THEORY OF COMPUTING, Volume 19 (10), 2023, pp. 1–44
www.theoryofcomputing.org

Separating k-Player from t-Player One-Way Communication, with Applications to Data Streams

Elbert Du Michael Mitzenmacher* David Woodruff
Guang Yang

Received September 25, 2021; Revised January 8, 2023; Published December 31, 2023

Abstract. In a k-party communication problem, the k players with inputs $x_1, x_2, \ldots, x_k$ want to evaluate a function $f(x_1, x_2, \ldots, x_k)$ using as little communication as possible. We consider the message-passing model, in which the inputs are partitioned in an arbitrary, possibly worst-case manner, among a smaller number t of players ($t < k$). The t-player communication cost of computing f can only be smaller than the k-player communication cost, since the t players can trivially simulate the k-player protocol. But how much smaller can it be? We study deterministic and randomized protocols in the one-way model, and provide separations for product input distributions, which are optimal for low error probability protocols. We also provide much stronger separations when the input distribution is non-product.

A preliminary version of this paper, by a subset of the authors, appeared in the Proceedings of the 46th International Colloquium on Automata, Languages and Programming, 2019 (ICALP'19) [24].

*Supported in part by NSF grants CCF-2101140, CNS-2107078, and DMS-2023528, and by a gift to the Center for Research on Computation and Society at Harvard University.

ACM Classification: F.1.3, F.2.3, F.2.1

AMS Classification: 68Q17, 68W25, 68W20

Key words and phrases: streaming, space complexity, hamming norm, approximation, algorithms with predictions, direct sum

© 2023 Elbert Du, Michael Mitzenmacher, David Woodruff, and Guang Yang
Licensed under a Creative Commons Attribution License (CC-BY) DOI: 10.4086/toc.2023.v019a010

A key application of our results is in proving lower bounds for data stream algorithms. In particular, we give an optimal $\Omega(\varepsilon^{-2}\log(N)\log\log(mM))$ bits of space lower bound for the fundamental problem of $(1\pm\varepsilon)$-approximating the number $\|x\|_0$ of non-zero entries of an n-dimensional vector x after m integer updates each of magnitude at most M, and with success probability $\geq 2/3$, in a strict turnstile stream. We additionally prove the matching $\Omega(\varepsilon^{-2}\log(N)\log\log(T))$ space lower bound for the problem when we have access to a heavy hitters oracle with threshold T. Our results match the best known upper bounds when $\varepsilon \geq 1/\mathsf{polylog}(mM)$ and when $T = 2^{\mathsf{poly}(1/\epsilon)}$, respectively. It also improves on the prior $\Omega(\varepsilon^{-2}\log(mM))$ lower bound and separates the complexity of approximating L_0 from approximating the p-norm L_p for p bounded away from 0, since the latter has an $O(\varepsilon^{-2}\log(mM))$ bit upper bound.

1 Introduction

Consider a k-party communication problem, in which the players have inputs $x_1, x_2, \ldots, x_k$ and want to compute a function $f(x_1, x_2, \ldots, x_k)$ of their inputs using as little communication as possible. We consider the message-passing model, in which the inputs are partitioned in an arbitrary, possibly worst-case manner among a smaller number t of players. That is, we partition $\{1, 2, \ldots, k\}$ into t subsets $S_1, S_2, \ldots, S_t$ such that $\bigcup_{i=1}^{t} S_i = \{1, 2, \ldots, k\}$ and $S_i \cap S_j = \emptyset$ for every $1 \leq i < j \leq t$, and let the i-th player P_i hold the sequence of inputs $y_i := \left(x_{i_1}, x_{i_2}, \ldots, x_{i_{|S_i|}}\right)$. When we work in this model in the reduction from streaming, we will get $i_1 + |S_i| - 1 = i_2 + |S_i| - 2 = \cdots = i_{|S_i|}$. We are still interested in computing the original function f. The total communication required must be smaller than in the original k-player setting, since the t players can simulate the protocol involving the original k players. A natural question is: *how much smaller can the communication be?*

There are many communication models that are possible, but our main motivation for looking at this question comes from applications to data streams, see below, and so we are primarily interested in the *one-way number-in-hand* model. In this model, each of the t players can only see its own input. The first player composes a message m_1 based on its input y_1 and sends m_1 to the second player. The second player takes m_1 and its input y_2 to compute a message m_2 for the third player, and so on. The t-th (also the last) player, upon receiving the message m_{t-1} from the $(t-1)$-st player, computes the output of the protocol based on m_{t-1} and its own input y_t. We sometimes abuse notation and refer to the output as m_t. The total communication cost is the maximum of $\sum_{i=1}^{t} |m_i|$, where $|m_i|$ denotes the length of the i-th message and the maximum is taken over all possible inputs $y_1, \ldots, y_t$ (which is a partition of $\{x_1, \ldots, x_k\}$) and all random coin tosses of the players. For streaming applications we are especially interested in $\max_{i\in\{1,\ldots,t\}} |m_i|$.

To explain the connection to data streams, almost all known lower bound arguments on the memory required of a data stream algorithm are proven via communication complexity, or at least can be reformulated using communication complexity. The basic idea is to partition the elements of an input stream contiguously, consisting of say k elements, into a possibly smaller number t of players. Then one argues that if there is a data stream algorithm solving the problem, then the communication problem can be solved by passing the memory contents as messages from player to player. Note that this naturally gives rise to the one-way number-in-hand model. Since the total communication cost is $t \cdot S$, where S is the size of the memory of the streaming algorithm, if the randomized t-player communication complexity of the function f is CC_t, we must have $S \geq CC_t/t$. Many lower bounds in data streams are proven already with two players. However, it is known that for some functions more players are needed to obtain stronger lower bounds, such as for estimating the frequency moments in insertion only streams (see, e. g., [3, 25] and references therein).

One cannot help but ask *how powerful is communication complexity for proving data stream lower bounds*? Another natural question is: *for a given function f, which number t of players should one partition the stream into*? Yet another question is regarding the input distribution – should it be a product distribution for which the inputs to the players are chosen independently, or should the inputs be drawn from a non-product distribution to obtain the best space lower bounds? Since we are interested in the limits of using t players for establishing lower bounds for data stream algorithms, we allow the original k inputs (which correspond to the k elements in a stream) to be partitioned in the worst possible way for a t-player communication protocol, as this will give the strongest possible lower bound.

1.1 Our results

In this paper we study these communication questions and their connections to data streams.

We first make the simple observation that for non-product input distributions, the communication complexity can be arbitrarily smaller if we partition the k inputs into $t < k$ players. Indeed, consider the k-player set disjointness problem in which the i-th player, $1 \leq i \leq k$, has a set $S_i \subseteq [n]$, where for notational simplicity we define $[n] := \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$. The input distribution satisfies the promise that either (1) $S_i \cap S_j = \emptyset$ for every $1 \leq i < j \leq k$, or (2) there is a unique item $a \in [n]$ such that $a \in S_i$ for all $i \in [k]$, and for any other $a' \neq a$, there is at most one $i \in [k]$ for which $a' \in S_i$. It is well-known that the randomized communication complexity of this problem is $\Omega(n/k)$ [3, 9, 12], and that the bound holds even for multiple rounds of communication and when players share a common blackboard. However, if we look at $t < k$ players and an arbitrary, even if the worst-case mapping of the input sets $S_1, \ldots, S_k$ to the t players, then by the pigeonhole principle there exists a player who gets two input sets S_i, S_j with $i \neq j$. Now this player can locally determine the output of the function by checking if $S_i \cap S_j = \emptyset$. Thus with $t < k$ players the problem is solvable using $O(1)$ bits per player. This simple argument shows that for non-product distributions, there can be an arbitrarily large gap between the k-player and the t-player worst-case-partitioned randomized communication complexities. Note that this example applies to a symmetric problem, meaning that the k-player set disjointness problem is invariant under any one-to-one assignment of $x_1, \ldots, x_k$ to the k players.

Perhaps surprisingly, and this is one of the main messages of our work: for symmetric functions and product input distributions, we show that for any $t < k$, for deterministic one-way communication complexity or randomized one-way communication complexity with error probability $1/\mathsf{poly}(k)$, that is, the gap between the k-player and t-player communication complexities is at most a multiplicative $O(1)$ factor in maximum message length, or the maximum communication from a single player, and $O(k)$ in total communication. Further, this gap is tight, as there are problems for which the input distribution is a product distribution, and the t-player communication with $1/\mathsf{poly}(k)$ error probability is $O(\log k)$ for constant $t = O(1)$, while the k-player communication with $1/\mathsf{poly}(k)$ error probability is $\Omega(k \log k)$.

Thus, the answer for product input distributions is significantly different than what we saw for non-product distributions, even for symmetric functions.

We also show that for protocols with constant error and under product input distributions, the gap is at most a multiplicative $O(\log k)$ factor in message length and $O(k \log k)$ in total communication. Further, we show that there exists a symmetric function and input distribution which is product on any $k-1$ out of k inputs, for which this gap is best possible. We leave open the question of the existence of a symmetric function and product input distribution (on all k inputs rather than $k-1$ out of k) which realizes this gap for constant error protocols.

One takeaway message from our results is that when showing space lower bounds for data stream algorithms computing symmetric functions on product distributions, by looking at 2-player communication complexity (which is by far the most common communication setup), there is only an $O(1)$ factor loss for error probability $1/\mathsf{poly}(k)$ protocols, and an $O(\log k)$ factor loss for constant error protocols.

However, for non-product distributions, which are often needed to show hardness of approximation in data streams (such as for the frequency moments [3]), one may need to use as many as k players in order to obtain a non-trivial lower bound from communication complexity.

1.1.1 Data stream lower bounds:

As a key application of our lower bound techniques, we provide a space lower bound for $(1 \pm \varepsilon)$-approximating the *Hamming norm* in the strict turnstile model. This problem, which is also known as the L_0 *norm estimation* and denoted by T_ε, requires estimating $\|\mathbf{x}\|_0 := |\{i \mid x_i \neq 0\}|$ of a vector $\mathbf{x} = (x_1, \ldots, x_N)$ and outputting an estimate $\widetilde{F}$ for which $(1-\varepsilon)\|\mathbf{x}\|_0 \leq \widetilde{F} \leq (1+\varepsilon)\|\mathbf{x}\|_0$ with constant probability. The vector $\mathbf{x}$ is initialized to all zeros and undergoes a sequence of m updates each of the form $(i, v) \in [N] \times [\pm M]$, where $[\pm M] := \{0, \pm 1, \ldots, \pm M\}$ and each update (i, v) causes $x_i \leftarrow x_i + v$. In the strict turnstile model $x_i \geq 0$ holds for all i and at all points in the stream. We obtain an $\Omega\left(\varepsilon^{-2} \log(N) \log\log(mM)\right)$ bits of space lower bound for $(1 \pm \varepsilon)$-approximating the Hamming norm. This lower bound matches the best known upper bound $O\left(\varepsilon^{-2} \log(N) \left(\log(1/\varepsilon) + \log\log(mM)\right)\right)$ [16] for any $\varepsilon \geq 1/\mathsf{polylog}(mM)$. Note that $\varepsilon \geq 1/\mathsf{polylog}(mM)$ is required in order to obtain polylogarithmic space, and so is the most common setting of parameters.

Perhaps surprisingly, there is an upper bound of $O\left(\varepsilon^{-2} \log(mM)\right)$ bits of space for $(1 \pm \varepsilon)$-approximating L_p for $p > 0$ [15] (improving an earlier $O\left(\log^2 N\right)$ bound of [11]; see also a time-efficient version in [14]), and thus we provide a strict separation in the complexities for $p = 0$ and $p > 0$.

The Hamming norm has many applications, as it corresponds to estimating the number of distinct values, and can be used to estimate set union and intersection sizes (see [8] where it was introduced).

Lower Bounds in the Learning Augmented Setting Recently, there has been a growing interest in using machine learning to infer information about the stream that would be useful for solving certain problems in the streaming setting. In this learning augmented setting, we have access to an oracle (which in practice would have some degree of error and could be implemented with machine learning). Learned oracles have been used to develop improved algorithms for various problems, including frequency estimation [10], caching [19], scheduling [20], frequency moments [13], and more. A fairly comprehensive survey of learning augmented algorithms can be found in [21].

In our setting, the oracle provides an additional operation: we can give the oracle a coordinate, and the oracle will tell us whether the frequency of this coordinate at the end of the stream is at least T for a *threshold* T. We refer to this oracle as the *heavy hitters oracle*. Approximate heavy hitter oracles have been used for frequency estimation [10].

We derive a new method to prove space lower bounds even with a perfect heavy hitters oracle (that is, an oracle that can be accessed with no space cost which always answers correctly whether the frequency of the coordinate is at least T). We use this method to prove a lower bound of $\Omega\left(\varepsilon^{-2}\log(N)\log\log(T)\right)$ for approximating the L_0 norm, which is optimal when $T = 2^{\mathsf{poly}(1/\varepsilon)}$ as it matches the upper bound in [13]. To prove this, we prove and use a slightly modified version of the direct sum theorem for Viola's problem, which will be stated in the following section.

1.2 Notable changes from the conference version

In this version of our paper, we have substantially updated Section 7. First, the result was generalized to the setting with a heavy hitters oracle as described in the paragraph above. Second, the proof of the bound in the previous version used an incorrect reduction to gap-hamming. In this version, this issue was resolved by instead reducing to gap-orthogonality.

1.3 Technical overview

We first illustrate the idea behind showing there is no gap between k-player and 2-player deterministic one-way communication complexity. The first player P_1 of the k-player protocol pretends to be Alice, the first player of the 2-player protocol, to create the message m_1 as Alice would do and sends it to the second player P_2 of the k-player protocol. Having received this message m_1, P_2 enumerates over all possible inputs of P_1 until finding one which would cause P_1 to send m_1. Since the protocol is deterministic and it evaluates a function defined on a product domain,[1] meaning that it is a total function on a domain of the form $S_1 \times S_2 \times \cdots \times S_k$, the

[1]Note that while we will be working with non-product input distributions, the function is still defined on all inputs, including ones that occur with probability 0 in the distribution we are working with.

function value must be the same as long as P_1's input results in the same message m_1 to be sent. So P_2 can arbitrarily pick one of those inputs as its guess for P_1. Now P_2 has a guess x for P_1's input together with its own input y, and P_2 can simulate Alice in the 2-player protocol. This is feasible because the 2-player protocol works under any partitioning of the inputs. Then P_2 sends to the third player P_3 the message that Alice would send to Bob in the 2-player protocol, given that Alice had input (x, y). In case when every player P_i cannot figure out how many input items have been processed from its own input and the received message m_{i-1}, which is important for its simulation of the 2-player protocol, an additional logarithmic-many-bits index carrying this piece of information should be passed together with the simulated messages. In this way, the entire k-player protocol can be simulated and the per player communication equals to the communication of the 2-player protocol between Alice and Bob, sometimes plus the additional logarithmic many bits for the index. Moreover, both protocols are deterministic.

For the randomized case with a product input distribution, we first consider 2-player protocols with error probability $1/\mathsf{poly}(k)$.

We would like to run the same simulation as for deterministic protocols, except now it is unclear how the second player P_2 can reconstruct a valid input x for the first player P_1 from the first message m. A natural thing would be for P_2 to choose the input $x = x_m$ to P_1 for which the probability of sending m, given that P_1's input is x_m, is greatest. This is not correct though, since the overall probability of P_1 holding x_m and sending m may be less than the $1/\mathsf{poly}(k)$ error bound and the protocol could afford to be always wrong on such a combination of x_m and m. Thus we need some balancing between two probabilities: (i) the first player P_1 sends m on input x; and (ii) the protocol output is correct given that P_1 has input x and sends m.

The above naturally suggests that we should impose an input product distribution μ. Then it must be that for a good fraction of x, weighted according to μ, the k-player protocol is correct when the first player has input x and sends message m. Thus we can sample x from the conditional distribution on μ given that message m is sent. Here, for correctness, it is crucial that μ is a product distribution; this ensures for most settings of remaining player's inputs (weighted according to μ), for most choices of x (weighted according to μ) giving rise to m, the function evaluated on the inputs is the same, and x can be sampled independently of remaining inputs. Once we have sampled x, and given that the second player has private input y in the k-player protocol, we can then have the second player pretend to be Alice of a randomized 2-player protocol with input (x, y), similar to the deterministic case. Ultimately, we will show that under distribution μ we obtain a protocol with total communication at most $O(k)$ times that of the 2-player protocol with error probability $1/\mathsf{poly}(k)$. The maximum message length, which is an important resource measure in our setting, blows up by at most an $O(1)$ multiplicative factor times that of the 2-player protocol, where the factor k comes from the number of invocations of the 2-player protocol.

We illustrate the optimality of the randomized reduction above by looking at the SUM-EQUAL problem studied by Viola [23]: in this problem each of k players holds an input $x_i \bmod p$, where $p = \Theta\left(k^{1/4}\right)$ is a prime, and they wish to determine whether $\sum_i x_i = 0$ or $1 \bmod p$. Viola shows this problem has randomized communication complexity $\Theta\left(k \log k\right)$, for both randomized protocols with constant error probability as well as deterministic protocols (and thus also randomized protocols with $1/\text{poly}(k)$ error probability). Moreover, for randomized protocols with $1/\text{poly}(k)$ error probability, Viola's $\Omega(k \log k)$ lower bound holds even for a product distribution on the inputs (where if $\sum_i x_i \bmod p \notin \{0, 1\}$ the output can be arbitrary). We observe that under any partition of the inputs into 2-players Alice and Bob, the problem can be solved with $O(\log k)$ bits with probability $1 - 1/\text{poly}(k)$ just by running an equality test on the sum modulo p of Alice and the negated sum modulo p of Bob. Thus, this illustrates that the factor $O(k)$ gap in total communication for protocols for product input distributions with $1/\text{poly}(k)$ error probability is *optimal*.

On the other hand, for constant error protocols and a product input distribution, there is a 2-player $O(1)$ bit upper bound in the public coin model which comes from running an equality test with constant error probability (since we measure error with respect to an input distribution, equality has an $O(1)$ upper bound with constant error).

We note that the k-player protocol has communication $\Omega(k \log k)$ for constant error protocols, which gives the $\Omega(k \log k)$ factor gap we claimed. The only downside is that the $\Omega(k \log k)$ lower bound holds for an input distribution which is product on $k - 1$ out of k players, rather than all k players. We leave it as an open question to give an optimal separation for product input distributions for constant error probability.

Given the importance of Viola's problem in showing separations, we next show a *direct sum theorem* for his problem, showing its communication complexity increases to $\Omega(km \log k)$ for solving a constant fraction of m independent copies. This additionally confirms that the $\Omega(k \log k)$ factor gap noted above is multiplicative and not additive. For technical reasons we require $m < k^c$ for a constant c, as discussed in Remark 6.6, but we suspect this may be an artifact of the proof.

To show the direct sum theorem for Viola's problem, one issue is that, unlike for two players where the technique of *information complexity* often provides direct sum theorems, for k-players the analogues are much weaker. A natural route would be to take Viola's *corruption bound*, argue it implies a high information bound, and then apply standard direct sum theorems for information. This approach does not give an information cost lower bound on private coin protocols, though one can fix it for two players using [5], which improves upon a bound in [6]. However, for k players similarly strong bounds are unknown. Another natural approach is to use the fact that if a problem has a corruption bound, then one immediately has a direct sum for it [4]. Again though, this is only for two players or the *number on forehead* model, and not for our setting.

Instead, our proof is inspired by Viola's rectangle argument for a single copy of the SUM-EQUAL problem, where each rectangle, restricted to the first $k - 1$ players, is a product distribution on which the protocol generates a message to the k-th player. We use a rectangle argument on multiple copies where the output is now a binary vector instead of a single bit. The main obstacle is that we must consider the Hamming distance between the protocol output and the correct answer in a vector space, which is much more involved than studying the error probability for a single instance. The intuition of our proof is that for every large rectangle, there must be linearly many copies that appear (almost) uniformly random in the last player's view. The above argument is fairly intricate, and involves several levels of conversion:

(i) a large rectangle implies large conditional entropy in many players' inputs;

(ii) the large entropy of all copies implies we have min-entropy at least 1 on many copies;

(iii) a random variable of min-entropy at least 1 can always be decomposed into a convex combination of uniform distributions over two elements;

(iv) the summation of $\omega(p^2)$ independent random variables that are each drawn from a uniform-over-two-element distribution turns out to be nearly uniform over $\mathbb{Z}_p$ due to a simple argument based on the primitive roots of unity, and hence many SUM-EQUAL copies look uniform to the last player.

Thus, the last player can hardly outperform a random guess. Note that it is insufficient to prove uniformity for many copies individually (which is not too hard using the same idea as in Viola's proof), since such a situation could be simulated with a much smaller rectangle with very small error. We instead perform our rectangle argument inductively to show most copies appear almost uniform, even if conditioned on previous copies.

This direct sum technique has further applications. One application of the direct sum technique, with slight modifications, is to prove a lower bound for approximating the Hamming norm in a strict turnstile stream. Using a result of [2], to show lower bounds for streaming algorithms in the strict turnstile model, it suffices to show lower bounds in the simultaneous communication model, where each player simultaneously sends a linear sketch to a referee who outputs the answer. To get the desired direct sum property, we have a chain of reductions leading to the SUM-EQUAL problem of which we compute the information complexity.

Specifically, we consider a composition of the Gap-Orthogonality problem on top of the SUM-EQUAL instances as well as an augmented index version of the composed problem. When we compose these problems, each coordinate of the Gap-Orthogonality problem becomes a SUM-EQUAL instance, and we show that in order to solve Gap Orthogonality, we must solve most of the SUM-EQUAL instances. Thus. we can use a direct sum to bound the information cost of the composed problem in a similar manner as in [25]. We then prove that approximating the Hamming norm reduces to the augmented index version of this, which allows us to bound its communication complexity and accordingly its streaming complexity.

In the augmented problem we additionally give a referee an index i and the answers to all copies j, with $j > i$. Similar augmentation has been studied for L_p-norms [15]. This allows us to reduce our communication problem to Hamming norm approximation, and ultimately prove our data stream lower bound.

2 Preliminaries

A function $f : \Sigma^k \to \Gamma$ is called a *k-party symmetric function* if for every $(x_1, x_2, \ldots, x_k) \in \Sigma^k$ and for every permutation σ over $\{1, 2, \ldots, k\}$, we have $f(x_1, \ldots, x_k) = f\left(x_{\sigma(1)}, \ldots, x_{\sigma(k)}\right)$.

A k-dimensional vector space S is called a *product space* if it can be represented as $S = S_1 \times S_2 \times \cdots \times S_k$. A distribution μ is called a *product distribution* if it is obtained by taking the product of k independent distributions, i. e., $\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_k$.

In the t-player communication complexity model, there are t computationally unbounded players, e. g., $P_1, \ldots, P_t$, required to compute a function $f : X_1 \times \cdots \times X_t \to Y$, where f is usually a t-party symmetric function. Each player P_i is given a private input $x_i \in X_i$ and follows a fixed protocol to exchange messages. For every input $(x_1, \ldots, x_t)$, the message transcript is denoted by $\Pi_t(x_1, \ldots, x_t)$ when all players follow the protocol Π_t (when Π_t is randomized, $\Pi_t(x_1, \ldots, x_t)$ is a random variable taking probabilities over players' random coins). A deterministic protocol Π_t computes f if there is a function Π_{out} such that $\Pi_{out}\left(\Pi_t^{(t)}(x_1, \ldots, x_t), x_t\right) \equiv f$, where $\Pi_t^{(t)}(x_1, \ldots, x_t)$ denotes P_t's view under the execution of Π_t on input $(x_1, \ldots, x_t)$ and for simplicity we let $\Pi_{out}(x_1, \ldots, x_t) := \Pi_{out}\left(\Pi_t^{(t)}(x_1, \ldots, x_t), x_t\right)$. A δ-error randomized protocol Π_t for f requires the existence of Π_{out} such that for all inputs $(x_1, \ldots, x_t)$, $\Pr_{\Pi_t}\left[\Pi_{out}(x_1, \ldots, x_t) = f(x_1, \ldots, x_t)\right] \geq 1 - \delta$. The *communication cost* of Π_t is the maximum size of $\Pi_t(x_1, \ldots, x_t)$ over all $x_1, \ldots, x_t$ and all random coins. The *t-player deterministic communication complexity*, denoted by $\mathbf{DCC}_t(f)$, is the cost of the best t-player deterministic protocol Π_t for f. Analogously, the *t-player δ-error randomized communication complexity*, denoted by $\mathbf{RCC}_{t,\delta}(f)$, is the cost of the best t-player δ-error randomized protocol Π_t for f with probability $1 - \delta$.

Given a k-party function $f : X_1 \times \cdots \times X_k \to Y$ and $t < k$, we define $\mathbf{DCC}_t(f)$ and $\mathbf{RCC}_{t,\delta}(f)$ under a *worst-case partition* of inputs. That is, let $f_t(z_1, \ldots, z_t) = f(x_1, \ldots, x_k)$ be defined for every partition $i_0 = 0 \leq i_1 \leq \cdots \leq i_t = k$ and $z_j := (x_{i_{j-1}+1}, \ldots, x_{i_j})$, and the t-player communication complexity of f is defined with respect to the worst choice of f_t, i. e., $\mathbf{DCC}_t(f) := \max_{f_t} \mathbf{DCC}_t(f_t)$ and $\mathbf{RCC}_{t,\delta}(f) := \max_{f_t} \mathbf{RCC}_{t,\delta}(f_t)$.

Given a t-party function f and its input distribution μ, we let $\mathbf{DCC}^{\mu}_{t,\delta}(f)$ denote the communication cost of the best t-player deterministic protocol Π_t computing f such that $\Pr_{x\sim\mu}\left[\Pi_{out}(x) \neq f(x)\right] \leq \delta$. Similarly we define $\mathbf{RCC}^{\mu}_{t,\delta}(f)$ for randomized protocols.

In the *one-way communication model* [22, 1, 17], [2] the i-th player sends exactly one message to the $(i + 1)$-st player for $i \in [t - 1]$ following Π_t, and then P_t announces the output of Π_t as specified by Π_{out}. Note that in this setting there are only $k - 1$ messages sent by $P_1, \ldots, P_{k-1}$, and we do not count the final output announced by P_t in the communication in order to best correspond to streaming algorithms. This is also known as a *sententious* protocol in previous work, e. g., [23]. We denote the *deterministic* and *randomized t-player one-way communication complexities of f* by $\overrightarrow{\mathbf{DCC}}_t(f)$ and $\overrightarrow{\mathbf{RCC}}_{t,\delta}(f)$, respectively.

In the *common reference string model* (a. k. a. *CRS model*), there is a sequence of public random coins, which is by default a uniformly random binary string, accessible to all players. The obvious advantage of communication in the CRS model is that players have access to the same random string and thus save the cost of synchronizing their private coins.

A streaming algorithm is an algorithm that scans the input $(x_1, \ldots, x_m) \in \Sigma^m$ as m stream input items in sequence, updates its internal memory of size $s = o\left(m \log |\Sigma|\right)$ (i. e., a streaming automaton with 2^s states, where the space cost of updating the internal memory is not accounted

[2]We are aware that there are errors in [17]. This does not affect our results in any way as we do not use any theorems from this work.

for), and finally outputs a function $f(x_1, \dots, x_m)$ evaluated on all input items. If the best[3] deterministic streaming algorithm computes f with s bits of memory and t passes over the data stream, then we say the *deterministic streaming complexity* of f is st, denoted by $\mathbf{DSC}(f) = st$. The *δ-error streaming complexity* of f is defined analogously (with reference to the best δ-error randomized streaming algorithm) and is denoted by $\mathbf{RSC}_\delta(f) = st$. In a popular and standard setting, a streaming algorithm scans the input stream in a *single pass* and only processes every input item once. The necessary amount of memory required by such single-pass algorithms is called the *single-pass deterministic/δ-error streaming complexity* and denoted by $\overrightarrow{\mathbf{DSC}}(f)$ and $\overrightarrow{\mathbf{RSC}}_\delta(f)$, respectively.

Note that every streaming algorithm can be naturally interpreted as a communication protocol where each party holds some (possibly an empty set of) input items on the stream and the messages capture the memory updates. The connection between streaming complexity and communication complexity trivially follows in the following lemma.

Lemma 2.1. *For every function f and error tolerance δ, for every $k \in \mathbb{N}$, it holds that:*

$$\mathbf{DSC}(f) \geq \frac{1}{k} \cdot \mathbf{DCC}_k(f), \quad \mathbf{RSC}_\delta(f) \geq \frac{1}{k} \cdot \mathbf{RCC}_{k,\delta}(f)$$

Furthermore, similar relations hold for single-pass streaming complexities versus k-player one-way communication complexities:

$$\overrightarrow{\mathbf{DSC}}(f) \geq \frac{1}{k-1} \cdot \overrightarrow{\mathbf{DCC}}_k(f), \quad \overrightarrow{\mathbf{RSC}}_\delta(f) \geq \frac{1}{k-1} \cdot \overrightarrow{\mathbf{RCC}}_{k,\delta}(f)$$

Next, we introduce the linear sketch model of communication. In this setting, we have n players, the last of whom is the referee, and the only protocols allowed are of the following form:

There is some matrix A such that if player i receives input x_i, they compute Ax_i and send the result to the referee. The referee then computes $\sum_{i=1}^{n} Ax_i$ and uses the result to compute the answer. We denote the randomized communication complexity of a function f in this model by $\mathbf{RCC}_{k,\delta}^{LIN}(f)$.[4]

Additionally, we let $\mathbf{D}_{k,\delta,\mu}(f)$ denote the communication complexity of f with k players and δ error under input distribution μ and $\mathbf{IC}_{k,\delta}(f)$ denote the information complexity of f with k players and δ error. Both of these complexities are considered in the linear sketch model. We extend the notion of information complexity from [7] to this setting by summing the information costs over all of the players and allowing some probability of returning an incorrect answer. That is, let I denote mutual information, and let $\Pi_{k,\delta}^{f}$ denote the set of k-player randomized protocols in the linear sketch model solving f with probability $1-\delta$. Additionally, let $\Pi(x_1, x_2, \dots, x_k)_i$ denote the message sent by the i^{th} player when we run the protocol Π on input $(x_1, x_2, \dots, x_k)$.

[3]The "best" such algorithm is the one with the minimal value of st on the input that maximizes st.

[4]Simultaneous communication models often define the referee as an additional player with no input. In this case, this is equivalent to our model except the $k-1$s in our proofs and bounds would become ks.

Then

$$\mathbf{IC}_{k,\delta}(f) := \min_{\Pi \in \Pi^f_{k,\delta}} \sum_{i=1}^{k} I\left(x_i, \Pi(x_1, x_2, \ldots, x_k)_i\right).$$

The following lemma relates $\mathbf{IC}_{k,\delta}(f)$ to $\mathbf{RCC}^{LIN}_{k,\delta}(f)$

Lemma 2.2. *For any function f,*

$$\mathbf{IC}_{k,\delta}(f) \leq \mathbf{RCC}^{LIN}_{k,\delta}(f)$$

Proof. Consider any protocol Π solving f in the linear sketch model, and let A be the matrix used in the protocol Π. Then, $\Pi(x_1, x_2, \ldots, x_k)_i = Ax_i$. If we let b_i be the number of bits used to represent Ax_i for each $i \in [k]$, we have $I(x_i, Ax_i) \leq b_i$ for every $i \in [k]$. In particular, if we let

$$\Pi := \mathrm{argmin}_{\Pi \in \Pi^f_{k,\delta}} \sum_{i=1}^{k} I\left(x_i, \Pi(x_1, x_2, \ldots, x_k)_i\right),$$

then

$$\mathbf{IC}_{k,\delta}(f) = \sum_{i=1}^{k} I\left(x_i, \Pi(x_1, x_2, \ldots, x_k)_i\right) \leq \sum_{i=1}^{k} b_i \leq \mathbf{RCC}^{LIN}_{k,\delta}(f)$$ □

Additionally, $\mathbf{IC}(f)$ is well-behaved in the sense that it satisfies the direct sum property. That is, letting f^m denote the problem where we solve m independent instances of f:

Theorem 2.3. *For any function f and any positive integer m,*

$$\mathbf{IC}_{k,\delta}(f^m) \geq m \cdot \mathbf{IC}_{k,\delta}(f)$$

where a δ probability of failure for f^m is defined to mean a δ probability of failure on each instance.

This follows from the direct sum theorem on two players and no error by grouping all but player i into the referee for each i and summing over the information complexities of the protocols for each i. Then, to deal with the δ probability of error, we simply force the protocols to be deterministic and consider the function only on the values for which it is correct.

We denote the randomized communication complexity of a function f in the linear sketch model given that the maximum frequency of any coordinate at the end of the stream is at most T by $\mathbf{RCC}^{LIN,T}_{k,\delta}(f)$. Similarly, in the other models, when we bound the frequency of the coordinates, we will write $\mathbf{D}^{LIN,T}_{k,\delta,\mu}$, $\mathbf{IC}^{T}_{k,\delta}$, and $\mathbf{RSC}^{T}_{k,\delta}$ for distributional complexity, information complexity, and randomized streaming complexity, respectively.

3 Communication complexity for functions on non-product spaces

Theorem 3.1. *For every $t \geq 2$, there is a t-party symmetric function $f : D \to \{0,1\}$ defined on $D \subseteq \{0,1\}^n = \left(\{0,1\}^{n/t}\right)^t$ such that for every error tolerance $\delta < 1/4$, $\overrightarrow{\mathbf{DCC}}_{t-1}(f) \leq t-1$ but $\mathbf{RCC}_{t,\delta}(f) = \Omega(n/t)$. In particular, as long as $t = O(1)$, we have $\overrightarrow{\mathbf{DCC}}_{t-1}(f) = O(1)$ and $\mathbf{RSC}_\delta(f) \geq \frac{1}{t} \cdot \mathbf{RCC}_{t,\delta}(f) = \Omega(n)$.*

Proof. Consider the t-party set disjointness problem $\textsc{Disj}_{n/t,t}$ defined as follows: there are t players $P_1, \ldots, P_t$ such that every player P_i holds a private indicator vector $\mathbf{x}_i \in \{0,1\}^{n/t}$ which represents a subset of $[n/t]$, i. e., $\textsc{Disj}_{n/t,t}(\mathbf{x}_1, \ldots, \mathbf{x}_t) = \vee_{j=1}^{n/t}\left(\wedge_{i=1}^{t} x_{i,j}\right)$, where $x_{i,j}$ denotes the j-th coordinate of $\mathbf{x}_i$. We consider the domain D such that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t \in \{0,1\}^{n/t}$ are either (1) pairwise disjoint, or (2) there exists a unique $j \in [n/t]$ such that $x_{i,j} = 1$ for all $i \in [t]$. Let f be the function that computes $\textsc{Disj}_{n/t,t}$ on domain D.

On the one hand, it is easy to verify that $\overrightarrow{\mathbf{DCC}}_{t-1}(f) \leq t-1$. Indeed, at least one of the $t-1$ players obtains two distinct indicator vectors and hence can itself decide the output of f. The communication is 1 bit per player to pass the result, and hence the total communication is bounded by $t-1$ since there are $t-1$ players.

On the other hand, the $\Omega(n/t)$ lower bound for $\mathbf{RCC}_{t,\delta}(f)$ follows from the known lower bound for multi-player set disjointness (see [3], which was improved to optimal in [9, 12]). The lower bound for $\mathbf{RSC}_\delta(f)$ immediately follows by Lemma 2.1. ◻

4 Deterministic communication and streaming complexity

We first show that 2-player one-way communication complexity is equivalent to the streaming complexity of single-pass streaming algorithms in the deterministic setting. In the following theorem, we assume for convenience that m is known to both players.

Theorem 4.1. *For every symmetric function $f : \Sigma^m \to \Gamma$, $\overrightarrow{\mathbf{DCC}}_2(f) \leq \overrightarrow{\mathbf{DSC}}(f) \leq \overrightarrow{\mathbf{DCC}}_2(f) + \log m$.*

Proof. Obviously, $\overrightarrow{\mathbf{DSC}}(f) \geq \overrightarrow{\mathbf{DCC}}_2(f)$ since a 2-player communication protocol can simulate a streaming algorithm. It remains to prove $\overrightarrow{\mathbf{DSC}}(f) \leq \overrightarrow{\mathbf{DCC}}_2(f) + \log m$.

Suppose the input stream is $(x_1, \ldots, x_m) \in \Sigma^m$, and for every partition into $(x_1, \ldots, x_i)$ and $(x_{i+1}, \ldots, x_m)$ there is a deterministic 2-player one-way protocol Π_2^i computing f. We design the deterministic single-pass streaming algorithm A for f by simulating 2-player one-way communication protocols under different partitions. The memory usage of A is therefore bounded by the maximum communication cost of the simulated 2-player protocols plus an index in $[m]$ recording the number of processed items.

Notice that when processing the item x_{i+1}, A has already processed $x_1, \ldots, x_i$ and has (m_i, i) in memory. A can thus reconstruct a compatible guess of $x_1'', \ldots, x_i''$ that would induce exactly the message m_i as in Π_2^i, and then sets the memory to be $(m_{i+1}, i+1)$ where m_{i+1} is the message sent in Π_2^{i+1} when P_1 has $(x_1'', \ldots, x_i'', x_{i+1})$ and P_2 has $(x_{i+2}, \ldots, x_m)$. Since Π_2^i is deterministic, it will always output the same answer when P_2 has input $(x_{i+2}, \ldots, x_m)$ and receives message m_i. Thus, if $x''_1, \ldots, x''_i$ would induce the same message m_i, then Π_2^i would produce the same answer regardless of whether P_1 had input $(x_1, \ldots, x_i)$ or $(x''_1, \ldots, x''_i)$. In particular, since Π_2^i computes f, this means $f(x_1, \ldots, x_m) = f(x''_1, \ldots, x''_i, x_{i+1}, \ldots, x_m)$. Thus, when we compute m_{i+1}, we still get some message that Π_2^{i+1} can use to correctly compute f alongside P_2's input $(x_{i+2}, \ldots, x_m)$.

A repeats this process for every $i = 1, \ldots, m-1$ and at the end it outputs $f(x_1, \ldots, x_m)$.

Therefore, we complete the proof with $\overrightarrow{\mathbf{DCC}}_2(f) \leq \overrightarrow{\mathbf{DSC}}(f) \leq \overrightarrow{\mathbf{DCC}}_2(f) + \log m$. $\square$

Note that the additional index i in the above simulation, which results in the additive $\log m$ term in the upper bound, indicates which 2-player protocol should be simulated in the reconstruction, and it is implicitly shared in the 2-player communication case when m is common knowledge.

When m is not known, the memory used for the index follows any previously agreed upon encoding, which uses $O(\log m)$ space. For functions that are well-defined for an arbitrary number of input items, e. g., the parity function, this index can be saved, and hence $\overrightarrow{\mathbf{DSC}}(f) = \overrightarrow{\mathbf{DCC}}_2(f)$.

For communication complexity among more players, we establish the following corollary.

Corollary 4.2. *For every k-party symmetric function f,*

$$(k-1) \cdot \overrightarrow{\mathbf{DCC}}_2(f) \leq \overrightarrow{\mathbf{DCC}}_k(f) \leq (k-1) \cdot \left(\overrightarrow{\mathbf{DCC}}_2(f) + \log k\right)$$

Proof. Combining Lemma 2.1 and Theorem 4.1, it follows that

$$\overrightarrow{\mathbf{DCC}}_k(f) \leq (k-1) \cdot \overrightarrow{\mathbf{DSC}}(f) \leq (k-1) \cdot \left(\overrightarrow{\mathbf{DCC}}_2(f) + \log k\right)$$

The other direction $\overrightarrow{\mathbf{DCC}}_k(f) \geq (k-1) \cdot \overrightarrow{\mathbf{DCC}}_2(f)$ holds by giving $z_j = \emptyset$ to every player $j \in \{2, \ldots, k-1\}$ in the k-player case, when the problem degenerates to 2-player communication but the same message has to be passed $k-1$ times. $\square$

Such a linear separation naturally extends to the communication complexity of t-player versus k-player protocols, as long as $2 \leq t < k$. Thus, the deterministic communication complexity grows *linearly* in the number of parties.

We remark that if every player must get a non-trivial input, i. e., at least one input element to the function, the linear growth remains for some but not all problems. For example, the communication complexity of the parity of k bits is linear in the number of players. However, to decide whether k elements in $[k]$ are distinct, the 2-player protocol requires communication $\log \binom{k}{k/2} \approx k - \log \sqrt{k}$, whereas the k-player worst-case communication grows sublinearly, i. e., for k players the communication is no more than $\sum_{i=1}^{k-1} \log \binom{k}{i} \ll (k-1) \cdot \log \binom{k}{k/2}$.

5 Communication complexity for functions on a product space

5.1 Separations for randomized communication complexity

In this section, we consider the communication cost of randomized multi-player protocols defined on product input distributions and present a $k \log k$ versus $t \log t$ separation between k-player and t-player communication complexity.

First we introduce the SUM-EQUAL problem (as used in Viola's work [23]).

Definition 5.1. The *k-player SUM-EQUAL over integers*, denoted by SUM-EQUAL_k, requires deciding whether $\sum_{i=1}^{k} x_i = 0$, where each player P_i is given an integer x_i as its private input together with the integer k as public input shared by all players. In the CRS model, an additional public random string is also known to all players. The *k-player SUM-EQUAL over* $\mathbb{Z}_m$, denoted by $\text{SUM-EQUAL}_{k,m}$, is defined similarly as SUM-EQUAL_k, except that the input items are drawn from $\mathbb{Z}_m$ and the summation is over $\mathbb{Z}_m$, for a publicly known m.

Lemma 5.2 ([23], Theorem 15 and Theorem 29)**.** *For every $k \in \mathbb{N}$, $0 \le \delta \le 1/3$, and in the CRS model, the k-player δ-error communication complexity of SUM-EQUAL satisfies:*

(a) For every $m \in \mathbb{N}$, $\overrightarrow{\mathbf{RCC}}_{k,\delta}(\text{SUM-EQUAL}_{k,m}) = O\left(k \log(k/\delta)\right)$.

(b) For every prime $p \in (k^{1/4}, 2k^{1/4})$, $\mathbf{RCC}_{k,\delta}(\text{SUM-EQUAL}_{k,p}) = \Omega\left(k \log k\right)$.[5]

In particular, $\mathbf{RCC}_{k,\delta}(\text{SUM-EQUAL}_{k,p}) = \Theta\left(k \log k\right)$ in the CRS model if $\delta = \Omega\left(1/\mathsf{poly}(k)\right)$.

Remark 5.3. Viola's lower bound for $\text{SUM-EQUAL}_{k,p}$ is proved for a non-product distribution μ' whose support covers exactly a $2/p$ fraction of the whole (product) input space. Specifically, μ' is defined as follows:

Definition 5.4. We define two distributions G and B:

$$\begin{cases} G := & \left(G_1, \ldots, G_{k-1}, \quad -\sum_{j=1}^{k-1} G_j\right) \\ B := & \left(B_1, \ldots, B_{k-1}, \quad 1 - \sum_{j=1}^{k-1} B_j\right) \end{cases}$$

where for each $i \in [k-1]$, G_i and B_i are chosen iid uniformly from $\mathbb{Z}_p$. Then, $\mu' := G/2 + B/2$ is drawn from each distribution with probability $\frac{1}{2}$.

Thus if a k-player protocol solves $\text{SUM-EQUAL}_{k,p}$ with error $\delta \le 1/k$ on a uniform distribution μ over the whole input space, then its error with respect to μ' is bounded by $\frac{1/k}{2/p} < k^{-3/4}$.

Notice that the two player version of $\text{SUM-EQUAL}_{k,p}$ degenerates to testing equality over $\mathbb{Z}_p$ whose upper bound is $O\left(\log(1/\delta) + \log\log k\right)$, see more details in Appendix A. By Lemma 5.2, the $\Omega(k)$ separation in Corollary 5.5 naturally follows.

[5]Viola's states the lower bound for constant δ, but it naturally holds for smaller δ (sometimes not tight).

Corollary 5.5. *For every prime $p \in (k^{1/4}, 2k^{1/4})$ and $\delta \leq 1/\mathsf{poly}(k)$, there is a product distribution μ such that $\mathbf{RCC}^{\mu}_{k,\delta}(\textsc{Sum-Equal}_{k,p}) = \Omega\left(k \log k\right)$, $\overrightarrow{\mathbf{RCC}}_{2,\delta}(\textsc{Sum-Equal}_{k,p}) = O\left(\log k\right)$.*

For a larger error tolerance, say δ is a constant, we have a stronger separation between k-party communication and t-party communication. However, the hard distribution is slightly non-product, that is, it is a product distribution on any $k-1$ out of the k players.

Corollary 5.6. *For every $k \in \mathbb{N}$, there is a k-party symmetric function f such that*

(a) For any product distribution μ, for every $2 \leq t \leq k$ and $0 \leq \delta \leq 1/3$, $\overrightarrow{\mathbf{RCC}}^{\mu}_{t,\delta}(f) = O\left(t \log(t/\delta)\right)$. In particular, $\overrightarrow{\mathbf{RCC}}^{\mu}_{2,\delta}(f) = O\left(\log(1/\delta)\right)$.

(b) There exists a distribution μ', which is product on any $k-1$ out of k players, for which $\mathbf{RCC}^{\mu'}_{k,\delta}(f) = \Omega\left(k \log k\right)$ as long as $\delta \leq 1/3$.

For $\delta \geq 1/\mathsf{poly}(t)$, the gap between $\mathbf{RCC}^{\mu}_{k,\delta}(f)$ and $\overrightarrow{\mathbf{RCC}}^{\mu}_{t,\delta}(f)$ is bounded as below:

$$\frac{\mathbf{RCC}^{\mu}_{k,\delta}(f)}{\overrightarrow{\mathbf{RCC}}^{\mu}_{t,\delta}(f)} = \Omega\left(\frac{k \log k}{t \log t}\right)$$

Proof. (a) If we plug in $k = t$ to part (a) of Lemma 5.2, we get $\overrightarrow{\mathbf{RCC}}_{k,\delta}(\textsc{Sum-Equal}_{t,m}) = O\left(t \log(t/\delta)\right)$ for every $m, t \in \mathbb{N}$ and $0 \leq \delta \leq 1/3$. Thus, $\textsc{Sum-Equal}_{t,m}$ satisfies (a).

(b) Part (b) of Lemma 5.2 tells us that for any $0 \leq \delta \leq 1/3$, we can take some prime $p \in (k^{1/4}, 2k^{1/4})$, and we have $\mathbf{RCC}_{k,\delta}(\textsc{Sum-Equal}_{k,p}) = \Omega(k \log k)$. Furthermore, as is noted in Remark 5.3, we actually have that this holds for a distribution μ' which is a product on the first $k-1$ players. As the Sum-Equal problem is symmetric with respect to all k players, the desired property follows immediately.

□

The outline of the proof of Corollary 5.6 was given in Section 1. That is, the upper bound in part (a) follows from applying $k = j$ in the first part of Lemma 5.2, while the lower bound in part (b) follows from the second part of Lemma 5.2.

5.2 Tightness of the communication complexity separation

The following theorem and corollary show tightness of our separations.

Theorem 5.7. *For every k-party function $f : \Sigma^k \to \Gamma$, product distribution μ over Σ^k, and error tolerance $\delta < 1/3$,* *the following holds:*

$$\overrightarrow{\mathbf{RCC}}^{\mu}_{k,\delta}(f) = \begin{cases} O\left(k\log k\right)\cdot \overrightarrow{\mathbf{RCC}}_{2,\delta}(f) & \text{if } \delta = \Omega\left(1\right) \\ O\left(k\right)\cdot \overrightarrow{\mathbf{RCC}}_{2,\delta}(f) + O\left(k\log k\right) & \text{if } \delta \le 1/k^{\Omega(1)} \end{cases}$$

When $\delta > 0$, we have that $\overrightarrow{\mathbf{RCC}}_{2,\delta}(f) = \Omega(\log k)$ and thus the following holds:

$$\frac{\overrightarrow{\mathbf{RCC}}^{\mu}_{k,\delta}(f)}{\overrightarrow{\mathbf{RCC}}_{2,\delta}(f)} \le O\left(k\cdot\left(1+\frac{\log k}{\log(1/\delta)}\right)\right) = \begin{cases} O\left(k\log k\right) & \text{if } \delta = \Omega\left(1\right) \\ O\left(k\right) & \text{if } \delta = 1/k^{\Omega(1)} \end{cases}$$

Proof. First we let Π_0 be the optimal δ-error 2-player one-way protocol Π_0 that computes f with communication $C = \overrightarrow{\mathbf{RCC}}_{2,\delta}(f)$, and construct a new protocol Π_2 by taking the majority of M independent parallel copies of Π_0 such that Π_2 has error $\varepsilon = \frac{\delta^2}{16k^2}$ and communication CM. Recall that Π_0 has $\delta < 1/3$, it suffices to let t and M be defined as in Lemma 5.8 below:

$$t = \left\lceil \log\frac{\delta}{16k^2}/\log\left(4\delta(1-\delta)\right)\right\rceil \tag{5.1}$$

$$M = 1 + 2t = 1 + 2\left\lceil\frac{\log(1/\delta) + 2\log k + 4}{\log(1/\delta) + \log(1/(1-\delta)) - 2}\right\rceil = \Theta\left(1 + \frac{\log k}{\log(1/\delta)}\right) \tag{5.2}$$

Lemma 5.8. *Let $t \in \mathbb{N}$ and $X_1, X_2, \ldots, X_{2t+1}$ be i.i.d. binary random variable such that $\Pr[X_i = 1] = \delta < 1/2$ for every $i \in [t]$, and let $Y = \mathbf{Majority}\{X_1, \ldots, X_{2t+1}\}$ be the majority of all X_i's. Then $\Pr[Y = 1] \le \varepsilon$ as long as $t \ge \log(\varepsilon/\delta)/\log(4\delta(1-\delta))$.*

Proof. For $0 < \delta < 1/2$ and $t \ge \log(\varepsilon/\delta)/\log(4\delta(1-\delta))$, we have

$$\begin{aligned}
\Pr\left[Y = 1\right] &= \Pr\left[\left|\{i \mid X_i = 1\}\right| \ge t+1\right] \\
&= \sum_{j=t+1}^{2t+1}\binom{2t+1}{j}\delta^j(1-\delta)^{2t+1-j} \\
&\le \sum_{j=t+1}^{2t+1}\binom{2t+1}{j}\delta^{t+1}(1-\delta)^t \\
&= \frac{2^{2t+1}}{2}\cdot\delta^{t+1}(1-\delta)^t = \left(4\delta(1-\delta)\right)^t\cdot\delta \\
&\le \frac{\varepsilon}{\delta}\cdot\delta = \varepsilon
\end{aligned}$$

The first inequality holds because $\delta < 1/2$ and hence $\delta^j(1-\delta)^{2t+1-j} \le \delta^{t+1}(1-\delta)^t$ for $j \ge t+1$. The second inequality holds because $4\delta(1-\delta) < 1$ for $\delta < 1/2$, and $\left(4\delta(1-\delta)\right)^t \le \left(4\delta(1-\delta)\right)^{\log(\varepsilon/\delta)/\log(4\delta(1-\delta))} = \varepsilon/\delta$. Thus, we have proved that $\Pr[Y = 1] \le \varepsilon$ for $t \ge \log(\varepsilon/\delta)/\log(4\delta(1-\delta))$. ∎

Note that Π_2 is still a 2-player one-way protocol but has communication CM. Furthermore, we remark that $CM = \Omega\left(\log k\right)$ for $\delta > 0$, since the error probability must be $\delta \geq 1/2^C$ if it is not zero, and hence $M = \Theta\left(1 + \frac{\log k}{\log(1/\delta)}\right) = \Omega\left(1 + \frac{\log k}{C}\right)$.

Second we prove that for every product input distribution μ over Σ^k, the k-party function f can be evaluated by a randomized k-player one-way protocol Π_k with communication $O\left(k \cdot (CM + \log k)\right)$ and error $\delta/2$ with respect to μ. The idea is that given the product input distribution μ, each player P_i acts as follows:

1. P_i first assumes that the received message m_{i-1} from P_{i-1} will lead to a correct answer with probability $\geq 1 - \frac{\delta}{4k}$ with respect to μ. When we make this assumption, we essentially have P_i consider the problem using the input P_{i-1} generated in their step 2 rather than the real input.

2. P_i samples a possible input $x'_1, \ldots, x'_{i-1}$ of previous players $P_1, \ldots, P_{i-1}$, such that if Alice gets input $(x'_1 \ldots, x'_{i-1})$ and sends m_{i-1}, then with probability $\geq 1 - \frac{\delta}{4k}$ the protocol Π_2 leads to the correct answer. The probability is taken over internal randomness and Bob's input following the marginal distribution of μ on the remaining players (here we use the condition that μ is a product distribution).

3. Finally, P_i sends a message (m_i, i) of length $CM + \log k = O\left(CM\right)$, where m_i is the message that Alice would send in Π_2 when her input is $(x'_1, \ldots, x'_{i-1}, x_i)$.

Now, we can bound the error probability recursively: Suppose player P_i receives the message m_{i-1} from P_{i-1}, generated from $(x_1, x_2, \ldots, x_{i-1})$. Then, suppose P_i generates $(x'_1, x'_2, \ldots, x'_{i-1})$ in step 2. Then Π_2 is correct on the input $X' = (x'_1, x'_2, \ldots, x'_{i-1}, x_i, x_{i+1}, \ldots, x_k)$ with probability $\geq 1 - \frac{\delta}{4k}$ by our choice of M. Furthermore, since Π_2 generates m_i from both $(x_1, x_2, \ldots, x_{i-1})$ and $(x'_1, x'_2, \ldots, x'_{i-1})$ and is deterministic, it produces the same answer on both X' and $X = (x_1, x_2, \ldots, x_k)$. Thus, the answer we get from Π_2 on input X' is also correct on input X with probability at least $1 - \frac{\delta}{2k}$.

Thus, we can union bound over all the players to get that the error probability of Π_k is bounded by $k \cdot (\frac{\delta}{2k}) = \delta/2$ with respect to μ. The fact that μ is a product distribution is used in the second step where the sampling process relies on that previous players' inputs are independently distributed from that of future players.

Thus we finish the proof and conclude that $\overrightarrow{\mathbf{RCC}}^{\mu}_{k,\delta}(f) \leq O\left(kCM\right)$. □

Notice that in the proof of Theorem 5.7, every message in Π_k has the length bounded by $O\left(CM\right)$, which gives an upper bound for the single-pass streaming complexity.

Corollary 5.9. *For every k-party function f and product input distribution μ, and for every $\delta < 1/3$,* $\mathbf{RSC}^{\mu}_{\delta}(f) \leq \overrightarrow{\mathbf{RSC}}^{\mu}_{\delta}(f) \leq O\left(1 + \frac{\log k}{\log(1/\delta)}\right) \cdot \overrightarrow{\mathbf{RCC}}_{2,\delta}(f)$.

6 A direct sum for Viola's problem

We next turn to our direct sum theorem for Viola's problem. This extends the results of the previous section by demonstrating that the gap is indeed multiplicative, where the results of the previous section do not rule out the possibility of an additive $\Omega(k \log k)$ gap between 2-player and k-player communication complexity. A slightly modified version of this result is used in our streaming result. Note that the theorem is proved for $\delta < 1/9$, but lower bounds for large error tolerance such as $\delta = 1/3$ can be obtained using a standard error amplification argument.

Theorem 6.1. *Let $F : \left(\mathbb{Z}_p^m\right)^k \to \{0,1\}^m$ be the k-party function computing m independent copies of* SUM-EQUAL$_{k,p}$, *where p is a prime between $k^{1/4}$ and $2k^{1/4}$ and $m = o(k^{1/4})$. For every error tolerance $\delta \in (0, 1/9)$, we say a protocol Π is* correct with probability $1-\delta$ *if there is a reconstruction function G such that for every fixed $i \in [m]$ and input $x \in \left(\mathbb{Z}_p^m\right)^k$, $G\big(i, \Pi_{out}(x)\big)$ equals the output of the i-th instance of* SUM-EQUAL$_{k,p}$ *with probability at least $1-\delta$, over the internal randomness of Π. Then the communication cost of any Π which is correct with probability $1-\delta$, is $\Omega\left(mk \log k\right)$.*

Proof. For simplicity of notation in the proof, we flip the output of F, so that it outputs 0 if the input to the corresponding SUM-EQUAL$_{k,p}$ instance sums to 0 in $\mathbb{Z}_p$, and F outputs 1 on instances with summation other than 0.

Let Π be an δ-error randomized protocol for F, and let $\Pi_{out}(x)$ denote the output of Π on input x. Here by "the δ-error protocol" we mean that the expected error rate of Π is bounded by δ, since both $\Pi_{out}(x)$ and $F(x)$ are binary vectors in $\{0,1\}^m$. Therefore,

$$\Pr_{i \in_R [m]} \left[\Pi_{out}(x)_i \neq F_i(x)\right] \leq \delta$$

where the input to F is partitioned as $x = \left(x^{(1)}, x^{(2)}, \ldots, x^{(m)}\right) \in \mathbb{Z}_p^{m \times k}$ such that $F_i(x) := \overline{\text{SUM-EQUAL}_{k,p}}(x^{(i)})$ computes the i-th instance of SUM-EQUAL$_{k,p}$ for each $i \in [m]$.

We abuse notation a little in this proof and let $|\cdot|$ denote the *Hamming weight* of a not necessarily binary vector, which measures the number of non-zero coordinates of the vector. Then,

$$\mathbf{E}\left[|\Pi_{out}(x) - F(x)|\right] \leq \delta m$$

To prove that $\mathbf{RCC}_{k,\delta}(F) = \max_x |\Pi(x)| = \Omega\left(mk \log k\right)$ for the optimal δ-error protocol Π, we will deduce a contradiction if Π uses $c < \gamma mk \log k$ bits of communication, for a constant $\gamma = (1-9\delta)/135 > 0$ and sufficiently large k. Thus, we can conclude a communication lower bound of $c \geq \gamma mk \log k = \Omega\left(mk \log k\right)$.

For the purposes of a contradiction, we first convert the randomized protocol Π into a deterministic protocol Π' that has small error with respect to a specific distribution $\mathcal{H}$. The deterministic protocol Π' is obtained by fixing all internal random coins of Π so that Π' has error rate at most δ for inputs drawn from $\mathcal{H}$.

$$\mathbf{E}_{X \sim \mathcal{H}}\left[\left|\Pi'_{out}(X) - F(X)\right|\right] \leq \delta m$$

Since Π' can never generate a transcript larger than the communication that Π uses in the worst case, i. e., $|\Pi'(X)| \le \max_x |\Pi(x)| = c$, it suffices to prove a communication lower bound for Π' on inputs drawn from $\mathcal{H}$.

By Markov's inequality, we have that for every positive constant $\varepsilon > 0$,

$$\Pr_{X\sim\mathcal{H}}\left[\left|\Pi'_{out}(X) - F(X)\right| > \varepsilon m\right] \le \frac{\mathbf{E}_{X\sim\mathcal{H}}\left[\left|\Pi'_{out}(X) - F(X)\right|\right]}{\varepsilon m} \le \frac{\delta}{\varepsilon} \tag{6.1}$$

Now we specify the distribution $\mathcal{H}$. Let G, B be defined as in Definition 5.4. Note that:

(a) $\textsc{Sum-Equal}_{k,p}(G) = 1$, $\textsc{Sum-Equal}_{k,p}(B) = 0$ and hence $F_i(G) = 0, F_i(B) = 1$.

(b) the first $k-1$ elements of G and B, denoted by G_{-k} and B_{-k}, follow the same distribution, i. e., the uniform distribution over $\mathbb{Z}_p^{k-1}$.

For convenience we can write $B = (G_{-k}, 1 + G_k)$.

Let $H := G/2 + B/2$ be a mixture of G and B and let $\mathcal{H}$ be m independent copies of H as below:

$$\mathcal{H} := H^m = (G/2 + B/2)^m$$

Since $B = (G_{-k}, 1 + G_k)$ and $H = G/2 + B/2$, we note that

$$\mathcal{H} = \sum_{v\in\{0,1\}^m} \frac{1}{2^m}(G^m_{-k}, v + G^m_k) = (G^m_{-k}, V + G^m_k),$$

where G^m_{-k} is uniformly distributed over $\mathbb{Z}_p^{m\times(k-1)}$, G^m_k is a vector in $\mathbb{Z}_p^m$ such that $G^m_k = -\sum_{j=1}^{k-1} G^m_j$, and V is a random variable that is uniform over $\{0,1\}^m$, that we will think of as an element in $\mathbb{Z}_p^m$. With the above notation of $\mathcal{H}, V$, we have

$$F(\mathcal{H}) = F(G^m_{-k}, V + G^m_k) = V$$

To prove the communication lower bound of a deterministic protocol Π' that has error probability $\le \delta$ w.r.t. $\mathcal{H}$, we recall the following protocol decomposition by monochromatic rectangles, c.f. Claim 24 in [23] or Lemma 1.16 in [18].

Claim 6.2 ([23], Claim 24). *A k-player (number-in-hand) deterministic protocol using communication $\le c$ partitions the inputs into $C \le 2^c$ sets of inputs $R^1, R^2, \ldots, R^C$ such that*

- *the protocol outputs the same value on inputs in the same set, and*
- *the sets are rectangles: each R^i can be written as $R^i = R^i_1 \times R^t_2 \times \ldots \times R^i_k$ where R^i_j is a subset of the inputs of Player j.*

For every $i \in [C]$ and rectangle R^i, we use the notation $R^i_{-j} := R^i_1 \times R^i_2 \times \cdots \times R^i_{j-1} \times R^i_{j+1} \times \cdots \times R^i_k$ to denote the projection of R^i on to the $k-1$ coordinates except the j-th one, for every $j \in [k]$. In particular, $R^i_{-k} := R^i_1 \times R^i_2 \times \cdots \times R^i_{k-1}$ denotes the first $k-1$ coordinates. Sometimes the index i of rectangle R^i is clear from context, for which we simply write R instead of R^i.

In what follows we show a contradiction when Π' has communication $c < \gamma m k \log k$ and hence there are $C \le 2^c < k^{\gamma m k}$ rectangles. The argument depends on the following lemma, which essentially guarantees that for every large rectangle, Π' is likely to make mistakes on more than εm coordinates.

Lemma 6.3. *For every rectangle R satisfying $\Pr_{X_{-k}\sim\mathcal{H}_{-k}}[X_{-k} \in R_{-k}] \ge \frac{1}{\alpha C} > \frac{1}{\alpha k^{\gamma m k}}$ for which $\alpha = p^{O(1)}$, there must be a set $L \subseteq [m]$ such that $|L| = (1-135\gamma)m$ and the conditional distribution $G_k^{(L)} \mid (G^m_{-k} \in R_{-k})$ is $\frac{|L|}{p}$-close to uniform over $\mathbb{Z}_p^{|L|}$; that is, the total variation distance between $G_k^{(L)} \mid (G^m_{-k} \in R_{-k})$ and the uniform distribution is at most $\frac{|L|}{p}$, where $G_k^{(L)}$ is the sequence of coordinates G_k corresponding to the instances of* SUM-EQUAL *in L.*

Lemma 6.3 implies the following claim:

Claim 6.4. *For every rectangle R on which Π' outputs $w \in \{0,1\}^m$, if $\Pr_{X_{-k}\sim\mathcal{H}_{-k}}[X_{-k} \in R_{-k}] \ge \frac{1}{\alpha C}$, then for γ, ε satisfying $1-135\gamma \ge 3\varepsilon$,*

$$\Pr_{X\sim\mathcal{H}}\Big[X \in R,\ |F(X)-w| \le \varepsilon m\Big] < \frac{1}{2}\Pr_{X\sim\mathcal{H}}[X \in R] \tag{6.2}$$

For compactness of the proof of Theorem 6.1 we defer the proofs of Claim 6.4 and Lemma 6.3 to the end of this section.

Let $\widetilde{\mathcal{R}}$ be the set of the C rectangles and $\mathcal{R} \subseteq \widetilde{\mathcal{R}}$ be the set of all large rectangles satisfying $\Pr_{X_{-k}\sim\mathcal{H}_{-k}}[X_{-k} \in R_{-k}] \ge \frac{1}{\alpha C} > \frac{1}{\alpha k^{\gamma m k}}$. Then for every rectangle $R \in \widetilde{\mathcal{R}}\backslash\mathcal{R}$,

$$\Pr_{X\sim\mathcal{H}}[X \in R] \le \Pr_{X_{-k}\sim\mathcal{H}_{-k}}[X_{-k} \in R_{-k}] < \frac{1}{\alpha C} \le \frac{1}{\alpha\big|\widetilde{\mathcal{R}}\backslash\mathcal{R}\big|}$$

Using Claim 6.4, we have

$$\begin{aligned}
&\Pr_{X\sim\mathcal{H}}\left[\left|\Pi'_{out}(X) - F(X)\right| \le \varepsilon m\right] \\
&= \sum_{R\in\widetilde{\mathcal{R}}}\Pr_{X\sim\mathcal{H}}\left[X \in R,\ \left|F(X) - \Pi'_{out}(R)\right| \le \varepsilon m\right] \\
&\le \sum_{R\in\mathcal{R}}\Pr_{X\sim\mathcal{H}}\left[X \in R,\ \left|F(X) - \Pi'_{out}(R)\right| \le \varepsilon m\right] + \sum_{R\in\widetilde{\mathcal{R}}\backslash\mathcal{R}}\Pr_{X\sim\mathcal{H}}[X \in R] \\
&\le \sum_{R\in\mathcal{R}}\frac{1}{2}\Pr_{X\sim\mathcal{H}}[X \in R] + \sum_{R\in\widetilde{\mathcal{R}}\backslash\mathcal{R}}\Pr_{X\sim\mathcal{H}}[X \in R] \\
&\le \frac{1}{2}\sum_{R\in\widetilde{\mathcal{R}}}\Pr_{X\sim\mathcal{H}}[X \in R] + \frac{1}{2}\cdot\left|\widetilde{\mathcal{R}}\backslash\mathcal{R}\right|\cdot\frac{1}{\alpha\left|\widetilde{\mathcal{R}}\backslash\mathcal{R}\right|} \le \frac{1}{2} + \frac{1}{2\alpha}
\end{aligned}$$

Combining it with (6.1), we have

$$1 - \frac{\delta}{\varepsilon} \leq \Pr_{X \sim \mathcal{H}}\left[\left|\Pi'_{out}(X) - F(X)\right| \leq \varepsilon m\right] \leq \frac{1}{2} + \frac{1}{2\alpha} \implies 1 - \frac{2\delta}{\varepsilon} \leq \frac{1}{\alpha}$$

However, the above inequality cannot be true if we set $\varepsilon = 3\delta$ and pick a constant $\alpha > 3$. Let $\gamma := (1-9\delta)/135$ be the constant for which we want to show $c \geq \gamma mk \log k = \Omega\left(mk \log k\right)$. Then $1 - 135\gamma = 9\delta \geq 3\varepsilon$ satisfies the condition in Claim 6.4 and $\alpha = O(1)$ satisfies the requirement in Lemma 6.3.

Thus we finish the contradiction argument and complete the proof of Theorem 6.1 with $\mathbf{RCC}_{k,\delta}(F) \geq \gamma mk \log k = \Omega\left(mk \log k\right)$. □

Proof of Claim 6.4. Recall that $G' := G_k^{(L)} \mid (G_{-k}^m \in R_{-k})$, G' is $|L|/p$ close to the uniform distribution by Lemma 6.3. Therefore for every fixed $u \in \mathbb{Z}_p^{|L|}$, letting $\overline{v}$ denote the complement of v (that is, we flip all of the bits in v),

$$\begin{aligned}
&\sum_{v \in \{0,1\}^{|L|} : |v| \leq \varepsilon m} \Pr\left[G' = u - v\right] \\
&= \frac{1}{2}\left(\sum_{v:|v| \leq \varepsilon m} \Pr\left[G' = u - v\right] + \sum_{v:|v| \geq |L| - \varepsilon m} \Pr\left[G' = u - \overline{v}\right]\right) \\
&\leq \frac{1}{2}\left(\sum_{v:|v| \leq \varepsilon m} \Pr\left[G' = u - v\right] + \sum_{v:|v| \geq |L| - \varepsilon m} \Pr\left[G' = u - v\right] + \frac{2|L|}{p}\right) \\
&< \frac{1}{2} \sum_{v \in \{0,1\}^{|L|}} \Pr\left[G' = u - v\right]
\end{aligned}$$

where the first inequality follows Lemma 6.3, and the last inequality holds since as long as G' is close to the uniform distribution and $|L| = (1 - 135\gamma)m \geq 3\varepsilon m$, there is

$$\sum_{v : \varepsilon m < |v| < |L| - \varepsilon m} \Pr\left[G' = u - v\right] = \Omega(1) > \frac{|L|}{p}$$

Recall that u_L and v_L denote u and v restricted to coordinates in the set L, u_{-L} and v_{-L} denote u and v restricted to coordinates not in L, and $G_k^{(-L)}$ denotes G_k restricted to coordinates not in L. We then apply the above inequality and get the following bound relating probabilities on a single coordinate conditional on the rest of the coordinates being contained in the rectangle R:

$$
\begin{aligned}
&\sum_{v\in\{0,1\}^m:|v-w|\le\varepsilon m} \Pr\left[G_k^m = u-v \,\middle|\, G_{-k}^m \in R_{-k}\right] \\
&\le \sum_{v\in\{0,1\}^m:|v_L-w_L|\le\varepsilon m} \Pr\left[G_k^m = u-v \,\middle|\, G_{-k}^m \in R_{-k}\right] \\
&= \sum_{v_L\in\{0,1\}^{|L|}:|v_L-w_L|\le\varepsilon m} \Pr\left[G_k^{(L)} = u_L-v_L \,\middle|\, G_{-k}^m \in R_{-k}\right] \\
&\quad\cdot \sum_{v_{-L}\in\{0,1\}^{m-|L|}} \Pr\left[G_k^{(-L)} = u_{-L}-v_{-L} \,\middle|\, G_{-k}^m \in R_{-k}, G_k^{(L)} = u_L - v_L\right] \\
&< \frac{1}{2}\sum_{v_L\in\{0,1\}^{|L|}} \Pr\left[G_k^{(L)} = u_L-v_L \,\middle|\, G_{-k}^m \in R_{-k}\right] \\
&\quad\cdot \sum_{v_{-L}\in\{0,1\}^{m-|L|}} \Pr\left[G_k^{(-L)} = u_{-L}-v_{-L} \,\middle|\, G_{-k}^m \in R_{-k}, G_k^{(L)} = u_L - v_L\right] \\
&= \frac{1}{2}\sum_{v\in\{0,1\}^m} \Pr\left[G_k^m = u-v \,\middle|\, G_{-k}^m \in R_{-k}\right] \qquad (6.3)
\end{aligned}
$$

The above inequality (6.3) implies (6.2) since:

$$
\begin{aligned}
&\Pr_{X\sim\mathcal{H}}\left[X\in R, |F(X)-w|\le\varepsilon m\right] \\
&= \Pr_{X_{-k}\sim\mathcal{H}_{-k}}\left[X_{-k}\in R_{-k}\right]\cdot \Pr_{X_k\sim\mathcal{H}_k}\left[X_k\in R_k, |F(X)-w|\le\varepsilon m \,\middle|\, X_{-k}\in R_{-k}\right] \\
&= \Pr\left[G_{-k}^m\in R_{-k}\right]\cdot\sum_{v\in\{0,1\}^m}\frac{1}{2^m}\Pr_{X_k\sim\mathcal{H}_k}\left[X_k\in R_k, |F(X)-w|\le\varepsilon m \,\middle|\, G_{-k}^m\in R_{-k}, F(\mathcal{H}) = v\right] \\
&= \Pr\left[G_{-k}^m\in R_{-k}\right]\cdot\sum_{v\in\{0,1\}^m}\frac{1}{2^m}\Pr\left[v+G_k^m\in R_k, |v-w|\le\varepsilon m \,\middle|\, G_{-k}^m\in R_{-k}\right] \\
&= \Pr\left[G_{-k}^m\in R_{-k}\right]\cdot\sum_{v\in\{0,1\}^m:|v-w|\le\varepsilon m}\frac{1}{2^m}\sum_{u\in R_k}\Pr\left[v+G_k^m=u \,\middle|\, G_{-k}^m\in R_{-k}\right] \\
&= \frac{1}{2^m}\Pr\left[G_{-k}^m\in R_{-k}\right]\cdot\sum_{u\in R_k}\sum_{v\in\{0,1\}^m:|v-w|\le\varepsilon m}\Pr\left[v+G_k^m=u \,\middle|\, G_{-k}^m\in R_{-k}\right] \\
&< \frac{1}{2^m}\Pr\left[G_{-k}^m\in R_{-k}\right]\cdot\sum_{u\in R_k}\frac{1}{2}\sum_{v\in\{0,1\}^m}\Pr\left[v+G_k^m=u \,\middle|\, G_{-k}^m\in R_{-k}\right] \\
&= \frac{1}{2}\Pr_{X\sim\mathcal{H}}\left[X\in R\right]
\end{aligned}
$$

Thus we have completed the proof of Claim 6.4. □

Proof of Lemma 6.3. We prove this lemma inductively for the indices in L, which we label as $1, 2, \ldots, \ell$. In what follows, let $\delta_i := \frac{i}{p}$ for every $i \in [\ell]$. Given that $\left(G_k^{(1)}, \ldots, G_k^{(\ell-1)}\right) \;\Big|\; G_{-k}^m \in R_{-k}$ is $\delta_{\ell-1}$-close to the uniform distribution over $\mathbb{Z}_p^{\ell-1}$, we will show that there exists another instance which, w.l.o.g., we label as $G_k^{(\ell)}$, for which $\left(G_k^{(1)}, \ldots, G_k^{(\ell-1)}, G_k^{(\ell)}\right) \;\Big|\; G_{-k}^m \in R_{-k}$ is δ_ℓ-close to uniform distribution over $\mathbb{Z}_p^\ell$.

The base case for $\ell = 0$ is trivial. In what follows we suppose that the conditional distribution $\left(G_k^{(1)}, \ldots, G_k^{(\ell-1)}\right) \;\Big|\; G_{-k}^m \in R_{-k}$ is already $\delta_{\ell-1}$-uniform and we do our induction for $G_k^{(\ell)}$.

First we fix $x \in \mathbb{Z}_p^{(\ell-1)\times(k-1)}$ for which $\Pr\left[\left(G_{-k}^{(1)}, \ldots, G_{-k}^{(\ell-1)}\right) = x \mid G_{-k}^m \in R_{-k}\right] \geq \frac{1}{\eta p^{(\ell-1)(k-1)}}$, where η is some value that will be specified later to get the desired property, and let $\mathcal{E}_x$ denote the event $\left(G_{-k}^{(1)}, \ldots, G_{-k}^{(\ell-1)}\right) = x$. Then we discuss the conditional distribution of the remaining instances given $\mathcal{E}_x$.

Let

$$J_x := \left\{ j \in [k-1] \;\middle|\; \Pr\left[G_j^m \in R_j \;\middle|\; \mathcal{E}_x\right] \geq \frac{1}{\beta k^{2\gamma m}} \right\},$$

Then

$$\Pr\left[G_{-k}^m \in R_{-k} \mid \mathcal{E}_x\right] = \prod_{j\in[k-1]} \Pr\left[G_j^m \in R_j \mid \mathcal{E}_x\right] \leq \prod_{j\in([k-1]\setminus J_x)} \frac{1}{\beta k^{2\gamma m}} = \left(\frac{1}{\beta k^{2\gamma m}}\right)^{k-1-|J_x|} \tag{6.4}$$

On the other hand, recalling that $\left(G_{-k}^{(1)}, \ldots, G_{-k}^{(\ell-1)}\right)$ is uniformly distributed and hence $\Pr[\mathcal{E}_x] = \frac{1}{p^{(\ell-1)(k-1)}}$, we have

$$\begin{aligned}
&\Pr\left[G_{-k}^m \in R_{-k} \mid \mathcal{E}_x\right] \\
&= \Pr\left[G_{-k}^m \in R_{-k}, \mathcal{E}_x\right] / \Pr[\mathcal{E}_x] \\
&= \Pr\left[\mathcal{E}_x \;\middle|\; G_{-k}^m \in R_{-k}\right] \cdot \Pr[G_{-k}^m \in R_{-k}] / \Pr[\mathcal{E}_x] \\
&\geq \frac{1}{\eta p^{(\ell-1)(k-1)}} \cdot \frac{1}{\alpha k^{\gamma m k}} \cdot \left(\frac{1}{p^{(\ell-1)(k-1)}}\right)^{-1} = \frac{1}{\eta \alpha k^{\gamma m k}}
\end{aligned} \tag{6.5}$$

Combining eqs. (6.4) and (6.5) and letting $\beta \geq (\eta\alpha)^{2/k}$, we can conclude

$$\begin{aligned}
&\left(\frac{1}{\beta k^{2\gamma m}}\right)^{k-1-|J_x|} \geq \frac{1}{\eta \alpha k^{\gamma m k}} \\
\implies\; &|J_x| \geq k - 1 - \frac{\gamma m k \log k + \log \eta\alpha}{2\gamma m \log k + \log \beta} \geq \frac{k \log k + \log \eta\alpha}{2 \log k + 2/k \log \eta\alpha} \geq \left(1 - \frac{\gamma}{2\gamma}\right) k - 1
\end{aligned}$$

Thus the size of J_x is at least $|J_x| \geq \left(1 - \frac{\gamma}{2\gamma}\right) k - 1 = \Omega(k)$.

For every $j \in J_x$, we have $\Pr\left[G_j^m \in R_j \;\middle|\; \mathcal{E}_x\right] \geq 1/\beta k^{2\gamma m}$ by definition of J_x and hence

$$\mathsf{H}\left[G_j^m \mid G_j^m \in R_j, \mathcal{E}_x\right] \geq \log\left(\frac{p^{m-\ell}}{\beta k^{2\gamma m}}\right) = (m-\ell)\log p - 2\gamma m \log k - \log \beta \tag{6.6}$$

The inequality in (6.6) follows by noting that G_j^m is initially uniform over $\mathbb{Z}_p^m$, which has p^m distinct values. When we condition on $\mathcal{E}_x$, we fix $\ell - 1$ of the coordinates, so there are $p^{m-\ell+1}$ possible values of G_j^m conditional on $\mathcal{E}_x$. Then, the distribution is still uniform so we must have at least $p^{m-\ell+1}/(\beta k^{2\gamma m})$ values of G_j^m satisfying $\mathcal{E}_x$ and $G_j^m \in R_j$. Thus, the entropy is at least

$$\mathsf{H}\left[G_j^m \mid G_j^m \in R_j, \mathcal{E}_x\right] \geq \log\left(\frac{p^{m-\ell+1}}{\beta k^{2\gamma m}}\right) > \log\left(\frac{p^{m-\ell}}{\beta k^{2\gamma m}}\right)$$

Note that for every $i \in [m]$, $G_j^{(i)}$ is uniform over $\mathbb{Z}_p$ as long as $j \in [k-1]$. Thus conditioned on $\mathcal{E}_x$ and $G_j^m \in R_j$, if there exists $a \in \mathbb{Z}_p$, $\Pr[G_j^{(i)} = a \mid G_j^m \in R_j, \mathcal{E}_x] = p_a > \frac{1}{2}$ then we have an upper bound for the conditional entropy of $G_j^{(i)}$:

$$\mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] \leq p_a \log \frac{1}{p_a} + (1-p_a)\log(p-1) < (1+\log(p-1))/2$$

Let $I_{j,x}$ be defined as

$$I_{j,x} := \left\{ i \in [m] \mid \mathsf{H}_\infty\left[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x\right] \geq 1 \right\} = \left\{ i \mid \forall a, \Pr\left[G_j^{(i)} = a \mid G_j^m \in R_j, \mathcal{E}_x\right] \leq \frac{1}{2} \right\}$$

where H_∞ refers to the min-entropy.

Then for all $i \in \overline{I_{j,x}} := \left(([m] \backslash [\ell-1]) \backslash I_{j,x}\right)$, $\mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] < (1+\log(p-1))/2$. Additionally for $i \in [\ell-1]$, $\mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] = 0$ since $G_j^{(i)}$ is already fixed in $\mathcal{E}_x$. Finally, for all $i \in I_{j,x}$, $\mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] \leq \log p$ since there are p possible values of $G_j^{(i)}$. As such, we can bound the conditional entropy as follows:

$$\mathsf{H}\left[G_j^m \mid G_j^m \in R_j, \mathcal{E}_x\right] \leq \sum_{i=1}^{m} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] \tag{6.7}$$

$$= \sum_{i \in I_{j,x}} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] + \sum_{i \notin I_{j,x} \cup \overline{I_{j,x}}} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] + \sum_{i \in \overline{I_{j,x}}} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] \tag{6.8}$$

$$= \sum_{i \in I_{j,x}} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] + 0 + \sum_{i \in \overline{I_{j,x}}} \mathsf{H}[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x] \tag{6.9}$$

$$\leq |I_{j,x}| \cdot \log p + (m - \ell + 1 - |I_{j,x}|)(1+\log(p-1))/2 \tag{6.10}$$

Here, (6.7) follows from the subadditivity of entropy, (6.8) splits the indices into 3 sets, and (6.9) and (6.10) use the statements we just showed above to upper bound the entropy of $G_j^{(i)}$ for i in each of the sets $I_{j,x}, \overline{I_{j,x}}$, and i in neither set.

Combining the above with the lower bound for $\mathsf{H}\left[G_j^m \mid G_j^m \in R_j, \mathcal{E}_x\right]$ in (6.6),

$$\begin{aligned}(m-\ell)\log p - 2\gamma m \log k - \log\beta &\le |I_{j,x}|\cdot \log p + (m-\ell+1-|I_{j,x}|)(1+\log(p))/2\\ \implies \left(\frac{\log p - 1}{2}\right)|I_{j,x}| &\ge (m-\ell)\left(\frac{\log p - 1}{2}\right) - 2\gamma m\log k - \frac{1+\log p}{2} - \log\beta\end{aligned}$$

Therefore, recalling that $p > k^{1/4}$ and for $\log\beta = o(\log p) = o(\log k)$, we have

$$|I_{j,x}| \ge m-\ell-\frac{4\gamma m\log k}{\log p - 1} - O\left(\frac{\log\beta}{\log p}\right) > m-\ell-\frac{4\gamma m\log k}{\frac{1}{4}\log k - 1} - o\left(1\right) > m-\ell-18\gamma m+1$$

Therefore, for every $x \in \mathbb{Z}_p^{(\ell-1)\times(k-1)}$ for which

$$\Pr\left[\left(G_{-k}^{(1)},\ldots,G_{-k}^{(\ell-1)}\right) = x \mid G_{-k}^m \in R_{-k}\right] \ge \frac{1}{\eta p^{(\ell-1)(k-1)}},$$

the size of $|J_x| \ge \left(1-\frac{\gamma}{2\gamma}\right)k-1 = \Omega\left(k\right)$; and for every $j\in J_x$, $|I_{j,x}| > m-\ell-18\gamma m+1$ and $\left|\overline{I_{j,x}}\right| = m-\ell+1-|I_{j,x}| < 18\gamma m$.

That is, these three bounds hold with probability at least $1-\frac{1}{\eta}$ by taking a union bound over all $x\in\mathbb{Z}_p^{(\ell-1)\times(k-1)}$ where

$$\Pr\left[\left(G_{-k}^{(1)},\ldots,G_{-k}^{(\ell-1)}\right) = x \mid G_{-k}^m \in R_{-k}\right] < \frac{1}{\eta p^{(\ell-1)(k-1)}}$$

for $x \sim \left(G_{-k}^{(1)},\ldots,G_{-k}^{(\ell-1)}\right) \,\Big|\, G_{-k}^m \in R_{-k}$. In what follows we abuse notation a little by assuming $\mathcal{X} := \left(G_{-k}^{(1)},\ldots,G_{-k}^{(\ell-1)}\right) \,\Big|\, G_{-k}^m \in R_{-k}$ is a distribution over $\mathbb{Z}_p^{(\ell-1)(k-1)}$ for which $\mathcal{X}$ satisfies all the above statements of J_x and $I_{j,x}$. This causes at most an additional loss of $\frac{1}{\eta}$ in the error probability.

Notice that the conditional distribution $\left(G_{-k}^{(1)},\ldots,G_{-k}^{(\ell-1)}\right) \,\Big|\, G_{-k}^m \in R_{-k}$ is indeed a product distribution since R is a rectangle. That is, letting $x = (x_1,\ldots,x_{k-1})$ where $x_j \in \mathbb{Z}_p^{\ell-1}$ for $j\in[k-1]$, then $\mathcal{E}_x$ can be decomposed into $k-1$ independent events $\mathcal{E}_{x_j}$, where each $\mathcal{E}_{x_j}$ denotes the event $\left(G_j^{(1)},\ldots,G_j^{(\ell-1)}\right) = x_j$ and $\mathcal{E}_x = \wedge_{j=1}^{k-1}\mathcal{E}_{x_j}$. Therefore the conditional distribution $G_j^m \mid \mathcal{E}_x$ is identical to $G_j^m \,\Big|\, \mathcal{E}_{x_j}$ since the distribution of G_j^m is independent from inputs of the remaining $k-2$ players (among the first $k-1$ players) in the product distribution. As a result, we have $\Pr\left[G_j^m \in R_j \,\middle|\, \mathcal{E}_x\right] = \Pr\left[G_j^m \in R_j \,\middle|\, \mathcal{E}_{x_j}\right]$ so that $\mathcal{E}_{x_j}$ and x_j fully determines whether $j \in J_x$ following the definition of J_x. Similarly we have $G_j^{(i)} \,\Big|\, \left\{G_j^m \in R_j, \mathcal{E}_x\right\}$ identical to $G_j^{(i)} \,\Big|\, \left\{G_j^m \in R_j, \mathcal{E}_{x_j}\right\}$, so that $I_{j,x}$ is also fully determined by x_j and $\mathcal{E}_{x_j}$.

Next we fix $j \in [k-1]$ and pick $x_j \in \mathbb{Z}_p^{\ell-1}$ for which $j \in J_x$ for x extended from x_j. Now we have $\mathcal{E}_{x_j}$ and $I_{j,x_j} := I_{j,x}$ containing all but a fraction of $< \frac{18\gamma m}{m-\ell+1}$ coordinates, since $\left|\overline{I_{j,x_j}}\right| < 18\gamma m$ out of the $m-\ell+1$ unfixed coordinates in total. Then for $X_j \sim \mathcal{U}_{\mathbb{Z}_p^{\ell-1}}$ and $\mathcal{I}(\cdot)$ denoting the indicator function,

$$\begin{aligned}
&\sum_{i=\ell}^{m} \mathcal{I}\left(\Pr_{X_j}\left[i \in \overline{I_{j,X_j}} \,\middle|\, j \in J_{X_j}\right] \geq \frac{1}{3}\right) \\
&\leq \sum_{i=\ell}^{m} 3 \Pr_{X_j}\left[i \in \overline{I_{j,X_j}} \,\middle|\, j \in J_{X_j}\right] \\
&= 3 \sum_{i=\ell}^{m} \sum_{x_j \in \mathbb{Z}_p^{\ell-1} : j \in J_{x_j}} \Pr[X_j = x_j] \cdot \mathcal{I}\left(i \in \overline{I_{j,x_j}}\right) \\
&= 3 \sum_{x_j \in \mathbb{Z}_p^{\ell-1} : j \in J_{x_j}} \Pr[X_j = x_j \mid j \in J_{X_j}] \cdot \sum_{i=\ell}^{m} \mathcal{I}\left(i \in \overline{I_{j,x_j}}\right) \\
&= 3 \sum_{x_j \in \mathbb{Z}_p^{\ell-1} : j \in J_{x_j}} \Pr[X_j = x_j \mid j \in J_{X_j}] \cdot \left|\overline{I_{j,x_j}}\right| < 54\gamma m
\end{aligned}$$

That is, for every fixed $j \in [k-1]$, there are at least $m-\ell+1-54\gamma m$ coordinates $i \in [m]$ satisfying $\Pr\left[i \in I_{j,X_j} \,\middle|\, j \in J_{X_j}\right] > \frac{2}{3}$, i. e., with probability $\frac{2}{3}$, $G_j^{(i)}$ satisfies $\mathsf{H}_\infty\left[G_j^{(i)} \mid G_j^m \in R_j, \mathcal{E}_x\right] \geq 1$ for a randomly selected x_j conditioned on that $j \in J_{x_j}$ specifies a big component in the rectangle. This is exactly the probability that the i-th coordinate $G_j^{(i)}$ of G_j^m can be decomposed into a convex combination of a uniform distribution over 2 elements.

Now we have at least $(m-\ell+1-54\gamma m)(k-1)$ pairs of $(i,j) \in \{\ell, \ell+1, \ldots, m\} \times [k-1]$ satisfying the above condition $\Pr\left[i \in I_{j,X_j} \,\middle|\, j \in J_{X_j}\right] > \frac{2}{3}$, which means at least one fixed i must appear in $\frac{(m-\ell+1-54\gamma m)(k-1)}{m-\ell+1} = \left(1 - \frac{54\gamma m}{m-\ell+1}\right)(k-1)$ many pairs for different $j \in [k-1]$ by a standard averaging argument. Without loss of generality we may assume $i = \ell$, and let $G'' := (G_1'', \ldots, G_k'')$ denote the conditional distribution of $G^{(\ell)}$, i. e., each $G_j'' := G_j^{(\ell)} \,\Big|\, \left\{G_j^m \in R_j, \mathcal{E}_x\right\}$ denotes the conditional distribution of $G_j^{(\ell)}$. Recalling that $|J_x| \geq \left(1 - \frac{\gamma}{2\gamma}\right)k - 1$, the number of elements in $|J_x|$ hit by those pairs containing ℓ is at least

$$\left(1 - \frac{\gamma}{2\gamma}\right)k - 1 + \left(1 - \frac{54\gamma m}{m-\ell+1}\right)(k-1) - (k-1) \geq \left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m-\ell+1}\right)k - 1 = \Omega(k)$$

We say the pair (i, j) is *good for* x if $j \in J_x$ and $i \in I_{j,x}$. Then recalling that $|J_x| \geq \left(1 - \frac{\gamma}{2\gamma}\right) k - 1$, the expected number of good (ℓ, j) over $x \sim \mathcal{X}$ is lower bounded as follows.

$$\begin{aligned}
&\mathbf{E}_x \left[\#\left\{j \in [k-1] \mid (\ell, j) \text{ is good for } x\right\}\right] \\
&= \mathbf{E}_x \left[\sum_{j=1}^{k-1} \mathcal{I}\left((\ell, j) \text{ is good for } x\right)\right] \\
&\geq \mathbf{E}_x \left[\sum_{j=1}^{k-1} \mathbf{E}_x \left[\mathcal{I}\left((\ell, j) \text{ is good for } x\right)\right]\right] = \mathbf{E}_x \left[\sum_{j=1}^{k-1} \Pr_x[\ell \in I_{j,x}, j \in J_x]\right] \\
&\geq \mathbf{E}_x \left[\sum_{j \in J_x} \Pr_x[\ell \in I_{j,x} \mid j \in J_x]\right] \\
&\geq \frac{2}{3} \cdot \left(\left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m - \ell + 1}\right) k - 1\right)
\end{aligned}$$

By a Chernoff bound it implies

$$\begin{aligned}
&\Pr_x \left[\#\left\{j \in [k-1] \mid (\ell, j) \text{ is good for } x\right\} \leq \frac{1}{3} \left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m - \ell + 1}\right) k\right] \\
&\leq \exp\left(-\Omega\left(\left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m - \ell + 1}\right) k\right)\right)
\end{aligned}$$

Let $\delta' = \exp\left(-\Omega\left(\left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m-\ell+1}\right) k\right)\right)$ be an upper bound of this error probability. Then with probability at least $1 - \delta'$, the conditional distribution G_j'' can be decomposed into a convex combination of uniform distributions over two distinct elements for at least $\frac{1}{3}\left(1 - \frac{\gamma}{2\gamma} - \frac{54\gamma m}{m-\ell+1}\right) k$ indices $j \in [k-1]$.

Next we show that conditioned on the above decomposition, which happens with probability $\geq 1 - \delta'$, the conditional distribution G_k'' is close to uniform by the following claim.

Claim 6.5 (Claim 31 in [23])**.** *Let p be a prime number. Let X be the sum of t independent random variables each uniform over $\{a_i, b_i\} \subset \mathbb{Z}_p$ for $a_i \neq b_i$. Then X modulo p is $\delta \leq 0.5\sqrt{p} \exp\left(-\Omega\left(t/p^2\right)\right)$ close to uniform.*

Remark 6.6. This claim is actually stated incorrectly in the source, where Viola claimed $O(t/p^2)$ instead of $-\Omega(t/p^2)$. However, their usage of the claim as well as their proof are consistent with the statement here. The proof of this claim simply considers primitive roots of unity and does some basic computation, but to see intuitively why this is the case, we can think of $X = ta_i + Y(b_i - a_i)$ where Y is binomial random variable with t trials of probability $1/2$. The probability mass of Y is concentrated within $O(\sqrt{t})$ of the mean, so as long as t is large enough for this range to be significantly larger than p, we will have a strong bound.

Plugging our parameters into the above claim and following exactly the same argument as in [23] (G''_k is δ''-close to uniform if every component in the above convex decomposition of G''_k is δ''-close to uniform), the statistical distance between $G''_k = -\sum_{j=1}^{k-1} G''_j$ and the uniform distribution over $\mathbb{Z}_p$ is bounded by

$$\begin{aligned}\delta'' &\le 0.5\sqrt{p}\exp\left(-\Omega\left(\frac{1}{3}\left(1-\frac{\gamma}{2\gamma}-\frac{54\gamma m}{m-\ell+1}\right)k/p^2\right)\right)\\ &= \exp\left(-\Omega\left(\left(1-\frac{\gamma}{2\gamma}-\frac{54\gamma m}{m-\ell+1}\right)\sqrt{k}\right)\right)\end{aligned}$$

Putting it all together, we conclude that $G_k^{(\ell)} \,\Big|\, \left\{G_{-k}^m \in R_{-k}, G^{(1)}, \ldots, G^{(\ell-1)}\right\}$ is close to uniform, which implies $\left(G_k^{(1)}, \ldots, G_k^{(\ell-1)}, G_k^{(\ell)}\right) \,\Big|\, G_{-k}^m \in R_{-k}$ is also close to uniform. Moreover, its statistical distance to uniform is bounded by

$$\delta_\ell \le \delta_{\ell-1} + \frac{1}{\eta} + \delta' + \delta''$$

Let $\eta = 2p$ and $\beta = 2 \ge (\eta\alpha)^{2/k} = 2^{O\left(\frac{\log k}{k}\right)}$ for $\alpha = p^{O(1)}$. Then for sufficiently large km the above induction argument goes through for $\ell \le (1-135\gamma)m$, with error δ', δ'' bounded by

$$\delta' = \exp\left(-\Omega\left(k\right)\right), \delta'' = \exp\left(-\Omega\left(\sqrt{k}\right)\right) \Longleftarrow 1-\frac{\gamma}{2\gamma}-\frac{54\gamma m}{m-\ell+1} \ge 0.1 \Longleftrightarrow \ell \le (1-135\gamma)m+1$$

Therefore the conditional distribution $\left(G_k^{(1)}, \ldots, G_k^{(\ell-1)}, G_k^{(\ell)}\right) \,\Big|\, G_{-k}^m \in R_{-k}$ is δ_ℓ-close to uniform for δ_ℓ bounded by $\frac{\ell}{p}$ as follows:

$$\delta_\ell \le \delta_{\ell-1} + \frac{1}{\eta} + \delta' + \delta'' \le \frac{\ell-1}{p} + \frac{1}{2p} + \exp\left(-\Omega\left(\sqrt{k}\right)\right) \le \frac{\ell}{p}$$

Thus we have proved the induction hypothesis for every $\ell \le (1-135\gamma)m$. Letting L be the first $(1-135\gamma)m$ indices as in the induction hypothesis, we complete the proof of Lemma 6.3 for $|L| = (1-135\gamma)m$ and statistical distance $\frac{|L|}{p}$. □

7 Lower bound for Hamming norm estimation

In this section we present a space lower bound for single-pass streaming algorithms for $(1\pm\varepsilon)$-approximating the Hamming norm L_0 in the strict turnstile model, which is denoted by T_ε as in Section 1.1.1.

Formally, in the Hamming norm estimation problem there is an underlying vector $(x_1, \ldots, x_N)$ which starts from the all zero vector and processes up to m updates each of the form $(i, v) \in [N] \times [\pm M]$. The update (i, v) means one should add v to the i-th coordinate x_i in the vector x.

After processing all m updates, we have $\|x\|_0 = |\{i \mid x_i \neq 0\}|$ and we want to output a number within $(1 \pm \varepsilon)\|x\|_0$ with probability $\geq 2/3$. We additionally assume all players have access to a heavy hitters oracle, which tells them whether the frequency of a given coordinate is greater than T. This is a generalization of the case without a heavy hitters oracle, where we simply let $T = mM$ and we know that all frequencies are guaranteed to be smaller. The strict turnstile model guarantees that $x_i \geq 0$ for all $i \in [N]$ at all positions in the stream, in which case it suffices to prove the space lower bound in the simultaneous communication model following the reduction in Theorem 4.1 of [2]. Furthermore, it is also guaranteed that for every $i \in [N]$, $x_i \leq \mathsf{poly}(n)$ at the end of the stream. In this setting, the algorithm of [16] approximates $\|x\|_0$ up to a $(1 \pm \varepsilon)$ factor with $O\left(\varepsilon^{-2}\log(N)\left(\log(1/\varepsilon) + \log\log(T)\right)\right)$ bits of space[6], as long as $\varepsilon > 0$.

We first note that solving distinct elements with a heavy hitters oracle reduces to solving distinct elements given a threshold on the frequency of the coordinates. As such, we will solve the complexity question of space complexity given a threshold T for the frequency.

Theorem 7.1. *The space complexity of $(1 \pm \epsilon)$ approximating L_0 with probability at least $2/3$ in a strict turnstile stream with access to a heavy hitters oracle with a threshold of $T > 1$ is $\Omega(\epsilon^{-2}\log N \log\log T)$.*

We note that the assumption $T > 1$ is necessary for this bound to be well defined. When $T = 1$, the heavy hitters oracle tells us exactly whether or not the frequency of a coordinate is 0 at the end of the stream, so the complexity is $\Theta(\log N)$. This lower bound follows as we need to write down the answer and the upper bound follows as we can directly count the elements with nonzero frequency.

To prove this theorem, we first prove the following lemma:

Lemma 7.2. *The space complexity of $(1 \pm \epsilon)$ approximating L_0 with probability at least $2/3$ in a strict turnstile stream with access to a heavy hitters oracle with a threshold of $T > 1$ is at least $RSC^T_{2/3}(T_\epsilon)$.*

Proof. Suppose we have an algorithm A which gives us a $(1 \pm \epsilon)$ approximation of L_0 in a strict turnstile stream with access to a heavy hitters oracle with a threshold of T.

Now, if we are given an input where the maximum frequency of any element is at most T, then we can go through our input and do exactly what A would do for everything other than calls to the heavy hitters oracle. If A would make a call to a heavy hitters oracle, we just treat the answer as 0 without making this query and proceed as A would.

Since we assumed the input has a maximum frequency of T, the heavy hitters oracle would return 0 for every element, so this would give us the same answer as A, and by correctness of A, it is a $(1 \pm \epsilon)$ approximation. □

Now, we will state and prove our main theorem:

[6]Indeed, their algorithm stores $O\left(\varepsilon^{-2}\log N\right)$ counters modulo primes that are each $O\left(\log(1/\varepsilon) + \log\log(T)\right)$ bits in magnitude, and it does not matter how large the values of x_i are at intermediate positions in the stream.

Theorem 7.3. *For error tolerance $\varepsilon < 1/3$ and $\varepsilon = \max\left\{\Omega\left(\sqrt{\frac{\log k}{k}}\right), \frac{1}{N^{0.49}}\right\}$, any single-pass streaming algorithm solving T_ε with probability $\geq 2/3$ in the strict turnstile model must use $\Omega\left(\varepsilon^{-2}\log(N)\log\log(T)\right)$ bits of space.*

First we introduce some supplementary problems that will be used in reductions:

Definition 7.4. In the c-GAP-ORT$_n$ problem, we have two players Alice and Bob. They each have as input a vector in $\{0,1\}^n$ and we wish to compute

$$c\text{-}\textsc{Gap-Ort}_n(x,y) = \begin{cases} 1, & \left|(\sum_{i\in n}\mathbf{XOR}(x_i,y_i)) - \frac{n}{2}\right| \geq 2c\sqrt{n}, \\ 0, & \left|(\sum_{i\in n}\mathbf{XOR}(x_i,y_i)) - \frac{n}{2}\right| \leq c\sqrt{n}, \end{cases}$$

and otherwise, it can return anything.

Definition 7.5. In the c-GAP-ORT-SUM-EQUAL$_{n,k}$ problem, we have k players. The i^{th} player has input $(x_{i,1}, x_{i,2}, \ldots, x_{i,n}) \in \mathbb{Z}^n$. Then we wish to compute

$$c\text{-}\textsc{Gap-Ort-Sum-Equal}_n(x,y) = \begin{cases} 1, & \left|\left(\sum_{j\in[n]}\mathbf{1}_{x_{1,j}+x_{2,j}+\cdots+x_{k,j}=0}\right) - \frac{n}{2}\right| \geq 2c\sqrt{n}, \\ 0, & \left|\left(\sum_{j\in[n]}\mathbf{1}_{x_{1,j}+x_{2,j}+\cdots+x_{k,j}=0}\right) - \frac{n}{2}\right| \leq c\sqrt{n}, \end{cases}$$

and in other cases it can return either 0 or 1. We let $\mathbf{1}_{x_{1,j}+x_{2,j}+\cdots+x_{k,j}=0}$ denote the indicator function which is 1 if $x_{1,j} + x_{2,j} + \cdots + x_{k,j} = 0$ and 0 otherwise.

We will often be working with the 2-player problem c-GAP-ORT-SUM-EQUAL$_{n,2}$, which we simply denote by c-GAP-ORT-SUM-EQUAL$_n$.

Additionally, we will let SUM-EQUAL$_{k,\delta}^{m,a}$ denote the problem where we have m independent instances of SUM-EQUAL$_{k,\delta}$, and to solve it, our protocol needs to be able to solve at least am of these instances correctly with probability at least $1-\delta$.

Definition 7.6. The AUG-INDEX-GOSE$_{n,k}^t$ problem consists of t independent instances of $10\epsilon\sqrt{n}$-GAP-ORT-SUM-EQUAL$_n$, denoted $g_1, g_2, \ldots g_t$, with k players and n coordinates each.[7] In this problem, the referee is asked to estimate g_i based on an index $i \in [t]$ together with the auxiliary information of $f_{i+1}, \ldots, f_t$, where we let $f_i \in [\pm n]$ is defined as follows:

Let a be the number of underlying SUM-EQUAL$_k$ instances in g_i outputting 1, and let b be the number of underlying SUM-EQUAL$_k$ instances in g_i outputting 0. Then, $f_i = a - b$.

To prove Theorem 7.3, we combine the following statements:

(1) T_ϵ reduces to AUG-INDEX-GOSE$_{n,k}^t$

[7] Note that AUG-INDEX-GOSE$_{n,k}^t$ implicitly depends on the value of ε even though we do not take ε as a parameter since in this work, we will only be using it with 1 value of ε.

(2) AUG-INDEX-GOSE$^t_{n,k}$ reduces to $10\epsilon\sqrt{n}$-GAP-ORT-SUM-EQUAL$^t_{n,k}$.

(3) $10\epsilon\sqrt{n}$-GAP-ORT-SUM-EQUAL$^t_{n,k}$ reduces to 1-GAP-ORT-SUM-EQUAL$^t_{n',k}$ for some n' which will be defined later.

(4) k-player communication complexity in the linear model is bounded by $k-1$ times the 2-player communication complexity in the linear model with a specific input distribution.

(5) For a particular hard input distribution μ, 1-GAP-ORT-SUM-EQUAL$^t_{n',2}$ reduces to SUM-EQUAL$^{tn',c}_2$.

(6) We can directly bound the communication complexity of SUM-EQUAL$^{tn',c}_2$ on said distribution.

Of these statements, (2) follows almost immediately from the definition of AUG-INDEX-GOSE and (4) follows by definition of the linear sketch model of communication. This will be explained in more detail when we combine all of the above parts to bound the complexity of $\mathbf{RSC}^T_{k,0,4}(T_\epsilon)$.

The following lemma proves (3)

Lemma 7.7. *for every $k \in \mathbb{N}$, $0 \le \delta \le 1/2$, and $n \ge \frac{c^2}{100\epsilon^2} = n'$,*

$$\mathbf{RCC}^{LIN,T}_{k,\delta}(10\epsilon\sqrt{n}\text{-}\textit{GAP-ORT-SUM-EQUAL}_{n,k}) \ge \mathbf{RCC}^{LIN,T}_{k,\delta}(c\text{-}\textit{GAP-ORT-SUM-EQUAL}_{n',k})$$

Proof of Lemma 7.7. Given $n' = \frac{c^2}{100\varepsilon^2}$ and an input instance of c-GAP-ORT-SUM-EQUAL$_{n'}$ with underlying SUM-EQUAL problems outputting $\mathbf{x}' \in \{0,1\}^{n'}$, we create the new input to $10\varepsilon\sqrt{n}$-GAP-ORT-SUM-EQUAL$_n$ by taking $100\varepsilon^2 n/c^2$ copies of each coordinate, with results of underlying problems being $\mathbf{x} \in \{-1,1\}^n$ where we map each output of 0 to -1. As a result, $\sum_{j=1}^n \mathbf{x}_j = \frac{100\varepsilon^2 n}{c^2} \cdot \sum_{j=1}^{n'} \mathbf{x}'_j$.

If $|\sum_j \mathbf{x}'_j| \le c\sqrt{n'}$, then $|\sum_j \mathbf{x}_j| \le 10\varepsilon n$, and on the other hand $|\sum_j \mathbf{x}'_j| \ge 2c\sqrt{n'}$ implies $|\sum_j \mathbf{x}_j| \ge 20\varepsilon n$.

Thus, any k-player δ-error simultaneous communication protocol for $10\varepsilon\sqrt{n}$-GAP-ORT-SUM-EQUAL$_n$ immediately implies a k-player δ-error simultaneous communication protocol for c-GAP-ORT-SUM-EQUAL$_{n'}$.

Since all we are doing is copying coordinates, this does not change the threshold. $\square$

Now, we prove (5)

Theorem 7.8. *Given some simultaneous communication protocol Π with two players that solves 1-GAP-ORT-SUM-EQUAL$_n$ when each SUM-EQUAL instance has input drawn from the distribution μ, where $\mu := (G/2 + B/2)^m$ consists of m independent copies of $G/2 + B/2$ (here and later, we use this notation* *to denote a random variable being drawn from G with probability $\frac{1}{2}$ and B with probability $\frac{1}{2}$), for G, B defined as follows:*

$$\begin{cases} G := & \left(G_1, \ldots, G_{k-1}, \quad -\sum_{j=1}^{k-1} G_j\right) \\ B := & \left(B_1, \ldots, B_{k-1}, \quad M - \sum_{j=1}^{k-1} B_j\right) \end{cases}$$

there exists a protocol Π' such that $\mathbf{RCC}_{2,\delta}^{LIN}(\Pi') \le O(\mathbf{RCC}_{2,\delta}^{LIN}(\Pi))$ which solves $\Omega(n)$ of the individual SUM-EQUAL *instances with probability at least $\frac{1+\alpha}{2}$ for some constant $\alpha > 0$.*

Proof. Suppose Π is a protocol that solves 1-GAP-ORT-SUM-EQUAL$_n$. Now, if Alice has input $X = (x_1, x_2, \ldots x_n)$ and Bob has input $Y = (y_1, y_2, \ldots y_n)$ to 1-GAP-ORT-SUM-EQUAL$_n$, we define a corresponding instance of 1-GAP-ORT$_n$ where Alice gets input $X' = (x'_1, x'_2, \ldots x'_n)$ and Bob gets input $Y' = (y'_1, y'_2, \ldots y'_n)$. We define $x'_i = 1 - y'_i$ iff $x_i + y_i = 0$ and $y'_i = 0$ with probability 1 if $y_i < M/2$, probability $\frac{1}{2}$ if $y_i = M/2$, and probability 0 otherwise, where $M = a!$ for the value a that is the largest integer such that $a! < T$.

When $X, Y \sim \mu$, each $x_i + y_i = 0$ with probability $\frac{1}{2}$ and y_i is equally likely to be $-x_i$ or $M - x_i$ so it is symmetric about $M/2$. Hence, $(X', Y') \sim \{0,1\}^{2n}$. Furthermore, the answer to the 1-GAP-ORT-SUM-EQUAL$_n$ instance is the same as the answer to the 1-GAP-ORT$_n$ instance by construction. Therefore, we can solve this instance of 1-GAP-ORT$_n$ by simply running Π.

Now, if we let $\mathcal{M}$ be the message sent by Alice to Bob in protocol Π, then

$$I(X'; \mathcal{M}, Y) \ge \mathbf{IC}_{2,\delta}(\text{1-Gap-Ort}_n) = \Omega(n)$$

since Bob can solve 1-GAP-ORT$_n$ where Alice has input X' and Bob has input Y' when he has access to $(\mathcal{M}, Y)$ by returning the answer to 1-GAP-ORT-SUM-EQUAL$_n$ using protocol Π with input Y after being sent the message $\mathcal{M}$.

We now note that X' is n iid uniformly random bits. As such,

$$I(X'; \mathcal{M}, Y) = \sum_{i=1}^{n} I(x'_i; \mathcal{M}, Y \mid x'_1, x'_2, \ldots x'_{i-1}) \ge \sum_{i=1}^{n} I(x'_i; \mathcal{M}, Y).$$

Each of these terms is upper bounded by 1, so in order for the sum to be $\Omega(n)$, there exists some constant $c > 0$ such that there are at leasts cn indices j such that $I(x'_j; \mathcal{M}, Y) \ge \alpha$ for some constant $\alpha > 0$.

Now, let

$$J = \{j \mid I(x'_j; \mathcal{M}, Y) \ge \alpha\}.$$

We claim that the transcript of Π must contain the solution to the j^{th} Sum-Equal instance with probability at least $\frac{1+\alpha}{2}$ for each j. To see this, we note that Bob has as input Y for 1-GAP-ORT-SUM-EQUAL$_n$ so he can compute y'_j. Then, we note that

$$H(x'_j \mid \mathcal{M}, Y) = H(x'_j) - I(x'_j; \mathcal{M}, Y) \le 1 - \alpha$$

Since $x'_j \in \{0,1\}$, let $\Pr[x'_j = 0 \mid \mathcal{M}, Y] = p$. Then, if $p = 0$, the entropy is 0 so this is satisfied for any $0 < \alpha \le 1$. If $p > 0$, we have

$$-(p \log p + (1-p)\log(1-p)) \le 1 - \alpha.$$

Since this is symmetric about $p = \frac{1}{2}$ and cannot be satisfied by $p = \frac{1}{2}$ since $\alpha > 0$, we assume WLOG that $p < \frac{1}{2}$, in which case the entropy monotonically decreases as p decreases. Now, we claim that we must have $p < \frac{1}{2} - \frac{\alpha}{2}$. It suffices to show that

$$-\left(\left(\frac{1-\alpha}{2}\right)\log\left(\frac{1-\alpha}{2}\right) + \left(\frac{1+\alpha}{2}\right)\log\left(\frac{1+\alpha}{2}\right)\right) \ge 1 - \alpha$$

Simplifying this expression yields the solution

$$0 < \alpha < 1.$$

Thus, for $0 < \alpha < 1$, we must have $p < \frac{1-\alpha}{2}$. If $\alpha = 1$, then the entropy is 0 so we must have $p = 0$. Thus, we get that $p \le \frac{1-\alpha}{2}$.

By symmetry, we thus have that either

$$\Pr[x'_j = 0 \mid \mathcal{M}, Y] \le \frac{1-\alpha}{2}$$

or

$$\Pr[x'_j = 0 \mid \mathcal{M}, Y] \ge \frac{1+\alpha}{2}.$$

In the former case, Bob lets $\hat{x}'_j = 1$ and in the latter case, Bob lets $\hat{x}'_j = 0$. Bob then computes y'_j from y_j. Then, if $\hat{x}'_j = y'_j$, Bob concludes that $x_j + y_j \neq 0$ and if $\hat{x}'_j = 1 - y'_j$, Bob concludes that $x_j + y_j = 0$. By construction, this succeeds with probability at least $\frac{1+\alpha}{2}$, and all we did was run Π and compute the value from the transcript. □

Corollary 7.9. *There exist constants $\alpha, c > 0$ such that*

$$\mathbf{D}^{LIN,T}_{2,\delta,\mu'}\left(1\text{-}\mathit{GAP\text{-}ORT\text{-}SUM\text{-}EQUAL}_{\epsilon^{-2}/100}\right) \ge \mathbf{D}^{LIN,T}_{2,(1+\alpha)/2,\mu}\left(\mathit{SUM\text{-}EQUAL}_2^{\epsilon^{-2}/100,c}\right)$$

Proof. This follows directly from Theorem 7.8. The protocol Π' solves $\Omega(n)$ of the individual sum-equal instances, so there is some constant c such that Π' solves at least cn of them with probability at least $\frac{1+\alpha}{2}$. Each instance of Sum-Equal corresponds to a single coordinate from 1-GAP-ORT-SUM-EQUAL so their frequencies must all be bounded by T as well. □

We can now prove (6)

Theorem 7.10. *When $\delta < \frac{1}{2}$ and a is some constant fraction,*

$$\mathbf{IC}^{T}_{k,\delta}(\textsc{Sum-Equal}^{n',a}_{k}) \geq \Omega(n'k \log\log T) \tag{7.1}$$

where $\textsc{Sum-Equal}^{n',a}_{k}$ *is the problem where we are given n' independent instances of* Sum-Equal *and we are asked to solve an' of them with probability $1-\delta$ each.*

The proof of this theorem can be found in Appendix B.

Corollary 7.11. *For the input distribution μ defined in the proof of Theorem 7.10, $\delta < \frac{1}{2}$, and $0 < a < 1$,*

$$\mathbf{D}^{LIN,T}_{2,\delta,\mu}(\textsc{Sum-Equal}^{n',a}_{2}) \geq \Omega(n' \log\log T)$$

Proof. If we plug $k = 2$ into (7.1), we get

$$\mathbf{D}^{LIN,T}_{2,\delta,\mu}(\textsc{Sum-Equal}^{n',a}_{2}) \geq \mathbf{IC}^{T}_{2,\delta}(\textsc{Sum-Equal}^{n',a}_{2}) \geq \Omega(n' \log\log T)$$

since by definition μ is the hard distribution from which we got the information complexity bound. □

And finally, we prove (1)

Theorem 7.12. $\mathbf{RCC}^{LIN,T}_{k,1/3}(T_\epsilon) \geq \mathbf{RCC}^{LIN,T}_{k,0.4}(\textit{Aug-Index-GOSE})$

Proof. Suppose we have a protocol that solves T_ϵ. Then, from the input of $\textsc{Aug-Index-GOSE}^{t}_{n,k}$, we construct an input to T_ϵ as follows:

For the i-th $10\varepsilon\sqrt{n}$-$\textsc{Gap-Ort-Sum-Equal}_n$ instance g_i in the $\textsc{Aug-Index-GOSE}^{t}_{n,k}$ problem, we construct 100^{i-1} distinct copies of every element in the input. We take the concatenation of all of these inputs as our input to T_ϵ. Thus the universe contains $N := n + 100 \cdot n + \cdots + 100^{t-1} \cdot n \leq 100^t n/99$ distinct elements in total, which is $N \leq n^{1.01}$ for sufficiently small t (and hence $1/N^{0.49} > 1/\sqrt{n}$). The final Hamming norm is a weighted sum $F' := \sum_{i=1}^{t} 100^{i-1} f'_i$. The advantage of F' (that is, the difference between the number of 1s and 0s in this stream) is hence $F := 2F' - N = \sum_{i=1}^{t} 100^{i-1} f_i$.

Then we invoke the simultaneous communication protocol for T_ε to estimate F', which returns a value $\widetilde{F'}$ satisfying $(1-\varepsilon)F' \leq \widetilde{F'} \leq (1+\varepsilon)F'$. Translating to the advantage we get $\left|\widetilde{F} - F\right| \leq 2\varepsilon F' \leq 2\varepsilon N$. From this approximated value $\widetilde{F}$, together with the index i and auxiliary information $f_{i+1}, \ldots, f_t$, we need to determine the output value of g_i. Since the influence of f_j with $j > i$ can be precisely removed from F before getting the approximated norm $\widetilde{F}$, in what follows it suffices to consider the estimation of g_t when the index is indeed $i = t$. Recall that $F = 100^{t-1} f_t + \sum_{i=1}^{t-1} 100^{i-1} f_i$, and thus $\widetilde{F}$ is also an approximation of $100^{t-1} f_t$ as long as the additive error $\sum_{i=1}^{t-1} 100^{i-1} f_i$ is bounded.

Let the input distribution to every f_i be padded from the 1-$\textsc{Gap-Ort-Sum-Equal}_{\epsilon^{-2}}$ distribution μ' as in Theorem 7.8, where the coordinates are iid bits drawn uniformly from $\{0, 1\}$. Thus, each f_i has expectation 0 and variance $25\epsilon^2 n^2$ It immediately follows by Chebyshev's inequality that $\Pr[|f_i| \geq 50\varepsilon n] \leq 1/100$. Similarly, $\Pr\left[|f_i| \geq 50^j \varepsilon n\right] \leq 1/100^j$. Therefore,

$$\Pr\left[\left|\sum_{i=1}^{t-1} 100^{i-1} f_i\right| > 100^{t-1}\varepsilon n\right] \leq \sum_{i=1}^{t-1} \Pr\left[|f_{t-i}| > 50^i \varepsilon n\right] \leq \sum_{i=1}^{t-1} \frac{1}{100^i} \leq \frac{1}{99} \tag{7.2}$$

where the first inequality holds because if $|f_{t-i}| \leq 50^i \varepsilon n$ for every i, then $\left|\sum_{i=1}^{t-1} \cdot 100^{i-1} f_i\right| \leq \sum_{i=1}^{t-1} 100^{i-1} \times 50^{t-i} \varepsilon n \leq \frac{50^t}{100} \varepsilon n \sum_{i=1}^{t-1} 2^i < 100^{t-1}\varepsilon n$.

Notice that as long as $\widetilde{F}$ is a $(1 \pm \varepsilon)$-approximation of F, we must have $\left|\widetilde{F} - F\right| \leq 2\varepsilon N$. Furthermore suppose that we return 0 if $\widetilde{F} < 15 \cdot 100^{t-1}\epsilon n$ and 1 if $\widetilde{F} \geq 15 \cdot 100^{t-1}\epsilon n$. Since we know that $N \leq 100^t n/99$, we have

$$2\varepsilon N \leq 2\epsilon 100^t n/99 < 3 \cdot 100^{t-1}\epsilon n.$$

So in particular, if T_ε succeeds, if $g_t = 0$, we have $|f_t| \leq 10 \cdot 100^{t-1}\epsilon n$, so $|F| = |\sum_{i=1}^t 100^{i-1} f_i| \leq 11 \cdot 100^{t-1}\epsilon n$ with probability at least $\frac{98}{99}$. Then, $|\tilde{F}| < 14 \cdot 100^{t-1}\epsilon n$ and our algorithm succeeds.

Similarly, if $g_t = 1$, we have $|f_t| \geq 20 \cdot 100^{t-1}\epsilon n$. Thus, with probability at least $\frac{98}{99}$, $|F| = |\sum_{i=1}^t 100^{i-1} f_i| \geq 19 \cdot 100^{t-1}\epsilon n$, so $|\tilde{F}| > 16 \cdot 100^{t-1}\epsilon n$ and our algorithm succeeds.

Thus, if T_ε succeeds with probability $\frac{2}{3}$, the above algorithm succeeds with probability $\frac{2}{3} \cdot \frac{98}{99} > 0.6$. Thus we can determine the value of g_t with probability ≥ 0.6. The thresholds stay the same because all we did to change the input was copy coordinates, which does not change the frequencies. Hence,

$$\mathbf{RCC}_{k,1/3}^{LIN,T}(\mathrm{T}_\varepsilon) \geq \mathbf{RCC}_{k,0.4}^{LIN,T}\left(\textsc{Aug-Index-GOSE}_{n,k}^t\right).$$

□

Finally, we must fill in the missing statements 2) and 4) to bound the streaming complexity $\mathbf{RSC}_{k,0.4}^T(T_\epsilon)$. To do this, we conclude as follows:

When we solve Aug-Index-GOSE, we claim that in order to solve Aug-Index-GOSE with probability at least 0.6 on every input, we must solve every instance of Gap-Ort-Sum-Equal with probability at least 0.6. This is statement 2) in the overview, and can be proved as follows:

Suppose we have a protocol that solves Aug-Index-GOSE with probability at least 0.6. Now, consider the j^{th} instance of Gap-Ort-Sum-Equal. Since the protocol succeeds with probability at least 0.6 on any input, we can simply consider any input to Aug-Index-GOSE where the index is j. By assumption, our protocol succeeds with probability at least 0.6, so it solves the j^{th} instance of Gap-Ort-Sum-Equal with probability at least 0.6. This holds for every j, so our protocol must solve every instance of Gap-Ort-Sum-Equal with probability at least 0.6.

To see why statement 4) in the overview is true, we wish to prove that

$$\mathbf{D}_{k,\delta,\mu}^{LIN,T}\left(\text{1-GAP-ORT-SUM-EQUAL}_{n',k}^{t}\right) \geq (k-1)\mathbf{D}_{2,\delta,\mu}^{LIN,T}\left(\text{1-GAP-ORT-SUM-EQUAL}_{n',2}^{t}\right)$$

and this holds by definition of the linear sketch model of communication: recall that a protocol consists of some matrix A where every player simply multiplies their input by A and sends the resulting vector to the referee. For any matrix A, number of bits communicated by each player will be the same regardless of the number of players, so we get this equivalence.

Thus, putting all of this together, we get the chain of inequalities:

$$\begin{aligned}
\mathbf{RCC}_{k,2/3}^{LIN,T}(T_\epsilon) &\geq \mathbf{RCC}_{k,0.4}^{LIN,T}\left(\text{AUG-INDEX-GOSE}_{n,k}^{t}\right) \\
&\geq \mathbf{RCC}_{k,0.4}^{LIN,T}\left(10\epsilon\sqrt{n}\text{-GAP-ORT-SUM-EQUAL}_{n,k}^{t}\right) \\
&\geq \mathbf{RCC}_{k,0.4}^{LIN,T}\left(\text{1-GAP-ORT-SUM-EQUAL}_{n',k}^{t}\right) \\
&\geq \mathbf{D}_{k,\delta,\mu}^{LIN,T}\left(\text{1-GAP-ORT-SUM-EQUAL}_{n',k}^{t}\right) \\
&\geq (k-1)\mathbf{D}_{2,\delta,\mu}^{LIN,T}\left(\text{1-GAP-ORT-SUM-EQUAL}_{n',2}^{t}\right) \\
&\geq (k-1)\mathbf{D}_{k,\delta,\mu}^{LIN,T}\left(\text{SUM-EQUAL}_{2}^{tn',c}\right) \\
&\geq \Omega(ktn' \log\log T) = \Omega(\epsilon^{-2} k \log n \log\log T)
\end{aligned}$$

so

$$\mathbf{RSC}_{k,0.4}^{T}(T_\epsilon) \geq \frac{1}{k}\mathbf{RCC}_{k,0.4}^{LIN,T}(T_\epsilon) \geq \Omega(\epsilon^{-2}\log n \log\log T) = \Omega(\epsilon^{-2} k \log N \log\log T)$$

□

A Communication upper bound for EQUALITY

The standard δ-error protocol solving the EQUALITY problem starts by sending and comparing the digest under a random hash function $h : [p] \to [q]$ where $q = O\left(\delta^{-1}\log p\right)$. For example, let q be a random prime drawn from the interval $[\delta^{-2}\log^2 p, 2\delta^{-2}\log^2 p]$ and let h compute a number modulo q. By the prime number theorem there are at least $2\sqrt{N}$ primes in the interval $[N, 2N]$, which implies the existence of $2\delta^{-1}\log(p)$ distinct primes in that range. For any two distinct numbers $x, y \in \mathbb{Z}_p$, since $z = x - y$ has no more than $\log|z| \leq \log p$ prime factors, the error probability of the protocol is bounded by the collision probability of h as follows:

$$\Pr_q[h(x) = h(y)] = \Pr_q[x \equiv y \pmod q] = \Pr_q[q|(x-y)] \leq \frac{\log p}{2\delta^{-1}\log p} < \delta$$

The communication is a message of the form $(h, h(x))$ (indeed $(q, x \bmod q)$ in the above example), whose length is at most $2\lceil\log q\rceil = O\left(\log(1/\delta) + \log\log p\right) = O\left(\log(1/\delta) + \log\log k\right)$

bits. In particular this is an upper bound for one-way communication protocols computing EQUALITY. Recalling that $p = \Theta\left(k^{1/4}\right)$, we can conclude

$$\mathbf{RCC}_{2,\delta}(f) \leq \overrightarrow{\mathbf{RCC}}_{2,\delta}(f) = O\left(\log(1/\delta) + \log\log k\right)$$

We note that the $1/\delta$ factor in q is unavoidable, since otherwise more than an δ fraction of numbers would share the same message and hence the collision probability, as well as the error probability, would exceed δ.

B The lower bound for SUM-EQUAL$_k^{m,a}$ over integers

Theorem 7.10 (restated). *Let Π be the δ-error simultaneous k-player protocol for solving the* SUM-EQUAL$_k^{m,a'}$ *problem, where $m \leq \frac{k \log\log T}{20 \log k}$ and the error tolerance $\delta \in (0, 1/6)$. The simultaneous communication complexity of Π is $\mathbf{RCC}_{k,\delta}^{LIN,T}(\Pi) = \Omega\left(mk \log\log T\right)$.*

Proof. To prove the $\Omega\left(mk \log\log T\right)$ lower bound we will deduce a contradiction if Π uses $c < \gamma mk \log\log T$ bits of communication, for a *sufficiently small* constant γ. By decreasing γ we may assume that k is arbitrarily large.

For the hard distribution we first introduce a magnitude bound a defined to be the largest integer such that $a! \leq \min(k^{1/8}, T)$. We define $M = a!$. Note that $\log a = \Omega(\log\log T)$ as $M \geq \frac{T}{a+1}$, so $a \log a = \Omega(\log M) = \Omega(\log T)$. Taking the log of both sides, we have $\log(a \log a) = \log a + \log\log a < 2 \log a$, so $\log a = \Omega(\log\log T)$.

Now we specify the distribution $\mathcal{H}$ for the SUM-EQUAL$_k$ instances. $\mathcal{H} := (G/2 + B/2)^m$ consists of m independent copies of $G/2 + B/2$, for G, B defined as follows:

$$\begin{cases} G := & \left(G_1, \ldots, G_{k-1}, \quad -\sum_{j=1}^{k-1} G_j\right) \\ B := & \left(B_1, \ldots, B_{k-1}, \quad M - \sum_{j=1}^{k-1} B_j\right) \end{cases}$$

where G_j, B_j are uniformly and independently chosen from $[a]$ for every $j \in [k-1]$. Note that:

(a) SUM-EQUAL$_k(G) = 0$, SUM-EQUAL$_k(B) = 1$;

(b) the first $k-1$ elements of G and B, denoted by G_{-k} and B_{-k}, are the same uniform distribution over $[a]^{k-1}$. Thus we can write $B = (G_{-k}, M + G_k)$

(c) for $j \in [k-1]$, the j-th player's input $\mathcal{H}_j$ is uniform over $[a]^m$ and independent from other players' input.

Besides $\mathcal{H}_k$, the referee gets in addition an index n uniformly drawn from $[m]$ together with the answers $Y^{(j)} =$ SUM-EQUAL$_k(X^{(j)})$ for $j = n+1, \ldots, m$. Let $\mathcal{H}'_n := (\mathcal{H}, Y^{(n+1)}, \ldots, Y^{(m)})$ and the hard input distribution is defined as $\mathcal{H}' := \sum_{n=1}^{m} \frac{1}{m} \cdot \mathcal{H}'_n$.

Now we derandomize the protocol Π by fixing the randomness and thus get an δ-error deterministic protocol Π' with respect to the above input distribution. That is, Π' outputs $\textsc{Sum-Equal}_k^{(n)} = \textsc{Sum-Equal}_k(X^{(n)})$ with probability $\geq 1 - \delta$.

By averaging, for at least $m/2$ choices of the index $n \in [m]$ and the restricted distribution $\mathcal{H}'_n$, the error of Π' is bounded by 2δ.

$$\Pr_{(X,Y)\sim\mathcal{H}'_n}\left[\Pi'_{out}(\Pi'(X,Y)) \neq \textsc{Sum-Equal}_k(X^{(n)})\right] \leq 2\delta \tag{B.1}$$

Then we introduce Lemma B.1 that lower bounds $I\left(X_{-k}^{(n)}; M_1, \ldots, M_{k-1}\right) \geq 0.1k \log a$ for protocols with small error. For compactness the proof of Lemma B.1 is deferred to the end of this section.

Lemma B.1. *For every n such that Π' errs with probability $\leq 1/3$ on input $(X,Y) \sim \mathcal{H}'_n$, on at least $a'm$ of the* Sum-Equal *instances, the mutual information between $X^{(n)}$ and $\Pi'(X,Y)$ must be $I\left(X_{-k}^{(n)}; M_1, \ldots, M_{k-1}\right) \geq 0.1k \log a$.*

Using Lemma B.1, it immediately follows that for $\delta \leq 1/6$ the protocol Π' must use $\Omega\left(mk \log a\right)$ bits of communication. Since

$$\begin{aligned}
\mathbf{RCC}_{k,\delta}^{sim}(\Pi') &\geq I\left(X_{-k}; M_1, \ldots, M_{k-1}\right) \\
&= \sum_{i=1}^{m} I\left(X_{-k}^{(i)}; M_1, \ldots, M_{k-1} \mid X_{-k}^{(1)}, \ldots, X_{-k}^{(i-1)}\right) \\
&= \sum_{i=1}^{m} I\left(X_{-k}^{(i)}; M_1, \ldots, M_{k-1}, X_{-k}^{(1)}, \ldots, X_{-k}^{(i-1)}\right) \\
&\geq \sum_{i=1}^{m} I\left(X_{-k}^{(i)}; M_1, \ldots, M_{k-1}\right) \\
&\geq \frac{a'm}{2} \cdot 0.1k \log a = \Omega\left(mk \log a\right)
\end{aligned}$$

since a' is some constant between 0 and 1. □

Proof of Lemma B.1. Suppose by contradiction that $I\left(X_{-k}^{(n)}; M_1, \ldots, M_{k-1}\right) < 0.1k \log a$ and recall that $m \leq \frac{k \log\log T}{20 \log k} \leq \frac{0.1k \log a}{\log(ka)}$ for $\log a = \Omega(\log\log T)$ and sufficiently large T,

$$I\left(X_{-k}^{(n)}; M_1, \ldots, M_{k-1}, X_k, Y^{(n+1)}, \ldots, Y^{(m)}\right) < 0.1k \log a + m \log(ka) < 0.2k \log a$$

Therefore, recalling that $I(A; B, C) = I(A; B \mid C)$ when A is independent from C and that

$X_j^{(n)}$ is independent from $X_1^{(n)}, \dots, X_{j-1}^{(n)}$,

$$\begin{aligned}
&\sum_{j=1}^{k-1} I\left(X_j^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right) \\
\leq &\sum_{j=1}^{k-1} I\left(X_j^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}, X_1^{(n)}, \dots, X_{j-1}^{(n)}\right) \\
= &\sum_{j=1}^{k-1} I\left(X_j^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)} \mid X_1^{(n)}, \dots, X_{j-1}^{(n)}\right) \\
\leq &I(X_{-k}^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}) < 0.2k \log a
\end{aligned}$$

As a result, there is $J \subseteq [k-1]$ and $|J| > k/2$ such that for every $j \in [k-1]$, it holds that

$$I\left(X_j^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right) < -1 + 0.5 \log a,$$

and hence

$$\begin{aligned}
&\mathsf{H}\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] \\
= &\mathsf{H}\left[X_j^{(n)}\right] - I\left(X_j^{(n)}; M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right) \\
> &\log a - (-1 + 0.5 \log a) = 1 + 0.5 \log a
\end{aligned} \tag{B.2}$$

Note that $\mathsf{H}_\infty\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] < 1$ implies the existence of $x \in [a]$ such that $\Pr\left[X_j^{(n)} = x \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] = p_x > \frac{1}{2}$, and hence it follows that

$$\begin{aligned}
\mathsf{H}\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] = &\sum_{i \in [a]} p_i \log \frac{1}{p_i} \\
\leq &p_x \log \frac{1}{p_x} + (1 - p_x) \log \frac{a-1}{1-p_x} \\
< &1 + 0.5 \log(a-1)
\end{aligned} \tag{B.3}$$

Thus, (B.2) ensures that $\mathsf{H}_\infty\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] \geq 1$ for every $j \in J$. In what follows, we prove that if $\mathsf{H}_\infty\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] \geq 1$ for every $j \in J$ and $|J| > k/2$, then the conditional distribution $B'_k := G'_k + M$ and $G'_k := -\sum_{j=1}^{k-1} X_j^{(n)} \,\Big|\, \left\{M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right\}$ have statistical distance $\leq k^{-1/8}$.

Notice that for $j \in J$ and $\mathsf{H}_\infty\left[X_j^{(n)} \mid M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right] \geq 1$, the conditional distribution $G'_j := X_j^{(n)} \,\Big|\, \left\{M_1, \dots, M_{k-1}, X_k, Y^{(n+1)}, \dots, Y^{(m)}\right\}$ is a convex combination of distributions uniform over two values. More specifically, $G'_j = \sum_{v_j} \alpha_{v_j} \cdot G^{[v_j]}$, where $\alpha_{v_j} \in (0, 1)$ and each $G^{[v_j]}$ is a random variable uniform over two values. For $j \notin J$, $G'_j = \sum_{v_j} \alpha_{v_j} \cdot G^{[v_j]}$ where $G^{[v_j]}$ is fixed, i. e., a random variable that equals one value with probability 1. For $v = (v_1, \dots, v_{k-1})$, let $\alpha_v = \prod_{j=1}^{k-1} \alpha_{v_j}$ and $G^{[v]} = \left(G^{[v_1]}, \dots, G^{[v_{k-1}]}, -\sum_{j=1}^{k-1} G^{[v_j]}\right)$, then G' can be decomposed as $G' = \sum_v \alpha_v \cdot G^{[v]}$.

Now for every $j \in J$ and $G^{[v_j]}$ uniform over $\{a_j, b_j\} \subset [a]$, we can assume w.l.o.g., $a_j < b_j$ and write $G^{[v_j]} = a_j + (b_j - a_j)Z_j$ where Z_j is uniform over $\{0, 1\}$. Since $b_j - a_j \in [a]$, among the $> k/2$ indices $j \in J$ for which $G^{[v_j]}$ takes two values, we must have $t \geq |J|/a > k/O(\log k) > \sqrt{k}$ indices J' such that for any $j \in J'$ the value $b_j - a_j$ is the same value M'.

Thus $G^{[v]}$ can be further decomposed into a convex combination of $G^{\{u\}}$ where, among the indices in J, only those in J' are not fixed. Fix any u and denote $G^{\{u\}}$ by G''. Let $S = \sum_{j \in J'} Z_j$ denote the sum of t uniform i.i.d. 0/1 random variables. Then we can write

$$G''_k = b + M'S$$
$$B''_k = b + M'S + M$$

Since $1 \leq M' < a$, M' divides M and hence $M = M'q$ for $q \in \mathbb{Z}$ and $q \leq M \leq k^{1/8}$. Now we can apply q times the shift-invariance of the binomial distribution, which is stated as follows:

Claim B.2 (Claim 39 in [23])**.** *Let S be the sum of t uniform, i.i.d. Boolean random variables. Then S and $S + 1$ have statistical distance $\leq O\left(1/\sqrt{t}\right)$.*

This yields that G''_k and B''_k have statistical distance

$$SD(G''_k, B''_k) = SD(M' \cdot S, M' \cdot (q + S)) \leq q \cdot O\left(1/\sqrt{\sqrt{k}}\right) \leq k^{1/8}/k^{1/4} = k^{-1/8}$$

Recalling that G' is just a convex combination of G'', the statistical distance between G'_k and $B'_k = G'_k + M$ is also bounded by $k^{-1/8}$. However, by definition of G'_k and B'_k we conclude that the referee cannot distinguish the two cases of $X^{(n)} \sim G$ and $X^{(n)} \sim B$ with advantage greater than $k^{-1/8} < 1/6$, which contradicts the condition that Π' has error probability $< 1/3$.

Therefore, $I\left(X_{-k}^{(n)}; M_1, \dots, M_{k-1}\right) \geq 0.1k \log a = \Omega\left(k \log a\right)$. $\square$

References

[1] FARID ABLAYEV: Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoret. Comput. Sci.*, 157(2):139–159, 1996. [doi:10.1016/0304-3975(95)00157-3] 9

[2] YUQING AI, WEI HU, YI LI, AND DAVID P. WOODRUFF: New characterizations in turnstile streams with applications. In *Proc. 31st Comput. Complexity Conf. (CCC'16)*, pp. 20:1–22. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. ACM DL. 8, 29

[3] ZIV BAR-YOSSEF, T. S. JAYRAM, RAVI KUMAR, AND D. SIVAKUMAR: An information statistics approach to data stream and communication complexity. *J. Comput. System Sci.*, 68(4):702–732, 2004. [doi:10.1016/j.jcss.2003.11.006] 3, 4, 12

[4] PAUL BEAME, TONIANN PITASSI, NATHAN SEGERLIND, AND AVI WIGDERSON: A direct sum theorem for corruption and the multiparty NOF communication complexity of set disjointness. In *Proc. 20th IEEE Conf. on Comput. Complexity (CCC'05)*, pp. 52–66. IEEE Comp. Soc., 2005. [doi:10.1109/CCC.2005.1] 7

[5] MARK BRAVERMAN AND ANKIT GARG: Public vs private coin in bounded-round information. In *Proc. 41st Internat. Colloq. on Automata, Languages, and Programming (ICALP'14)*, pp. 502–513. Springer, 2014. [doi:10.1007/978-3-662-43948-7_42] 7

[6] JOSHUA BRODY, HARRY BUHRMAN, MICHAL KOUCKÝ, BRUNO LOFF, FLORIAN SPEELMAN, AND NIKOLAY K. VERESHCHAGIN: Towards a reverse newman's theorem in interactive information complexity. *Algorithmica*, 76(3):749–781, 2016. Preliminary version in CCC'13. [doi:10.1007/s00453-015-0112-9] 7

[7] AMIT CHAKRABARTI, YAOYUN SHI, ANTHONY WIRTH, AND ANDREW CHI-CHIH YAO: Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd FOCS*, pp. 270–278. IEEE Comp. Soc., 2001. [doi:10.1109/SFCS.2001.959901] 10

[8] GRAHAM CORMODE, MAYUR DATAR, PIOTR INDYK, AND S. MUTHUKRISHNAN: Comparing data streams using Hamming norms (how to zero in). *IEEE Trans. Knowledge Data Engr.*, 15(3):529–540, 2003. [doi:10.1109/TKDE.2003.1198388] 5

[9] ANDRÉ GRONEMEIER: Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and Disjointness. In *Proc. 26th Symp. Theoret. Aspects of Comp. Sci. (STACS'09)*, pp. 505–516. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2009. [doi:10.4230/LIPIcs.STACS.2009.1846, arXiv:0902.1609] 3, 12

[10] CHEN-YU HSU, PIOTR INDYK, DINA KATABI, AND ALI VAKILIAN: Learning-based frequency estimation algorithms. In *Int. Conf. Learning Representations (ICLR'19)*. ICLR, 2019. OpenReview. 5

[11] PIOTR INDYK: Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. [doi:10.1145/1147954.1147955] 4

[12] T. S. JAYRAM: Hellinger strikes back: A note on the multi-party information complexity of AND. In *Proc. 13th Internat. Workshop on Randomization and Computation (RANDOM'09)*, pp. 562–573. Springer, 2009. [doi:10.1007/978-3-642-03685-9_42] 3, 12

[13] TANQIU JIANG, YI LI, HONGHAO LIN, YISONG RUAN, AND DAVID P. WOODRUFF: Learning-augmented data stream algorithms. In *Int. Conf. Learning Representations (ICLR'20)*. ICLR, 2020. OpenReview. 5

[14] DANIEL M. KANE, JELANI NELSON, ELY PORAT, AND DAVID P. WOODRUFF: Fast moment estimation in data streams in optimal space. In *Proc. 43rd STOC*, pp. 745–754. ACM Press, 2011. [doi:10.1145/1993636.1993735, arXiv:1007.4191] 4

[15] DANIEL M. KANE, JELANI NELSON, AND DAVID P. WOODRUFF: On the exact space complexity of sketching and streaming small norms. In *Proc. 21st Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'10)*, pp. 1161–1178. SIAM, 2010. ACM DL. 4, 8

[16] DANIEL M. KANE, JELANI NELSON, AND DAVID P. WOODRUFF: An optimal algorithm for the distinct elements problem. In *Proc. 29th Symp. on Principles of Database Systems (PODS'10)*, pp. 41–52. ACM Press, 2010. [doi:10.1145/1807085.1807094] 4, 29

[17] ILAN KREMER, NOAM NISAN, AND DANA RON: On randomized one-round communication complexity. *Comput. Complexity*, 8(1):21–49, 1999. Errata (2001). [doi:10.1007/s000370050018] 9

[18] EYAL KUSHILEVITZ: Communication complexity. *Adv. in Computers*, 44:331–360, 1997. [doi:10.1016/S0065-2458(08)60342-3] 19

[19] THODORIS LYKOURIS AND SERGEI VASSILVITSKII: Competitive caching with machine learned advice. *J. ACM*, 68(4):24:1–25, 2021. Preliminary version in PMLR 2018. [doi:10.1145/3447579, arXiv:1802.05399] 5

[20] MICHAEL MITZENMACHER: Scheduling with predictions and the price of misprediction. In *Proc. 11th Innovations in Theoret. Comp. Sci. Conf. (ITCS'20)*, pp. 14:1–18. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPICS.ITCS.2020.14] 5

[21] MICHAEL MITZENMACHER AND SERGEI VASSILVITSKII: Algorithms with predictions. In TIM ROUGHGARDEN, editor, *Beyond the Worst-Case Analysis of Algorithms*, pp. 646–662. Cambridge Univ. Press, 2020. [doi:10.1017/9781108637435.037] 5

[22] CHRISTOS H. PAPADIMITRIOU AND MICHAEL SIPSER: Communication complexity. *J. Comput. System Sci.*, 28(2):260–269, 1984. Preliminary version in STOC'82. [doi:10.1016/0022-0000(84)90069-2] 9

[23] EMANUELE VIOLA: The communication complexity of addition. *Combinatorica*, 35(6):703–747, 2015. Preliminary version in SODA'13. [doi:10.1007/s00493-014-3078-3] 6, 9, 14, 19, 27, 28, 40

[24] DAVID P. WOODRUFF AND GUANG YANG: Separating k-player from t-player one-way communication, with applications to data streams. In *Proc. 46th Internat. Colloq. on Automata, Languages, and Programming (ICALP'19)*, pp. 1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ICALP.2019.97] 1

[25] DAVID P. WOODRUFF AND QIN ZHANG: Tight bounds for distributed functional monitoring. In *Proc. 44th STOC*, pp. 941–960. ACM Press, 2012. [doi:10.1145/2213977.2214063] 3, 8

AUTHORS

Elbert Du
Ph. D. student
Department of Computer Science
Harvard University
Cambridge, MA, USA
edu1@g.harvard.edu
https://sites.google.com/view/elbert-du/about

Michael Mitzenmacher
Professor
Department of Computer Science
Harvard University
Cambridge, MA, USA
michaelm@eecs.harvard.edu
https://www.eecs.harvard.edu/~michaelm/

David Woodruff
Associate Professor
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
dwoodruf@andrew.cmu.edu
http://www.cs.cmu.edu/~dwoodruf/

Guang Yang
Research Director
Tree-Graph Blockchain Innovation Center of Shanghai and
Tree-Graph Blockchain Innovation Center of Xiang River Hunan
Conflux Foundation
Shanghai, China
guang.research@gmail.com
https://sites.google.com/site/guangyangresearch/home

ABOUT THE AUTHORS

ELBERT DU is currently a Ph. D. student studying Computer Science at Harvard University. He was first introduced to the world of academic mathematics when he was in fifth grade, and he began attending the late Professor Paul Sally Jr.'s Young Scholars' Program at the University of Chicago. Elbert is now interested in studying complexity, differential privacy, and adaptive data analysis. In his spare time, Elbert enjoys reading, solving chess puzzles, and playing video games.

MICHAEL MITZENMACHER is a Professor of Computer Science at Harvard University. He is the coauthor of a well-known textbook on randomized algorithms and probabilistic techniques in computer science with Eli Upfal. He is an ACM and IEEE Fellow.

DAVID WOODRUFF is an associate professor in the Computer Science Department at Carnegie Mellon University. He works on the foundations of data science, specifically in data streams, machine learning, randomized numerical linear algebra, sketching and sparse recovery.

GUANG YANG is currently the research director at Conflux, a startup blockchain project initiated by Fan Long and Andrew Yao. Before joining Conflux, he was an assistant professor at Institute of Computing Technology (ICT), Chinese Academy of Sciences.