THEORY OF COMPUTING, Volume 12 (10), 2016, pp. 1–35
www.theoryofcomputing.org

Robust Lower Bounds for Communication and Stream Computation

Amit Chakrabarti* Graham Cormode† Andrew McGregor‡

Received June 28, 2014; Revised May 27, 2016; Published August 22, 2016

Abstract: We study the communication complexity of evaluating functions when the input data is randomly allocated (according to some known distribution) amongst two or more players, possibly with information overlap. This naturally extends previously studied variable partition models such as the best-case and worst-case partition models. We aim to understand whether the hardness of a communication problem holds for almost every allocation of the input, as opposed to holding for perhaps just a few atypical partitions.

A key application is to the heavily studied data stream model. There is a strong connection between our communication lower bounds and lower bounds in the data stream model that are "robust" to the ordering of the data. That is, we prove lower bounds for when the order of the items in the stream is chosen not adversarially but rather uniformly (or near-uniformly) from the set of all permutations. This random-order data stream model has attracted recent interest, since lower bounds here give stronger evidence for the inherent hardness of streaming problems.

A short version of this article is available in the Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC'08), ACM, pp. 641–650 [6]. Compared to the conference presentation, this version considerably expands the detail of the discussion and in the proofs, and substantially changes some of the proof techniques.

*Supported in part by NSF Awards CCF-0448277 and IIS-0916565, and a McLane Family Fellowship.

†Supported in part by European Research Council grant ERC-2014-CoG 647557, the Yahoo Faculty Research and Engagement Program and a Royal Society Wolfson Research Merit Award.

‡This work was supported by NSF Awards CCF-0953754, CCF-1320719, and a Google Research Award.

ACM Classification: F.2.2

AMS Classification: 68Q25

Key words and phrases: communication complexity, lower bounds, data streams

 DOI: 10.4086/toc.2016.v012a010

Our results include the first random-partition communication lower bounds for problems including multi-party set disjointness and gap-Hamming-distance. Both are tight. We also extend and improve previous results for a form of pointer jumping that is relevant to the problem of selection (in particular, median finding). Collectively, these results yield lower bounds for a variety of problems in the random-order data stream model, including estimating the number of distinct elements, approximating frequency moments, and quantile estimation.

1 Introduction

Since its introduction in 1979 by Yao, communication complexity [44, 32] has proven to be a powerful framework for proving lower bounds in a variety of settings, including the cell-probe and data stream models, circuit and decision tree complexity and VLSI design. The majority of results in this area involve a fixed-partition model of communication complexity, where the goal is for two or more players to evaluate a function of an input that has been partitioned between them in a particular way, e. g., computing $f(x,y)$ when one player holds x and the other has y. Many explicit functions can be shown to require a large amount of communication to evaluate when the input is partitioned between the players in this manner. These imply lower bounds for various models of computation, via arguments that such partitions necessarily arise in the course of the computation.

To a lesser extent, variable-partition models, such as best-case and worst-case partition, have also been studied; see, e. g., [2, 33, 36] and [32, Chap. 7] for a survey. For example, understanding the best-case partition complexity, where the data is partitioned in the most advantageous manner (subject to constraints such as each player receiving an equal amount of the input), is important for understanding various problems in VLSI design [2]. Another kind of worst-case partition arises when the corresponding bits of two equal-length input strings are written on opposite sides of opaque cards (the "two-sided card model" [14, 37]). However, a natural question that, to the best of our knowledge, has not been explored to date, is what happens when the input is partitioned amongst the players *at random*. In other words, does evaluating a given function require significant communication for only a few pathological partitions or does such a requirement apply to an overwhelming fraction of all partitions?

In this paper we initiate a study of communication complexity under random partitions of the input. In fact, we consider more general allocations of the input to the players, possibly allowing information overlap, where bits of data may be known to more than one player. A particularly interesting case is when each *token* of data is given to a player chosen uniformly at random; this provides a convenient way to count "bad" partitions. We consider a communication lower bound to be *robust* if it applies to all but a small fraction of possible partitions. One can think of our work as a form of average-case analysis. However, it is important to note that our work stands in contrast to the usual notion of distributional complexity: rather than considering a random input, we consider worst-case inputs allocated randomly amongst the players.

Data stream computation. A strong motivation for our study is the goal of proving robust lower bounds for problems in the data stream model. The data stream model has enjoyed significant attention in recent years owing to some influential work in the late 1990s [3, 26, 16]. Study of this model has thrived both because of the rich theoretical questions it raises and its applicability to numerous real world applications such as network monitoring and query planning in databases. Consequently, it is important to understand the complexity of problems not just in worst-case but also in “average-case” settings. To this end we prove lower bounds in the setting that the ordering of tokens in the data stream is chosen not adversarially but randomly, from the set of all permutations. Arguably, such a lower bound provides a stronger indication that a problem cannot be solved efficiently in the data stream model than a “fragile” lower bound that might depend on a clever adversarial ordering. (For further, more detailed, justification see the recent papers [23, 8]).

Random-order data streams were considered by Munro and Paterson [35] in one of the first studies of the data stream model. In recent years there has a been a resurgence of interest in this model for a variety of reasons [8, 13, 23, 22, 24, 43]. Uniform or near-uniform orderings can arise in a number of ways, such as when processing a stream of samples that are drawn independently from a non-time-varying distribution. For problems such as quantile estimation and finding frequent items it has been shown that there is a considerable difference between processing random-order stream and adversarial streams. In particular, streaming algorithms to find the median using polylogarithmic space require exponentially fewer passes if the stream is ordered randomly [23, 8].

In this paper, we use robust lower bounds on communication complexity in order to deduce robust data stream lower bounds. Once the communication bounds have been shown, the data stream bounds follow by simple reductions to appropriate instances of communication. Where such bounds were known before, our method yields cleaner proofs and tighter bounds. It also yields a number of new bounds for random-order data streams.

Our results and overview. We begin in Section 2 with a formal definition of our model and introduce some techniques and terminology. We prove the following results.

- *Multi-Party Set Disjointness:* We consider the problem of t-way set disjointness where each entry of the relevant $t \times n$ matrix is given to one of p players chosen uniformly at random. If $p = \Omega(t^2)$ then we show that any randomised protocol requires $\Omega(n/t)$ communication. See Section 3.
- *Pointer Jumping and Selection:* We consider a natural variant of tree pointer jumping, called weight-based tree pointer jumping, that is related to the problem of selection. In this problem, instead of an explicit pointer at each node, we have a binary string at each node whose weight encodes the pointer. We consider trees of depth $p+1$ and show that if the bits of these strings are distributed uniformly between two players, then, for every constant $\varepsilon > 0$, any p-round randomised protocol requires $\Omega(n^{(2+\varepsilon)^{-p}})$ bits of communication. See Section 4 for details of this two-player result and a generalization to more than two players.
- *Hamming Distance and Index:* For $x, y \in \{0,1\}^n$, let $\Delta(x,y) := |\{i \in [n] : x_i \neq y_i\}|$ denote the Hamming distance between x and y. We show that, for some constant c, any protocol that can distinguish between the cases $\Delta(x,y) \leq n/2 - c\sqrt{n}$ and $\Delta(x,y) \geq n/2 + c\sqrt{n}$ requires $\Omega(n)$ communication if the $2n$ input bits are split uniformly between two players. We also show that a one-way protocol for the index problem—$\text{INDEX}(x, j) := x_j$, with $x \in \{0,1\}^n$, $j \in [n]$—requires $\Omega(n)$ communication if the $n+1$ tokens (j being a single token) are split uniformly between two players. See Section 5.

The above communication lower bounds lead to lower bounds for a number of data stream problems in the random-order model. In Section 6, we deduce such bounds, many of which are tight, for approximating frequency moments, the number of distinct values, entropy, information divergences, selection, and graph connectivity. Two of these bounds deserve particular emphasis. For the k-th frequency moment, we obtain a robust lower bound of $\Omega(n^{1-3/k})$ for $k \geq 3$, where n is the universe size, which comes close to the optimal $\Omega(n^{1-2/k})$ bound under adversarial ordering. For the problem of finding the median of a stream of length m, our framework greatly simplifies the proof of an $\Omega(\log\log m)$ lower bound [8] on the number of passes required to achieve polylogarithmic space. Note that in a multi-pass algorithm, the data is seen in the same order in each pass. Further, our pass-space tradeoff for this problem improves the results of [8]. For instance, with two passes, we obtain a space lower bound of $\Omega(m^{1/10})$ as compared with their $\Omega(m^{3/80})$.

2 Notation and preliminaries

We summarise some notation that we use throughout the paper. We use "log" and "ln" to denote base-2 and natural logarithms, respectively. Define the *weight* $|x|$ of a Boolean vector $x \in \{0,1\}^N$ to be $|\{i : x_i = 1\}|$. Let $\mathbf{e}_i$ denote the vector that is 1 at location i and 0 elsewhere. For random variables X and Y, let $\mathbb{E}[X]$ denote the expectation and $\mathrm{H}(X)$ the entropy of X, $\mathrm{H}(X \mid Y)$ the conditional entropy of X given Y, and $\mathrm{I}(X : Y)$ the mutual information between X and Y. We use some basic results from information theory at certain points in this paper; the textbook by Cover and Thomas [12] is a good reference for all such results. We write $X \sim \mu$ to indicate that X is drawn from the probability distribution μ, and $X \equiv Y$ to indicate that X and Y have the same distribution. We denote the product of the distributions μ and ν by $\mu \otimes \nu$. We use the notation $X \in_R S$ to denote that the random variable X is uniform over the set S.

There are a large number of natural notions of "distance" between two probability distributions μ and ν. In this paper, we use three of them: the total variation distance

$$\mathrm{D_{TV}}(\mu,\nu) = \frac{1}{2}\|\mu - \nu\|_1 ,$$

the Hellinger distance

$$\mathrm{h}(\mu,\nu) = \frac{1}{\sqrt{2}}\|\sqrt{\mu} - \sqrt{\nu}\|_2 ,$$

where "$\sqrt{\cdot}$" denotes the pointwise positive square root, and the Kullback-Leibler divergence $\mathrm{D_{KL}}(\mu\|\nu)$, which is also known as relative entropy. Unlike the first two of these "distances," the third is not a metric.

The Binomial distribution with parameters n (number of trials) and p (success probability) is denoted $\mathcal{B}(n,p)$. For an integer k, $\binom{S}{k}$ denotes the set of all k-subsets of S and 2^S denotes the power set of S. We say that a real quantity Q' is an (ε,δ)-approximation for Q if $\Pr[|Q'-Q| > \varepsilon Q] \leq \delta$. For a real value $x \in [0,1]$ we let $\mathrm{H}_b(x) := -x\log x - (1-x)\log(1-x)$ denote the binary entropy function; for continuity we define $\mathrm{H}_b(0) = \mathrm{H}_b(1) = 0$.

2.1 The communication model

Traditionally, a two-party communication problem (between Alice and Bob, say) is formalised as a function, or partial function, on a domain of the form $X \times Y$, where the finite set X (resp. Y) is the set of Alice's (resp. Bob's) possible inputs. For our purposes, it is helpful to think of the input domain represented differently. We shall think of an input as an m-tuple of *tokens*, where the tokens are given to the players according to a random *allocation* drawn from a known distribution. Thus, it will help to represent the input domain as $X_1 \times X_2 \times \cdots \times X_m$, where X_i is the set of possible values for the i-th token. Typically, each X_i will be either the set $\{0,1\}$ or the set $[N] := \{1,2,\ldots,N\}$, for some positive integer N. An allocation amongst p players is then a function $\sigma : [m] \to 2^{[p]}$.

A natural and interesting special case of an allocation is a *split*, where each token is given to exactly one player selected at random (not necessarily uniformly) from amongst all players. It will be convenient to think of splits as functions $\sigma : [m] \to [p]$. A further special case is that of a *uniform split*, where each token is equally likely to go to each of the players. We let $\mathcal{U}_p$ denote the probability distribution of a uniform split amongst p players.

Definition 2.1 (Communication Problems and Protocols)**.** A *random-allocation communication problem* for p players consists of a function $f : X_1 \times \cdots \times X_m \to Z$ and a probability distribution ν on allocations $\sigma : [m] \to 2^{[p]}$. A traditional communication problem is a special case, where ν is supported on a single allocation (that is typically a split). Protocols, unless explicitly qualified otherwise, are assumed to be randomised, with the players having access to private as well as public coins. (For a formal definition of a "protocol," we refer the reader to a standard textbook, such as Kushilevitz and Nisan [32].) For a random-allocation protocol P, let $P(x,\sigma)$ denote the (possibly random) *transcript* of P, and $\mathrm{out}(P,x,\sigma)$ the output of P, on input x allocated according to σ. For a traditional protocol, where σ has only one possible value, we drop σ from these notations.

Definition 2.2 (Error, Cost, Complexity)**.** Let P be a protocol for a random-allocation communication problem (f,ν). We define the error

$$\mathrm{err}(P,f,\nu) \ := \ \max_x \Pr[\mathrm{out}(P,x,\sigma) \neq f(x)]\,,$$

where the probability is taken over $\sigma \sim \nu$ and the (public and private) coins used by the protocol. If μ is a distribution on the inputs to f, we define the distributional error

$$\mathrm{err}_\mu(P,f,\nu) \ := \ \Pr[\mathrm{out}(P,X,\sigma) \neq f(X)]\,,$$

where $X \sim \mu$ and $\sigma \sim \nu$. Let $\mathrm{cost}(P) := \max |P(x,\sigma)|$ denote the communication cost of P, where this maximum is taken over x, σ, and the random coins of P. We define the δ-error communication complexity of (f,ν) to be

$$\mathrm{R}_\delta(f,\nu) \ := \ \min\{\mathrm{cost}(P) : \mathrm{err}(P,f,\nu) \leq \delta\}$$

and the δ-error μ-distributional complexity to be

$$\mathrm{R}_{\mu,\delta}(f,\nu) \ := \ \min\{\mathrm{cost}(P) : \mathrm{err}_\mu(P,f,\nu) \leq \delta\}\,.$$

Let $\mathrm{R}^{\to}$ and R^k denote the restrictions of these notions to one-way and k-round protocols, respectively (the notion of a "round" will be made precise later, when we use it). For traditional communication problems, where there is a deterministic and well-known input allocation, we drop ν from these notations.

Informally, a communication lower bound is *robust* if it applies to $\mathrm{R}_\delta(f,\nu)$ or $\mathrm{R}_{\mu,\delta}(f,\nu)$ for some high-entropy distribution ν, such as the aforementioned $\mathcal{U}_p$.

2.2 Technique preliminaries

In this section we introduce some of the main techniques that we use to establish our results. These are all based on considering random input in addition to random splits.

The notion of information complexity has been used on many occasions in the study of communication protocols [11, 4, 9, 29]. Loosely speaking, information complexity is used to establish a direct sum result, which reduces the problem of lower bounding the complexity of a "compound" problem (here, disjointness) to that of lower bounding the complexity of a simpler "base" problem (here, the AND function). The direct sum result follows from a *simulation argument*, where we design a protocol for the base problem that randomly pads its input to generate an artificial input for the compound problem and then simulates a protocol for the compound problem. Here, for our robust lower bounds for set disjointness, we need to consider information complexity in a setting that allows both public and private coins. This is a subtle matter: we must condition on the public coin to have a meaningful notion of information complexity. At the same time, we must be careful about how the public coin is used in the simulation argument, ensuring that we do not introduce undesirable correlations in the random padding.

Definition 2.3 (Information cost and complexity)**.** Let P^R be a protocol that uses a public random string R (in addition to any private random strings that players use) and let μ be a distribution on its inputs. We define

$$\mathrm{icost}_\mu(P^R) \ := \ \mathrm{I}(X : P^R(X) \mid R)\,, \quad \text{where} \quad X \sim \mu\,.$$

Let D be a random variable, possibly correlated with X, but independent from R and any private randomness used in P. We define the D-conditional μ-information cost

$$\mathrm{icost}_\mu(P^R \mid D) \ := \ \mathrm{I}(X : P^R(X) \mid D, R)\,.$$

For each information cost measure above, we define a corresponding information complexity measure in the natural way, e. g., for a communication problem f,

$$\mathrm{IC}_{\mu,\delta}(f) \ := \ \inf\{\mathrm{icost}_\mu(P) : \mathrm{err}(P,f) \le \delta\}\,,$$

where P ranges over protocols that are allowed both private and public coins.

We also consider random inputs $X \sim \mu$ in another setting. Some of our lower bounds will use a reduction from a communication problem in the fixed-partition model to one where the allocation $\sigma \sim \nu$. In these reductions, the players choose σ using public random bits, but then distributing the input tokens according to σ would seem to necessitate communicating a large fraction of the data and this would render the reduction useless. The solution is to use distributional lower bounds on fixed-partition problems. This suggests that the players may "guess" data that they do not know. Unfortunately, the issue that arises is that this guessing may be correlated to the distribution of σ. However, the following lemma connects us back to the "usual" situation, when inputs and allocations are independent of each other, provided this correlation is sufficiently weak.

Lemma 2.4. *If a protocol P satisfies $\Pr_{(x,\sigma)\sim\xi}[\mathrm{out}(P,x,\sigma) \ne f(x)] \le \delta$, for some joint distribution ξ, then for all input distributions μ and allocation distributions ν,*

$$\mathrm{err}_\mu(P,f,\nu) \ \le \ \delta + \mathrm{D_{TV}}(\mu \otimes \nu, \xi)\,.$$

Proof. Simply observe that

$$\mathrm{err}_\mu(P,f,\nu) \;=\; \Pr_{\substack{x\sim\mu,\\ \sigma\sim\nu}}[\mathrm{out}(P,x,\sigma)\neq f(x)] \;\le\; \Pr_{(x,\sigma)\sim\xi}[\mathrm{out}(P,x,\sigma)\neq f(x)] + \mathrm{D_{TV}}\left(\mu\otimes\nu,\xi\right). \qquad \square$$

2.3 Preliminary lemmas

We collect together a few basic results that we appeal to at various points in the paper. The first result is a sharp lower bound on the communication complexity of the INDEX problem. In this problem, Alice holds a string $x\in\{0,1\}^n$ and Bob holds $j\in[n]$. The goal is for Bob to learn x_j. See, e. g., Ablayev [1] for a proof of the following result.

Lemma 2.5. *Let $n>0$ be an integer and $\delta\in(0,1/2)$ be a real number. Then we have*

$$\mathrm{R}_\delta^{\rightarrow}(\mathrm{INDEX}) \ge (1-\mathrm{H}_b(\delta))n\,.$$

The second result is a well-known calculation giving a pair of probability bounds about what is often called the "birthday problem."

Lemma 2.6. *For $t,p\in\mathbb{N}$, let $\alpha(t,p)$ denote the probability that t independent random variables, each drawn uniformly from $[p]$, do not take t distinct values. Then*

$$1-e^{-t(t-1)/(2p)} \;\le\; \alpha(t,p) \;=\; 1-\prod_{i=1}^{t-1}\left(1-\frac{i}{p}\right) \;\le\; \frac{t(t-1)}{2p}\,.$$

The third result upper bounds the total variation distance between binomial distributions with similar parameters. The proof of this lemma is presented in Appendix A.

Lemma 2.7. *There exists a constant $c_1>0$ such that for all $q\in[1/2,1)$, $r\in(0,1)$, and $a,w\in\mathbb{N}$,*

$$\frac{w}{\sqrt{v}}\le r \quad\Longrightarrow\quad \mathrm{D_{TV}}\left(\mathcal{B}(a,q),\mathcal{B}(a-w,q)\right) \;\le\; c_1 r\sqrt{\ln\frac{2}{r}}\,,$$

where $v=aq(1-q)$ is the variance of $\mathcal{B}(a,q)$. In order to define the total variation distance above, we treat the binomial distribution $\mathcal{B}(n,p)$ as a distribution on the set of all non-negative integers, rather than just $\{0,1,\dots,n\}$.

3 Multi-party set disjointness

Let $\mathrm{DISJ}_{n,t}:\{0,1\}^{nt}\to\{0,1\}$ denote the following problem. The input is an (nt)-tuple of bits denoted $\{x_{ij}\}_{i\in[t],j\in[n]}$, to be thought of as the entries of a $t\times n$ Boolean matrix. The input satisfies a *unique intersection promise*, namely, each column of the matrix has weight in $\{0,1,t\}$ and at most one column has weight t. The desired output is $\bigvee_{j=1}^n\bigwedge_{i=1}^t x_{ij}$. Gronemeier [20] culminated a line of work [3, 4, 9] on this problem, showing that $\mathrm{R}_\delta(\mathrm{DISJ}_{n,t})=\Omega(n/t)$ under a t-player split where each player receives one row of the matrix.

Let $\text{AND}_t : \{0,1\}^t \to \{0,1\}$ be shorthand for $\text{DISJ}_{1,t}$. That is, the input is a t-tuple of bits $x = (x_1,\dots,x_t)$ that satisfies the promise $|x| \in \{0,1,t\}$. The desired output is $\bigwedge_{i=1}^t x_i$. Let $D \in_R [t]$ and $X \in_R \{0, \mathbf{e}_D\}$. Denote the resulting joint distribution of (X,D) by λ and the marginal distribution of X by μ. The lower bound of [20] follows by carefully analysing $\text{IC}_{\mu,\delta}(\text{AND}_t \mid D)$ and using the direct sum techniques of Bar-Yossef et al. [4] to link this quantity with $\text{IC}_{\mu^n,\delta}(\text{DISJ}_{n,t} \mid D^n)$.

Here, we consider the random-allocation communication problem $(\text{DISJ}_{n,t}, \mathcal{U}_p)$ for some suitably large number, p, of players. We now prove a robust lower bound on its complexity by extending the earlier techniques.

Lemma 3.1. *Let $\delta' = \delta + \alpha(t,p)$. Then*

$$\text{R}_\delta(\text{DISJ}_{n,t}, \mathcal{U}_p) \;\ge\; n \cdot \text{IC}_{\mu,\delta'}(\text{AND}_t \mid D)\,.$$

Proof. Let P^R be a minimum-cost δ-error protocol for $(\text{DISJ}_{n,t}, \mathcal{U}_p)$ that uses a public random string R, possibly in addition to private randomness. Then $\text{cost}(P^R) = \text{R}_\delta(\text{DISJ}_{n,t}, \mathcal{U}_p)$. Consider n independent pairs of random variables $(X_1,D_1),\dots,(X_n,D_n)$, each drawn from λ. Then $X := X_1X_2\dots X_n \sim \mu^n$ is a suitable random input for $\text{DISJ}_{n,t}$. Let $S \sim \mathcal{U}_p$ be a random split. Then, by standard information theoretic arguments, we have

$$\begin{aligned}
\text{cost}(P^R) &= \max_{x,\sigma} |P^R(x,\sigma)| \;\ge\; \text{H}(P^R(X,S)) \\
&\ge \text{I}(X : P^R(X,S) \mid D_1D_2\dots D_n, R, S) \\
&\ge \sum_{j\in[n]} \text{I}(X_j : P^R(X,S) \mid D_1D_2\dots D_n, R, S) \qquad (3.1) \\
&= \sum_{j\in[n]} \mathbb{E}_d[\text{I}(X_j : P^R(X,S) \mid D_j, R, S, D_{-j} = d)]\,,
\end{aligned}$$

where (3.1) holds because the X_js are independent even after conditioning on $D_1D_2\dots D_n, R$, and S. Here, D_{-j} denotes the vector $(D_1,\dots,D_{j-1},D_{j+1},\dots,D_n)$ and the final expectation is over d drawn uniformly from $[t]^{[n]\setminus\{j\}}$. To finish the proof, it suffices to show that

$$c_{j,d} \;:=\; \text{I}(X_j : P^R(X,S) \mid D_j, R, S, D_{-j} = d) \;\ge\; \text{IC}_{\mu,\delta'}(\text{AND}_t \mid D)\,,$$

for each $j \in [n]$ and each $d \in [t]^{[n]\setminus\{j\}}$. To this end, we shall design a certain δ'-error t-party traditional protocol $Q^{R,S}_{j,d}$ for AND_t, parametrised by j and d, that uses (R,S) as a public random string. Further, for each possible value (ρ,σ) of (R,S), the transcript $Q^{\rho,\sigma}_{j,d}(X_j)$ will either be constant or be distributed identically to $(P^R(X,\sigma) \mid R = \rho, D_{-j} = d)$, and the players will know which case they are in based on σ alone. Then, as required, we shall have

$$\text{IC}_{\mu,\delta'}(\text{AND}_t \mid D) \;\le\; \text{icost}_\mu(Q^{R,S}_{j,d} \mid D_j) \;=\; \text{I}(X_j : Q^{R,S}_{j,d}(X_j) \mid D_j, R, S) \;\le\; c_{j,d}\,.$$

The protocol $Q^{\rho,\sigma}_{j,d}$ works as follows. On input $x = (x_1,\dots,x_t) \in \{0,1\}^t$, the t players create a random virtual input $\{Z_{ik}\}_{i,k} \in \{0,1\}^{t\times n}$ for $\text{DISJ}_{n,t}$, pretend that this input has been split according to σ amongst p virtual players, and then, if possible, simulate the behaviour of these virtual players when they execute

P^ρ on the virtual input. The virtual input is obtained by embedding x into the j-th column of a random Boolean matrix drawn from $(\mu^n|D_{-j} = d)$. To wit:

$$Z_{ik} \in_R \begin{cases} \{x_i\}, & \text{if } k = j, \\ \{0\}, & \text{if } k \neq j \text{ and } d(k) \neq i, \\ \{0,1\}, & \text{if } k \neq j \text{ and } d(k) = i. \end{cases}$$

Therefore, the simulation is possible iff σ assigns each of the inputs $(Z_{1j}, \ldots, Z_{tj})$ to a distinct virtual player; we shall say that σ *ramifies* if this condition is met. If σ does not ramify, the protocol ends immediately (note that all players know σ so this happens without any communication), leading to a constant empty transcript and an error probability of 1. If σ does ramify, then Player i plays the role of that virtual player who is assigned Z_{ij} by σ. The crucial observation that makes this role-playing possible is that all the *other* bits assigned to that virtual player are available to Player i, because they are either set to 0 or can be drawn uniformly at random from $\{0,1\}$ using Player i's private coin. All virtual players who are not assigned any of the inputs $\{Z_{ij}\}_{i \in [t]}$ are simulated by Player 1 (say). Thus, if σ ramifies, then

$$Q_{j,d}^{\rho;\sigma}(X_j) \equiv (P^R(X,\sigma) \mid R = \rho, D_{-j} = d).$$

Finally, $Q_{j,d}^{R,S}$ is indeed a δ'-error protocol, because

$$\mathrm{err}(Q_{j,d}^{R,S}, \text{AND}_t) \;\le\; \Pr[\sigma \text{ does not ramify}] + \mathrm{err}(P^R, \text{DISJ}_{n,t}, \mathcal{U}_p) \;=\; \alpha(t,p) + \delta \;=\; \delta'. \qquad \square$$

Lemma 3.2. *There exists a constant $c > 0$ such that, for all $\delta \in (0, 1/10)$ and $t \ge 2$,*

$$\mathrm{IC}_{\mu,\delta}(\text{AND}_t \mid D) \ge c/t.$$

Proof. This result can *almost* be deduced from Gronemeier [20], except for the subtlety introduced by public coins. Specifically, from the work of Gronemeier we can deduce that for a *private* coin traditional protocol P such that $\mathrm{err}(P, \text{AND}_t) \le 1/10$, we have $\mathrm{icost}_\mu(P \mid D) = \Omega(1/t)$.

To complete the proof, we show that public coins cannot help reduce information complexity.[1] Consider a general δ-error protocol Q^S for AND_t that uses a public random string S (recall that Definition 2.1 allows such a protocol to use both public and private coins). Let P be a private coin protocol in which Player 1 generates S privately and announces it to all players, following which the players simulate Q^S. Clearly, $\mathrm{err}(P, \text{AND}_t) = \mathrm{err}(Q^S, \text{AND}_t)$. The transcript of P on input X is precisely $(S, Q^S(X))$. Thus,

$$\mathrm{icost}_\mu(P \mid D) = \mathrm{I}(X : S, Q^S(X) \mid D) = \mathrm{I}(X : S \mid D) + \mathrm{I}(X : Q^S(X) \mid D, S) = \mathrm{icost}_\mu(Q^S \mid D),$$

where the second step uses the chain rule and the third step uses the independence of S from (X,D). Combining this with the private-coin lower bound completes the proof. $\square$

Putting together Lemmas 2.6, 3.1 and 3.2 yields the following theorem.

[1] This observation is folklore but we have included a proof for the sake of completeness. Note that this situation is in contrast to standard communication complexity, where the public-coin complexity *could* be smaller than the private-coin complexity.

Theorem 3.3. *For all $\delta \in (0, 1/20)$, $t = t(n) \geq 2$, and $p \geq 10t^2$, we have the robust lower bound*

$$\mathrm{R}_\delta(\mathrm{DISJ}_{n,t}, \mathcal{U}_p) = \Omega(n/t) .$$

We note that in order to get this kind of robust lower bound for $\mathrm{DISJ}_{n,t}$ under $\mathcal{U}_p$ that increases linearly with n, we *must* make p, the number of players, as large as $\Omega(t^2)$ for constant δ. This is because when an input x such that $\mathrm{DISJ}_{n,t}(x) = 1$ is allocated to p players, with probability $\alpha(t, p)$ there exists a player that receives at least two tokens from the all-ones column. Therefore, a simple $O(p)$-communication protocol, where each player announces whether or not they have received two 1s from the same column, has error probability at most $1 - \alpha(t, p)$. By Lemma 2.6, we now have $\mathrm{R}_\delta(\mathrm{DISJ}_{n,t}, \mathcal{U}_p) = O(p)$ for $p \leq t(t-1)/(2\ln(1/\delta)) = O(t^2)$.

4 Pointer jumping and selection

We now consider the *tree pointer jumping* problem $\mathrm{TPJ}_{k,t}$, defined as follows. (In reading this section, it will help to think of t as growing and k as fixed.)

Definition 4.1 (The tree pointer jumping function)**.** Consider a complete k-level t-ary tree, T, rooted at v_0. We use the convention that the leaves are at level 1 and the root at level k. The input is a function $\phi : V(T) \to [t]$, with $\phi(v) \in \{0, 1\}$ if v is a leaf of T. We shall call such an input a "k-input" and shall sometimes view it as a labelling of $V(T)$. Define $g(v)$ to be the $\phi(v)$-th child of v if v is an internal node, and $\phi(v)$ if v is a leaf. The desired output is $\mathrm{TPJ}_{k,t}(\phi) := g^{(k)}(v_0) = g(g(\cdots g(v_0)\cdots)) \in \{0, 1\}$.

There are at least two natural ways to make a traditional communication problem out of $\mathrm{TPJ}_{k,t}$, both of which are of interest to us. The first way is to have two players, Alice and Bob, with Alice (resp. Bob) receiving the values of $\phi(v)$ for odd-level (resp. even-level) vertices v; recall that leaves are at level 1. The second way is to have k players, with Player i receiving the values of $\phi(v)$ for vertices v on level i. When speaking of communication problems, we shall use $\mathrm{TPJ}_{k,t}$ to denote the former, and $\mathrm{M\text{-}TPJ}_{k,t}$ to denote the latter ("M" for "multi-player"). For $k = 2$, the two definitions coincide and we obtain the well-studied INDEX problem, for which strong one-way lower bounds are known [1], with numerous implications for stream computation. In particular, Guha and McGregor [23] use a reduction from INDEX to obtain a tight (up to logarithmic factors) space lower bound for estimating the median of a randomly ordered stream of numbers in one pass. This lower bound was subsequently extended to multiple passes by Chakrabarti, Jayram and Pătraşcu [8] via a rather different (and intricate) proof.

As a consequence of the robust communication lower bounds we prove in this section, we obtain a considerably simpler and improved multi-pass streaming lower bound for median finding.[2] The five theorems in this section can be organised into two parallel chains of implications, each consisting of three stages and culminating in a lower bound for the MEDIAN problem, as follows.

Stage 1: We prove a multi-round lower bound on the communication complexity of an appropriate "source problem," which is either $\mathrm{M\text{-}TPJ}_{k,t}$, as in Theorem 4.4 or $\mathrm{TPJ}_{k,t}$, as in Theorem 4.9.

[2]Our results, like the earlier ones [23, 8], apply to the more general problem of selection.

Stage 2: We reduce the source problem to an intermediate problem that we call *weight-based tree pointer jumping*, or $\text{W-TPJ}_{k,n}$, defined below. At this stage, we have a *robust* lower bound for $\text{W-TPJ}_{k,n}$, under an allocation distribution that depends on the source problem we started with. These reductions appear as Theorems 4.8 and 4.10 below.

Stage 3: Finally, we reduce $\text{W-TPJ}_{k,n}$ to the MEDIAN problem, as in Theorem 4.3, obtaining a robust lower bound for the latter. This reduction does not depend on the choice of the source problem.

The precise notion of a "round" is crucial here, and is different for the two parallel chains of implications. When using the two-player problem $\text{TPJ}_{k,t}$ as the source, a round consists of a single message, from either Alice or Bob. The player that does not know $\phi(v_0)$ speaks first. When using the multi-player problem $\text{M-TPJ}_{k,t}$ as the source, a round consists of one message from each of the k players, speaking in the fixed order Player 1, ..., Player k (recall that Player 1 holds the labels of the leaf nodes).

Definition 4.2 (Cost and Complexity, Multi-Round)**.** Fix one of the two notions of a "round," as described above. We define the notations $\mathrm{R}^k_{\mu,\delta}(f,\nu)$, etc., as in Definition 2.2, with protocols restricted to k rounds. The cost of a round is the maximum possible *total* number of bits communicated by the players who speak in that round. The cost of a protocol is the *maximum* cost of a single round.

The next three subsections are organised thus. We first present the Stage 3 reduction, then the Stage 1 and Stage 2 theorems for the implication chain that starts with $\text{M-TPJ}_{k,t}$, and then deal with the chain that starts from $\text{TPJ}_{k,t}$. We choose to present the M-TPJ chain first, and in greater detail, because it ultimately implies stronger lower bounds for data stream computation. Furthermore, the Stage 1 theorem in this chain (Theorem 4.4) is a fundamental and interesting result in communication complexity in its own right that, to the best of our knowledge, has not been proven before.

4.1 Weight-based TPJ and a reduction to selection

We now define the problem $\text{W-TPJ}_{k,n}$ mentioned above. It is closely related to $\text{TPJ}_{k,t}$ and $\text{M-TPJ}_{k,t}$ (with n determined by k and t); as before, the input can be thought of as a labelling of a complete k-level t-ary tree. However, the labels are presented differently: instead of specifying $\phi(v)$ directly, the input specifies a binary string $x_v \in \{0,1\}^{a_i}$ for each level-i node of T, where the lengths a_i are parameters to be fixed later, and the Hamming weight of x_v implicitly determines $\phi(v)$. If v is a leaf ($i = 1$), then $a_i = 1$ and $\phi(v) = x_v = |x_v|$. Otherwise, $|x_v|$ uniquely determines $\phi(v)$ via the following equation.

$$|x_v| \;=\; \frac{a_i}{2} + \left(\phi(v) - \frac{t+1}{2}\right) b_{i-1}\,, \tag{4.1}$$

where b_i is the total length of all strings associated with nodes in the subtree rooted at a level-i node, i. e., $b_i = a_i + t b_{i-1}$ and $b_1 = 1$. We will only define $\text{W-TPJ}_{k,n}$ on inputs such that each $|x_v|$ determines a value $\phi(v)$ in the range $\{1,\ldots,t\}$. In particular, each a_i will need to be "large enough" so that Equation (4.1) is feasible. Let $x \in \{0,1\}^n$ be the concatenation of all the strings x_v. We then define the partial function $\text{W-TPJ}_{k,n}(x) := \text{TPJ}_{k,t}(\phi)$, where ϕ is determined by x as just described.

The next theorem completes Stage 3 in the above proof outline. The reduction from W-TPJ to MEDIAN used in its proof is along similar lines to one by Guha and McGregor [23].

Theorem 4.3. *Let* $\text{MEDIAN}_{m,N}$ *denote the random allocation communication problem where the input consists of m tokens* $(x_1,\ldots,x_m) \in [N]^m$ *and the desired output is the median of this collection of tokens. For any* $\delta > 0$*, any allocation distribution* ν*, and any number* $p \geq 1$ *of rounds of communication, we have*

$$\mathrm{R}^p_\delta(\text{MEDIAN}_{n,\Theta(n)},\nu) \geq \mathrm{R}^p_\delta(\text{W-TPJ}_{p+1,n},\nu)\,.$$

Proof. Let $k = p+1$. We reduce W-TPJ to MEDIAN. Let T be a complete k-level t-ary tree as usual, and let $x = \{x_v\}_{v\in V(T)}$ be an input to $\text{W-TPJ}_{k,n}$. Our reduction will associate a pair of integers $(\alpha(v),\beta(v))$ with each $v \in V(T)$ such that the following properties are satisfied.

1. For each leaf v, we have $\alpha(v) \equiv 0 \pmod 2$ and $\beta(v) \equiv 1 \pmod 2$.
2. For each strict descendant v of each internal node u, we have $\alpha(u) < \alpha(v) < \beta(v) < \beta(u)$.
3. If v_i and v_j are the i-th and j-th children of u, with $i < j$, then $\beta(v_i) < \alpha(v_j)$.

Further, it will associate a multiset $A(v)$ with each $v \in V(T)$ as follows. If v is a level-i node, then $A(v)$ consists of $a_i - |x_v|$ copies of $\alpha(v)$ and $|x_v|$ copies of $\beta(v)$. The properties above, together with Equation (4.1), ensure that

$$\text{W-TPJ}(x) \;=\; \mathrm{median}\left(\bigcup_{v\in V(T)} A(v)\right) \bmod 2\,;$$

this can be justified by a straightforward induction on k. The reduction itself works by having each player generate one element of $\bigcup_{v\in V(T)} A(v)$ per bit of x allocated to her. This is done in the natural way: if the bit in question corresponds to a node v, then she generates the element $\alpha(v)$ if the bit's value is 0 and $\beta(v)$ if the bit's value is 1.

It remains to demonstrate that suitable values $(\alpha(v),\beta(v))$ satisfying the above properties exist. Here is an explicit construction. We use the notation $v[i_k,\ldots,i_j]$ to denote the i_j-th child of $v[i_k,\ldots,i_{j-1}]$, with $v[\,]$ being the root of T. Set $B = 2\lceil (t+2)/2\rceil$ and let $\langle h_k,h_{k-1},\ldots,h_1\rangle_B$ denote the quantity $\sum_{i=1}^k B^{i-1}h_i$, i. e., a base-B representation. We now set

$$\alpha(v) = \langle i_k,\ldots,i_{j+1},0,0,\ldots,0\rangle_B \quad\text{and}\quad \beta(v) = \langle i_k,\ldots,i_{j+1},t+1,0,\ldots 0\rangle_B\,,$$

for each internal node $v = v[i_k,\ldots,i_{j+1}]$ at level j. For each leaf node $v = v[i_k,\ldots,i_2]$, let

$$\alpha(v) = \langle i_k,\ldots,i_2,0\rangle_B \quad\text{and}\quad \beta(v) = \langle i_k,\ldots,i_2,1\rangle_B\,.$$

One can easily verify that this construction has the properties claimed. □

4.2 A robust multi-player lower bound

We now fill in Stages 1 and 2 of our proof outline, using M-TPJ as our source problem, and deriving a robust lower bound for W-TPJ. Both problems involve $(p+1)$ players, for $p \geq 1$. Recall that, in this case, a "round" consists of one message from each player, in the order Player 1, ..., Player $(p+1)$. We start by obtaining the following traditional (i. e., "fragile") bounded-round lower bound for M-TPJ.

Theorem 4.4. *Let μ_k denote the uniform distribution over k-inputs (as introduced in Definition 4.1). Then, for $p = p(t) \geq 1$, we have*

$$\mathrm{R}^{p}_{\mu_{p+1},1/3}(\text{M-TPJ}_{p+1,t}) = \Omega(t/p^2)\,.$$

To prove this, we define an appropriate notion of information cost that is concerned only with the information revealed in the *first round* of a multi-round protocol's execution. We then use this notion to establish an appropriate *round elimination lemma*, à la Miltersen et al. [34] and Sen [38], which in turn implies the above theorem.

Definition 4.5 (First-round information cost)**.** Let P be a multi-round, multi-player, private-coin protocol and μ an input distribution for P. Let $P^1(x,R)$ denote the concatenation of all messages sent by the players during the *first round* of P, where R denotes the concatenation of the random strings used by the players. Then, we define the first round μ-information cost of P as follows.

$$\mathrm{icost}^1_{\mu}(P) \;=\; \mathrm{I}(X : P^1(X,R))\,, \quad \text{where} \quad X \sim \mu\,.$$

As a precursor to our round elimination lemma, we prove the following multi-round analogue of a lemma of Sen [38, Lemma 1].

Lemma 4.6 (Uninformative round lemma)**.** *Suppose a k-input Boolean function f has an r-round k-player private-coin protocol P, in which each round costs at most C. Then, for any input distribution μ, f has an $(r-1)$-round k-player deterministic protocol Q such that*

$$\mathrm{err}_{\mu}(Q,f) \;\leq\; \mathrm{err}_{\mu}(P,f) + \sqrt{\frac{\ln 2}{2}\cdot \mathrm{icost}^1_{\mu}(P)} \;\leq\; \mathrm{err}_{\mu}(P,f) + \sqrt{\mathrm{icost}^1_{\mu}(P)}\,,$$

and where each round costs at most C.

Proof. Without loss of generality, we may assume that each player uses two *independent* random strings in P: one to generate his first-round message, and another to generate all subsequent messages.[3] We proceed under this assumption. Let Q_m denote the $(r-1)$-round protocol obtained by fixing the first round's communication in P to m. Define the function g by

$$g(x,m) \;=\; \Pr[\mathrm{out}(Q_m,x) \neq f(x)]\,, \tag{4.2}$$

where the probability is over the collection of second random strings used the players.

Define random variables X and M, where $X \sim \mu$, and M is the first-round communication generated from X according to P; let λ denote the resulting joint distribution of (X,M). Let β denote the distribution of M. We then have

$$\mathrm{D_{TV}}(\lambda, \mu\otimes\beta) \;\leq\; \sqrt{\frac{\ln 2}{2}\cdot \mathrm{D_{KL}}(\lambda\|\mu\otimes\beta)} \;=\; \sqrt{\frac{\ln 2}{2}\cdot \mathrm{I}(X:M)} \;=\; \sqrt{\frac{\ln 2}{2}\cdot \mathrm{icost}^1_{\mu}(P)}\,, \tag{4.3}$$

[3] To see why, consider a particular player who uses a random string R to generate his messages, the first such message being M. This player can instead draw two independent random strings R and R', using R to generate M, and then R' to draw from the conditional distribution $R \mid M$. Finally, he can use R' in place of R while generating all his remaining messages. It is easy to see that the distribution of messages so generated is identical to that in the original protocol.

where the first two steps are basic information theory (the inequality is often credited to Pinsker).

We can express the distributional errors of P and Q_m in terms of g, by averaging Equation (4.2) in two ways:

$$\mathrm{err}_\mu(P,f) = \mathbb{E}_{(X,M)\sim\lambda}[g(X,M)]; \quad \mathrm{err}_\mu(Q_m,f) = \mathbb{E}_{X\sim\mu}[g(X,m)].$$

Thus, we have

$$\begin{aligned}
\mathbb{E}_{m\sim\beta}[\mathrm{err}_\mu(Q_m,f)] &= \mathbb{E}_{(X,M)\sim\mu\otimes\beta}[g(X,M)] \\
&\le \mathbb{E}_{(X,M)\sim\lambda}[g(X,M)] + \mathrm{D_{TV}}(\lambda,\mu\otimes\beta) \\
&\le \mathrm{err}_\mu(P,f) + \sqrt{\frac{\ln 2}{2}\cdot \mathrm{icost}^1_\mu(P)}\,,
\end{aligned}$$

where the first inequality holds because $|g(x,m)| \le 1$ for all x and m, and the second inequality uses Equation (4.3). Choose m to minimise $\mathrm{err}_\mu(Q_m,f)$, and fix the random strings used by the players in Q_m so as to minimise the μ-distributional error of the resulting deterministic protocol, Q. Then $\mathrm{err}_\mu(Q,f)$ is upper-bounded as desired. □

Lemma 4.7 (Round elimination for M-TPJ)**.** *Let $p \ge 2$ be an integer, let K and ε be positive reals. Let μ_k denote the uniform distribution over k-inputs. Let $\mathcal{A}(p,K,\varepsilon)$ denote the statement "M-TPJ$_{p+1,t}$ has a deterministic p-round protocol in which each round uses at most t/K^2 bits of communication in total, and whose distributional error under μ_{p+1} is at most ε." Then $\mathcal{A}(p,K,\varepsilon) \Rightarrow \mathcal{A}(p-1,K,\varepsilon+1/K)$.*

Proof. Let P be a protocol whose existence is asserted by $\mathcal{A}(p,K,\varepsilon)$. Based on P, we shall construct t private-coin protocols $Q_1,\ldots,Q_t$, each for M-TPJ$_{p,t}$. Let T be a $(p+1)$-level t-ary tree, and let $T_1,\ldots,T_t$ denote the p-level subtrees hanging off the root, v_0. Recall, from Definition 4.1 that a $(p+1)$-input can be thought of as a function from $V(T)$ to $[t]$, or equivalently, as a labelling of $V(T)$ using labels from $[t]$. Given a p-input ϕ and an integer $i \in [t]$, let $\phi^{(i)}$ denote the random $(p+1)$-input obtained as follows. Treat ϕ as a function from $V(T_i)$ to $[t]$. Choose independent random inputs $\psi_j : V(T_j) \to [t]$, for $j \in [t]\setminus\{i\}$, each distributed according to μ_p. Then put

$$\phi^{(i)}(v) = \begin{cases} i\,, & \text{if } v = v_0\,, \\ \phi(v)\,, & \text{if } v \in V(T_i)\,, \\ \psi_j(v)\,, & \text{if } v \in V(T_j) \text{ where } j \ne i\,. \end{cases}$$

Let ξ_i denote the distribution of $\Phi^{(i)}$, where $\Phi \sim \mu_p$. Notice that ξ_i is identical to μ_{p+1} conditioned on the label of the root being i.

Here is how the protocol Q_i works. On input ϕ, the players use private randomness to construct $\phi^{(i)}$ (note that this is possible because of an appropriate product structure of $\phi^{(i)}$), and then simulate P on this input, using a virtual "Player $p+1$," who can be locally simulated by each real player, because his input, i, is common knowledge. Clearly, Q_i only errs when its call to P errs. Therefore, we have

$$\frac{1}{t}\sum_{i=1}^{t} \mathrm{err}_{\mu_p}(Q_i, \text{M-TPJ}_{p,t}) = \frac{1}{t}\sum_{i=1}^{t} \mathrm{err}_{\xi_i}(P, \text{M-TPJ}_{p+1,t}) = \mathrm{err}_{\mu_{p+1}}(P, \text{M-TPJ}_{p+1,t}) \le \varepsilon\,. \tag{4.4}$$

Let M denote the concatenation of the messages generated *in the first round* by Players $1,\ldots,p$ when the protocol P runs on input $X \sim \mu_{p+1}$, defined on the tree T. For $i \in [t]$, let X_i denote the portion of X that corresponds to the labelling of the subtree T_i. Then we have

$$\frac{t}{K^2} \ \geq\ |M| \ \geq\ \mathrm{I}(X:M) \ \geq\ \sum_{i=1}^{t} \mathrm{I}(X_i:M) \ =\ \sum_{i=1}^{t} \mathrm{icost}^1_{\mu_p}(Q_i)\,, \tag{4.5}$$

where the rightmost inequality uses the independence of $\{X_i\}_{i\in[t]}$. Combining (4.4) and (4.5), we have

$$\frac{1}{t}\sum_{i=1}^{t} \mathrm{err}_{\mu_p}(Q_i, \text{M-TPJ}_{p,t}) + \sqrt{\frac{1}{t}\sum_{i=1}^{t} \mathrm{icost}^1_{\mu_p}(Q_i)} \ \leq\ \varepsilon + \frac{1}{K}\,.$$

Using the concavity of the square root function, plus an averaging argument, we now conclude that

$$\exists\, i \in [t]:\ \mathrm{err}_{\mu_p}(Q_i, \text{M-TPJ}_{p,t}) + \sqrt{\mathrm{icost}^1_{\mu_p}(Q_i)} \ \leq\ \varepsilon + \frac{1}{K}\,.$$

Applying Lemma 4.6 to this particular Q_i gives us the desired protocol, thereby establishing the truth of $\mathcal{A}(p-1, K, \varepsilon + 1/K)$. □

We now have the tools we need to prove Theorem 4.4.

Proof of Theorem 4.4. Suppose that

$$\mathrm{R}^p_{\mu_{p+1},1/3}(\text{M-TPJ}_{p+1,t})$$

is not lower bounded as stated. Specifically, using a standard error-reduction argument, we may assume that

$$\mathrm{R}^p_{\mu_{p+1},1/6}(\text{M-TPJ}_{p+1,t}) \leq t/(6p)^2\,.$$

By the easy direction of Yao's minimax lemma, we have $\mathcal{A}(p, 6p, 1/6)$, where the predicate $\mathcal{A}$ is as defined in Lemma 4.7. Applying that lemma repeatedly, we conclude $\mathcal{A}(1, 6p, 1/6 + (p-1)/6p)$, which implies $\mathcal{A}(1, 6p, 1/3)$. Notice that $\text{M-TPJ}_{2,t}$ is just the INDEX problem with a t-bit input. We have just shown that this problem has a one-round protocol with error at most $1/3$ under the uniform distribution and communication cost at most $t/(6p)^2 \leq t/36$. Since $\mathrm{H}_b(1/3) < 12/13$, this contradicts Lemma 2.5. □

Now that we have the desired Stage 1 lower bound, we move on to Stage 2, proving the following robust lower bound. In our proof, we use a reduction from TPJ that introduces a slight correlation between input and split, and then appeal to Lemma 2.4 to correct for this.

Theorem 4.8. *Let $p = p(n) \geq 1$ and let $\mathcal{V}_{p+1}$ be the (non-uniform) split distribution that gives each token to Player 1 with probability $1/2$ and to Player i with probability $\gamma := 1/(2p)$ for each $i \in \{2,\ldots,p+1\}$. Then*

$$\mathrm{R}^p_{1/24}(\text{W-TPJ}_{p+1,n}, \mathcal{V}_{p+1}) \ =\ \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \Big/ q(n,p)\right), \quad \textit{where} \quad q(n,p) = p^2\left(cp^3 \log n\right)^{\frac{2^p-1}{(p-1)2^{p+1}+2}}$$

for some large constant c. Note that $q(n,p) = \mathrm{polylog}(n)$ *for constant p.*

Thus, for every constant $\varepsilon > 0$*, when p is large enough, we have*

$$\mathrm{R}^{p}_{1/24}(\textsc{w-tpj}_{p+1,n}, \mathcal{V}_{p+1}) \;=\; \Omega\!\left(n^{(2+\varepsilon)^{-p}}\right).$$

Proof. Let P be a protocol for $(\textsc{w-tpj}, \mathcal{V}_{p+1})$ such that $\mathrm{err}(P, \textsc{w-tpj}, \mathcal{V}_{p+1}) \le 1/24$. We will use P to construct a protocol Q for M-TPJ such that $\mathrm{err}_\mu(Q, \textsc{m-tpj}_{p+1,t}) \le 1/3$, where μ is an arbitrary distribution with the property that, for an instance $\phi \sim \mu$, we have $\phi(v) \in_R \{0,1\}$ for each leaf node v. Note that, in particular, this will imply $\mathrm{err}_{\mu_{p+1}}(Q, \textsc{m-tpj}_{p+1,t}) \le 1/3$. The result will then follow by invoking Theorem 4.4.

In Q, the players first use public randomness to transform an input ϕ for M-TPJ into an input x for W-TPJ together with a random split of its tokens. They then proceed to simulate P on this instance. Recall the notation a_i and b_i from the start of Section 4.1. We set

$$a_i \;:=\; (cp^3 t^{2(p+2)} \log n)^{2^{i-1}-1} t^{-2(3\cdot 2^{i-1}-i-2)} \tag{4.6}$$

for some large constant c to be determined. For each node v, the players use the following *public coin* randomised procedure to determine a bit string x_v and an allocation of its bits to the players in P.

If v is an internal node at level i: Choose random integers

$$d_{1v} \sim \mathcal{B}\left(\frac{a_i}{2}, 1-\gamma\right) \quad \text{and} \quad d_{0v} \sim \mathcal{B}\left(\frac{a_i}{2}, 1-\gamma\right),$$

as well as a set

$$S_v^{-i} \in_R \binom{[a_i]}{d_{1v}+d_{0v}}.$$

Let $S_v^{-i} = S_v^1 \cup \ldots \cup S_v^{i-1} \cup S_v^{i+1} \cup \ldots \cup S_v^{p+1}$ be a random partition where, for each $k \in S_v^{-i}$,

$$\Pr\left[k \in S_v^j\right] \;=\; \begin{cases} \gamma/(1-\gamma), & \text{if } j \neq 1, \\ 1/(2(1-\gamma)), & \text{if } j = 1, \end{cases}$$

and put $S_v^i = [a_i] \setminus S_v^{-i}$. Player j will be allocated the values $\{x_{v,k} : k \in S_v^j\}$. Randomly set d_{1v} of the bits $\{x_{v,k} : k \in S_v^{-i}\}$ to 1 and the remaining d_{0v} bits to 0. Notice that all of this is done without reference to the input ϕ.

Player i uses ϕ to determine a target weight $|x_v|$ for the string x_v, based on Equation (4.1). Notice that many of the bits of x_v have already been fixed by the construction so far. Player i sets the free bits in such a way as to achieve this target weight, i. e., she randomly sets $|x_v| - d_{1v}$ of the bits $\{x_{v,k} : k \in S_v^i\}$ to 1 and the remaining bits to 0. Note that this requires $d_{1v} \le |x_v| \le a_i - d_{0v}$; if this condition fails to hold, the protocol *aborts* and outputs a uniform random bit.

If v is a leaf node: In this case x_v is a single bit. Allocate this bit to a random player, with Player 1 being chosen with probability $1/2$ and every other player being chosen with probability γ. If the bit is allocated to Player 1, she sets $x_v = \phi(v)$. Otherwise, the players set $x_v \in_R \{0,1\}$.

This completes the description of Q. Because $\mathcal{V}_{p+1}$ allocates each token to the first player with probability $1/2$, and ϕ assigns a uniformly random bit to each leaf, we have

$$\Pr[\text{W-TPJ}(x) = \text{TPJ}(\phi) \mid \text{the protocol } Q \text{ does not abort}] \;=\; \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \;=\; \frac{3}{4}\,.$$

It remains to show that the bit string x and the allocation σ generated in the reduction are sufficiently close to being independent. Note that the marginals are correct: we do have $\sigma \sim \mathcal{V}_{p+1}$ and, for each leaf v, the value of x_v is indeed chosen according to a uniform setting of $\phi(v)$. The issue is that the joint distribution is not a product distribution. However, note that had d_{1v} and d_{0v} been chosen according to $\mathcal{B}(|x_v|, 1-\gamma)$ and $\mathcal{B}(a_i - |x_v|, 1-\gamma)$, respectively, then σ and x would have been independent, and furthermore, the protocol would not abort. For each internal node v at level i, let

$$\tilde{A}_v \;:=\; \mathcal{B}\left(\frac{a_i}{2}, 1-\gamma\right), \qquad \tilde{B}_v \;:=\; \mathcal{B}\left(\frac{a_i}{2}, 1-\gamma\right),$$
$$A_v \;:=\; \mathcal{B}\left(|x_v|, 1-\gamma\right), \qquad B_v \;:=\; \mathcal{B}\left(a_i - |x_v|, 1-\gamma\right).$$

Then it suffices to show that the product of the distributions $\tilde{A}_v$ and $\tilde{B}_v$, over all internal nodes v, is sufficiently close to the corresponding product of A_v and B_v. Using Lemma 2.7 with the fact that $||x_v| - a_i/2| \le tb_{i-1}$, we can bound the total variation distance in terms of a_i and b_i as follows,

$$\begin{aligned} \mathrm{D_{TV}}\left(\bigotimes_v(\tilde{A}_v \otimes \tilde{B}_v), \bigotimes_v(A_v \otimes B_v)\right) \;&\le\; \sum_v \mathrm{D_{TV}}\left(\tilde{A}_v, A_v\right) + \sum_v \mathrm{D_{TV}}\left(\tilde{B}_v, B_v\right) \\ &\le\; O(\sqrt{\log n}) \sum_{i=2}^{p+1} \frac{t^{p+2-i} b_{i-1}}{\sqrt{a_i/p}}\;, \end{aligned}$$

where the first inequality follows from the triangle inequality. Noting that $b_{i-1} \le 2a_{i-1}$ and by substituting in the value for a_i, we get

$$\frac{b_{i-1}}{\sqrt{a_i/p}} \;\le\; \frac{2a_{i-1}}{\sqrt{a_i/p}} \;=\; \frac{2\sqrt{p}(cp^3 t^{2(p+2)} \log n)^{2^{i-2}-1} t^{-2(3\cdot 2^{i-2}-i-1)}}{(cp^3 t^{2(p+2)} \log n)^{2^{i-2}-1/2} t^{-(3\cdot 2^{i-1}-i-2)}} \;=\; \frac{2}{\sqrt{c}pt^{p+2-i}\sqrt{\log n}}\,.$$

Therefore,

$$\mathrm{D_{TV}}\left(\bigotimes_v(\tilde{A}_v \otimes \tilde{B}_v), \bigotimes_v(A_v \otimes B_v)\right) \;\le\; O(\sqrt{\log n}) \sum_{i=2}^{p+1} \frac{\sqrt{p}\cdot t^{p+2-i} b_{i-1}}{\sqrt{a_i}} \;=\; O(1) \sum_{i=2}^{p+1} \frac{1}{\sqrt{c}p} \;=\; \frac{O(1)}{\sqrt{c}}$$

and the distance can be made less than $1/24$ for sufficiently large constant c. By Lemma 2.4,

$$\mathrm{err}_\mu(Q, \text{M-TPJ}_{p+1,t}) \;\le\; \frac{1}{4} + \frac{1}{24} + \mathrm{err}(P, \text{W-TPJ}_{p+1,n}, \mathcal{V}_{p+1}) \;\le\; \frac{1}{3}\,.$$

As noted above, this implies the same upper bound on $\mathrm{err}_{\mu_{p+1}}(Q, \text{M-TPJ}_{p+1,t})$. Therefore, by Theorem 4.4,

$$\mathrm{R}^p_{1/24}(\text{W-TPJ}_{p+1,n}, \mathcal{V}_{p+1}) \;=\; \Omega(t/p^2)\,.$$

Note that

$$n \;=\; b_{p+1} \;\le\; 2(cp^3 t^{2(p+2)} \log n)^{2^p-1} t^{-2(3\cdot 2^p - p - 3)} \;=\; 2(cp^3 \log n)^{2^p-1} t^{(p-1)2^{p+1}+2},$$

and hence,

$$t \;=\; \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \Big/ \left(cp^3 \log n\right)^{\frac{2^p-1}{(p-1)2^{p+1}+2}}\right). \qquad \square$$

4.3 A robust two-player lower bound

Finally, we revisit Stages 1 and 2 of our proof outline, this time using the 2-player problem $\mathrm{TPJ}_{k,t}$ as our source problem. Now a "round" consists of one message from either Alice or Bob. The traditional (fragile) lower bound that we need for Stage 1 can be deduced from the work of Klauck et al. [31], who in fact studied the problem in the more general *quantum* communication setting. The underlying intuition is, once again, round elimination.

Theorem 4.9. *For $p = p(t) \geq 1$, we have*

$$\mathrm{R}^{p}_{\mu_{p+1},1/3}(\mathrm{TPJ}_{p+1,t}) = \Omega(t/p^2)\,,$$

where μ_k is the uniform distribution over k-inputs (as introduced in Definition 4.1). □

For Stage 2, we obtain the following robust lower bound for W-TPJ, using a proof that closely parallels that of Theorem 4.8: as before, our reduction from TPJ introduces a slight correlation between input and split, and we use Lemma 2.4 to correct for this.

Theorem 4.10. *For each $p = p(n) \geq 1$,*

$$\mathrm{R}^{p}_{1/24}(\mathrm{W\text{-}TPJ}_{p+1,n}, \mathcal{U}_2) \;=\; \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \Big/ q(n,p)\right)\,, \quad \textit{where} \quad q(n,p) = p^2\left(cp^2\log n\right)^{\frac{2^p-1}{(p-1)2^{p+1}+2}}$$

for some large constant c. Thus, for every constant $\varepsilon > 0$, when p is large enough, we have

$$\mathrm{R}^{p}_{1/24}(\mathrm{W\text{-}TPJ}_{p+1,n}, \mathcal{U}_2) \;=\; \Omega\big(n^{(2+\varepsilon)^{-p}}\big)\,.$$

Proof. Let P be a protocol for $(\mathrm{W\text{-}TPJ}, \mathcal{U}_2)$ such that $\mathrm{err}(P, \mathrm{W\text{-}TPJ}, \mathcal{U}_2) \leq 1/24$. We will use P to construct a protocol Q for TPJ that works with probability at least $2/3$ on any instance ϕ when $\phi(v) \in_R \{0,1\}$ for each leaf node v. In Q, Alice and Bob first use public randomness to construct an input x for W-TPJ together with a random split of its tokens. They then proceed to simulate P on this instance. We first define

$$a_i \;:=\; (cp^2 t^{2(p+2)}\log n)^{2^{i-1}-1}\, t^{-2(3\cdot 2^{i-1}-i-2)} \tag{4.7}$$

for some large constant c. For each node v, the players use the following *public coin* randomised procedure to determine a bit string x_v and an allocation of its bits to the players in P.

If v is an internal node at level i: Choose random integers

$$d_{1v} \sim \mathcal{B}\left(\frac{a_i}{2}, 1/2\right) \quad \text{and} \quad d_{0v} \sim \mathcal{B}\left(\frac{a_i}{2}, 1/2\right)\,,$$

as well as a set

$$S_v \in_R \binom{[a_i]}{d_{1v}+d_{0v}}\,.$$

First assume i is even. Alice determines $\{x_{v,k} : k \in S_v\}$ and, uniformly at random, sets d_{1v} of these tokens to 1 and the remaining d_{0v} tokens to 0. Notice that all of this is done without reference to the input ϕ. Bob then uses ϕ to determine a target weight $|x_v|$ for the string x_v, based on Equation (4.1). Notice that many of the bits of x_v have already been fixed by the construction so far. Bob sets the free bits in such a way as to achieve this target weight, i. e., he randomly sets $|x_v| - d_{1v}$ of the bits $\{x_{v,k} : k \notin S_v\}$ to 1 and the remaining bits to 0. Note that this requires $d_{1v} \le |x_v| \le a_i - d_{0v}$; if this condition fails to hold, the protocol *aborts* and outputs a uniform random bit. If i is odd then Alice and Bob's roles are reversed.

If v is a leaf node: In this case x_v is a single bit. Allocate this bit to a random player, with Alice and Bob being chosen with equal probability. If the bit is allocated to Alice, she sets $x_v = \phi(v)$. Otherwise, Bob sets $x_v \in_R \{0,1\}$.

This completes the description of Q. Because $\mathcal{U}_2$ allocates each token to Alice with probability $1/2$, and ϕ assigns a uniformly random bit to each leaf, we have

$$\Pr[\text{W-TPJ}(x) = \text{TPJ}(\phi) \mid \text{the protocol } Q \text{ does not abort}] \;=\; \frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} \;=\; \frac{3}{4}\,.$$

It remains to show that the bit string x and the allocation σ generated in the reduction are sufficiently close to being independent. As in the proof of Theorem 4.8, we note that the marginals are correct: we do have $\sigma \sim \mathcal{U}_2$ and, for each leaf v, the value of x_v is indeed chosen according to a uniform setting of $\phi(v)$. The issue, as before, is that the joint distribution is non-product. However, note that had d_{1v} and d_{0v} been chosen according to $\mathcal{B}(|x_v|, 1/2)$ and $\mathcal{B}(a_i - |x_v|, 1/2)$, respectively, then σ and x would have been independent, and furthermore, the protocol would not abort. For each internal node v at level i, let

$$\tilde{A}_v := \mathcal{B}\left(\frac{1}{2}a_i, \frac{1}{2}\right), \quad \tilde{B}_v := \mathcal{B}\left(\frac{1}{2}a_i, \frac{1}{2}\right), \quad A_v := \mathcal{B}\left(|x_v|, \frac{1}{2}\right), \quad B_v := \mathcal{B}\left(a_i - |x_v|, \frac{1}{2}\right).$$

Hence, we need to show that the product of the distributions $\tilde{A}_v$ and $\tilde{B}_v$, over all internal nodes v, is sufficiently close to that of all A_v and B_v. Using Lemma 2.7, we can bound the total variation distance in terms of a_i and b_i as follows,

$$\begin{aligned} \mathrm{D_{TV}}\left(\bigotimes_v (\tilde{A}_v \otimes \tilde{B}_v), \bigotimes_v (A_v \otimes B_v)\right) &\le \sum_v \mathrm{D_{TV}}\left(\tilde{A}_v, A_v\right) + \sum_v \mathrm{D_{TV}}\left(\tilde{B}_v, B_v\right) \\ &\le O(\sqrt{\log n}) \sum_{i=2}^{p+1} \frac{t^{p+2-i} b_{i-1}}{\sqrt{a_i}} \end{aligned}$$

where the first inequality follows from the triangle inequality. Noting that $b_{i-1} \le 2a_{i-1}$ and substituting in the value for a_i, the distance can be made less than $1/24$ for sufficiently large constant c. By Lemma 2.4,

$$\mathrm{err}_\mu(Q, \text{TPJ}_{p+1,t}) \;\le\; \frac{1}{4} + \frac{1}{24} + \mathrm{err}(P, \text{W-TPJ}_{p+1,n}, \mathcal{U}_2) \;\le\; \frac{1}{3}\,.$$

Therefore, by Theorem 4.9,

$$\mathrm{R}^p_{1/24}(\text{TPJ}_{p+1,n}, \mathcal{U}_2) \;=\; \Omega(t/p^2)\,.$$

Note that

$$n \;=\; b_{p+1} \;\le\; 2(cp^2 t^{2(p+2)} \log n)^{2^p - 1} t^{-2(3\cdot 2^p - p - 3)} \;=\; 2(cp^2 \log n)^{2^p - 1} t^{(p-1)2^{p+1}+2},$$

and hence,

$$t = \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \big/ \left(cp^2 \log n\right)^{\frac{2^p-1}{(p-1)2^{p+1}+2}}\right). \qquad \square$$

5 Hamming distance and index

In this section, we prove robust lower bounds for the fundamental communication problems INDEX and GAP-HAMMING-DISTANCE.

5.1 Hamming distance

The GAP-HAMMING-DISTANCE problem (henceforth, GHD) was first formally stated in the context of data stream lower bounds [30, 42, 27]. The central goal is to determine whether the Hamming distance between two binary strings is "low" or "high," with a certain gap (given by a parameter, G) between the demarcations of "low" and "high." To be precise, define the function $\Delta : \{0,1\}^{2n} \to \mathbb{Z}$ by

$$\Delta(x) := |\{i \in [n] : x_{2i} \neq x_{2i-1}\}|, \quad \text{for } x \in \{0,1\}^{2n}.$$

For $G \in \mathbb{R}^+$, we then define $\text{GHD}_{n,G} : \{0,1\}^{2n} \to \{0,1,\star\}$ by

$$\text{GHD}_{n,G}(x) := \begin{cases} 0, & \text{if } \Delta(x) \geq n/2+G, \\ 1, & \text{if } \Delta(x) \leq n/2-G, \\ \star, & \text{otherwise,} \end{cases}$$

where "$\star$" can be interpreted as "undefined." Equivalently, a computation problem corresponding to the function $\text{GHD}_{n,G}$ can be thought of as a promise problem, where we are promised that $\Delta(x)$ does not fall between $n/2-G$ and $n/2+G$. Traditional (i. e., "fragile") communication lower bounds for this problem, where Alice receives $x_1, x_3, \ldots, x_{2n-1}$ and Bob receives $x_2, x_4, \ldots, x_{2n}$, have been heavily studied recently. In particular, Chakrabarti and Regev [10] show that $\mathrm{R}(\text{GHD}_{n,G}) = \Theta(\min\{n, n^2/G^2\})$; see, also, Sherstov [39] and Vidick [41].

For a number of reasons (in particular, the data stream applications) the problem is most interesting when we set $G = \Theta(\sqrt{n})$. We shall prove an optimal robust lower bound for the problem in this setting.

Theorem 5.1. *There exists a constant $c_3 > 0$ such that*

$$\mathrm{R}_{1/4}(\text{GHD}_{n,c_3\sqrt{n}}, \mathcal{U}_2) = \Omega(n).$$

Remark 5.2. It is worth pointing out that c_3 needs to be sufficiently small for the theorem to hold. In fact, for a sufficiently large constant c_4, we have $\mathrm{R}_{1/4}(\text{GHD}_{n,c_4\sqrt{n}}, \mathcal{U}_2) = 0$. This is in contrast to the case of the standard fixed-partition version of GHD, which remains hard at all gaps in $\Theta(\sqrt{n})$.

The theorem will be proved by a reduction from the GHD problem in the standard setting where Alice and Bob hold $x_1, x_3, \ldots, x_{2n-1}$ and $x_2, x_4, \ldots, x_{2n}$ respectively. Before presenting the actual proof, it may be useful to consider a proof attempt that will *not* work. Specifically, suppose Alice and Bob use public randomness to determine a random split σ. If $\sigma(2i-1) = 1$ and $\sigma(2i) = 2$, then Alice and Bob already know the relevant bits of x. If $\sigma(2i-1)=2$ and $\sigma(2i)=1$, then Alice could use x_{2i-1} in place of x_{2i} and Bob could use x_{2i} in place of x_{2i-1} since this will not change the Hamming distance. However, if $\sigma(2i-1)=\sigma(2i)$ then the relevant player will not know both x_{2i-1} and x_{2i} so suppose that he or she picks the unknown bit randomly. If y is the resulting bit string upon which the protocol is being simulated and $\mathrm{GHD}_{n,\sqrt{n}}(x)\in\{0,1\}$, then it can be shown that $\mathrm{GHD}_{n,\sqrt{n}}(x)=\mathrm{GHD}_{n,c\sqrt{n}}(y)$ with probability 0.99 if $c>0$ is a sufficiently small constant.

Hence, it would *appear* that we can conclude that any protocol for $\mathrm{GHD}_{n,c\sqrt{n}}$ in the random-allocation setting can be used to solve $\mathrm{GHD}_{n,\sqrt{n}}$ in the traditional setting and therefore imply a lower bound. Unfortunately this is incorrect. The correctness guarantee of a protocol in the random-allocation setting is that for any input y, if the bits of y are partitioned uniformly at random, then the protocol should be correct with the required probability. This guarantee naturally still applies if y is chosen according to some distribution and the bits are partitioned uniformly at random, but only if the distribution of y is independent of the partitioning, which is not the case in the above attempt at a proof, since

$$\Pr[y_{2i-1}\oplus y_{2i}=1] \;=\; \begin{cases}1/2 & \text{if } \sigma(2i-1)=\sigma(2i)\,,\\ \Delta(x)/n & \text{if;} \sigma(2i-1)\neq\sigma(2i)\,.\end{cases}$$

We address this issue in the proof by padding the original instance of GHD with extra coordinates such that the weak correlation in pairs of bits that are split between the different players is masked by the random bits in the extra coordinates. There will be a small correlation between the resulting string and the random allocation but, by setting parameters appropriately, we will be able to make this correlation arbitrarily small.

Proof of Theorem 5.1. Let $y\in\{0,1\}^{2n}$ be an instance of GHD in the standard setting where y_i is known to Alice if i is odd and known to Bob if i is even. Define g by $\Delta(y)=n/2+g$. Distinguishing whether $g\geq\sqrt{n}$ or $g<-\sqrt{n}$ requires $\Omega(n)$ bits of communication [10] in the standard setting. This bound also applies if we promise $\sqrt{n}\leq|g|\leq 10\sqrt{n}$ since the lower bound applies even when $y\in_R\{0,1\}^{2n}$ and under this distribution the promise is satisfied with constant probability. Henceforth, we assume $\sqrt{n}\leq|g|\leq 10\sqrt{n}$.

Using y and some public randomness, Alice and Bob generate an instance $z\in\{0,1\}^{2m}$ in the robust setting, together with a split σ, where $m=cn$ for some large constant c. The pair $\mathcal{I}=(z,\sigma)$ will have the properties that z and σ are nearly independent, and that $\mathrm{GHD}_{n,\sqrt{n}}(y)=\mathrm{GHD}_{m,\sqrt{n}}(z)$.

It will be helpful to generate (z,σ) by first generating five sets S_0,S_1,D_0,D_1, and D_{input} that partition $[m]$. For $b\in\{0,1\}$, S_b includes i if the $(2i-1)$-th token and the $(2i)$-th token are received by the *same* player, i. e., $\sigma(2i-1)=\sigma(2i)$, and $z_{2i-1}\oplus z_{2i}=b$. For $b\in\{0,1\}$, D_b includes i if the $(2i-1)$-th token and the $(2i)$-th token are received by *different* players and $z_{2i-1}\oplus z_{2i}=b$. The set D_{input} encodes a further n index pairs that will be received by different players and where

$$z_{2i-1}\oplus z_{2i}=y_{2\pi(i)-1}\oplus y_{2\pi(i)}\,,$$

where π is a random bijection between $[n]$ and D_{input}.

To determine S_0,S_1,D_0,D_1, and D_{input}, we proceed as follows. Using public randomness, Alice and Bob choose independent random integers $s_0,s_1\sim\nu$ where ν is the distribution $\mathcal{B}(m/2,1/2)$ conditioned on the event that the value is at most $(m-n)/2$. Let (S_0, S_1, E, D_0, D_1) be a random partition of $[m]$ conditioned on

$$|S_0| = s_0, \quad |S_1| = s_1, \quad |E| = n, \quad |D_0| = (m-n)/2 - s_0, \quad \text{and} \quad |D_1| = (m-n)/2 - s_1.$$

Alice and Bob then choose (z, σ) uniformly at random conditioned on the choice of the above sets:

- For $i \in D_0 \cup D_1 \cup S_0 \cup S_1$, Alice and Bob know $z_{2i-1} \oplus z_{2i} = b$ and hence the pair $(z_{2i-1}, z_{2i}) \in_R \{(0,b),(1,1-b)\}$ can be determined using only public randomness.
- For $i \in D_{\text{input}}$, let r_i be a public random bit. If $\sigma(2i-1) = 1$ and $\sigma(2i) = 2$ then Alice sets $z_{2i-1} = r_i \oplus y_{2\pi(i)-1}$ and Bob sets $z_{2i} = r_i \oplus y_{2\pi(i)}$. If $\sigma(2i-1) = 2$ and $\sigma(2i) = 1$ then Alice sets $z_{2i} = r_i \oplus y_{2\pi(i)-1}$ and Bob sets $z_{2i-1} = r_i \oplus y_{2\pi(i)}$. In each case, z_{2i-1} and z_{2i} are chosen uniformly at random conditioned on $z_{2i-1} \oplus z_{2i} = y_{2\pi(i)-1} \oplus y_{2\pi(i)}$.

Observe that $\Delta(z) = |S_1| + |D_1| + \Delta(y) = (m-n)/2 + n/2 + g = m/2 + g$. Therefore $\text{GHD}_{n,\sqrt{n}}(y) = \text{GHD}_{m,\sqrt{n}}(z)$ as required.

Suppose P is a robust protocol with failure probability δ. In the above reduction, z and σ are not generated independently but we will show that P still has a failure probability at most $\delta + 1/5$ on the inputs we generate. To do this, we consider a second distribution over instances $\mathcal{I}'$; this distribution is purely for the sake of analysis and Alice and Bob will not have sufficient information to generate instances according to this distribution. To generate $\mathcal{I}'$, choose $s_0 \sim \mathcal{B}(m/2 - g, 1/2)$ and $s_1 \sim \mathcal{B}(m/2 + g, 1/2)$ and let (S_0, S_1, D_0, D_1) be a random partition of $[m]$ conditioned on

$$|S_0| = s_0, \quad |S_1| = s_1, \quad |D_0| = m/2 - g - s_0, \quad \text{and} \quad |D_1| = m/2 + g - s_1,$$

and we set the values of $\sigma(2i-1), \sigma(2i), z_{2i-1}$, and z_{2i} uniformly at random conditioned on the choice of these sets. Note that the distribution of $\mathcal{I}$ and $\mathcal{I}'$ is identical conditioned on the values of s_0 and s_1. Furthermore, $\mathcal{I}'$ is distributed according to a product distribution and hence P outputs $\text{GHD}_{m,\sqrt{n}}(z) = \text{GHD}_{n,\sqrt{n}}(y)$ with probability at least $1-\delta$ on this distribution. Hence, if δ' is the failure probability of P on the distribution of $\mathcal{I}$, we have

$$\begin{aligned}
\delta' &\le \delta + \mathrm{D_{TV}}\Big(\nu, \mathcal{B}(m/2-g, 1/2)\Big) + \mathrm{D_{TV}}\Big(\nu, \mathcal{B}(m/2+g, 1/2)\Big) \\
&\le \delta + 2\Pr[\mathcal{B}(m/2, 1/2) \ge (m-n)/2] + \mathrm{D_{TV}}\Big(\mathcal{B}(m/2, 1/2), \mathcal{B}(m/2-g, 1/2)\Big) \\
&\quad + \mathrm{D_{TV}}\Big(\mathcal{B}(m/2, 1/2), \mathcal{B}(m/2+g, 1/2)\Big) \\
&\le \delta + 1/10 + 1/20 + 1/20 \le \delta + 1/5.
\end{aligned}$$

where the last line follows from the Chernoff bound and from Lemma 2.7 (since $g^2/m \le 100n/m$), by ensuring that $m/n = c$ is sufficiently large. □

5.2 Index

For our purposes, we define the INDEX_n problem over inputs $x \in [n] \times \{0,1\}^n$ as follows: $\text{INDEX}_n(x) := x_j$ where $j := x_0$. Traditionally, one considers the worst-case partition where Alice (the player who speaks) holds $x_1 \dots x_n$ and Bob holds j. The resulting problem is one of the most basic in communication complexity, and strong randomised lower bounds are known for it in this setting [1]. In this fixed-partition model, INDEX_n can be thought of as a special case of $\text{DISJ}_{n,2}$, where one string is of the form $\mathbf{e}_i$. This is no longer the case under uniform splits, since the zeros in $\mathbf{e}_i$ get spread between the players, and leak information about which indices are not of interest.

We prove a robust lower bound for a generalisation of INDEX that allocates multiple copies of the tokens $(x_0, \dots, x_n)$ amongst two players. This generalisation is needed for proving subsequent data stream bounds. For positive integers a and b, let $\text{INDEX}_n^{a,b}$ denote the problem where the input consists of a copies of each x_i (for $i \in [n]$) and b copies of x_0, with $x = (x_0, \dots, x_n)$ being an input for INDEX_n. The (partial) function $\text{INDEX}_n^{a,b}$ takes the value $\text{INDEX}_n(x)$ on such an input. Let ν_p denote the distribution of a random split obtained by independently giving each input token to Player 1 with probability p, and to Player 2 otherwise.

Theorem 5.3. *Let a, b, and p constants such that a and b are positive integers and $0 < p < 1$. We have*

$$\mathrm{R}_\delta^{\rightarrow}(\text{INDEX}_n^{a,b}, \nu_p) = \Omega(n)\,,$$

where $\delta = (1-p)^b p^a / 4$.

Proof. The proof is by reduction from INDEX_n when Alice holds $x_1 \dots x_n \in \{0,1\}^n$ and Bob holds index $x_0 = j$. Let μ be the uniform distribution over all possible inputs. By Lemma 2.5, any one-way protocol succeeding with probability at least $1/2 + (1-p)^b p^a/4$ (for a, b, p positive constants) for instances of INDEX_n drawn from μ requires $\Omega(n)$ bits to be communicated.

Suppose there exists a one-way protocol P with the property that $\text{err}(P, \text{INDEX}_n^{a,b}, \nu_p) \le (1-p)^b p^a/4$. We use P to create the following (traditional) protocol Q for INDEX_n. Let $x = (x_0, \dots, x_n)$ denote the input given to Alice and Bob in Q, and let $\hat{x}$ denote the corresponding input to the players in $\text{INDEX}_n^{a,b}$. Alice and Bob agree on a split $\sigma \sim \nu_p$, of the tokens in $\hat{x}$, using public coins. Consider the events

$$\begin{aligned} B_0 &= \text{“a copy of } x_0 \text{ is allocated to Player 1,” and} \\ B_i &= \text{“a copy of } x_i \text{ is allocated to Player 2,” for } i \in [n]. \end{aligned}$$

Alice and Bob behave as follows in the protocol Q. If B_0 occurs, then Bob outputs a uniform random bit. Otherwise, for each $i \in [n]$ such that B_i occurs, they jointly choose a (public) random bit $r_i \in_R \{0,1\}$ and set all bits of $\hat{x}$ that are copies of x_i equal to r_i. The remaining bits of $\hat{x}$ (i. e., those that are copies of x_i such that B_i does not occur) are left unchanged. Alice and Bob then simulate P on this updated input $\hat{x}$, playing the roles of Player 1 and Player 2 respectively. Clearly, $\text{cost}(Q) \le \text{cost}(P)$.

Let $B = B_0 \vee B_j$ (recall that $j = x_0$). If B occurs, then the output of Q is a random bit. Otherwise, Q is correct whenever P is. Note that $\Pr[B] = 1 - (1-p)^b p^a$. Thus, the correctness probability of Q is

$$\frac{\Pr[B]}{2} + \Pr[\neg B \wedge (P \text{ is correct})] \;\ge\; \frac{\Pr[B]}{2} + \Pr[P \text{ is correct}] - \Pr[B] \;\ge\; \frac{1}{2} + \frac{(1-p)^b p^a}{4}\,,$$

which implies $\text{cost}(Q) = \Omega(n)$, and hence, $\text{cost}(P) = \Omega(n)$. $\square$

6 Robust lower bounds for data stream computation

We use our results on communication complexity from the previous sections to derive robust lower bounds for a number of problems in the data stream model. The connection between random-allocation communication complexity and robust bounds in the data stream model is a natural extension of the connection between fixed-partition communication complexity and the basic data stream model where the data is ordered adversarially. In particular, an r-pass, s-space data stream algorithm for evaluating a function f on a set S of tokens presented in (uniform) random order yields an r-round, p-player communication protocol for evaluating $f(S)$ for certain ways of partitioning S into p subsets $S_1, \ldots, S_p$, with the i-th player receiving S_i. To be precise, each token is placed in one subset chosen independently, but not necessarily uniformly, from $S_1, \ldots, S_p$.

The communication protocol then works as follows. The i-th player (uniformly) randomly permutes S_i to generate stream s_i and the players emulate the algorithm on the concatenated stream $\langle s_1 | s_2 | \ldots | s_p \rangle$. This emulation requires $O(rps)$ bits of communication. Given a lower bound on the complexity of the communication problem, this allows us to deduce a lower bound for the data stream problem.

6.1 Frequency moments

The (estimations of) various frequency moments are some of the most well-studied problems in the data stream model [3]. Suppose the stream comprises a sequence of m values $a_j \in [n]$. Define $f_i = |\{j : a_j = i\}|$. The k-th frequency moment (k not necessarily integral) is

$$F_k := \sum_{i \in [n]} f_i^k .$$

We consider constant $k \geq 3$. It is known that any $O(1)$-pass algorithm that returns a $(1/2, 1/4)$-approximation of F_k requires $\tilde{\Omega}(n^{1-2/k})$ space[4] and that this is tight under worst-case orderings [28, 9]. However, it was observed that for random orderings and $m = \tilde{\Omega}_\varepsilon(an)$ ($a > 1$) there exists a single pass $\tilde{O}((n/a)^{1-2/k})$-space algorithm that (ε, δ)-approximates F_k [22]. The following theorem shows a lower bound on the space usage in the random-order case. The proof combines Theorem 3.3 with a variation of the reduction used in Alon, Matias, and Szegedy [3, Theorem 3.2].

Theorem 6.1. *An r-pass algorithm giving a $(1/10, 1/10)$-approximation for F_k of a randomly ordered stream requires $\Omega(n^{1-3/k}/r)$ space. For streams where $m = \Omega(an)$, for some integer $a > 1$, $\Omega(n^{1-3/k}/(a^3 r))$ space is required.*

Proof. Suppose there exists an r-pass, $(1/10, 1/10)$-approximation algorithm for F_k that uses s bits of space. Set $t = (5n/4)^{1/k}$. Let $x = \{x_{ij}\}_{i \in [t], j \in [n]}$ be an instance for $\text{DISJ}_{n,t}$ that satisfies the unique intersection promise (as discussed at the start of Section 3). Consider a uniform random split of the nt input tokens between $p = 20t^2$ players. Let the player who receives the token for x_{ij}, generate the value j if $x_{ij} = 1$ and define S_j to be the multiset of values generated by the j-th player. Note that the sets $S_1, \ldots, S_p$ are a random partition of $S = S_1 \cup \cdots \cup S_p$. Furthermore $F_k(S) \geq t^k = 5n/4$ if $\text{DISJ}_{n,t}(x) = 1$

[4] The $\tilde{O}(\cdot)$ and $\tilde{\Omega}(\cdot)$ notations used in this section suppress logarithmic dependencies on the stream length, m, the universe size, n, and the inverse error probability, δ^{-1}.

and $F_k(S) \le n$ if $\mathrm{DISJ}_{n,t}(x) = 0$. Using the template at the start of Section 6 and appealing to Theorem 3.3, we can deduce that $rps = \Omega(n/t)$. Therefore $s = \Omega(n^{1-3/k})$ as required.

To prove the second part of the theorem, the reduction from $\mathrm{DISJ}_{n,t}$ proceeds as before but we also add a copies of $[n]$ randomly distributed between the p players. This is achieved using public randomness. Now, if $\mathrm{DISJ}_{n,t}(x) = 1$, then $F_k \ge t^k$, but if $\mathrm{DISJ}_{n,t}(x) = 0$, then $F_k \le (a+1)^k n$. If we now choose $t = (5n/4)^{1/k}(a+1)$, a $(1/10, 1/10)$-approximation to F_k distinguishes the two cases. The resulting lower bound on the space is $\Omega(n/(rpt)) = \Omega(n^{1-3/k}/(a^3 r))$. □

6.2 Distinct elements and entropy: Space/approximation tradeoffs

The number of distinct elements in a stream is $F_0 := |\{i \in [n] : f_i \neq 0\}|$, and the empirical entropy is $H := \sum_{i\in[n]}(f_i/m)\log(m/f_i)$; recall that m denotes the length of the stream. Within this subsection let us think of the "input size" parameters m and n as fixed, and the approximation parameter ε as varying. One-pass, $\tilde{O}(\varepsilon^{-2})$-space, (ε, δ)-approximation algorithms are known for both problems [7, 25, 17, 5, 18]. We prove that the known algorithms are essentially tight even under random order. These results follow from Theorem 5.1 and the reductions in [7, Theorem 2] and [42, Section 3.2].

Theorem 6.2. *Let $k \ge 0$ be a constant with $k \neq 1$. Then, any r-pass $(\varepsilon, 1/10)$-approximation for F_k of a randomly ordered stream requires $\Omega(\varepsilon^{-2}/r)$ space. Also, any r-pass $(\varepsilon, 1/10)$-approximation for H of a randomly ordered stream requires $\Omega(\varepsilon^{-2}/(\log^2\varepsilon^{-1}\cdot r))$ space.*

Proof. Suppose there exists an r-pass, $(\varepsilon, 1/10)$-approximation algorithm for H that uses $s(\varepsilon)$ bits of space. Let $x \in \{0,1\}^{2n}$ be an instance of $\mathrm{GHD}_{n,G}$ (where n and G are determined by ε below) and consider a uniform random split of the $2n$ tokens between two players. Let the player who receives the token for x_i generate the value $(\lceil i/2 \rceil, x_i)$. Define S_A and S_B to be the multisets of values generated by Alice and Bob respectively. Note that S_A and S_B are a random partition of $S := S_A \cup S_B$. Furthermore,

$$H \;=\; \frac{\Delta}{n}\log(2n) + \frac{n-\Delta}{n}\log n \;=\; \frac{\Delta}{n} + \log n\,,$$

where $\Delta = \Delta(x) = |\{i \in [n] : x_{2i} \neq x_{2i-1}\}|$. Hence, any algorithm which can $(\varepsilon, 1/10)$-approximate H can also distinguish the cases $\Delta(x) \le n/2 - G$ and $\Delta(x) \ge n/2 + G$, provided $\varepsilon < G/(n\log n)$. For any ε that is less than c_3, we set $n = c_3^2\varepsilon^{-2}/(4\log^2\varepsilon^{-1})$ and $G = c_3\sqrt{n}$, where c_3 is the constant from Theorem 5.1. This ensures $\varepsilon < G/(n\log n)$. Using the template at the start of Section 6 and appealing to Theorem 5.1, we can deduce that $rs(\varepsilon) = \Omega(n) = \Omega(\varepsilon^{-2}/\log^2\varepsilon^{-1})$.

The frequency moments lower bound is similar: the same reduction ensures that if $\Delta(x) = n/2 + g$, then

$$F_k = 2^k(n - \Delta(x)) + 1^k \cdot 2\Delta(x) = 2^k n + (2 - 2^k)\cdot\Delta(x) = (2^k + 1 - 2^{k-1})n + (2 - 2^k)\cdot g\,.$$

Then either

$$\frac{F_k}{(2^k + 1 - 2^{k-1})n} \le 1 - \frac{|2 - 2^k|G}{(2^k + 1 - 2^{k-1})n} \quad\text{or}\quad \frac{F_k}{(2^k + 1 - 2^{k-1})n} \ge 1 + \frac{|2 - 2^k|G}{(2^k + 1 - 2^{k-1})n}\,.$$

Setting $n = (\varepsilon^{-1}c_3|2 - 2^k|/(2^k + 1 - 2^{k-1}))^2$ and $G = c_3\sqrt{n}$ ensures that the value of F_k in each case differs by a factor of at least $1 + \varepsilon$ and hence the communication lower bound of $\Omega(n)$ entails a space lower bound of $\Omega(\varepsilon^{-2}/r)$ for constant k. □

6.3 Selection

Selection, including median-finding, is one of the earliest-studied problems in the data stream model [35] and has been the focus of several works [19, 23, 8]. The following result improves upon the previous best single and multi-pass lower bounds [23, 8]. As an example, our theorem implies a $\tilde{\Omega}(m^{1/10})$ space lower bound for 3-pass algorithms whereas the best previous result was $\tilde{\Omega}(m^{3/80})$ [8].

Theorem 6.3. *Any p-pass algorithm to return the median of a length-m randomly ordered stream which succeeds with probability at least $3/4$ requires*

$$\Omega\left(m^{1/((p-1)2^{p+1}+2)}/q(m,p)\right)$$

space where the function q, defined in Theorem 4.8, is $\operatorname{polylog}(m)$ *for any constant $p \geq 1$.*

Proof. Using the template at the start of Section 6, the theorem is immediate from Theorem 4.8. Note that this is an example where we consider a robust communication bound in which the player receiving a specific token is not chosen uniformly. However, as long as the tokens are distributed independently, the *order* of the concatenated stream is chosen uniformly at random as required. □

We note that a weaker bound follows from Theorem 4.10. The reason that a reduction from the two-player result is weaker (despite the apparent similarity between Theorem 4.10 and Theorem 4.8) stems from the different definition of communication rounds. In the multi-player setting, p streaming passes corresponds to p rounds, whereas in the two-player setting, p streaming passes corresponds to $2p-1$ rounds. Hence, a reduction from the two-party setting would result in occurrences of p in the exponent in the above theorem being replaced by occurrences of $2p-1$.

6.4 Graph streaming

We now consider bounds on solving graph problems given a stream of edges in random order. These problems are known to require a large amount of space when edges are presented in arbitrary order [15, 26]. The problems are no easier when the edge order is randomized. Using Theorem 3.3 and Theorem 5.3 and reductions similar in spirit to those in [15, 26] we show the following results.

Theorem 6.4. *An r-pass algorithm that, given a stream of edges in random order, determines whether the resulting graph is connected requires $\Omega(n/r)$ space.*

Proof. We consider a reduction from $\text{DISJ}_{n/2,2}$ where tokens corresponding to each bit are uniformly distributed between p players. We present a lower bound on the communication required between p players to determine whether a graph is connected when the edges of the graph are randomly partitioned between the p players. The stream lower bound follows immediately from the template at the start of Section 6. Let $x = \{x_{ij}\}_{i\in[2],j\in[n/2]}$ be an instance of $\text{DISJ}_{n/2,2}$. Based on x we define the following graph $G_x = (L \cup R, E_1 \cup E_2 \cup E_3)$ where $L = \{l_1, \ldots, l_{n/2}\}$, $R = \{r_1, \ldots, r_{n/2}\}$ and the edge set includes

$$\begin{aligned}
E_1 &= \{(l_i, r_i) : i \in [n/2]\}, \\
E_2 &= \{(l_j, l_{j+1}) : j \in [n/2], x_{1,j} = 0\}, \\
E_3 &= \{(r_j, r_{j+1}) : j \in [n/2], x_{2,j} = 0\},
\end{aligned}$$

where $l_{n/2+1} = l_1$ and $r_{n/2+1} = r_1$. It is easy to see that G_x is disconnected iff there exists j such that $x_{1,j} = x_{2,j} = 1$. To perform the reduction, the players replace the token corresponding to each $x_{i,j}$ if appropriate. Note that the edges of $E_2 \cup E_3$ are randomly partitioned between the players because the relevant tokens were randomly partitioned. Using public randomness, the players can decide on a random partition of E_1. In this way the entire edge set of G_x is randomly partitioned between the p players. Setting $p = 40$ and appealing to Theorem 3.3 gives the required result. □

Theorem 6.5. *Given a stream of edges in random order, a single pass algorithm that distinguishes between when the distance between two given vertices (known at the start of the stream) is at most* 1 *or at least* $t+1$ *requires* $\Omega(\mathrm{ex}(n-2, C_3, \ldots, C_{t+1}))$ *space, where* $\mathrm{ex}(n-2, C_3, \ldots, C_{t+1})$ *is the maximum number of edges of a graph on* $n-2$ *vertices that does not include any cycles of length strictly less than* $t+2$.

A well-known result in extremal graph theory is that $\mathrm{ex}(n, C_3, \ldots, C_{t+1}) = \Omega(n^{1+1/t})$, and it has long been conjectured that $\mathrm{ex}(n, C_3, \ldots, C_{t+1}) = \Omega(n^{1+2/t})$ for $t \geq 2$; see, e. g., Simonovits [40].

Proof. Let $G = (V, E)$ be a graph on $n-2$ vertices with $m = \mathrm{ex}(n-2, C_3, \ldots, C_{t+1})$ edges such that the shortest cycle has length at least $t+2$. Let $e_1, \ldots, e_m$ be some arbitrary ordering of the edges in G. Let a, b be two vertices not in V. Consider an instance $x \in [m] \times \{0,1\}^m$ of INDEX where one copy of each x_i ($i \in [m]$) and two copies of x_0 are distributed uniformly between two players. Consider the reduction in which, for $i \geq 1$, each x_i is ignored if $x_i = 0$ and replaced by an edge e_i with unit weight if $x_i = 1$. Suppose that $x_0 = j$ and that $e_j = (u_j, v_j)$. With probability $1/2$ replace the first copy of x_0 by (a, u_j) and the second copy by (b, v_j); these edges have zero weight. With the remaining probability, replace them in the reverse order. In this way we define a graph G' on vertices $V \cup \{a, b\}$ where the distance between a and b is 1 if $x_j = 1$ and at least $t+1$ if $x_j = 0$. Hence, any protocol that distinguishes between the distance being 1 and at least $t+1$ also determines the value of x_j. Appealing to Theorem 5.3 gives the required result. □

6.5 Information divergences

It is common to interpret streams as defining a distribution over a finite set of tokens: the frequencies can be rescaled to define an empirical probability distribution. Over such empirical distributions, we desire to compute various statistical measures. The next theorem extends a result by Guha et al. [21] on the approximation of information divergences. The results follows from Theorem 5.3 using a variant of their reduction.

Theorem 6.6. *Let a be an even positive integer. Given a randomly ordered stream defining two empirical distributions p and q on* $[n]$, $\Omega(n)$ *space is required to produce an estimate* $\hat{h}$ *such that*

$$\frac{\mathrm{h}^2(p,q)}{\sqrt{(a+1)/2}} \leq \hat{h} \leq \sqrt{(a+1)/2} \cdot \mathrm{h}^2(p,q)$$

holds with probability at least $1 - 2^{-a-3}$. *Here,* h^2 *denotes squared Hellinger distance.*

Proof. We consider a reduction from INDEX. Let $k \in [n]$, $x_1 \ldots x_n \in \{0,1\}^n$ be an instance of INDEX. Consider the random allocation where a copies of each x_i are uniformly distributed between the two players and k is revealed to a player chosen uniformly at random. The players transform this input into a set of tokens $\langle p,i \rangle$ and $\langle q,i \rangle$ as follows:

1. Using public randomness, the players generate n random binary strings $y^1,\ldots,y^n \in \{p,q\}^a$ where each string has the same number of each symbol. Suppose Alice receives d_i copies of the token for x_i. Then, Alice generates a copy of $\langle y^i_j, i\rangle$ for each $j \le d_i$ if $x_i = 1$. Similarly, Bob generates a copy of $\langle y^i_j, i\rangle$ for each $d_i < j \le a$ if $x_i = 1$. Note that if $d_i = 0$ or $d_i = a$, then one of the players will not know x_i. But in this case, that player will be generating the empty set.

2. The player receiving the token for k generates a copy of $\langle q,k \rangle$.

3. Additionally, the players generate $a/2+1$ copies of $\langle p,n+1 \rangle$ and $a/2$ copies of $\langle q,n+1 \rangle$. These are uniformly distributed between the players.

In this way, for each $i \in [n]$ such that $x_i = 1$, $a/2$ copies of $\langle p,i \rangle$ and $a/2$ copies of $\langle q,i \rangle$ have been generated. Additionally, one copy of $\langle q,k \rangle$, $a/2+1$ copies of $\langle p,n+1 \rangle$ and $a/2$ copies of $\langle q,n+1 \rangle$ have been generated. Therefore,

$$\mathrm{h}^2(p,q) = \begin{cases} \frac{1}{m}\left(\left(\sqrt{a/2}-\sqrt{a/2+1}\right)^2+1\right), & \text{if } x_k = 0, \\ \frac{2}{m}\left(\sqrt{a/2}-\sqrt{a/2+1}\right)^2, & \text{if } x_k = 1, \end{cases}$$

where $m = a/2+1+(a/2)|\{i \in [n] : x_i = 1\}|$. Furthermore, the ratio of these quantities is

$$\frac{\left(\sqrt{a/2}-\sqrt{a/2+1}\right)^2+1}{2\left(\sqrt{a/2}-\sqrt{a/2+1}\right)^2} = \frac{1}{2}+\frac{1}{2\left(\sqrt{a/2}-\sqrt{a/2+1}\right)^2} > \frac{a+1}{2}.$$

Hence, an algorithm producing an estimate satisfying the stated bounds would be sufficient to solve the instance of INDEX. Therefore, by Theorem 5.3, any stream algorithm that returns such an estimate requires $\Omega(n)$ space. □

Acknowledgements

We are grateful to the anonymous referees for a number of comments that were a great help to us in improving the presentation of the paper and fixing some subtle issues in the proof of Theorem 3.3.

A Distance between binomial distributions

We prove the following technical lemma, giving an upper bound on the total variation distance between two binomial distributions that have roughly the same number of trials. The lemma essentially shows that the total variation distance between two binomial distributions $\mathcal{B}(a,q)$ and $\mathcal{B}(a-w,q)$ is small as long as w is small compared to the standard deviation of $\mathcal{B}(a,q)$.

Lemma A.1 (Restatement of Lemma 2.7). *There exists a constant $c_1 > 0$ such that for all $q \in [1/2, 1)$, $r \in (0,1)$, and $a \in \mathbb{N}$,*

$$\frac{w}{\sqrt{v}} \le r \quad \Longrightarrow \quad \mathrm{D_{TV}}\left(\mathcal{B}(a,q), \mathcal{B}(a-w,q)\right) \;\le\; c_1 r\sqrt{\ln\frac{2}{r}}\,,$$

where $v = aq(1-q)$ is the variance of $\mathcal{B}(a,q)$. In order to define the total variation distance above, we treat $\mathcal{B}(n,p)$ as a distribution on the set of all non-negative integers, rather than just $\{0,1,\ldots,n\}$.

Proof. Let $\gamma = 1-q$ and

$$\alpha_i = \binom{a}{i} q^i (1-q)^{a-i} \quad \text{and} \quad \beta_i = \binom{a-w}{i} q^i (1-q)^{a-w-i}\,.$$

By an application of the Chernoff bounds,

$$\Pr\left[|\mathcal{B}(a,q) - aq| \ge \sqrt{4\ln(2/r)\cdot v}\right] \;\le\; r\,, \quad \text{and}$$
$$\Pr\left[|\mathcal{B}(a-w,q) - aq| \ge wq + \sqrt{4\ln(2/r)\cdot v}\right] \;\le\; r\,.$$

Let $t = 4\sqrt{\ln(2/r)\cdot v}$ and note that

$$t \;=\; \sqrt{4\ln(2/r)\cdot v} + \sqrt{4\ln(2/r)\cdot v} \;\ge\; wq + \sqrt{4\ln(2/r)\cdot v}\,.$$

because $wq \le \sqrt{v}$ and $\sqrt{4\ln(2/r)} \ge 1$. Therefore we may bound the total variation distance as follows.

$$\begin{aligned}
\mathrm{D_{TV}}\left(\mathcal{B}(a,q), \mathcal{B}(a-w,q)\right) \;&=\; \sum_i |\alpha_i - \beta_i| \\
&\le\; 2r + \sum_{\substack{i\in aq\pm t\\ \alpha_i\ge\beta_i}} (\alpha_i - \beta_i) + \sum_{i\in aq\pm t:\beta_i\ge\alpha_i} (\beta_i - \alpha_i) \\
&=\; 2r + \sum_{i\in aq\pm t:\alpha_i\ge\beta_i} \alpha_i(1-\beta_i/\alpha_i) + \sum_{i\in aq\pm t:\beta_i\ge\alpha_i} \beta_i(1-\alpha_i/\beta_i) \\
&\le\; 2r + \left(\sum_{\substack{i\in aq\pm t\\ \alpha_i\ge\beta_i}} \alpha_i\right) \max_{\substack{i\in aq\pm t\\ \alpha_i\ge\beta_i}} \left(1 - \frac{\beta_i}{\alpha_i}\right) + \left(\sum_{\substack{i\in aq\pm t\\ \beta_i\ge\alpha_i}} \beta_i\right) \max_{\substack{i\in aq\pm t\\ \beta_i\ge\alpha_i}} \left(1 - \frac{\alpha_i}{\beta_i}\right) \\
&\le\; 2r + 2 - \min_{i\in aq\pm t} \left(\frac{\beta_i}{\alpha_i}\right) - \min_{i\in aq\pm t} \left(\frac{\alpha_i}{\beta_i}\right)\,,
\end{aligned}$$

where the last line follows because $\sum_i \alpha_i = \sum_i \beta_i = 1$ and α_i, β_i are positive.

For $i \in aq \pm t$, we have $a - i \in \gamma a \mp t$, and thus

$$\begin{aligned}
\alpha_i/\beta_i \;&=\; \frac{a!(a-w-i)!}{(a-i)!(a-w)!}(1-q)^w \\
&=\; \frac{a(a-1)\ldots(a-w+1)}{(a-i)(a-i-1)\ldots(a-w-i+1)}(1-q)^w \\
&\ge\; \left(\frac{a\gamma}{a-i}\right)^w \;\ge\; \left(\frac{a\gamma}{\gamma a + t}\right)^w \;=\; \left(1 - \frac{t}{\gamma a + t}\right)^w \;\ge\; 1 - \frac{wt}{\gamma a + t}\,,
\end{aligned}$$

and

$$\begin{aligned}
\beta_i/\alpha_i &= \frac{(a-w)!(a-i)!}{(a-w-i)!a!}(1-q)^{-w} \\
&= \frac{(a-i)(a-i-1)\dots(a-w-i+1)}{a(a-1)\dots(a-w+1)}(1-q)^{-w} \\
&\geq \left(\frac{a-w-i}{\gamma a}\right)^w \geq \left(\frac{\gamma a-w-t}{\gamma a}\right)^w = \left(1-\frac{w+t}{\gamma a}\right)^w \geq 1-\frac{w^2+tw}{\gamma a}.
\end{aligned}$$

Therefore,

$$\mathrm{D_{TV}}\left(\mathcal{B}(a,q),\mathcal{B}(a-w,q)\right) \leq 2r+\frac{wt}{\gamma a+t}+\frac{w^2+tw}{\gamma a} = O(1)\cdot r\sqrt{\ln(2/r)},$$

where the last line follows since $w^2 \leq tw = 4w\sqrt{\ln(2/r)\cdot v} \leq 4rv\sqrt{\ln(2/r)} < 4r\gamma a\sqrt{\ln(2/r)}$. □

References

[1] FARID M. ABLAYEV: Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoret. Comput. Sci.*, 157(2):139–159, 1996. Preliminary version in ICALP'93. [doi:10.1016/0304-3975(95)00157-3] 7, 10, 23

[2] ALFRED V. AHO, JEFFREY D. ULLMAN, AND MIHALIS YANNAKAKIS: On notions of information transfer in VLSI circuits. In *Proc. 15th STOC*, pp. 133–139. ACM Press, 1983. [doi:10.1145/800061.808742] 2

[3] NOGA ALON, YOSSI MATIAS, AND MARIO SZEGEDY: The space complexity of approximating the frequency moments. *J. Comput. System Sci.*, 58(1):137–147, 1999. Preliminary version in STOC'96. [doi:10.1006/jcss.1997.1545] 2, 7, 24

[4] 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. Preliminary version in FOCS'02. [doi:10.1016/j.jcss.2003.11.006] 6, 7, 8

[5] ZIV BAR-YOSSEF, T. S. JAYRAM, RAVI KUMAR, D. SIVAKUMAR, AND LUCA TREVISAN: Counting distinct elements in a data stream. In *Proc. 6th Internat. Workshop on Randomization and Computation (RANDOM'02)*, volume 2483 of *LNCS*, pp. 1–10. Springer, 2002. [doi:10.1007/3-540-45726-7_1] 25

[6] AMIT CHAKRABARTI, GRAHAM CORMODE, AND ANDREW MCGREGOR: Robust lower bounds for communication and stream computation. In *Proc. 40th STOC*, pp. 641–650. ACM Press, 2008. [doi:10.1145/1374376.1374470] 1

[7] AMIT CHAKRABARTI, GRAHAM CORMODE, AND ANDREW MCGREGOR: A near-optimal algorithm for computing the entropy of a stream. *ACM Trans. Algorithms*, 6:51:1–51:21, 2010. Preliminary version in SODA'07. [doi:10.1145/1798596.1798604] 25

[8] AMIT CHAKRABARTI, T. S. JAYRAM, AND MIHAI PĂTRAŞCU: Tight lower bounds for selection in randomly ordered streams. In *Proc. 19th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'08)*, pp. 720–729. ACM Press, 2008. ACM DL. 3, 4, 10, 26

[9] AMIT CHAKRABARTI, SUBHASH KHOT, AND XIAODONG SUN: Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proc. 18th IEEE Conf. on Computational Complexity (CCC'03)*, pp. 107–117. IEEE Comp. Soc. Press, 2003. [doi:10.1109/CCC.2003.1214414] 6, 7, 24

[10] AMIT CHAKRABARTI AND ODED REGEV: An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012. Preliminary versions in STOC'11 and ECCC. [doi:10.1137/120861072, arXiv:1009.3460] 20, 21

[11] 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. Press, 2004. [doi:10.1109/SFCS.2001.959901] 6

[12] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory (2nd ed.)*. Wiley-Interscience, 2006. 4

[13] ERIK D. DEMAINE, ALEJANDRO LÓPEZ-ORTIZ, AND J. IAN MUNRO: Frequency estimation of internet packet streams with limited space. In *Proc. 10th Ann. European Symp. on Algorithms (ESA'02)*, volume 2461 of *LNCS*, pp. 348–360. Springer, 2002. [doi:10.1007/3-540-45749-6_33] 3

[14] JEFF EDMONDS AND RUSSELL IMPAGLIAZZO: Towards time-space lower bounds on branching programs. Unpublished manuscript, 1993. Available on author's website. 2

[15] JOAN FEIGENBAUM, SAMPATH KANNAN, ANDREW MCGREGOR, SIDDHARTH SURI, AND JIAN ZHANG: Graph distances in the data-stream model. *SIAM J. Comput.*, 38(5):1709–1727, 2008. Preliminary version in SODA'05. [doi:10.1137/070683155] 26

[16] JOAN FEIGENBAUM, SAMPATH KANNAN, MARTIN STRAUSS, AND MAHESH VISWANATHAN: An approximate L^1 difference algorithm for massive data streams. *SIAM J. Comput.*, 32(1):131–151, 2002. Preliminary version in FOCS'99. [doi:10.1137/S0097539799361701] 2

[17] PHILIPPE FLAJOLET AND G. NIGEL MARTIN: Probabilistic counting algorithms for data base applications. *J. Comput. System Sci.*, 31(2):182–209, 1985. [doi:10.1016/0022-0000(85)90041-8] 25

[18] SUMIT GANGULY: Counting distinct items over update streams. *Theoret. Comput. Sci.*, 378(3):211–222, 2007. Preliminary version in ISAAC'05. [doi:10.1016/j.tcs.2007.02.031] 25

[19] MICHAEL B. GREENWALD AND SANJEEV KHANNA: Space-efficient online computation of quantile summaries. In *ACM SIGMOD Internat. Conf. on Managment of Data*, pp. 58–66. ACM Press, 2001. [doi:10.1145/375663.375670] 26

[20] ANDRE GRONEMEIER: Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness. In *Proc. 26th Symp. Theoretical Aspects of Comp. Sci. (STACS'09)*, volume 3 of *LIPIcs*, pp. 505–516. Dagstuhl Publishing, 2009. [doi:10.4230/LIPIcs.STACS.2009.1846, arXiv:0902.1609] 7, 8, 9

[21] SUDIPTO GUHA, PIOTR INDYK, AND ANDREW MCGREGOR: Sketching information divergences. *Machine Learning*, 72(1):5–19, 2008. [doi:10.1007/s10994-008-5054-x] 27

[22] SUDIPTO GUHA AND ANDREW MCGREGOR: Space-efficient sampling. In *Proc. 11th Internat. Conf. on Artificial Intelligence and Stat. (AISTATS'07)*, volume 2 of *JMLR Proceedings*, pp. 171–178. JMLR, 2007. JMLR.org. 3, 24

[23] SUDIPTO GUHA AND ANDREW MCGREGOR: Stream order and order statistics: Quantile estimation in random-order streams. *SIAM J. Comput.*, 38(5):2044–2059, 2009. Preliminary version in ICALP'07. [doi:10.1137/07069328X] 3, 10, 11, 26

[24] SUDIPTO GUHA, ANDREW MCGREGOR, AND SURESH VENKATASUBRAMANIAN: Streaming and sublinear approximation of entropy and information distances. *ACM Trans. Algorithms*, 5(4):35:1–35:16, 2009. Preliminary version in SODA'06. [doi:10.1145/1597036.1597038, arXiv:cs/0508122] 3

[25] NICHOLAS J. A. HARVEY, JELANI NELSON, AND KRZYSZTOF ONAK: Sketching and streaming entropy via approximation theory. In *Proc. 49th FOCS*, pp. 489–498. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.76] 25

[26] MONIKA RAUCH HENZINGER, PRABHAKAR RAGHAVAN, AND SRIDHAR RAJAGOPALAN: Computing on data streams. In JAMES M. ABELLO AND JEFFREY SCOTT VITTER, editors, *External Memory Algorithms*, volume 50 of *DIMACS Ser. in Discr. Math. and Theoret. Comput. Sci.*, pp. 107–118. Amer. Math. Soc., 1999. ACM DL. 2, 26

[27] PIOTR INDYK AND DAVID P. WOODRUFF: Tight lower bounds for the distinct elements problem. In *Proc. 44th FOCS*, pp. 283–288. IEEE Comp. Soc. Press, 2003. [doi:10.1109/SFCS.2003.1238202] 20

[28] PIOTR INDYK AND DAVID P. WOODRUFF: Optimal approximations of the frequency moments of data streams. In *Proc. 37th STOC*, pp. 202–208. ACM Press, 2005. [doi:10.1145/1060590.1060621] 24

[29] T. S. JAYRAM, RAVI KUMAR, AND D. SIVAKUMAR: Two applications of information complexity. In *Proc. 35th STOC*, pp. 673–682. ACM Press, 2003. [doi:10.1145/780542.780640] 6

[30] T. S. JAYRAM, RAVI KUMAR, AND D. SIVAKUMAR: The one-way communication complexity of Hamming distance. *Theory of Computing*, 4(6):129–135, 2008. [doi:10.4086/toc.2008.v004a006] 20

[31] HARTMUT KLAUCK, ASHWIN NAYAK, AMNON TA-SHMA, AND DAVID ZUCKERMAN: Interaction in quantum communication and the complexity of set disjointness. *IEEE Trans. Inform. Theory*, 53(6):1970–1982, 2007. Preliminary version in STOC'01. [doi:10.1109/TIT.2007.896888] 18

[32] EYAL KUSHILEVITZ AND NOAM NISAN: *Communication Complexity*. Cambridge Univ. Press, 1997. 2, 5

[33] TAK WAH LAM AND WALTER L. RUZZO: Results on communication complexity classes. *J. Comput. System Sci.*, 44(2):324–342, 1992. Preliminary version in SCT'89. [doi:10.1016/0022-0000(92)90025-E] 2

[34] PETER BRO MILTERSEN, NOAM NISAN, SHMUEL SAFRA, AND AVI WIGDERSON: On data structures and asymmetric communication complexity. *J. Comput. System Sci.*, 57(1):37–49, 1998. Preliminary version in STOC'95. [doi:10.1006/jcss.1998.1577] 13

[35] J. IAN MUNRO AND MIKE PATERSON: Selection and sorting with limited storage. *Theoret. Comput. Sci.*, 12(3):315–323, 1980. Preliminary version in FOCS'78. [doi:10.1016/0304-3975(80)90061-4] 3, 26

[36] 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] 2

[37] PAVEL PUDLÁK AND JIŘÍ SGALL: An upper bound for a communication game related to time-space tradeoffs. In *The Mathematics of Paul Erdős I*, pp. 399–407. Springer, 2013. Preliminary version (1995) in ECCC. [doi:10.1007/978-1-4614-7258-2_24] 2

[38] PRANAB SEN AND SRINIVASAN VENKATESH: Lower bounds for predecessor searching in the cell probe model. *J. Comput. System Sci.*, 74(3):364–385, 2008. Preliminary version in CCC'03. [doi:10.1016/j.jcss.2007.06.016, arXiv:cs/0309033] 13

[39] ALEXANDER A. SHERSTOV: The communication complexity of gap Hamming distance. *Theory of Computing*, 8(8):197–208, 2012. Preliminary version in ECCC. [doi:10.4086/toc.2012.v008a008] 20

[40] MIKLÓS SIMONOVITS: Extremal graph theory. In LOWELL W. BEINEKE AND ROBIN J. WILSON, editors, *Selected Topics in Graph Theory*, volume 2, pp. 161–200. Academic Press, 1983. 27

[41] THOMAS VIDICK: A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-Hamming-distance problem. *Chicago J. Theor. Comput. Sci.*, 2012(1), 2012. Preliminary version in ECCC. [doi:10.4086/cjtcs.2012.001] 20

[42] DAVID P. WOODRUFF: Optimal space lower bounds for all frequency moments. In *Proc. 15th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'04)*, pp. 167–175. ACM Press, 2004. ACM DL. 20, 25

[43] DAVID P. WOODRUFF: The average-case complexity of counting distinct elements. In *Proc. 13th Internat. Conf. on Database Theory (ICDT'09)*, volume 361 of *ACM Internat. Conf. Proc. Series*, pp. 284–295. ACM Press, 2009. [doi:10.1145/1514894.1514928] 3

[44] ANDREW CHI-CHIH YAO: Some complexity questions related to distributive computing (preliminary report). In *Proc. 11th STOC*, pp. 209–213. ACM Press, 1979. [doi:10.1145/800135.804414] 2

AUTHORS

Amit Chakrabarti
Dartmouth College, NH
ac@cs.dartmouth.edu
http://www.cs.dartmouth.edu/~ac/

Graham Cormode
University of Warwick, UK
g.cormode@warwick.ac.uk
http://dimacs.rutgers.edu/~graham/

Andrew McGregor
University of Massachusetts, Amherst, MA
mcgregor@cs.umass.edu
https://people.cs.umass.edu/~mcgregor/

ABOUT THE AUTHORS

AMIT CHAKRABARTI grew up in Bangalore, received a B. Tech. from IIT Bombay, and a Ph. D. from Princeton University. While broadly interested in theoretical computer science, he is especially fascinated by impossibility phenomena in algorithms and mathematics. During his high school days, he was known to cite Abel's theorem on algebraically solving the general quintic as his favourite mathematical result. This has not changed yet. He enjoys playing Scrabble at the international tournament level, and various other games with his son at a sub-amateur level.

GRAHAM CORMODE grew up in England, received a B. A. from the University of Cambridge, and a Ph. D. from the University of Warwick. His research interests within computer science span algorithms, databases, networks, and security. In his vanishingly small spare time, he likes to solve mysteries.

ANDREW MCGREGOR grew up in Scotland, received a B. A. and M. Math from the University of Cambridge, and a Ph. D. from the University of Pennsylvania. He is interested in a broad range of problems in theoretical computer science and information theory. He sometimes cares about constants but mainly when they are in the exponent. He once played Scrabble against an opponent of international renown; who won will be left as a mystery.