every elements of $|\Psi_8|$ is a subset of $\{1, ..., n\}$ with at most $8\sigma^{(l+\frac{3}{4})\mu}n$ elements, therefore $|\Psi_8| \leq \binom{n}{8\sigma^{(l+\frac{3}{4})\mu}n}$. To estimate the binomial coefficient we use the follwoing well-known fact that can be proved e.g. by using Stirling's formula: there is a $c' > 0$ so that for all $0 < \rho < 1$ if n is sufficiently large then $\binom{n}{\rho n} \leq e^{c'\rho n|\log\rho|}$. The inequality is applicable in our case since Lemma 12 guarantees that μ has positive upper and lower bounds independent of n.

We get the following: $|\Psi_8| \leq e^{c'8\sigma^{(l+\frac{3}{4})\mu}|\log(8\sigma^{(l+\frac{3}{4})\mu})|n} \leq 2^{\sigma^{(l+\frac{5}{8})\mu}n} = 2^{\kappa n}$ provided that n is sufficently large with respect to σ. We have $\mathcal{L}_i(\mathtt{stem}(F_i,\chi),\mathtt{core}(F_j,\eta)) \leq 2^{\kappa n}$ for all $\eta \in R_1$. Since $R_1 \subseteq R$, $|R_1| \geq (1-\gamma)|R|$, $\gamma < \frac{1}{8}$ and $R = \mathtt{fan}(D,F_i,\chi)$ this implies (34.11).

(35.11). According to (42.13) we have $|\mathtt{core}(F_i,\chi)| \geq \sigma^{(l+\frac{1}{2})\mu}n$. Since $\tau = 1-\frac{1}{40k}$, $l \leq 2k$, $\kappa = \sigma^{(l+\frac{5}{8})\mu}$ and $\sigma < 1$ we have that $\kappa^\tau < \sigma^{(l+\frac{1}{2})\mu}$ that is $|\mathtt{core}(F_i,\chi)| \geq \kappa^\tau n$. *Q.E.D.*(Lemma 11).

A probabilistic algorithm for the element distinctness problem. In this section we give an upper bound on the time necessary for the solutionof the element distinctness problem. Our computational model now is a random access machine in the usual much narrower sense of the word than the one that we have used for our lower bound results. That is, we assume that the read and write memory of the machine is also consists of registers of the same sizes as the input registers. Now the machine cannot change its state in an arbitrary way it can only perform arithmetic and logical operations on the contents of to registers and can only access a register if its address is the content of a distinguished register. An exact defintion of this RAM is given e.g. in [AHU]. According to the defintion given there we assume that the content of each register is a nonnegtive integer. Our additional assumption will be an upper bound on this integer and an upper bound on the total number of registers. It is important for our algorithm that not only the arithmetic operations addition and multiplication can be proformed between the contents of registers (assuming that the result of the operation is not greater then the maximal number allowed in a register), but also the operation $[\frac{x}{y}]$, provided that $y \neq 0$. (This makes it possible to store and recover efficiently a sequence of positive integers where some of them are much smaller than the allowed maximal size. In this case we may have to store several integers in a single register to minimize the size of the needed memory.) We also assume that the machine has a program which is stored in some registers. (From our point of view it is irrelevant whether the contents of these registers can be changed or not).

If we allow as an additional operation for the machine to ask for a random bit which appears as the content of a distinguished register then we will call the machine a probabilistic true random acess machine. Each access for a random bit will be counted as a time unit. We assume that the random bits provided to the machine during a computation are generated by randomizing of a sequence of mututally independent random variables with $0,1$-values.

Theorem 5. *For all $\delta > 0, \theta > 0, c \geq 1$ there is a $k > 0$ so that if n is sufficiently large, then there is a probabilistic true random access machine with a program contained in a constant number of registers and with n read only input registers and with at most δn registers of read and write memory so that each of the input registers*

and each of the registers of the read and write memory contain $[c \log n]$ *bits and the following holds. For any input* χ *the machine gives a* $0,1$ *output* $\mathtt{out}(\chi)$ *in time* kn *so that with a probability of at least* $1-\theta$ *we have*

$\mathtt{out}(\chi) = 1$ *iff there exist* $1 \leq i < j \leq n$ *so that* $\chi(i) = \chi(j)$

Proof. According to the statement of the theorem δn is the number of registers in the R/W memory where δ can be an arbitrarily small constant. However it is sufficient to prove the theroem for the case when $\delta > c_1$, where c_1 is a sufiiciently large absolute constant. Indeed assume that this modified version of the theroem is true and c_1 is fixed. We want to prove the original version. We cut the interval $[1,n]$ into disjoint subintervals $I_1, ..., I_t$ whose lengths is about $\frac{1}{4}\delta c_1^{-1} n$. For each pair of intervals I_i, I_j using the algorithm provided by the modified theorem we may check in time $k'n$ whether there are $x, y \in I_j \cup I_j$, $x \neq y$ so that $\chi(x) = \chi(y)$. The theorem is applicable because the number of working registers δn is now larger than c_1 times the number of input registers, since the input is contained in only $|I_i| + |I_j| \leq \delta c^{-1} n$ registers. We may also assume that we get the correct answer with a probability of at least $1 - \frac{\theta}{t^2}$. If we do this for all of the possible pairs I_i, I_j then we will get the answer in time $t^2 k'$ with a probability of at least $1-\theta$. Since t remains below a bound depending only on c_1 and δ, this proves the original version of the theorem. In the remaining part of this section we give the proof of the modified version, so the word theorem will refer to the modified form.

Definitions. 1. Assume that t is a positive integer and $A \subseteq B$ are finite sets, and h is a function defined on B with values in $\{1, .., t\}$. We say that h is a (t, A) dispersed hash-function on B if the number of elements $a \in A$ with $|h^{-1}(h(a)) \cap A| = 1$ is at least $\frac{1}{2}|A|$.

2. Assume that s, n, are positive integers p is a prime and $d_0, ..., d_{s-1}$ are integers in the interval $[0, p)$. We define a function $h_{s,n,p,d_0,...,d_{s-1}}$ whose domain is the set $\{0, 1, ..., n^s - 1\}$ in the following way. Suppose that $x \in \{0, 1, ..., n^s - 1\}$. We may write x uniquely in the form of $x = \sum_{i=0}^{s-1} b_i n^i$ where b_i is an integer and $0 \leq b_i < n$ for $i = 0, 1, ..., s-1$. Let $h_{s,n,p,d_0,...,d_{s-1}}(x)$ be the least positive residue of $\sum_{i=0}^{s-1} d_i b_i$ modulo p.

Lemma 17. *There is a* $c_2 > 1$ *so that for all positive integer* c *if* n *is sufficiently large* $B = \{0, 1, ..., n^c - 1\}$, $A \subseteq B$, $|A| \leq n$, p *is a prime between* $c_2 n$ *and* $2c_2 n$ *and* $d_0, ..., d_{c-1}$ *is a random sequence of integers with* $0 \leq d_i < n$ *taken with uniform distribution from the set of all sequences with these properties, then with a probability of at least* $\frac{1}{2}$ *we have that* $h_{c,n,p,d_0,...,d_{s-1}}$ *is a* $(2c_2 n, A)$ *dispersed hash function on* B.

Proof. Let T be the number of pairs $\langle a_1, a_2\rangle \in A \times A$ so that $a_1 \neq a_2$ and $h(a_1) = h(a_2)$ where $h = h_{c,n,p,d_0,\ldots,d_{c-1}}$. We estimate $E(T)$. Let $a_1, a_2 \in A$, $a_1 \neq a_2$ be fixed and assume that $a_j = \sum_{i=0}^{c-1} b_{i,j} n^i$, where $0 \leq b_{i,j} < n$. Since the sequences $\langle b_{i,1} | i = 0, \ldots, c-1\rangle$ and $\langle b_{i,2} | i = 0, \ldots, c-1\rangle$ are different we have that the distribution of $(\sum_{i=0}^{c} d_i b_{i,1}) - (\sum_{i=0}^{c} d_i b_{i,2}) = \sum_{i=0}^{c} d_i (b_{i,1} - b_{i,2})$ is uniform modulo p. Therefore $P(h(a_1) = h(a_2)) = \frac{1}{p} \leq \frac{1}{c_2 n}$. This implies that $E(T) \leq \frac{1}{c_2 n}\binom{|A|}{2} \leq \frac{|A|}{c_2}$ since $|A| \leq n$.

By Markov's inequality we have that $P(T > \frac{1}{4}|A|) < \frac{4}{c_2}$ and so if c_2 is sufficiently large, then $P(T > \frac{1}{4}|A|) \geq \frac{1}{2}$. We claim that $T \leq \frac{1}{4}|A|$ implies that h is a $(2c_2 n, A)$ dispersed hash function on B. Indeed in this case there are at most $\frac{1}{4}|A|$ pairs (a_1, a_2) with $h(a_1) = h(a_2)$. These pairs can cover at most $\frac{1}{2}|A|$ elements of A. Therefore for the remaining $\frac{1}{2}|A|$ elements a of A we have $|h^{-1}(h(a)) \cap A| = 1$. *Q.E.D.*(Lemma 17)

Now we continue the proof of the theorem. First we describe the algorithm in the more general random access machine model that we have used for the lower bound proofs. Then we show thst it can be implemented on a true random access machine as well.

The algorithm will have two phases. The time requirement for each phase is $\frac{k}{2}n$.

Phase I. We will construct a sequence of sets $U_0 = \{1, \ldots, n\} \supseteq U_1 \supseteq \ldots \supseteq U_i \supseteq \ldots$. Each susbset of n can be represented by n bits. We will always keep the set U_i that we have constructed the last time in the working memory, and discard all U_j, $j = 0, \ldots, i-1$.

We describe the construction of the sequence U_i by recursion on i. $U_0 = \{1, \ldots, n\}$. Assume that U_{i-1} has been already constructed, for some $i \geq 1$ and it is in the working memory. We will apply Lemma 17 later, when we prove the correctness of the algorithm, with $A \to \{\eta(x) | x \in U_{i-1}\}$. Let h be the hash function from the lemma taken at random as described there. We randomize the bits of the numbers $d_0, \ldots, d_{c-1}$. (The necessary time is $O(c(1 + \log n))$.) The possible values of h are the integers $1, \ldots, 2c_2 n$. We reserve 2 bits for each of these integers in our working memory. Then we go along all of the input registers in U_{i-1} and for register x we compute $h(\eta(x))$ (this takes a constant number of steps for each fixed x). Assume $a = h(\eta(x))$. Using the two bits reserved for the number a we "count" how many times the value a has been attained as $h(\eta(y))$ for some register y which has been already inspected. "Counting" however means now that we want to distinguish only the possibilities "0", "1", "more than 1". Clearly we can do this with the two bits.

After we went along all of the registers in U_{i-1} we will know what are those integers $i \in \{1, \ldots, 2c_2 n\}$ which were taken as a value $h(\eta(x))$ exactly once. If we go along the registers again and compute $h(\eta(x))$ again for all $x \in U_{i-1}$ we will know also the set V_i of those registers x where these values are taken, that is, the set of all registers x so that $h(\eta(x)) \neq h(\eta(y))$ for any $y \neq x$, $y \in U_{i-1}$. (This set can be represented by n bits and as we go along the registers we get the individual bits.) We know that if

there are two identical contents in the registers belonging to the set U_{i-1} they are not contained in the registers of V_i. U_i will be the complement of V_i in U_{i-1}.

The described steps, including "going along" the elements of a set X of registers do not cause any problem in the more general model since the set is uniquely determined by the state of the working memory so the machine can access the registers in $X \subseteq \{1, ..., n\}$ say according to the linear oredring of the natural numbers. However in the case of the true random access model we have to prove that "going along" of the elements of X in the claimed amount of time is possible if the set X is given in a suitable representation which can be constructed in time linear in $|X|$. We will return to this question later.

We continue the construction of the sets U_i until either

(a) we get a U_i with $|U_i| < \frac{n}{\log n}$ or

(b) we reach time $\frac{k}{2}n$. In this case U_r will denote the last set U_i which was completely constructed.

Phase II. If Phase I has been terminated through case (a) then we read the contents of all of the registers in U_i and because of the size of U_i they will fit in the working memory so we know whether there are two identical numbers among these contents. If there are such two contents then the output of our algorithm is 1 otherwise it is 0. (If we work with a true RAM then we may use bucket sorting to decide whether there are two identical contents among the registers belonging to U after we have copied the contents of these registers in the workspace.)

If Phase I was terminated through (b), then we will repeatedly do the following until we run out of time:

First we take a random element x of U_r (by randomizing an integer s between 1 and $|U_r|$ with uniform distribution and then take the s-th element of U_r). Then by going along the elements of U_r we check whether there is an $y \in U_r$, $y \neq x$ so that $h(\eta(x)) = h(\eta(y))$. If we found such a y then the output of our algorithm is 1.

If the algorithm haven't decided that the output is 1 (by repeating the proceedure decribed above) before we get to time kn then the output is 0.

Proof of the correctness of the algorithm. We prove by induction on i that

(48) *for all* $i \geq 1$, *if* $x \in U_{i-1} - U_i$, *then* $\eta(x) \neq \eta(y)$ *for any* $y \neq x$, $y \in \{1, ..., n\}$.

Assume $i \geq 1$, $x \in U_{i-1} - U_i$ and that the inductive assumption holds for $i-1$. By the inductive hypothesis $\eta(x) \neq \eta(y)$ for any $y \in U_j$ where $j < i-1$, and equivalently the same holds for any $y \in \{1, ..., n\} - U_{i-1}$. For all $y \in U_{i-1}$, $x \neq y$ we have $h(\eta(x)) \neq h(\eta(y))$ and so again $\eta(x) \neq \eta(y)$. (This is true even for $i = 1$.) Therefore if Phase I terminates through case (a) then if there are identical elements they must be contained in U_i and so our algorithm finds them.

To show that the algorithm gives the right answer with high probability even if it terminates through case (b) we will prove that:

(49) *if Phase I terminates through case (b) then with a probability of at least* $1-\frac{\theta}{2}$ *there is a* $D \subseteq U_r$, $|D| \geq \frac{1}{10}|U_r|$ *so that*

(*) for all $x \in D$ *there is a* $y \in D$, $y \neq x$ *so that* $\eta(x) = \eta(y)$.

If there is a set D, $|D| \geq \frac{1}{10}|U_r|$ with property (*) then Phase II will find identical elements in the input in $c_4|U_r|$ steps with a probability of at least $1-\frac{\theta}{2}$ if k is sufficiently large with repsect to θ but it does not depend on n. More precisely after i repetition of the cycle of Phase II, the probability that identical elements has not been found yet will be at most $(1-\frac{1}{10})^i$. Therefore (49) implies that the probability that the algorithm does not give the right answer is less than $\frac{\theta}{2}+\frac{\theta}{2}=\theta$.

Proof of (49). If there is a set U_j and a subset $D \subseteq U_j$, $|D| \geq \frac{1}{10}|U_j|$ with property (*) then the analogue of this assertion holds for all of the sets $U_{j+1}, U_{j+2}, ...$ with the the same set D. This is a consequence of $U_i \supseteq U_{i+1} \supseteq U_{i+2} \supseteq ....$ Therefore if Phase I terminates through case (b) but U_r does not contain a set D, with property (*) then

(50) *Phase I is terminated through case (b) and none of the sets* U_j, $j = 0, ..., r$ *contain a* $D \subseteq U_j$, $|D| \geq \frac{1}{10}|U_j|$ *with property (*).*

We will complete the proof by showing that the probability of (50) is smaller than $\frac{\theta}{2}$. Let H_i be the event $|U_i| < \frac{2}{3}|U_{i-1}|$. (50) implies that for $i = 1, ..., r$

(51) $P(H_i|X) \geq \frac{1}{2}$, *for any event* X *in the Boolean algebra generated by* $H_1, ..., H_{i-1}$.

Indeed if we apply the Lemma 17 with $A \to \{\eta(x)|x \in U_{i-1}\} = Z$ as we promised at the description of the algorithm, then with a probability of at least $\frac{1}{2}$, h is a $(2c_2n, Z)$ dispersed hash function. We show that if a h is picked which is $(2c_2n, Z)$ dispersed then $|U_i| < \frac{2}{3}|U_{i-1}|$.

(50) implies that $|Z| > \frac{9}{10}|U_{i-1}|$, therefore by the definition of the $(2c_2, n)$ dispersion we have that there is a $Z' \subseteq Z$, $|Z'| \geq \frac{1}{2}|Z| \geq \frac{9}{20}|U_{i-1}|$ so that $h^{-1}(h(z)) \cap Z = \{z\}$ for all $z \in Z'$. Let $B = \eta^{-1}(Z)$. (50) implies that B may contain at most $\frac{1}{10}|U_{i-1}|$ registers which are in U_i. $B \subseteq U_{i-1}$, $|B| \geq |Z| \geq \frac{9}{20}|U_{i-1}|$ therefore $|U_{i-1} - U_i| \geq \frac{9}{20}|U_{i-1}| - \frac{1}{10}|U_{i-1}| > \frac{1}{3}|U_{i-1}|$ which completes the proof of (51).

Let g be an integer sufficiently large with respect to θ so that k is sufficiently large with respect to g and let G_j be the following event:

There are more than $g(j+1)$ elements i of the set $\{1, ..., r\}$ so that $(\frac{2}{3})^j n \geq |U_i| > (\frac{2}{3})^{j+1}n$. Clearly (51) implies that $P(G_j) \leq 2^{-g(j+1)}$. Therefore the probability of $\exists j G(j)$ is at most $\sum_{j=0}^{\infty} 2^{-g(j+1)} = (1-2^{-g})^{-1} - 1$. Since g is sufficiently large with respect to θ we have that $P(\exists j, G(j)) < \frac{\theta}{2}$. We claim that $\forall j, G(j)$ implies that the algorithm terminates Phase I through case (a) which will complete the proof of (49). Indeed if $\forall j, G(j)$ then we may give an upper bound on the total computational

time of Phase I. The time that the algorithm spends on computing sets U_{i+1} so that $(\frac{2}{3})^j n \geq |U_i| > (\frac{2}{3})^{j+1} n$ is at most $O(1)g(j+1)|U_i| \leq O(1)g(j+1)(\frac{2}{3})^j n$. Therefore the total time of Phase I is less than $O(1)gn \sum_{j=0}^{\infty}(j+1)(\frac{2}{3})^j n < \frac{k}{2}n$ if k is suffciently large with respect to g. Therefore Phase I terminates through case (a), which completes the proof of the theorem for the more general RAM model.

As we have seen during the description of the algorithm the only step which cannot be realized easily on a true RAM is the representaion of a subset H of $\{1, ..., n\}$ with $O(1)n$ bits in a way that

(i) the representaion can be constructed in time $O(1)|H|$ if the elements of the set are given one by one to the machine in an increasing order, and

(ii) the elements of the set H can be generated from the representation one-by-one in an increasing order, in time $O(1)|H|$.

At a conceptual level we will represent the set $H = \{h_1, ..., h_s\}$, where $1 \leq h_1 < ... < h_s \leq n$ by the sequence d_i $i = 1, ..., h_s$, where $d_1 = h_1$ and $d_i = h_i - h_{i-1}$ for $i = 2, ..., s$. First we note that the total number of bits in the binary representation of the integers $d_1, ..., d_s$ is at most $O(n)$. This is a consequence of the fact that if $w_1, ..., w_s$ are arbitrary nonnegative real numbers and $\sum_{i=1}^{s} w_i \leq n$, then $\sum_{i=1}^{s} \log_2 w_i \leq \sum_{i=1}^{s}(s-1)\log \frac{n}{s-1} \leq n$. (This can be proved by the usual methods for finding the extreme values functions of several variables.)

Our problem is now the following: we have a sequence of positive integers $d_1, ..., d_s$, $1 \leq d_i \leq n$ and we have to make a representation of them in time $O(s)$ if we get them in the given order so that we can produce them again one-by-one in the same order. It is enough to show that we can solve the problem with the additional requirement that $\prod_{i=1}^{s}(d_i + 1) \leq n$, since we may break up the original sequence into maximal disjoint subintervals with this additional property. It follows that from two consecutive intervals at least on will satisfy the inequality $\prod_{i=1}^{s}(d_i + 1) \geq \sqrt{n}$. Consequently if for each subproblem we use a constant number of registers then the total memory requirement is only a constant times larger then the total number of bits in the binary representations of the numbers $d_1, ..., d_s$. The problem with the restriction $\prod_{i=1}^{s}(d_i + 1) \leq n$ can be solved e.g. by storing the rational representation of the finite continued fraction

$$d_1 + \cfrac{1}{d_2 + \cfrac{1}{\cdots\cdots}}$$

generated by sequence $d_1, ..., d_s$. The assumption $\prod_{i=1}^{s}(d_i + 1) \leq n$ implies that the binary representations of the nominator and the denominator of this rational number have at most $O(\log n)$ bits. This is a consequence of the fact that the continued fraction is a rational function $\frac{P}{Q}$ of $d_1, ..., d_s$, where both P and Q are multilinear

polynomials of the variables $d_1, ..., d_s$ with coefficients 0 or 1. Therefore the number of terms in both P and Q is at most 2^s, $s \leq \log n$ and each term is at most $\prod_{i=1}^{s} d_i \leq n$. This implies that both the construction of the continued fraction and the reverse process can be done by rational arithmetic so that for each d_i in both direction we need only a constant number of arithmetic operations. $Q.E.D.$(Theorem 5)

References

[AHU] A. V. Aho, J. E. Hopcroft, J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, 1975.

[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

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

[BFKLT] A. Borodin, M. Fischer, D. Kirckpatrick, N. Lynch, and M. Tompa, A time-space tradeoff for sorting on non-oblivious machines, J. Comput. System Sci., 22 (1981), pp. 351-364.

[BFMUW] A. Borodin, F. Fich, F. Meyer auf der Heide, E. Upfal, and A. Wigderson, A time-space tradeoff for element distinctness, SIAM J. Comput., 16 (1987), pp. 97-99.

[Bo] B. Bollobás, Combinatorics, Cambridge University Press, 1986. p 129. Section 16, Theorem 5.

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

[K] M. Karchmer, Two time-space tradeoffs for elements ditinctness, Theoret. Comput. Sci., 47 (1986), pp. 237-246.

[PPST] W. J. Paul, N. Pippenger, E. Szemerédi, W. T. Trotter, On Determinism versus Non-determinism and Related Problems, 24th Annual Symposium on Foundations of Computer Science, 1983, pp. 429-438

[PR] J. Pagter, T. Rauhe, Optimal Time-Space Trade-Offs for Sorting, 39th Annual Symposium on Foudations of Computer Science, 1998. pp. 264-268,

[Y] A. C. Yao, Near-optimal Time-space Tradeoff for Element Distinctness.

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'