# Simulation Beats Richness: New Data-Structure Lower Bounds

Arkadev Chattopadhyay
Tata Institute of Fundamental Research
Mumbai, India
arkadev.c@tifr.res.in

Michal Koucký
Charles University
Prague, Czech Republic
koucky@iuuk.mff.cuni.cz

Bruno Loff
INESC-Tec and University of Porto
Porto, Portugal
bruno.loff@gmail.com

Sagnik Mukhopadhyay
KTH Royal Institute of Technology
Stockholm, Sweden
sagnik@kth.se

## ABSTRACT

We develop a technique for proving lower bounds in the setting of asymmetric communication, a model that was introduced in the famous works of Miltersen (STOC'94) and Miltersen, Nisan, Safra and Wigderson (STOC'95). At the core of our technique is a novel simulation theorem: Alice gets a $p \times n$ matrix $x$ over $\mathbb{F}_2$ and Bob gets a vector $y \in \mathbb{F}_2^n$. Alice and Bob need to evaluate $f(x \cdot y)$ for a Boolean function $f : \{0,1\}^p \to \{0,1\}$. Our simulation theorems show that a deterministic/randomized communication protocol exists for this problem, with cost $C \cdot n$ for Alice and $C$ for Bob, if and only if there exists a deterministic/randomized *parity decision tree* of cost $\Theta(C)$ for evaluating $f$.

As applications of this technique, we obtain the following results:

(i) The first strong lower-bounds against randomized data-structure schemes for the Vector-Matrix-Vector product problem over $\mathbb{F}_2$. Moreover, our method yields strong lower bounds even when the data-structure scheme has tiny advantage over random guessing.

(ii) The first lower bounds against randomized data-structures schemes for two natural Boolean variants of Orthogonal Vector Counting.

(iii) We construct an asymmetric communication problem and obtain a deterministic lower-bound for it which is provably better than any lower-bound that may be obtained by the classical Richness Method of Miltersen et al.. This seems to be the first known limitation of the Richness Method in the context of proving deterministic lower bounds.

## CCS CONCEPTS

• **Theory of computation → Communication complexity**; **Cell probe models and lower bounds**;

## KEYWORDS

Communication complexity, data structures, lifting theorem, simulation theorem, richness method, vector-matrix-vector product

**ACM Reference Format:**
Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. 2018. Simulation Beats Richness: New Data-Structure Lower Bounds. In *Proceedings of 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC'18).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3188745.3188874

## 1 INTRODUCTION

A central question in theoretical computer science is proving lower bounds on the time needed to solve various algorithmic problems. For general computation this is extremely difficult; indeed, over the past many decades there has been only limited progress in this area despite great effort. One of the main available techniques to prove such lower bounds is the analysis of the flow of information during computation. The area of communication complexity is devoted entirely to the analysis of this information flow.

Data structure problems are computational problems having a well structured form, where information bottlenecks can often be established via communication complexity. In a *static* data structure problem we have a domain $\mathcal{D}$ of possible data, a domain $\mathcal{Q}$ of possible queries and a function $f : \mathcal{D} \times \mathcal{Q} \to \mathcal{A}$ where $f(x;y)$ represents the answer to query $y$ on data $x$. The goal is to store the data $x$ in memory, using space as efficiently as possible, so that given a query $y$ we can evaluate $f(x;y)$ quickly.[1] A major theme of research is to understand the space-query tradeoffs inherent to such problems.

This paper explores this theme in data structures with problems related to matrix-vector multiplication. In the vector-matrix-vector problem $\mathsf{VMV}_{n\times n}$, the data are matrices $x \in \mathbb{F}^{n^2}$ over some field $\mathbb{F}$, queries are pairs of vectors $(q,y) \in \mathbb{F}^n \times \mathbb{F}^n$, and the solicited answers are $f(x;(q,y)) = q \cdot x \cdot y$. In the orthogonal vector counting problem $\mathsf{OVC}_{n\times n}$, the data is also a matrix $x \in \mathbb{F}^{n^2}$, the query is a single vector $y \in \mathbb{F}^n$ and $f(x;y)$ counts the number of zeros in $x \cdot y$, i.e., the number of rows of $x$ which are orthogonal to $y$; we will actually consider two different variants of OVC which have a 1-bit output. The mod-3 orthogonal vector counting $\mathsf{OVC}^3_{n\times n}$ is a variant of $\mathsf{OVC}_{n\times n}$ where $f(x;y) = 1$ if the number of rows of $x$ which are orthogonal to $y$ is a multiple of 3, and $f(x;y) = 0$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*STOC'18, June 25–29, 2018, Los Angeles, CA, USA*
© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.
ACM ISBN 978-1-4503-5559-9/18/06...$15.00
https://doi.org/10.1145/3188745.3188874

[1] In dynamic data structures we also allow certain updates to the data $x$. In this work, we will only be concerned with static data structure problems.

otherwise. The orthogonal gap-majority problem $\mathsf{OGMaj}_{n\times n}$ is a promise variant of $\mathsf{OVC}_{n\times n}$, where we have $f(x;y) = 1$ if at least $\frac{n}{2} + \sqrt{n}$ of the rows of $x$ are orthogonal to $y$, and $f(x;y) = 0$ if no more than $\frac{n}{2} - \sqrt{n}$ of the rows of $x$ are orthogonal to $y$, with the promise that we are in one of the two cases.

We are interested in the complexity of these data-structure problems in Yao's *cell-probe model* [54]. In this model the data is represented in a memory consisting of $s$ cells, each cell storing $w$ bits. We do not charge for the preprocessing time to create the data structure in memory for given $x$, but we charge for the time to answer a query $y$. The cost of the query is the number of memory cells we have to read (probe) in order to answer the query. This model is one of the most general data structure models; in particular, any lower bound on the number of probes to answer a query immediately translates into a lower bound on the time to answer a query in models such as the word-RAM.

The problems we study are closely related to previous work on matrix-vector product. Henzinger et al. [24], and Larsen and Williams [31] study the matrix-vector product and the vector-matrix-vector product over the *Boolean semiring*, in its relation to fine-grained complexity and conditional lower bounds. In particular, Henzinger et al. conjecture that there are no *truly subcubic* algorithms to solve the online version of matrix-vector multiplication (OMV). Assuming this conjecture, they are able to establish tight lower bounds for over a dozen different dynamic problems, establishing the central importance that OMV enjoys in this area. Indeed, unconditional lower bounds for some versions of matrix-vector multiplication have been recently established. Frandsen et al. [17] study the matrix-vector multiplication over finite fields, and give a lower bound $\Omega(\min\{\frac{n\log|\mathbb{F}|}{\log s}, n^2\})$ on the number of cell-probes for deterministic data structures, where $|\mathbb{F}|$ is the field size. Clifford, Grønlund and Larsen [14] improved this to $\Omega(\min\{\frac{n\log|\mathbb{F}|}{\log(s/n^2)}, n^2\})$ in the randomized setting even with error $1 - |\mathbb{F}|^{n/4}$ for fields of size $|\mathbb{F}| = n^{\Omega(1)}$ and $w = \Theta(\log|\mathbb{F}|)$.

Interestingly, while there exist several hardness results for different versions of matrix-vector multiplication problem, before this work there were no strong hardness results known for the VMV problem over the field $\mathbb{F}_2$, even though it is a natural variant of the matrix-vector product, which has been well-understood even since the seminal paper of Miltersen et al. [34].

And, indeed, it is not clear why the hardness of matrix-vector product (MVP) should carry over to VMV unabated. For example, while Larsen and Williams [31] give a surprising data structure for VMV over the Boolean semiring which uses only $O(n^{3/2}/\sqrt{w})$ cell probes to answer a query, their upper-bounds for MVP in the same setting require a larger number $O(n^{7/4}/\sqrt{w})$ of cell probes.

The problem of counting orthogonal vectors has been widely studied in the context of fine-grained complexity [9, 16, 50], although in that setting the dimension of the input vectors is much smaller than the number of vectors, and these two are comparable in our setting.

## 1.1 Data-Structure Lower-Bounds

We study the VMV, OVC and OGMaj problems over the field $\mathbb{F}_2 = \mathrm{GF}[2]$. We establish the following new lower-bounds against randomized data-structure schemes:

**Theorem I.** *There exists a real constant $\varepsilon > 0$ such that:*

(a) *Any randomized data-structure scheme for $\mathsf{VMV}_{n\times n}$ that uses $s$ cells, each storing $w \le n$ bits, must either make $t \ge \frac{\varepsilon n}{\log\frac{sw}{n}}$ probes, or have success probability $\rho \le \frac{1}{2} + 2^{-\varepsilon n}$.*

(b) *Any randomized data-structure scheme for $\mathsf{OVC}^3_{n\times n}$ that uses $s$ cells, each storing $w \le n$ bits, must either make $t \ge \frac{\varepsilon n}{\log\frac{sw}{n}}$ probes, or have success probability $\rho \le \frac{2}{3} + 2^{-\varepsilon n}$.* [2]

(c) *Any randomized data-structure scheme for $\mathsf{OGMaj}_{n\times n}$ that uses $s$ cells, each storing $w \le n$ bits, must either make $t \ge \frac{\varepsilon n}{\log\frac{sw}{n}}$ probes, or have success probability $\rho \le 1 - \varepsilon$.*

The above lower bounds are optimal when the cell size is $w = n$, as each data-structure problem above has a deterministic scheme using $s = \frac{n}{\log n}$ cells of $w = n$ bits, where the number of probes required to solve a given query is only $O(\frac{n}{\log n})$ [4, 49]. Such a large word size may naturally occur in settings such as external memory models, although it is not really seen in the usual scenario where we want to implement the data structure on a random-access machine.

Intuitively, one would guess that the true complexity of the VMV problem is actually $\frac{n^2}{w\log s}$. However, it is a major open problem in the field of data structures to prove a lower-bound for *any* static data structure problem where the number of probes is shown to be $\omega(\log|Q|)$. We *do not* solve that open problem in this paper. Indeed, it is well known that any purely communication complexity based approach, such as ours and most past techniques, is doomed to give bounds at best $\Theta(\log|Q|)$. What we do develop is a novel general technique for establishing strong lower bounds, that are also the best possible using communication complexity method alone, for natural 1-bit output problems based on matrix-vector multiplication. Previous techniques do not seem to yield such bounds for this important class of problems.

In their seminal paper, Miltersen et al. [34] study the span problem, where we need to store a vector space $V$ and decide, given a query $y$, whether $y \in V$. This is equivalent to determining whether $x\cdot y = 0$ if the matrix $x$ is chosen so that $V = \ker x$; i.e. we want to know if the number of rows of $x$ orthogonal to $y$ is $n$, or not. For this problem, [34] show lower bound on the number of probes, similar to our own, but in the randomized setting with just one-sided constant error — the data-structure scheme is allowed to err only when $x \cdot y = 0$.

When first thinking about the VMV problem, one soon realizes that there is a one-sided error randomized reduction from the span problem to the VMV problem, and that this might be enough to give us a one-sided error lower-bound to the VMV problem.[3] But, it turns out, the error of the reduction is on the wrong side, and

[2] Note that the success probability of $\frac{1}{2}$ is achievable by random guessing, so this probability bound is not optimal.

[3] The reduction is simple and works in the communication setting: in order to know if $x \cdot y = 0$, Alice and Bob use a protocol for VMV to compute $q \cdot x \cdot y$ for a shared random vector $q$; if $x \cdot y = 0$, then also $q \cdot x \cdot y = 0$, but if $x \cdot y \neq 0$, then $q \cdot x \cdot y = 1$ with probability exactly $\frac{1}{2}$ over the choice of $q$. Alas, the reduction may err precisely in the case when $x \cdot y = 1$, so the lower-bound of [34] does not apply.

this does not allow us to derive any lower-bound for VMV from the [34] lower-bound for the span problem. To our knowledge, our randomized lower-bound for VMV is also the first *deterministic* lower-bound for VMV.

Note that this one-sided error lower-bound of Miltersen et al. for the span problem immediately implies the same lower-bound for the OVC problem (although not for the OGMaj problem); however, it can be shown that there is a *two-sided error* randomized data-structure scheme for the span problem where the number of probes is $O(\frac{n}{w})$, and this implies that our randomized lower-bounds for OVC cannot possibly work for the span problem.

☞ Rather remarkably, this difference between the span problem and counting orthogonal vectors may be explained by the fact that the randomized parity decision-tree complexity of the (negated) Boolean OR function is $O(1)$, but is $\Omega(n)$ for the mod-3 function. To understand why this is relevant, we need to make a detour into asymmetric communication complexity, and explain how Theorem I is proven.

## 1.2 Our Tool: An Asymmetric Simulation Theorem

To prove our data-structure lower bounds of Theorem I, we develop a technique of independent interest for proving lower bounds on asymmetric communication complexity. The *asymmetric* setting is distinguished from the *usual* setting of two-party communication complexity by the following:

- One player's input is much larger than the other player's.
- The two players have different communication budgets, so we may talk about $[a, b]$-protocols where Alice communicates $\le a$ bits and Bob communicates $\le b$ bits. Typically the player with the large input has a higher budget.
- Only one of the players needs to learn the output, typically the player with the smaller input. This makes a difference, for example, when the task is to compute a function with an output which is larger than the communication budget.

Asymmetric communication complexity was introduced explicitly by Miltersen [33],[4] and later studied more systematically in the work of Miltersen et al. [34]. In both these works, it was also shown that a lower-bound for a communication problem in this setting implies a similar lower-bound for the corresponding data-structure problem. All our lower-bounds are based on this relationship. While asymmetric communication complexity was primarily motivated by its application to proving lower bounds for data-structures [26, 35–37] and streaming algorithms [5, 51], it is indeed a communication model of independent interest (see for example [38]). Despite the significant interest, there were very few general techniques developed for proving lower bounds in this model. Two such techniques appeared in the original work of Miltersen et al. The first is the Richness Method for primarily proving deterministic and randomized, one-sided-error lower-bounds. The second is the round-elimination technique for two-sided error protocols, that gives strong bounds only when the number of rounds involved is quite limited. Other techniques developed are more ingenious and problem specific, like the tour de force of Patrascu [35] for proving strong bounds on lopsided Disjointness. In this work, we develop a novel and reasonably widely applicable technique that yields strong lower bounds for randomized complexity even with unrestricted number of rounds of communication. Moreover, we exhibit a function for which our technique provides strong deterministic lower bounds that the Richness Method provably cannot yield.

Our technique is based on a recent trend seen in *symmetric* communication complexity, of proving *lifting theorems*, sometimes known as *simulation theorems*. Such theorems show, for some carefully chosen two-player function $g(x; y)$, called the *gadget*, that the communication complexity of a composed function $f \circ g = f(g(x_1; y_1), \ldots, g(x_p; y_p))$, under some setting, is proportional to a corresponding measure of complexity on $f$ multiplied by the communication complexity of $g$.

For example, in the paper [21], building on the work of [39], the authors have shown that — taking the gadget $g$ to be the indexing function — the deterministic communication complexity of $f \circ g$ equals, up to constant factors, to the deterministic query-complexity of $f$ times $\log n$, and used this to show a separation between the deterministic communication complexity and the partition number, which was a longstanding open problem at the time. This result was improved in a recent work of the authors [12], and independently by [52]. Lifting theorems, by now, have numerous other applications, such as monotone-circuit lower-bounds [20, 27, 29, 39, 41, 45], small-depth circuit lower-bounds [10, 43], proof-complexity lower-bounds [6, 25], and separations of complexity classes in communication complexity [15, 19, 21, 22]. Many of these developments have happened recently and indeed, in FOCS 2017, a workshop [32] was devoted entirely to such results and their applications.

In this work, we prove two simulation theorems — a deterministic simulation theorem and a randomized simulation theorem. Our gadget is the matrix-vector product ($\mathsf{MVP}_{p\times n}$), so Alice gets a $p \times n$ matrix $x$, and Bob gets a single $n$-bit vector $y$, and we ask them to compute $F(x; y) = f \circ \mathsf{MVP}_{p\times n}(x, y) = f(x \cdot y)$, where $f$ is a function of $p$ bits.[5]

It is easy to see that this can be done with $O(d \cdot n)$ bits of communication from Alice, and $O(d)$ bits from Bob, where $d$ is the smallest depth of a parity decision-tree (PDT) for $f$. If the PDT is randomized, we get a randomized protocol, if the PDT is deterministic, we get a deterministic protocol. To simulate a parity query $q \cdot (x \cdot y)$, Alice sends $q \cdot x \in \{0, 1\}^n$ to Bob, and Bob then replies with $(q \cdot x) \cdot y \in \{0, 1\}$.

Our simulation theorems show that this relatively naive protocol is, indeed, optimal up to constant factors.

**Theorem II** (Main Tool). *Let $n, p \le m = \frac{n}{1000}$ and $C < \frac{m}{100}$ be natural numbers and let $f : \{0, 1\}^p \to \mathcal{Z}$ be an arbitrary (possibly*

[4] However the notion appears implicitly in earlier work [1, 53].

[5] Lifting theorems are generally proven for a symmetrically composed function $f \circ g^p$ which is defined as $f \circ g^p(x_1, \ldots, x_p; y_1, \ldots, y_p) = f(g(x_1, y_1), \ldots, g(x_p, y_p))$. The matrix-vector product can bee seen as an asymmetric composition, i.e. $f \circ g^{p\times 1}$, defined as $f \circ g^{p\times 1}(x_1, \ldots, x_p; y) = f(g(x_1, y), \ldots, g(x_p, y))$, where $g$ is the inner-product function. This is more subtle because in the asymmetric composition case, all the $x$'s participate with the same $y$. Although previous lifting theorems have been proven with asymmetric budgets [e.g. 27], ours is the first lifting theorem to work with an asymmetric composition.

*partial) function. Consider communication protocols where Alice gets an input $x \in \{0,1\}^{p\times n}$ and Bob gets an input $y \in \{0,1\}^n$.*

*(a) If there exists a deterministic two-player $[C \cdot n, C]$-protocol for computing $f \circ \mathsf{MVP}_{p\times n}(x,y)$, then there exists a deterministic parity decision-tree which on input $z$ outputs $f(z)$, and makes $\leq 40 \cdot C$ parity queries to $z$.*

*(b) If there exists a randomized two-player $[C \cdot n, C]$-protocol for computing $f \circ \mathsf{MVP}_{p\times n}(x,y)$ with success probability $\rho$, then there exists a randomized parity decision-tree which on input $z$ outputs $f(z)$ with success probability $\geq \rho - 2^{-m}$, and makes $\leq 200 \cdot C$ parity queries to $z$.*

A few remarks are in order. First, Theorem II is also the first instance of *any* simulation theorem extracting a randomized PDT from a randomized communication protocol. Second, for deterministic protocols in the symmetric two party and multiparty settings for XOR functions, Hatami et al. [23] and Yao [55] do prove theorems lifting parity decision-tree complexity. But both results incur polynomial loss in the process of lifting. To the best of our knowledge, Theorem II *is* the first lifting theorem that characterizes parity decision-tree complexity so tightly — up to constant factors. On the other hand, the gadget size in [23, 55] are constant whereas our gadgets are polynomially large w.r.t to the arity of the outer function $f$. Obtaining such tight simulation theorems, w.r.t. decision tree complexity measures in general as in Theorem II, with constant gadget size is a fundamental open problem in communication complexity.

☞ We will then prove the data-structure lower-bounds (b) and (c) of Theorem I by showing lower-bounds against randomized parity decision-trees. We will show that the randomized parity decision-tree complexity of the mod-3 function is high, and it easily follows from the work of [8, 44, 47] that the randomized parity decision-tree complexity of gap-majority is high as well. However, the randomized PDT complexity of (negated) OR is $O(1)$, which is what prevents our lower-bound from applying to the span problem mentioned above.

The lower-bound for the VMV problem — Theorem I (a) — does not directly follow from the above simulation theorems. Instead, it is proven by a *simulation-type* argument: one shows that a short protocol for the VMV problem would give us a parity decision-tree for solving a certain task, and then show that this task cannot be solved efficiently.

The proof of our simulation theorems is inspired by several previous works, most notably the recent work of Göös, Pitasi and Watson [22]. However the peculiarities of the asymmetric setting call for substantial development of more ideas. In particular, we make use of a novel notion, which we call *linear* min-entropy, and of a variant thereof, which we call *smooth* linear min-entropy. We believe these two notions are interesting in their own right, and should find other uses. Implementing the simulation theorems using these notions requires delicate technical work. These are the main technical contributions of this submission.

### 1.3 Beating the Richness Method

The Miltersen et al. [34] paper presented two techniques for proving lower-bounds in the asymmetric settings — the richness technique, and the round-elimination technique [see also 42]. Pătraşcu and Thorup [37] later proved a direct-sum theorem for the Richness technique.

The round-elimination technique method only works in situations where the number of rounds is small — typically sub-logarithmic. To the authors' knowledge, the Richness technique and its extension in [37] is essentially the only general method known for proving deterministic unbounded-round lower-bounds in the asymmetric setting. Even those lower-bounds which are proven in the two-sided error randomized asymmetric setting — lower-bounds such as [35], which cannot be shown by the richness technique because it is limited to proving one-sided error lower-bounds — the same lower-bound (up to constant factors) can be shown in the deterministic setting using the richness technique.

Given this state of affairs, it would be tempting to think, for example, that a deterministic (or one-sided error) lower-bound for the VMV problem might exist which completely circumvents our approach based on simulation theorems. However, this might actually not be the case: we show that, at least in some situations, our simulation theorem proves a deterministic lower-bound which cannot be proven by the richness technique of Bro Miltersen et al.:

**Theorem III.** *There exists a promise problem $F : \{0,1\}^{p\times n} \times \{0,1\}^n \to \{0,1\}$ such that:*

- *Theorem II (a) implies that any deterministic $[a,b]$-protocol for $F$ has $a = \Omega(n^2)$ or $b = \Omega(n)$;*
- *However, $F$ has a randomized zero-error $[O(n), O(1)]$-protocol.*

Since any lower-bound proven by the richness technique also gives a lower-bound against randomized protocols with one-sided error (and thus zero-error; and this consideration applies to [37], as well), it follows that the above lower-bound cannot be proven via the richness method, or its extension in [37] — it is the first known lower-bound in deterministic (unbounded-round) asymmetric communication complexity for which this is the case.

## 2 OVERVIEW OF OUR TECHNIQUES

All our lower-bounds follow from the well-known connection between data structures and communication complexity, which first explicitly appeared in [33]: if we have a data-structure scheme for $f(x;y)$, then we obtain a protocol for the communication problem where Alice gets the data $x$, Bob gets the query $y$, and they must communicate to compute $f(x;y)$. Hence we will prove the lower-bounds for data structures of Theorem I, by proving lower-bounds for asymmetric communication problems.

In turn, our communication complexity lower-bounds are all shown by first proving a lower-bound against parity decision-trees, and then *lifting* these lower-bounds to communication complexity, by use of Theorem II, which is the main technical contribution of this paper. We will thus begin by sketching the proof of Theorem II in Section 2.1; we then sketch the proofs of the data-structure lower-bounds in Section 2.2. We made an effort to include the full proof of at least one theorem within the 10-page limit. We opted for Theorem III, whose full proof appears in Section 2.3.

## 2.1 Proving Theorem II

To explain how we prove our simulation theorems, it is worthwhile to give a general overview of how previous simulation theorems have been proven — the discussion broadly applies to all of [3, 12, 18, 19, 21, 39, 48, 52] and [22].

We are given a protocol for a composed function $f \circ g$ — $g$ takes a pair $(x, y)$ of inputs and produces a $p$-bit string, which is then fed to $f$. We wish to construct a decision-tree for computing $f(z)$ when given query access to $z$. The general strategy is to find a leaf in the protocol tree where $z$ is represented, meaning that the rectangle $A \times B$ associated with said leaf is such that $z \in g(A \times B)$; this way, we may output the label which the protocol assigns to that rectangle, and it should equal $f(z)$. In the randomized case we will actually want a specific distribution on such rectangles, but let's set that aside for now.

So we go down the protocol tree, keeping in mind a rectangle $A \times B$. As long as we haven't queried $z$, we need to make sure that *every* $z$ is represented in $g(A \times B)$; once we have made *some* queries to $z$, then every $z'$ which is consistent with those queries must be represented in $g(A \times B)$.

If the gadget $g$ is well-chosen, it becomes feasible to enforce this invariant. For example, if $g = (\mathsf{IP}_n)^p$ is the $p$-fold inner-product of $n$-bit strings,[6] there are two known properties which, if true of $A$ and $B$ both, ensure that every $z$ is represented in $g(A \times B)$ — one such property is called *thickness* and is used in [3, 12, 21, 39], and another is called *density*, and is used in [18, 19, 22, 48]. It is worthwhile to briefly review these notions.

For $\delta \in [0, 1]$, a set $A \subseteq \{0, 1\}^{p \times n}$ is called $\delta$*-thick*, if for every $a = (a_1, \ldots, a_p) \in A$ and every $i \in [p]$, there exist $\geq 2^{\delta n}$-many different $a'_i$ such that $(a_1, \ldots, a_{i-1}, a'_i, a_{i+1}, \ldots, a_p) \in A$; $A$ is called $\delta$*-dense*, if for every $I \subseteq [p]$ of size $|I| = k \geq 1$, the distribution $(\mathbf{x})_I$, obtained by picking a uniformly-random $\mathbf{x} \in A$ and projecting onto the coordinates in $I$, has min-entropy $\geq \delta\ kn$.

The *thickness* of $A$ is then the largest $\delta$ for which it is $\delta$-thick, and the *density* of $A$ is the largest $\delta$ for which it is $\delta$-dense. We may also say that $A$ is $\delta$-thick or $\delta$-dense with respect to a set $S \subseteq [p]$ of coordinates, if we replace $[p]$ with $S$ in the above definitions.

In order to find the desired leaf in the protocol tree, and thus prove the simulation theorem, the decision tree goes down the protocol tree while being careful to preserve one such property (density or thickness) as an invariant. As the rectangle becomes smaller, and we are at risk of loosing our invariant, we must have a means of restoring it by querying some coordinates of $z$. We then focus only on those inputs $(x, y)$ such that $g(x, y)$ is consistent with the outcome of these queries, and it is important that our property (e.g. thickness or density) still holds with respect to those coordinates which we did not query yet.

All of the simulation theorems just mentioned follow this general pattern, and so do the simulation theorems proven in this paper. But, even after having a good understanding of this general framework, it is not *apriori* clear how to proceed when the inner gadget is the matrix-vector product, nor how to connect such results to the vector-matrix-vector problem which is not itself a composed function.

Let $g$ be the matrix-vector product over $\mathbb{F}_2$, so that $g(x, y) = x \cdot y$ where $x$ is a $p \times n$ matrix and $y$ is an $n$-bit vector. The first thing to observe, when using $g$ as a gadget, is that if Alice has a matrix $x \in \{0, 1\}^{p \times n}$ and Bob a vector $y \in \{0, 1\}^n$, then they are able to make a "parity query" to $g(x, y) = x \cdot y \in \{0, 1\}^p$ by having Alice send only $n$ bits and Bob send only 1 bit: to compute $q \cdot x \cdot y$, Alice sends over $q \cdot x \in \{0, 1\}^n$, and then Bob computes and returns $(q \cdot x) \cdot y \in \{0, 1\}$. So it follows that the communication complexity of $f(x \cdot y)$ is upper-bounded by the randomized parity decision-tree complexity of $f$.

This seems to make the properties of density and thickness unsuitable for carrying out the above strategy. Indeed, it is easy to construct, for example, a dense set $A$ such that $g(A \times \{0, 1\}^n)$ is missing some vectors — indeed, if $A$ is the set of matrices such that the bitwise XOR of all the rows is the zero vector, then every $g(A \times \{0, 1\}^n)$ is missing all vectors with odd Hamming weight. On the other hand, thickness is a property which is difficult to preserve, and it would seem that if we were able to preserve this property as an invariant in our construction, we would obtain a simulation theorem for normal decision trees, not parity decision trees. However, we cannot obtain such a simulation from the above protocol for making a "parity query" to $x \cdot y$. So we had to devise a different property for our invariant. We call it *linear min-entropy*.

**Definition.** The *linear min-entropy* of a set $A$ of $p \times n$ matrices is the maximum $\eta \in [0, 1]$ such that, for every $k' \times p$ matrix $Q'$, the distribution $Q' \cdot \mathbf{x}$ — obtained by picking a uniformly random $\mathbf{x}$ in $A$, and then outputting the product $Q' \cdot \mathbf{x} \in \{0, 1\}^{k' \times n}$ — has min-entropy $\geq \eta\ k'n$.

So, in some sense, we require a certain min-entropy from the linear combinations of the rows of a random matrix from $A$. We will also need to look at a variant of this notion, called *smooth linear min-entropy*, which is the maximum linear min-entropy among all subsets $A' \subseteq A$ which preserve all but an exponentially-small fraction of $A$.

As one may see, it is a property stronger than density, as one demands a lower-bound on the min-entropy of *any* linear combination of coordinates, and not just of the coordinates themselves.

It will then happen that if $A$ has linear min-entropy at least $\frac{4}{5}$, say, and $|B| \geq 2^{\frac{9}{10}n}$, then every $z \in \{0, 1\}^p$ is represented in $g(A \times B)$. We will show something even stronger, a result which we call *pruning lemma*: for any such $A$ and $B$, we may remove an exponentially-small fraction of $A$ and $B$, to obtain a subrectangle $A' \times B' \subseteq A \times B$, such that every $z$ appears in every row and column of the $g(A' \times B')$ communication matrix[7] in roughly equal proportion. Meaning every row and every column of the $g(A' \times B')$ communication matrix will be (roughly) equally split among the different $z \in \{0, 1\}^p$.

The pruning lemma is then used to show a result called *entropy-restoring partition*. It can be considered the heart of the proof of the simulation theorems in this paper. This result shows how one

[6] Note that, $g$ here is a function which outputs a $p$-bit string. We maintain this convention through out the paper.

[7] I.e., the matrix with rows indexed by $A'$, columns indexed by $B'$, and with the $(x, y)$ entry equal to $x \cdot y$.

may take a set $A \subseteq \{0,1\}^{p\times n}$, such that the smooth linear min-entropy of $A$ is not too high ($\leq \frac{9}{10}$), but where the linear min-entropy of $A$ is still somewhat high ($\geq \frac{4}{5}$), and partition $A$ into subsets $A^{\dagger}, A_1, A_2, \ldots$, with $A^{\dagger}$ very small, such that in each $A_i$ we have fixed some linear combination of rows (of the matrices in $A_i$), and where each $A_i$ has large linear min-entropy ($\geq \frac{9}{10}$) on the remaining (linearly-independent) linear combinations of the rows. Furthermore, if we have a large set $B \subseteq \{0,1\}^n$ of vectors ($|B| \geq 2^{\frac{9}{10}n}$), we may do this in a way that for each $x \in A_i$, the values of $x \cdot y$, for $y \in B$ are equidistributed among the various possible $z \in \{0,1\}^p$ — this means that when the linear decision-tree queries the $k$ coordinates of $z$ corresponding to the rows which were fixed, $B$ will be cutoff by no more than $2^{-k}$. This is the main technical device which allows us to maintain a rectangle $A\times B$ where $A$ has large linear min-entropy, and $B$ is large, as we go down the protocol tree in our simulation theorem. On its own, the entropy-restoring partition suffices for proving our deterministic simulation theorem — Theorem II (a); in fact, the existence of a single part $A_1$ of the entropy-restoring partition is enough for the deterministic simulation theorem, whereas the randomized simulation theorem needs the full entropy-restoring partition.

To prove the randomized simulation theorem, Theorem II (b), we will use a crucial insight from [22]. Suppose $\bar{\pi}$ is a randomized protocol for $f \circ g$ which is the convex combination of several deterministic protocols $\pi$. Then a good approach to proving a randomized simulation theorem is the following: in order to obtain a decision-tree for $f$, it suffices to be able to approximate, for each deterministic protocol $\pi$, the distribution $\pi^{-1}(z)$ obtained by running $\pi$ on a random input $(x,y)$ such that $g(x,y) = z$. We want to do this by making few queries to $z$, and for this purpose [22] proves a result called the *inverse-marginals lemma*. Our version of this lemma states that if $A$ has large linear min-entropy and $B$ is large, then for any $z \in \{0,1\}^p$, if we choose a uniformly random $(x,y) \in A \times B$ among those such that $x \cdot y = z$, then the $x$-marginal will be close to a uniform distribution on $A$ and the $y$-marginal will be close to a uniform distribution on $B$.

To illustrate how this is used, suppose that we are simulating $\pi$ on a rectangle $A \times B$, and it was Alice's turn to communicate, and she would send bit $b$ when $x \in A_b$ — for the partition $A = A_0 \cup A_1$; then if one were to pick a uniformly-random input in $g^{-1}(z) \cap A \times B$, then Alice would send $b = 0$ with probability roughly $\frac{|A_0|}{|A|}$ and send $b = 1$ with probability roughly $\frac{|A_1|}{|A|}$. This heuristic allows us to construct a randomized parity decision-tree which will produce, on input $z \in \{0,1\}^p$, a transcript of the protocol which is exponentially close, in statistical distance, to the transcript which we would obtain if we had run the protocol on a uniformly-random input from $g^{-1}(z)$ — which is enough to prove Theorem II (b).

## 2.2 The Data-Structure Lower-Bounds

The data-structure lower-bounds (b) and (c) of Theorem I follow from lower-bounds against randomized parity decision-trees, by using Theorem II (b) and the connection between data structures and asymmetric communication complexity.

It is intuitive that counting mod-3 should be hard for parity decision-trees. This is shown by making use of the *polynomial discrepancy lemma* of [10]. The polynomial discrepancy lemma says that the Mod3 function (roughly) equally splits the zero set of any linear form over $\mathbb{F}_2$.[8] This will imply that any randomized parity decision-tree for Mod3 will succeed with probability $\leq \frac{1}{3} + 2^{-\Omega(n)}$. If counting mod-3 is hard, then so is counting in general, which gives us the lower-bound for OVC — Theorem I (b).

By way of binary search we can use a single majority $\log n$ times to count exactly. This would easily give us I (c), but with a $\log n \cdot \log\log n$ factor loss. However it follows from [8, 44, 47] that the randomized parity decision-tree complexity of $\sqrt{n}$-gap-majority is $\Omega(n)$, and this implies Theorem I (c).

The data-structure lower-bound of Theorem I (a) does not seem to follow directly from a lower-bound on randomized PDTs for some function. The VMV problem is quite different to a composed problem — in a composed problem both players know the outer function $f$ and the lower-bound depends on $f$ having large PDT complexity; we may think of the VMV problem as if only Bob knew the outer function — $q$ is a parity which is given as Bob's input. But we can still prove the lower-bound by an interesting analogy: instead of proving a lower-bound for randomized PDTs trying to compute a certain function, we instead prove a lower-bound for randomized PDTs trying to succeed at the following task:

**Lemma** (Impossible task)**.** *Suppose we have a randomized parity decision-tree running in time $t$ which, on every input $z \in \{0,1\}^p$, outputs a pair $(q,b) \in \{0,1\}^p \times \{0,1\}$ such that both:*

- *$q$ is (always) linearly-independent of the set $Q$ of parity queries made, and*
- *with probability $\rho$ over the choice of $q$, we have $q \cdot z = b$.*

*Then either $t \geq p$ or $\rho = 1/2$.*

Then, analogously to the simulation theorem — Theorem II (b) — we prove that any randomized communication protocol for VMV, succeeding with probability $\rho$, would give us a randomized parity decision-tree for the above task, succeeding with probability $\geq \rho - 2^{-\Omega(n)}$. This establishes a lower-bound on the asymmetric randomized communication complexity of VMV, which then gives us Theorem I (a).

## 2.3 A Lower-Bound Beating the Richness Method

The *Richness-Method* of Bro Miltersen et al [34], is a method for proving lower-bounds for the communication complexity of asymmetric problems. It relies on the following definition:

**Definition 2.1** (Richness)**.** A two-player problem $F : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ is said to be $(u,v)$-rich with respect to $z \in \{0,1\}$, if there exists $X \subseteq \mathcal{X}$ with $|X| \geq u$, such that for every $x \in X$ there exists $Y_x \subseteq \mathcal{Y}$ with $|Y| \geq v$, such that $F(x,y) = z$ for every $y \in Y_x$.

The Richness Method then consists of two steps: (a) Show that $F$ is $(u,v)$-rich with respect to some $z \in \{0,1\}$. (b) Show that $F$ does not have any $z$-monochromatic rectangles of size $u' \times v'$, where both $u' \geq u/2^{a+b+2}$ and $v' \geq v/2^{b+2}$. i.e., any such large rectangle must intersect $F^{-1}(1-z)$.

[8] Indeed, the lemma holds for any low-degree polynomial over $\mathbb{F}_2$, not just linear forms, hence the name *polynomial* discrepancy lemma.

It will then follow that $F$ does not have any deterministic $[a, b]$-protocols. But something stronger will then also follow: that $F$ does not have any randomized $[a, b]$-protocols, which are allowed to err whenever $F(x, y) = z$ (for the same $z$ for which the two properties above were shown), but not when $F(x, y) \neq z$. I.e., any lower-bound proven using the richness method will give a one-sided-error lower-bound. This follows from the cellebrated *Richness Lemma*:

**Lemma 2.2** (Richness Lemma [34]). *Let $F$ be a $(u, v)$-rich problem with respect to $z$. If $F$ has a randomized one-sided error $[a, b]$-protocol, erring only on inputs $(x, y) \in F^{-1}(z)$, then there is a $z$-monochromatic rectangle of $F$ of dimensions at least $u/2^{a+b+2} \times v/2^{b+2}$.*

In particular, any lower-bound proven using the richness method also shows a lower-bound for zero-error ("ZPP") protocols. Then our goal is to construct a problem with short zero-error protocols, but for which we can prove a large deterministic lower-bound. We start by showing the following:

**Theorem 2.3.** *There exists a promise problem $Z$, having zero-error randomized query complexity $O(1)$, but deterministic parity decision-tree complexity $\Omega(n)$.*

We may now use Theorem II (a) to lift the deterministic PDT lower-bound to the setting of asymmetric communication complexity:

**Corollary 2.4.** *Any deterministic $[a, b]$-protocol for $Z \circ \mathsf{MVP}_{n\times n}$ has $a = \Omega(n^2)$ or $b = \Omega(n)$, but there is a randomized, zero-error $[a, b]$-protocol for $Z \circ \mathsf{MVP}_{n\times n}$ with $a = O(n)$, $b = O(1)$.*

Since the promise problem $F = Z \circ \mathsf{MVP}_{n\times n}$ has zero-error $[O(n), O(1)]$-protocol, it then follows that the richness method cannot give a lower-bound against $[a, b]$-protocols computing $F$, that achieves $a = \omega(n)$. We thus established Theorem III.

Proof of Theorem 2.3. Let $Z^{-1}(0) \subseteq \{0, 1\}^{2n}$ be the set of binary strings which have $z_i = 0$ whenever $i$ is odd, and $z_i = 1$ for at least $\frac{n}{10}$-many even coordinates $i$. Let $Z^{-1}(1) \subseteq \{0, 1\}^{2n}$ have instead $z_i = 0$ whenever $i$ is *even*, and $z_i = 1$ for at least $\frac{n}{10}$-many *odd* coordinates $i$; let $Z : \{0, 1\}^{2n} \to \{0, 1\}$ be the corresponding promise problem. Also, let $\Delta \subseteq \{0, 1\}^n$ be a binary linear code with distance $\geq \frac{n}{10}$ and constant rate $\rho = \frac{\log|\Delta|}{n} > 0$. E.g. a Justesen code [28].

The upper-bound is trivial: the zero-error algorithm queries a pair $x_{2i}x_{2i+1}$; if it equals 00, the algorithm answers "I don't know", which happens with only constant probability, and otherwise the algorithm knows the answer.

The lower-bound rests on the following:

**Claim 2.5.** *Any vector space $V \subseteq \mathbb{F}_2^{2n}$ disjoint from $Z^{-1}(0)$ or disjoint from $Z^{-1}(1)$ must have codimension $\geq \rho \cdot n$.*

The proof of this claim is akin to the Hamming bound for codes. Let us prove the codimension lower-bound assuming $V$ is disjoint from $Z^{-1}(0)$; the other case is proven in the same way.

Define the set $\Delta_0 \subseteq \{0, 1\}^{2n}$ by placing the bits of the $\Delta$-codewords at the even positions, and setting the odd positions to zero. For $c \in \{0, 1\}^{2n}$, let $B^{(0)}(c) = c + \Delta_0$ be the set of words obtained from $c$ by bitwise-XORing a word from $\Delta_0$.

Suppose we had $B^{(0)}(v') \cap B^{(0)}(v'') \neq \varnothing$ for distinct $v', v'' \in V$, say $v' + \delta' = v'' + \delta''$. Then it would follow that $v \overset{\text{def}}{=} v' - v'' = \delta'' - \delta'$ is both in $V$, since $v' - v''$ is in $V$, and in $\Delta_0$, since $\delta'' - \delta'$ is in $\Delta_0$. But $\Delta_0 \subseteq Z^{-1}(0)$, since the distance of the code $\Delta$ is at least $\frac{n}{10}$. Hence, by contradiction, we must conclude that $B^{(0)}(v')$ and $B^{(0)}(v'')$ are disjoint for every distinct $v', v'' \in V$.

It then holds that $|V| \leq 2^{2n}/|\Delta_0|$ which is $\leq 2^{2n-\rho\, n}$ since the code $\Delta$ has rate $\rho$. So $V$ has co-dimension $\geq \rho n$. This proves the claim.

Now take any deterministic parity decision-tree of depth $t < \rho \cdot n$. Consider what happens when every query $q_1, \ldots, q_t$ is answered 0. Suppose without loss of generality that the parity-decision-tree answers 1. Let $V \subseteq \{0, 1\}^{2n}$ be the subspace defined by the linear equations $q_i \cdot x = 0$. Then $V$ has co-dimension $< \rho n$, and so $V \cap Z^{-1}(0) \neq \varnothing$; but this means that the given tree does not correctly compute $Z$. □

## ACKNOWLEDGMENTS

The authors would like to thank Kasper Green Larsen for a helpful and encouraging exchange, and also Bruno Bauwens and Vladimir Podolsky for their help settling the VMV problem in the Boolean ring (whose solution is to be included in the journal version of the paper).

Arkadev Chattopadhyay was supported by the Ramanujan Fellowship of the DST, India.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 616787. Michal Koucký was partially supported by the Center of Excellence CE-ITI (P202/12/G061 of GA ČR).

Bruno Loff was funded by FCT posdoctoral grant number SFRH/BPD/116010/2016.

A large part of this work done while Sagnik Mukhopadhyay was a graduate student of TIFR and a recipient of a TCS fellowship there.

## REFERENCES

[1] Miklós Ajtai. 1988. A lower bound for finding predecessors in Yao's cell probe model. *Combinatorica* 8, 3 (1988), 235–247. https://doi.org/10.1007/BF02126797

[2] Noga Alon, Oded Goldreich, and Yishay Mansour. 2003. Almost k-wise independence versus k-wise independence. *Inform. Process. Lett.* 88, 3 (2003), 107–110.

[3] Anurag Anshu, Naresh B Goud, Rahul Jain, Srijita Kundu, and Priyanka Mukhopadhyay. 2017. Lifting randomized query complexity to randomized communication complexity. *arXiv:1703.07521* (2017).

[4] V. Z. Arlazarov, E. A. Dinic, M. A. Kronrod, and I. A. Faradzev. 1970. On economical construction of the transitive closure of a directed graph. *Soviet Mathematics Doklady* (1970), 11(5):1209–1210.

[5] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. 2010. Lower Bounds for Sparse Recovery. In *Proceedings of the 21st SODA*. 1190–1197. https://doi.org/10.1137/1.9781611973075.95

[6] Paul Beame, Trinh Huynh, and Toniann Pitassi. 2010. Hardness amplification in proof complexity. In *Proceedings of the 42nd STOC*. 87–96.

[7] Mark Braverman. 2011. Poly-logarithmic independence fools bounded-depth boolean circuits. *Commun. ACM* 54, 4 (2011), 108–115. https://doi.org/10.1145/1924421.1924446

[8] Amit Chakrabarti and Oded Regev. 2012. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput.* 41, 5 (2012), 1299–1317.

[9] Timothy M. Chan and Ryan Williams. 2016. Deterministic APSP, orthogonal vectors, and more: Quickly derandomizing razborov-smolensky. In *Proceedings of the 27th SODA*. 1246–1255.

[10] Arkadev Chattopadhyay. 2007. Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits. In *Proceedings of the 48th FOCS*. 449–458. https://doi.org/10.1109/FOCS.2007.30

[11] Arkadev Chattopadhyay. 2008. *Ciruits, Communication and Polynomials*. Ph.D. Dissertation. McGill University.

[12] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. 2017. Simulation Theorems via Pseudorandom Properties. *arXiv:1704.06807* (2017).

[13] Benny Chor and Oded Goldreich. 1988. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.* 17, 2 (1988), 230–261.

[14] Raphael Clifford, Allan Grønlund, and Kasper Green Larsen. 2015. New unconditional hardness results for dynamic and online problems. In *Proceedings of the 56th FOCS*. 1089–1107.

[15] Matei David, Toniann Pitassi, and Emanuele Viola. 2009. Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory* 1, 2 (2009).

[16] Holger Dell and John Lapinskas. 2017. Fine-grained reductions from approximate counting to decision. *CoRR* abs/1707.04609 (2017). arXiv:1707.04609 http://arxiv.org/abs/1707.04609

[17] Gudmund Skovbjerg Frandsen, Johan P Hansen, and Peter Bro Miltersen. 2001. Lower bounds for dynamic algebraic problems. *Information and Computation* 171, 2 (2001), 333–349.

[18] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. 2017. Query-to-communication Lifting for $\mathsf{P}^{\mathsf{NP}}$. In *Proceedings of the 32nd CCC*.

[19] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. 2015. Rectangles are nonnegative juntas. In *Proceedings of the 47th STOC*. 257–266.

[20] Mika Göös and Toniann Pitassi. 2014. Communication lower bounds via critical block sensitivity. In *Proceedings of the 46th STOC*. 847–856.

[21] Mika Göös, Toniann Pitassi, and Thomas Watson. 2015. Deterministic communication vs. partition number. In *Proceedings of the 56th FOCS*.

[22] Mika Göös, Toniann Pitassi, and Thomas Watson. 2017. Query-to-Communication Lifting for BPP. In *Proceedings of the 58th FOCS*.

[23] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. 2016. Structure of protocols for XOR functions. In *Proceedings of the 57th FOCS*. 282–288.

[24] Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. 2015. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings of the 47th STOC*. 21–30.

[25] Trinh Huynh and Jakob Nordstrom. 2012. On the virtue of succinct proofs: Amplifying communication complexity hardness to time-space trade-offs in proof complexity. In *Proceedings of the 44th STOC*. 233–248.

[26] T. S. Jayram, Subhash Khot, Ravi Kumar, and Yuval Rabani. 2004. Cell-probe lower bounds for the partial match problem. *J. Comput. Syst. Sci.* 69, 3 (2004), 435–447. https://doi.org/10.1016/j.jcss.2004.04.006

[27] Jan Johannsen. 2001. Depth Lower Bounds for Monotone Semi-Unbounded Fan-in Circuits. *ITA* 35, 3 (2001), 277–286. https://doi.org/10.1051/ita:2001120

[28] Jørn Justesen. 1972. Class of constructive asymptotically good algebraic codes. *IEEE Transactions on Information Theory* 18, 5 (1972), 652–656.

[29] Mauricio Karchmer and Avi Wigderson. 1990. Monotone Circuits for Connectivity Require Super-Logarithmic Depth. *SIAM J. Discrete Math.* 3, 2 (1990), 255–265. https://doi.org/10.1137/0403021

[30] Eyal Kushilevitz and Noam Nisan. 1997. *Communication Complexity*. Cambridge University Press, New York, NY, USA. 0.

[31] Kasper Green Larsen and Ryan Williams. 2017. Faster online matrix-vector multiplication. In *Proceedings of the 28th SODA*. 2182–2189.

[32] Raghu Meka and Toniann Pitassi (Eds.). 2017. *Hardness Escalation in Communication Complexity and Query Complexity, Workshop at 58th FOCS*. https://raghumeka.github.io/workshop.html

[33] Peter Bro Miltersen. 1994. Lower Bounds for Union-Split-Find related problems on random access machines. In *Proceedings of the 26th STOC*. 625–634. https://doi.org/10.1145/301250.301330

[34] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. 1998. On Data Structures and Asymmetric Communication Complexity. *J. Comput. System Sci.* 57, 1 (1998), 37–49. http://dx.doi.org/10.1006/jcss.1998.1577

[35] Mihai Pătraşcu. 2011. Unifying the Landscape of Cell-Probe Lower Bounds. *SIAM J. Comput.* 40, 3 (2011), 827–847. https://doi.org/10.1137/09075336X

[36] Mihai Pătraşcu and Mikkel Thorup. 2006. Time-space trade-offs for predecessor search. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*. 232–240. https://doi.org/10.1145/1132516.1132551

[37] Mihai Pătraşcu and Mikkel Thorup. 2009. Higher Lower Bounds for Near-Neighbor and Further Rich Problems. *SIAM J. Comput.* 39, 2 (2009), 730–741. https://doi.org/10.1137/070684859

[38] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. 2015. How to compress asymmetric communication. In *Proceedings of the 30th CCC*. 102–123.

[39] Ran Raz and Pierre McKenzie. 1999. Separation of the Monotone NC Hierarchy. *Combinatorica* 19, 3 (1999), 403–435.

[40] Renato Renner and Stefan Wolf. 2004. Smooth Rényi entropy and applications. In *International Symposium on Information Theory*. IEEE, 233.

[41] Robert Robere, Toniann Pitassi, Benjamin Rossman, and Stephen A Cook. 2016. Exponential lower bounds for monotone span programs. In *Proceedings of the 57th FOCS*. 406–415.

[42] Pranab Sen and Srinivasan Venkatesh. 2008. Lower bounds for predecessor searching in the cell probe model. *J. Comput. System Sci.* 74, 3 (2008), 364–385.

[43] Alexander A. Sherstov. 2009. Separating AC0 from Depth-2 Majority Circuits. *SIAM J. Comput.* 38, 6 (2009), 2113–2129. https://doi.org/10.1137/08071421X

[44] Alexander A. Sherstov. 2012. The Communication Complexity of Gap Hamming Distance. *Theory of Computing* 8, 1 (2012), 197–208.

[45] Dmitry Sokolov. 2017. Dag-like communication and its applications. In *International Computer Science Symposium in Russia*. 294–307.

[46] Umesh Vazirani. 1986. *Randomness, Adverseries and Computation*. Ph.D. Dissertation. University of California, Berkeley.

[47] Thomas Vidick. 2012. A concentration inequality for the overlap of a vector on a large set. *Chicago Journal OF Theoretical Computer Science* 1 (2012), 1–12.

[48] Thomas Watson. 2017. A $\mathsf{ZPP}^{\mathsf{NP}}$ Lifting Theorem. *Unpublished preprint* (2017).

[49] Ryan Williams. 2007. Matrix-vector multiplication in sub-quadratic time (some preprocessing required). In *Proceedings of the SODA*, Vol. 7. 995–1001.

[50] Ryan Williams and Huacheng Yu. 2014. Finding orthogonal vectors in discrete structures. In *Proceedings of the 25th SODA*. 1867–1877.

[51] David P. Woodruff. 2014. Sketching as a Tool for Numerical Linear Algebra. *Foundations and Trends in Theoretical Computer Science* 10, 1-2 (2014), 1–157. https://doi.org/10.1561/0400000060

[52] Xiaodi Wu, Penghui Yao, and Henry S. Yuen. 2017. Raz-McKenzie simulation with the inner product gadget. *Electronic Colloquium on Computational Complexity (ECCC)* 24 (2017), 10. https://eccc.weizmann.ac.il/report/2017/010

[53] B. Xiao. 1992. *New bounds in cell probe model*. Ph.D. Dissertation. UC San Diego.

[54] Andrew Chi-Chih Yao. 1979. Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *Proceedings of the 11h STOC*. 209–213. https://doi.org/10.1145/800135.804414

[55] Penghui Yao. 2015. Parity decision tree complexity and 4-party communication complexity of XOR-functions are polynomially equivalent. *arXiv:1506.02936* (2015).