An Optimal Lower Bound for Distinct Elements in the Message Passing Model

David P. Woodruff
IBM Research Almaden
dpwoodru@us.ibm.com

Qin Zhang
Indiana University Bloomington
qzhangcs@indiana.edu

Abstract

In the message-passing model of communication, there are k players each with their own private input, who try to compute or approximate a function of their inputs by sending messages to one another over private channels. We consider the setting in which each player holds a subset S_i of elements of a universe of size n, and their goal is to output a $(1+\epsilon)$-approximation to the total number of distinct elements in the union of the sets S_i with constant probability, which can be amplified by independent repetition. This problem has applications in data mining, sensor networks, and network monitoring. We resolve the communication complexity of this problem up to a constant factor, for all settings of n, k and ϵ, by showing a lower bound of $\Omega(k \cdot \min(n, 1/\epsilon^2) + k \log n)$ bits. This improves upon previous results, which either had non-trivial restrictions on the relationships between the values of n, k, and ϵ, or were suboptimal by logarithmic factors, or both.

1 Introduction

Estimating the number F_0 of distinct elements of the union of datasets held by different servers is a fundamental problem in distributed streaming with applications to monitoring network traffic, detecting denial of service attacks, designing query response strategies, OLAP, data warehousing, and data integration.

The problem is well-studied in the streaming model, in which there is a stream $i_1, \ldots, i_m$ of indices $i_j \in [n]$, and $F_0 =: |\{i_1, \ldots, i_m\}|$. Since computing F_0 exactly requires linear space even with randomization [1], one usually allows the algorithm to output an approximation $\tilde{F}_0 \in [F_0, (1+\epsilon)F_0]$ with probability at least $2/3$, where $0 < \epsilon < 1$ is an input parameter. We refer to $\tilde{F}_0$ as a $(1+\epsilon)$-approximation. This probability can then be amplified with independent repetition. The problem was introduced in the theory community by Flajolet and Martin [16], and revived by Alon, Matias, and Szegedy [1]. Since then there has been a large body of work on understanding its space complexity [8, 3, 13, 14, 4, 19].

While these works resolve the complexity of estimating F_0 on a single stream, they are not entirely realistic in practice since data is often shared across multiple servers each with their own stream. This may occur in networks for which routers with limited memory share network traffic information, or in sensor networks in which low-end devices collect data which is then aggregated by a centralized server.

These scenarios motivate the *distributed functional monitoring* model of Cormode, Muthukrishnan, and Yi [9]. In this model there are k players, also known as sites, $P_1, \ldots, P_k$, each holding a set $S_1, \ldots, S_k \subseteq [n]$, respectively. The players want to design a low-communication protocol so that one of the players obtains a $(1+\epsilon)$-approximation to $F_0(S_1, \ldots, S_k) =: |\cup_{i=1}^k S_i|$ with constant probability. The communication is point-to-point, meaning that each pair of players has a private communication channel and messages sent between the two players are not seen by other players. This is also referred to as the message-passing model [12, 24, 25, 27, 6, 18]. One of the main goals is to minimize the total number of bits exchanged between the players, i.e., the communication complexity of accomplishing this task.

In some settings the network, instead of having a connection between all pairs of players, has an arbitrary graph topology and so communication is more restricted. In this paper we focus on the *coordinator model* [12] in which the players communicate with a centralized coordinator by sending and receiving messages on private inputs. Lower bounds in the coordinator model imply lower bounds for the message-passing model in which every node can directly communicate with every other node, as the coordinator model is the most basic topology in which we only assume that every pair of players is connected.

Since there is a streaming algorithm with $O(\min(n, 1/\epsilon^2) + \log n)$ bits of space for obtaining a $(1+\epsilon)$-approximation $\tilde{F}_0$ [19], this implies an $O(k \min(n, 1/\epsilon^2) + k \log n)$ bit communication proto-

col in which the players consecutively run the streaming algorithm on their input and pass the state of the algorithm to the next player. Cormode, Muthukrishnan and Yi [9] showed an $\Omega(k)$ lower bound for this problem via a non-trivial argument, while work of Arackaparambil, Brody and Chakrabarti [2] combined with work of Chakrabarti and Regev [7] showed an $\Omega(1/\epsilon^2)$ lower bound. These bounds were improved to $\Omega(k/(\epsilon^2 \log(\epsilon^2 k)))$ under the assumption that the number k of players exceeds $1/\epsilon^2$ [25].[1]

The starting point of our work is that it is difficult to make assumptions on what n, k, and ϵ ought to be in practice. For instance, in some applications $1/\epsilon^2$ may be quite large if say $\epsilon = .01$ approximation is desired. It may therefore be unreasonable to assume that the number k of players is larger than $1/\epsilon^2 = 10000$, as is done in [25]. Thus, for this regime the best known lower bound remains $\Omega(k + 1/\epsilon^2)$ [2, 9, 7]. On the other hand, if ϵ is a large constant, then the lower bound is $\Omega(k + \log n)$ [9], while one could hope for $\Omega(k \cdot \log n)$. Thus, already for both extremes of settings of ϵ, it is unclear what the right bound should be.

Our Results: We resolve the communication complexity of approximating the number of distinct elements in the message-passing model up to a constant factor. The following is our main theorem.

THEOREM 1.1. *For any setting of $n, k, 1/\epsilon = \Omega(1)$, any private-coin randomized protocol in the message-passing model for estimating F_0 up to a factor of $1 + \epsilon$ with probability at least $2/3$ requires $\Omega(k \cdot \min(n, 1/\epsilon^2) + k \log n)$ bits of communication.*

Given the upper bound above, our lower bound is simultaneously optimal, up to a constant factor, in all parameters n, k, and ϵ.

As a number of problems reduce to estimating F_0, we also obtain tight communication lower bounds for, e.g., estimating rarity [11] or the Klee's measure problem [23] in the message-passing model.

As a corollary of our techniques, we improve the direct sum theorem for randomized public-coin complexity for computing the OR in the message-passing model. Formally, for an arbitrary 2-player function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, in the k-OR-f problem the i-th site has a vector $x_i \in \{0,1\}^n$ and the coordinator has a vector $y \in \{0,1\}^n$. The goal is to compute $\bigvee_{i \in [k]} f(x_i, y)$. Philips, Verbin, and Zhang [24] showed that $R^{1/20,\text{pub}}(k\text{-OR-}f) = \Omega(k \cdot R^{1/3,\text{pub}}(f)/\log^2 k)$, which was improved by the authors to $R^{1/20,\text{pub}}(k\text{-OR-}f) = \Omega(k \cdot R^{1/3,\text{pub}}(f)/\log k)$ [27]. Here we show that $R^{1/20,\text{pub}}(k\text{-OR-}f) = \Omega(k \cdot R^{1/3,\text{pub}}(f))$, which is optimal. This implies optimal bounds for cycle-freeness, connectivity, counting the number of connected components, testing bipartiteness, and testing triangle-freeness; see [27] where plugging in our direct sum theorem improves the previous lower bound for these problems by a $\log k$ factor.

Technique for the $\Omega(k \log n)$ Bound: As noted by Philips, Verbin, and Zhang [24], there are only a few techniques for proving lower bounds in the message-passing model. Perhaps surprisingly, they provably cannot be applied to the following communication problem k-OR-NEQ, in which the coordinator C has a string $y \in \{0,1\}^n$, each player P_i has a string $x_i \in \{0,1\}^n$, and their goal is to decide if there exists an i for which $x_i \neq y$. We prove an $\Omega(k \log n)$ lower bound for the k-OR-NEQ problem. Via a standard reduction, this already improves the $\Omega(k)$ lower bound for F_0 shown in Theorem 5.2 of [10].

The reason this bound cannot be shown using existing techniques is that they either require proving a distributional communication lower bound [24], or they prove a lower bound on the information cost [6]. For the 2-player 2-NEQ problem of deciding whether two strings are not equal, for any distribution μ there is an upper bound of $O(\log 1/\delta)$, where δ is the error probability over inputs drawn from μ [21]. This implies an $O(k \log k)$ communication protocol for any distribution on inputs for k-OR-NEQ. Similarly, under any distribution there is a protocol for 2-NEQ with zero-error and information cost $O(1)$ [5], implying a protocol for k-OR-NEQ with $O(k)$ information cost. These gaps essentially arise because the hard direction of Yao's minimax principle holds only for public-coin protocols. While for 2 players this only amounts to an additive difference of $O(\log n)$ in communication, we show for k players it results in an $O(k \log n)$ difference.

The idea of our proof is to take a k-player protocol and look only at inputs for which k-OR-NEQ evaluates to 0, that is, all players have the same input. We would then like to find a player P_i for which the communication with the coordinator C is typically small, over a random such input, and build a protocol for 2-NEQ between P_i and C, using that the output of k-OR-NEQ is determined by the output of 2-NEQ between P_i and C. However, we have no bound on the communication between C and P_i when their inputs are not equal. What we can do, though, is terminate the protocol if the communication between C and P_i

[1] The conference version of this paper also requires $k > 1/\epsilon^2$ and claims an $\Omega(k/\epsilon^2)$ bound, but in working out the proof in that paper the actual bound obtained is $\Omega(k/(\epsilon^2 \log(\epsilon^2 k)))$.

becomes too large. In this case we either know their inputs are unequal, or their inputs are one of the few inputs causing the communication between C and P_i to be large. For a randomized protocol, though, the induced 2-player protocol must succeed with constant probability on all inputs, not only a large fraction. We use the self-reducibility of 2-NEQ, that we can injectively map instances of 2-NEQ on $(n-1)$-bit strings into inputs on n-bit strings, and remove the inputs causing large communication between C and P_i.

Technique for the $\Omega(k\epsilon^{-2})$ Bound: We recall the approach in [26] for achieving an $\Omega(k\epsilon^{-2}/\log(\epsilon^2 k))$ bound under the assumption that $k > 1/\epsilon^2$, and argue there is an inherent reason this assumption was necessary. The coordinator C was given $y \in \{0,1\}^r$, and each player P_i was given $x_i \in \{0,1\}^r$, where $r = \Theta(\epsilon^{-2})$. Each (x_i, y) pair was chosen from a hard distribution μ for the 2-player disjointness problem 2-DISJ, in which 2-DISJ$(x_i, y) = 1$ iff there exists a j for which $x_{i,j} = y_j = 1$. Notice that the same input y is used in all k 2-player 2-DISJ instances, though the choice of x_i is drawn independently for different i from the marginal distribution conditioned on y. Distribution μ was such that $\mathbf{Pr}[\text{2-DISJ}(x_i, y) = 1] = \epsilon^{-2}/k$. Further, the reduction to F_0 was designed so that the F_0 value was equal to the number of i for which 2-DISJ$(x_i, y) = 1$. The $\Omega(k\epsilon^{-2}/\log(\epsilon^2 k))$ lower bound came from the fact that a correct k-player protocol for F_0 must effectively solve many of the 2-player 2-DISJ instances, each requiring $\Omega(\epsilon^{-2})$ communication.

This approach requires $k > \epsilon^{-2}$ since ϵ^{-2}/k is the probability that 2-DISJ$(x_i, y) = 1$, which must be at most one. Moreover, we need $\Theta(\epsilon^{-2})$ of the 2-player 2-DISJ instances to evaluate to 1, since then by anticoncentration of the binomial distribution (the sum of the output bits of the 2-DISJ instances to the k players) this number will deviate from its expectation by $\Theta(\epsilon^{-1})$, which is why a $(1+\epsilon)$-approximation to F_0 can detect this deviation. To handle $k < \epsilon^{-2}$ we instead consider a 2-player problem 2-SUM in which each player holds inputs to $t = \epsilon^{-2}/k$ independent 2-DISJ instances on sets of size k, in which the output of each instance is 0 or 1 with probability 1/2. The goal is to decide if at least $t/2 + \Omega(\sqrt{t})$ instances are equal to 1, or at most $t/2 - O(\sqrt{t})$ instances are equal to 1. Note that this is a promise problem, and can be viewed as a 2-player problem of solving the majority of t 2-DISJ instances, given that there is a gap. Such a problem may be easier than solving all t of the 2-DISJ instances, which is hard by standard direct sum theorems. We show a reduction to previous work by the authors [26, 25] in the context of estimating F_2 in the blackboard model, showing 2-SUM has randomized communication complexity $\Omega(tk) = \Omega(\epsilon^{-2})$. We note that when $k = \epsilon^{-2}$, this problem reduces to that in [26], while if $k = 1$ it can be seen as the Gap-Thresh(AND) problem, in which given two strings $x, y \in \{0,1\}^r$, is the problem of deciding if the number of coordinates j for which $x_j \wedge y_j = 1$ is at least $r/2 + \sqrt{r}$, or at most $r/2 - \sqrt{r}$ (this is equivalent to the Gap-Hamming problem [7]).

To prove our k-player lower bound, the coordinator C holds an input $(y^1, \ldots, y^t)$ to 2-SUM, where each y^i is an input to a 2-DISJ instance. Each of the players P_i holds an input $(x_i^1, \ldots, x_i^t)$ to 2-SUM chosen from a marginal distribution conditioned on $(y^1, \ldots, y^t)$. While the reduction to F_0 is similar to that in [26, 25], we need a new argument which shows why, from the transcript of the protocol for F_0, one can solve the 2-SUM instance between C and P_i for many i. This requires new arguments since solving a 2-SUM instance only reveals information about the majority of t bits, provided there is a gap, and one needs to argue that if most of these majorities were not learned very well, the sum of them across the k players would not be concentrated well enough to approximate F_0.

Finally, for $k > \epsilon^{-2}$ we improve the $\Omega(k\epsilon^{-2}/\log(\epsilon^2 k))$ lower bound of [26, 25] by a more careful reduction of a k-player problem to a 2-player problem. Usually, one first chooses C's input y to some 2-player problem (e.g., 2-DISJ or 2-SUM), and then one independently samples the inputs $x_i, \ldots, x_k$ to the players from the marginal distribution conditioned on y. Hence, each (x_i, y) is chosen from some distribution μ for the 2-player problem. One argues that typically the transcript for the k-player protocol reveals information about the answer to the 2-player problem for some player for which the communication cost is roughly a $1/k$ fraction of the overall communication. This contradicts a lower bound for the 2-player problem under distribution μ. We instead randomly choose a player P_i and plant an instance (x_i, y) to the 2-player problem under a different distribution μ' between C and P_i. The distribution of (x_j, y) for $j \neq i$ is still μ (so it is important the marginal distributions μ and μ' of the 2-player problem are the same). We argue that the k-player protocol cannot tell we have done this, and so it solves the 2-player problem with low communication under μ'. We can thus choose μ' to obtain a stronger lower bound.

2 Preliminaries

The computational models. We will work in the coordinator model, where we have k players (we call sites, to be consistent with the literature on the coordinator model) $P_1, \ldots, P_k$ and one coordinator.

Each site P_i has an input x_i, and the coordinator has no input. They want to jointly compute some function $f(x_1, \ldots, x_k)$ defined on the union of their inputs. There is a two-way communication channel between each site and the coordinator (which is not seen by other sites), and each site can only talk to the coordinator. The goal is to minimize the communication cost.

We can view the computation in terms of rounds. At each round the coordinator picks a site P_i to communicate, by sending P_i a message based on all the previous messages received from the k sites, and then P_i replies with a message based on its input x_i and all previous messages received from the coordinator.

We note that in the proofs, for reduction purposes we will introduce intermediate problems in which the coordinator will be given an input, but for the original problem, that is, the F_0 problem, the input for the coordinator is always an empty set.

Communication complexity. In the two-party communication complexity model, we have two parties, Alice and Bob. Alice has an input x and Bob an input y, and they want to jointly compute a function $f(x, y)$ by communicating with each other according to a protocol Π. Let $\Pi(x, y)$ be the transcript of the protocol running on the input $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

1. The *deterministic communication complexity (we will abbreviate communication complexity as CC)* of a function f, denoted by $D(f)$, is defined to be $\min_\Pi \max\{|\Pi(x, y)| \mid (x, y) \in \mathcal{X} \times \mathcal{Y}\}$, where $\Pi(x, y) = f(x, y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

2. The *δ-error randomized CC* of f, denoted by $R^\delta(f)$, is defined to be $\min_\Pi \max\{|\Pi(x, y)| \mid (x, y) \in \mathcal{X} \times \mathcal{Y}\}$, where $\mathbf{Pr}[\Pi(x, y) = f(x, y)] \geq 1 - \delta$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Let $R^{\delta,\mathrm{pub}}(f)$ be the public coin δ-error randomized CC where players are allowed to use public coins.

3. The *δ-error μ-distributional CC* of f, denoted by $D^\delta_\mu(f)$, is defined to be $\min_\Pi \max\{|\Pi(x, y)| \mid (x, y) \in \mathcal{X} \times \mathcal{Y}\}$, where $\mathbf{Pr}[\Pi(X, Y) = f(X, Y)] \geq 1 - \delta$ when $(X, Y) \sim \mu$.

4. The *expected δ-error μ-distributional CC* of f, denoted by $\mathsf{ED}^\delta_\mu(f)$, is $\min_\Pi \mathbf{E}_{(X,Y)\sim\mu} |\Pi(X, Y)|$, where $\mathbf{Pr}[\Pi(X, Y) = f(X, Y)] \geq 1 - \delta$ when $(X, Y) \sim \mu$.

These definitions readily generalize from the two-party communication setting to the multi-party setting.

LEMMA 2.1. (YAO'S LEMMA [28]) *In the k-party communication game, for any function f, any input distribution μ, and any $\delta > 0$, it holds that $R^\delta(f) \geq D^\delta_\mu(f)$. Moreover, when $k = 2$, there exists an input distribution τ for which $R^{\delta,pub}(f) = D^\delta_\tau(f)$.*

When $k = 2$, the lemma was proved in [28]. We can easily extend the first part of the lemma to the general k-party communication game, see, e.g., [27]. We have included a proof in Appendix A for completeness.

Conventions. Let $[n] = \{1, 2, \ldots, n\}$. All logarithms are base 2. We often identify sets with their corresponding characteristic vectors when there is no confusion. All bounds are in terms of bits.

3 An $\Omega(k \log n)$ Lower Bound

In this section we prove an $\Omega(k \log n)$ communication lower bound for F_0 in the coordinator model. We first introduce a problem called k-OR-NEQ and analyze its randomized communication complexity, and then prove a lower bound for F_0 by a reduction. At the end, using similar techniques we will also show a general result for k-OR-f for any 2-player problem f.

3.1 The 2-NEQ Problem In the 2-NEQ^n problem, we have Alice and Bob. Alice has an input $x \in \{0, 1\}^n$ and Bob has an input $y \in \{0, 1\}^n$. They output 1 if $x \neq y$, and 0 otherwise. The superscript n on 2-NEQ denotes the size of the input, which we will need to keep track of. The following theorem can be found in [21], Chapter 3.2.

THEOREM 3.1. *$R^{1/3}(2\text{-}NEQ^n) = c_E \cdot \log n$, for an absolute constant c_E.*

3.2 The k-OR-NEQ Problem The k-OR-NEQ problem is defined in the coordinator model. The i-th site has a vector $x_i \in \{0, 1\}^n$, and the coordinator has a vector $y \in \{0, 1\}^n$. The goal is to compute $\bigvee_{i\in[k]} 2\text{-NEQ}^n(x_i, y)$.

THEOREM 3.2. *$R^{1/20}(k\text{-}OR\text{-}NEQ) = \Omega(k \log n)$.*

Proof. First consider NO instances of k-OR-NEQ: such an instance has the form that each of the k sites together with the coordinator has the same input vector u, for some $u \in \{0, 1\}^n$. We identify the NO instance with the vector u.

We prove the theorem by contradiction. Suppose that there is a randomized protocol $\mathcal{P}'$ with communication cost $o(k \log n)$ for k-OR-NEQ. Then by a Markov inequality, there exists a site P_I ($I \in [k]$) for which for at least a $1/2$ fraction of NO instances u, at least a $99/100$ fraction of random strings r have the property that the communication between the coordinator and P_I on u with random string r is at most $\alpha \log n$, for an arbitrary small constant $\alpha > 0$. Since $\mathcal{P}'$ succeeds on

each input u with probability at least $19/20$, by a union bound, we have that for at least a $1/2$ fraction of NO instances u, a $99/100 - 1/20 > 9/10$ fraction of random strings r have the property that the communication between the coordinator and P_I on u with random string r is at most $\alpha \log n$, and $\mathcal{P}'$ outputs the correct answer. Let $S \subseteq \{0,1\}^n$ be this set of NO instances u.

We perform a reduction from 2-NEQ^{n-1} to k-OR-NEQ. Let g be an arbitrary injection between $\{0,1\}^{n-1}$ and S. In 2-NEQ^{n-1}, let $x \in \{0,1\}^{n-1}$ be Alice's input, and $y \in \{0,1\}^{n-1}$ be Bob's input. Alice and Bob construct a protocol $\mathcal{P}$ for 2-NEQ^{n-1} using the protocol $\mathcal{P}'$ for k-OR-NEQ as follows.

1. Alice simulates the site P_I with input $g(x)$.
2. Bob simulates the remaining $k-1$ sites and the coordinator by assigning all of them the input $g(y)$.
3. They run $\mathcal{P}'$ on the resulting input, denoted by z, for k-OR-NEQ.

Note that Bob can simulate any communication between P_i ($i \neq I$) and the coordinator without any actual communication, and the communication between Alice and Bob is equal to the communication between P_I and the coordinator. During the run of $\mathcal{P}'$, if the total communication between the coordinator and P_I exceeds $\alpha \log n$, they *early-terminate* the protocol, meaning they stop the protocol once its communication exceeds $\alpha \log n$ (otherwise we say the protocol *normally-terminates*). They run $\mathcal{P}'$ on z a total of c_R times for a large enough constant c_R, which can be chosen independently of α, using independent private randomness each time. At the end, if more than a $1/10$ fraction of the runs are early-terminated, then they output "$x \neq y$". Otherwise, they output the majority of the outcomes of the runs of $\mathcal{P}'$, without counting those that early-terminate.

Now we show that the resulting protocol $\mathcal{P}$ computes 2-NEQ^{n-1} correctly with probability at least $2/3$.

First, if $x = y$, then $g(x) = g(y)$, that is, the resulting input z for k-OR-NEQ is a NO instance. Notice that by our choice of P_I, with probability $99/100$ over the randomness of $\mathcal{P}'$, the communication between P_I and the coordinator is at most $\alpha \log n$, that is, the protocol will normally-terminate. By a Chernoff bound, for a large enough constant c_R, with probability at least $99/100$, less than a $1/10$ fraction of the c_R runs will early-terminate. Moreover, $\mathcal{P}'$ computes k-OR-NEQ correctly with error probability at most $1/10$ on a run which is normally-terminated (by our choice of site P_I). The process of running the protocol c_R times and then taking the majority of the outcomes, without counting those that early-terminate, will only increase the latter success probability. Therefore, protocol $\mathcal{P}$ computes 2-NEQ^{n-1} correctly with probability at least $1 - 1/100 - 1/10 > 2/3$.

Second, if $x \neq y$, then $g(x) \neq g(y)$, that is, the resulting input z for k-OR-NEQ is not a NO instance. We analyze two cases.

1. If for at least a $4/5$ fraction of random strings of $\mathcal{P}'$, the communication between the coordinator and P_I on z is at most $\alpha \log n$, then for each run, $\mathcal{P}'$ normally-terminates and outputs correctly with probability at least $4/5 - 1/20 > 2/3$. Running the protocol c_R times and taking the majority of the outcomes, without counting those that early-terminate, only increases the success probability.

2. If for at least a $1/5$ fraction of random strings of $\mathcal{P}'$, the communication between the coordinator and P_I on z exceeds $\alpha \log n$, then for a large enough number c_R of repetitions of $\mathcal{P}'$, where c_R is a constant chosen independently of α, we have that by a Chernoff bound, with probability at least $99/100 > 2/3$, at least a $1/10$ fraction of the runs will early-terminate. Alice and Bob can detect such an event and declare that $x \neq y$.

Finally, since α is unconstrained, by choosing $\alpha = c_E/(2c_R)$, the communication cost of $\mathcal{P}$ for 2-NEQ^{n-1} is at most $c_R \cdot \alpha \cdot \log n = c_E \log n/2 < c_E \log(n-1) = R^{1/3}(\text{2-NEQ}^{n-1})$ (Theorem 3.1). We have therefore reached a contradiction.

3.3 A Lower Bound of F_0 There is a simple reduction from k-OR-NEQ to approximating F_0 up to a constant factor (a $(1+\epsilon)$-approximation with $1+\epsilon < 3/2$ suffices). By results in coding theory (c.f. [1], Section 3.3), there exists a family $\mathcal{G}$ consisting of $t = 2^n$ subsets of $[n/\iota]$ (for a constant ι), each of cardinality $n/(4\iota)$, such that for any two $a, b \in \mathcal{G}$, it holds that $|a \cap b| \leq n/(8\iota)$.

Now given an input $(x_1, \ldots, x_k, y) \in \{0,1\}^n \times \cdots \times \{0,1\}^n$ for k-OR-NEQ, we construct an input for F_0. Let h be an arbitrary bijection between $\{0,1\}^n$ and elements in $\mathcal{G}$. The k sites and the coordinator run a $(1+\epsilon)$-approximation protocol for F_0, for $1 + \epsilon < 3/2$, on input $(h(x_1), \ldots, h(x_k))$. Note that if k-OR-NEQ$(x_1, \ldots, x_k, y) = 1$, then we have $F_0(h(x_1), \ldots, h(x_k)) \geq n/(4\iota) + n/(8\iota)$; and if k-OR-NEQ$(x_1, \ldots, x_k, y) = 0$, then we have $F_0(h(x_1), \ldots, h(x_k)) = n/(4\iota)$. Therefore we can use a $(1+\epsilon)$-approximation to F_0 to solve k-OR-NEQ. The following theorem is a direct consequence of this reduction and Theorem 3.2.

THEOREM 3.3. *For $1+\epsilon < 3/2$, it holds that $R^{1/20}((1+\epsilon)$-approximate $F_0) = \Omega(k \log n)$.*

3.4 The k-OR-f Problem In this section we generalize Theorem 3.2 to k-OR-f for an arbitrary 2-player function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. The k-OR-f problem is defined in the coordinator model. The i-th site has a vector $x_i \in \{0,1\}^n$, and the coordinator has a vector $y \in \{0,1\}^n$. The goal is to compute $\bigvee_{i\in[k]} f(x_i, y)$. We have the following theorem.

THEOREM 3.4. $R^{1/20,pub}(k\text{-}OR\text{-}f) = \Omega(k \cdot R^{1/3,pub}(f))$.

REMARK 1. *Note that in Theorem 3.4, we use the public coin communication complexity, thus this theorem cannot be directly applied to $f = 2$-NEQ for proving an $\Omega(k \log n)$ lower bound, since $R^{1/3,pub}(2\text{-}NEQ) = O(1)$ (see, e.g., [21], Chapter 3.2). But this theorem is sufficient for proving an $\Omega(nk)$ lower bound for k-OR-DISJ ($f = 2$-DISJ, see its definition in Section 4.1), which has applications to many basic statistic and graph problems [27], e.g., ℓ_∞, graph connectivity, bipartiteness, etc.*

Proof. (of Theorem 3.4) Let τ be an input distribution for f such that $D_\tau^{1/3}(f) = R^{1/3,\text{pub}}(f)$. By Yao's Lemma (Lemma 2.1) such a distribution always exists. Let τ_1, τ_0 be the induced distributions of τ on YES instances and NO instances, respectively. We can write $\tau = \lambda\tau_0 + (1-\lambda)\tau_1$ for some $0 \le \lambda \le 1$.

We prove by contradiction. Assume that $R^{1/20,\text{pub}}(k\text{-OR-}f) = o(k \cdot R^{1/3,\text{pub}}(f)) = o(k \cdot D_\tau^{1/3}(f))$, and let $\mathcal{P}'$ be such a protocol realizing $R^{1/20,\text{pub}}(k\text{-OR-}f)$. Note that $\mathcal{P}'$ succeeds on every input with probability at least 19/20, over its randomness. We will show that we can get a deterministic protocol $\mathcal{P}$ for f on input distribution τ with distributional communication cost less than $D_\tau^{1/3}(f)$, resulting in a contradiction.

First, note that if the input $(X_1, \ldots, X_k, Y)$ for k-OR-f is distributed so that $(X_i, Y) \sim \tau_0$ for all $i \in [k]$ (that is, a distribution on the NO instances of k-OR-f, denoted by $\tau_0^{(k)}$), then by a Markov inequality, there must be a site P_i ($i \in [k]$) for which with probability 99/100 over the distribution $\tau_0^{(k)}$ and the randomness of $\mathcal{P}'$, the communication between P_i and the coordinator is at most $\alpha \cdot D_\tau^{1/3}(f)$ (for some arbitrarily small constant $\alpha > 0$). Let P_I denote such a site.

The reduction consists of two steps, during which we allow Alice and Bob to use randomness, which we will fix at the end.

Input reduction. Given an input $(A, B) \sim \tau$, Alice and Bob construct an input $(X_1, \ldots, X_k, Y)$ for k-OR-f.

1. Alice assigns the site P_I the input $X_I = A$.
2. Bob assigns inputs for the remaining $k-1$ sites and the coordinator: He assigns the coordinator with an input $Y = B$, and then independently samples $X_1, \ldots, X_{I-1}, X_{I+1}, \ldots, X_k$ from the marginal distribution $\tau_0|Y$, and assigns them to the remaining $k-1$ sites.

Note that we have k-OR-$f(X_1, \ldots, X_k, Y) = f(A, B)$.

Constructing a protocol $\mathcal{P}$ for f using a protocol $\mathcal{P}'$ for k-OR-f. Alice and Bob run $\mathcal{P}'$ on $(X_1, \ldots, X_k, Y)$ for k-OR-f a total of c_R times for a large enough constant c_R using independent private randomness each time, where c_R is chosen independently of α. During each run of $\mathcal{P}'$, Alice simulates P_I, and Bob simulates the remaining $k-1$ sites and the coordinator. Note that Bob can simulate any communication between P_i ($i \ne I$) and the coordinator without any actual communication, and the communication between Alice and Bob is equal to the communication between P_I and the coordinator. In each of the c_R runs, if the total communication between Alice and Bob exceeds $\alpha D_\tau^{1/3}(f)$, then they *early-terminate* that run (otherwise we again say the run *normally-terminates*). At the end, if more than a 1/10 fraction of runs early-terminate, they output YES, otherwise they output the majority of the outcomes of the runs (without counting those that early-terminate).

Now we show that $\mathcal{P}$ succeeds on input $(A, B) \sim \tau$ for f with error probability at most 1/12.

First, it succeeds on the distribution τ_0 (on NO instances) with error probability at most 1/12. This is because in each run, $\mathcal{P}'$ normally-terminates with probability 99/100 over the input distribution $\tau_0^{(k)}$ and the randomness of the protocol, by our choice of P_I. Moreover, since $\mathcal{P}'$ is correct with error probability at most 1/20 on each input, by a union bound, with error probability at most $1/20 + 1/100 < 1/12$ over the input distribution $\tau_0^{(k)}$ and the randomness of $\mathcal{P}'$, $\mathcal{P}'$ normally-terminates and outputs the correct answer. Running the protocol c_R times and then taking the majority of the outcomes ,without counting those that early-terminate, will not decrease the success probability.

We next consider the distribution τ_1 (on YES instances). First, $\mathcal{P}'$ succeeds on every input with probability at least 19/20 over its randomness, and therefore this holds for every input created for $\mathcal{P}'$ using (A, B) in the support of τ_1 to assign to P_I and the coordinator. The only case we have to take care of is the early-termination of a run. Fix an input created for $\mathcal{P}'$ using (A, B) in the support of τ_1. Suppose that $\mathcal{P}'$ early-terminates with probability at most 1/5 over the

randomness of the protocol. Then by a union bound, with probability $(1-1/20)-1/5 > 2/3$, $\mathcal{P}'$ outputs a correct answer on each run. We can run the protocol c_R times (for a large enough constant c_R) and then take the majority of the outcomes, without counting those that early-terminate, to reduce the error probability to $1/12$. Otherwise, if $\mathcal{P}'$ early-terminates with probability more than $1/5$ over the randomness of the protocol, then after running $\mathcal{P}'$ a total of c_R times, for a sufficiently large constant c_R, by a Chernoff bound, with error probability at most $1/100 < 1/12$, at least a $1/10$ fraction of the runs will early-terminate. Alice and Bob can detect such an event and output YES.

Since τ is a linear combination of τ_0 and τ_1, $\mathcal{P}$ succeeds on input $(A,B) \sim \tau$ with error probability at most $1/12$. The communication cost of the protocol $\mathcal{P}$ is at most $c_R \cdot \alpha \cdot D_\tau^{1/3}(f) < D_\tau^{1/3}(f)/4$ (by choosing $\alpha = 1/(8c_R)$, which we can do since c_R is chosen independently of α). Finally, we use two Markov inequalities to fix all the randomness used in the reduction, such that the resulting deterministic protocol $\mathcal{P}$ succeeds with error probability $4 \cdot 1/12 = 1/3$ on input distribution ν, and its communication cost is less than $4 \cdot D_\tau^{1/3}(f)/4 = D_\tau^{1/3}(f)$. We have reached a contradiction.

4 An $\Omega(k \cdot \min\{n, 1/\epsilon^2\})$ Lower Bound

In this section we prove an $\Omega(k/\epsilon^2)$ lower bound for F_0. We will focus on $\Omega(1) \le k \le O(1/\epsilon^2)$, since an $\Omega(k/(\epsilon^2 \log(\epsilon^2 k)))$ lower bound for $k \ge \Omega(1/\epsilon^2)$ was already shown in [25]. In Section 4.5 we note that in fact, we can also achieve $\Omega(k/\epsilon^2)$ for the case when $k \ge \Omega(1/\epsilon^2)$, by a better embedding argument.

For the case when $\Omega(1) \le k \le O(1/\epsilon^2)$, we start with a problem called 2-SUM, whose expected distributional communication complexity can be obtained by a reduction from another problem called 2-BTX (stands for k-BLOCK-THRESH-XOR). Next, we use a reduction from 2-SUM to prove a distributional communication complexity lower bound for a problem called k-SUM. Finally, we prove a lower bound for F_0 by a reduction from k-SUM.

We will set the universe size in this proof to be $n = \Theta(1/\epsilon^2)$, and prove an $\Omega(k/\epsilon^2)$ lower bound. If $n = \omega(1/\epsilon^2)$, then we can simply use a subset of the universe of size $\Theta(1/\epsilon^2)$. If $n = o(1/\epsilon^2)$, then we can still use the same proof with an approximation parameter $\epsilon' = 1/\sqrt{n} > \epsilon$ (that is, we can prove the lower bound for an even larger error), and obtain an $\Omega(k/(\epsilon')^2) = \Omega(kn)$ lower bound.

We fix $\beta \triangleq 1/4$ in this section.

4.1 The 2-DISJ Problem

In the 2-DISJ problem, Alice has an input $X = (X_1, \ldots, X_L) \in \{0,1\}^L$, and Bob has an input $Y = (Y_1, \ldots, Y_L) \in \{0,1\}^L$. The output 2-DISJ$(X,Y) = 0$ if $\sum_{\ell \in [L]} X_\ell \wedge Y_\ell = 0$, and 2-DISJ$(X,Y) = 1$ if $\sum_{\ell \in [L]} X_\ell \wedge Y_\ell \ge 1$. We define an input distribution μ for 2-DISJ:

μ : For each $\ell \in [L]$, choose $D_\ell \in \{0,1\}$ uniformly at random. If $D_\ell = 0$, then set $X_\ell = 0$, and choose $Y_\ell \in \{0,1\}$ uniformly at random. Otherwise if $D_\ell = 1$, then set $Y_\ell = 0$, and choose $X_\ell \in \{0,1\}$ uniformly at random. The choices for different $\ell \in [L]$ are independent. Finally, pick a *special coordinate* $M \in [L]$ uniformly at random, and reset $(X_M, Y_M) \in \{0,1\}^2$ uniformly at random.

Note that when $(X,Y) \sim \mu$, we have $\mathbf{Pr}[\text{2-DISJ}(X,Y) = 1] = 1/4 = \beta$.

4.2 The 2-SUM Problem

In the 2-SUM problem, Alice and Bob have inputs $X = (X^1, \ldots, X^t)$ and $Y = (Y^1, \ldots, Y^t)$, respectively. They want to approximate 2-SUM$(X,Y) = \sum_{j \in [t]} \text{2-DISJ}(X^j, Y^j)$ up to an additive error of $\sqrt{\beta t}$. We define an input distribution ν for 2-SUM.

ν: For each $j \in [t]$, we independently pick $(X^j, Y^j) \sim \mu$.

In [25], Section 4.4, a similar problem called k-BTX problem was considered. When $k = 2$, 2-BTX can be stated as follows: There are two parties Alice and Bob. Alice has an input $X = (X^1, \ldots, X^t)$ and Bob has an input $Y = (Y^1, \ldots, Y^t)$. Independently for each j, $(X^j, Y^j) \sim \mu$. Thus $(X,Y) \sim \nu$. Let M^j be the index of the special coordinate when sampling (X^j, Y^j) from μ. The problem 2-BTX is:

$$\text{2-BTX}(X,Y) = \begin{cases} 1, & \text{if } \left|\sum_{j \in [t]} X^j_{M^j} \oplus Y^j_{M^j} - \frac{t}{2}\right| \ge 4\sqrt{\beta t}, \\ 0, & \text{if } \left|\sum_{j \in [t]} X^j_{M^j} \oplus Y^j_{M^j} - \frac{t}{2}\right| \le 2\sqrt{\beta t}, \\ *, & \text{otherwise,} \end{cases}$$

where $*$ means that the output can be arbitrary.

The following theorem for 2-BTX is an easy consequence of Corollary 1 in [25]. [2]

THEOREM 4.1. $\mathsf{ED}_\nu^{\delta_1}(2\text{-}BTX) = \Omega(tL)$, *for a sufficiently small constant* δ_1.

[2] Corollary 1 in [25] states that any randomized protocol that computes 2-BTX on input distribution ν with error probability δ for a sufficiently small constant δ has communication complexity $\Omega(tL)$. We can replace the randomized protocol with any deterministic protocol. We can also terminate the deterministic protocol when the communication exceeds $C \cdot \mathsf{ED}_\nu^\delta(\text{2-BTX})$ for an arbitrarily large constant C, which only introduces an additional (arbitrarily small) error of $1/C$. Thus if $\mathsf{ED}_\nu^\delta(\text{2-BTX}) = o(tL)$, then we also have $D_\nu^{\delta+1/C}(\text{2-BTX}) = o(tL)$.

The following theorem can be shown by a simple reduction from 2-BTX to 2-SUM.

THEOREM 4.2. $\mathsf{ED}_{\nu}^{\delta_2}(2\text{-}SUM) = \Omega(tL)$, *for a sufficiently small constant* δ_2.

Proof. To show the desired communication cost, we just need to show that if we have a protocol $\mathcal{P}$ for 2-SUM on input distribution ν with error probability $\delta_2 = \delta_1/2$, then by running $\mathcal{P}$ twice we can solve 2-BTX on input distribution ν with error probability δ_1.

To see that this can be done, Alice and Bob first run protocol $\mathcal{P}$ on (X, Y), obtaining a value W_1, which approximates $\sum_{j\in[t]} \text{AND}(X^j_{M^j}, Y^j_{M^j})$ up to an additive error $\sqrt{\beta t}$. Next, Alice and Bob flip all bits of X and Y, obtaining $\bar{X}, \bar{Y}$, respectively, and then they run $\mathcal{P}$ on $(\bar{X}, \bar{Y})$, obtaining a value W_2, which approximates $\sum_{j\in[t]} \text{AND}(\bar{X}^j_{M^j}, \bar{Y}^j_{M^j})$ up to an additive error $\sqrt{\beta t}$. Finally, $t - (W_1 + W_2)$ approximates $\sum_{j\in[t]} X^j_{M^j} \oplus Y^j_{M^j}$ up to an additive error $\sqrt{\beta t}+\sqrt{\beta t} = 2\sqrt{\beta t}$, and therefore solves 2-BTX.

4.3 The k-SUM Problem The k-SUM problem is defined in the coordinator model. Each site P_i ($i \in [k]$) has an input X_i, and the coordinator has an input Y. The $k+1$ parties want to approximate $\sum_{i\in[k]} \text{2-SUM}(X_i, Y)$ up to an additive error $\sqrt{\beta k t}$. We define an input distribution ψ for k-SUM.

ψ: We first choose $(X_1, Y) \sim \nu$, and then independently choose $X_2, \ldots, X_k \sim \nu|Y$ (the distribution of A conditioned on $B = Y$, when $(A, B) \sim \nu$).

We show that any protocol that computes k-SUM well must effectively compute many of the 2-SUM(X_i, Y) values well, and then prove a lower bound for k-SUM using a reduction from 2-SUM. If not otherwise specified, probabilities, expectations and variances below are taken over the input distribution ψ to k-SUM.

Let $X_i = (X_i^1, \ldots, X_i^t)$ and $Y = (Y^1, \ldots, Y^t)$. By definition of distribution ν, we have $(X_i^j, Y^j) \sim \mu$ for each pair (i, j), and these are independent for different values of j. Let $Z_i^j = \text{2-DISJ}(X_i^j, Y^j) \sim$ Bernoulli(β), and let $Z_i = \sum_{j\in[t]} Z_i^j$. We have the following observation, based on the rectangle-property of communication, whose proof can be found in [25]: [3]

OBSERVATION 1. *Conditioned on* Π, $Z_1, \ldots, Z_k$ *are independent.*

The following definition characterizes the usefulness of a protocol transcript $\Pi = \pi$.

[3] In [25], Z_i's are bits, but the proof also works for general random variables Z_i, as long as they are independent.

DEFINITION 1. *We say a transcript* π *is* weak *if* $\sum_{i\in[k]} \mathbf{Var}(Z_i \mid \Pi = \pi) \geq \beta k t/c_0$ *where* $c_0 = 200/\delta_2^2$, *and* strong *otherwise.*

LEMMA 4.1. *Let* Π *be the transcript of any deterministic protocol that computes* k*-SUM on input distribution* ψ *with error probability* δ_3 *for a sufficiently small constant* δ_3. *Then* $\mathbf{Pr}[\Pi$ *is strong*$] \geq 1 - \delta_2/10$.

The proof in the high level is similar to Lemma 3 of [25]. The differences are (1) $Z_1, \ldots, Z_k$ are integers rather than bits, and (2) we also have a different setting of parameters. In particular, we cannot use the anti-concentration result (Fact 1 in [25]) directly. We instead use the Berry-Esseen theorem together with some additional conditions.

Let $\kappa = \sqrt{c_\kappa \log k}$ for a sufficiently large constant c_κ. Let ξ_i be the indicator variable of the event that $|Z_i - \beta t| \leq \kappa\sqrt{t}$. Let $\xi = \xi_1 \wedge \xi_2 \wedge \ldots \wedge \xi_k$. We have the following simple claim.

CLAIM 1. $\mathbf{Pr}[\xi = 1] \geq 0.99$.

Proof. For each $i \in [k]$, note that $Z_i = \sum_{j\in[t]} Z_i^j$ and $Z_i^j \sim$ Bernoulli(β). We apply the Chernoff-Hoeffding bound for each $i \in [k]$ and get $\mathbf{Pr}[\xi_i = 1] \geq 1 - e^{-\kappa^2/3} \geq 1 - e^{-c_\kappa/3\cdot\log k}$. The claim follows by a union bound on ξ_i ($i \in [k]$), by choosing a large enough constant c_κ.

We need two more definitions and an auxiliary lemma.

DEFINITION 2. *We say a transcript* π *is* rare$^+$ *if* $\Pi = \pi$, $\mathbf{E}[Z_i \mid \Pi = \pi] \geq 4\beta t$ *and* rare$^-$ *if* $\mathbf{E}[Z_i \mid \Pi = \pi] \leq \beta t/4$. *In both cases we say* π *is* rare. *Otherwise we say it is* normal.

DEFINITION 3. *We say* $Z = \{Z_1, Z_2, \ldots, Z_k\}$ *is a* joker$^+$ *if* $\sum_{i\in[k]} Z_i \geq 2\beta t k$, *and a* joker$^-$ *if* $\sum_{i\in[k]} Z_i \leq \beta t k/2$. *In both cases we say* Z *is a* joker.

The following lemma is similar to Lemma 2 in [25]. We include a proof for completeness.

LEMMA 4.2. *Let* Π *be the transcript of any deterministic protocol that computes* k*-SUM on input distribution* ψ *with error probability* δ_3 *for a sufficiently small constant* δ_3, *then* $\mathbf{Pr}[\Pi$ *is normal*$] \geq 1 - \delta_2/20$.

Proof. First, $\sum_{i\in[k]} Z_i = \sum_{i\in[k]} \sum_{j\in[t]} Z_i^j$, and $Z_i^j \sim$ Bernoulli(β) for all $i \in [k], j \in [t]$. Applying a Chernoff bound on random variables Z_i^j's, we have

$$\mathbf{Pr}[Z \text{ is a joker}^+] = \mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq 2\beta t k\right] \leq e^{-\Omega(tk)}.$$

We next use Observation 1, and apply another Chernoff bound on $Z_i \mid \Pi = \pi$. Note that $(Z_i \mid \Pi = \pi) \in [0, t]$ for all $i \in [k]$. Let $\bar{Z} = \mathbf{E}\left[\sum_{i\in[k]} Z_i \;\middle|\; \Pi = \pi\right]$.

$$\begin{aligned}
& \mathbf{Pr}[Z \text{ is a joker}^+ \mid \Pi \text{ is rare}^+] \\
\geq & \sum_\pi \mathbf{Pr}\left[\Pi = \pi \mid \Pi \text{ is rare}^+\right] \\
& \times \mathbf{Pr}\left[Z \text{ is a joker}^+ \mid \Pi = \pi, \Pi \text{ is rare}^+\right] \\
= & \sum_\pi \mathbf{Pr}\left[\Pi = \pi \mid \Pi \text{ is rare}^+\right] \\
& \times \mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq 2\beta tk \;\middle|\; \bar{Z} \geq 4\beta tk, \Pi = \pi\right] \\
\geq & \sum_\pi \mathbf{Pr}\left[\Pi = \pi \mid \Pi \text{ is rare}^+\right]\left(1 - e^{-\Omega(k)}\right) \\
= & \left(1 - e^{-\Omega(k)}\right).
\end{aligned}$$

Finally by Bayes' theorem, we have

$$\begin{aligned}
& \mathbf{Pr}[\Pi \text{ is rare}^+] \\
= & \frac{\mathbf{Pr}[Z \text{ is a joker}^+] \cdot \mathbf{Pr}[\Pi \text{ is rare}^+ \mid Z \text{ is a joker}^+]}{\mathbf{Pr}[Z \text{ is a joker}^+ \mid \Pi \text{ is rare}^+]} \\
\leq & \frac{e^{-\Omega(tk)}}{1 - e^{-\Omega(k)}} \leq e^{-\Omega(tk)}.
\end{aligned}$$

Similarly, we can also show that $\mathbf{Pr}[\Pi \text{ is rare}^-] \leq e^{-\Omega(tk)}$. Therefore $\mathbf{Pr}[\Pi \text{ is rare}] \leq e^{-\Omega(tk)} \leq \delta_2/20$.

Now we prove Lemma 4.1.

Proof. Let $\bar{Z} = \mathbf{E}\left[\sum_{i\in[k]} Z_i \;\middle|\; \Pi\right]$. We first show there exists a constant $\delta_\ell = \delta_\ell(c_\ell)$ such that

(4.1)
$$\mathbf{Pr}\left[\sum_{i\in[k]} Z_i \leq \bar{Z} + 2\sqrt{\beta tk} \;\middle|\; \Pi \text{ is normal \& weak}\right] \geq \delta_\ell,$$

(4.2)
$$\mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq \bar{Z} + 4\sqrt{\beta tk} \;\middle|\; \Pi \text{ is normal \& weak}\right] \geq \delta_\ell.$$

The first inequality is a simple application of Chernoff-Hoeffding. Recall $\mathbf{E}\left[\sum_{i\in[k]} Z_i \;\middle|\; \Pi \text{ is normal}\right] \leq 4\beta tk$. Using Observation 1, we have

$$\begin{aligned}
& \mathbf{Pr}\left[\sum_{i\in[k]} Z_i \leq \bar{Z} + 2\sqrt{\beta tk} \;\middle|\; \Pi \text{ is normal}\right] \\
\geq & 1 - \mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq \bar{Z} + 2\sqrt{\beta tk} \;\middle|\; \Pi \text{ is normal}\right] \\
\geq & 1 - e^{-\frac{8\sqrt{\beta tk}^2}{\bar{Z}}} \geq 1 - e^{-2} \\
\geq & \delta_\ell. \quad \text{(for a sufficiently small constant } \delta_\ell)
\end{aligned}$$

Now we prove the second inequality. We will need the following version of the Berry-Esseen theorem.

THEOREM 4.3. (BERRY-ESSEEN) *Let $X_1, X_2, \ldots, X_k$ be independent random variables with $\mathbf{E}[X_i] = 0, \mathbf{E}[X_i^2] = \sigma_i^2$, and $\mathbf{E}[|X_i|^3] = \rho_i < \infty$. Also, let $S_k = \sum_{i\in[k]} X_i / \sqrt{\sum_i \sigma_i^2}$ be the normalized sum. Denote F_k the cumulative distribution function of S_k, and Φ the cumulative distribution function of the standard normal distribution. Then there exists an absolute constant c such that*

$$\sup_{x\in\mathbb{R}} |F_k(x) - \Phi(x)| \leq c \cdot \sum_{i\in[k]} \rho_i \Big/ \left(\sum_{i\in[k]} \sigma_i^2\right)^{3/2}.$$

In our application, we define $X_i = (Z_i|\Pi, \xi = 1) - \mathbf{E}[Z_i|\Pi, \xi = 1]$. Thus $\mathbf{E}[X_i] = 0$, and for all $i \in [k]$,

$$\rho_i \leq |X_i|^3 \leq (2\kappa\sqrt{t})^3, \tag{4.3}$$

by the definition of ξ.

Let $\sigma^2 = \sum_{i\in[k]} \sigma_i^2 = \mathbf{Var}[\sum_{i\in[k]} X_i] = \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi, \xi = 1\right]$. For a weak Π, we have

$$\begin{aligned}
& \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}\right] \\
= & \sum_{i\in[k]} \mathbf{Var}\left[Z_i \mid \Pi \text{ is weak}\right] \quad \text{(by Observation 1)} \\
\geq & \beta tk/c_0. \quad \text{(by definition of a weak } \Pi)
\end{aligned}$$

We next bound $\mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}, \xi = 1\right]$ using $\mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}\right]$. We first define a few events. Let η be the minimum value for which $Z_i \in [\beta t - \eta \cdot \sqrt{t}, \beta t + \eta \cdot \sqrt{t}]$ for all $i \in [k]$. By Chernoff-Hoeffding and the union bound, this holds with probability at least $1 - k \cdot e^{-\eta^2/3}$. Define F_y to be the event that $\eta \in [y, 2y]$. We have

$$
\begin{aligned}
&(4.4)\quad \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}\right] \\
&= \mathbf{E}\left[\left(\sum_{i\in[k]} Z_i - \bar{Z}\right)^2 \mid \Pi \text{ is weak}, \xi = 1\right] \cdot \mathbf{Pr}[\xi = 1] \\
&\quad + \sum_{j=\log\kappa}^{\infty} \mathbf{E}\left[\left(\sum_{i\in[k]} Z_i - \bar{Z}\right)^2 \mid \Pi \text{ is weak}, \xi = 0, F_{2^j}\right] \\
&(4.5)\quad \cdot \mathbf{Pr}[F_{2^j}] \\
&\leq \mathbf{E}\left[\left(\sum_{i\in[k]} Z_i - \bar{Z}\right)^2 \mid \Pi \text{ is weak}, \xi = 1\right] \\
&(4.6)\quad + \sum_{j=\log\kappa}^{\infty} (k \cdot 16 \cdot 4^j \cdot t) \cdot (k \cdot e^{-4^j/3}) \\
&\leq \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}, \xi = 1\right] \\
&(4.7)\quad + t/\mathrm{poly}(k),
\end{aligned}
$$

where

1. (4.4) → (4.5) is due to the law of total expectation.

2. (4.5) → (4.6) since (1) F_{2^j} holds with probability at most $k \cdot e^{-4^j/3}$ by a Chernoff bound and a union bound over all $i \in [k]$; (2) conditioned on F_{2^j}, $\sum_{i\in[k]} Z_i - \bar{Z}$ is bounded by the maximum deviation of $\sum_{i\in[k]} Z_i$, which is at most $2 \cdot 2 \cdot 2^j \cdot \sqrt{t}$, where one of the 2's comes from the definition of F_y, that is, $\eta \leq 2y$.

3. (4.6) → (4.7) holds if we choose constant c_κ (recall that $\kappa = \sqrt{c_\kappa \log k}$) large enough. This is because

$$
\begin{aligned}
&\sum_{j=\log\kappa}^{\infty} (k \cdot 16 \cdot 4^j \cdot t) \cdot (k \cdot e^{-4^j/3}) \\
\leq\ & \sum_{j=\log\kappa}^{\infty} t \cdot 16k^2 \cdot e^{-4^j/4} \\
\leq\ & 2 \cdot t \cdot 16k^2 \cdot e^{-2^{\log\kappa^2}/4} \\
\leq\ & 2 \cdot t \cdot 16k^2 \cdot e^{-c_\kappa/4 \cdot \log k} \leq t/\mathrm{poly}(k).
\end{aligned}
$$

We therefore have

$$
\begin{aligned}
\sigma^2 &= \sum_{i\in[k]} \sigma_i^2 \\
&= \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}, \xi = 1\right] \\
&\geq \mathbf{Var}\left[\sum_{i\in[k]} Z_i \mid \Pi \text{ is weak}\right] - t/\mathrm{poly}(k) \\
&(4.8)\quad \geq \beta tk/(2c_0).
\end{aligned}
$$

Conditioned on $\xi = 1$ and weak Π, by Theorem 4.3, using (4.3) and (4.8) we get

$$
\begin{aligned}
\sup_{x\in\mathbb{R}} |F_k(x) - \Phi(x)| &\leq c \cdot \sum_{i\in[k]} \rho_i \Big/ \left(\sum_{i\in[k]} \sigma_i^2\right)^{3/2} \\
&\leq c \cdot \frac{\sum_{i\in[k]} (2\kappa\sqrt{t})^3}{(\beta tk/(2c_0))^{3/2}} \\
&= 8c/(\beta/(2c_0))^{3/2} \cdot \kappa^3/\sqrt{k} \\
&(4.9)\quad \leq c_B,
\end{aligned}
$$

for an arbitrarily small constant c_B, given $k \geq c'_B\kappa^6 = c'_B \cdot c_\kappa^3 \log^3 k$ for a large enough constant c'_B.

Using (4.9), and the fact that for a standard normal random variable x, $\mathbf{Pr}[x \geq c_\sigma] \geq c_N(c_\sigma)$ for any constant c_σ, where $c_N(c_\sigma)$ is a constant depending on c_σ, we have

$$
\mathbf{Pr}\left[\sum_{i\in[k]} X_i \geq 4\sqrt{c_0} \cdot \sigma \mid \Pi \text{ is weak}\right] \geq \delta'_\ell,
$$

for a sufficiently small constant δ'_ℓ. Consequently,

$$
\begin{aligned}
&\mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq \bar{Z} + 4\sqrt{\beta tk} \mid \Pi \text{ is weak}\right] \\
\geq\ & \mathbf{Pr}\left[\sum_{i\in[k]} Z_i \geq \bar{Z} + 4\sqrt{\beta tk} \mid \Pi \text{ is weak}, \xi = 1\right] \\
& \times \mathbf{Pr}[\xi = 1] \\
\geq\ & 0.99\delta'_\ell \geq \delta_\ell,
\end{aligned}
$$

for a sufficiently small constant δ_ℓ. In the last equality we have used Claim 1.

By (4.1) and (4.2), it is easy to see that given that Π is normal, it cannot be weak with probability more than $\delta_2/20$, since otherwise by Lemma 4.2 and the analysis above, the error probability of the protocol will be at least $(1 - \delta_2/20) \cdot \delta_2/20 \cdot \delta_\ell > \delta$, for a sufficiently small

constant error δ, violating the success guarantee of the lemma. Therefore,

$$\begin{aligned} & \mathbf{Pr}[\Pi \text{ is normal}] \\ \geq\ & \mathbf{Pr}[\Pi \text{ is normal and strong}] \\ \geq\ & \mathbf{Pr}[\Pi \text{ is normal}]\ \mathbf{Pr}[\Pi \text{ is strong} \mid \Pi \text{ is normal}] \\ \geq\ & (1-\delta_2/20)\cdot(1-\delta_2/20) \\ \geq\ & (1-\delta_2/10). \end{aligned}$$

Now we perform a reduction from 2-SUM to k-SUM.

LEMMA 4.3. *Suppose there exists a deterministic protocol $\mathcal{P}'$ which computes k-SUM on input distribution ψ with error probability δ_3, for a sufficiently small constant δ_3, and communication $o(C)$. Then there exists a deterministic protocol $\mathcal{P}$ that computes 2-SUM on input distribution ν with error probability δ_2 and expected communication $o(C/k)$.*

Proof. Given protocol $\mathcal{P}'$, Alice and Bob can solve 2-SUM on input $(A,B) \sim \nu$ as follows. They first construct an input $(X_1,\ldots,X_k,Y) \sim \psi$ for k-SUM using (A,B). We call this step *input reduction*. They then run protocol $\mathcal{P}'$ on $(X_1,\ldots,X_k,Y)$. Finally, they use the resulting protocol transcript to solve 2-SUM on input (A,B). In the input reduction, for convenience, we allow Alice and Bob to use both public and private randomness. We will fix all the randomness at the end of the argument.

Input reduction.

1. Alice and Bob pick a random player P_I ($I \in [k]$) using public randomness.
2. Alice simulates P_I. She assigns P_I the input $X_I = A$.
3. Bob simulates and constructs inputs for the remaining $(k-1)$ players and the coordinator. He assigns the coordinator the input $Y = B$. Next, he uses private randomness to generate $X_1,\ldots,X_{I-1},X_{I+1},\ldots,X_k$ independently according to $\nu|Y$, and assigns them to $P_1,\ldots,P_{I-1},P_{I+1},\ldots,P_k$, respectively.

The resulting $(X_1,\ldots,X_k,Y)$ is distributed according to ψ.

Next, Alice and Bob run $\mathcal{P}'$ on $(X_1,\ldots,X_k,Y)$. By Lemma 4.1, with probability $1-\delta_2/10$, we obtain a strong $\Pi=\pi$. For a strong $\Pi=\pi$, by a Markov inequality, for at least a $(1-\delta_2/10)$ fraction of $i \in [k]$, it holds that $\mathbf{Var}[Z_i \mid \Pi=\pi] \leq \beta t/c_1$, where $c_1 = (\delta_2/10)\cdot c_0 = 20/\delta_2$. Let G^π be the collection of such i. For a strong $\Pi=\pi$, and an $i \in G^\pi$, by Chebyshev's inequality, we have

$$\mathbf{Pr}\left[\,|Z_i - \mathbf{E}[Z_i \mid \Pi=\pi]| \leq \sqrt{c_1}\cdot\sqrt{\beta t/c_1}\ \middle|\ \Pi=\pi\right] \geq 1-1/c_1.$$

Since I is chosen randomly from $[k]$, by a union bound, with probability $1-\delta_2/10-\delta_2/10-1/c_1 = 1-\delta_2/4$ over the input distribution ν and the randomness used in the input reduction, we get a strong $\Pi=\pi$ and $I \in G^\pi$ is such that $|Z_I - \mathbf{E}[Z_i \mid \Pi=\pi]| \leq \sqrt{\beta t}$. That is, we approximate $Z_I = 2\text{-SUM}(A,B)$ up to an additive error $\sqrt{\beta t}$.

We next analyze the communication cost. Since I is chosen randomly from $[k]$, and conditioned on Y, X_i ($i \in [k]$) are independent and identically distributed, the expected communication between player P_I and the coordinator (or equivalently, the expected communication between Alice and Bob in the simulation) is equal to the total communication among the k players and the coordinator divided by a factor of k, which is $o(C/k)$, where the expectation is taken over the input distribution ν and the choice of I.

Finally we use two Markov inequalities to fix all the randomness used in the reduction, such that the resulting deterministic protocol $\mathcal{P}$ succeeds with probability $1-\delta_2$ on input distribution ν, and the expected communication cost is $o(C/(4k)) = o(C/k)$.

Combining Lemma 4.3 and Theorem 4.2, we have the following theorem for k-SUM.

THEOREM 4.4. *$D^{\delta_3}_{\psi}(k\text{-SUM}) = \Omega(ktL)$, for a sufficiently small constant δ_3.*

4.4 A Lower Bound for F_0 when $\Omega(1) \leq k \leq O(1/\epsilon^2)$ In this section we set $\delta = \delta_3/2$, $c_L = 1000/\delta$, $L = c_L k$, $\gamma = 1/(12c_L)$, and $t = 1/(\epsilon^2 k)$. We define an input distribution ζ for the F_0 problem.

ζ: We choose $(X_1,\ldots,X_k,Y) \sim \psi$, and write $X_i = (X_i^1,\ldots,X_i^t)$ where $X_i^j \in \{0,1\}^L$. Next, for each $i \in [k], j \in [t]$ and $\ell \in [L]$, we assign an item $(j-1)L+\ell$ to site P_i if $X^j_{i,\ell} = 1$.

Note that the size of the universe of items is $n = tL = 1/(\epsilon^2 k)\cdot c_L k = \Theta(1/\epsilon^2)$.

THEOREM 4.5. *Suppose that $\Omega(1) \leq k \leq O(1/\epsilon^2)$. Then $R^{1/3}((1+\epsilon)\text{-approximate } F_0) = \Omega(k/\epsilon^2)$.*

We first show a reduction from k-SUM to F_0.

LEMMA 4.4. *Any deterministic protocol $\mathcal{P}$ which computes a $(1+\gamma\epsilon)$-approximation to F_0, for a sufficiently small constant $\gamma > 0$, on the above input distribution ζ with error probability δ and communication C can be used to compute k-SUM on input distribution ν with error probability $2\delta(=\delta_3)$ and communication C.*

Let $B \sim \zeta$. Let M_i^j ($i \in [k], j \in [t]$) be the index of the special coordinate when sampling (X_i^j, Y^j) from

μ. We prove the lemma by establishing a relationship between $F_0(B)$ and

$$\begin{aligned} & k\text{-SUM}(X_1,\ldots,X_k,Y) \\ = & \sum_{j\in[t]}\left(\sum_{i\in[k]} 2\text{-DISJ}(X_i^j,Y^j)\right) \\ = & \sum_{j\in[t]}\left(\sum_{i\in[k]} \text{AND}(X^j_{i,M_i^j},Y^j_{i,M_i^j})\right). \end{aligned}$$

Let $N^j = \{i \mid \text{AND}(X^j_{i,M_i^j},Y^j_{i,M_i^j}) = 1\}$. Let $U^j = \left|N^j\right|$, and let $U = \sum_{j\in[t]} U^j$. Thus $U = k\text{-SUM}(X_1,\ldots,X_k,Y)$. Let $R^j = \left|\{M_i^j \mid i \in N^j\}\right|$, and let $R = \sum_{j\in[t]} R^j$. Let $Q^j = \left|\{\cup_{i\in[k]} X_i^j \setminus Y^j]\}\right|$, and let $Q = \sum_{j\in[t]} Q^j$. For convenience, in the remainder of the paper when we write $\mathbf{E}[Q]$ and $\mathbf{E}[R]$, we actually mean $\mathbf{E}[Q \mid Y = y]$ and $\mathbf{E}[R \mid Y = y]$ given a fixed $Y = y$ (the input of the coordinator).

We start by proving a technical lemma. Roughly speaking, it shows that $F_0(B)$ is tightly concentrated around $\mathbf{E}[Q]+\mathbf{E}[R]$, and $\mathbf{E}[R]$ has a fixed relationship with $U = k\text{-SUM}(X_1,\ldots,X_k,Y)$.

LEMMA 4.5. *With probability $1-2\delta$, it holds that $F_0(B) = \mathbf{E}[Q]+\mathbf{E}[R]+\kappa_1 = \mathbf{E}[Q]+(1-\lambda)U+\kappa_1$, where $|\kappa_1| \le 1/(4\epsilon)$, and λ is a fixed constant in $[0, 5/(4c_L)]$.*

We prove this lemma by two claims. Please see the paragraph below Lemma 4.4 for the definitions of relevant notations (Q, R, U, etc.), and the values of parameters (c_L, L, etc.) at the beginning of Section 4.4.

CLAIM 2. *With probability $1-e^{-\Omega(k)}$, it holds that $F_0(B) = \mathbf{E}[Q]+R+\kappa_0$, where $|\kappa_0| \le 1/(8\epsilon)$.*

Proof. We first consider $j = 1$. Recall in the input distribution ψ, we have $(X_i^1, Y^1) \sim \mu$ for all $i \in [k]$. Let $D^1_{i,\ell}$ ($i \in [k], \ell \in [L]$) be the random variable which is chosen from $\{0,1\}$ uniformly at random when sampling (X_i^1, Y^1) from μ: If $D^1_{i,\ell} = 0$, then $X^1_{i,\ell} = 0$ and Y^1_ℓ is chosen from $\{0,1\}$ uniformly at random; and if $D^1_{i,\ell} = 1$, then $Y^1_\ell = 0$ and $X^1_{i,\ell}$ is chosen from $\{0,1\}$ uniformly at random.

For any $i \in [k]$ and $\ell \in [L]$, we consider the probability that $X^1_{i,\ell} = 1$ conditioned on $Y^1_\ell = 0$.

$$\begin{aligned} & \mathbf{Pr}[X^1_{i,\ell} = 1 \mid Y^1_\ell = 0] \\ = & \mathbf{Pr}[D^1_{i,\ell} = 0]\cdot\mathbf{Pr}[X^1_{i,\ell} = 1 \mid Y^1_\ell = 0, D^1_{i,\ell} = 0] \\ & +\mathbf{Pr}[D^1_{i,\ell} = 1]\cdot\mathbf{Pr}[X^1_{i,\ell} = 1 \mid Y^1_\ell = 0, D^1_{i,\ell} = 1] \\ \ge & 1/2\cdot\mathbf{Pr}[X^1_{i,\ell} = 1 \mid Y^1_\ell = 0, D^1_{i,\ell} = 1] \\ \ge & 1/2\cdot\mathbf{Pr}[\ell \ne M_i^1]\cdot\mathbf{Pr}[X^1_{i,\ell} = 1 \mid D^1_{i,\ell} = 1, \ell \ne M_i^1] \\ & -1/2\cdot\mathbf{Pr}[\ell = M_i^1] \\ = & 1/4\cdot(1-1/L) - 1/(2L) \\ \ge & 1/5. \end{aligned}$$

By a Chernoff bound, for an ℓ such that $Y^1_\ell = 0$, $\sum_{i\in[k]} X^1_{i,\ell} \ge 1$ with probability $1-e^{-\Omega(k)}$.

Similarly, we can show that, for each j, for each $\ell \in [L]$ such that $Y^j_\ell = 0$, it holds that $\sum_{i\in[k]} X^j_{i,\ell} \ge 1$ with probability at least $1-e^{-\Omega(k)}$. Therefore, we have

$$\mathbf{Var}[Q] \le \sum_{j\in[t]}\sum_{\ell\in[L]}(1-e^{-\Omega(k)})\cdot e^{-\Omega(k)} \le c_L kte^{-\Omega(k)}.$$

By Chebyshev's inequality, it holds that

$$\begin{aligned} & \mathbf{Pr}[|Q-\mathbf{E}[Q]| > \sqrt{kt}/8] \\ \le & \frac{\mathbf{Var}[Q]}{kt/64} < \frac{c_L kte^{-\Omega(k)}}{kt/64} \le e^{-\Omega(k)}. \end{aligned}$$

Note that items corresponding to $\bigcup_{i\in[k]} X_i^j$ are different from those corresponding to $\bigcup_{i\in[k]} X_i^{j'}$ ($j' \ne j$), we thus have

$$\begin{aligned} & F_0(B) = \sum_{j\in[t]} F_0(X_1^j,\ldots,X_k^j) \\ = & \sum_{j\in[t]}\left(\left|\{\cup_{i\in[k]} X_i^j \setminus Y^j]\}\right| + \left|\{\cup_{i\in[k]} X_i^j \bigcap Y^j]\}\right|\right) \\ = & Q+R \\ = & (\mathbf{E}[Q]+\kappa_0)+R, \end{aligned}$$

where $|\kappa_0| \le \sqrt{kt}/8 = 1/(8\epsilon)$ with probability $1-e^{-\Omega(k)}$.

We next analyze the value R.

CLAIM 3. *With probability $1-\delta/2$, we have $|R-\mathbf{E}[R]| \le \sqrt{\beta kt}/4$, where $\mathbf{E}[R] = (1-\lambda)U$ for a fixed constant $0 \le \lambda \le 5/(4c_L)$.*

Proof. For a vector $V \in \{0,1\}^L$, let $wt_1(V) = \{\ell \mid V_\ell = 1\}$.

We first consider $j = 1$. For a fixed $Y^1 = y$, for each $i \in [k]$, we consider the probability $\mathbf{Pr}[M_i^1 = \ell \mid Y^1 = y]$

for an $\ell \in [L]$. By Bayes' theorem,

$$\begin{aligned}&\mathbf{Pr}[M_i^1 = \ell \mid Y^1 = y]\\ =\ &\frac{\mathbf{Pr}[Y^1 = y \mid M_i^1 = \ell] \cdot \mathbf{Pr}[M_i^1 = \ell]}{\mathbf{Pr}[Y^1 = y]}.\end{aligned}$$

Let $s_\ell = \left|\{\ell' \mid Y_{\ell'}^1 = 1, \ell' \neq \ell\}\right|$. Then

$$\mathbf{Pr}[Y^1 = y \mid M_i^1 = \ell] = \frac{1}{2} \cdot \left(\frac{1}{4}\right)^{s_\ell} \cdot \left(\frac{3}{4}\right)^{L-1-s_\ell}.$$

Now we can write

$$\begin{aligned}&\mathbf{Pr}[M_i^1 = \ell \mid Y^1 = y]\\ =\ &\frac{\frac{1}{2} \cdot \left(\frac{1}{4}\right)^{s_\ell} \cdot \left(\frac{3}{4}\right)^{L-1-s_\ell} \cdot \frac{1}{L}}{\sum_{\ell' \in [L]} \left(\frac{1}{L} \cdot \frac{1}{2} \cdot \left(\frac{1}{4}\right)^{s_{\ell'}} \cdot \left(\frac{3}{4}\right)^{L-1-s_{\ell'}}\right)}.\end{aligned} \tag{4.10}$$

Let $s_\ell = s$ if $y_\ell = 1$. Then $s_\ell = s + 1$ if $y_\ell = 0$. Thus we can write (4.10) as

$$\mathbf{Pr}[M_i^1 = \ell \mid y_\ell = 1] = \frac{1}{wt_1(y) + (L - wt_1(y))/3}, \tag{4.11}$$

which is a fixed value for a fixed $Y^1 = y$.

We can view the value of R^1 as a result of a bin-ball game: we view $\{M_i^1 \mid i \in N^1\}$ as the balls and $\{\ell \mid Y_\ell^1 = 1\}$ as the bins. We throw balls to bins uniformly at random, and the result of the game is the number of non-empty bins at the end. We can think balls are thrown to bins uniformly at random, since given a fixed $Y^1 = y$, $\mathbf{Pr}[M_i^1 = \ell \mid y_\ell = 1]$ is fixed for all $i \in [k]$ by Equation (4.11), and $\mathbf{Pr}[\ell \in N^1 \mid M_i^1 = \ell, y_\ell = 1] = 1/2$ for all $i \in [k]$ by our choice of the input distribution.

Let $U^1 = \left|N^1\right|$ be the number of balls, and $V^1 = wt_1(Y^1)$ be the number of bins. By Chernoff bounds, with probability $1 - e^{-\Omega(L)} - e^{-\Omega(k)} = 1 - e^{-\Omega(k)}$ (recall that we have set $L = c_L k$ for a constant c_L), we have $V^1 \geq L/4 - 0.01L \geq L/5$ and $U^1 \leq 2\beta k = k/2$ (recall that we have set $\beta = 1/4$). By Fact 1 and Lemma 1 in [20], we have

1. $\mathbf{E}[R^1] = V^1(1-(1-1/V^1)^{U^1}) = (1-\lambda)U^1$ for some fixed constant $0 \leq \lambda \leq U^1/(2V^1) \leq 5/(4c_L)$.
2. $\mathbf{Var}[R^1] < 4(U^1)^2/V^1 \cdot (1 - e^{-\Omega(k)}) + k^2 \cdot e^{-\Omega(k)} \leq 5k/c_L + k^2 \cdot e^{-\Omega(k)} \leq 6k/c_L$.

By the same argument we can show that $\mathbf{E}[R^j] = (1-\lambda)U^j$ and $\mathbf{Var}[R^j] < 6k/c_L$ for all $j \in [t]$. Using the fact that $R^1, \ldots, R^t$ are independent by our choice of the input distribution, we get $\mathbf{E}[R] = \sum_{j\in[t]} \mathbf{E}[R^j] = (1-\lambda)\sum_{j\in[t]} U^j = (1-\lambda)U$, and $\mathbf{Var}[R] = \sum_{j\in[t]} \mathbf{Var}[R^j] < 6kt/c_L$. By Chebyshev's inequality, we have (recall we set $c_L = 1000/\delta$)

$$\mathbf{Pr}[|R - \mathbf{E}[R]| > \sqrt{\beta kt}/4] \leq \frac{\mathbf{Var}[R]}{\beta kt/16} < \frac{6kt/c_L}{kt/64} \leq \delta/2.$$

Therefore with probability $1 - \delta/2$, the claim holds.

By Claim 2, Claim 3, and the fact that $\mathcal{P}$ computes F_0 correctly with error probability δ, we have that with probability $1 - e^{-\Omega(k)} - \delta/2 - \delta \geq 1 - 2\delta$,

$$F_0(B) = \mathbf{E}[Q] + (1-\lambda)U + \kappa_1,$$

where $|\kappa_1| \leq \sqrt{\beta kt}/4 + |\kappa_0| \leq 1/(8\epsilon) + 1/(8\epsilon) = 1/(4\epsilon)$.

Proof. (of Lemma 4.4) Given a W which is a $(1+\gamma\epsilon)$-approximation to $F_0(B)$, by Lemma 4.5, with probability $1 - 2\delta$ it holds that $W = \mathbf{E}[Q] + (1-\lambda)U + \kappa_2$, where

$$|\kappa_2| \leq \gamma\epsilon \cdot F_0(B) + |\kappa_1| \leq \gamma\epsilon tL + |\kappa_1| \leq 1/(3\epsilon)$$

(for a small enough constant γ). Now the coordinator who holds Y can approximate U using W: $U = \frac{W - \mathbf{E}[Q]}{1-\lambda}$. The additive error of this approximation is at most $|\kappa_2|/(1-\lambda) < 1/(2\epsilon) = \sqrt{\beta kt}$. Thus $\mathcal{P}$ can be used to compute k-SUM correctly with probability at least $1 - 2\delta$.

Finally, Theorem 4.5 follows immediately from Lemma 4.4, Theorem 4.4, Lemma 2.1, and the fact that $R^{1/3}(f) = \Theta(R^\delta(f))$ for any constant $\delta \leq 1/3$. By our choices of t and L, $\Omega(ktL) = \Omega(k/\epsilon^2)$.

4.5 An Improved Lower Bound for F_0 when $k \geq \Omega(1/\epsilon^2)$

As mentioned, in [25] a lower bound of $\Omega(k/(\epsilon^2 \log(\epsilon^2 k)))$ was shown for F_0 when $k \geq \Omega(1/\epsilon^2)$. Here we make an observation that we actually can improve it to $\Omega(k/\epsilon^2)$, by a better embedding argument.

In [25], a problem called k-APPROX-SUM is defined in the coordinator model. This problem is similar to a degenerate case (when $t = 1$) of the k-SUM problem defined in Section 4.3, but with a different input distribution. In k-APPROX-SUM, the coordinator has input Y and each site P_i has input X_i. We choose $(X_1, Y) \sim \tau_\eta$, and then independently choose $X_2, \ldots, X_k \sim \tau_\eta | Y$. Here τ_η is an input distribution for 2-DISJ, defined as follows.

- τ_η: Let $n = \Theta(1/\epsilon^2)$. Let $\ell = (n+1)/4$. With probability η, x and y are random subsets of $[n]$ such that $|x| = |y| = \ell$ and $|x \cap y| = 1$. And with probability $1-\eta$, x and y are random subsets of $[n]$ such that $|x| = |y| = \ell$ and $x \cap y = \emptyset$.

Let $\eth_\eta$ be this input distribution for k-APPROX-SUM. Let $Z_i = 2\text{-DISJ}(X_i, Y)$. The goal of k-APPROX-SUM is to approximate $\sum_{i\in[k]} Z_i$ up to an additive error of $\sqrt{\eta k}$.

The following reduction from 2-DISJ to k-APPROX-SUM was established in [25]. The lower

bound for F_0 follows by another reduction from k-APPROX-SUM (Lemma 8 in [25]), and a lower bound for 2-DISJ: for any $\eta \le 1/4$, $\mathsf{ED}^{\eta/100}_{\tau_\eta}(\text{2-DISJ}) = \Omega(n)$. We refer readers to [25] for details.

For convenience, we set $\rho \triangleq 1/(\epsilon^2 k)$ in the remaining of the section.

LEMMA 4.6. ([25], LEMMA 9) [4] *Suppose that there exists a deterministic protocol $\mathcal{P}'$ which computes k-APPROX-SUM on input distribution $\eth_\rho$ with error probability δ (for a sufficiently small constant δ) and communication C, then there exists a deterministic protocol $\mathcal{P}$ that computes 2-DISJ on input distribution τ_ρ with error probability $\rho/100$ and expected communication $O(\log(\epsilon^2 k) \cdot C/k)$, where the expectation is taken over the input distribution τ_ρ.*

In this paper we improve this lemma to the following.

LEMMA 4.7. *Suppose that there exists a deterministic protocol $\mathcal{P}'$ which computes k-APPROX-SUM on input distribution $\eth_\rho$ with error probability δ (for a sufficiently small constant δ) and communication C, then there exists a deterministic protocol $\mathcal{P}$ that computes 2-DISJ on input distribution $\tau_{1/4}$ with error probability $1/400$ and communication $O(C/k)$.*

Note that we have shaved a $\log(\epsilon^2 k)$ factor in the communication in the reduction. This is how we manage to improve the lower bound of F_0 when $k \ge \Omega(1/\epsilon^2)$ by a $\log(\epsilon^2 k)$ factor. The improvement comes from the observation that we do not need a totally symmetric distribution of $X_1, \ldots, X_k$. We have also replaced "expected communication" by "(worst-case) communication" in the last sentence of the lemma, which only helps since the worst-case communication cost is always at least the expected communication cost.

For completeness, in the proof of Lemma 4.7 we first repeat part of the proof for Lemma 4.6 in [25], and then address the modifications. We need the following definition and theorem from [25] for k-APPROX-SUM. Let $\delta_1 > 0$ be a sufficiently small constant. In k-APPROX-SUM, for a fixed transcript $\Pi = \pi$, let $q_i^\pi = \mathbf{Pr}[Z_i = 1 \mid \Pi = \pi]$.

DEFINITION 4. ([25]) *Given an input $(x_1, \ldots, x_k, y)$ for k-APPROX-SUM and a transcript $\Pi = \pi$, let $z_i = \text{2-DISJ}(x_i, y)$ and $z = \{z_1, \ldots, z_k\}$. Define $\Pi(z) \triangleq \Pi(x_1, \ldots, x_k, y)$. Let c_0 be a large enough constant. We say π is* good *for z if $\Pi(z) = \pi$, and for at least a $1-\delta_1$ fraction of $\{i \in [k] \mid z_i = 1\}$, it holds that $q_i^\pi > \rho/c_0$; and for at least $1-\delta_1$ fraction of $\{i \in [k] \mid z_i = 0\}$, it holds that $q_i^\pi < \rho/c_0$.*

THEOREM 4.6. ([25]) *Let Π be the transcript of any deterministic protocol for k-APPROX-SUM on input distribution $\eth_\rho$ with error probability δ for some sufficiently small constant δ, then $\mathbf{Pr}_{\eth_\rho}[\Pi \text{ is good}] \ge 1 - \delta_1$.*

Proof. (of Lemma 4.7) In 2-DISJ, Alice has A and Bob has B, where $(A, B) \sim \tau_\rho$. The reduction again consists of two phases. During the reduction, for convenience, we will use public and private randomness, which we will fix at the end.

Input reduction phase. Alice and Bob construct an input for k-APPROX-SUM using A and B. They pick a random site P_I ($I \in [k]$) using public randomness. Alice assigns P_I the input $X_I = A$; Bob assigns the coordinator with input $Y = B$, and constructs inputs for the remaining $k-1$ sites as follows: for each $i \in [k] \backslash I$, Bob samples an X_i according to $\tau_\rho | Y$ using independent private randomness, and assigns it to P_i. Let $Z_i = \text{2-DISJ}(X_i, Y)$. Note that $\{X_1, \ldots, X_k, Y\} \sim \eth_\rho$. Intuitively speaking, we "embed" the input (A, B) for 2-DISJ between P_I and the coordinator in the k-APPROX-SUM.

Simulation phase. Alice simulates P_I, and Bob simulates the remaining $k-1$ sites and the coordinator. They run protocol $\mathcal{P}'$ on $\{X_1, \ldots, X_k, Y\} \sim \eth_\rho$ to compute k-APPROX-SUM with error probability δ. By Theorem 4.6, for a $1-\delta_1$ fraction of $Z = z$ over the input distribution $\eth_\rho$ and $\pi = \Pi(z)$, it holds that for a $1-\delta_1$ fraction of $\{i \in [k] \mid z_i = 0\}$, $q_i^\pi < \rho/c_0$, and a $1-\delta_1$ fraction of $\{i \in [k] \mid z_i = 1\}$, $q_i^\pi > \rho/c_0$. Now they output 1 if $q_I^\pi > \rho/c_0$, and 0 otherwise.

Now we show the correctness of the protocol and analyze the communication cost. Since P_I is chosen randomly among the k sites, and conditioned on Y, all X_i ($i \in [k]$) are independent and identically distributed, $\mathcal{P}'$ computes $Z_I = \text{2-DISJ}(X, Y)$ on input distribution τ_ρ correctly with error probability at most $\delta_1 + \delta_1 \le 2\delta_1$, and the expected communication between P_I and the coordinator is C/k. Both the probability and the expectation are taken over the input distribution τ_ρ. By a Markov inequality, with error probability at most $2\delta_1 + \delta_2$ (for an arbitrarilyy small constant δ_2) over the input distribution, it holds that the communication between P_I and the coordinator is at most $\kappa_3 C/k$ for a large enough constant κ_3. Finally, using two Markov inequalities, we can fix all the randomness used in the reduction, and the resulting deterministic protocol $\mathcal{P}''$ for 2-DISJ under input distribution τ_ρ has communication cost $4\kappa_3 C/k$ and error probability $4(2\delta_1 + \delta_2)$.

We next modify the embedding. We change the

[4] The original Lemma 9 in [25] is in fact a combination of this lemma and Lemma 8 in [25].

input distribution for Alice and Bob from $(A, B) \sim \tau_\rho$ to $(A, B) \sim \tau_{1/4}$. Let $Z'_I = \text{2-DISJ}(A, B) \sim \text{Bernoulli}(1/4)$. We still perform the same input reduction and simulation as $\mathcal{P}''$. The only modification on $\mathcal{P}''$ to obtain the final protocol $\mathcal{P}$ is that in the simulation phase, when the communication between Alice and Bob (equivalently, between P_I and the coordinator) exceeds $4\kappa_3 C/k$, we early-terminate the protocol and output "error".

Obviously, the communication cost of $\mathcal{P}$ is always bounded by $4\kappa_3 C/k = O(C/k)$. Now we analyze the error probability of $\mathcal{P}$. Let $\mathsf{TV}(\zeta, \zeta')$ be the total variation distance between distributions ζ and ζ', which is defined to be $\max_{A \subseteq \mathcal{X}} |\zeta(A) - \zeta'(A)|$, where $\mathcal{X}$ is the union of supports of distributions ζ, ζ'. The observation is that if we change $Z_I \sim \text{Bernoulli}(\rho)$ to $Z'_I \sim \text{Bernoulli}(1/4)$, the total variation distance between $(Z_1, \ldots, Z_I, \ldots, Z_k)$ (all $Z_i \sim \text{Bernoulli}(\rho)$) and $(Z_1, \ldots, Z'_I, \ldots, Z_k)$ is at most

$$\max\{\mathsf{TV}(\text{Binomial}(k, \eta), \text{Binomial}(k-1, \eta)), \\ \mathsf{TV}(\text{Binomial}(k, \eta), \text{Binomial}(k-1, \eta)+1)\},$$

which can be bounded by $O(1/\sqrt{\eta k}) = O(\epsilon)$ (see, e.g., Fact 2.4 of [17]). Therefore, under the input distribution $\tau_{1/4}$, $\mathbf{Pr}[\mathcal{P} \text{ early-terminates}] \leq \mathcal{O}(\epsilon)$, and $\mathbf{Pr}[\mathcal{P} \text{ is correct} \mid \mathcal{P}\text{normally-terminates}] \leq 4(2\delta_1 + \delta_2) + O(\epsilon)$. Thus the total error probability can be bounded by $(4(2\delta_1 + \delta_2) + O(\epsilon)) < 1/400$, by choosing constants δ_1, δ_2 sufficiently small.

References

[1] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In *Proc. ACM Symposium on Theory of Computing*, 1996.

[2] C. Arackaparambil, J. Brody, and A. Chakrabarti. Functional monitoring without monotonicity. In *Proc. International Colloquium on Automata, Languages, and Programming*, 2009.

[3] Z. Bar-Yossef, T. S. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. Counting distinct elements in a data stream. In *RANDOM*, 2002.

[4] K. S. Beyer, P. J. Haas, B. Reinwald, Y. Sismanis, and R. Gemulla. On synopses for distinct-value estimation under multiset operations. In *SIGMOD Conference*, pages 199–210, 2007.

[5] M. Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

[6] M. Braverman, F. Ellen, R. Oshman, T. Pitassi, and V. Vaikuntanathan. Tight bounds for set disjointness in the message passing model. *CoRR*, abs/1305.4696, 2013.

[7] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proc. ACM Symposium on Theory of Computing*, 2011.

[8] E. Cohen. Size-estimation framework with applications to transitive closure and reachability. *J. Comput. Syst. Sci.*, 55(3):441–453, 1997.

[9] G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed functional monitoring. *ACM Transactions on Algorithms*, 7(2):21, 2011.

[10] G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed functional monitoring. *ACM Transactions on Algorithms*, 7(2), Article 21, 2011. Preliminary version in SODA'08.

[11] M. Datar and S. Muthukrishnan. Estimating rarity and similarity over data stream windows. In *ESA*, pages 323–334, 2002.

[12] D. Dolev and T. Feder. Determinism vs. nondeterminism in multiparty communication complexity. *SIAM J. Comput.*, 21(5):889–895, 1992.

[13] M. Durand and P. Flajolet. Loglog counting of large cardinalities (extended abstract). In *ESA*, pages 605–617, 2003.

[14] C. Estan, G. Varghese, and M. E. Fisk. Bitmap algorithms for counting active flows on high-speed links. *IEEE/ACM Trans. Netw.*, 14(5):925–937, 2006.

[15] W. Feller. Generalization of a probability limit theorem of cramer. *Trans. Amer. Math. Soc*, 54(3):361–372, 1943.

[16] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31(2):182–209, 1985.

[17] P. Gopalan, R. Meka, O. Reingold, and D. Zuckerman. Pseudorandom generators for combinatorial shapes. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 253–262, New York, NY, USA, 2011. ACM.

[18] Z. Huang, B. Radunovic, M. Vojnovic, and Q. Zhang. Communication complexity of approximate maximum matching in distributed graph data, 2013.

[19] D. M. Kane, J. Nelson, and D. P. Woodruff. An optimal algorithm for the distinct elements problem. In *PODS*, pages 41–52, 2010.

[20] D. M. Kane, J. Nelson, and D. P. Woodruff. An optimal algorithm for the distinct elements problem. In *Proc. ACM Symposium on Principles of Database Systems*, pages 41–52, 2010.

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

[22] J. Matousek and J. Vondrák. *The probabilistic method*. Lecture Notes, 2008.

[23] A. McGregor, A. Pavan, S. Tirthapura, and D. P. Woodruff. Space-efficient estimation of statistics over sub-sampled streams. In *PODS*, pages 273–282, 2012.

[24] J. M. Phillips, E. Verbin, and Q. Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *Proc. ACM-SIAM Symposium on Discrete Algorithms*, 2012.

[25] D. P. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. *CoRR*, abs/1112.5153, 2011.

[26] D. P. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 941–960, New York, NY, USA, 2012. ACM.

[27] D. P. Woodruff and Q. Zhang. When distributed computation is communication expensive. In *DISC, to appear*, 2013.

[28] A. C. Yao. Probabilistic computations: Towards a unified measure of complexity. In *Proc. IEEE Symposium on Foundations of Computer Science*, 1977.

A Proof for Lemma 2.1

Proof. If Π is a δ-error protocol then for all possible inputs $x^1, \ldots, x^k$ to the k players, let R be the randomness used by the k players,

$$\mathbf{Pr}_R[\Pi(x^1, \ldots, x^k) = f(x^1, \ldots, x^k)] \geq 1 - \delta,$$

which implies for any distribution μ on $(x^1, \ldots, x^k)$ that

$$\mathbf{Pr}_{R,(x^1,\ldots,x^k)\sim\mu}[\Pi(x^1, \ldots, x^k) = f(x^1, \ldots, x^k)] \geq 1 - \delta,$$

which implies there is a fixing of the randomness of the players so that

$$\mathbf{Pr}_{(x^1,\ldots,x^k)\sim\mu}[\Pi(x^1, \ldots, x^k) = f(x^1, \ldots, x^k)] \geq 1 - \delta,$$

which implies $D^{\delta}_{\mu}(f)$ is at most $R^{\delta}(f)$.