Electronic Colloquium on Computational Complexity, Report No. 111 (2016)

Strong Fooling Sets for Multi-Player Communication with Applications to Deterministic Estimation of Stream Statistics

Amit Chakrabarti[*] Sagar Kale[†]

July 19, 2016

Abstract

We develop a paradigm for studying multi-player deterministic communication, based on a novel combinatorial concept that we call a *strong fooling set*. Our paradigm leads to optimal lower bounds on the per-player communication required for solving multi-player EQUALITY problems in a private-message setting. This in turn gives a very strong—$O(1)$ versus $\Omega(n)$—separation between private-message and one-way blackboard communication complexities.

Applying our communication complexity results, we show that for deterministic data streaming algorithms, even loose estimations of some basic statistics of an input stream require large amounts of space. For instance, approximating the frequency moment F_k within a factor α requires $\Omega(n/\alpha^{1/(1-k)})$ space for $k < 1$ and roughly $\Omega(n/\alpha^{k/(k-1)})$ space for $k > 1$. In particular, approximation within any *constant* factor α, however large, requires *linear* space, with the trivial exception of $k = 1$. This is in sharp contrast to the situation for randomized streaming algorithms, which can approximate F_k to within $(1 \pm \varepsilon)$ factors using $\widetilde{O}(1)$ space for $k \leqslant 2$ and $o(n)$ space for all finite k and all constant $\varepsilon > 0$. Previous linear-space lower bounds for deterministic estimation were limited to small factors α, such as $\alpha < 2$ for approximating F_0 or F_2.

We also provide certain space/approximation tradeoffs in a deterministic setting for the problems of estimating the empirical entropy of a stream as well as the size of the maximum matching and the edge connectivity of a streamed graph.

[*] Department of Computer Science, Dartmouth College. Email: ac (at) cs.dartmouth.edu. Work supported in part by NSF Award CCF-1217375.

[†] Department of Computer Science, Dartmouth College. Email: sag (at) cs.dartmouth.edu. Work supported in part by NSF Award CCF-1217375.

ISSN 1433-8092

1 Introduction

This paper introduces a new paradigm for studying multi-player number-in-hand deterministic communication complexity, uses the paradigm to obtain some new communication lower bounds centered around the EQUALITY problem, and applies these results to derive a number of lower bounds in the data streaming model. These latter results address a very basic topic in the theory of data stream algorithms.

Data stream algorithms offer a convincing demonstration of the power of randomization. A large number of problems that call for summarization of "big data" admit remarkably efficient *sublinear*-space streaming algorithms, provided such algorithms are allowed to make random decisions. Restricted to determinism, such sublinear-space solutions usually do not exist. This dichotomy between determinism and randomization is among the first lessons one learns in the study of streaming algorithms. Indeed, as shown in the pioneering work of Alon, Matias, and Szegedy [3], one encounters it in the most basic problem of estimating the number of distinct elements in a stream, as well as the more general problem of estimating the stream's frequency moments.

Applying our nuanced understanding of the multi-player EQUALITY problem, we establish very sharp contrasts between the deterministic and randomized space complexities of several important problems for data streams. The problems we study are already known to admit very efficient and accurate randomized estimations. For most of them, past work shows that *accurate* deterministic estimations require linear space. An important takeaway from our present work is that even *loose* deterministic estimations require linear space: randomization is even more crucial to these problems than was previously realized.

Returning to the technical level of communication complexity, our lower-bounding paradigm, which drives these results, should be of independent interest. It sharply separates the "blackboard" and "private-message" communication models, an issue previously highlighted by Gál and Gopalan [24].

Contributions to Data Streaming. Take the case of estimating F_0, the number of distinct elements, in a stream of elements from the universe $[n] := \{1, \ldots, n\}$. The randomized algorithm of Kane et al. [27] makes one pass over the stream, using $O(\varepsilon^{-2} + \log n)$ bits of space, and computes an estimate that lies in $[(1-\varepsilon)F_0, (1+\varepsilon)F_0]$ with probability $\geqslant \frac{2}{3}$. In contrast, Alon et al. [3] show that, for $\alpha = 1.1$, a deterministic estimator that returns a value in $[\alpha^{-1}F_0, \alpha F_0]$ using $O(1)$ passes must use $\Omega(n)$ bits of space. One consequence of our results is that this $\Omega(n)$ space bound holds for *every* constant α, no matter how large.

We go on to show that for every frequency moment F_k, apart from $k = 1$, deterministic constant-pass constant-factor estimation requires $\Omega(n)$ space. Stated in full (Section 4), our results give detailed tradeoffs between the space usage, the number of passes, and the quality of estimation for computing the frequency moments and the empirical entropy of a stream, and for estimating graph parameters (Section 5) such as the size of a maximum matching and the edge connectivity, when the input is a stream of undirected edges.

Contributions to Communication Complexity. At a philosophical level, our results ultimately derive from the phenomenon that, in a communication complexity setting, determining equality between strings is very hard deterministically but easy with randomization. However, this alone does not lead to the strong streaming lower bounds we are claiming. In particular, the standard two-player EQUALITY (EQ) problem does not yield these results. Instead, we study multi-player

EQ problems with certain strong promises (Section 3) that greatly separate YES-instances from NO-instances. In one version, each of t players is given an n-bit string: these strings are either all equal or all distinct. Another variant gives the players fixed-sized sets that are either all equal or have large "spread," i.e., large union size.

To effectively lower-bound the complexities of these promise versions of EQ, we need to consider *discreet protocols*, wherein players can only communicate via private messages. This is in contrast to the better-studied *blackboard* protocols, which allow players to write their messages on a public blackboard. The latter model makes our promise-EQ problems too easy. Gál and Gopalan [24] had previously shown a separation between the discreet and one-way blackboard models, but for an artificial[1] function and with a bespoke proof for the discreet lower bound. Here, we separate these models using what may be *the* natural separating problem (q.v. the end of Section 3). More importantly, we introduce an abstract paradigm for studying discreet protocols, based on a novel combinatorial concept that we call a *strong fooling set* (Definition 2.7) and a key technical result that we call the "strong fooling set bound" (Lemma 2.8).

Other Contributions. In the process of designing reductions from our promise-EQUALITY problems to obtain our data streaming lower bounds, we establish a sharp concentration result for power sums of a collection of independent binomial random variables. This result (Lemma 4.4), though basic-looking and proved via elementary means, appears to be novel. Indeed, other seemingly basic questions about moments of the binomial distribution seem to have been addressed only recently [8].

Other Related Work. The general topic of the communication complexity of EQUALITY problems, in their many variants, has occupied researchers for decades, beginning with Yao's seminal work [40]. Some key results for 2-player EQ address its amortized complexity [22], simultaneous-message complexity [6, 9, 10], direct sum properties [15, 34], round complexity [13], and information complexity [12]. Multi-player versions of EQ, which are interesting only in a number-in-hand setting, have been receiving attention recently, as has the discreet model of communication. Liang and Vaidya [31] consider the basic version of distinguishing all-equal inputs from not-all-equal ones. This is easily shown to require $nt/2$ bits of deterministic communication in total; they show a nontrivial upper bound of cnt for a specific constant $c < 1$. A combinatorial construction of Alon, Moitra, and Sudakov [4] improves this to $c = 1/2 + \varepsilon$. Chattopadhyay, Radhakrishnan, and Rudra [16] consider another (also non-promise) variant: ELEMENT-DISTINCTNESS, where the players must distinguish all-distinct inputs from not-all-distinct ones. They study the effect of the communication topology—who may send messages to whom—on the complexities of this, and other, problems. Our lower bounds in this work hold regardless of topology.

In the topology-restricted setting, the *coordinator* model has attracted a lot of study [17, 7, 35]. Building on techniques developed for this model, Woodruff and Zhang [37, 39] gave near-optimal lower bounds for estimating stream statistics and graph parameters in a randomized distributed setting. Separately, they gave strong lower bounds for randomized *exact* computation of such statistics under general topology [38]. This complements our streaming results, which concern *deterministic* weakly-approximate computation.

Our results on estimation of graph parameters contribute to a fast-developing body of literature on streaming and sketching algorithms for graphs. We discuss the relevant background in Section 5.

[1] The Gál–Gopalan function is arguably artificial as a *communication* problem. It was used in aid of space lower bounds for the (natural) data-streaming problem of estimating the length of the longest increasing sequence.

2 Definitions and Preliminaries: Communication Complexity

All our logarithms are to the base 2. The notation $f\colon A \rightsquigarrow B$ denotes a partial function f with domain A and codomain B; formally this is a function $f\colon A \to B \cup \{\star\}$, where $\star \notin B$ is a special do-not-care value. We say that f *is constant* on $A' \subseteq A$ if $|f(A') \setminus \{\star\}| \leqslant 1$.

Let $\mathcal{X}_1, \ldots, \mathcal{X}_t$ be finite sets. Put $\boldsymbol{\mathcal{X}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_t$. Consider a communication game between players $\textsc{plr}_1, \ldots, \textsc{plr}_t$ given by a partial function $f\colon \boldsymbol{\mathcal{X}} \rightsquigarrow \mathcal{Z}$. An input for this game is a tuple $\mathbf{x} = (x_1, \ldots, x_t)$. At the start of the game, $\textsc{plr}_i$ receives the *input fragment* $x_i \in \mathcal{X}_i$, for each i; the players then communicate according to a deterministic protocol Π that ends with an output $\Pi^o(\mathbf{x}) \in \mathcal{Z}$. We say that Π computes f if

$$\forall\, \mathbf{x} \in \boldsymbol{\mathcal{X}} :\ f(\mathbf{x}) \neq \star \implies \Pi^o(\mathbf{x}) = f(\mathbf{x})\,.$$

In the above context, a *discreet protocol* is a sequence $(s_1, r_1, M_1), \ldots, (s_L, r_L, M_L)$, where the jth tuple describes the action during the jth *step* of the protocol: the sender, $\textsc{plr}_{s_j}$, sends a *private* message to the receiver, $\textsc{plr}_{r_j}$, using the message function $M_j\colon \{0,1\}^* \times \mathcal{X}_{s_j} \to \{0,1\}^*$. Specifically, if h_j is the concatenation of the messages received by $\textsc{plr}_{s_j}$ during the first $j-1$ steps, then $M_j(h_j, x_{s_j})$ is the message she sends in the jth step. We require that the range of each M_j be a prefix code; this makes such concatenations self-punctuating. Notice that our discreet protocols are "oblivious" in the sense that the communication pattern (including L, the number of steps) is input-independent. The final (Lth) message is defined to be the output of the protocol, using some canonical map from $\{0,1\}^*$ to $\mathcal{Z}$; the "receiver" r_L is a dummy value.

Let Π be a discreet protocol formalized as above. For an input $\mathbf{x}$ and a player $\textsc{plr}_i$, the *local transcript* $\Pi_i(\mathbf{x})$ is defined to be concatenation (in order of occurrence) of the messages received and sent by $\textsc{plr}_i$ when Π is run on $\mathbf{x}$. The protocol Π is *B-bounded* if, for all i and $\mathbf{x}$, $|\Pi_i(\mathbf{x})| \leqslant B$, i.e., each player sends and receives a total of at most B bits. We define

$$\begin{aligned} \operatorname{cost}(\Pi) &:= \max\nolimits_{i,\mathbf{x}} |\Pi_i(\mathbf{x})| = \min\{B :\ \Pi \text{ is } B\text{-bounded}\}\,; \\ \mathrm{DD}(f) &:= \min\{\operatorname{cost}(\Pi) :\ \text{discreet protocol } \Pi \text{ computes } f\}\,. \end{aligned}$$

Another, better-studied, kind of protocol for such multi-player communication games is a *blackboard protocol*. Here, players communicate by writing messages on a blackboard visible to all players. The communication history determines which player will write the next bit on the blackboard, and the bit written is a function of this history and the writer's input fragment. Let $\Pi_i^{\mathrm{w}}(\mathbf{x})$ denote the concatenation of all messages written by $\textsc{plr}_i$ when a blackboard protocol Π is run on an input $\mathbf{x}$. We define

$$\begin{aligned} \operatorname{cost}(\Pi) &:= \max\nolimits_{i,\mathbf{x}} |\Pi_i^{\mathrm{w}}(\mathbf{x})|\,; \\ \mathrm{BB}(f) &:= \min\{\operatorname{cost}(\Pi) :\ \text{blackboard protocol } \Pi \text{ computes } f\}\,; \\ \operatorname{cost}^{\mathrm{tot}}(\Pi) &:= \max\nolimits_{\mathbf{x}} \left(|\Pi_1^{\mathrm{w}}(\mathbf{x})| + \cdots + |\Pi_t^{\mathrm{w}}(\mathbf{x})|\right)\,; \\ \mathrm{BB}^{\mathrm{tot}}(f) &:= \min\{\operatorname{cost}^{\mathrm{tot}}(\Pi) :\ \text{blackboard protocol } \Pi \text{ computes } f\}\,. \end{aligned}$$

Notice that discreet protocols can be thought of as special cases of blackboard protocols: ones that are oblivious and wherein each message is ignored by all players except its receiver. This natural simulation translates a B-bounded t-player discreet protocol Π into a blackboard protocol Π' with $\operatorname{cost}(\Pi') \leqslant B$ and $\operatorname{cost}^{\mathrm{tot}}(\Pi') \leqslant tB/2$.

2.1 Warm-Up: A Weak Fooling Set Bound

We recall the well-known concepts of *rectangles* and *fooling sets* used in the analysis of deterministic two-player protocols [30] and, by an easy extension [24, 20], to deterministic t-player blackboard protocols. A (combinatorial) rectangle in $\boldsymbol{\mathcal{X}}$ is a set of the form $\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_t$, where each $\mathcal{Y}_i \subseteq \mathcal{X}_i$. The *span* of a set $\boldsymbol{\mathcal{F}} \subseteq \boldsymbol{\mathcal{X}}$, denoted $\operatorname{span}(\boldsymbol{\mathcal{F}})$, is the minimal rectangle that includes $\boldsymbol{\mathcal{F}}$.

Let Π be some t-player blackboard protocol on input space $\boldsymbol{\mathcal{X}}$. The *transcript* of Π on input $\mathbf{x} \in \boldsymbol{\mathcal{X}}$ is defined to be $(\Pi_1^{\mathrm{w}}(\mathbf{x}), \ldots, \Pi_t^{\mathrm{w}}(\mathbf{x}))$. Two inputs that generate the same transcript are *equivalent*: this relation partitions $\boldsymbol{\mathcal{X}}$ into equivalence classes.

Lemma 2.1 (Rectangle Property (folklore))**.** *Each equivalence class of* Π *is a rectangle in* $\boldsymbol{\mathcal{X}}$*. In particular, if* $\boldsymbol{\mathcal{F}} \subseteq \boldsymbol{\mathcal{X}}$ *lies within an equivalence class, then so does* $\operatorname{span}(\boldsymbol{\mathcal{F}})$*. Consequently, if* Π *computes a partial function* f*, then* f *is constant on* $\operatorname{span}(\boldsymbol{\mathcal{F}})$*.* □

Let $f\colon \boldsymbol{\mathcal{X}} \rightsquigarrow \mathcal{Z}$ specify a communication game and let $\boldsymbol{\mathcal{F}} \subseteq \boldsymbol{\mathcal{X}}$. We say that $\boldsymbol{\mathcal{F}}$ is a *K-weak-fooling set* for f if, for all $\boldsymbol{\mathcal{F}}' \subseteq \boldsymbol{\mathcal{F}}$ with $|\boldsymbol{\mathcal{F}}'| > K$, f is nonconstant on $\operatorname{span}(\boldsymbol{\mathcal{F}}')$. The standard notion of "fooling set" used in a number of two-player communication complexity lower bounds would be a 1-weak-fooling set under this terminology.

It follows, using Lemma 2.1, that if f and $\boldsymbol{\mathcal{F}}$ are as above, and Π is a blackboard protocol that computes f, then no subset of $\boldsymbol{\mathcal{F}}$ of size $> K$ can lie within an equivalence class of Π. So Π must have at least $|\boldsymbol{\mathcal{F}}|/K$ equivalence classes. This gives us the following basic lower bounds, which (we emphasize) we are stating for contrast with what is to follow.

Lemma 2.2 (Weak fooling set bound)**.** *Suppose that* $f\colon \boldsymbol{\mathcal{X}} \rightsquigarrow \mathcal{Z}$ *specifies a t-player communication game and that f has a K-weak-fooling set* $\boldsymbol{\mathcal{F}}$*. Then*

$$\mathrm{BB}^{tot}(f) \geqslant \log \frac{|\boldsymbol{\mathcal{F}}|}{K}\,; \quad \mathrm{DD}(f) \geqslant \frac{2}{t} \log \frac{|\boldsymbol{\mathcal{F}}|}{K}\,.$$

Proof. The former bound is immediate by counting. The latter follows from the aforementioned simulation of discreet protocols in the blackboard model. □

2.2 The Strong Fooling Set Bound for Discreet Protocols

Having set the stage, we now focus on discreet protocols. We shall analyze such protocols by using more nuanced notions of equivalence of inputs, and then strengthening the above notion of weak fooling sets.

Let Π be a t-player discreet protocol on input space $\boldsymbol{\mathcal{X}}$. Inputs $\mathbf{x}$ and $\mathbf{y}$ are *i-equivalent* if $\Pi_i(\mathbf{x}) = \Pi_i(\mathbf{y})$; they are *equivalent* if they are i-equivalent for all $i \in [t]$. These relations partition the input space $\boldsymbol{\mathcal{X}}$ into the *i-equivalence classes* and the *equivalence classes* of Π, respectively. Clearly, $\Pi^o(\mathbf{x}) = \Pi^o(\mathbf{y})$ whenever $\mathbf{x}$ and $\mathbf{y}$ are equivalent.

Let $\boldsymbol{\mathcal{G}} \subseteq \boldsymbol{\mathcal{X}}$ be nonempty. A *neighborhood within* $\boldsymbol{\mathcal{G}}$ is a t-tuple $\mathcal{N} = (\boldsymbol{\mathcal{H}}_1, \ldots, \boldsymbol{\mathcal{H}}_t)$ where each $\boldsymbol{\mathcal{H}}_i \subseteq \boldsymbol{\mathcal{G}}$ and its *core*, defined by $\operatorname{core}(\mathcal{N}) := \boldsymbol{\mathcal{H}}_1 \cap \cdots \cap \boldsymbol{\mathcal{H}}_t$, is nonempty. For an input $\mathbf{x} = (x_1, \ldots, x_t)$, we put $\operatorname{proj}_i \mathbf{x} := x_i$. We extend this notation to sets, defining $\operatorname{proj}_i \boldsymbol{\mathcal{G}} := \{\operatorname{proj}_i \mathbf{x} : \mathbf{x} \in \boldsymbol{\mathcal{G}}\}$. We define the width and the span of the neighborhood $\mathcal{N}$ as follows:

$$\begin{aligned} \operatorname{wid}(\mathcal{N}) &:= \min\{|\boldsymbol{\mathcal{H}}_1|, \ldots, |\boldsymbol{\mathcal{H}}_t|\}\,, \\ \operatorname{span}(\mathcal{N}) &:= \operatorname{proj}_1 \boldsymbol{\mathcal{H}}_1 \times \cdots \times \operatorname{proj}_t \boldsymbol{\mathcal{H}}_t\,. \end{aligned}$$

The definitions immediately imply that $\operatorname{core}(\mathcal{N}) \subseteq \operatorname{span}(\mathcal{N}) \subseteq \boldsymbol{\mathcal{X}}$ and that $\operatorname{core}(\mathcal{N}) \subseteq \boldsymbol{\mathcal{G}}$.

We say that a protocol Π is *smooth* on $\mathcal{N}$ if, for each $i \in [t]$, $\boldsymbol{\mathcal{H}}_i$ lies within an i-equivalence class of Π. This notion allows us to generalize the rectangle property from Lemma 2.1.

Lemma 2.3 (Generalized Rectangle Property)**.** *Let Π be a discreet protocol on input space $\boldsymbol{\mathcal{X}}$. Let $\mathcal{N}$ be a neighborhood within $\boldsymbol{\mathcal{X}}$ such that Π is smooth on $\mathcal{N}$. Then* $\operatorname{span}(\mathcal{N})$ *lies within an equivalence class of Π. Consequently, if Π computes a partial function f, then f is constant on* $\operatorname{span}(\mathcal{N})$*.*

To see that the above is a generalization, consider the neighborhood $(\boldsymbol{\mathcal{G}}, \ldots, \boldsymbol{\mathcal{G}})$, where $\boldsymbol{\mathcal{G}}$ lies within an equivalence class of some blackboard protocol.

The following helper lemma will help us prove the generalized rectangle property.

Lemma 2.4. *With respect to a protocol Π, suppose that inputs $\mathbf{x} := (x_1, \ldots, x_t)$ and $\mathbf{y} := (y_1, \ldots, y_t)$ are i-equivalent, for some $i \in [t]$. Then $\mathbf{x}$ and $\mathbf{x}' := (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_t)$ are equivalent.*

Proof. Think of Π as a "virtual" two-player protocol between PLR_i and the rest of the players, combined into a single entity. The local transcript $\Pi_i(\mathbf{x})$ is the transcript of this virtual protocol on input $\mathbf{x}$. Since $\Pi_i(\mathbf{x}) = \Pi_i(\mathbf{y})$, Lemma 2.1 (the usual rectangle property) applied to this virtual protocol tells us that $\Pi_i(\mathbf{x}) = \Pi_i(\mathbf{x}')$.

Consider a switch of input from $\mathbf{x}$ to $\mathbf{x}'$, and consider an arbitrary $j \in [t]$ with $j \neq i$. Since $\Pi_i(\mathbf{x}) = \Pi_i(\mathbf{x}')$, the switch affects neither the input fragment nor any messages received by PLR_j. Therefore $\Pi_j(\mathbf{x}) = \Pi_j(\mathbf{x}')$. We conclude that $\mathbf{x}$ and $\mathbf{x}'$ are equivalent. □

Proof of Lemma 2.3. Fix an input $\mathbf{x} := (x_1, \ldots, x_t) \in \operatorname{core}(\mathcal{N})$. We shall prove that every input in $\operatorname{span}(\mathcal{N})$ is equivalent to $\mathbf{x}$. Let $\mathbf{y} = (y_1, \ldots, y_t) \in \operatorname{span}(\mathcal{N})$. Put

$$\mathbf{x}^i := (y_1, \ldots, y_{i-1}, x_i, \ldots, x_t)\,, \quad \text{for } i \in [t+1]\,.$$

Since Π is smooth on $\mathcal{N}$, for $i \in [t]$, the set $\{\Pi_i(\mathbf{x}') : \mathbf{x}' \in \boldsymbol{\mathcal{H}}_i\}$ is a singleton; let π_i be its lone element.

We shall prove by induction on i that $\mathbf{x}$ and $\mathbf{x}^i$ are equivalent. This will imply that $\mathbf{x}^1 = \mathbf{x}$ and $\mathbf{x}^{t+1} = \mathbf{y}$ are equivalent, as required. The base case, $i = 1$, is trivial. Since $\mathbf{y} \in \operatorname{span}(\mathcal{N})$, we have $y_i \in \operatorname{proj}_i \boldsymbol{\mathcal{H}}_i$ and there is some $\mathbf{z} := (z_1, \ldots, z_{i-1}, y_i, z_{i+1}, \ldots, z_t) \in \boldsymbol{\mathcal{H}}_i$. So $\Pi_i(\mathbf{z}) = \pi_i$. Also, $\mathbf{x} \in \operatorname{core}(\mathcal{N}) \subseteq \boldsymbol{\mathcal{H}}_i$, so $\Pi_i(\mathbf{x}) = \pi_i$. Thus, $\mathbf{z}$ and $\mathbf{x}$ are i-equivalent.

Using Lemma 2.4, we get that $\mathbf{x}$ and $(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_t) =: \mathbf{x}'$ are equivalent (paste the ith coordinate of $\mathbf{z}$, which is y_i, into that of $\mathbf{x}$). By the inductive hypothesis, $\mathbf{x}$ and $\mathbf{x}^i$ are equivalent. So $\mathbf{x}'$ and $\mathbf{x}^i$ are equivalent, and hence, i-equivalent. Using Lemma 2.4 again, we get that $\mathbf{x}^i$ and $(y_1, \ldots, y_i, x_{i+1}, \ldots, x_t) = \mathbf{x}^{i+1}$ are equivalent. This completes the inductive step. □

By a pigeonhole argument, a low-cost blackboard protocol entails a large rectangle lying within an equivalence class.[2] We prove the following stronger result for discreet protocols.

Lemma 2.5. *Let Π be a B-bounded t-player discreet protocol on input space $\boldsymbol{\mathcal{X}}$. Let $\boldsymbol{\mathcal{G}} \subseteq \boldsymbol{\mathcal{X}}$. Then there exists a neighborhood $\mathcal{N}$ within $\boldsymbol{\mathcal{G}}$ such that Π is smooth on $\mathcal{N}$ and* $\operatorname{wid}(\mathcal{N}) \geqslant |\boldsymbol{\mathcal{G}}|/(t2^B)$*.*

Proof. We begin with an observation, readily proved by counting.

Observation 2.6. *Suppose the finite set S is partitioned into L blocks and $s \in_R S$ is picked uniformly at random. For every real $A > 0$,* $\Pr[$*s lies in a block of size* $< |S|/(AL)] < 1/A$.

For each $i \in [t]$, the i-equivalence classes of Π partition $\boldsymbol{\mathcal{X}}$, and hence $\boldsymbol{\mathcal{G}}$, into at most 2^B blocks. Pick $\mathbf{x} \in_R \boldsymbol{\mathcal{G}}$ uniformly at random. Put $[\![\mathbf{x}]\!]_i := \{\mathbf{y} \in \boldsymbol{\mathcal{G}} : \mathbf{y} \text{ is } i\text{-equivalent to } \mathbf{x}\}$. Then

[2] Such a rectangle is often said to be "monochromatic."

$\mathcal{N}_{\mathbf{x}} := (\llbracket \mathbf{x} \rrbracket_1, \ldots, \llbracket \mathbf{x} \rrbracket_t)$ is a neighborhood within $\boldsymbol{\mathcal{G}}$ (its core is nonempty because it contains $\mathbf{x}$) on which Π is smooth. By the above observation and a union bound, we have

$$\begin{aligned} \Pr\left[\operatorname{wid}(\mathcal{N}_{\mathbf{x}}) \geqslant \frac{|\boldsymbol{\mathcal{G}}|}{t2^B}\right] &= 1 - \Pr\left[\exists\, i : \ \left|\llbracket \mathbf{x} \rrbracket_i\right| < \frac{|\boldsymbol{\mathcal{G}}|}{t2^B}\right] \\ &\geqslant 1 - \sum_{i=1}^{t} \Pr\left[\left|\llbracket \mathbf{x} \rrbracket_i\right| < \frac{|\boldsymbol{\mathcal{G}}|}{t2^B}\right] \\ &> 1 - \sum_{i=1}^{t} \frac{1}{t} = 0\,. \end{aligned}$$

Thus, by the probabilistic method, the lemma follows. □

We now strengthen our earlier notion of weak fooling sets and prove our main technical lemma, a stronger communication lower bound in terms of these "strong" fooling sets.

Definition 2.7. Let $f\colon \boldsymbol{\mathcal{X}} \rightsquigarrow \mathcal{Z}$ specify a communication game and let $\boldsymbol{\mathcal{F}} \subseteq \boldsymbol{\mathcal{X}}$. We say that $\boldsymbol{\mathcal{F}}$ is a *K-fooling set* for f if, for every neighborhood $\mathcal{N}$ within $\boldsymbol{\mathcal{F}}$,

$$\operatorname{wid}(\mathcal{N}) > K \implies f \text{ is nonconstant on } \operatorname{span}(\mathcal{N})\,.$$

Lemma 2.8 (Strong fooling set bound)**.** *Suppose that $f\colon \boldsymbol{\mathcal{X}} \rightsquigarrow \mathcal{Z}$ specifies a t-player communication game and that f has a K-fooling set $\boldsymbol{\mathcal{F}}$. Then*

$$\mathrm{DD}(f) \geqslant \log \frac{|\boldsymbol{\mathcal{F}}|}{tK}\,.$$

Proof. Let Π be a B-bounded discreet protocol for f. By Lemma 2.5, there exists a neighborhood $\mathcal{N}$ within $\boldsymbol{\mathcal{F}}$ with $\operatorname{wid}(\mathcal{N}) \geqslant |\boldsymbol{\mathcal{F}}|/(t2^B)$ such that Π is smooth on $\mathcal{N}$. By Lemma 2.3, f is constant on $\operatorname{span}(\mathcal{N})$. In view of Definition 2.7, we must have $\operatorname{wid}(\mathcal{N}) \leqslant K$, which implies

$$\frac{|\boldsymbol{\mathcal{F}}|}{t2^B} \leqslant K\,.$$

The lemma follows by rearranging the above inequality. □

3 Three Lower Bounds for Multi-Player Equality

We shall now use our strong fooling set bound to analyze two promise versions of the EQUALITY problem, as alluded to in Section 1. Each of our problems is given by a partial function of the form $f\colon \mathcal{X}^t \rightsquigarrow \{0,1\}$, where $\mathcal{X} = \{0,1\}^n$. In the Equal-vs-Distinct problem, the goal is to distinguish the case when all players hold the same n-bit string from the case when no two players hold the same string. This is formalized by the following function:

$$\text{EQ-DIST}_{n,t}(x_1,\ldots,x_t) = \begin{cases} 1\,, & \text{if } x_1 = \cdots = x_t\,, \\ 0\,, & \text{if } x_i \neq x_j \text{ whenever } 1 \leqslant i < j \leqslant t\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

In the Equal-vs-Spread problem, each player receives a $\lceil \beta n \rceil$-subset of $[n]$ and they must distinguish the case when all of these subsets are equal from the case when these subsets are sufficiently spread out. Formally, we interpret $\mathcal{X}$ as the power set $2^{[n]}$ and use the following function:

$$\textsc{eq-sprd}_{n,t}^{\beta,\gamma}(x_1,\ldots,x_t) = \begin{cases} 1\,, & \text{if } |x_1| = \cdots = |x_t| = \lceil \beta n \rceil \text{ and } x_1 = \cdots = x_t\,, \\ 0\,, & \text{if } |x_1| = \cdots = |x_t| = \lceil \beta n \rceil \text{ and } |x_1 \cup \cdots \cup x_t| \geqslant \gamma n\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

This problem is nontrivial when the parameters satisfy $0 < \beta < \gamma \leqslant 1$ and $\gamma n > \lceil \beta n \rceil$.

We now give strong, essentially optimal, communication lower bounds for discreet deterministic protocols that solve these problems. In each case, when the problem's input space is $\mathcal{X}^t$ and $x \in \mathcal{X}$ is an input fragment, we denote the input $(x, x, \ldots, x) \in \mathcal{X}^t$ by $x^{\otimes t}$.

The following observation will aid some of our estimations.

Observation 3.1. *For all integral values $0 \leqslant k \leqslant \ell \leqslant n$,*

$$\binom{n}{k} \Big/ \binom{\ell}{k} = \frac{n}{\ell} \cdot \frac{n-1}{\ell-1} \cdot \cdots \cdot \frac{n-k+1}{\ell-k+1} \geqslant \left(\frac{n}{\ell}\right)^k .$$

Theorem 3.2 (Lower bound for Equal-vs-Distinct)**.** $\mathrm{DD}(\textsc{eq-dist}_{n,t}) \geqslant n - 2\log t$.

Proof. Put $f := \textsc{eq-dist}_{n,t}$. We claim that the set $\boldsymbol{\mathcal{F}} := \{x^{\otimes t} : x \in \{0,1\}^n\}$ is a $(t-1)$-fooling set for f. Indeed, let $\mathcal{N} = (\boldsymbol{\mathcal{H}}_1, \ldots, \boldsymbol{\mathcal{H}}_t)$ be a neighborhood within $\boldsymbol{\mathcal{F}}$ such that $\mathrm{wid}(\mathcal{N}) > t-1$. Since $f(\mathbf{x}) = 1$ for all $\mathbf{x} \in \boldsymbol{\mathcal{F}}$, our earlier observations that $\varnothing \neq \mathrm{core}(\mathcal{N}) \subseteq \boldsymbol{\mathcal{F}}$ and $\mathrm{core}(\mathcal{N}) \subseteq \mathrm{span}(\mathcal{N})$ imply that f takes the value 1 at some point in $\mathrm{span}(\mathcal{N})$. On the other hand, consider the point $\mathbf{y} = (y_1, \ldots, y_t)$ constructed by the following procedure:

- Having chosen $y_1, \ldots, y_{i-1}$, where $1 \leqslant i \leqslant t$, choose an arbitrary fragment y_i such that $y_i^{\otimes t} \in \boldsymbol{\mathcal{H}}_i \setminus \{y_1^{\otimes t}, \ldots, y_{i-1}^{\otimes t}\}$. This choice is possible because $|\boldsymbol{\mathcal{H}}_i| \geqslant \mathrm{wid}(\mathcal{N}) > t-1 \geqslant i-1$.

The construction ensures that $\mathbf{y} \in \mathrm{span}(\mathcal{N})$ and $f(\mathbf{y}) = 0$. Therefore f is nonconstant on $\mathrm{span}(\mathcal{N})$, proving the claim.

Applying Lemma 2.8,

$$\mathrm{DD}(f) \geqslant \log \frac{|\boldsymbol{\mathcal{F}}|}{t(t-1)} \geqslant \log \frac{2^n}{t^2} = n - 2\log t\,. \qquad \square$$

Theorem 3.3 (Lower bound for Equal-vs-Spread)**.** *For all values $0 < \beta < \gamma \leqslant 1$ and sufficiently large n, if $t \geqslant \gamma n$, then $\mathrm{DD}(\textsc{eq-sprd}_{n,t}^{\beta,\gamma}) \geqslant (\beta \log(1/\gamma))n - \log t$.*

Proof. Put $f := \textsc{eq-sprd}_{n,t}^{\beta,\gamma}$ and $w = \lceil \beta n \rceil$. We claim that the set $\boldsymbol{\mathcal{F}} := \{x^{\otimes t} : x \in \{0,1\}^n, |x| = w\}$ is a $\binom{\lfloor \gamma n \rfloor}{w}$-fooling set for f. Indeed, let $\mathcal{N} = (\boldsymbol{\mathcal{H}}_1, \ldots, \boldsymbol{\mathcal{H}}_t)$ be a neighborhood within $\boldsymbol{\mathcal{F}}$ such that $\mathrm{wid}(\mathcal{N}) > \binom{\lfloor \gamma n \rfloor}{w}$. Since $f(\mathbf{x}) = 1$ for all $\mathbf{x} \in \boldsymbol{\mathcal{F}}$, as in the proof of Theorem 3.2, f takes the value 1 at some point in $\mathrm{span}(\mathcal{N})$. On the other hand, consider the point $\mathbf{y} = (y_1, \ldots, y_t)$ constructed by the following procedure:

- Having chosen $y_1, \ldots, y_{i-1}$, where $1 \leqslant i \leqslant t$, let $U_{i-1} := y_1 \cup \cdots \cup y_{i-1}$. If $|U_{i-1}| \geqslant \gamma n$, then choose an arbitrary fragment y_i such that $y_i^{\otimes t} \in \boldsymbol{\mathcal{H}}_i$.
- Otherwise, let $B_i := \{x \in \{0,1\}^n : x \subseteq U_{i-1}, |x| = w\}$. Choose an arbitrary fragment $y_i \notin B_i$ such that $y_i^{\otimes t} \in \boldsymbol{\mathcal{H}}_i$. This choice is possible because

$$|B_i| = \binom{|U_{i-1}|}{w} \leqslant \binom{\lfloor \gamma n \rfloor}{w} < \mathrm{wid}(\mathcal{N}) \leqslant |\boldsymbol{\mathcal{H}}_i|\,.$$

Then $\mathbf{y} \in \text{span}(\mathcal{N})$. Notice that $|U_i| > |U_{i-1}|$ whenever the second case occurs while choosing y_i. We make $t \geqslant \gamma n$ choices in total, which ensures that $|U_t| \geqslant \gamma n$, implying that $f(\mathbf{y}) = 0$. Therefore f is nonconstant on $\text{span}(\mathcal{N})$, proving the claim.

Applying Lemma 2.8 and Observation 3.1,

$$\text{DD}(f) \geqslant \log \frac{|\boldsymbol{\mathcal{F}}|}{t\binom{\lfloor \gamma n \rfloor}{w}} = \log \frac{\binom{n}{w}}{t\binom{\lfloor \gamma n \rfloor}{w}} \geqslant \log \frac{\left(\frac{n}{\lfloor \gamma n \rfloor}\right)^w}{t} \geqslant w \log \frac{1}{\gamma} - \log t\,. \qquad \square$$

Theorem 3.3 gives a lower bound for "large" t. In particular, if the spread threshold γn is to be $\Omega(n)$, then we have to take $t = \Omega(n)$ in order to apply the theorem. We now give an alternate lower bound for EQ-SPRD that holds in a different parameter regime, where t could be "small."

Theorem 3.4. *For all values $t \geqslant 2$, $\beta > 0$, $\gamma = \beta t(1 - e\beta t) > \beta$, and sufficiently large integral n, we have* $\text{DD}(\text{EQ-SPRD}_{n,t}^{\beta,\gamma}) \geqslant 2e\beta^2 n - 2\log t - \Theta(1)$.

We shall prove this result by a reduction from EQ-DIST, using a coding-style argument. The idea is to encode strings in $\{0,1\}^N$, for some suitable N, as fixed-sized subsets of $[n]$ that are pairwise "far apart." Define an (r,s,n)-packing to be set system $\mathcal{C} \subseteq 2^{[n]}$ such that

- for all $A \in \mathcal{C}$, $|A| = s$, and
- for all $A, B \in \mathcal{C}$ with $A \neq B$, $|A \cap B| \leqslant r$.

We shall need the following bound, which can be inferred from Proposition 2.1 in Erdös, Frankl, and Füredi [19].

Lemma 3.5. *For all values $0 \leqslant r \leqslant s \leqslant n$, there exists an (r,s,n)-packing $\mathcal{C}$ such that*

$$|\mathcal{C}| \geqslant \binom{n}{r} \Big/ \binom{s}{r}^2 . \qquad \square$$

Proof of Theorem 3.4. Let $\mathcal{C}$ be a maximum-sized (r,s,n)-packing, with $s = \lceil \beta n \rceil$ and $r = 2\lceil e\beta s \rceil$. By Lemma 3.5, Observation 3.1, and the estimation $\binom{s}{r} \leqslant (es/r)^r$,

$$|\mathcal{C}| \geqslant \frac{\binom{n}{r}}{\binom{s}{r}^2} \geqslant \left(\frac{n}{s}\right)^r \left(\frac{r}{es}\right)^r = \left(\frac{nr}{es^2}\right)^r .$$

By our choice of parameters,

$$r \log \frac{nr}{es^2} \geqslant r \log \frac{2e\beta sn}{es^2} = r \log \frac{2\beta n}{\lceil \beta n \rceil} = r\left(1 - \Theta\left(\frac{1}{n}\right)\right) = 2e\beta^2 n - \Theta(1)\,.$$

Therefore, we can find an injection from $\{0,1\}^N$ to $\mathcal{C}$ provided $N \leqslant 2e\beta^2 n - \Theta(1)$. We choose the largest possible N satisfying this bound and fix such an injection.

To solve $\text{EQ-DIST}_{N,t}$, the players encode their respective input fragments using this injection and then solve $\text{EQ-SPRD}_{n,t}^{\beta,\gamma}$ on the encoded input. Recall that $\gamma = \beta t(1 - e\beta t)$. We now argue that this reduction is correct. A 1-input for $\text{EQ-DIST}_{N,t}$ is, rather obviously, mapped to a 1-input for $\text{EQ-SPRD}_{n,t}^{\beta,\gamma}$. Suppose a 0-input for $\text{EQ-DIST}_{N,t}$ maps to the input $(x_1, \ldots, x_t)$. Then $x_1, \ldots, x_t$ are distinct sets in $\mathcal{C}$. By the packing property,

$$|x_1 \cup \cdots \cup x_t| \geqslant ts - \binom{t}{2} r = ts - t(t-1)\lceil e\beta s \rceil \geqslant ts - t^2 e\beta s = \lceil \beta n \rceil t(1 - e\beta t) \geqslant \gamma n\,,$$

where the second inequality holds once n (and hence, s) is sufficiently large. So $\text{EQ-SPRD}_{n,t}^{\beta,\gamma}(x_1, \ldots, x_t) = 0$, which proves the correctness of the reduction.

Appealing to Theorem 3.2, we obtain $\mathrm{DD}(\text{EQ-SPRD}_{n,t}^{\beta,\gamma}) \geqslant \mathrm{DD}(\text{EQ-DIST}_{N,t}) \geqslant N - 2\log t \geqslant 2e\beta^2 n - 2\log t - \Theta(1)$, as required. □

We conclude this section with some commentary on the lower bounds that we have just shown.

Optimality. Suppose that $t = \mathrm{poly}(n)$. The trivial protocol, where PLR_1 sends his input to PLR_2, shows that $\mathrm{DD}(\text{EQ-DIST}_{n,t}) \leqslant n + 1$. This shows that Theorem 3.2 is tight up to lower order terms. Another trivial protocol, where players efficiently encode $\lceil \beta n \rceil$-subsets of $[n]$ and each sends his input to the "next" player, shows that

$$\mathrm{DD}(\text{EQ-SPRD}_{n,t}^{\beta,\gamma}) \leqslant 2\log \binom{n}{\lceil \beta n \rceil} \leqslant 2H(\beta)n\,,$$

for $\beta < 1/2$. Thus, Theorem 3.3 and Theorem 3.4 are both asymptotically tight in their dependence on n. Moreover, when $\gamma/\beta = O(1)$, Theorem 3.3 is also tight in its dependence on β.

Separation Between Models. Our lower bounds also demonstrate a separation between one-way blackboard protocols and discreet protocols, an issue highlighted by Gál and Gopalan [24]. Consider the following blackboard protocol for $\text{EQ-DIST}_{n,t}$. Partition $[n]$ into $t-1$ blocks, each of size at most $\lceil n/(t-1) \rceil$. For each $j \in [t-1]$, PLR_j announces his input fragment restricted to the jth block, provided all blocks from 1 to $j-1$ of his input fragment agree with previously announced blocks (if not, PLR_j ends the protocol with output 0). If this protocol reaches PLR_t, he knows the entirety of PLR_{t-1}'s input fragment. He outputs 1 if his own fragment agrees with this, and outputs 0 otherwise. Based on the promise, this is a correct protocol for $\text{EQ-DIST}_{n,t}$. This protocol has max-cost $O(n/t)$, which is $O(1)$ when $t = \Theta(n)$. Yet, $\mathrm{DD}(\text{EQ-DIST}_{n,\Theta(n)}) = n - \Theta(\log n)$.

4 Stream Statistics

In the data stream model, an input consists of m elements of $[n]$ arriving in the form of a stream σ that may be read in one or more passes by a *streaming* algorithm. Formally, σ is a sequence $(a_1, a_2, \ldots, a_m)$, where each $a_j \in [n]$. The stream σ defines a frequency vector $\mathbf{f} = \mathbf{f}(\sigma) = (f_1, \ldots, f_n)$, where $f_i = |\{j \in [m] : a_j = i\}|$ for each $i \in [n]$. Stream statistics problems involve computing some function of $\mathbf{f}$, e.g., frequency moments and empirical entropy, which we consider in this section. The kth frequency moment and the empirical entropy are defined, respectively, as $F_k(\mathbf{f}) := \sum_{i\in[n]} f_i^k$ and $\text{ENT}(\mathbf{f}) = m^{-1} \sum_{i\in[n]} f_i \log(m/f_i)$. Note that $F_1(\mathbf{f}) = m$ and $F_0(\mathbf{f})$ is the number of distinct elements in σ.

Our focus is on *deterministic* streaming algorithms. An s-space p-pass streaming algorithm is one that uses $s = s(m,n)$ bits of space to process its input, which it reads in $p = p(m,n)$ passes. Consider such an algorithm $\mathcal{A}$. We denote its output, on input σ, by $\mathcal{A}(\sigma)$. By splitting σ into $t = t(m,n)$ sub-streams, we obtain a communication problem for which $\mathcal{A}$ naturally gives rise to a $2ps$-bounded discreet protocol, for every t. For a real quantity $\alpha = \alpha(m,n) \geqslant 1$, we say that $\mathcal{A}$ is an α-estimator for a quantity $Q(\sigma)$ if

$$\exists\, \kappa, \lambda \geqslant 1 \; (\kappa\lambda \leqslant \alpha \text{ and } \forall\, \sigma \; (\kappa^{-1} Q(\sigma) \leqslant \mathcal{A}(\sigma) \leqslant \lambda Q(\sigma))\,)\,.$$

The main results in this section are lower bounds that trade off α against the product ps, for algorithms estimating frequency moments and empirical entropy. For F_k ($k \neq 1$), when $\alpha = O(1)$, we obtain the strongest possible bound: $ps = \Omega(n)$. We also give simple estimators (upper bounds) for F_k and ENT; for moments of "lower" order ($0 \leqslant k < 1$) these simple estimators show that our lower bounds are tight even in their dependence on α.

Throughout this section, asymptotic expressions may hide constants depending on k. For readability, we ignore floors and ceilings. This does not affect our (asymptotic) bounds.

4.1 Warm-Up: Distinct Elements and Basic Lower Bounds for Other Moments

We shall first obtain lower bounds for estimating F_k ($k \neq 1$) by reduction from the communication game $\textsc{eq-sprd}^{\beta,\gamma}_{n,t}$, for certain values of t, β, γ and invoking Theorems 3.3 and 3.4 to lower-bound $\mathrm{DD}(\textsc{eq-sprd}^{\beta,\gamma}_{n,t})$. The bound we obtain is tight for F_0 (the distinct elements problem). Our bounds here are also tight for all k when $\alpha = O(1)$. In the next section, we use a more complicated analysis to obtain a tighter tradeoff between ps and α.

Theorem 4.1. *For each $k \in [0,1)$, every deterministic s-space p-pass α-estimator for F_k satisfies $ps = \Omega(\max\{n^{1-k}/\alpha, n/\alpha^{2/(1-k)}\})$. In particular, at $k = 0$ we have $ps = \Omega(n/\alpha)$.*

Proof. Let $\mathbf{x} = (x_1, \ldots, x_t)$ be an input for $\textsc{eq-sprd}^{\beta,\gamma}_{n,t}$. For each $j \in [t]$, $\textsc{plr}_j$ turns his input fragment $x_j \in \{0,1\}^n$ into the stream of indices j where $x_j = 1$. The concatenations of the t such streams has frequency vector $\mathbf{f} = x_1 + \cdots + x_t$. Note that

$$\textsc{eq-sprd}^{\beta,\gamma}_{n,t}(\mathbf{x}) = 1 \implies F_k(\mathbf{f}) = \beta t^k n\,; \tag{1}$$

$$\textsc{eq-sprd}^{\beta,\gamma}_{n,t}(\mathbf{x}) = 0 \implies F_0(\mathbf{f}) \geqslant \gamma n\,. \tag{2}$$

Also, $F_k(\mathbf{f}) \geqslant F_0(\mathbf{f})$. Thus, an α-estimator for F_k can separate these two cases provided $\gamma/(\beta t^k) > \alpha$. If such an estimator uses s bits of space and p passes then, as argued at the start of Section 4, we have $2ps \geqslant \mathrm{DD}(\textsc{eq-sprd}^{\beta,\gamma}_{n,t})$. It remains to invoke a suitable communication lower bound.

Set $\gamma = 1/e$, $t = \gamma n$, and $\beta = \gamma/(\alpha t^k) - 1/n$. This ensures that $\gamma/(\beta t^k) > \alpha$ and optimizes the lower bound from Theorem 3.3, giving $ps \geqslant \frac{1}{2}((n/e)^{1-k}(\log e)/\alpha - \log n) = \Omega(n^{1-k}/\alpha)$.

We could instead apply Theorem 3.4 to estimate $\mathrm{DD}(\textsc{eq-sprd}^{\beta,\gamma}_{n,t})$. We set $t = (2\alpha)^{1/(1-k)}$ and $\beta < 1/(2et)$; the theorem then requires $\gamma = \beta t(1 - e\beta t)$. Note that $\gamma/(\beta t^k) > \alpha$, as required. Applying the theorem gives $ps \geqslant e\beta^2 n - \log t - \Theta(1) = \Omega(n/t^2) = \Omega(n/\alpha^{2/(1-k)})$. □

For frequency moments F_k of "higher" order ($k > 1$), we can follow a similar proof template, but it takes more work to analyze the effect of the "spread" case in the Equal-vs-Spread problem.

Theorem 4.2. *For each $k > 1$, every deterministic s-space p-pass α-estimator for F_k has $ps = \Omega(n/\alpha^{2k/(k-1)})$.*

Proof. As in Theorem 4.1, we reduce from $\textsc{eq-sprd}^{\beta,\gamma}_{n,t}$. An input $\mathbf{x}$ for $\textsc{eq-sprd}^{\beta,\gamma}_{n,t}$ turns into a stream with frequency vector $\mathbf{f}$ satisfying eqs. (1) and (2). We now need a good *upper* bound on $F_k(\mathbf{f})$ when eq. (2) applies. For this, we invoke the following technical lemma.

Lemma 4.3. *Let $g\colon \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function and let $\mathbf{f} \in \{0, 1, \ldots, t\}^n$ where $t \geqslant 2$. Suppose that $F_1(\mathbf{f}) = m$, $F_0(\mathbf{f}) \geqslant r$, and $rt \geqslant m$. Then*

$$\sum_{i=1}^{n} g(f_i) \leqslant \ell g(t) + (r - \ell) g(1)\,,$$

where $\ell = \lceil (m - r)/(t - 1) \rceil$.

The proof, given in Section 6, is via a shifting argument that redistributes the "mass" in $\mathbf{f}$, subject to the constraints on $F_1(\mathbf{f})$ and $F_0(\mathbf{f})$, and uses convexity and Karamata's inequality to analyze the effect of this redistribution on $\sum g(f_i)$.

We use Lemma 4.3 with $g(x) = x^k$, which is convex because $k > 1$. For our frequency vector $\mathbf{f}$, we have $m = \beta tn$ and, thanks to eq. (2), we may use $r = \gamma n$. We take $\gamma = \beta t(1 - e\beta t)$ as required by Theorem 3.4. This gives $\ell = \lceil e\beta^2 t^2 n/(t-1) \rceil \leqslant 3\beta^2 tn$. Also, $r - \ell \leqslant r \leqslant \beta tn$. Thus,

$$F_k(\mathbf{f}) = \sum_{i=1}^{n} f_i^k \leqslant \ell t^k + (r - \ell)1^k \leqslant 3\beta^2 tn \cdot t^k + \beta tn = \beta tn(3\beta t^k + 1)\,.$$

On the other hand, when eq. (1) applies, we have $F_k(\mathbf{f}) = \beta t^k n$. The gap between these two cases is at least

$$\beta t^k n \Big/ \Big(\beta tn(3\beta t^k + 1)\Big) = t^{k-1} \Big/ (3\beta t^k + 1)\,.$$

Setting $t = (2\alpha)^{1/(k-1)}$ and $\beta < 1/(3t^k)$ makes the above gap greater than α. Therefore, an s-space p-pass α-estimator for F_k gives a $2ps$-bounded discreet protocol for $\text{EQ-SPRD}_{n,t}^{\beta,\gamma}$. By Theorem 3.4, we get $ps \geqslant e\beta^2 n - \log t - \Theta(1) = \Omega(n/\alpha^{2k/(k-1)})$, as required. □

4.2 Stronger Lower Bounds for Frequency Moments

We shall now improve the lower bounds in Theorems 4.1 and 4.2, obtaining a tighter dependence on α. From a data-streaming perspective, the two new lower bounds for F_k estimation given in this section—one for $k > 1$ and one for $k < 1$—are the main theorems of this paper.

The improvements ultimately stem from sufficiently sharp concentration bounds for power sums of binomial random variables. Let $Y_1, \ldots, Y_n$ be independent random variables, each with binomial distribution $\mathcal{B}(t, q)$. Let $Z = Z(n, t, q, k) = Y_1^k + \cdots + Y_n^k$ be the kth power sum of this collection. In Section 6.1, we prove the following concentration bounds for Z.

Lemma 4.4. *For each $k > 1$, there exist $b, c > 0$ such that the following holds. For each $q \in (0, 1/(2e^2))$, there exist integers n_0 and t_0 such that, for all $n \geqslant n_0$ and $t \geqslant t_0$,*

$$\Pr[Z > bq^k t^k n] \leqslant \exp\left(-\frac{cq^k tn}{(\log\log(1/q))^2}\right)\,.$$

Lemma 4.5. *For each $k \geqslant 0$, each $q \in (0, 1)$, and each integer $t \geqslant 32/q$, there exists an integer n_0 such that, for all $n \geqslant n_0$,*

$$\Pr[Z < q^k t^k n/2^{k+1}] \leqslant \exp(-qtn/32)\,.$$

The upper tail bound, Lemma 4.4, does not follow from Chernoff-Hoeffding and Azuma-Hoeffding inequalities [5]; those give a much weaker upper bound of the form $\exp(-\Theta(n))$. Indeed, even a bound of $\exp(-\Theta(tn))$ would not be strong enough for our purposes. We need to understand how the coefficient in front of tn depends on q, and this seems to require a delicate partitioning of the large deviation event.

In contrast, the lower tail bound, Lemma 4.5, does follow from standard Chernoff bounds.

Our improved lower bounds for F_k estimation are obtained by reducing from $\text{EQ-DIST}_{N,t}$, using what we shall call an *F_k-separating mapping*, defined as follows. A function $R\colon \{0,1\}^N \to \{0,1\}^n$ is said to be F_k-separating with parameters (t, α) if

$$\frac{\min\{F_k(t \cdot R(x)) : x \in \{0,1\}^N\}}{\max\{F_k(R(x_1) + \cdots + R(x_t)) : \text{EQ-DIST}_{N,t}(x_1, \ldots, x_t) = 0\}} > \alpha\,, \quad \text{when } k > 1\,, \tag{3}$$

and

$$\frac{\min\{F_k(R(x_1)+\cdots+R(x_t)) : \textsc{eq-dist}_{N,t}(x_1,\ldots,x_t)=0\}}{\max\{F_k(t\cdot R(x)) : x\in\{0,1\}^N\}} > \alpha\,, \quad \text{when } k<1\,. \tag{4}$$

Suppose that such a mapping exists and that we have an s-space p-pass α-estimator for F_k over the universe $[n]$. Then, following the template from Section 4.1, a team of t players can solve $\textsc{eq-dist}_{N,t}$ by mapping their inputs to $\{0,1\}^n$ via R and converting the mapped inputs to streams over $[n]$. Applying Theorem 3.2, we obtain $2ps \geqslant \mathrm{DD}(\textsc{eq-dist}_{N,t}) = N - 2\log t$.

Theorem 4.6. *For each $k>1$ and $\alpha\geqslant 1$, every deterministic s-space p-pass α-estimator for F_k satisfies $ps = \Omega(n/(\alpha^{k/(k-1)}(\log\log\alpha)^2))$.*

Proof. We follow the outline above. It remains to prove the existence of an F_k-separating mapping for a large enough $N = N(n,t,\alpha)$ and a not-too-large t.

We construct the mapping R at random, as follows. Generate a random $2^N\times n$ matrix whose entries are independent Bernoulli random variables, each equal to 1 with probability q. We shall fix q later. Then, for each $x\in\{0,1\}^N$, define $R(x)$ to be the xth row of this matrix. We claim that with positive probability both of the following events occur:

$$\mathcal{E}_1 := \left\{\min\{F_k(t\cdot R(x)) : x\in\{0,1\}^N\} \geqslant qt^kn/2\right\},$$
$$\mathcal{E}_2 := \left\{\max\{F_k(R(x_1)+\cdots+R(x_t)) : \textsc{eq-dist}_{N,t}(x_1,\ldots,x_t)=0\} \leqslant bq^kt^kn\right\}.$$

Noting that $\mathbb{E}|R(x)| = qn$ for each $x\in\{0,1\}^N$, a standard Chernoff bound followed by a union bound gives $\Pr[\neg\mathcal{E}_1] = \Pr\left[\exists\, x\in\{0,1\}^N : |R(x)|<qn/2\right] \leqslant 2^N\exp(-qn/8)$. On the other hand, for each choice of distinct $x_1,\ldots,x_t\in\{0,1\}^N$, the quantity $F_k(R(x_1)+\cdots+R(x_t))$ is the kth power sum of n independent binomial random variables. By Lemma 4.4 and a union bound,

$$\Pr[\neg\mathcal{E}_2] \leqslant \binom{2^N}{t}\exp\left(-\frac{cq^ktn}{(\log\log(1/q))^2}\right) \leqslant 2^{Nt}\exp\left(-\frac{cq^ktn}{(\log\log(1/q))^2}\right),$$

for all large enough n and t. Therefore, setting $N = c'q^kn/(\log\log(1/q))^2$ for an appropriate constant c' ensures $\Pr[\neg\mathcal{E}_1\vee\neg\mathcal{E}_2]<1$.

Thus, there exists a specific R at which both $\mathcal{E}_1$ and $\mathcal{E}_2$ occur. For this R, the left-hand side of eq. (3) is at least $(qt^kn/2)/(bq^kt^kn) = 1/(2bq^{k-1})$. We set $q = O(1/\alpha^{1/(k-1)})$ so that this ratio exceeds α. Then eq. (3) is satisfied and R is F_k-separating.

Taking $t=n$ (say) gives us the bound $ps = \Omega(N-\log t) = \Omega(n/(\alpha^{k/(k-1)}(\log\log\alpha)^2))$. □

Next, we handle frequency moments of "lower" order, in a similar fashion.

Theorem 4.7. *For each $k\in[0,1)$ and $\alpha\geqslant 1$, every deterministic s-space p-pass α-estimator for F_k satisfies $ps = \Omega(n/\alpha^{1/(1-k)})$.*

Proof. Again, we prove the existence of an F_k-separating mapping for a large enough $N = N(n,t,\alpha)$, using the same random construction of R. The events of interest are

$$\mathcal{E}_1 := \left\{\max\{F_k(t\cdot R(x)) : x\in\{0,1\}^N\} \leqslant 2qt^kn\right\},$$
$$\mathcal{E}_2 := \left\{\min\{F_k(R(x_1)+\cdots+R(x_t)) : \textsc{eq-dist}_{N,t}(x_1,\ldots,x_t)=0\} \geqslant q^kt^kn/2^{k+1}\right\}.$$

Since $\mathbb{E}|R(x)| = qn$ for each $x\in\{0,1\}^N$, a standard Chernoff bound followed by a union bound gives $\Pr[\neg\mathcal{E}_1] = \Pr\left[\exists\, x\in\{0,1\}^N : |R(x)|>2qn\right] \leqslant 2^N\exp(-qn/3)$. As before, for each choice of distinct $x_1, \ldots, x_t \in \{0,1\}^N$, the quantity $F_k(R(x_1) + \cdots + R(x_t))$ is the kth power sum of n independent binomial random variables. By Lemma 4.5 and a union bound, $\Pr[\neg\mathcal{E}_2] \leqslant 2^{Nt} \exp(-qtn/32)$, for $t \geqslant 32/q$. Therefore, setting $N = qn/64$ ensures $\Pr[\neg\mathcal{E}_1 \vee \neg\mathcal{E}_2] < 1$.

Thus, there exists a specific R at which both $\mathcal{E}_1$ and $\mathcal{E}_2$ occur. For this R, the left-hand side of eq. (4) is at least $(q^k t^k n/2^{k+1})/(2qt^k n) = 1/(q^{1-k}2^{2+k})$. We set $q = O(1/\alpha^{1/(1-k)})$ so that this ratio exceeds α. Then eq. (4) is satisfied and R is F_k-separating.

Again, taking $t = n$ and applying Theorem 3.2 gives $ps = \Omega(N - \log t) = \Omega(n/(\alpha^{1/(1-k)}))$. □

4.3 A Lower Bound for Empirical Entropy

We now turn to the estimation of $\text{ENT}(\mathbf{f})$, the empirical entropy of the input stream. Using the template established in Section 4.2 leads to the following space/approximation tradeoff.

Theorem 4.8. *For every $\varepsilon > 0$ and $\alpha \in [1, o(\log n)]$, every deterministic s-space p-pass α-estimator for $\text{ENT}(\mathbf{f})$ satisfies $ps = \Omega(n^{1/((1+\varepsilon)\alpha)})$.*

Proof. We reduce from $\text{EQ-DIST}_{N,t}$ using a separating mapping whose existence we prove using the same random construction $R\colon \{0,1\}^N \to \{0,1\}^n$ as in the proofs of Theorems 4.6 and 4.7. This causes the players to estimate $\text{ENT}(R(x_1) + \cdots + R(x_t))$, given an input $\mathbf{x} = (x_1, \ldots, x_t)$ for $\text{EQ-DIST}_{N,t}$. Note that, when $\text{EQ-DIST}_{N,t}(\mathbf{x}) = 1$, i.e., $\mathbf{x} = x^{\otimes t}$, then this entropy equals $\text{ENT}(t \cdot R(x)) = \text{ENT}(R(x)) = \log|R(x)|$.

Define the events

$$\begin{aligned}\mathcal{E}_1 &:= \left\{ \max\{\text{ENT}(R(x)) : x \in \{0,1\}^N\} \leqslant \log(2qn) \right\}, \\ \mathcal{E}_2 &:= \left\{ \min\{\text{ENT}(R(x_1) + \cdots + R(x_t)) : \text{EQ-DIST}_{N,t}(x_1, \ldots, x_t) = 0\} \geqslant (\log n)/(1 + \varepsilon/2) \right\}.\end{aligned}$$

By a Chernoff and a union bound, $\Pr[\neg\mathcal{E}_1] \leqslant 2^N \exp(-qn/3)$. On the other hand, by a tail estimate analogous to Lemma 4.5, $\Pr[\neg\mathcal{E}_2] \leqslant 2^{Nt} \exp(-\Omega_\varepsilon(qtn))$, for all large enough t and n. The required tail estimate is formally proved as Lemma 6.4 in Section 6.1.

Setting $N = cqn$, for an appropriate constant c, ensures that $\Pr[\neg\mathcal{E}_1 \vee \neg\mathcal{E}_2] < 1$. So there exists a specific R at which both $\mathcal{E}_1$ and $\mathcal{E}_2$ occur. Using this R, and setting $q = n^{-1+1/((1+\varepsilon)\alpha)}$ gives a gap of $(\log n)/((1 + \varepsilon/2)\log(2qn)) > \alpha$ in the entropy values corresponding to the cases $\text{EQ-DIST}_{N,t}(\mathbf{x}) = 0$ and $\text{EQ-DIST}_{N,t}(\mathbf{x}) = 1$. Thus, an s-space p-pass α-estimator for entropy requires $ps = \Omega(N - 2\log t)$. Taking $t = n$ (say) gives $ps = \Omega(N) = \Omega(n^{1/((1+\varepsilon)\alpha)})$, as required. □

4.4 Some Simple Upper Bounds

Here, we give (very) simple deterministic estimators for the frequency moments. These already suffice to show that our tradeoff lower bounds for F_k are optimal for $0 \leqslant k < 1$ and in the correct ballpark for $k > 1$.

Recall that the input stream is a sequence $(a_1, \ldots, a_m)$ with each $a_j \in [m]$. Crucially, it is an "insert-only" stream, as opposed to a "turnstile" stream. This restriction is reasonable, since all our lower bounds were proved in this same model.

Consider the following one-pass estimator for F_k. Divide the universe $[n]$ into n/β *buckets* of size β each; Let $B_i \subseteq [n]$ be the ith bucket. Maintain the quantity $Q_i := \sum_{j \in B_i} f_j$ for each $i \in [n/\beta]$. This requires only $O(n \log m/\beta)$ bits of space. At the end of the stream,

- if $k > 1$, output $\beta \sum_{i=1}^{n/\beta} (Q_i/\beta)^k$;
- if $0 \leqslant k < 1$, output $\sum_{i=1}^{n/\beta} Q_i^k$.

Let S_i be the contribution of items landing in bucket B_i to the sum defining F_k, i.e., $S_i = \sum_{j \in B_i} f_j^k$. A shifting argument based on convexity and Karamata's inequality (akin to the argument in the proof of Lemma 6.5) can be used to show that

$$\beta(Q_i/\beta)^k \leqslant S_i \leqslant Q_i^k = \beta(Q_i/\beta)^k \beta^{k-1}\,, \quad \text{when } k > 1\,,$$
$$Q_i^k \beta^{1-k} = \beta(Q_i/\beta)^k \geqslant S_i \geqslant Q_i^k\,, \quad \text{when } 0 \leqslant k < 1\,.$$

Therefore, our algorithm is a β^{k-1}-estimator for F_k when $k > 1$ and β^{1-k}-estimator for F_k when $k < 1$. Choosing $\beta = \alpha^{1/(k-1)}$ in the former case and $\beta = \alpha^{1/(1-k)}$ in the latter case gives us an α-estimator in each case.

For estimating F_0, we can dispense with maintaining the quantities Q_i and simply maintain one bit per bucket indicating whether or not $Q_i > 0$.

Finally, we can further cut down the space usage if we are allowed $p > 1$ passes. Simply divide the universe $[n]$ into p equal-sized *ranges* and, for each $\ell \in [p]$, use the ℓth pass to estimate the contribution of the ℓth range to F_k, using the above one-pass procedure.

Putting all of this together, we obtain the following collection of results.

Theorem 4.9. *For integers $p \geqslant 1$, and reals $k \geqslant 0$ and $\alpha \geqslant 1$, there is a family of deterministic p-pass α-estimators for F_k, with the following guarantees on their space usage, s.*

- *When $k = 0$, we have $ps = \lceil n/\alpha \rceil + O(\log n)$.*
- *When $0 < k < 1$, we have $ps = O(n \log m / \alpha^{1/(1-k)})$.*
- *When $k = 1$, at $p = 1$ we have $s \leqslant \lceil \log m \rceil$, trivially.*
- *When $k > 1$, we have $ps = O(n \log m / \alpha^{1/(k-1)})$.* □

Next, we turn to the estimation of the empirical entropy of a stream. Notice that the lower bound in Theorem 4.8 becomes trivial once $\alpha = \Omega(\log n)$. We give another simple estimator to show that this behavior is to be expected.

Theorem 4.10. *There is a deterministic $O(\log m)$-space 2-pass $(1 + \log n)$-estimator for the empirical entropy of a stream.*

Proof. Let m and $\mathbf{f}$ denote the stream's length and its frequency vector, as usual. An element $j \in [n]$ is called a *majority* in the stream if $f_j > m/2$.

The algorithm is as follows. In the first pass, use either the Misra-Gries algorithm with one counter [33] or (equivalently) the Boyer-Moore algorithm [11] to identify a *majority candidate* $M \in [n]$. If the stream does have a majority, then M is guaranteed to be that majority. In the second pass, count f_M precisely to determine whether this is the case. Let $\gamma = f_M/m$.

If the stream has no majority, then $\mathrm{ENT}(\mathbf{f}) \geqslant 1$. Output 1 in this case. Since $\mathrm{ENT}(\mathbf{f}) \leqslant \log n$, this is a $(\log n)$-estimate. Otherwise, let $\mathbf{f}_{-M}$ denote the vector $\mathbf{f}$ restricted to the coordinates in $[n] \setminus \{M\}$. An application of the chain rule for entropy gives us

$$\mathrm{ENT}(\mathbf{f}) = H(\gamma) + (1-\gamma)\mathrm{ENT}(\mathbf{f}_{-M})\,,$$

where $H(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function. Output $H(\gamma)$ in this case. Noting that $\gamma > 1/2$, the approximation ratio is

$$1 + \frac{(1-\gamma)\mathrm{ENT}(\mathbf{f}_{-M})}{H(\gamma)} \leqslant 1 + \frac{(1-\gamma)\log n}{(1-\gamma)\log\frac{1}{1-\gamma}} \leqslant 1 + \log n\,, \text{ since } \gamma \geqslant 1/2. \qquad \square$$

5 Graph Streams

A number of important data stream problems are graph-theoretic. The input graph $G_n = (V, E)$, where $|V| = n$, is described as a stream of edges (the edge-arrival model, our default). It is usually interesting, and nontrivial, to achieve space $\tilde{O}(n)$ for most standard graph computations [23]. We focus on two particular graph problems: maximum matching size estimation (MMSE), described next, and a variant of edge connectivity, described in Section 5.2.

5.1 Maximum Matching Size

The MMSE problem asks for an estimate of the number of edges in a maximum cardinality matching (MCM). For this problem, it is also natural to consider the vertex-arrival model, where the input is a bipartite graph $G_n = (V_1, V_2, E)$, with $|V_2| = n$ and $|V_1| = O(n)$, and the stream lists each vertex $u \in V_1$ with all its neighbors in V_2. This potentially makes an algorithm's task easier, so lower bounds proven in this model are stronger. We prove lower bounds in the vertex-arrival model for s-space p-pass α-estimators for MMSE; our bounds trade off α against the product ps. A closely-related problem, which we call the MCM problem, is that of outputting a large matching.

Previous Work on MMSE. An algorithm that just maintains a maximal matching using $n\lceil\log n\rceil$ space is a 2-estimator for MMSE. Another simple algorithm that just maintains a simple randomized sketch and uses $O(\operatorname{polylog} n)$ space is a $O(\sqrt{n})$-estimator. Kapralov et al. [29] gave a $O(\operatorname{polylog} n)$-estimator over randomly-ordered streams that uses $O(\operatorname{polylog} n)$ space. Esfandiari et al. [21] gave a one-pass randomized $O(\nu)$-estimator that uses $O(\nu n^{2/3})$ space and a two-pass randomized $O(\nu)$-estimator that uses $O(\nu\sqrt{n})$ space for graphs with *arboricity* $\nu = o(\sqrt{n})$, which is defined as $\nu := \max_{U \subseteq V} E(U)/(|U| - 1)$, where $E(U)$ is the set of edges with both endpoints in U. Even when randomization is allowed, no $o(n)$-space α-estimator is known, where α is a constant.

Esfandiari et al. [21] also gave a $\Omega(\sqrt{n})$ space lower bound for randomized one-pass $(3/2 - \varepsilon)$-estimators and $\Omega(n)$ space lower bound for deterministic one-pass $(3/2 - \varepsilon)$-estimators; the latter bound should be compared with that in Theorem 5.2. They obtain these lower bounds by reducing from a communication problem known as *Boolean Hidden Matching (BHM)*, using lower bounds given by Gavinsky et al. [25]. Bury and Schwiegelshohn [14] show that for any constant $\beta \geqslant 2$, one-pass randomized $(1 + 1/(3\beta/2 - 1))$-estimators need space $\Omega(n^{1-1/\beta})$. Note that substituting $\beta = 2$ recovers the lower bound by Esfandiari et al. [21]. They achieve this generalization by using the reduction given by Esfandiari et al. [21] and the lower bound given by Verbin and Yu [36] for the communication problem *Boolean Hidden Hypermatching (BHHM)*, which is a generalization of BHM.

Previous Work on MCM. Since an MCM can have $\Omega(n)$ size, an algorithm needs $\Omega(n \log n)$ space just to output a large matching. As noted earlier, outputting a maximal matching is a 2-approximation algorithm for MCM. We note that no better one-pass deterministic or randomized approximation using $o(n^2)$ space is known. Feigenbaum et al. [23] were the first to study MCM in the streaming model, and they gave a $(3/2 + \varepsilon)$-approximation algorithm. Improving this, McGregor [32] gave a randomized $(1 + \varepsilon)$-approximation algorithm, and Ahn and Guha [1] gave a linear-programming based, deterministic, $(1 + \varepsilon)$-approximation algorithm. Each of the algorithms just mentioned uses $O(n \operatorname{polylog} n)$ space and $O_\varepsilon(1)$ passes.

On the lower-bound side, Goel, Kapralov, and Khanna [26] showed that a one-pass randomized MCM algorithm achieving approximation better than 3/2 must use $n^{1+\Omega(1/\log\log n)}$ space, even in the vertex-arrival model; Kapralov [28] showed that this bound in fact applies to algorithms achieving approximation better than $e/(e-1)$. For exact one-pass algorithms, $\Omega(n^2)$ space is required even if randomization is allowed; this can be proved by a simple reduction from the two-party communication problem of INDEX.

Maximum Matching in the Simultaneous Message (SM) model. Dobzinski, Nisan, and Oren [18] consider a related problem in the SM model. To elaborate, there are n players, and together they hold a bipartite graph $G_n = (V_1, V_2, E)$, with $|V_1| = |V_2| = n$. Each player gets as input the set of neighbors of a vertex in V_1. They send a possibly randomized message to a coordinator simultaneously who has to output a perfect matching, say M, possibly containing edges not in E. The goal is to maximize $|M \cap E|$. They give lower bounds for the maximum message size, say ℓ, where the maximum is taken over all players. They show that for deterministic α-approximation protocols, $\ell = \Omega(n/\alpha)$, and for randomized α-approximation protocols, for any constant $\varepsilon > 0$, $\ell = \Omega(n^{1/2-\varepsilon}/\alpha)$.[3] A careful examination shows that their lower-bound proof for deterministic protocols works for the problem when the coordinator has to just estimate the maximum matching size. Our communication lower-bound techniques for discreet protocols also extend to SM protocols with essentially no change. So the lower bound obtained in Theorem 5.2 below also applies to SM protocols and discreet protocols for MMSE defined appropriately as a communication problem. Thus it generalizes the deterministic lower bound by Dobzinski et al. from a star communication topology (for SM protocols) to arbitrary topology. This, in particular, yields data streaming lower bounds.

Our Results. First, we define a variant of the Equal-vs-Spread problem that we call Equal-vs-Distinct-Representatives. There are $t = \lfloor \gamma n \rfloor$ players. Each player receives a $\lceil \beta n \rceil$-subset of $[n]$ (where $\beta < \gamma$) and they must distinguish the case when all of these subsets are equal from the case when each player can pick a representative element from her subset so that these representatives are distinct. Formally, we use the following function:

$$\text{EQ-DR}^{\beta}_{n,t}(x_1, \ldots, x_t) = \begin{cases} 1\,, & \text{if } |x_1| = \cdots = |x_t| = \lceil \beta n \rceil \text{ and } x_1 = \cdots = x_t\,, \\ 0\,, & \text{if } |x_1| = \cdots = |x_t| = \lceil \beta n \rceil \text{ and } \exists\, g \colon [t] \to x_1 \cup \cdots \cup x_t \\ & \quad \text{such that } g \text{ is injective and } g(i) \in x_i \text{ for each } i \in [t]\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

It is not hard to see that the proof of Theorem 3.3 applies to an analysis of this problem as well, after the substitution $\gamma = t/n$, due to the way in which $\mathbf{y} \in \text{span}(\mathcal{N})$ is constructed in that proof. This leads to the following lower bound.

Theorem 5.1. *For all values $0 < \beta < 1$, $\varepsilon > 0$, and sufficiently large n, if $(\beta + \varepsilon)n \leqslant t < n$, then we have $\text{DD}(\text{EQ-DR}^{\beta}_{n,t}) \geqslant (\beta \log(n/t))n - \log t$.*

We reduce $\text{EQ-DR}^{\beta}_{n,t}$ to MMSE to get lower bounds for α-estimators for MMSE. This improves the $\Omega(n)$ space lower bound for one-pass $(3/2 - \varepsilon)$-estimators, where $\varepsilon > 0$, due to Esfandiari et al. [21]. Though the following theorem is stated for streaming algorithms for MMSE, it is a special case of a more general communication result as noted earlier.

Theorem 5.2. *For a deterministic s-space p-pass α-estimator for MMSE, we have $ps \geqslant ((n/e\alpha)(\log e) - \log n)/2$.*

[3] We can show that their bound can be improved to $\Omega(n/\alpha^2)$ by tweaking the parameters in their proof slightly.

Proof. We reduce from $\text{EQ-DR}^{\beta}_{n,t}$ setting t and β later. The theorem can be proved in the vertex arrival model with $V_1 = \{u_1, \ldots, u_t\}$ and $V_2 = [n]$. Note that the lower bounds for vertex-arrival model also apply for the edge-arrival model. For $i \in [t]$, PLR_i adds edges $\{\{u_i, j\} : j \in x_i\}$. Call the resulting graph G_n. When $\text{EQ-DR}^{\beta}_{n,t}(x_1, \ldots, x_t) = 1$, an MCM in G_n has size βn. When $\text{EQ-DR}^{\beta}_{n,t}(x_1, \ldots, x_t) = 0$, an MCM in G_n has size t, because the definition of $\text{EQ-DR}^{\beta}_{n,t}$ guarantees existence of an injective mapping g from $[t]$ to $[n]$. So, if β and t are such that $t/\beta n > \alpha$, then a deterministic s-space p-pass α-estimator for MMSE can be used to give a $2ps$-bounded discreet protocol $\text{EQ-DR}^{\beta}_{n,t}$. Setting $t = n/e$ and $\beta = 1/(\alpha e) - 1/n$ optimizes the lower bound we get by Theorem 5.1, i.e., we get $2ps \geqslant (\beta \log(n/t))n - \log t = (1/(\alpha e) - 1/n)(\log e)n - \log n + \log e$; after simplification, this gives us the desired result. □

We note that we *can* reduce $\text{EQ-SPRD}^{\beta,\gamma}_{n,t}$ to MMSE by making each player add $\lceil \beta n \rceil$ vertices, but that reduction gives a much weaker bound than that in Theorem 5.2.

5.2 Edge Connectivity

We divert from multi party to two party communication complexity in this section. The dynamic graph connectivity problem XCONN is as follows. There are two players, Alice and Bob, who get inputs E_A and E_B which are sets of edges on the vertex set $[n]$. For two sets S and T, denote by $S \oplus T$ the set $(S \cup T) \setminus (S \cap T)$. Alice and Bob communicate to determine whether the graph $E_A \oplus E_B$ is connected.

We reduce $\text{EQ}_{n^2/4}$ to XCONN, where EQ is the well-known two-party equality problem. Alice adds a complete graph on $[n/2]$, Bob adds a complete graph on $[n] \setminus [n/2]$, and they encode the inputs for $\text{EQ}_{n^2/4}$ within the edges in $[n/2] \times ([n] \setminus [n/2])$. In case of equality, XCONN will evaluate to false, otherwise XCONN will evaluate to true; hence, communication complexity of XCONN is at least $n^2/4$.

There is a randomized protocol for XCONN. Alice can send Bob the sketch for connectivity given by Ahn, Guha, and McGregor [2] of size $O(n \log^3 n)$. Bob can solve XCONN using this sketch. This separates the randomized and deterministic communication complexity of XCONN.

By using error correcting codes (ECC), we can show that even the following version of XCONN with a strong promise is hard. Alice and Bob get inputs E_A and E_B with the promise that the graph $(E_A \cup E_B) \setminus (E_A \cap E_B)$ is disconnected or $(n/2-1)$-connected, i.e., at least $n/2-1$ edges need to be removed to disconnect it. We reduce from EQ_{N^2} where $N = \Omega(n)$. We use a binary ECC of size 2^{N^2}, block length $n^2/4$, and distance $n/2-1$. By Shannon's construction, we can construct such an ECC with $N = \Omega(n)$. Then we use the same construction as in the reduction from $\text{EQ}_{n^2/4}$ to XCONN to get E_A and E_B. In case of equality, $E_A \oplus E_B$ will be disconnected (no edge from $[n/2]$ to $[n] \setminus [n/2]$). In case of inequality, there will be at least $n/2-1$ edges from $[n/2]$ to $[n] \setminus [n/2]$. Since $E_A \oplus E_B$ has a complete graph within $[n/2]$ and within $[n] \setminus [n/2]$, it is at least $(n/2-1)$-connected. Hence, the communication complexity of strong-promise version of XCONN is at least $N^2 = \Omega(n^2)$.

6 Proofs of Technical Lemmas

In this section, we prove the technical lemmas used in obtaining results in earlier sections. Our proof of an upper-tail bound on power sums of independent binomial random variables is particularly instructive and worth understanding for its own sake. Though basic-looking, this result appears to be novel. As noted before, other seemingly basic questions about moments of the binomial distribution were addressed only recently [8].

6.1 Power Sums of Binomial Random Variables

Let $Y_1, \ldots, Y_n$ be independent random variables, each with binomial distribution $\mathcal{B}(t,q)$. Define $Z = Z(n,t,q,k) = Y_1^k + \cdots + Y_n^k$, the kth power sum of this collection. We shall prove the following theorem, which was used as a key technical lemma in establishing Theorem 4.6, our lower bound for higher-order frequency moments.

Theorem 6.1 (Restatement of Lemma 4.4)**.** *For each $k > 1$, there exist $b, c > 0$ such that the following holds. For each $q \in (0, 1/(2e^2))$, there exist integers n_0 and t_0 such that, for all $n \geqslant n_0$ and $t \geqslant t_0$,*

$$\Pr[Z > bq^k t^k n] \leqslant \exp\left(-\frac{cq^k tn}{(\log\log(1/q))^2}\right).$$

Though Z is a sum of independent, bounded random variables, this tail bound does not follow from Chernoff-Hoeffding and Azuma-Hoeffding inequalities [5]; those give a much weaker upper bound of the form $\exp(-\Theta(n))$. Indeed, even a bound of $\exp(-\Theta(tn))$ would not be strong enough to prove Theorem 4.6. We need to understand how the coefficient in front of tn depends on q, and this seems to require a delicate partitioning of the large deviation event.

Our proof proceeds by partitioning the random variables $\{Y_j\}$—or rather, their indices—into buckets: indices j placed in a "high" bucket correspond to Y_js whose realization is much higher than the expected value of qt. Intuitively, the higher the bucket, the fewer the indices we expect to land in that bucket, but also, the fewer the indices we can *afford* to have land in that bucket so that the power sum Z stays small. The trick is to define buckets and events involving their sizes with parameters chosen carefully enough to balance these two effects.

We shall make use of the following form of the Chernoff bound.

Lemma 6.2. *Suppose that $0 < q_1 < q < \kappa$. Let $Y \sim \mathcal{B}(t, q_1)$ for some integer $t \geqslant 1$. Then*

$$\Pr[Y \geqslant \kappa t] \leqslant \exp\left(-\kappa t \ln\frac{\kappa}{eq} - qt\right) \leqslant \exp\left(-\kappa t \ln\frac{\kappa}{eq}\right). \qquad \square$$

Proof of Lemma 4.4. Put $S(\kappa, \lambda) = \{j \in [n] : \kappa t \leqslant Y_j \leqslant \lambda t\}$. Let $\theta \in (2^{-k}, 1/2)$ be a parameter to be determined. Define the events

$$\begin{aligned}\mathcal{E}_0 &:= \left\{|S(\theta,1)| > q^k n\right\}, \\ \mathcal{E}_m &:= \left\{\left|S\left(\theta^{k^m}, \theta^{k^{m-1}}\right)\right| > m^{-2} q^k \theta^{-k^m} n\right\}, \quad \text{for each integer } m \geqslant 1.\end{aligned}$$

We choose θ so that there exists an integer $M \geqslant 0$ such that $\theta^{k^M} = e^2 q$. Indeed, since $q < 1/(2e^2)$, θ is the unique member of the sequence $e^2 q, (e^2 q)^{1/k}, (e^2 q)^{1/k^2}, \ldots$ that lies in $(2^{-k}, 1/2)$. If none of the events $\mathcal{E}_0, \ldots, \mathcal{E}_M$ occur, then

$$\begin{aligned}Z &\leqslant \left(\sum_{j \in S(\theta,1)} Y_j^k\right) + \left(\sum_{m=1}^{M} \sum_{j \in S(\theta^{k^m}, \theta^{k^{m-1}})} Y_j^k\right) + \left(\sum_{j \in S(0, e^2 q)} Y_j^k\right) \\ &\leqslant |S(\theta,1)| t^k + \left(\sum_{m=1}^{M} \left|S\left(\theta^{k^m}, \theta^{k^{m-1}}\right)\right| (\theta^{k^{m-1}} t)^k\right) + |S(0, e^2 q)|\,(e^2 qt)^k \\ &\leqslant q^k n t^k + \left(\sum_{m=1}^{M} (m^{-2} q^k \theta^{-k^m} n)(\theta^{k^m} t^k)\right) + n(e^2 qt)^k\end{aligned}$$

$$\begin{aligned}
&= \left(1 + e^{2k} + \sum_{m=1}^{M} m^{-2}\right) q^k t^k n \\
&\leqslant b q^k t^k n\,,
\end{aligned}$$

for some constant b depending on k, where the last step uses the convergence of $\sum_{m=1}^{\infty} m^{-2}$.

We now seek good bounds on the probabilities of these events $\mathcal{E}_m$. For this, we first bound the tails of $|S(\kappa,\lambda)|$. Note that $|S(\kappa,\lambda)| \leqslant |S(\kappa,1)| = \sum_{j=1}^{n} \mathbb{1}_{\{Y_j \geqslant \kappa t\}} \sim \mathcal{B}(n, \Pr[Y_1 \geqslant \kappa t])$. Invoking Lemma 6.2, we obtain that, for $e^2 q \leqslant \kappa \leqslant 1$,

$$\Pr[Y_1 \geqslant \kappa t] \leqslant \exp\left(-\kappa t \ln \frac{\kappa}{eq}\right) \leqslant \exp(-\kappa t)\,. \tag{5}$$

By design, $e^2 q \leqslant \theta^{k^m} \leqslant 1$ for each m with $0 \leqslant m \leqslant M$. So, by another invocation of Lemma 6.2, for each $m \in [M]$, we have

$$\begin{aligned}
\Pr[\mathcal{E}_m] &\leqslant \Pr\left[\left|S(\theta^{k^m},1)\right| > m^{-2} q^k \theta^{-k^m} n\right] \\
&= \Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j \geqslant \theta^{k^m} t\}} > m^{-2} q^k \theta^{-k^m} n\right] \\
&\leqslant \exp\left(-m^{-2} q^k \theta^{-k^m} n \cdot \ln \frac{m^{-2} q^k \theta^{-k^m}}{e \cdot \exp(-\theta^{k^m} t)}\right), && \text{using Lemma 6.2 and (5)}\,, \\
&= \exp(-m^{-2} q^k n t + n x \ln(e/x))\,, && \text{where } x = m^{-2} q^k \theta^{-k^m} \in (0,1)\,, \\
&\leqslant \exp(-M^{-2} q^k t n + n)\,, && \text{using } x \ln(e/x) \leqslant 1\,.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr[\mathcal{E}_0] &= \Pr\left[|S(\theta,1)| > q^k n\right] \\
&= \Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j \geqslant \theta t\}} > q^k n\right] \\
&\leqslant \exp\left(-q^k n \cdot \ln \frac{q^k}{e \cdot \exp(-\theta t)}\right), && \text{using Lemma 6.2 and (5)}\,, \\
&= \exp(-\theta q^k n t + n q^k \ln(e/q^k)) \\
&\leqslant \exp(-\theta q^k t n + n)\,, && \text{using } q^k \ln(e/q^k) \leqslant 1\,.
\end{aligned}$$

Therefore, applying a union bound,

$$\Pr[Z > b q^k t^k n] \leqslant \Pr\left[\bigvee_{m=0}^{M} \mathcal{E}_m\right] \leqslant \sum_{m=0}^{M} \Pr[\mathcal{E}_m] \leqslant M \cdot \exp(-M^{-2} q^k t n + n) + \exp(-\theta q^k t n + n)\,.$$

By definition, $M = \log_k \log_{1/\theta}(1/e^2 q) \leqslant \log_k \log(1/q)$. So, taking n and t large enough,

$$\Pr[Z > b q^k t^k n] \leqslant \exp(-c' M^{-2} q^k t n) \leqslant \exp\left(-\frac{c q^k t n}{(\log\log(1/q))^2}\right),$$

for suitably chosen constants c' and c, dependent on k alone and not on q, t, and n. □

Next, we bound the lower tail of the distribution of the same random variable, Z. This turns out to be considerably more straightforward than bounding the upper tail. The lower-tail bound was used as a key lemma in the proof of Theorem 4.7.

Lemma 6.3 (Restatement of Lemma 4.5)**.** *For each $k \geqslant 0$, each $q \in (0,1)$, and each integer $t \geqslant 32/q$, there exists an integer n_0 such that, for all $n \geqslant n_0$,*

$$\Pr[Z < q^k t^k n/2^{k+1}] \leqslant \exp(-qtn/32)\,.$$

Proof. By a Chernoff bound, $\Pr[Y_j < qt/2] \leqslant \exp(-qt/8)$, for $j \in [n]$. Using Lemma 6.2,

$$\begin{aligned}\Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j<qt/2\}} > n/2\right] &\leqslant \exp\left(-(n/2)\cdot\ln\frac{1/2}{e\cdot\exp(-qt/8)}\right)\\ &= \exp(-(n/2)\cdot(qt/8-\ln(2e)))\\ &\leqslant \exp(-(n/2)\cdot(qt/16))\,, \qquad \text{because } qt/16 \geqslant 2 \text{ and } \ln(2e) \leqslant 2,\\ &= \exp(-qtn/32)\,. \qquad (6)\end{aligned}$$

Notice that $\sum_{j=1}^{n} \mathbb{1}_{\{Y_j \geqslant qt/2\}} \geqslant n/2$ implies $Z \geqslant (qt/2)^k n/2$. Therefore,

$$\begin{aligned}1-\Pr[Z < q^k t^k n/2^{k+1}] &= \Pr[Z \geqslant (qt/2)^k n/2]\\ &\geqslant \Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j\geqslant qt/2\}} \geqslant n/2\right]\\ &= 1-\Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j<qt/2\}} > n/2\right]\\ &\geqslant 1-\exp(-qtn/32)\,,\end{aligned}$$

where we use Equation (6) in the last step. Rearranging the above gives the desired bound. □

In the spirit of the above tail bounds, we establish a tail bound for the empirical entropy of an ensemble of independent binomial random variables.

Lemma 6.4. *For any constant $\varepsilon > 0$, $q \in (0,1)$, and $t \geqslant 32/q$, we have $\Pr[\textsc{ent}((Y_1,\ldots,Y_n)) < (\log n)/(1+\varepsilon)] \leqslant \exp(-\Omega(qtn))$.*

Proof. For ease of exposition, we shall instead prove the following weaker form of the lemma: For each $q \in (0,1)$ and $t \geqslant 32/q$, $\Pr[\textsc{ent}((Y_1,\ldots,Y_n)) < (\log(n/4))/8] \leqslant \exp(-qtn/40)$.

By a Chernoff bound, $\Pr[\sum_{j\in[n]} Y_j > 2qtn] \leqslant \exp(-qtn/3)$, and $\Pr[Y_j \notin [qt/2, 2qt]] \leqslant \exp(-qt/8)$, for $j \in [n]$. Now, using Lemma 6.2,

$$\begin{aligned}\Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\{Y_j\notin[qt/2,2qt]\}} > n/2\right] &\leqslant \exp\left(-(n/2)\cdot\ln\frac{1/2}{e\cdot\exp(-qt/8)}\right)\\ &= \exp(-(n/2)\cdot(qt/8-\ln(2e)))\\ &\leqslant \exp(-(n/2)\cdot(qt/16))\,, \qquad \text{because } qt/16 \geqslant 2 \text{ and } \ln(2e) \leqslant 2,\\ &= \exp(-qtn/32)\,.\end{aligned}$$

If $\sum_{j\in[n]} Y_j \leqslant 2qtn$ and $\sum_{j=1}^{n} \mathbb{1}_{\{Y_j \in [qt/2, 2qt]\}} \geqslant n/2$, then

$$\text{ENT}((Y_1, \dots, Y_n)) \geqslant \frac{n}{2}\frac{qt/2}{2qtn}\log\frac{qtn/2}{2qt} = \frac{1}{8}\log\frac{n}{4}\,.$$

Hence,

$$\Pr[\text{ENT}((Y_1, \dots, Y_n)) \geqslant (\log(n/4))/8] \geqslant \Pr\left[\sum_{j\in[n]} Y_j \leqslant 2qtn \wedge \sum_{j=1}^{n} \mathbb{1}_{\{Y_j \in [qt/2,2qt]\}} \geqslant n/2\right]\,,$$

which, after taking complements of the events, gives

$$\Pr[\text{ENT}((Y_1, \dots, Y_n)) < (\log(n/4))/8] \leqslant \Pr\left[\sum_{j\in[n]} Y_j > 2qtn \vee \sum_{j=1}^{n} \mathbb{1}_{\{Y_j \notin [qt/2,2qt]\}} > n/2\right]\,.$$

Using the bounds on the probabilities for the events on the right hand side of the inequality above, we get the desired result. □

6.2 A Convexity Lemma

Our proof of Theorem 4.2 and the analysis of our algorithms leading to Theorem 4.9 relied on bounds on the ℓ_k-norm of a vector under certain technical conditions on its coordinates. We now give a detailed proof of one such bound, namely, Lemma 4.3.

Lemma 6.5 (Restatement of Lemma 4.3)**.** *Let $g\colon \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function with $g(0) = 0$, and let $\mathbf{f} \in \{0, 1, \dots, t\}^n$ where $t \geqslant 2$. Suppose that $F_1(\mathbf{f}) = m$, $F_0(\mathbf{f}) \geqslant r$, and $rt \geqslant m$. Then*

$$\sum_{i=1}^{n} g(f_i) \leqslant \ell g(t) + (r-\ell)g(1)\,, \tag{7}$$

where $\mathbf{f} = (f_1, \dots, f_n)$ and $\ell = \lceil (m-r)/(t-1)\rceil$.

Proof. Assume WLOG that $f_1 \geqslant f_2 \geqslant \cdots \geqslant f_n$. Note that, by the conditions of the lemma, $\ell \leqslant \lceil (rt-r)/(t-1)\rceil = r$. Let $\mathbf{f}^* = (f_1^*, \dots, f_n^*)$ be the unique vector such that

$$f_1^* = \cdots = f_{\ell-1}^* = t\,, \quad f_{\ell+1}^* = \cdots = f_r^* = 1\,, \quad f_{r+1}^* = \cdots = f_n^* = 0\,, \quad \text{and} \quad F_1(\mathbf{f}^*) = m\,.$$

We can see that $\mathbf{f}^* \in \{0, 1, \dots, t\}^n$ and $f_1^* \geqslant f_2^* \geqslant \cdots \geqslant f_n^*$ from the computation

$$f_\ell^* = m - (\ell-1)t - (r-\ell) = t - \left(\left\lceil \frac{m-r}{t-1}\right\rceil - \frac{m-r}{t-1}\right)(t-1)\,.$$

We now claim that the vector $\mathbf{f}^*$ majorizes $\mathbf{f}$. Assuming this claim, we then have

$$\sum_{i=1}^{n} g(f_i) \leqslant \sum_{i=1}^{n} g(f_i^*) = (\ell-1)g(t) + g(f_\ell^*) + (r-\ell)g(1) + (n-r)g(0) \leqslant \ell g(t) + (r-\ell)g(1)\,,$$

where the first step uses Karamata's inequality and the convexity of g, and the last step uses the facts that g is nondecreasing and $g(0) = 0$. This proves eq. (7), as required.

To prove our claim, we first note that $\sum_{i=1}^{n} f_i^* = m = \sum_{i=1}^{n} f_i$, by construction. Suppose, to the contrary, that $\mathbf{f}^*$ does not majorize $\mathbf{f}$. Then let $j \in [n-1]$ be the smallest index such that

$\sum_{i=1}^{j} f_i > \sum_{i=1}^{j} f_i^*$. Now, j cannot be less than ℓ because $f_i^* = t$ for $i < \ell$. Also, j cannot be greater than r because $\sum_{i=1}^{r} f_i^* = m$. Thus, $\ell \leqslant j \leqslant r$. We then have

$$\begin{aligned}
F_1(\mathbf{f}) &\geqslant \sum_{i=1}^{j} f_i + \sum_{i=j+1}^{r} f_i\,, && \text{since } F_0(\mathbf{f}) \geqslant r, \\
&> \sum_{i=1}^{j} f_i^* + \sum_{i=j+1}^{r} f_i\,, && \text{by assumption,} \\
&\geqslant \sum_{i=1}^{j} f_i^* + \sum_{i=j+1}^{r} 1\,, && \text{because } F_0(\mathbf{f}) \geqslant r \text{ implies that } f_{j+1}, \ldots, f_r \geqslant 1, \\
&= \sum_{i=1}^{n} f_i^*\,, && \text{because } j \geqslant \ell, \text{ implying } f_{j+1}^* = \cdots = f_r^* = 1, \\
&= m\,,
\end{aligned}$$

which is a contradiction. This completes the proof of the claim, and hence, the lemma. □

References

[1] K. J. Ahn and S. Guha. Linear programming in the semi-streaming model with application to the maximum matching problem. *Inf. Comput.*, 222:59–79, 2013.

[2] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *Proc. 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 459–467, 2012.

[3] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999. Preliminary version in *Proc. 28th Annual ACM Symposium on the Theory of Computing*, pages 20–29, 1996.

[4] N. Alon, A. Moitra, and B. Sudakov. Nearly complete graphs decomposable into large induced matchings and their applications. In *Proc. 44th Annual ACM Symposium on the Theory of Computing*, pages 1079–1090, 2012.

[5] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley-Interscience, New York, NY, 2000.

[6] A. Ambainis. Communication complexity in a 3-computer model. *Algorithmica*, 16(3):298–301, 1996.

[7] C. J. Arackaparambil, J. Brody, and A. Chakrabarti. Functional monitoring without monotonicity. In *Proc. 36th International Colloquium on Automata, Languages and Programming*, pages 95–106, 2009.

[8] Árpád Bényi and S. M. Manago. A recursive formula for moments of a binomial distribution. *The College Mathematics Journal*, 36(1):68–72, 2005.

[9] L. Babai and P. G. Kimmel. Randomized simultaneous messages: Solution of a problem of Yao in communication complexity. In *Proc. 12th Annual IEEE Conference on Computational Complexity*, pages 239–246, 1997.

[10] R. Bottesch, D. Gavinsky, and H. Klauck. Equality, revisited. In *Proc. 40th International Symposium on Mathematical Foundations of Computer Science*, pages 127–138, 2015.

[11] R. S. Boyer and J. S. Moore. MJRTY — a fast majority vote algorithm. Technical Report ICSCA-CMP-32, University of Texas at Austin, 1982.

[12] M. Braverman. Interactive information complexity. In *Proc. 44th Annual ACM Symposium on the Theory of Computing*, pages 505–524, 2012.

[13] J. Brody, A. Chakrabarti, R. Kondapally, D. P. Woodruff, and G. Yaroslavtsev. Certifying equality with limited interaction. In *Proc. 18th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 545–581, 2014.

[14] M. Bury and C. Schwiegelshohn. Sublinear estimation of weighted matchings in dynamic data streams. In *Proc. 23rd Annual European Symposium on Algorithms*, pages 263–274, 2015.

[15] A. Chakrabarti, Y. Shi, A. Wirth, and A. C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[16] A. Chattopadhyay, J. Radhakrishnan, and A. Rudra. Topology matters in communication. In *Proc. 55th Annual IEEE Symposium on Foundations of Computer Science*, pages 631–640, 2014.

[17] G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed functional monitoring. In *Proc. 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1076–1085, 2008.

[18] S. Dobzinski, N. Nisan, and S. Oren. Economic efficiency requires interaction. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 233–242, 2014.

[19] P. Erdős, P. Frankl, and Z. Füredi. Families of finite sets in which no set is covered by the union of r others. *Israel J. Math.*, 51:79–89, 1985.

[20] F. Ergün and H. Jowhari. On distance to monotonicity and longest increasing subsequence of a data stream. In *Proc. 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 730–736, 2008.

[21] H. Esfandiari, M. T. Hajiaghayi, V. Liaghat, M. Monemizadeh, and K. Onak. Streaming algorithms for estimating the matching size in planar graphs and beyond. In *Proc. 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1217–1233, 2015.

[22] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan. Amortized communication complexity. *SIAM J. Comput.*, 24(4):736–750, 1995. Preliminary version in *Proc. 32nd Annual IEEE Symposium on Foundations of Computer Science*, pages 239–248, 1991.

[23] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. *Theor. Comput. Sci.*, 348(2–3):207–216, 2005. Preliminary version in *Proc. 31st International Colloquium on Automata, Languages and Programming*, pages 531–543, 2004.

[24] A. Gál and P. Gopalan. Lower bounds on streaming algorithms for approximating the length of the longest increasing subsequence. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 294–304, 2007.

[25] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008.

[26] A. Goel, M. Kapralov, and S. Khanna. On the communication and streaming complexity of maximum bipartite matching. In *Proc. 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 468–485, 2012.

[27] D. M. Kane, J. Nelson, and D. P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *Proc. 21st Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1161–1178, 2010.

[28] M. Kapralov. Better bounds for matchings in the streaming model. In *Proc. 24th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1679–1697, 2013.

[29] M. Kapralov, S. Khanna, and M. Sudan. Approximating matching size from random streams. In *Proc. 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 734–751, 2014.

[30] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 1997.

[31] G. Liang and N. H. Vaidya. Multiparty equality function computation in networks with point-to-point links. In *Proc. 18th International Colloquium on Structural Information and Communication Complexity*, pages 258–269, 2011.

[32] A. McGregor. Finding graph matchings in data streams. In *Proc. 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 170–181, 2005.

[33] J. Misra and D. Gries. Finding repeated elements. *Sci. Comput. Program.*, 2(2):143–152, 1982.

[34] M. Molinaro, D. Woodruff, and G. Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *Proc. 24th Annual ACM-SIAM Symposium on Discrete Algorithms*, page to appear, 2013.

[35] J. M. Phillips, E. Verbin, and Q. Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *Proc. 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 486–501, 2012.

[36] E. Verbin and W. Yu. The streaming complexity of cycle counting, sorting by reversals, and other problems. In *Proc. 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 11–25, 2011.

[37] D. P. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. In *Proc. 43rd Annual ACM Symposium on the Theory of Computing*, pages 941–960, 2012.

[38] D. P. Woodruff and Q. Zhang. When distributed computation is communication expensive. In *Proc. 27th International Symposium on Distributed Computing*, pages 16–30, 2013.

[39] D. P. Woodruff and Q. Zhang. An optimal lower bound for distinct elements in the message passing model. In *Proc. 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 718–733, 2014.

[40] A. C. Yao. Some complexity questions related to distributive computing. In *Proc. 11th Annual ACM Symposium on the Theory of Computing*, pages 209–213, 1979.