To appear in *Combinatorica*

Invited to SIAM Journal on Computing, FOCS special issue

Conference version in 48th IEEE Symp. on Foundations of Computer Science (FOCS), 2007

One-way multiparty communication lower bound for pointer jumping with applications

Emanuele Viola* Avi Wigderson†

September 9, 2009

Abstract

In this paper we study the one-way multiparty communication model, in which every party speaks exactly once in its turn. For every k, we prove a tight lower bound of $\Omega\left(n^{1/(k-1)}\right)$ on the probabilistic communication complexity of pointer jumping in a k-layered tree, where the pointers of the i-th layer reside on the forehead of the i-th party to speak. The lower bound remains nontrivial even for $k = (\log n)^{1/2-\epsilon}$ parties, for any constant $\epsilon > 0$. Previous to our work a lower bound was known only for $k = 3$ (Wigderson, see [BHK01]), and in restricted models for $k > 3$ [DJS98, Gro06, Cha07a, BC08, Bro09]. Our results have the following consequences to other models and problems, extending previous work in several directions.

The one-way model is strong enough to capture *general* (not one-way) multiparty protocols with a bounded number of rounds. Thus we generalize two problem areas previously studied in the 2-party model (cf. [PS84, DGS87, NW93]). The first is a rounds hierarchy: we give an exponential separation between the power of r and $2r$ rounds in general probabilistic k-party protocols, for any k and r. The second is the relative power of determinism and nondeterminism: we prove an exponential separation between nondeterministic and deterministic communication complexity for general k-party protocols with r rounds, for any k, r.

The pointer jumping function is weak enough to be a special case of the well-studied disjointness function. Thus we obtain a lower bound of $\Omega\left(n^{1/(k-1)}\right)$ on the probabilistic complexity of k-set disjointness in the one-way model, which was known only for $k = 3$ parties. Our result also extends a similar lower bound for the weaker simultaneous model, in which parties simultaneously send one message to a referee [BPSW05].

Finally, we infer an exponential separation between the power of any two different orders in which parties send messages in the one-way model, for every k. Previous results [NW93, BHK01] separated orders based on who speaks first.

Our lower bound technique, which handles functions of high discrepancy over cylinder intersections, provides a "party-elimination" induction, based on a restricted form of a direct-product result, specific to the pointer jumping function.

*The author is supported by NSF grant CCF-0845003. This work was partially done while the author was a postdoctoral fellow at the Institute for Advanced Study, supported by NSF grant CCR-0324906.

†The author is supported by NSF grant CCR-0324906.

1 Introduction

In Yao's standard communication complexity model [Yao79], two parties each hold an input, and they attempt to compute (or approximate) a given function of these two inputs by exchanging at most c bits of communication (cf. the excellent book [KN97]). This model has been one of the most extensively studied in computational complexity theory, and captures essential features of diverse computational settings, from Turing machines, VLSI, and distributed computation, to linear programming and auctions. Many techniques for proving strong lower bounds are known.

This model was generalized by Chandra, Furst, and Lipton [CFL83], to the *multiparty model* (often called "number-on-the-forehead"). In k-party communication complexity party i is assigned input $x_i \in X_i = \{0,1\}^n$ $(i = 1, \ldots, k)$, and again the parties have to compute (or approximate) a function $f : X_1 \times X_2 \times \cdots \times X_k \to \{0,1\}$ of all k inputs by exchanging c bits of communication. However, the input $x_i \in X_i$ to the i-th party (figuratively) resides on that party's forehead, and so (formally) each party knows *all but* its own input. The overlapping information of the parties allows this model to capture more complex settings, like multi-tape Turing machines [BNS92], branching programs [CFL83], constant-depth circuits with modular gates [HG91] and more. In the multiparty model, the celebrated work by Babai, Nisan, and Szegedy [BNS92] exhibits explicit functions that require large communication for as many as $k = k(n) = (1-\epsilon)\log n$ parties (see also [BHK01, CT93, Raz00, VW08]). Improving their result remains an outstanding challenge: even for the most restricted form of this model, namely *simultaneous multiparty protocols* [BGKL03] (in which the parties do not interact with each other but each party simultaneously sends a message to the referee, and the referee, who does not see any of the input, announces the answer) it is consistent with the current state of knowledge that every explicit function is computable by $k = \log_2 n$ parties and $c = \log_2 n$ communication. Strong lower bounds for any $k = \log^{\omega(1)} n$ would have significant consequences for circuit lower bounds via results by Yao [Yao90], Beigel and Tarui [BT94], and Håstad and Goldmann [HG91].

In this paper we obtain several new results on natural variants of and questions about the multiparty model, which we describe in the next subsections with their background.

1.1 Rounds in communication complexity.

Papadimitriou and Sipser [PS84] initiated the study of how 2-party communication complexity is affected by limiting the parties to r rounds of messages. This natural question was taken up by several researchers: Duris, Galil, and Schnitger [DGS87] obtained an exponential separation between r and $r+1$ rounds in 2-party protocols, and their bound was later improved by Nisan and Wigderson [NW93]. No separation was known for $k > 2$ parties. In this work we obtain the first such result, stated next. For the definition or rounds, see Section 2.

Theorem 1.1 (Rounds hierarchy)**.** *For every r and k there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that requires $n^{\Omega(1)}$ communication for k-party r-round protocols, but that can be computed with $O(\log n)$ communication by k-party r'-round protocols, if $r'(k-1) \geq r \cdot k$.*

For k-party r-round protocols we also obtain an exponential separation between *nondeterministic* and deterministic protocols. Again, this seems to be the first such separation for $k > 2$ parties (see [PS84, DGS87, NW93] for separations for $k = 2$).

Theorem 1.2 (Nondeterminism vs. determinism)**.** *For every r and k there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that requires $n^{\Omega(1)}$ communication for k-party r-round protocols, but that can be computed with $O(\log n)$ communication by nondeterministic k-party 2-round protocols.*

Since the preliminary publication of our results [VW07], much progress has been made on separating nondeterminism from determinism in multiparty communication complexity (with no restriction on the number of rounds). Non-explicit separations for a large number of parties are given in [BDPW07]. Recent progress on explicit separations [She09, She08b, Cha07b, LS09, CA08, DPV08, BHN09] was sparked by the work by Sherstov [She09, She08b], see his survey [She08a]. The strongest explicit separation has been obtained by David, Pitassi, and Viola [DPV08] and holds for up to $k = (1-\epsilon)\log n$ parties. We note that in this separation the nondeterministic protocol uses 2 rounds, thus improving on Theorem 1.2.

The two theorems stated will be derived as corollaries to a new communication complexity lower bound in the one-way model, which we describe next.

1.2 The one-way model and pointer jumping.

In the one-way model, the k parties speak in a given order, and each party speaks only once. This is an interesting model that arises in several contexts, such as oblivious branching programs and streaming algorithms [AMS99, BYJKS04]. In this work we establish a new one-way lower bound for the *pointer jumping* function. To define this function, we think of a regular tree[1] of depth k. The input to the pointer jumping function specifies a $\{0,1\}$-value for each leaf, and for every internal node a pointer to one of its children. The output of the pointer jumping function is the bit obtained by following the pointers from the root to a leaf (see Figure 1). Party i knows all pointers except those from layer i to layer $i+1$.

The pointer jumping function naturally captures "inherently sequential" computation and inference. It has been studied by a number of researchers in various models and contexts, including constant-depth circuits, proof complexity, PRAMs, and communication complexity (e. g., [PS84, NW93, NBY97, BIP98, DJS98, BHK01, Gro06, Cha07a, BC08, Bro09]).

In particular, attempts to prove lower bounds for pointer jumping in the one-way multiparty model have existed for over a decade. For $k = 3$ this was achieved by Wigderson (cf. appendix in [BHK01]), who proved an $\Omega(\sqrt{n})$ lower bound via extremal set theory arguments. For $k > 3$ strong lower bounds were obtained for various restrictions of the one-way model, such as the "conservative" [DJS98] and the "myopic" model [Gro06], see also [Cha07a, BC08, Bro09]. The main result of this paper makes progress in the general one-way model, obtaining a tight lower bound for any constant number k of parties. Our lower bound remains nontrivial as long as the number of parties is less than $(\log n)^{1/2-\Omega(1)}$.

Theorem 1.3. *The pointer jumping function $TPJ_k : (\{0,1\}^n)^k \to \{0,1\}$ requires at least $n^{1/(k-1)}/k^{O(k)}$ bits of communication in the one-way k-party model.*

We also apply Theorem 1.3 to the study of the effect of the order in which parties speak. Nisan and Wigderson [NW93] demonstrate an exponential gap between the one-way complexities depending on who speaks first for $k = 3$ parties. Babai, Hayes, and Kimmel [BHK01, Corollary 6.4] extend this to $k \leq (1-\epsilon)\log n$ parties. Their result gives a function that is hard if party 1

[1] It makes sense to consider this function over arbitrary directed graphs, but the case of trees already suffices for all of our applications.

speaks first, and is easy otherwise. We prove a different order separation by exhibiting a function (namely, pointer jumping) that is hard for exactly one order, a result which does not seem to follow from [BHK01] (albeit we do not quite handle as many parties as [BHK01]).

Corollary 1.4 (Order separation)**.** *For any k and any order π for the k parties to speak, there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that requires $n^{\Omega(1)}$ communication if the parties speak in the order π, but can be computed with $O(\log n)$ communication if the parties speak in any other order.*

The separation remains nontrivial even for $k = (\log n)^{1/2-\epsilon}$ parties, for any constant $\epsilon > 0$.

1.3 Discrepancy and disjointness.

Another reason for our interest in the one-way model is that, for some basic functions, it is the strongest model in which multiparty communication complexity lower bounds are known. Specifically, following [BNS92], lower bounds in the general model are typically obtained by giving an upper bound on the *discrepancy* of the target function over cylinder intersections, a quantity which measures the bias of the function over large cylinder intersections (we will always be referring to discrepancy over cylinder intersections). Although we know of functions with exponentially small discrepancy [BNS92, BHK01], such as generalized inner product [BNS92] (see also [CT93, Raz00, VW08]), this quantity is large for some important functions such as pointer jumping, making it necessary to use a different technique.

Another fundamental example of a function with high discrepancy is the *disjointness* function. Here, each of the k parties is assigned a subset of $\{1, 2, \ldots, n\}$, and the parties wish to determine whether or not their sets share a common element. The study of the disjointness function is central to communication complexity, and is motivated both by the naturalness of the function and by its interesting consequences in proof complexity and circuit complexity. For example, in propositional proof complexity theory, it is shown in [BPS07] that sufficiently strong ($c = \omega(\log^4 n)$) lower bounds for k-party disjointness would solve the major open problem of obtaining proof size lower bounds for a family of proof systems which capture certain semi-definite optimization techniques, and are known as "tree-like, degree-$(k-1)$ threshold systems." In circuit complexity theory, it can be shown, using techniques from [RW93, Vio07], that sufficiently strong ($\log^d n$ for a certain constant d) lower bounds for k-party disjointness for any $k = \Omega(\log n)$ would separate polynomial-size constant-depth circuits (with "and," "or," and "not" gates) from polynomial-size depth-2 circuits with "and" gates and an arbitrary symmetric gate at the root.

The difficulty of obtaining such lower bounds for the general multiparty model has caused researchers to try more restricted models first. For disjointness, an $\Omega\left(n^{1/(k-1)}\right)$ lower bound was known in the simultaneous model [BPSW05]. In the stronger one-way model an $n^{\Omega(1)}$ lower bound was only known for $k = 3$ (cf. appendix in [BHK01] and [BPSW05]).[2] In this work we extend these previous results by obtaining an $\Omega\left(n^{1/(k-1)}\right)$ lower bound for one-way protocols.

Corollary 1.5. *The disjointness function $DISJ_{k,n} : (\{0,1\}^n)^k \to \{0,1\}$ requires at least $n^{1/(k-1)}/k^{O(k)}$ bits of communication in the one-way k-party model.*

[2] We note that in [BPSW05] they also prove an $n^{\Omega(1)}$ bound for 3-party disjointness in a stronger model where the first party speaks once, and then the other parties interact arbitrarily. This stronger model also appears in [NW93, BHK01] (for other functions, and in the latter with several parties).

We note that this result is a corollary to our lower bound for pointer jumping. It is easy to see that when the underlying graph is a tree, pointer jumping is a special case of disjointness.

Since the preliminary publication of our results [VW07], there has been exciting progress on both communication lower bounds for disjointness and the applications mentioned above in this section. This progress largely overlaps with the progress on separating nondeterminism from determinism, mentioned earlier after Theorem 1.2. In particular, Lee and Shraibman [LS09] and Chattopadhyay and Ada [CA08] prove an $\Omega\left(n^{1/k+1}/2^{2^k}\right)$ lower bound for k-party disjointness in the general model, which is non-trivial for any $k = o(\log \log n)$. More recently, Beame and Huynh-Ngoc [BHN09] prove an $\Omega\left(2^{\sqrt{(\log n)/k}-k}\right)$ lower bound for k-party disjointness in the general model, which remains non-trivial for any $k = o(\log^{1/3})$.

1.4 Techniques

We now summarize how we obtain our lower bound for pointer jumping (Theorem 1.3). The bound is obtained by induction on k, the number of parties. Carrying through this induction is somewhat involved, and in particular we will need to consider non-Boolean variants of pointer jumping, in which one has to chase pointers on several trees. Specifically, we will work with the function $TPJ_k^{m,d} : (\{0,1\}^n)^k \to \{0,1\}^m$, which we think of as pointer jumping on a forest of m d-regular trees of depth k. Again, the input to $TPJ_k^{m,d}$ specifies a $\{0,1\}$-value for the leaves, and for every internal node a pointer to one of its d children. The output of the pointer jumping function is the m-bit string obtained by following the paths from the m roots to the leaves. It is trivial to compute $TPJ_k^{m,d}$ correctly with probability $\exp(-m) := 2^{-m}$ over the choice of a random input (any output will do). It will be important for us to show that computing $TPJ_k^{m,d}$ is difficult even if all we require is a success probability of $\exp(-o(m))$ over the choice of a random input. Specifically, we prove by induction on k that the following statement holds for every choice of m and d (cf. Theorem 3.2), which may be viewed as a direct product theorem in this model, specific to pointer jumping:

> (I) No k-party one-way protocol using communication $o(m \cdot d)$ can compute $TPJ_k^{m,d}$ with probability $\exp(-o(m))$ over the choice of a random input.

Proving the inductive step in (I) is the main technical contribution of this work. Our main tool is a *party-reduction* theorem (Theorem 3.3). This theorem shows how to transform a given protocol for $TPJ_k^{m,d}$ into another protocol for pointer jumping with $k-1$ parties on $m' := d \cdot m$ trees, $TPJ_{k-1}^{m',d}$. For example, in the special case in which we start with a Boolean ($m = 1$) function with k parties, we construct a protocol for a non-Boolean ($m' = d$) function with $k-1$ parties. Our party-reduction theorem gives a general transformation for protocols in which the first party speaks only once, and thus it applies to other models beyond one-way, and other functions besides pointer jumping.

This transformation is obtained as follows. We fix a message for the first party that maximizes the success probability of the protocol. Thinking of the input as $(x_1, y) \in X_1 \times Y$, where $Y = X_2 \times \cdots \times X_k$, we note that such fixing decreases the probability of success over the choice of y (however, this loss will be $\exp(-o(d \cdot m)) = \exp(-o(m'))$, which is good enough for (I)), but does not decrease the success probability over the choice of x_1, because the message does not depend on x_1. This bigger success probability over x_1 allows us to run the protocol $r \approx d$ times for different choices of pointers $x_1 \in X_1$ and still tolerate the loss in the error (recall that $d = m'/m$). At this

point, we argue that the r outputs of the protocol let us compute $(1-o(1))\cdot m'$ out of the m' bits of $TPJ_{k-1}^{m',d}$ that we need. To argue this, we proceed by showing that only with exponentially small probability do the pointers (given by all r runs of the protocol) hit fewer than $(1-o(1))\cdot m'$ bits. By a suitable choice of parameters, this probability can be made smaller than the probability that all r runs of the protocol are correct. Thus, with sufficiently high probability, all r runs of the protocol are correct *and* their accumulated pointers hit at least $(1-o(1))\cdot m'$ bits. When this happens, we can compute each of these $(1-o(1))\cdot m'$ bits by outputting the corresponding output bit of the protocol.

It remains to deal with the $o(m)'$ bits we know nothing about. To compute these, we simply guess at random. The probability of guessing them correctly will be $\exp(-o(m'))$, which is good enough for (I). This concludes our proof overview (cf. Remark 4.3 for the history of this approach).

To clarify the parameters in the above discussion we remark that, intuitively, $(1-o(1))\cdot m'$ bits is as much information as we can hope to obtain from the protocol: there may well be a fraction of $o(m)'$ indices such that the protocol never gives a meaningful answer if a pointer falls there; such a protocol would satisfy the inductive hypothesis because its success probability would be $\exp(-o(m))$, but its outputs give no information on the value of $o(m)'$ bits.

Finally, we point out that the idea of employing a party-reduction theorem is already present in the work [DJS98], where it is used to derive lower bounds in communication models that are more restricted than the one-way model considered here. Party-reduction is also employed in lower bounds in [NW93, BHK01, BPSW05] which proceed by eliminating the first party and then analyzing the arbitrary (not one-way) communication between the remaining parties.

Organization. This paper is organized as follows. After some preliminaries in Section 2, in Section 3 we formally state our lower bounds for pointer jumping, including the party-reduction theorem, and then we derive our lower bound assuming the party-reduction theorem. In Section 4 we prove the party-reduction theorem. The consequences of our lower bound for general protocols with bounded number of rounds (Theorems 1.1 and 1.2), relative power of different orders in the one-way model (Corollary 1.4), and disjointness (Corollary 1.5), are given in Section 5. Finally, in Section 6 we discuss some applications of our results to streaming models of computation.

2 Background on communication complexity

We use standard definitions of communication protocols, which we now briefly recall. In a k-party protocol, k parties wish to collaboratively compute a function $f : X_1 \times X_2 \times \cdots \times X_k \to D$ on input $x = (x_1, x_2, \ldots, x_k)$. The function f is known to each party, and the i-th party sees all pieces of x *except* x_i. The parties communicate by broadcasting messages, and the final party is supposed to communicate the output of the function. The communication complexity of a protocol is the total length of the messages exchanged; note that we count the final message (which if the protocol is correct is the value of the function) towards communication.

In an *r-round protocol*, the parties speak in any order $\pi \in [k]^r$, which is chosen adaptively during the protocol: at any round, the party to speak is a function of the messages exchanged before that round. We stress that r may be much larger than k, and that parties may speak more than once. We now define more restricted models where the order is fixed a priori, i. e., the parties always speak in the same order regardless of the messages exchanged and of the input. A *one-way*

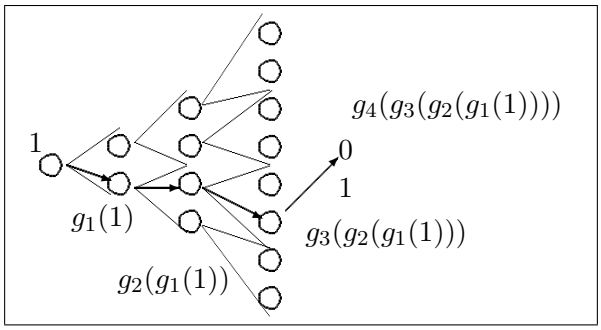

Figure 1: TPJ function ($k = 4, m = 1, d = 2$).

protocol with order π is a $(r = k)$-round protocol in which the parties always speak in the order π, and π is a permutation of $[k]$. A (standard) *one-way protocol* is a one-way protocol with order π where π is the identity permutation $\pi = (1, 2, 3, \ldots, k)$.

For concreteness, we note that a 1-party one-way protocol $P : X_1 \to \{0,1\}^m$ is simply a fixed string $P(x_1) = P \in \{0,1\}^m$, independent of the input. A 2-party $c-$bit one-way protocol $P : X_1 \times X_2 \to \{0,1\}^m$ is given by two functions f_1 and f_2, where $P(x_1, x_2) = f_2(f_1(x_2), x_1)$, and the output length of each function is bounded by c. The function f_1 computes the message of the first party, which depends on x_2 only, while f_2 computes the (output) message of the second party, which depends on x_1 and the message $f_1(x_2)$ sent by the first party. Similarly, a 3-party c-bit one-way protocol $P : X_1 \times X_2 \times X_3 \to \{0,1\}^m$ is given by three functions f_1, f_2, and f_3, where $P(x_1, x_2, x_3) = f_3(f_2(f_1(x_2, x_3), x_1, x_3), x_1, x_2)$, and the output length of each function is bounded by c.

A *probabilistic* or *randomized protocol* is a probability distribution over deterministic protocols. A *nondeterministic protocol* P is given by a collection of protocols P_y such that $P(x) = z$ if and only if there exists y such that $P_y(x) = z$. The communication used by P is $|y|$ plus the maximum over y of the communication used by P_y.

3 Lower bound for pointer jumping

We now define the pointer jumping function and then state our main result. The pointer jumping function can be defined in different ways, and in this work we adopt the following variant. We think of a forest of m d-regular trees of depth k. The input to the pointer jumping function specifies a $\{0,1\}$-value for each leaf, and for every internal node a pointer to one of its d children. (Specifically, for every level the j-th internal node at that level has its d children labelled from the interval $[d \cdot (j-1) + 1, d \cdot j]$.) The output of the pointer jumping function is the m-bit string obtained by following the paths from the m roots to the leaves, see Figure 1. Compared to other definitions (e. g., [BHK01]), ours has the advantage of having a shorter input length (thus giving a stronger lower bound), and more directly implying a lower bound for disjointness (cf. Corollary 5.3).

Definition 3.1 (Pointer jumping function)**.** For integers k, m, d, the pointer jumping function is

$$TPJ_k^{m,d} : [d]^m \times [d]^{m \cdot d} \times [d]^{m \cdot d^2} \times \cdots \times [d]^{m \cdot d^{k-2}} \times \{0,1\}^{m \cdot d^{k-1}} \to \{0,1\}^m,$$

where for every i, and for $j \in [m]$ we have

$$TPJ_k^{m,d}(g_1, g_2, \ldots, g_k)_j := (g_k \circ g_{k-1} \circ \cdots \circ g_1)(j) = g_k(g_{k-1}(\cdots(g_1(j))\cdots),$$

where $g_k : [m \cdot d^{k-1}] \to \{0,1\}$ and, for $i < k$, $g_i \in BF_d^{m \cdot d^{i-1}}$ using the definition

$$BF_d^t := \{g : [t] \to [t \cdot d] \text{ such that } g(j) \in [d \cdot (j-1) + 1, d \cdot j]\}.$$

Note that the function $TPJ_k^{m,d}$ has input length $n = m \cdot (d^{k-1} + d^{k-2} \cdot \log d + d^{k-3} \cdot \log d + \cdots + d^0 \cdot \log d) = O(m \cdot d^{k-1})$. This function is Boolean for $m = 1$, while for $m > 1$ its output is an m-bit string corresponding to the outputs of m Boolean pointer jumping functions on independent inputs. Recall from Section 2 that party i knows all the input $(g_1, g_2, \ldots, g_k)$ except g_i, and that in a one-way protocol players speak once and in the natural order $\pi = (1, 2, \ldots, k)$.

We now state our lower bound for pointer jumping. Here and throughout the paper we let $\exp(x) := 2^x$ and $\log = \log_2$. Also, all probabilities are taken over the uniform distribution.

Theorem 3.2. *For every $k, m, d \geq 1$, and for $\mu_k := (k+1)^{-100k}$, there is no one-way k-party protocol P such that*

$$\Pr_{x \in X}\left[P(x) = TPJ_k^{m,d}(x)\right] \geq \exp(-\mu_k \cdot m)$$

and P uses communication at most $\mu_k \cdot m \cdot d$.

In particular, every one-way k-party protocol for the Boolean pointer jumping function $TPJ_k^{1,d} : \{0,1\}^n \to \{0,1\}$ must use communication at least $d/k^{O(k)} = n^{1/(k-1)}/k^{O(k)}$.

Note that, in particular, for every constant $\epsilon > 0$ and every $k \leq \log^{1/2-\epsilon} n$ we have a lower bound of $\exp\left(\log^{\Omega(1)} n\right)$. Also note that the $TPJ_k^{1,d}$ function can be computed with $d+1$ bits of communication – simply by letting the first party communicate the d possible outputs of the function for each value of the input on its forehead, and then letting the second party announce the output of the function. Thus, for constant k our lower bound is tight up to constant factors; but in general we do not know whether our $d/k^{O(k)}$ bound is tight or not. Finally, we note that the bound in the above theorem is *distributional*, i. e., it limits the fraction of inputs on which any low-communication protocol can compute $TPJ_k^{m,d}$ correctly. By an averaging argument (see, e. g., [KN97, Theorem 3.20]), this implies a lower bound for randomized protocols as well.

The basic intuition behind the proof of Theorem 3.2 is that if we can solve m problems correctly (i. e., compute the m output bits of $TPJ_k^{m,d}$) with some success probability and communication, then by repeating the protocol roughly d times we should be able to solve $m \cdot d = m'$ problems (i. e., compute the m' output bits of $TPJ_{k-1}^{m',d}$) with a d-fold increase of communication and an exponential (in d) decrease in success probability.[3]

We use the idea of the previous paragraph inductively to reach a contradiction. Specifically, the proof of Theorem 3.2 proceeds by induction on k. For every k, we prove the theorem for every setting of the other parameters. For the inductive step, we show how to turn a given protocol for the function with k parties into another protocol for the function with $k-1$ parties on a bigger

[3] Our initial version of this argument was, intuitively, that being able to compute *most* of the m output bits of $TPJ_k^{m,d}$ implies the ability to compute *most* of the m' output bits of $TPJ_{k-1}^{m',d}$. The proof of this was complex, and the simplification here derives from an idea in [BARdW08] which was pointed out to us by Oded Regev. We will stress in the proof where their idea is used.

forest (for example, in the special case in which we start with a Boolean function with k parties, we construct a protocol for a non-Boolean function with $k-1$ parties). This transformation is provided by our party-reduction theorem which we state next (see Definition 3.1 for the notation BF_d^m).

Theorem 3.3 (Main theorem: party reduction)**.** *Let m, d be integers and $m' := d \cdot m$. Let $f' : X \to \{0,1\}^{m'}$ be any function, and consider the function $f : BF_d^m \times X \to \{0,1\}^m$ defined as $f(g,x)_j := f'(x)_{g(j)}$ where $g \in BF_d^m$.*

Suppose that there is a protocol $P(g,x)$ using communication $\mu \cdot m \cdot d$ such that

$$\Pr_{g \in BF_d^m, x \in X}[P(g,x) = f(g,x)] \geq \exp(-\mu \cdot m),$$

where g is on the forehead of a party who speaks first and only once, and $\mu < 1/2$.

Then there is a protocol $P'(x)$ using communication $\mu' \cdot m' \cdot d$ such that

$$\Pr_{x \in X}\left[P'(x) = f'(x)\right] \geq \exp(-\mu' \cdot m'),$$

where

$$\mu' = 100\mu \cdot \log(1/\mu)$$

and P' can be written as $P'(x) = h\left(P(\tilde{g}_1, x), P(\tilde{g}_2, x), \ldots, P(\tilde{g}_r, x)\right)$ for some fixed function $h : \{0,1\}^{m \cdot r} \to \{0,1\}^{m'}$ and fixed inputs $\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_r \in BF_d^m$ (and some r). In particular, if $X = X_2 \times \cdots \times X_k$ and P is a one-way k-party protocol, then P' is a one-way $(k-1)$-party protocol.

Proof of Theorem 3.2 from Theorem 3.3. To get a sense of the parameters, note that the statement we intend to prove is monotone in μ_k. Specifically, for any k, m, and d, if there is no protocol P such that $\Pr_{x \in X}\left[P(x) = TPJ_k^{m,d}(x)\right] \geq \exp(-\mu_k \cdot m)$ and P uses communication $\mu_k \cdot m \cdot d$, where $\mu_k = \alpha$, then the same is true for any choice of smaller $\mu_k := \beta < \alpha$.

We proceed by induction on k.

Base case $k = 1$. In this case the communication does not play a role and P is a fixed string $P \in \{0,1\}^m$, whereas $TPJ_1^{m,d}(x) = x$. The probability that a random input m-bit string x equals a fixed string P is $\exp(-m) < \exp(-\mu_1 \cdot m)$, and thus the base case is proved.

Inductive step. We prove the inductive step $k \geq 2$ by contradiction. Suppose that there is a protocol for $TPJ_k^{m,d}$ with parameter $\mu_k = (k+1)^{-100k} < 1/2$. We apply Theorem 3.3 to obtain a protocol for $TPJ_{k-1}^{d \cdot m, d}$ with parameter μ_{k-1} which equals

$$\begin{aligned} 100\mu_k \cdot \log(1/\mu_k) &= 100(k+1)^{-100k} \cdot 100k \cdot \log(k+1) \\ &\leq (k+1)^{-100k} \cdot k^{100} \leq k^{-100k} \cdot k^{100} = k^{-100(k-1)}. \end{aligned}$$

This contradicts the inductive hypothesis for $(k-1)$.

The "in particular" part of the theorem follows from the easily verifiable fact that the input length of $TPJ_k^{1,d}$ is $n = O(d^{k-1})$ (cf. the discussion after Definition 3.1). □

In the next section we prove Theorem 3.3. The reader may want to refer to Figure 2 while reading the proof.

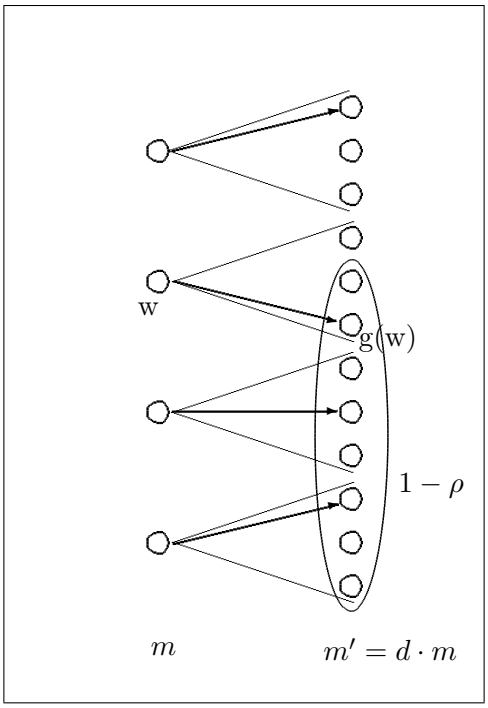

Figure 2: Variables of the proof of Theorem 3.3.

4 Proof of the party-reduction Theorem 3.3

We assume without loss of generality that $d \geq 1/\mu$. This is because if $d < 1/\mu$ then the protocol P uses fewer than m bits of communication, and since we count the output length of P towards communication, the protocol cannot even output an m-bit string, in which case the hypothesis of the theorem is always false and the theorem vacuously true.

From the hypothesis and a Markov argument[4] it follows that

$$\Pr_{x\in X}\left[\Pr_{g\in BF_d^m}[P(g,x) = f(g,x)] \geq \exp(-2\mu\cdot m)\right] \geq \exp(-\mu\cdot m)/(1+\exp(-\mu\cdot m)) \geq \exp(-\mu\cdot m)/2.$$

The message sent by the first party in P does not depend on g, since g lies on the forehead of the first party. Using this and the fact that the communication of P is $\mu\cdot m\cdot d = \mu\cdot m'$, we can fix a particular message a for the first party such that

$$\Pr_{x\in X}\left[\Pr_{g\in BF_d^m}[P_a(g,x) = f(g,x)] \geq \exp(-2\mu\cdot m)\right] \geq \exp(-\mu\cdot m')\cdot\exp(-\mu\cdot m)/2 \geq \exp(-3\mu\cdot m'), \tag{1}$$

where P_a denotes the $(k-1)$-party $(\mu\cdot m')$-bit protocol obtained from P by fixing the first party's message to a (and we are using that $\mu\cdot m' \geq 1$ because $d \geq 1/\mu$). We are now in the position to

[4] Specifically, suppose that $\Pr_{X,Y}[\phi(X,Y)] \geq \gamma$. Let us call x *heavy* if $\Pr_Y[\phi(x,Y)] \geq \gamma'$, and suppose that $\Pr_X[X \text{ is heavy}] = p$. Then we have $\gamma \leq \gamma'(1-p) + p = \gamma' + p(1-\gamma')$, and thus $p > (\gamma-\gamma')/(1-\gamma')$. (The case $\gamma := \exp(-\mu\cdot m)$, and $\gamma' := \gamma^2$ is used in the proof.)

describe the protocol P'. We define it as a randomized protocol; a deterministic protocol can be obtained by fixing its random choices.

Definition of P'. Let

$$r := 5 \cdot \lceil \log(1/\mu) \rceil \cdot d.$$

On input x, the protocol P' first chooses r independent inputs $g_1, g_2, \ldots, g_r$, then runs the protocol P_a on all inputs $(g_1, x), (g_2, x), \ldots, (g_r, x)$. Informally, to compute its j-th output bit, P' uses the output of P_a if one of the g_i happens to point to j, and otherwise it tosses a coin. More formally, to compute the j-th output bit, P' considers the first pair $u \in [r], w \in [m]$ such that $g_u(w) = j$. If any such u, w exist, then it outputs $P_a(g_u, x)_w$. If no such u, w exist, it outputs a random and uniform bit. It may be the case that P_a gives inconsistent answers on different pairs u, w such that $g_u(w) = j$. While this is addressed in the definition above by using the first such pair, this case of inconsistent pairs u, w is in fact irrelevant to our analysis below. This is because the analysis below gives the desired probability bound just considering the event in which all outputs are correct and hence consistent.

Complexity of P'. It is easy to verify that the randomized protocol P' can be written as $P'(x) = h\left(P_a(g_1, x), P_a(g_2, x), \ldots, P_a(g_r, x)\right)$, for a function h which depends on the g's. Since P' runs r times the protocol P_a, and P_a uses communication $\mu \cdot m'$, we see that P' uses communication $\mu \cdot m' \cdot r = \mu \cdot m' \cdot 5 \cdot \lceil \log(1/\mu) \rceil \cdot d \leq 10\mu \cdot \log(1/\mu) \cdot m' \cdot d$, as required. A deterministic protocol can be obtained by fixing the randomness $\tilde{g}_1 = g_1, \tilde{g}_2 = g_2, \ldots, \tilde{g}_r = g_r$ so as to maximize the success probability of P'.

Analysis of P'. Call an input $x \in X$ *good* if $\Pr_{g \in BF_d^m}\left[P_a(g, x) = f(g, x)\right] \geq \exp(-2\mu \cdot m)$. Our claim is the following.

Claim 4.1. *For every good $x \in X$, $\Pr_{g_1, \ldots, g_r}[P'(x) = f'(x)] \geq \exp(-4\mu \cdot m \cdot r)$.*

Before proving this claim, let us see how the theorem follows from it. Note that, by Equation (1), a random $x \in X$ is good with probability at least $\exp(-3\mu \cdot m')$. Combining this with Claim 4.1 we have that

$$\begin{aligned} \Pr_{x \in X, g_1, \ldots, g_r}[P'(x) = f'(x)] &\geq \exp(-3\mu \cdot m' - 4\mu \cdot m \cdot r) \\ &\geq \exp(-7\mu \cdot m \cdot r) \geq \exp(-70\mu \cdot \log(1/\mu) \cdot m') \quad (2) \end{aligned}$$

as required.

Proof of Claim 4.1. The choice of $g_1, \ldots, g_r$ corresponds to the choice of $t := m \cdot r = 5 \cdot \lceil \log(1/\mu) \rceil \cdot m'$ pointers $g_u(w) \in [m']$, where $u \leq r$ and $w \leq m$. Let us call a pointer $g_u(w)$ *good* if P_a correctly computes the corresponding bit, that is, if $P_a(g_u, x)_w = f'(x)_{g_u(w)}$.

Let us pick a parameter

$$\rho \in [4\mu, 5\mu]$$

such that $\rho \cdot d$ is an integer (such a ρ is guaranteed to exist because $d \geq 1/\mu$). Now consider the following two events.

I *All the $g_u(w)$ are good*: For all t pointers $g_u(w)$ we have $P_a(g_u, x)_w = f'(x)_{g_u(w)}$, and

II *Cover a $1-\rho$ fraction*: there are at least $(1-\rho)\cdot m'$ indices $j \in [m']$ for which there is a pointer $g_u(w)$ such that $g_u(w) = j$.

Whenever both events (I) and (II) happen, by definition the protocol will compute correctly a $1-\rho$ fraction of the output bits of f', and guess at random for the other ρ fraction. We prove below that the probability that both events (I) and (II) happen is at least $\exp(-2\mu\cdot m\cdot r)/2$. Also, the probability that the $\rho\cdot m'$ random guesses are all correct is $\exp(-\rho\cdot m')$, and thus we have, for every good x,

$$\begin{aligned}\Pr_{g_1,\ldots,g_r}[P'(x) = f'(x)] &\geq \Pr[(I) \text{ and } (II)]\cdot\exp(-\rho\cdot m')\\ &\geq \exp(-2\mu\cdot m\cdot r - 1 - \rho\cdot m')\\ &\geq \exp(-4\mu\cdot m\cdot r) \qquad (\text{since } \rho\leq 5\mu\leq \mu\cdot r/d),\end{aligned}$$

which proves Claim 4.1.[5]

It remains to prove that the probability that both events (I) and (II) happen is at least $\exp(-2\mu\cdot m\cdot r)/2$. First, observe that event (I) happens with probability at least $\exp(-2\mu\cdot m\cdot r)$. This is because x is good and therefore, for every $u\leq r$, with probability at least $\exp(-2\mu\cdot m)$ (over g_u) all pointers $g_u(1), g_u(2), \ldots, g_u(m)$ are good (cf. Equation (1)). As these events (for different values of u) are independent, the probability bound follows. So we only need to show that event (II) does not happen with probability at most $\exp(-2\mu\cdot m\cdot r)/2$. (We are using that $\Pr[(I) \text{ and } (II)] \geq \Pr[(I)] - \Pr[\text{not } (II)]$.)

Claim 4.2 (Cover a $1-\rho$ fraction w.h.p.)**.** *The probability that there are at most $(1-\rho)\cdot m'$ indices $j\in[m']$ for which there is a pointer $g_u(w)$ such that $g_u(w) = j$ is at most* $\exp(-2\mu\cdot m\cdot r)/2$.

Proof of Claim 4.2. We do a union bound over all subsets of $[m']$ of size $(1-\rho)\cdot m'$. Fix any such set S, and let $1-\rho_1, 1-\rho_2, \ldots, 1-\rho_m$ be the fraction of elements in each block of d indices, i. e., $1-\rho_i := |\{j\in[d\cdot(i-1)+1, d\cdot i]\cap S\}|/d$. The probability that all pointers fall in S is $\prod_{i\leq m}(1-\rho_i)^r$. Since the average of the $(1-\rho_i)'s$ is $(1-\rho)$, by the arithmetic mean vs. geometric mean inequality we have that the probability that all pointers fall in S is

$$\prod_{i\leq m}(1-\rho_i)^r \leq (1-\rho)^{m\cdot r}.$$

By a union bound, the probability that there is a set $S\subseteq[m']$ of size $(1-\rho)\cdot m'$ such that all the pointers fall inside S is at most

$$\begin{aligned}\binom{m'}{\rho\cdot m'}\cdot(1-\rho)^{m\cdot r} &\leq (e/\rho)^{\rho\cdot m'}\cdot\exp(-\rho\cdot m\cdot r)\\ &\leq \exp(\log(1/\mu)\cdot\rho\cdot m')\cdot\exp(-\rho\cdot m\cdot r) \qquad (\text{since } \rho\geq 4\mu)\\ &\leq \exp(\mu\cdot m\cdot r - 4\mu\cdot m\cdot r) \qquad (\text{since } \rho\in[4\mu, 5\mu] \text{ and } r = 5\cdot\lceil\log(1/\mu)\rceil\cdot d)\\ &= \exp(-3\mu\cdot m\cdot r),\end{aligned}$$

[5] We point out that this idea of guessing at random the missing ρ fraction of bits is borrowed from [BARdW08]. Just like in that paper, the exponentially small success probability is significant for our setting of parameters. Previously, without using this idea, we had to cope with errors which led to more complex analysis.

where in the first inequality we use that $\binom{n}{k} \leq (e \cdot n/k)^k$ and that $(1-\rho) \leq e^{-\rho} \leq 2^{-\rho} =: \exp(-\rho)$. □

□

Remark 4.3. We thank Oded Regev for pointing out an idea from the paper [BARdW08] which led to a simplification of the proof of our main party-reduction theorem. Our previous proof (unpublished) gave up on recovering all the m' bits of f', and only aimed to recover a $1-\rho$ fraction. This parameter ρ then appeared in the induction hypothesis leading to a more cumbersome analysis. Guessing at random the ρ fraction of the bits of f' on which we have no information allows us to have the invariant that we compute all the output bits of f', and leads to a simplified analysis.

5 Consequences of our lower bound for pointer jumping

In this section we restate and prove some consequences of our lower bound for pointer jumping. We start with the consequences of our lower bound for general protocols with bounded number of rounds: the rounds hierarchy Theorem 1.1 and the separation of nondeterminism from determinism, Theorem 1.2. Then we prove a separation between the power of different orders in the one-way model, Corollary 1.4. Finally, we give our lower bound for the disjointness function, Corollary 1.5. All these results follow *easily* from our main lower bound for pointer jumping, Theorem 3.2.

We start with the consequences regarding general protocols with a bounded number of rounds. The following corollary combines Theorems 1.1 and 1.2 from the Introduction.

Corollary 5.1 (Separations for a bounded number of rounds)**.** *For every r and k there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that requires $n^{\Omega(1)}$ communication for k-party r-round protocols, but that can be computed with $O(\log n)$ communication by both*

1. *k-party r'-round protocols, whenever $r'(k-1) \geq r \cdot k$, and*
2. *nondeterministic k-party 2-round protocols.*

Proof. Let $k' := k \cdot r = O(1)$, and consider the pointer jumping function $TPJ_{k'} : \{0,1\}^n \to \{0,1\}$. Let us think of the input pointers to this function as arranged in r blocks of k layers each, where the first block cointains the first k layers, the second block the layers between $k+1$ and $2 \cdot k$ and so on. Now consider the following partition of the inputs to the k parties: party i knows all pointers except those leaving the i-th layer in each block. We claim that this function, under this partition of the inputs, gives the separation.

Lower bound: The lower bound follows from our results in the one-way model (Theorem 3.2). Specifically, we use the fact that any r-round k-party protocol for the $TPJ_{k'}$ function above can be simulated by a one-way k'-party protocol, under the partition of the inputs described above (recall that $k' = k \cdot r$). The simulation is natural and works as follows. At any round j, if party i speaks in the original r-round protocol, we simulate this by letting party $(j-1) \cdot k + i$ speak in the one-way model (i. e., party i in the j-th block). Note that, by our choice of the partition of the inputs, this party $(j-1) \cdot k + i$ in the one-way protocol has access to all the information that party i had access to in the original protocol. This shows that the simulation can indeed be carried out and proves the lower bound.

Upper bound (1): Recall that evaluating the function amounts to following k' pointers, i. e., outputting the node reached after following the pointers. Since party i knows all pointers except those in the i-th layer in each block, it is not hard to see that in one round the parties can always follow $k-1$ pointers. Thus, to compute the function they only need r' rounds for $r'(k-1) \geq r \cdot k$. Since at each round they only need to communicate the node reached up to that point, and this takes $O(\log n)$ bits, the upper bound is proved.

Upper bound (2): Using nondeterminism, the first party guesses the piece of the input on its forehead, and announces the whole path in the graph from the starting node to a leaf. Note that all the steps in this path are correct except possibly those corresponding to the first layer in each block, which the first party does not see but guessed. As the graph has $k \cdot r = O(1)$ layers, this takes $O(\log n)$ communication. The second party, seeing the first party's input and the path, can verify the correctness of the steps corresponding to the first layer in each block, and thus check if the leaf announced by the first party is the correct one. □

The function witnessing the separation between nondeterminism and determinism in Theorem 5.1 can in fact be computed with $O(\log n)$ communication both by nondeterministic and by co-nondeterministic protocols. By a result of [AUY83],[6] such functions can always be computed with polylogarithmic communication by deterministic protocols, if we put no bound on the number of rounds. Thus, our exponential separation in Theorem 5.1 provides a sharp contrast in the case of bounded number of rounds.

Another consequence of our results is the following separation between the power of different orders in which the parties speak.

Corollary 1.4 (Order separation)**.** *For any k and any order π for the k parties to speak, there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that requires $n^{\Omega(1)}$ communication if the parties speak in the order π, but can be computed with $O(\log n)$ communication if the parties speak in any other order.*

Proof. Without loss of generality, let us assume that $\pi = (1, 2, \ldots, k)$. Consider the k-party pointer jumping function $TPJ_k : \{0,1\}^n \to \{0,1\}$; by Theorem 3.2, its one-way communication complexity is $n^{1/(k-1)}/k^{O(k)}$ if the parties speak in the order $1, 2, \ldots, k$. For the upper bound, fix a different order, and suppose that party $j > i$ speaks before party i. Then party j can announce the node reached after following the first j pointers, which takes $O(\log n)$ bits. Later, party i can annouce the value of the function. This is because it knows the pointers in every layer except the i-th, and thus in particular the pointers in the layers after the j-th, since $j > i$. □

In the rest of this section we explain how we obtain a lower bound for the disjointness function.

Definition 5.2 (Disjointness function)**.** For integers k, n the disjointness function is $DISJ_{k,n} : (\{0,1\}^n)^k \to \{0,1\}$ where $DISJ(x_1, x_2, \ldots, x_k) = 0$ if and only if there is $j \in [n]$ such that $(x_i)_j = 1$ for every $i \in [k]$, where $(x_i)_j$ denotes the j-th bit of the string $x_i \in \{0,1\}^n$.

Note that $DISJ(x_1, x_2, \ldots, x_k) = 1$ if and only if the k sets encoded by the strings x_i have nonempty intersection.

[6] Although [AUY83] only considers 2-party protocols, their result extends to any k.

Corollary 5.3. *There is a universal constant $c > 0$ such that the following is true. For every $k, d \geq 1$, and for $\eta_k := (k+1)^{-c \cdot k}$, there is a distribution D over the inputs of $DISJ_{k,n}$ such that there is no one-way k-party protocol P achieving*

$$\Pr_{x \in D}[P(x) = DISJ_{k,n}(x)] \geq \exp(-\eta_k)$$

and using communication at most $\eta_k \cdot d$.

In particular, every one-way k-party protocol for the disjointness function $DISJ_{k,n} : (\{0,1\}^n)^k \to \{0,1\}$ must use communication at least $n^{1/(k-1)}/k^{O(k)}$.

Proof. We reduce an instance of the pointer jumping function $TPJ_k^{1,d}$ to an instance of the disjointness function $DISJ_{k,n}$ where $n = d^{k-1}$. The corollary then follows from Theorem 3.2. We now explain this reduction, which amounts to a simple blockwise encoding of the truth-tables of the functions g. Specifically, given a $TPJ_k^{1,d}$ instance $g_1, g_2, \ldots, g_k$ we construct the corresponding $DISJ_{k,n}$ instance $x_1, x_2, \ldots, x_k$ as follows. The input x_k equals the truth-table of the binary function g_k mapping $[d^{k-1}]$ to $\{0,1\}$. To specify x_i for $i < k$, let us think of the integers up to $n = d^{k-1}$ written in d-ary notation, and of g_i as a map from d-ary integers of length $i-1$ into d-ary integers of length i, i. e., $g_i : [d]^{i-1} \to [d]^i$. Now, we set to 1 exactly those bits of x_i whose index is of the form $g_i(a) \circ b$, $a \in [d]^{i-1}, b \in [d]^{k-i}$. For example, for x_1 we set to 1 all bits whose index has the most significant digit equal to $g_1(1)$. It can be verified that for such x_i's we have $TPJ_k^{1,d}(g_1, g_2, \ldots, g_k) = 1 - DISJ_{k,n}(x_1, x_2, \ldots, x_k)$. □

6 Streaming models and branching programs

In this section we discuss some consequences of our lower bounds for *streaming models* of computation. Following Alon, Matias, and Szegedy [AMS99], we consider the problem of approximating, with limited space, the *frequency moments* of a sequence $A = (a_1, \ldots, a_m)$ of m elements $a_i \in \{1, \ldots, n\}$. The paper [AMS99] establishes several bounds on approximating various moments in the model where we have one-way access to the sequence. Our results only apply to the frequency moment F_∞, defined below.

[AMS99] also considered the extension of the one-way model where one allows for *several* one-way passes on the input sequence, but in the same order. This model was also studied by some of the many papers that followed [AMS99], for example in the paper [BYJKS04] which also gives sharper space bounds.

Here, we consider a further extension of the one-way streaming model, where we are allowed to access the sequence in *any* order, as long as we read at most $c \cdot n$ inputs (the order is fixed and independent of the actual sequence). We can think, for example, of approximating the moments of a matrix of n elements, where we first access them in row order, and then in column order. This overlap of information is reminiscent of the multiparty model, and in fact using our results we are able to prove a space lower bound for approximating the moment F_∞, which is simply the maximum over $i \in \{1, \ldots, n\}$ of the number of times that element i appears in the sequence (cf. [AMS99]); for example, $F_\infty = 1$ if all elements in the sequence are distinct.

Corollary 6.1. *For every c there is an $\epsilon > 0$ such that the following holds. Let M be a randomized algorithm that uses space s and, given a sequence A of n elements in $\{1, \ldots, n\}$, reads $c \cdot n$ elements in an arbitrary but fixed order (independent of the sequence). Suppose that M outputs with probability $1 - \epsilon$ a number F_∞^* such that $|F_\infty^* - F_\infty| \leq \epsilon \cdot F_\infty$. Then M must use space $s \geq n^\epsilon$.*

We prove Corollary 6.1 via branching programs, a model of interest in its own right. We note that branching programs are equivalent to the streaming model of Corollary 6.1. Let us start by briefly recalling the definition of a branching program. Because of the application to streaming models, we work with branching programs over n inputs from a non-necessarily Boolean domain D.

Definition 6.2 (Branching program)**.** A *branching program* on n inputs from a domain D is a directed, acyclic graph whose internal nodes are labeled with queries of the form "input i?", whose edges are labeled with $|D|$ possible query answers, and whose leaves are labeled with real numbers. A branching program computes a function from D^n to the reals in the natural way, by following the path from the root to a leaf. The *length* of a branching program is the length of a longest path from the root to a leaf, and the *width* is the maximum, over all d, of the number of nodes at distance d from the root. A branching program is *oblivious* if any path from the root to a leaf queries the same inputs in the same order.

We observe that the streaming algorithm M in Corollary 6.1 can be implemented by an oblivious branching program of length $c \cdot n$ and width 2^s (with inputs over the domain $\{1, \ldots, n\}$).

The next lemma states that branching programs can be simulated by multiparty protocols. This simulation motivated the introduction of the multiparty model [CFL83]. We make the straightforward observation that the simulation also holds for one-way protocols, which allows us to use the lower bounds in this paper to obtain branching program lower bounds.

Proposition 6.3 ([CFL83]; Simulating branching programs by multiparty protocols)**.** *Let $p : D^n \to \Re$ be an oblivious branching program of width w and length $c \cdot n$. For $k = 2c$ and a constant α that depends on c only, there is a partition $H_0, H_1, \ldots, H_k$ of $[n]$ where each set $H_i, i \geq 1$, has size at least $\alpha \cdot n$, and the branching program p can be simulated by a one-way k-party protocol exchanging $(k-1) \cdot \log w + 1$ bits of communication, where party i has the inputs indexed by H_i on its forehead (and the inputs indexed by H_0 are seen by everybody).*

Proof sketch. Divide up the sequence of $c \cdot n$ indices accessed by the program into k blocks of $n/2$ elements each. Start by setting H_i, $i \geq 1$, equal to the complement of indices in the i-th block (clearly this complement has size at least $n/2$). To make the H_i disjoint, apply iteratively the marriage theorem (a.k.a. Hall's theorem). (Alternatively, assign each index to a random set among those that contain it, and argue that with high probability the resulting sets will be large.) Finally, set $H_0 := [n] - \cup_{i \geq 1} H_i$. □

We are now in the position to present the proof of Corollary 6.1.

Proof of Corollary 6.1. The algorithm M can be seen as an oblivious branching program of length $c \cdot n$ and width 2^s (with inputs over the domain $\{1, \ldots, n\}$). By Proposition 6.3, there is a partition $H_0, H_1, \ldots, H_k$ of $[n]$ where $k = 2c$ and each set $H_i, i \geq 1$ has size at least $\alpha \cdot n$, and the branching program can be simulated by a one-way k-party protocol using communication $(k-1) \cdot s + 1$ (on that partition). Similarly to the approach in [AMS99], we now argue that such a protocol can be used to compute the k-party disjointness function $DISJ_{k,\alpha \cdot n} : (\{0,1\}^{\alpha \cdot n})^k \to \{0,1\}$. Given an input $x_1, \ldots, x_k \subseteq [\alpha \cdot n]$ of the disjointness function, we construct the sequence as follows. For every i, we set $|x_i|$ of the elements indexed by H_i to the elements in the input set $x_i \subseteq [\alpha \cdot n]$, in arbitrary order. The remaining $|H_0| + \sum_i |H_i| - |x_i|$ elements of the sequence can be, say, assigned to distinct elements from $\{\alpha \cdot n + 1, \ldots, n\}$. It is easy to see that the moment F_∞ of this sequence

is at most $k-1$ if $DISJ_{k,\alpha\cdot n}(x_1,\ldots,x_k)=1$, and otherwise it is k. Thus, for a sufficiently small constant ϵ depending on k, approximating F_∞ lets us compute disjointness, and by Corollary 5.3 we conclude that $s \geq n^\epsilon$. □

Finally, we mention that we do not know how to get hardness of approximation results for other moments in our model.

7 Open problems

One question, in proof complexity theory, is whether a size lower bound for tree-like threshold systems already follows from our one-way lower bound (cf. Section 1.3 and [BPS07]). This seems to require a "better than known" simulation of such proof systems by multiparty protocols, specifically a simulation that uses only few rounds. Another question is how tight is our one-way lower bound for pointer jumping if k is not constant. A small step in this direction would be understanding whether one can improve Theorem 3.3 so as to have η' linear in η, as opposed to quasi-linear. This would allow to establish lower bounds for pointer jumping of the form $n^{1/(k-1)}/2^{O(k)}$, as opposed to our $n^{1/(k-1)}/k^{O(k)}$.

Acknowledgments. We thank Noam Nisan for helpful conversations, and in particular for suggesting the application to round separation. We are grateful to Oded Regev for pointing out an idea from the paper [BARdW08] which led to a considerable simplification of the proof of our main party-reduction theorem (see Remark 4.3). We are especially grateful to László Babai for extensive comments and pointers to previous work. Finally, we would like to thank Ronald de Wolf and the anonymous referees for helpful comments.

References

[AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. System Sci.*, 58(1, part 2):137–147, 1999. 2, 14, 15

[AUY83] Alfred V. Aho, Jeffrey D. Ullman, and Mihalis Yannakakis. On notions of information transfer in vlsi circuits. In *15th ACM symposium on Theory of computing (STOC)*, pages 133–139. ACM Press, 1983. 13

[BARdW08] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *49th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 477–486. IEEE Computer Society, 2008. 7, 11, 12, 16

[BC08] Joshua Brody and Amit Chakrabarti. Sublinear communication protocols for multiparty pointer jumping and a related lower bound. In *25th Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 145–156, 2008. 1, 2

[BDPW07] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from nondeterministic nof multiparty communication complexity. In *34th*

International Colloquium on Automata, Languages and Programming (ICALP), pages 134–145. Springer, 2007. 2

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. Comput.*, 33(1):137–166, 2003. 1

[BHK01] László Babai, Thomas P. Hayes, and Peter G. Kimmel. The cost of the missing bit: communication complexity with help. *Combinatorica*, 21(4):455–488, 2001. 1, 2, 3, 5, 6

[BHN09] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of AC^0. In *50th Symposium on Foundations of Computer Science (FOCS)*, 2009. To appear. 2, 4

[BIP98] Paul Beame, Russell Impagliazzo, and Toniann Pitassi. Improved depth lower bounds for small distance connectivity. *Comput. Complexity*, 7(4):325–345, 1998. 2

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. 1, 3

[BPS07] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for lovász–schrijver systems and beyond follow from multiparty communication complexity. *SIAM J. Comput.*, 37(3):845–869, 2007. 3, 16

[BPSW05] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and the multiparty nof communication complexity of set disjointness. In *20th Annual Conference on Computational Complexity (CCC)*, pages 52–66. IEEE, 2005. 1, 3, 5

[Bro09] Joshua Brody. The maximum communication complexity of multiparty pointer jumping. In *Proc. 24th Conference on Computational Complexity (CCC)*. IEEE, 2009. 1, 2

[BT94] R. Beigel and J. Tarui. On acc. *Comput. Complexity*, 4(4):350–3–66, 1994. Special issue devoted to the 4th Annual McGill Workshop on Complexity Theory. Preliminary version in FOCS '91. 1

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. System Sci.*, 68(4):702–732, 2004. 2, 14

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity*, Technical Report TR08-002, 2008. 2, 4

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *15th Annual Symposium on Theory of Computing (STOC)*, pages 94–99, 1983. 1, 15

[Cha07a] Amit Chakrabarti. Lower bounds for multi-player pointer jumping. In *Proceedings of the 22nd Annual Conference on Computational Complexity*. IEEE, June 13–16 2007. 1, 2

[Cha07b] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *48th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 449–458. IEEE, 2007. 2

[CT93] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993. 1, 3

[DGS87] Pavol Duris, Zvi Galil, and Georg Schnitger. Lower bounds on communication complexity. *Inf. Comput.*, 73(1):1–22, 1987. 1

[DJS98] Carsten Damm, Stasys Jukna, and Jiří Sgall. Some bounds on multiparty communication complexity of pointer jumping. *Comput. Complexity*, 7(2):109–127, 1998. 1, 2, 5

[DPV08] Matei David, Toniann Pitassi, and Emanuele Viola. Improved separations between nondeterministic and randomized multiparty communication. In *12th Workshop on Randomization and Computation (RANDOM)*. Springer, 2008. To appear in Transactions on Computation Theory. 2

[Gro06] Andre Gronemeier. Nof-multiparty information complexity bounds for pointer jumping. In *31st Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 4162 of *Lecture Notes in Computer Science*, pages 459–470. Springer, 2006. 1, 2

[HG91] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. 1

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997. 1, 7

[LS09] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009. 2, 4

[NBY97] Noam Nisan and Ziv Bar-Yossef. Pointer jumping requires concurrent read. In *29th ACM Symposium on Theory of Computing (STOC)*, pages 549–558 (electronic), 1997. 2

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. Comput.*, 22(1):211–219, 1993. 1, 2, 3, 5

[PS84] Christos H. Papadimitriou and Michael Sipser. Communication complexity. *J. Comput. System Sci.*, 28(2):260–269, 1984. 1, 2

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Comput. Complexity*, 9(2):113–122, 2000. 1, 3

[RW93] Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993. 3

[She08a] Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008. 2

[She08b] Alexander A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *40th Annual Symposium on the Theory of Computing (STOC)*, pages 85–94. ACM, 2008. 2

[She09] Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. 2

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. 3

[VW07] Emanuele Viola and Avi Wigderson. One-way multiparty communication lower bound for pointer jumping with applications. In *48th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2007. 2, 4

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. *Theory of Computing*, 4:137–168, 2008. 1, 3

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *11th ACM symposium on theory of computing (STOC)*, pages 209–213. ACM Press, 1979. 1

[Yao90] A. C. Yao. On acc and threshold circuits. In *Proc. 31st Ann. IEEE Symp. Found. Comput. Sci.*, pages 619–627, 1990. 1