Robust Lower Bounds for Communication and Stream Computation

Amit Chakrabarti[*]   Graham Cormode[†]   Andrew McGregor[‡]

April 5, 2011

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 [34, 37]. 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.

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 [8, 24] 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.

[*]Dartmouth College. `ac@cs.dartmouth.edu`. Work supported by an NSF CAREER award and by Dartmouth College startup funds.

[†]AT&T Labs–Research. `graham@research.att.com`

[‡]University of Massachusetts, Amherst. `mcgregor@cs.umass.edu`. Work supported by NSF CAREER Award CCF-0953754.

1   Introduction

Since its introduction in 1979 by Yao, communication complexity [33, 43] has proven to be a powerful technique 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 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 can 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, 34, 37] and [33, 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 VSLI 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, 38]). 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, 16, 27]. 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 [8, 24]).

Random-order data streams were considered by Munro and Paterson [36] 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, 23–25, 42]. 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 consid-

erable difference between processing random-order stream and adversarial streams. In particular, streaming algorithms to find the median using polylog space require exponentially fewer passes if the stream is ordered randomly [8, 24].

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 much 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 Disjointess:* We consider the problem of t-party 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 randomized 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 t-ary trees of depth $p+1$ and show that if the bits of these strings are distributed uniformly between multiple players, we require about $\Omega(n^{(2+\varepsilon)^{-p}})$ bits of communication for a p-round protocol. See Section 4.

- *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 one-way 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 a wide variety of lower bounds for 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 kth frequency moment, we obtain a robust lower bound of $\Omega(n^{1-3/k})$, which comes close to the optimal $\Omega(n^{1-2/k})$ bound under adversarial ordering. For the problem of median finding, our framework greatly simplifies the proof of a recent $\Omega(\log\log n)$ lower bound [8] on the number of passes required to achieve polylogarithmic space. Further, our pass-space tradeoff for this problem greatly improves the results of [8]: for instance, with two passes, we obtain a space lower bound of $\Omega(n^{1/10})$ as opposed to their $\Omega(n^{3/80})$.

Subsequent Developments: While our results for the Multi-Party Set Disjointess problem are in a sense tight, the lower bound for estimating the kth frequency moment of a randomly ordered stream has been improved since the preliminary announcement of our results in STOC 2008. Andoni et al. [4] showed an $\Omega(n^{1-2.5/k})$ lower bound and this was improved to $\Omega(n^{1-2/k})$ by Guha and Huang [21].

2 Notation and Preliminaries

We summarize some notation that we need repeatedly. 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: $\mathbb{E}[X]$ denotes the expectation and $H(X)$ the entropy of X, $H(X \mid Y)$ the conditional entropy of X given Y and $I(X : Y)$ 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 $\mu \otimes \nu$.

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 $D_{TV}(\mu, \nu) = \frac{1}{2}\|\mu - \nu\|_1$, the Hellinger distance $h(\mu, \nu) = \frac{1}{\sqrt{2}}\|\sqrt{\mu} - \sqrt{\nu}\|_2$, where "$\sqrt{\cdot}$" denotes the pointwise positive square root, and the Kullback-Leibler divergence $D_{KL}(\mu\|\nu)$, 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 $\mathscr{B}(n, p)$. The notation $X \in_R S$ indicates that X is chosen uniformly at random from the set S. 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 Q' is an (ε, δ)-approximation for Q if $\Pr[|Q' - Q| > \varepsilon Q] \leq \delta$.

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 ith 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 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 $\mathscr{U}_p$ 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 randomized, 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 [33].) For a random-allocation protocol P, let $P(x, \sigma)$ denote the (possibly random) *transcript* of P, and $\text{out}(P, x, \sigma)$ 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, v). We define the error

$$\text{err}(P, f, v) \; := \; \max_x \Pr[\text{out}(P, x, S) \neq f(x)],$$

where the probability is taken over $S \sim v$ 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

$$\text{err}_\mu(P, f, v) \; := \; \Pr[\text{out}(P, X, S) \neq f(X)],$$

where $X \sim \mu$ and $S \sim v$. Let $\text{cost}(P) := \max_{x, \sigma} |P(x, \sigma)|$ denote the communication cost of P. We define the δ-error communication complexity of (f, v) to be

$$R_\delta(f, v) \; := \; \min\{\text{cost}(P) : \text{err}(P, f, v) \leq \delta\}$$

and the δ-error μ-distributional complexity to be

$$R_{\mu, \delta}(f, v) \; := \; \min\{\text{cost}(P) : \text{err}_\mu(P, f, v) \leq \delta\}.$$

Let $R^\rightarrow$ 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, we drop v from these notations.

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

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 [6, 9, 11, 30]. 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 extend the methods of Bar-Yossef et al. [6] to handle public coin protocols. 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). For a traditional private coin protocol P and a distribution μ on its inputs, we define

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

If D is a random variable (possibly correlated with X), we define the D-conditional μ-information cost

$$\text{icost}_\mu(P \mid D) \; := \; I(X : P(X) \mid D).$$

5

We extend these notions to public coin protocols thus: if P^R is a public coin protocol that uses a public random string R, we define

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

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

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

We write IC^{pub} for the information complexity of public coin protocols.

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 partition $\sigma \sim \nu$. 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 $\Pr_{(x,\sigma)\sim\xi}[\text{out}(P, x, \sigma) \ne f(x)] \le \delta$, for some joint distribution ξ, then*

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

Proof. Simply observe that

$$\text{err}_\mu(P, f, \nu) \ = \ \Pr_{x\sim\mu,\sigma\sim\nu}[\text{out}(P, x, \sigma) \ne f(x)] \ \le \ \Pr_{(x,\sigma)\sim\xi}[\text{out}(P, x, \sigma) \ne f(x)] + \text{D}_{\text{TV}}(\mu \otimes \nu, \xi). \qquad \square$$

2.3 Preliminary Lemmas

We collect together a couple of 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 fact.

Fact 2.5. *Let $n > 0$ be an integer and $\delta \in (0, \frac{1}{2})$ be a real number. Then $\text{R}^{\text{pub},\rightarrow}(\text{INDEX}) \ge (1 - \text{H}_b(\delta))n$, where $\text{H}_b(x) := -x\log x - (1-x)\log(1-x)$ is the binary entropy function.*

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

Lemma 2.6. *For all $q \in [1/2, 1)$, there exist constants $c_1, c_2 > 0$ such that, for $a \in \mathbb{N}$ sufficiently large and any $w \in [a]$,*

$$\text{D}_{\text{TV}}(\mathscr{B}(a, q), \mathscr{B}(a - w, q)) \ \le \ c_1 w \sqrt{\frac{\ln a}{(1 - q)a}},$$

and for $q = 1/2 + \delta$,

$$\text{D}_{\text{TV}}(\mathscr{B}(a, 1/2), \mathscr{B}(a, q)) \ \le \ c_2 \delta^2 a.$$

6

3 Multi-Party Set Disjointness

Let $\text{DISJ}_{n,t} : \{0,1\}^{nt} \to \{0,1\}$ 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 [3,6,9] on this problem, showing that $R_\delta(\text{DISJ}_{n,t}) = \Omega(n/t)$ 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 $x = (x_1,\ldots,x_t)$ 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. [6] to link this quantity with $\text{IC}_{\mu^n,\delta}(\text{DISJ}_{n,t} \mid D^n)$.

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

Fact 3.1 (Birthday problem). *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*

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

Lemma 3.2. *Let $\delta' = \delta + \alpha(t,p)$. Then*

$$R_\delta(\text{DISJ}_{n,t}, \mathscr{U}_p) \geq n \cdot \text{IC}^{\text{pub}}_{\mu,\delta'}(\text{AND}_t \mid D).$$

Proof. Let P be an optimal δ-error protocol for $(\text{DISJ}_{n,t}, \mathscr{U}_p)$, i.e., a protocol that achieves $\text{cost}(P) = R_\delta(\text{DISJ}_{n,t}, \mathscr{U}_p)$. Consider n independent pairs of random variables $(X_1, D_1),\ldots,(X_n, D_n)$, each drawn from λ. Then $X := X_1 X_2 \ldots X_n \sim \mu^n$ is a suitable random input for $\text{DISJ}_{n,t}$. Let $S \sim \mathscr{U}_p$ be a random split. Then, by standard information theoretic arguments, we have

$$\begin{aligned}
\text{cost}(P) = \max_{x,\sigma} |P(x,\sigma)| &\geq \text{H}(P(X,S)) \\
&\geq \text{I}(X : P(X,S) \mid D_1 D_2 \ldots D_n, S) \\
&\geq \sum_{j\in[n]} \text{I}(X_j : P(X,S) \mid D_1 D_2 \ldots D_n, S) \qquad (1)\\
&= \sum_{j\in[n]} \mathbb{E}_d[\text{I}(X_j : P(X,S) \mid D_j, S, D_{-j} = d)],
\end{aligned}$$

where (1) holds because the X_js are independent even after conditioning on $D_1 D_2 \ldots D_n$ and S. Here, D_{-j} denotes the vector $(D_1,\ldots,D_{j-1}, D_{j+1},\ldots,D_n)$ and the final expectation is over d drawn uniformly from $[t]^{[n]\setminus\{j\}}$. To finish the proof, it suffices to show that

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

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 $Q^S_{j,d}$ for AND_t, parametrised by j and d, that uses S as a public random string. Further, for each

7

possible value σ of S, the transcript $Q^{\sigma}_{j,d}(X_j)$ is either constant or distributed identically to $(P(X,\sigma) \mid D_{-j} = d)$. Then, as required, we shall have

$$\mathrm{IC}^{\mathrm{pub}}_{\mu,\delta'}(\mathrm{AND}_t \mid D) \leq \mathrm{icost}^{\mathrm{pub}}_{\mu}(Q^S_{j,d} \mid D_j) = \mathrm{I}(X_j : Q^S_{j,d}(X_j) \mid D_j, S) \leq c_{j,d}.$$

The protocol $Q^{\sigma}_{j,d}$ works as follows. On input $x = (x_1, \ldots, x_t) \in \{0,1\}^t$, the players create a random virtual input $\{Z_{ik}\}_{i,k} \in \{0,1\}^{t \times n}$ for $\mathrm{DISJ}_{n,t}$, 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 on the virtual input. The virtual input is obtained by embedding x into the jth column of a random Boolean matrix drawn from $(\mu^n \mid D_{-j} = d)$. 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 players abort, 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 $Q^{\sigma}_{j,d}(X_j) \equiv (P(X,\sigma) \mid D_{-j} = d)$. Finally, $Q^S_{j,d}$ is indeed a δ'-error protocol, because

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

Lemma 3.3. *If $\delta \leq 1/20$, then $\mathrm{IC}^{\mathrm{pub}}_{\mu,\delta}(\mathrm{AND}_t \mid D) = \Omega(1/t)$.*

Proof. From the work of Gronemeier [20] we can deduce that for a *private* coin traditional protocol P such that $\mathrm{err}(P, \mathrm{AND}_t) \leq 1/10$, we have $\mathrm{icost}_{\mu}(P \mid D) = \Omega(1/t)$. Now, consider a public coin δ-error protocol Q^S for AND_t that uses a public random string S. For each possible value σ of S, define $c_{\sigma} := \mathrm{icost}_{\mu}(Q^{\sigma} \mid D)$, so that $\mathbb{E}_{\sigma}[c_{\sigma}] = \mathrm{icost}^{\mathrm{pub}}_{\mu}(Q^S \mid D)$ and $\mathbb{E}_{\sigma}[\mathrm{err}(Q^{\sigma}, \mathrm{AND}_t)] \leq \delta$.

Suppose $\delta \leq 1/20$. Call a particular split σ "good" if

$$\mathrm{err}(Q^{\sigma}, \mathrm{AND}_t) \leq 2\delta \leq 1/10.$$

By Markov's inequality, $\Pr[\sigma \text{ is good}] \geq 1/2$. For each good σ, considering the private coin protocol Q^{σ} shows $c_{\sigma} = \Omega(1/(t \log t))$. Thus, $\mathbb{E}_{\sigma}[c_{\sigma}] = \Omega(1/(t \log t))$. We conclude that

$$\mathrm{IC}^{\mathrm{pub}}_{\mu,\delta}(\mathrm{AND}_t \mid D) = \Omega(1/t). \qquad \square$$

Putting together Fact 3.1, Lemma 3.2 and Lemma 3.3 yields the following theorem.

Theorem 3.4. *For $\delta \leq 1/40$ and $p \geq 20t^2$, we have the robust lower bounds $\mathrm{R}_{\delta}(\mathrm{DISJ}_{n,t}, \mathcal{U}_p) = \Omega(n/t)$.*

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 $\Omega(t^2)$. 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 Fact 3.1, 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 $\text{TPJ}_{k,t}$, 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. 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 $\text{TPJ}_{k,t}(\phi) := g^{(k)}(v_0) = g(g(\cdots g(v_0)\cdots)) \in \{0, 1\}$.

There are at least two natural ways to make a traditional communication problem out of $\text{TPJ}_{k,t}$, 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; we use the convention 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 $\text{TPJ}_{k,t}$ to denote the former, and $\text{M-TPJ}_{k,t}$ 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 [24] 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 recently extended to multiple passes by Chakrabarti, Jayram and Pătrașcu [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 proof of a multi-pass streaming lower bound for median finding,[1] and in fact improve upon previous bounds. The five theorems in this section can be organized 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 $\text{M-TPJ}_{k,t}$, as in Theorem 4.4 or $\text{TPJ}_{k,t}$, as in Theorem 4.9.

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 $\text{R}^k_{\mu,\delta}(f, v)$, 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.

---

[1]Our results, like the earlier ones [8, 24], apply to the more general problem of selection.

The next three subsections are organized thus. We first present the Stage 3 reduction, then the Stage 1 and Stage 2 theorems for the implication chain that starts with M-TPJ$_{k,t}$, and then deal with the chain that starts from 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 W-TPJ$_{k,n}$ mentioned above. It is closely related to TPJ$_{k,t}$ and 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 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:

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

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 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 Eq. (2) is feasible. Let $x \in \{0,1\}^n$ be the concatenation of all the strings x_v. We then define the partial function W-TPJ$_{k,n}(x) := $ 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 [24].

Theorem 4.3. *Let* MEDIAN$_{m,N}$ *denote the random-partition 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* $R_\delta^p(\text{MEDIAN}_{n,\Theta(n)}, \nu) \geq R_\delta^p(\text{W-TPJ}_{k,n}, \nu)$. $\square$

Proof. 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 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 ith and jth 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 Eq. (2), ensure that

$$\text{median}\left(\bigcup_{v \in V(T)} A(v) \right) \equiv \text{W-TPJ}(x) \pmod 2;$$

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 $\alpha(v) = \langle i_k, \ldots, i_{j+1}, 0, 0, \ldots, 0 \rangle_B$ and $\beta(v) = \langle i_k, \ldots, i_{j+1}, t+1, 0, \ldots 0 \rangle_B$, 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 $\alpha(v) = \langle i_k, \ldots, i_2, 0 \rangle_B$ and $\beta(v) = \langle i_k, \ldots, i_2, 1 \rangle_B$. One can easily verify that this construction has the properties claimed. $\square$

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 each fixed $p \geq 1$, we have $R^p_{\mu_{p+1}, 1/3}(\text{M-TPJ}_{p+1,t}) = \Omega(t/p^2)$.*

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. [35] and Sen [39], 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.

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

As a precursor to our round elimination lemma, we prove the following multi-round analogue of a lemma of Sen [39, 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*

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

and where each round costs at most c.

Proof. By a standard transformation, 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. 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

$$g(x, m) \;=\; \Pr[\text{out}(Q_m, x) \neq f(x)], \tag{3}$$

11

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 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

$$\mathrm{D_{TV}}\left(\lambda, \mu \otimes \beta\right) \;\leq\; \sqrt{\frac{\ln 2}{2} \cdot \mathrm{D_{KL}}(\lambda \| \mu \otimes \beta)} \;=\; \sqrt{\frac{\ln 2}{2} \cdot \mathrm{I}(X : M)} \;=\; \sqrt{\frac{\ln 2}{2} \cdot \mathrm{icost}_\mu^1(P)}, \qquad (4)$$

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 Eq. (3) in two ways:

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

Thus, we have

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

where the first inequality holds because $|g(x, m)| \leq 1$ for all x amd m, and the second inequality uses Eq. (4). Choose m to minimize $\mathrm{err}_\mu(Q_m, f)$, and fix the random strings used by the players in Q_m so as to minimize the μ-distributional error of the resulting deterministic protocol, Q. Then $\mathrm{err}_\mu(Q, f)$ is upper-bounded as desired. $\qquad\square$

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

Proof. Let P be a protocol whose existence is asserted by $\mathscr{A}(p, K, \varepsilon)$. 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

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

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

$$\frac{1}{t}\sum_{i=1}^{t}\mathrm{err}_{\mu_p}(Q_i,\text{M-TPJ}_{p,t}) \;=\; \frac{1}{t}\sum_{i=1}^{t}\mathrm{err}_{\xi_i}(P,\text{M-TPJ}_{p+1,t}) \;=\; \mathrm{err}_{\mu_{p+1}}(P,\text{M-TPJ}_{p+1,t}) \;\le\; \varepsilon. \tag{5}$$

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

$$\frac{t}{(Kp)^2} \;\ge\; |M| \;\ge\; \mathrm{I}(X:M) \;\ge\; \sum_{i=1}^{t}\mathrm{I}(X_i:M) \;=\; \sum_{i=1}^{t}\mathrm{icost}_{\mu_p}^1(Q_i), \tag{6}$$

where the rightmost inequality uses the independence of $\{X_i\}_{i\in[t]}$. Combining (5) and (6), we have

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

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

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

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

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

Proof of Theorem 4.4. Suppose that $\mathrm{R}_{\mu_{p+1},1/3}^{p}(\text{M-TPJ}_{p+1,t})$ is not lower bounded as stated. Specifically, using a standard error-reduction argument, we may assume that $\mathrm{R}_{\mu_{p+1},1/6}^{p}(\text{M-TPJ}_{p+1,t}) \le t/(6p)^2$. By the easy direction of Yao's minimax lemma, we have $\mathscr{A}(p,6,1/6)$, where the predicate $\mathscr{A}$ is as defined in Lemma 4.7. Applying that lemma repeatedly, we conclude $\mathscr{A}(1,6,1/6+(p-1)/6p)$, which implies $\mathscr{A}(1,6,1/3)$. Notice that M-TPJ$_{2,t}$ is just the INDEX problem with a t-bit input, and 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 $t/36$. Since $\mathrm{H}_b(1/3) < 12/13$, this contradicts Fact 2.5. $\qquad\square$

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 $\mathscr{V}_{p+1}$ be the (non-uniform) split distribution that gives each token to Player 1 with probability $\frac{1}{2}$ and to Player i with probability $\gamma := 1/(2p)$ for each $i \in \{2,\ldots,p+1\}$. Then, we have*

$$\mathrm{R}_{1/24}^{p}(\text{W-TPJ}_{p+1,n},\mathscr{V}_{p+1}) \;=\; \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \cdot (\log n)^{\frac{-1}{2(p-1)}} \cdot p^{-2}\right).$$

Thus, for any constant $\varepsilon > 0$, for n and p large enough with $p = O(\log\log n)$, we have

$$\mathrm{R}_{1/24}^{p}(\text{W-TPJ}_{p+1,n},\mathscr{V}_{p+1}) \;=\; \Omega\left(n^{(2+\varepsilon)^{-p}}\right).$$

Proof. Let P be a protocol for $(\text{W-TPJ}, \mathcal{V}_{p+1})$ such that $\text{err}(P, \text{W-TPJ}, \mathcal{V}_{p+1}) \leq \frac{1}{24}$. We will use P to construct a protocol Q for M-TPJ such that $\text{err}_\mu(Q, \text{M-TPJ}_{p+1,t}) \leq 1/3$, 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 $\text{err}_{\mu_{p+1}}(Q, \text{M-TPJ}_{p+1,t}) \leq 1/3$.

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. We put

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

for some large constant c to be determined. For each node v, the players use the following *public coin* randomized 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 $d_{1v} \sim \mathscr{B}\left(\frac{a_i}{2}, 1-\gamma\right)$ and $d_{0v} \sim \mathscr{B}\left(\frac{a_i}{2}, 1-\gamma\right)$, 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 Eq. (2). 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 $d_{1v} \leq |x_v| \leq a_i - d_{0v}$; 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 $\frac{1}{2}$ 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

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

It remains to show that x and σ 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 $\mathscr{B}(|x_v|, 1-\gamma)$ and $\mathscr{B}(a_i - |x_v|, 1-\gamma)$, 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

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

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 2.6, we can bound the total variation distance

in terms of a_i and b_i as follows,

$$\text{D}_{\text{TV}}\left(\bigotimes_v(\tilde{A}_v \otimes \tilde{B}_v), \bigotimes_v(A_v \otimes B_v)\right) \leq \sum_v \text{D}_{\text{TV}}\left(\tilde{A}_v, A_v\right) + \sum_v \text{D}_{\text{TV}}\left(\tilde{B}_v, B_v\right) \leq O(\sqrt{\log n}) \sum_{i=2}^{p+1} \frac{t^{p+2-i} b_{i-1}}{\sqrt{a_i}},$$

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

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

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

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

Note that

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

and hence,

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

where the last line follows using the fact that $p^{\frac{1}{2(p-1)}} = O(1)$ for $p \geq 2$. $\qquad\square$

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 $\text{TPJ}_{k,t}$ 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. [32], who in fact studied the problem in the more general *quantum* communication setting. The underlying intuition is, once again, round elimination.

Theorem 4.9. *We have* $\text{R}^p_{\mu,1/3}(\text{TPJ}_{p+1,t}) = \Omega(t/p^2)$, *where* μ *is the uniform distribution over inputs.* $\qquad\square$

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. *We have*

$$\text{R}^p_{1/24}(\text{W-TPJ}_{p+1,n}, \mathcal{U}_2) = \Omega\left(n^{\frac{1}{(p-1)2^{p+1}+2}} \cdot (\log n)^{\frac{-1}{2(p-1)}} \cdot p^{-2}\right).$$

Thus, for any constant $\varepsilon > 0$, for n and p large enough with $p = O(\log\log n)$, we have

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

15

Proof. Let P be a protocol for $(\text{W-TPJ}, \mathcal{U}_2)$ such that $\text{err}(P, \text{W-TPJ}, \mathcal{U}_2) \leq \frac{1}{24}$. 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

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

for some large constant c. For each node v, the players use the following *public coin* randomized 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 $d_{1v} \sim \mathscr{B}\left(\frac{a_i}{2}, 1/2\right)$ and $d_{0v} \sim \mathscr{B}\left(\frac{a_i}{2}, 1/2\right)$, 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 Eq. (2). 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 $\{x_{v,k} : k \in S_v\}$ to 1 and the remaining bits to 0. Note that this requires $d_{1v} \leq |x_v| \leq a_i - d_{0v}$; 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

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

It remains to show that x and σ 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 $\mathscr{B}(|x_v|, 1/2)$ and $\mathscr{B}(a_i - |x_v|, 1/2)$, 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

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

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 2.6, we can bound the total variation distance in terms of a_i and b_i as follows,

$$\text{D}_{\text{TV}}\left(\bigotimes_v (\tilde{A}_v \otimes \tilde{B}_v), \bigotimes_v (A_v \otimes B_v)\right) \leq \sum_v \text{D}_{\text{TV}}(\tilde{A}_v, A_v) + \sum_v \text{D}_{\text{TV}}(\tilde{B}_v, B_v) \leq O(\sqrt{\log n}) \sum_{i=2}^{p+1} \frac{t^{p+2-i} b_{i-1}}{\sqrt{a_i}}$$

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

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

16

Therefore, by Theorem 4.9,

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

Note that

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

and hence,

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

5 Hamming Distance and Index

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

5.1 Hamming Distance

The GAP-HAMMING-DISTANCE problem (henceforth, GHD) was first formally stated in the context of data stream lower bounds [28, 31, 41]: 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

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

For $G \in \mathbb{R}^+$, we then define

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

where "$\star$" can be interpreted as "undefined." Equivalently, a computation problem corresponding to the function GHD_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 $R(\text{GHD}_G(x)) = \Omega(\min\{n, n^2/G^2\})$.

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. The main idea is to create an instance of GHD in the fixed partition model, and then pad this with carefully chosen random bits so that the resulting split appears almost uniform.

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

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

Proof. We reduce the traditional one-way INDEX problem to our GHD problem. Suppose Alice holds a string $x \in \{0,1\}^{n'}$ with $n' = c_2 n$ and Bob holds $j \in [n']$, where $c_2 < 1$ will be a constant to be fixed later. By Fact 2.5, we know $R^{\text{pub},\to}_{0.49}(\text{INDEX}) = \Omega(n)$.

Suppose there exists a one-way protocol P such that

$$\text{err}_\mu(P, \text{GHD}_{c_3\sqrt{n}}, \mathscr{U}_2) \le 1/4,$$

where μ is the uniform distribution over inputs. Let $r \in_R \{-1,1\}^{n'}$ be determined by public random bits. Define the indicator random variables T_{i1} and T_{i2} for the events "$\sum_{i=1}^{n'} r_i x_i > 0$" and "$r_j > 0$," respectively. It can be shown (see [31] for a proof) that, for some constant $c_1 > 0$,

$$\Pr[T_{i1} = T_{i2}] = \begin{cases} 1/2 - c_1/\sqrt{n'}, & \text{if } \text{INDEX}(x,j) = 0, \\ 1/2 + c_1/\sqrt{n'}, & \text{if } \text{INDEX}(x,j) = 1. \end{cases}$$

The players now generate an instance y of GHD using shared randomness. They first pick a split $\sigma \sim \mathscr{U}_2$. For each i such that $\sigma(2i) \neq \sigma(2i-1)$, with probability $p = c_2^{1/4}$, the players set (y_{2i}, y_{2i-1}) based on T_{i1} and T_{i2}: since T_{i1} is known to Alice, she sets whichever input bit was allocated to her as T_{i1}, and Bob similarly uses T_{i2}. Otherwise, set $(y_{2i}, y_{2i-1}) \in_R \{0,1\}^2$. Define $\Delta = \Delta(y) = |\{i : y_{2i} \neq y_{2i-1}\}|$.

Claim 5.2. *For sufficiently small c_2,*

$$(x_j = 0) \Rightarrow \Pr\left[\frac{\Delta}{n} > \frac{1}{2} + \frac{c_1}{5\sqrt{n'}}\right] \geq 0.99, \quad \text{and} \quad (x_j = 1) \Rightarrow \Pr\left[\frac{\Delta}{n} < \frac{1}{2} - \frac{c_1}{5\sqrt{n'}}\right] \geq 0.99.$$

Proof. Let t be the number of times Alice and Bob insert bits from T into their constructed strings. Note that $\mathbb{E}[t] = pn$ and, by an application of the Chernoff bound, for sufficiently large n, we have $\Pr[t \leq np/2] \leq 1/1000$. Now,

$$x_j = 1 \implies \Pr\left[\frac{\Delta}{n} \leq \frac{1}{2} - \frac{c_1}{5\sqrt{n'}}\right] = \Pr\left[\frac{\Delta}{n} - \mathbb{E}\left[\frac{\Delta}{n}\right] \leq -\frac{c_1}{5\sqrt{n'}}\right] \leq \exp\left(\frac{-c_1^2}{25c_2^{1/4}}\right), \quad \text{and}$$

$$x_j = 0 \implies \Pr\left[\frac{\Delta}{n} \geq \frac{1}{2} - \frac{2c_1}{5\sqrt{n'}}\right] = \Pr\left[\frac{\Delta}{n} - \mathbb{E}\left[\frac{\Delta}{n}\right] \geq +\frac{c_1}{5\sqrt{n'}}\right] \leq \exp\left(\frac{-c_1^2}{25c_2^{1/4}}\right).$$

Hence the claim holds true for sufficiently small c_2. $\qquad\square$

While σ is not fully independent of y, it has sufficient independence, as shown by the following claim:

Claim 5.3. *For sufficiently small c_2, with probability at least $5/8$, P answers $\text{GHD}_{c_3\sqrt{n}}$ correctly on y.*

Proof. Let μ_p be the distribution over $y \in \{0,1\}^{2n}$. For $p = 0$ both y and the partition are uniformly and independently chosen. We argue that $|\mu_p - \mu_0| \leq 1/8$ for sufficiently small c_2 and so by Lemma 2.4, P would answer $\text{GHD}_{c_3\sqrt{n}}$ with probability at least $3/4 - 1/8 = 5/8$ as required.

Define $I = \{i : \sigma(2i) \neq \sigma(2i-1)\}$. For $i \notin I$, $(y_{2i-1}, y_{2i}) \in_R \{0,1\}^2$ under both μ_0 and μ_p. For $i \in I$, define the probability that a pair of bits differ as

$$q = \Pr_{\mu_p}[y_{2i} \neq y_{2i-1} \mid i \in I] = \frac{1}{2} - \frac{pc_1}{\sqrt{n'}}.$$

Then

$$|\mu_p - \mu_0| = \sum_y \left|\Pr_{\mu_p}[y] - \Pr_{\mu_0}[y]\right| \leq \sum_y \left|2^{-n}q^\Delta(1-q)^{n-\Delta} - 2^{-2n}\right| = \sum_{d \in [n]} \binom{n}{d}\left|q^d(1-q)^{n-d} - 2^{-n}\right|.$$

By appealing to Lemma 2.6, we can make this smaller than $1/8$ by choosing c_2 sufficiently small. $\qquad\square$

Hence, if $c_3\sqrt{n} \leq nc_1/(5\sqrt{n'})$, i.e., if $c_3 \geq c_1/(5\sqrt{c_2})$, then the desired robust linear lower bound on $R_{1/4}^{\rightarrow}(\text{GHD}_{c_3\sqrt{n}}, \mathscr{U}_2)$ must hold. For otherwise, by Claim 5.2, $\text{GHD}_{c_3\sqrt{n}}$ on y reveals $\text{INDEX}(x,j)$ with probability at least $5/8 - 1/100 > 51/100$. This completes the proof of Theorem 5.1. $\qquad\square$

5.2 Index

For our purposes, we define the INDEX problem over inputs $x \in [n] \times \{0,1\}^n$ as follows: $\text{INDEX}(x) := x_j$ where $j := x_0$. Traditionally, one considers the worst-case partition where Alice (the player who speaks) holds $x_1 \ldots x_n$ and Bob holds j. The resulting problem is one of the most basic in communication complexity, and strong randomized lower bounds are known for it in this setting [1]. In this fixed-partition model, INDEX 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 INDEX that allocates multiple copies of the tokens $(x_0, \ldots, x_n)$ amongst two players. This generalization is needed for proving subsequent data stream bounds. For positive integers a and b, let $\text{INDEX}^{a,b}$ 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 $x = (x_0, \ldots, x_n)$ being an input for INDEX. The (partial) function $\text{INDEX}^{a,b}$ takes the value $\text{INDEX}(x)$ 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 5.4. *Let a and b be constants. We have $\text{R}_\delta^{\rightarrow}(\text{INDEX}^{a,b}, \nu_p) = \Omega(n)$, where $\delta = (1-p)^b p^a / 4$.*

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

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

$$B_0 = \text{``a copy of } x_0 \text{ is allocated to Player 1''}, \text{ and}$$
$$B_i = \text{``a copy of } x_i \text{ is allocated to Player 2''}, \text{ for some } i \in [n].$$

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, $\text{cost}(Q) \leq \text{cost}(P)$.

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

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

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 random order yields an r-round, p-player communication protocol for evaluating $f(S)$ when S is randomly partitioned into p subsets $S_1, \ldots, S_p$ and the ith player receives S_i: the ith player randomly permutes S_i to generate stream s_i and the players emulate the algorithm on the concatenated stream $\langle s_1 | s_2 | \ldots | s_p \rangle$. The 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 kth frequency moment 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[2] and that this is tight under worst-case orderings [9, 29]. However, it was observed that for random orderings and $m = \tilde{\Omega}_\varepsilon(an)$ there exists a single pass $\tilde{O}((n/a)^{1-2/k})$-space algorithm that (ε, δ)-approximates F_k [23]. The following theorem shows a lower bound on the space usage in the random-order case. The proof combines Theorem 3.4 with a variation of the reduction used in [3, Theorem 3.2].

Theorem 6.1. *Any constant pass $(1/10, 1/10)$-approximation for F_k of a randomly ordered stream requires $\Omega(n^{1-3/k})$ space. If we assume that $m = \Omega(an)$ then $\Omega(n^{1-3/k}/a^3)$ 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. Let $x = \{x_{ij}\}_{i \in [t], j \in [n]}$ be an instance for $\text{DISJ}_{n,t}$ that satisfies the unique intersection promise. Set $t = (5n/4)^{1/k}$ and consider a uniform random split of the nt 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 multi-set of values generated by the jth player. Note that the sets $S_1, \ldots, S_p$ are a random partition of $S = S_1 \cup \ldots \cup S_p$. Furthermore $F_k(S) \geq t^k = 5n/4$ if $\text{DISJ}_{n,t}(x) = 1$ and $F_k(S) \leq n$ if $\text{DISJ}_{n,t}(x) = 0$. Using the template at the start of Section 6 and appealing to Theorem 3.4, we can deduce that $rps = \Omega(n/t)$ and therefore $s = \Omega(n^{1-3k})$ as required.

To prove the second part of the theorem, the reduction from $\text{DISJ}_{n,t}$ 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 $\text{DISJ}_{n,t}(x) = 1$, then $F_k \geq t^k$, but if $\text{DISJ}_{n,t}(x) = 1$, then $F_k \leq (a+1)^k n$. 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 $\Omega(n/t^3) = \Omega(n^{1-3/k}/a^3)$. $\qquad\square$

6.2 Distinct Elements and Entropy

The number of distinct elements in a stream is $F_0 := |\{i \in [n] : f_i \neq 0\}|$, and the empirical entropy is $H := \sum_{i \in [n]} (f_i/m) \log(m/f_i)$. One-pass, $\tilde{O}(\varepsilon^{-2})$-space, (ε, δ)-approximation algorithms are known for both

---

[2]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}.

problems [5, 7, 17, 18, 26]. 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 [41, Section 3.2].

Theorem 6.2. *Let $k \geq 0$ be a constant, and suppose $k \neq 1$. Then, a one-pass (ε, δ)-approximation for F_k of a randomly ordered stream requires $\Omega(\varepsilon^{-2})$ space. Also, a one-pass (ε, δ)-approximation for H of a randomly ordered stream requires $\Omega(\varepsilon^{-2}/\log^2 \varepsilon^{-1})$ space.*

Proof. Suppose there exists a single pass, $(1/10, 1/10)$-approximation algorithm for H that uses s bits of space. Let $x \in \{0,1\}^{2n}$ be an instance of GHD_G 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,

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

where $\Delta = \Delta(x) = |\{i \in [2n] : x_{2i} \neq x_{2i-1}\}|$. Hence, any algorithm which can (ε, δ)-approximate H can also distinguish the cases $\Delta(x) \leq n/2 - G$ and $\Delta(x) \geq n/2 + G$, provided $\varepsilon = O(G/(n\log n))$. For a fixed ε, we set $n = O(\varepsilon^{-2}/\log^2 \varepsilon^{-1})$ and $G = \Theta(\sqrt{n})$. This ensures $\varepsilon = O(G/(n\log n))$. Using the template at the start of Section 6 and appealing to Theorem 5.1, we can deduce that $s = \Omega(n) = \Omega(\varepsilon^{-2}/\log^2 \varepsilon^{-1})$.

The distinct elements case is similar: the same reduction ensures that $F_0 = n + \Delta(x)$, so either $F_0 \leq 3n/2 - G$ or $F_0 \geq 3n/2 + G$. Setting $\varepsilon = O(G/n)$ and $G = \Theta(\sqrt{n})$ means that the communication lower bound of $\Omega(n)$ entails a space lower bound of $\Omega(\varepsilon^{-2})$. This extends to all F_k, since we have $F_k = 2^k(n - \Delta(x)) + 1^k\Delta(x)$. Choosing $\varepsilon = O(G/n)$ and $G = \Theta(\sqrt{n})$ is again sufficient to force a space lower bound of $\Omega(\varepsilon^{-2})$ for any algorithm which can (ε, δ) approximate F_k for any constant $k \neq 1$. $\qquad\square$

6.3 Selection

Selection, including median-finding, is one of the earliest-studied problems in the data stream model [36] and been the focus of several recent works [8, 19, 24]. The following result improves upon the previous best single and multi-pass lower bounds [8, 24]. 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 $\Omega\left(m^{1/((p-1)2^{p+1}+2)} \cdot (\log m)^{-1/(2(p-1))} \cdot p^{-2}\right)$ space.*

Proof. Using the template at the start of Section 6, the theorem is immediate from Theorem 4.8. $\qquad\square$

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 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. Using Theorem 3.4 and Theorem 5.4 and reductions from [15, 27] it is possible to show the following results.

Theorem 6.4. *Given a stream of edges in random order, $\Omega(n)$ space is required by any constant pass algorithm that determines if the resulting graph is connected. Furthermore, any single pass algorithm that returns a t-approximation of the distance between two nodes requires $O(\mathrm{ex}(n-2,C_3,\ldots,C_{t+1}))$ space where $\mathrm{ex}(n-2,C_3,\ldots,C_{t+1})$ is the maximum size of a of graph on $n-2$ nodes 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 $\mathrm{ex}(n,C_3,\ldots,C_{t+1}) = \Omega(n^{1+2/t})$; see, e.g., [40].

Proof. For the first part of the theorem we consider a reduction from $\mathrm{DISJ}_{n/2,2}$ 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 comments at the start of Section 6. Let $x = \{x_{ij}\}_{i\in[2],j\in[n/2]}$ be an instance of $\mathrm{DISJ}_{n/2,2}$. Based on x we define the following bipartite 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

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

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 replaces 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 $p = 80$ and appealing to Theorem 3.4 gives the required result.

For the second part of the result, let $G = (V,E)$ be a graph on $n-2$ nodes with $m = \mathrm{ex}(n-2,C_3,\ldots,C_{t+1})$ nodes 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 s,t be two nodes not in V. Consider an instance $x\in\{0,1\}^{m+1}$ 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 an edge e_i with unit weight if $x_i = 1$. Suppose $x_0 = j$ and that $e_j = (u_j,v_j)$. With probability $1/2$ replace the first copy of x_0 by (s,u_j) and the second copy by (t,v_j). These edges have zero weight. Otherwise replace them in the reverse order. In this way we define a graph G' on nodes $V\cup\{s,t\}$ where the distance between s and t is 1 if $x_j = 1$ and at least t if $x_j = 0$. Hence, any protocol that distinguishes between the distance being 1 and at least t also determines the value of x_j. Appealing to Theorem 5.4 gives the required result. $\qquad\square$

6.5 Information Divergences

The next theorem extends a result by Guha et al. [22] on the approximation of information divergences. The results follows from Theorem 5.4 using a variant of the reduction from [22].

Theorem 6.5. *Given a randomly ordered stream defining two empirical distributions p and q on $[n]$, $\Omega(n)$ space is required to find an $\sqrt{1/2+a/2}$ multiplicative approximation to the squared Hellinger distance $h^2(p,q)$ with probability at least $1-2^{-a-3}$ (for some even $a\in\mathbb{N}^+$.)*

Proof. We consider a reduction from INDEX. Let $j \in [n]$, $x_1 \ldots x_n \in \{0,1\}^n$ be an instance of INDEX. Consider the random allocation where a copies of each x_i are uniformly distributed between the two players and x_0 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 $y^1, \ldots, y^n \in \{0,1\}^a$ where each string has weight exactly $a/2$. Suppose Alice and Bob receives d_1 and $d_2 = a - d_1$ copies of the token for x_i for $i \in [n]$. If $x_i = 1$, Alice generates $|\{j \le d_1 : y^i_j = 1\}|$ copies of $\langle p,i \rangle$ and $|\{j \le d_1 : y^i_j = 0\}|$ copies of $\langle q,i \rangle$. If $x_i = 1$, Bob generates $|\{a - d_2 < j \le a : y^i_j = 1\}|$ copies of $\langle p,i \rangle$ and $|\{a - d_2 < j \le a : y^i_j = 0\}|$ copies of $\langle q,i \rangle$. Note that if d_1 or d_2 is zero, then the relevant player does not need to know the value of x_i to perform this reduction.

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

3. Additionally, the players generate $a + 1$ copies of $\langle p, n+1 \rangle$ and a 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 $\langle q,j \rangle$, $a+1$ copies of $\langle p, n+1 \rangle$ and a copies of $\langle q, n+1 \rangle$ have been generated. Therefore,

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

where $m = a + a|\{i \in [n] : x_i = 1\}|$. Furthermore,

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

Hence, a $\left(\sqrt{1/2 + a/2}\right)$-approximation would be sufficient to solve the instance of INDEX. Therefore, by Theorem 5.4, any stream algorithm that returns such an estimate requires $\Omega(n)$ space. $\square$

References

[1] F. Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 175(2):139–159, 1996.

[2] A. V. Aho, J. D. Ullman, and M. Yannakakis. On notions of information transfer in VLSI circuits. In *ACM Symposium on Theory of Computing*, pages 133–139, 1983.

[3] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[4] A. Andoni, A. McGregor, K. Onak, and R. Panigrahy. Better bounds for frequency moments in random-order streams. *CoRR*, abs/0808.2222, 2008.

[5] Z. Bar-Yossef, T. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. Counting distinct elements in a data stream. In *Proc. 6th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 1–10, 2002.

[6] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[7] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 328–335, 2007.

[8] A. Chakrabarti, T. Jayram, and M. Pătraşcu. Tight lower bounds for selection in randomly ordered streams. In *ACM-SIAM Symposium on Discrete Algorithms*, 2008.

[9] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *IEEE Conference on Computational Complexity*, pages 107–117, 2003.

[10] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of GAP-HAMMING-DISTANCE. In *ACM Symposium on Theory of Computing*, 2011. to appear.

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

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[13] E. D. Demaine, A. López-Ortiz, and J. I. Munro. Frequency estimation of internet packet streams with limited space. In *European Symposium on Algorithms*, pages 348–360, 2002.

[14] J. Edmonds and R. Impagliazzo. Manuscript. Unpublished, 1994.

[15] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. Graph distances in the data-stream model. *SIAM J. Comput.*, 38(5):1709–1727, 2008.

[16] J. Feigenbaum, S. Kannan, M. Strauss, and M. Viswanathan. An approximate L^1 difference algorithm for massive data streams. *SIAM Journal on Computing*, 32(1):131–151, 2002.

[17] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *J. Comput. Syst. Sci.*, 31(2):182–209, 1985.

[18] S. Ganguly. Counting distinct items over update streams. *Theor. Comput. Sci.*, 378(3):211–222, 2007.

[19] M. Greenwald and S. Khanna. Efficient online computation of quantile summaries. In *ACM International Conference on Management of Data*, pages 58–66, 2001.

[20] A. Gronemeier. Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the and-function and disjointness. In *Symposium on Theoretical Aspects of Computer Science*, pages 505–516, 2009.

[21] S. Guha and Z. Huang. Revisiting the direct sum theorem and space lower bounds in random order streams. In *International Colloquium on Automata, Languages and Programming*, 2009.

[22] S. Guha, P. Indyk, and A. McGregor. Sketching information divergences. *Mach. Learn.*, 72(1-2):5–19, 2008.

[23] S. Guha and A. McGregor. Space-efficient sampling. In *AISTATS*, pages 169–176, 2007.

[24] S. Guha and A. McGregor. Stream order and order statistics: Quantile estimation in random-order streams. *SIAM Journal on Computing*, 38(5):2044–2059, 2009.

[25] S. Guha, A. McGregor, and S. Venkatasubramanian. Streaming and sublinear approximation of entropy and information distances. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 733–742, 2006.

[26] N. J. A. Harvey, J. Nelson, and K. Onak. Sketching and streaming entropy via approximation theory. In *IEEE Symposium on Foundations of Computer Science*, pages 489–498, 2008.

[27] M. R. Henzinger, P. Raghavan, and S. Rajagopalan. Computing on data streams. *External memory algorithms*, pages 107–118, 1999.

[28] P. Indyk and D. P. Woodruff. Tight lower bounds for the distinct elements problem. *IEEE Symposium on Foundations of Computer Science*, pages 283–288, 2003.

[29] P. Indyk and D. P. Woodruff. Optimal approximations of the frequency moments of data streams. In *ACM Symposium on Theory of Computing*, pages 202–208, 2005.

[30] T. S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *ACM Symposium on Theory of Computing*, pages 673–682, 2003.

[31] T. S. Jayram, R. Kumar, and D. Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.

[32] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *ACM Symposium on Theory of Computing*, pages 124–133, 2001.

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

[34] T. W. Lam and W. L. Ruzzo. Results on communication complexity classes. *J. Comput. Syst. Sci.*, 44(2):324–342, 1992.

[35] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.

[36] J. I. Munro and M. Paterson. Selection and sorting with limited storage. *Theor. Comput. Sci.*, 12:315–323, 1980.

[37] C. H. Papadimitriou and M. Sipser. Communication complexity. *J. Comput. Syst. Sci.*, 28(2):260–269, 1984.

[38] P. Pudlák and J. Sgall. An upper bound for a communication game related to time-space tradeoffs. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(10), 1995.

[39] P. Sen. Lower bounds for predecessor searching in the cell probe model. In *IEEE Conference on Computational Complexity*, pages 73–83, 2003.

[40] M. Simonovits. Extremal graph theory. *Selected topics in graph theory*, 2:161–200, 1983.

[41] D. P. Woodruff. Optimal space lower bounds for all frequency moments. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 167–175, 2004.

[42] D. P. Woodruff. The average-case complexity of counting distinct elements. In *International Conference in Database Theory*, pages 284–295, 2009.

[43] A. C. Yao. Some complexity questions related to distributive computing (preliminary report). *ACM Symposium on Theory of Computing*, pages 209–213, 1979.

A Variational Distance between Binomial Distributions

Proof of Lemma 2.6. Let $\gamma = 1 - q$. For the first part of the lemma we may assume that $w = O(\sqrt{a\gamma \ln(a/w)})$ because otherwise the bound is trivial. Then, by an application of the Chernoff bounds, there exists a constant c_1' such that

$$\max\left(\Pr\left[|\mathscr{B}(a,q) - aq| \geq c_1'\sqrt{a\gamma \ln(a/w)}\right], \Pr\left[|\mathscr{B}(a-w,q) - aq| \geq c_1'\sqrt{a\gamma \ln(a/w)}\right]\right) \leq \frac{w}{\sqrt{a}}.$$

Let $t = c_1'\sqrt{a\gamma \ln(a/w)}$. Then,

$$
\begin{aligned}
\mathrm{D_{TV}}\left(\mathscr{B}(a,q), \mathscr{B}(a-w,q)\right) &\leq \frac{2w}{\sqrt{a}} + \sum_{r=aq-t}^{aq+t}\left|\binom{a}{r}q^r(1-q)^{a-r} - \binom{a-w}{r}q^r(1-q)^{a-w-r}\right| \\
&\leq \frac{2w}{\sqrt{a}} + \max_{r\in aq\pm t}\left|\frac{a!(a-w-r)!}{(a-r)!(a-w)!}(1-q)^w - 1\right| \\
&= \frac{2w}{\sqrt{a}} + \max_{r\in aq\pm t}\left|\frac{a(a-1)\ldots(a-w+1)}{(a-r)(a-r-1)\ldots(a-w-r+1)}(1-q)^w - 1\right| \\
&\leq \frac{2w}{\sqrt{a}} + \max\left\{\left|\left(\frac{a}{a-aq+t}\gamma\right)^w - 1\right|, \left|\left(\frac{a-w+1}{a-aq-t-w+1}\gamma\right)^w - 1\right|\right\} \\
&= \frac{2w}{\sqrt{a}} + \max\left\{\left|\left(1 - \frac{t}{\gamma a + t}\right)^w - 1\right|, \left|\left(1 + \frac{qw-q+t}{\gamma a - t - w + 1}\right)^w - 1\right|\right\} \\
&\leq \frac{2w}{\sqrt{a}} + \max\left\{\frac{tw}{\gamma a + t}, \exp\left(\frac{qw^2 - qw + tw}{\gamma a - t - w + 1}\right) - 1\right\} \\
&= O(1)\cdot w\sqrt{\ln(a)/(\gamma a)}.
\end{aligned}
$$

For the second part of lemma, we proceed in a similar fashion. By Chernoff bounds, there exists a constant c_2' such that

$$\max\left(\Pr\left[|\mathscr{B}(a,1/2) - a/2| \geq c_2'\sqrt{a\ln(\delta a)^{-1}}\right], \Pr\left[|\mathscr{B}(a,q') - a/2| \geq c_2'\sqrt{a\ln(\delta a)^{-1}}\right]\right) \leq \delta^2 a,$$

where we have assumed that $\delta a = O(\sqrt{a\ln(\delta^{-1}a^{-1})})$, since otherwise the bound is trivial. Let $s = c_2'\sqrt{a\ln(\delta a)^{-1}}$. Then,

$$
\begin{aligned}
\mathrm{D_{TV}}\left(\mathscr{B}(a,1/2), \mathscr{B}(a,q)\right) &\leq \sum_{r\in[a]}\binom{a}{r}|1/2^a - q^r(1-q)^{a-r}| \\
&= 2\delta^2 a + \max_{r\in a/2\pm s}\left|(1+2\delta)^r(1-2\delta)^{a-r} - 1\right| \\
&\leq 2\delta^2 a + \max_{u\in\pm s}\left|(1+2\delta)^{1/2+u/a}(1-2\delta)^{1/2-u/a} - 1\right|^a \\
&\leq 2\delta^2 a + \max_{u\in\pm s}\left|(1-4\delta^2)^{1/2}\left(1 + \frac{4\delta}{1-2\delta}\right)^{u/a} - 1\right|^a \\
&= O(1)\cdot\left(\delta^2 a + \delta\sqrt{a\ln(\delta a)^{-1}}\right). \qquad \square
\end{aligned}
$$