On the Power of Multiple Anonymous Messages*

Badih Ghazi Noah Golowich† Ravi Kumar Rasmus Pagh‡ Ameya Velingker
Google Research
Mountain View, CA
badihghazi@gmail.com, nzg@mit.edu, ravi.k53@gmail.com,
pagh@itu.dk, ameyav@google.com

Abstract

An exciting new development in differential privacy is the *shuffled* model, in which an anonymous channel enables non-interactive, differentially private protocols with error much smaller than what is possible in the local model, while relying on weaker trust assumptions than in the central model. In this paper, we study basic counting problems in the shuffled model and establish separations between the error that can be achieved in the single-message shuffled model and in the shuffled model with multiple messages per user.

For the problem of *frequency estimation* for n users and a domain of size B, we obtain:

- A nearly tight lower bound of $\tilde{\Omega}(\min(\sqrt[4]{n}, \sqrt{B}))$ on the error in the single-message shuffled model. This implies that the protocols obtained from the amplification via shuffling work of Erlingsson et al. (SODA 2019) and Balle et al. (Crypto 2019) are essentially optimal for single-message protocols. A key ingredient in the proof is a lower bound on the error of locally-private frequency estimation in the low-privacy (aka high ε) regime. For this we develop new techniques to extend the results of Duchi et al. (FOCS 2013; JASA 2018) and Bassily & Smith (STOC 2015), whose techniques were restricted to the high-privacy case.
- Protocols in the *multi-message* shuffled model with $\mathrm{poly}(\log B, \log n)$ bits of communication per user and $\mathrm{poly}\log B$ error, which provide an exponential improvement on the error compared to what is possible with single-message algorithms. This implies protocols with similar error and communication guarantees for several well-studied problems such as heavy hitters, d-dimensional range counting, M-estimation of the median and quantiles, and more generally sparse non-adaptive statistical query algorithms.

For the related *selection* problem on a domain of size B, we prove:

- A nearly tight lower bound of $\Omega(B)$ on the number of users in the single-message shuffled model. This significantly improves on the $\Omega(B^{1/17})$ lower bound obtained by Cheu et al. (Eurocrypt 2019), and when combined with their $\tilde{O}(\sqrt{B})$-error multi-message protocol, implies the first separation between single-message and multi-message protocols for this problem.

*A 1-page abstract based on this work will be presented at the Symposium on Foundations of Responsible Computing (FORC) 2020.

†MIT EECS. Supported at MIT by a Fannie & John Hertz Foundation Fellowship, an MIT Akamai Fellowship, and an NSF Graduate Fellowship. This work was done while at Google Research.

‡Visiting from BARC and IT University of Copenhagen.

Contents

1 Introduction

With increased public awareness and the introduction of stricter regulation of how personally identifiable data may be stored and used, user privacy has become an issue of paramount importance in a wide range of practical applications. While many formal notions of privacy have been proposed (see, e.g., [LLV07]), *differential privacy (DP)* [DMNS06, DKM$^+$06] has emerged as the gold standard due to its broad applicability and nice features such as composition and post-processing (see, e.g., [DR$^+$14b, Vad17] for a comprehensive overview). A primary goal of DP is to enable processing of users' data in a way that (i) does not reveal substantial information about the data of any single user, and (ii) allows the accurate computation of functions of the users' inputs. The theory of DP studies what trade-offs between privacy and accuracy are feasible for desired families of functions.

Most work on DP has been in the *central* (a.k.a. *curator*) setup, where numerous private algorithms with small error have been devised (see, e.g., [BLR08, DNR$^+$09, DR14a]). The premise of the central model is that a curator can access the raw user data before releasing a differentially private output. In distributed applications, this requires users to transfer their raw data to the curator — a strong limitation in cases where users would expect the entity running the curator (e.g., a government agency or a technology company) to gain little information about their data.

To overcome this limitation, recent work has studied the *local* model of DP [KLN$^+$08] (also [War65]), where each individual message sent by a user is required to be private. Indeed, several large-scale deployments of DP in practice, at companies such as Apple [Gre16, App17], Google [EPK14, Sha14], and Microsoft [DKY17], have used local DP. While estimates in the local model require weaker trust assumptions than in the central model, they inevitably suffer from significant error. For many types of queries, the estimation error is provably larger than the error incurred in the central model by a factor growing with the square root of the number of users.

Shuffled Privacy Model. The aforementioned trade-offs have motivated the study of the *shuffled* model of privacy as a middle ground between the central and local models. While a similar setup was first studied in cryptography in the work of Ishai et al. [IKOS06] on cryptography from anonymity, the shuffled model was first proposed for privacy-preserving protocols by Bittau et al. [BEM$^+$17] in their Encode-Shuffle-Analyze architecture. In the shuffled setting, each user sends one or more messages to the analyzer using an *anonymous* channel that does not reveal where each message comes from. This kind of anonymization is a common procedure in data collection and is easy to explain to regulatory agencies and users. The anonymous channel is equivalent to all user messages being randomly shuffled (i.e., permuted) before being operated on by the analyzer, leading to the model illustrated in Figure 1; see Section 2.2 for a formal description of the shuffled model. In this work, we treat the shuffler as a black box, but note that various efficient cryptographic implementations of the shuffler have been considered, including onion routing, mixnets, third-party servers, and secure hardware (see, e.g., [IKOS06, BEM$^+$17]). A comprehensive overview of recent work on anonymous communication can be found on Free Haven's Selected Papers in Anonymity website[1].

[1] `https://www.freehaven.net/anonbib/`

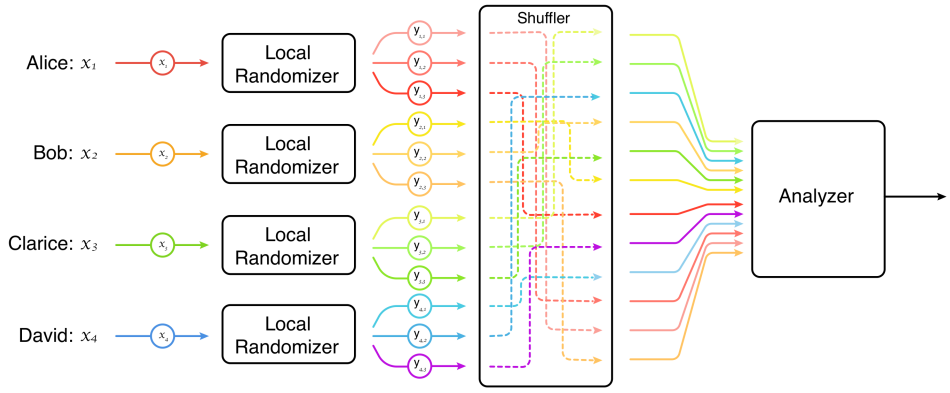


Figure 1: Computation in the shuffled model consists of local randomization of inputs in the first stage, followed by a shuffle of all outputs of the local randomizers, after which the shuffled output is passed on to an analyzer.

The DP properties of the shuffled model were first analytically studied, independently, in the works of Erlingsson et al. [EFM+19] and Cheu et al. [CSU+19]. Protocols within the shuffled model are non-interactive and fall into two categories: *single-message* protocols, in which each user sends one message (as in the local model), and *multi-message* protocols, in which a user can send more than one message. In both variants, the messages sent by all users are shuffled before being passed to the analyzer. The goal is to design private protocols in the shuffled model with as small error and total communication as possible. An example of the power of the shuffled model was established by Erlingsson et al. [EFM+19] and extended by Balle et al. [BBGN19c], who showed that every local DP algorithm directly yields a single-message protocol in the shuffled model with significantly better privacy.

1.1 Results

In this work, we study several basic problems related to *counting* in the shuffled model of DP. In these problems, each of n users holds an element from a domain of size B. We consider the problems of frequency estimation, variable selection, heavy hitters, median, and range counting and study whether it is possible to obtain (ε, δ)-DP[2] in the shuffled model with accuracy close to what is possible in the central model, while keeping communication low.

The *frequency estimation* problem (a.k.a. *histograms* or *frequency oracles*) is at the core of all the problems we study. In the simplest version, each of n users gets an element of a domain $[B] := \{1, \ldots, B\}$ and the goal is to estimate the number of users holding element j, for any query element $j \in [B]$. Frequency estimation has been extensively studied in DP where in the central model, the smallest possible error is $\Theta(\min(\log(1/\delta)/\varepsilon, \log(B)/\varepsilon, n))$ (see, e.g., [Vad17, Section 7.1]). By contrast, in the local model of DP, the smallest possible error is known to be $\Theta(\min(\sqrt{n \log(B)}/\varepsilon, n)$ under the assumption that $\delta < 1/n$ [BS15]. Confirm this

In the high-level exposition of our results given below, we let n and B be any positive integers, $\varepsilon > 0$ be any constant, and $\delta > 0$ be inverse polynomial in n. This assumption on ε and δ covers a regime of parameters that is relevant in practice. We will also make use of tilde notation (e.g., $\tilde{O}$, $\tilde{\Theta}$) to indicate the possible suppression of multiplicative factors that are polynomial in $\log B$ and $\log n$.

Single-Message Bounds for Frequency Estimation. For the frequency estimation problem, we show the following results in the shuffled model where each user sends a single message.

Theorem 1.1 (Informal version of Theorems 3.1 & 3.4)**.** *The optimal error of private frequency estimation in the single-message shuffled model is* $\tilde{\Theta}(\min(\sqrt[4]{n}, \sqrt{B}))$.

The main contribution of Theorem 1.1 is the lower bound. To prove this result, we obtain improved bounds on the error needed for frequency estimation in local DP in the weak privacy regime where ε is around $\ln n$. The upper bound in Theorem 1.1 follows by combining the recent result of Balle et al. [BBGN19c] (building on the earlier result of Erlingsson et al. [EFM+19]) with RAPPOR [EPK14] and B-ary randomized response [War65] (see Section 1.2 and Appendix A for more details).

Theorem 1.1 implies that in order for a single-message differentially private protocol to get error $o(n)$ one needs to have $n = \omega\left(\frac{\log B}{\log \log B}\right)$ users; see Corollary 3.2. This improves on a result of Cheu et al. [CSU+19, Corollary 32], which gives a lower bound of $n = \omega(\log^{1/17} B)$ for this task.

Single-Message Bounds for Selection. It turns out that the techniques that we develop to prove the lower bound in Theorem 1.1 can be used to get a nearly tight $\Omega(B)$ lower bound on the number of users necessary to solve the *selection* problem. In the selection problem[3], each user $i \in [n]$ is given an arbitrary subset of $[B]$, represented by the indicator vector $x_i \in \{0,1\}^B$, and the goal is for the analyzer to output an index $j^* \in [B]$ such that

$$\sum_{i \in [n]} x_{i,j^*} \geqslant \max_{j \in [B]} \sum_{i \in [n]} x_{i,j} - \frac{n}{10}. \tag{1}$$

[2] Formally stated in Definition 2.1.

[3] Sometimes also referred to as *variable selection*.

In other words, the analyzer's output should be the index of a domain element that is held by an approximately maximal number of users. The choice of the constant 10 in (1) is arbitrary; any constant larger than 1 may be used.

The selection problem has been studied in several previous works on differential privacy, and it has many applications to machine learning, hypothesis testing and approximation algorithms (see [DJW13, SU17, Ull18] and the references therein). Our work improves an $\Omega(B^{1/17})$ lower bound in the single-message shuffled model due to Cheu et al. [CSU+19]. For $\varepsilon = 1$, the exponential mechanism [MT07] implies an $(\varepsilon, 0)$-DP algorithm for selection with $n = O(\log B)$ users in the central model, whereas in the local model, it is known that any $(\varepsilon, 0)$-DP algorithm for selection requires $n = \Omega(B \log B)$ users [Ull18]. Variants of the selection problem appear in several natural statistical tasks such as feature selection and hypothesis testing (see, e.g., [SU17] and the references therein).

Theorem 1.2 (Informal version of Theorem 3.22)**.** *For any single-message differentially private protocol in the shuffled model that solves the selection problem given in Equation (1), the number n of users should be $\Omega(B)$.*

The lower bound in Theorem 1.2 nearly matches the $O(B \log B)$ upper bound on the required number of users that holds even in the local model (and hence in the single-message shuffled model) and that uses the B-randomized response [War65, Ull18]. Cheu et al. [CSU+19] have previously obtained a multi-message protocol for selection with $O(\sqrt{B})$ users, and combined with this result Theorem 1.2 yields the first separation between single-message and multi-message protocols for selection.

Multi-Message Protocols for Frequency Estimation. We next present (non-interactive) multi-message protocols in the shuffled model of DP for frequency estimation with only *polylogarithmic* error and communication. This is in strong contrast with what is possible for any protocol in the single-message shuffled setup where Theorem 1.1 implies that the error has to grow polynomially with $\min(n, B)$, even with unbounded communication. In addition to error and communication, a parameter of interest is the query time, which is the time to estimate the frequency of any element $j \in [B]$ from the data structure constructed by the analyzer.

Theorem 1.3 (Informal version of Theorems 4.1 & 4.2)**.** *There is a private-coin (resp., public-coin) multi-message protocol in the shuffled model for frequency estimation with error $\tilde{O}(1)$, total communication of $\tilde{O}(1)$ bits per user, and query time $\tilde{O}(n)$ (resp., $\tilde{O}(1)$).*

Combining Theorems 1.1 and 1.3 yields the first separation between single-message and multi-message protocols for frequency estimation. Moreover, Theorem 1.3 can be used to obtain multi-message protocols with small error and small communication for several other widely studied problems (e.g., heavy hitters, range counting, and median and quantiles estimation), discussed in Section 1.4. Finally, Theorem 1.3 implies the following consequence for statistical query (SQ) algorithms with respect to a distribution $\mathcal{D}$ on $\mathcal{X}$ (see Appendix B for the basic definitions). We say that a non-adaptive SQ algorithm $\mathcal{A}$ making at most B queries $q : \mathcal{X} \to \{0, 1\}$ is *k-sparse* if for each $x \in \mathcal{X}$, the Hamming weight of the output of the queries is at most k. Then, under the assumption that users' data is drawn i.i.d. from $\mathcal{D}$, the algorithm $\mathcal{A}$ can be efficiently simulated in the shuffled model as follows:

Corollary 1.4 (Informal version of Corollary B.1)**.** *For any non-adaptive k-sparse SQ algorithm $\mathcal{A}$ with B queries and $\beta > 0$, there is a (private-coin) shuffled model protocol satisfying (ε, δ)-DP whose output has total variation distance at most β from that of $\mathcal{A}$, such that the number of users is $n \leqslant \tilde{O}\left(\frac{k}{\varepsilon\tau} + \frac{1}{\tau^2}\right)$, and the per-user communication is $\tilde{O}\left(\frac{k^2}{\varepsilon^2}\right)$, where $\tilde{O}(\cdot)$ hides logarithmic factors in $B, n, 1/\delta, 1/\varepsilon$, and $1/\beta$.*

Corollary 1.4 improves upon the simulation of non-adaptive SQ algorithms in the *local model* [KLN+08], for which the number of users must grow as $\frac{k}{\varepsilon^2\tau^2}$ as opposed to $\frac{1}{\tau^2} + \frac{k}{\varepsilon\tau}$ in the shuffled model. We emphasize that the main novelty of Corollary 1.4 is in the regime that $k^2/\varepsilon^2 \ll B$; in particular, though prior work on low-communication private summation in the shuffled model [CSU+19, GMPV19, BBGN20] implies an algorithm for simulating $\mathcal{A}$ with roughly the same bound on the number of users n as in Corollary 1.4 and communication $\Omega(B)$, it was unknown whether the communication could be reduced to have logarithmic dependence on B, as in Corollary 1.4.

	Local		Local + shuffle	Shuffled, single-message	Shuffled, multi-message	Central
Expected max. error	$\tilde{O}(\sqrt{n})$	$\tilde{\Omega}(\sqrt{n})$	$\tilde{O}(\min(\sqrt[4]{n}, \sqrt{B}))$	$\tilde{\Omega}(\min(\sqrt[4]{n}, \sqrt{B}))$	$\tilde{\Theta}(1)$	$\tilde{\Theta}(1)$
Communication per user	$\Theta(1)$	any	$O(B)$ (err $\sqrt[4]{n}$) $\tilde{O}(1)$ (err $\sqrt{B}$)	any	$\tilde{\Theta}(1)$	n.a.
References	[BNST17]	[BS15]	[War65, EPK14, BBGN19c]	Theorems 3.4 & 3.1	Theorem 4.1	[MT07, SU17]

Table 1: Upper and lower bounds on expected maximum error (over all B queries, where the sum of all frequencies is n) for frequency estimation in different models of DP. The bounds are stated for fixed, positive privacy parameters ε and δ, and $\tilde{\Theta}/\tilde{O}/\tilde{\Omega}$ asymptotic notation suppresses factors that are polylogarithmic in B and n. The communication per user is in terms of the total number of bits sent. In all upper bounds, the protocol is symmetric with respect to the users, and no public randomness is needed. References are to the first results we are aware of that imply the stated bounds.

1.2 Overview of Single-Message Lower Bounds

We start by giving an overview of the lower bound of $\tilde{\Omega}(\min\{n^{1/4}, \sqrt{B}\})$ in Theorem 1.1 on the error of any single-message frequency estimation protocol. We first focus on the case where $n \leqslant B^2$ and thus $\min\{n^{1/4}, \sqrt{B}\} = n^{1/4}$. The main component of the proof in this case is a lower bound of $\tilde{\Omega}(n^{1/4})$ for frequency estimation for $(\varepsilon_L, \delta_L)$-*local* DP protocols[4] when $\varepsilon_L = \ln(n) + O(1)$. While lower bounds for local DP frequency estimation were previously obtained in the seminal works of Bassily and Smith [BS15] and Duchi, Jordan and Wainwright [DJW18], two critical reasons make them less useful for our purposes: (i) their dependence on ε_L is sub-optimal when $\varepsilon_L = \omega(1)$ (i.e., low error regime) and (ii) they only apply to the case where $\delta_L = 0$ (i.e., pure privacy).[5] We prove new error bounds in the low error and approximate privacy regime in order to obtain our essentially tight lower bound in Theorem 1.1 for single-message shuffled protocols. We discuss these and outline the proof next.

Let R be an $(\varepsilon_L, \delta_L)$-locally differentially private randomizer. The general approach [BS15, DJW18] is to show that if V is a random variable drawn uniformly at random from $[B]$ and if X is a random variable that is equal to V with probability parameter $\alpha \in (0, 1)$, and is drawn uniformly at random from $[B]$ otherwise, then the mutual information between V and the local randomizer output $R(X)$ satisfies

As a student of this field: I should be able to teach this paragraph

$$I(V; R(X)) \leqslant \frac{\log B}{4n}. \tag{2}$$

Once (2) is established, the chain rule of mutual information implies that $I(V; R(X_1), \ldots, R(X_n)) \leqslant \frac{\log B}{4}$, where $X_1, \ldots, X_n$ are independent and identically distributed given V. Fano's inequality [CT91] then implies that the probability that any analyzer receiving $R(X_1), \ldots, R(X_n)$ correctly guesses V is at most $1/4$; on the other hand, an $\Omega(\alpha n)$-accurate analyzer must be able to determine V with high probability since its frequency in the dataset $X_1, \ldots, X_n$ is roughly αn, greater than the frequency of all other $v \in [B]$. This approach thus yields a lower bound of $\Omega(\alpha n)$ on frequency estimation.

To prove the desired $\tilde{\Omega}(n^{1/4})$ lower bound using this approach, it turns out we need a bound of the form

$$I(V; R(X)) \leqslant \tilde{O}(\alpha^4 n e^{\varepsilon_L}), \tag{3}$$

where both $\delta_L > 0$ and $\varepsilon_L = \omega(1)$. (We will in fact choose $\alpha = \tilde{\Theta}(n^{-3/4})$ and $\varepsilon_L = \ln(n) + O(1)$; as we will discuss later, (3) is essentially tight in this regime.)

Limitations of Previous Approaches We first state the existing upper bounds on $I(V; R(X))$, which only use the privacy of the local randomizer. Bassily and Smith [BS15, Claim 5.4] showed an upper bound of $I(V; R(X)) \leqslant O(\varepsilon_L^2 \alpha^2)$ with $\varepsilon_L = O(1)$ and $\delta_L = o(1/(n \log n))$, which thus satisfies (2) with $\alpha = \Theta\left(\sqrt{\frac{\log B}{\varepsilon_L^2 n}}\right)$. For $\delta_L = 0$,

[4] Note that we use the subscripts in ε_L and δ_L to distinguish the privacy parameters of the *local* model from the ε and δ parameters (without a subscript) of the shuffled model.

[5] As we discuss in Remark 3.1, generic reductions [CSU+19, BNS18] showing that one can efficiently simulate an approximately differentially private protocol (i.e., with $\delta_L > 0$) with a pure differentially private protocol (i.e., with $\delta_L = 0$) are insufficient to obtain tight lower bounds.

Duchi et al. [DJW18] generalized this result to the case $\varepsilon_L \geqslant 1$, proving that[6] $I(V; R(X)) \leqslant O(\alpha^2 e^{2\varepsilon_L})$. Both of these bounds are weaker than (3) for the above setting of α and ε_L.

However, proving the mutual information bound in (3) turns out to be impossible if we only use the privacy of the local randomizers! In fact, the bound can be shown to be *false* if all we assume about R is that it is $(\varepsilon_L, \delta_L)$-locally differentially private for some $\varepsilon_L \approx \ln n$ and $\delta_L = n^{-O(1)}$. For instance, it is violated if one takes R to be R_{RR}, the local randomizer of the B-randomized response [War65]. Consider for example the regime where $B \leqslant n \leqslant B^2$, and the setting where $R_{\mathrm{RR}}(v)$ is equal to v with probability $1 - B/n$, and is uniformly random over $[B]$ with the remaining probability of B/n. In this case, the local randomizer $R_{\mathrm{RR}}(\cdot)$ is $(\ln(n) + O(1), 0)$-differentially private. A simple calculation shows that $I(V; R_{\mathrm{RR}}(X)) = \tilde{\Theta}(\alpha)$. Whenever $\alpha \ll 1/\sqrt{n}$, which is the regime we have to consider in order to obtain any non-trivial lower bound[7] in the single-message shuffled model, it holds that $\alpha \gg \alpha^4 n \exp(\ln(n))$, thus contradicting (3) (see Remark 3.4). The insight derived from this counterexample is actually crucial, as we describe in our new technique next.

Mutual Information Bound from Privacy and Accuracy Departing from previous work, we manage to prove the stronger bound (3) as follows. Inspecting the counterexample based on the B-randomized response outlined above, we first observe that any analyzer in this case must have error at least $\Omega(\sqrt{B})$, which is larger than αn, the error that would be ruled out by the subsequent application of Fano's inequality! This led us to appeal to accuracy, in addition to privacy, when proving the mutual information upper bound. We thus leverage the additional available property that the local randomizer R can be combined with an analyzer A in such a way that the mapping $(x_1, \ldots, x_n) \mapsto A(R(x_1), \ldots, R(x_n))$ computes the frequencies of elements of every dataset $(x_1, \ldots, x_n)$ accurately, i.e., to within an error of $O(\alpha n)$. At a high level, our approach for proving the bound in (3) then proceeds by:

(i) Proving a structural property satisfied by the randomizer corresponding to any accurate frequency estimation protocol. Namely, we show in Lemma 3.15 that if there is an accurate analyzer, the total variation distance between the output of the local randomizer on any given input, and its output on a uniform input, is close to 1.

(ii) Using the $(\varepsilon_L, \delta_L)$-DP property of the randomizer along with the structural property in (i) in order to upper-bound the mutual information $I(V; R(X))$.

We believe that the application of the structural property in (i) to proving bounds of the form (3) is of independent interest. As we further discuss below, this property is, in particular, used (together with privacy of R) to argue that for most inputs $v \in [B]$, the local randomizer output $R(v)$ is unlikely to equal a message that is much less likely occur when the input is uniformly random than when it is v. Note that it is somewhat counter-intuitive that accuracy is used in the proof of this fact, as one way to achieve very accurate protocols is to ensure that $R(v)$ is equal to a message which is unlikely when the input is any $u \neq v$. We now outline the proofs of (i) and (ii) in more detail.

The gist of the proof of (i) is an anti-concentration statement. Let v be a fixed element of $[B]$ and let X be a random variable uniformly distributed on $[B]$. Assume that the total variation distance $\Delta(R(v), R(X))$ is not close to 1, and that a small fraction of the users have input v while the rest have uniformly random inputs. Let $\mathcal{Z}$ denote the range of the local randomizer R. First, we consider the special case where $\mathcal{Z}$ is $\{0, 1\}$. Then the distribution of the shuffled outputs of the users with v as their input is in bijection with a binomial random variable with parameter $p := \mathbb{P}[R(v) = 1]$, and the same is true for the distribution of the shuffled outputs of the users with uniform random inputs X (with parameter $q := \mathbb{P}[R(X) = 1]$). Then, we use the anti-concentration properties of binomial random variables in order to argue that if $|p - q| = \Delta(R(v), R(X))$ is too small, then with nontrivial probability the shuffled outputs of the users with input v will be indistinguishable from the shuffled outputs of the users with uniform random inputs. This is then used to contradict the supposed accuracy of the analyzer. To deal with the general case where the range $\mathcal{Z}$ is any finite set, we repeatedly apply the data processing inequality for total variation distance in order to reduce to the binary case (Lemma 3.20). The full proof appears in Lemma 3.15.

Equipped with the property in (i), we now outline the proof of the mutual information bound in (ii). Denote by

- $\mathcal{T}_v$ the set of messages *much more likely* to occur when the input is v than when it is uniform,
- $\mathcal{Y}_v$ the set of messages *less likely* to occur when the input is v than when it is uniform.

[6] This bound is not stated explicitly in [DJW18], though [DJW18, Lemma 7] proves a similar result whose proof can readily be modified appropriately.

[7] i.e., any stronger lower bound than what holds even in the *local* model

Note that the union $\mathcal{T}_v \cup \mathcal{Y}_v$ is *not* the entire range $\mathcal{Z}$ of messages; in particular, it does not include messages that are *a bit more likely* to occur when the input is v than when it is uniform.[8] On a high level, it turns out that the mutual information $I(V; R(X))$ will be large, i.e., $R(X)$ will reveal a significant amount of information about V, if either of the following events occurs:

(a) There are too many inputs $v \in [B]$ such that the mass $\mathbb{P}[R(X) \in \mathcal{Y}_v]$ is small. Intuitively, for such v, the local randomizer R fails to "hide" the fact that a uniform input X is v given that X indeed equals v and $R(X) \in \mathcal{Y}_v$.

(b) There are too many inputs $v \in [B]$ such that the mass $\mathbb{P}[R(v) \in \mathcal{T}_v]$ is large. Such inputs make it too likely that $X = v$ given that $R(X) \in \mathcal{T}_v$, which makes it more likely in turn that $V = v$.

We first note that the total variation distance $\Delta(R(v), R(X))$ is upper-bounded by $\mathbb{P}[R(X) \in \mathcal{Y}_v]$. On the other hand, the accuracy of the protocol along with property (i) imply that $\Delta(R(v), R(X))$ is close to 1. By putting these together, we can conclude that event (a) does not occur (see Lemma 3.15 for more details).

To prove that event (b) does not occur, we use the $(\varepsilon_L, \delta_L)$-DP guarantee of the local randomizer R. Namely, we will use the inequality $\mathbb{P}[R(v) \in \mathcal{S}] \leqslant e^{\varepsilon_L} \cdot \mathbb{P}[R(X) \in \mathcal{S}] + \delta$ for various subsets $\mathcal{S}$ of $\mathcal{Z}$. Unfortunately, setting $\mathcal{S} = \mathcal{T}_v$ does not lead to a good enough upper bound on $\mathbb{P}[R(v) \in \mathcal{T}_v]$; indeed, for the local randomizer $R = R_{\mathrm{RR}}$ corresponding to the B-ary randomized response, we will have $\mathcal{T}_v = \{v\}$ for $n \gg B$, and so $\mathbb{P}[R(v) \in \mathcal{T}_v] = 1 - B/n \approx 1$ for any v. Thus, to establish (b), we need to additionally use the accuracy of the analyzer A (i.e., property (i) above), together with a careful double-counting argument to enumerate the probabilities that $R(v)$ belongs to subsets of $\mathcal{T}_v$ of different granularity (with respect to the likelihood of occurrence under input v versus a uniform input). For the details, we refer the reader to Section 3.3 and Lemma 3.14.

Having established the above lower bound for locally differentially private estimation in the low-privacy regime, the final step is to apply a lemma of Cheu et al. [CSU+19] (restated as Lemma 3.5 below), stating that any lower bound for $(\varepsilon + \ln(n), \delta)$-locally differentially private protocols implies a lower bound for (ε, δ)-differentially private protocols in the single-message shuffled model (i.e., we take $\varepsilon_L = \varepsilon + \ln(n)$). Moreover, for $\varepsilon_L = \ln(n) + O(1)$ and $\alpha = \tilde{\Theta}(n^{-3/4})$, we observe that (3) implies (2), and thus a lower bound of $\tilde{\Omega}(\alpha n) = \tilde{\Omega}(n^{1/4})$ for frequency estimation in the single-message shuffled model follows. Finally, we point out that while the above outline focused on the case where $n \leqslant B^2$, it turns out that this is essentially without loss of generality as the other case where $n > B^2$ can be reduced to the former (see Lemma 3.10).

Tightness of Lower Bounds The lower bounds sketched above are nearly tight. The upper bound of Theorem 1.1 follows from combining existing results showing that the single-message shuffled model provides privacy amplification of locally differentially private protocols [EFM+19, BBGN19c], with known locally differentially private protocols for frequency estimation [War65, EPK14, DJW18, BBGN19c]. In particular, as recently shown by Balle et al. [BBGN19c], a pure $(\varepsilon_L, 0)$-differentially private local randomizer yields a protocol in the shuffled model that is $\left(O\left(e^{\varepsilon_L}\sqrt{\frac{\log(1/\delta)}{n}}\right), \delta\right)$-differentially private and that has the same level of accuracy.[9] Then:

- When combined with RAPPOR [EPK14, DJW18], we get an upper bound of $\tilde{O}(n^{1/4})$ on the error.
- When combined with the B-randomized response [War65, ASZ19], we get an error upper bound of $\tilde{O}(\sqrt{B})$.

The full details appear in Appendix A. Put together, these imply that the minimum in our lower bound in Theorem 1.1 is tight (up to logarithmic factors). It also follows that the mutual information bound in Equation (3) is tight (up to logarithmic factors) for $\varepsilon_L = \ln(n) + O(1)$ and $\alpha = n^{-3/4}$ (which is the parameter settings corresponding to the single-message shuffled model); indeed, a stronger bound in Equation (3) would lead to larger lower bounds in the single-message shuffled model thereby contradicting the upper bounds discussed in this paragraph.

Lower Bound for Selection: Sharp Bound on Level-1 Weight of Probability Ratio Functions We now outline the proof of the nearly tight lower bound on the number of users required to solve the *selection* problem in the

[8] For clarity of exposition in this overview, we refrain from quantifying the likelihoods in each of these cases; for more details on this, we refer the reader to Section 3.3.

[9] Note that we cannot use the earlier amplification by shuffling result of [EFM+19], since it is only stated for $\varepsilon_L = O(1)$ whereas we need to amplify a much less private local protocol, having an ε_L close to $\ln n$.

single-message shuffled model (Theorem 1.2). The main component of the proof in this case is a lower bound of $\Omega(B)$ users for selection for $(\varepsilon_L, \delta_L)$-*local* DP protocols when $\varepsilon_L = \ln(n) + O(1)$.

In the case of local $(\varepsilon_L, 0)$-DP (i.e., pure) protocols, Ullman [Ull18] proved a lower bound $n = \Omega\left(\frac{B \log B}{(\exp(\varepsilon_L)-1)^2}\right)$. There are two different reasons why this lower bound is not sufficient for our purposes:

1. It does not rule out DP protocols with $\delta_L > 0$ (i.e., approximate protocols), which are necessary to consider for our application to the shuffled model.
2. For the low privacy setting of $\varepsilon_L = \ln(n) + O(1)$, the bound simplifies to $n = \tilde{\Omega}(B/n^2)$, i.e., $n = \tilde{\Omega}(B^{1/3})$, weaker than what we desire.

To prove our near-optimal lower bound, we remedy both of the aforementioned limitations by allowing positive values of δ_L and achieving a better dependence on ε_L. As in the proof of frequency estimation, we reduce proving Theorem 1.2 to the task of showing the following mutual information upper bound:

$$I((L, J); R(X_{L,J})) \leqslant \tilde{O}\left(\frac{1}{B}\right) + O(\delta_L(B + n)), \tag{4}$$

where L is a uniform random bit, J is a uniform random coordinate in $[B]$, and $X_{L,J}$ is uniform over the subcube $\{x \in \{0,1\}^B : x_J = L\}$. Indeed, once (4) holds and $\delta_L < o(1/(Bn))$, the chain rule implies that the mutual information between all users' messages and the pair (L, J) is at most $O\left(\frac{n \ln(B)}{B}\right)$. It follows by Fano's inequality that if $n = o(B)$, no analyzer can determine the pair (L, J) with high probability (which any protocol for selection must be able to do).

For any message z in the range of R, define the Boolean function $f_z(x) := \frac{\mathbb{P}[R(x)=z]}{\mathbb{P}[R(X_{L,J})=z]}$ where $x \in \{0,1\}^B$. Let $\mathbf{W}^1[f]$ denote the level-1 Fourier weight of a Boolean function f. To prove inequalities of the form (4), the prior work of Ullman [Ull18] shows that $I((L, J); R(X_{L,J}))$ is determined by $\mathbf{W}^1[f_z]$, up to normalization constants. In the case where $\delta_L = 0$ and $\varepsilon_L = \ln(n) + O(1)$, $f_z \in [0, e^{\varepsilon_L}]$, and by Parseval's identity $\mathbf{W}^1[f_z] \leqslant O(e^{2\varepsilon_L})$ for any message z, leading to

$$I((L, J); R(X_{L,J})) \leqslant O\left(\frac{e^{2\varepsilon_L}}{B}\right). \tag{5}$$

Unfortunately, for our choice of $\varepsilon_L = \ln(n) + O(1)$, (5) is weaker than (4).

To show (4), we depart from the previous approach in the following ways:

(a) We show that the functions f_z take values in $[0, O(e^{\varepsilon_L})]$ for *most* inputs x; this uses the $(\varepsilon_L, \delta_L)$-local DP of the local randomizer R.

(b) Using the *Level*-1 *inequality* from the analysis of Boolean functions [O'D14] (see Theorem 3.26 below), we upper bound $\mathbf{W}^1[g_z]$ by $O(\varepsilon_L)$, where g_z is the truncation of f_z defined by $g_z(x) = f_z(x)$ if $f_z(x) \leqslant O(n)$, and $g_z(x) = 0$ otherwise.

(c) We bound $I((L, J); R(X_{L,J}))$ by $\mathbf{W}^1[g_z]$, using the fact f_z is sufficiently close to its truncation g_z.

The above line of reasoning, formalized in Section 3.5, allows us to show

$$I((L, J); R(X_{L,J})) \leqslant O\left(\frac{\varepsilon_L}{B} + \delta \cdot (B + e^{\varepsilon_L})\right),$$

which is sufficient to establish that (4) holds.

Having proved a lower bound on the error of any $(\varepsilon + \ln n, \delta)$-local DP protocol for selection with $\varepsilon = O(1)$, the final step in the proof is to apply a lemma of [CSU+19] to deduce the desired lower bound in the single-message shuffled model.

1.3 Overview of Multi-Message Protocols

An important consequence of our lower bound in Theorem 1.1 is that one cannot achieve an error of $\tilde{O}(1)$ using *single-message* protocols. This in particular rules out any approach that uses the following natural two-step recipe for getting a private protocol in the shuffled model with accuracy better than in the local model:

1. Run any known locally differentially private protocol with a setting of parameters that enables high-accuracy estimation at the analyzer, but exhibits low privacy locally.

2. Randomly shuffle the messages obtained when each user runs step 1 on their input, and use the privacy amplification by shuffling bounds [EFM+19, BBGN19c] to improve the privacy guarantees.

Thus, shuffled versions of the B-randomized response [War65, ASZ19], RAPPOR [EPK14, DJW18, ASZ19], the Bassily–Smith protocol [BS15], TreeHist and Bitstogram [BNST17], and the Hadamard response protocol [ASZ19, AS19], will still incur an error of $\Omega(\min(\sqrt[4]{n}, \sqrt{B}))$. Local protocols!!

Moreover, although the single-message protocol of Cheu et al. [CSU+19] for binary aggregation (as well as the multi-message protocols given in [GPV19, BBGN19a, GMPV19, BBGN19b] for the more general task of real-valued aggregation) can be applied to the one-hot encodings of each user's input to obtain a multi-message protocol for frequency estimation with error $\tilde{O}(1)$, the communication per user would be $\Omega(B)$ bits, which is clearly undesirable.

Recall that the main idea behind (shuffled) randomized response is for each user to send their input with some probability, and random noise with the remaining probability. Similarly, the main idea behind (shuffled) Hadamard response is for each user to send a uniformly random index from the support of the Hadamard codeword corresponding to their input with some probability, and a random index from the entire universe with the remaining probability. In both protocols, the user is sending a message that either depends on their input or is noise; this restriction turns out to be a significant limitation. Our main insight is that multiple messages allows users to simultaneously send both types of messages, leading to a sweet spot with exponentially smaller error or communication.

Our protocols. We design a multi-message version of the private-coin Hadamard response of Acharya et al. [ASZ19, AS19] where each user sends a small *subset* of indices sampled uniformly at random from the support of the Hadamard codeword corresponding to their input, and in addition sends a small subset of indices sampled uniformly at random from the entire universe $[B]$. To get accurate results it is crucial that a subset of indices is sampled, as opposed to just a single index (as in the local model protocol of [ASZ19, AS19]). We show that in the regime where the number of indices sampled from inside the support of the Hadamard codeword and the number of noise indices sent by each user are both logarithmic, the resulting multi-message algorithm is private in the shuffled model, and it has polylogarithmic error and communication per user (see Theorem 4.1, Lemmas 4.4, 4.5, and 4.6 for more details).

A limitation of our private-coin algorithm outlined above is that the time for the analyzer to answer a single query is $\tilde{O}(n)$. This might be a drawback in applications where the analyzer is CPU-limited or where it is supposed to produce real-time answers. In the presence of public randomness, we design an algorithm that remedies this limitation, having error, communication per user, and query time all equal to $\tilde{O}(1)$. Furthermore, the frequency estimates of this algorithm have one-sided error, and never underestimate the frequency of an element. This algorithm is based on a multi-message version of randomized response combined in a delicate manner with the Count Min data structure [CM05a] (for more details, see Section 4.2). Previous work [BS15, BNST17] on DP have used Count Sketch [CCFC02], which is a close variant of Count Min, to go from frequency estimation to heavy hitters. In contrast, our use of Count Min has the purpose of reducing the amount of communication per user.

1.4 Applications

Heavy Hitters. Another algorithmic task that is closely related to frequency estimation is computing the *heavy hitters* in a dataset distributed across n users, where the goal of the analyzer is to (approximately) retrieve the identities and counts of all elements that appear at least τ times, for a given threshold τ. It is well-known that in the central DP model, it is possible to compute τ-heavy hitters for any $\tau = \tilde{\Theta}(1)$ whereas in the local DP model, it is possible to compute τ-heavy hitters if and only if $\tau = \tilde{\Theta}(\sqrt{n})$. By combining with known reductions (e.g., from Bassily et al. [BNST17]), our multi-message protocols for frequency estimation yield multi-message protocols for computing the τ-heavy hitters with $\tau = \tilde{\Theta}(1)$ and total communication of $\tilde{\Theta}(1)$ bits per user (for more details, see Appendix D).

Range Counting. In range counting, each of the n users is associated with a point in $[B]^d$ and the goal of the analyzer is to answer arbitrary queries of the form: given a rectangular box in $[B]^d$, how many of the points lie

in it?[10] This is a basic algorithmic primitive that captures an important family of database queries and is useful in geographic applications. This problem has been well-studied in the central model of DP, where Chan et al. [CSS11] obtained an upper bound of $(\log B)^{O(d)}$ on the error (see Section 1.5 for more related work). It has also been studied in the local DP model [CKS19]; in this case, the error has to be at least $\Omega(\sqrt{n})$ even for $d = 1$.

We obtain private protocols for range counting in the multi-message shuffled model with exponentially smaller error than what is possible in the local model (for a wide range of parameters). Specifically, we give a private-coin multi-message protocol with $(\log B)^{O(d)}$ messages per user each of length $O(\log n)$ bits, error $(\log B)^{O(d)}$, and query time $\tilde{O}(n \log^d B)$. Moreover, we obtain a public-coin protocol with similar communication and error but with a much smaller query time of $\tilde{O}(\log^d B)$ (see Section 5 for more details).

We now briefly outline the main ideas behind our multi-message protocols for range counting. We first argue that even for $d = 2$, the total number of queries is $\Theta(B^2)$ and the number of possible queries to which a user positively contributes is also $\Theta(B^2)$. Thus, direct applications of DP algorithms for aggregation or for frequency estimation would result in polynomial error and polynomial communication per user. Instead, we combine our multi-message protocol for frequency estimation (Theorem 1.3) with a communication-efficient implementation, in the multi-message shuffled model, of the space-partitioning data structure used in the central model protocol of Chan et al. [CSS11]. The idea is to use a collection $\mathcal{B}$ of $O(B \log^d B)$ d-dimensional rectangles in $[B]^d$ (so-called *dyadic intervals*) with the property that an arbitrary rectangle can be formed as the disjoint union of $O(\log^d B)$ rectangles from $\mathcal{B}$. Furthermore, each point in $[B]^d$ is contained in $O(\log^d B)$ rectangles from $\mathcal{B}$. This means that it suffices to release a private count of the number of points inside each rectangle in $\mathcal{B}$ — a frequency estimation task where each user input contributes to $O(\log^d B)$ buckets. To turn this into a protocol with small maximum communication in the shuffled model, we develop an approach analogous to the matrix mechanism [LHR$^+$10, LM12]. We argue that the transformation of the aforementioned central model algorithm for range counting into a private protocol in the multi-message shuffled model with small communication and error is non-trivial and relies on the specific protocol structure. In fact, the state-of-the-art range counting algorithm of Dwork et al. [DNRR15] in the central model does not seem to transfer to the shuffled model.

M-Estimation of Median. A very basic statistic of any dataset of real numbers is its *median*. For simplicity, suppose our dataset consists of real numbers lying in $[0, 1]$. It is well-known that there is no DP algorithm for estimating the *value* of the median of such a dataset with error $o(1)$ (i.e., outputting a real number whose absolute distance to the true median is $o(1)$) [Vad17, Section 3]. This is because the median of a dataset can be highly sensitive to a single data point when there are not many individual data points near the median. Thus in the context of DP, one has to settle for weaker notions of median estimation. One such notion is *M-estimation*, which amounts to finding a value $\tilde{x}$ that approximately minimizes $\sum_i |x_i - \tilde{x}|$ (recall that the median is the minimizer of this objective). This notion has been studied in previous work on DP including by [Lei11, DJW18] (for more on related work, see Section 1.5 below). Our private range counting protocol described above yields a multi-message protocol with communication $\tilde{O}(1)$ per user and that M-estimates the median up to error $\tilde{O}(1)$, i.e., outputs a value $y \in [0, 1]$ such that $\sum_i |x_i - y| \leqslant \min_{\tilde{x}} \sum_i |x_i - \tilde{x}| + \tilde{O}(1)$ (see Theorem E.1 in Appendix E). Beyond M-estimation of the median, our work implies private multi-message protocols for estimating *quantiles* with $\tilde{O}(1)$ error and $\tilde{O}(1)$ bits of communication per user (see Appendix E for more details).

1.5 Related Work

Shuffled Privacy Model. Following the proposal of the Encode-Shuffle-Analyze architecture by Bittau et al. [BEM$^+$17], several recent works have sought to formalize the trade-offs in the shuffled model with respect to standard local and central DP [EFM$^+$19, BBGN19c] as well as devise private schemes in this model for tasks such as secure aggregation [CSU$^+$19, BBGN19c, GPV19, BBGN19a, GMPV19, BBGN19b]. In particular, for the task of *real* aggregation, Balle et al. [BBGN19c] showed that in the single-message shuffled model, the optimal

[10] We formally define range queries as a special case of counting queries in Section 5.

error is $\Theta(n^{1/6})$ (which is better than the error in the local model which is known to be $\Theta(n^{1/2})$).[11] By contrast, recent follow-up work gave multi-message protocols for the same task with error and communication of $\tilde{O}(1)$ [GPV19, BBGN19a, GMPV19, BBGN19b][12]. Our work is largely motivated by the aforementioned body of works demonstrating the power of the shuffled model, namely, its ability to enable private protocols with lower error than in the local model while placing less trust in a central server or curator.

Wang et al. [WXD+19] recently designed an extension of the shuffled model and analyzed its trust properties and privacy-utility tradeoffs. They studied the basic task of frequency estimation, and benchmarked several algorithms, including one based on single-message shuffling. However, they did not consider improvements through multi-message protocols, such as the ones we propose in this work. Very recently, Erlingsson et al. [EFM+20] studied multi-message ("report fragmenting") protocols for frequency estimation in a practical shuffled model setup. Though they make use of a sketching technique, like we do, their methods cannot be parameterized to have communication and error polylogarithmic in n and B (which our Theorem 1.3 achieves). This is a result of using an estimator (based on computing a mean) that does not yield high-probability guarantees.

Private Frequency Estimation, Heavy Hitters, and Median. Frequency estimation and its extensions (considered below) has been extensively studied in concrete computational models including data structures, sketching, streaming, and communication complexity, (e.g., [MG82, CCFC02, EV03, CM05a, CM05b, CH08, MP80, MRL98, GK+01, GGI+02, YZ13, KLL16]). Heavy hitters and frequency estimation have also been studied extensively in the standard models of DP, e.g., [War65, HKR12, BS15, BNST17, WBLJ17, BNS18, AS19]. The other problems we consider in the shuffled model, namely, range counting, M-estimation of the median, and quantiles, have been well-studied in the literature on data structures and sketching [CY20] as well as in the context of DP in the central and local models. Dwork and Lei [DL09] initiated work on establishing a connection between DP and robust statistics, and gave private estimators for several problems including the median, using the paradigm of propose-test-release. Subsequently, Lei [Lei11] provided an approach in the central DP model for privately releasing a wide class of M-estimators (including the median) that are statistically consistent. While such M-estimators can also be obtained indirectly from non-interactive release of the density function [WZ10], the aforementioned approach exhibits an improved rate of convergence. Furthermore, motivated by risk bounds under privacy constraints, Duchi et al. [DJW18] provided private versions of information-theoretic bounds for minimax risk of M-estimation of the median.

Frequency estimation can be viewed as the problem of distribution estimation in the ℓ_∞ norm where the distribution to be estimated is the empirical distribution of a dataset $(x_1, \ldots, x_n)$. Some works [YB17, KBR16] have established tight lower bounds for locally differentially private distribution estimation in the weak privacy setting with loss instead given by either ℓ_1 or ℓ_2^2. However, their techniques proceed by using Assouad's method [DJW18] and are quite different from the approach we use for the ℓ_∞ norm in the proof of Theorem 1.1 (specifically, in the proof of Theorem 3.3).

We also note that an anti-concentration lemma qualitatively similar to our Lemma 3.15 was used by Chan et al. [CSS12, Lemma 3] to prove lower bounds on private aggregation, but they operated in a multi-party setting with communication limited by a sparse communication graph. After the initial release of this paper, Ghazi et al. [GGK+20] proved a similar anti-concentration lemma to establish a lower bound on private summation for protocols with short messages. The lemmas in both of these papers do not apply to the more general case of frequency estimation with an arbitrary number B of buckets, as is the case throughout this paper.

Range Counting. Range counting queries have also been an important subject of study in several areas including database systems and algorithms (see [Cor11] and the references therein). Early works on differentially private frequency estimation , e.g., [Dwo06, HLM12], apply naturally to range counting, though the approach of summing up frequencies yields large errors for queries with large ranges.

[11] Although the single-message real summation protocol of Balle et al. [BBGN19c] uses the B-ary randomized response, when combined with their lower bound on single-message protocols, it does not imply any lower bound on single-message frequency estimation protocols. The reason is that their upper bound doe not use the ℓ_∞ error bound for the B-ary randomized response as a black box.

[12] A basic primitive in these protocols is a "split-and-mix" procedure that goes back to the work of Ishai et al. [IKOS06].

For $d = 1$, Dwork et al. [DNPR10] obtained an upper bound of $O\left(\frac{\log^2 B}{\varepsilon}\right)$ and a lower bound of $\Omega(\log B)$ for obtaining $(\varepsilon, 0)$-DP. Chan et al. [CSS11] extended the analysis to d-dimensional range counting queries in the central model, for which they obtained an upper bound of roughly $(\log B)^{O(d)}$. Meanwhile, a lower bound of Muthukrishnan and Nikolov [MN12] showed that for $n \approx B$, the error is lower bounded by $\Omega\left((\log n)^{d-O(1)}\right)$. Since then, the best-known upper bound on the error for general d-dimensional range counting has been $(\log B + \log(n)^{O(d)})/\varepsilon$ [DNRR15], obtained using ideas from [DNPR10, CSS11] along with a k-d tree-like data structure. We note that for the special case of $d = 1$, it is known how to get a much better dependence on B in the central model, namely, exponential in $\log^* B$ [BNS13, BNSV15].

Xiao et al. [XWG10] showed how to obtain private range count queries by using Haar wavelets, while Hay et al. [HRMS10] formalized the method of maintaining a hierarchical representation of data; the aforementioned two works were compared and refined by Qardaji et al. [QYL13]. Cormode et al. [CKS19] showed how to translate many of the previous ideas to the local model of DP. We also note that the matrix mechanism of Li et al. [LHR$^+$10, LM12] also applies to the problem of range counting queries. An alternate line of work for tackling multi-dimensional range counting that relied on developing private versions of k-d trees and quadtrees was presented by Cormode et al. [CPS$^+$12].

Secure Multi-Party Computation. If we allow user interaction in the computation of the queries, then there is a rich theory, within cryptography, of *secure multi-party computation* (SMPC) that allows $f(x_1, \ldots, x_n)$ to be computed without revealing anything about x_i except what can be inferred from $f(x_1, \ldots, x_n)$ itself (see, e.g., the book of Cramer et al. [CDN15]). Kilian et al. [KMSZ08] studied SMPC protocols for heavy hitters, obtaining near-linear communication complexity with a multi-round protocol. In contrast, all results in this paper are about *non-interactive* (single-round) protocols in the shuffled-model (in the multi-message setting, all messages are generated at once). Though generic SMPC protocols can be turned into differentially private protocols (see, e.g., Section 10.2 in [Vad17] and the references therein), they almost always use multiple rounds, and often have large overheads compared to the cost of computing $f(x_1, \ldots, x_n)$ in a non-private setting.

1.6 Organization

We start with some notation and background in Section 2. In Section 3, we prove our lower bounds for single-message protocols in the shuffled model; corresponding upper bounds can be found in Appendix A. In Section 4, we present and analyze our multi-message protocols for frequency estimation (with missing proofs in Appendix C). In Section 5, we give our multi-message protocols for range counting. We conclude with some interesting open questions in Section 6. The proof of Corollary 1.4 is given in Appendix B. The reduction from frequency estimation to heavy hitters appears in Appendix D. The reduction from range counting to M-estimation of the median and quantiles is given in Appendix E.

2 Preliminaries

Notation. For any positive integer B, let $[B] = \{1, 2, \ldots, B\}$. For any set $\mathcal{Y}$, we denote by $\mathcal{Y}^*$ the set consisting of sequences of elements of $\mathcal{Y}$, i.e., $\mathcal{Y}^* = \bigcup_{n \geqslant 0} \mathcal{Y}^n$. Suppose $\mathcal{S}$ is a multiset whose elements are drawn from a set $\mathcal{X}$. With a slight abuse of notation, we will write $\mathcal{S} \subset \mathcal{X}$ and for $x \in \mathcal{X}$, we write $m_{\mathcal{S}}(x)$ to denote the *multiplicity* of x in $\mathcal{S}$. For an element $x \in \mathcal{X}$ and a non-negative integer k, let $k \times \{x\}$ denote the multiset with k copies of x (e.g., $3 \times \{x\} = \{x, x, x\}$). For a positive real number a, we use $\log(a)$ to denote the logarithm base 2 of a, and $\ln(a)$ to denote the natural logarithm of a. Let $\mathrm{Bin}(n, p)$ denote the binomial distribution with parameters $n > 0$ and $p \in (0, 1)$.

2.1 Differential Privacy

We now introduce the basics of differential privacy that we will need. Fix a finite set $\mathcal{X}$, the space of reports of users. A *dataset* is an element of $\mathcal{X}^*$, namely a tuple consisting of elements of $\mathcal{X}$. Let $\mathrm{hist}(X) \in \mathbb{N}^{|\mathcal{X}|}$ be the

histogram of X: for any $x \in \mathcal{X}$, the xth component of $\mathrm{hist}(X)$ is the number of occurrences of x in the dataset X. We will consider datasets X, X' to be *equivalent* if they have the same histogram (i.e., the ordering of the elements $x_1, \ldots, x_n$ does not matter). For a multiset $\mathcal{S}$ whose elements are in $\mathcal{X}$, we will also write $\mathrm{hist}(\mathcal{S})$ to denote the histogram of $\mathcal{S}$ (so that the xth component is the number of copies of x in $\mathcal{S}$).

Let $n \in \mathbb{N}$, and consider a dataset $X = (x_1, \ldots, x_n) \in \mathcal{X}^n$. For an element $x \in \mathcal{X}$, let $f_X(x) = \frac{\mathrm{hist}(X)_x}{n}$ be the *frequency* of x in X, namely the fraction of elements of X which are equal to x. Two datasets X, X' are said to be *neighboring* if they differ in a single element, meaning that we can write (up to equivalence) $X = (x_1, \ldots, x_{n-1}, x_n)$ and $X' = (x_1, \ldots, x_{n-1}, x'_n)$. In this case, we write $X \sim X'$. Let $\mathcal{Z}$ be a set; we now define the differential privacy of a randomized function $P : \mathcal{X}^n \to \mathcal{Z}$:

Definition 2.1 (Differential privacy [DMNS06, DKM+06])**.** A randomized algorithm $P : \mathcal{X}^n \to \mathcal{Z}$ is (ε, δ)-*differentially private* if for every pair of neighboring datasets $X \sim X'$ and for every set $\mathcal{S} \subset \mathcal{Z}$, we have

$$\mathbb{P}[P(X) \in \mathcal{S}] \leqslant e^{\varepsilon} \cdot \mathbb{P}[P(X') \in \mathcal{S}] + \delta,$$

where the probabilities are taken over the randomness in P. Here, $\varepsilon \geqslant 0, \delta \in [0, 1]$.

We will use the following compositional property of differential privacy.

Lemma 2.1 (Post-processing, e.g., [DR14a])**.** *If P is (ε, δ)-differentially private, then for every randomized function A, the composed function $A \circ P$ is (ε, δ)-differentially private.*

2.2 Shuffled Model

We briefly review the *shuffled model* of differential privacy [BEM+17, EFM+19, CSU+19]. The input to the model is a dataset $(x_1, \ldots, x_n) \in \mathcal{X}^n$, where item $x_i \in \mathcal{X}$ is held by user i. A protocol in the shuffled model is the composition of three algorithms:

- The *local randomizer* $R : \mathcal{X} \to \mathcal{Y}^*$ takes as input the data of one user, $x_i \in \mathcal{X}$, and outputs a sequence $(y_{i,1}, \ldots, y_{i,m_i})$ of *messages*; here m_i is a positive integer.
- The *shuffler* $S : \mathcal{Y}^* \to \mathcal{Y}^*$ takes as input a sequence of elements of $\mathcal{Y}$, say $(y_1, \ldots, y_m)$, and outputs a random permutation, i.e., the sequence $(y_{\pi(1)}, \ldots, y_{\pi(m)})$, where $\pi \in S_m$ is a uniformly random permutation on $[m]$. The input to the shuffler will be the concatenation of the outputs of the local randomizers.
- The *analyzer* $A : \mathcal{Y}^* \to \mathcal{Z}$ takes as input a sequence of elements of $\mathcal{Y}$ (which will be taken to be the output of the shuffler) and outputs an answer in $\mathcal{Z}$ which is taken to be the output of the protocol P.

We will write $P = (R, S, A)$ to denote the protocol whose components are given by R, S, and A. The main distinction between the shuffled and local model is the introduction of the shuffler S between the local randomizer and the analyzer. Similar to the local model, in the shuffled model the analyzer is untrusted; hence privacy must be guaranteed with respect to the input to the analyzer, i.e., the output of the shuffler. Formally, we have:

Definition 2.2 (Differential privacy in the shuffled model, [EFM+19, CSU+19])**.** A protocol $P = (R, S, A)$ is (ε, δ)-*differentially private* if, for any dataset $X = (x_1, \ldots, x_n)$, the algorithm

$$(x_1, \ldots, x_n) \mapsto S(R(x_1), \ldots, R(x_n))$$

is (ε, δ)-differentially private.

Notice that the output of $S(R(x_1), \ldots, R(x_n))$ can be simulated by an algorithm that takes as input the *multiset* consisting of the union of the elements of $R(x_1), \ldots, R(x_n)$ (which we denote as $\bigcup_i R(x_i)$, with a slight abuse of notation) and outputs a uniformly random permutation of them. Thus, by Lemma 2.1, it can be assumed without loss of generality for privacy analyses that the shuffler simply outputs the multiset $\bigcup_i R(x_i)$. For the purpose of analyzing accuracy of the protocol $P = (R, S, A)$, we define its *output* on the dataset $X = (x_1, \ldots, x_n)$ to be $P(X) := A(S(R(x_1), \ldots, R(x_n)))$. We also remark that the case of *local differential privacy*, formalized in Definition 2.3, is a special case of the shuffled model where the shuffler S is replaced by the identity function.

Definition 2.3 (Local differential privacy [KLN+08])**.** A protocol $P = (R, A)$ is (ε, δ)-*differentially private in the local model* (or (ε, δ)-*locally differentially private*) if the function $x \mapsto R(x)$ is (ε, δ)-differentially private in the sense of Definition 2.1. We say that the *output* of the protocol P on an input dataset $X = (x_1, \ldots, x_n)$ is $P(X) := A(R(x_1), \ldots, R(x_n))$.

3 Single-Message Lower and Upper Bounds

In this section, we prove Theorem 1.1, which determines (up to polylogarithmic factors) the accuracy of frequency estimation in the single-message shuffled model. Using similar techniques, we also prove Theorem 1.2, which establishes a tight (up to polylogarithmic factors) lower bound on the number of users required to solve the selection problem in the single-message shuffled model. Our theorems give tight versions (see Corollary 3.2) of Corollaries 30 and 32 of [CSU+19], which were each off from the respective optimal bounds by a polynomial of degree 17. We will use the following definition throughout this section:

Definition 3.1 ((α, β)-accuracy)**.** Let $\mathcal{Z}$ be a finite set, let $B \in \mathbb{N}$, and let $e_v \in \{0,1\}^B$ be the binary indicator vector with $(e_v)_j = 1$ if and only if $j = v$. We say that a (randomized) protocol $P : [B]^n \to [0,1]^B$ for frequency estimation is (α, β)-*accurate* if for each dataset $X = (x_1, \ldots, x_n) \in [B]^n$, we have that

$$\mathbb{P}_P \left[\max_{j \in [B]} \left| P(X)_j - \frac{1}{n} \sum_{i=1}^{n} (e_{x_i})_j \right| \leqslant \alpha \right] \geqslant 1 - \beta.$$

Often we will either have $P = (R, A)$ for a local randomizer R and an analyzer A (corresponding to the local model) or $P = (R, S, A)$ (corresponding to the shuffled model). In such a case, we will slightly abuse notation and refer to the local randomizer $R : [B] \to \mathcal{Z}$ as (α, β)-*accurate* if there exists an analyzer $A : \mathcal{Z}^n \to [0,1]^B$ such that the corresponding local or shuffled-model protocol is (α, β)-accurate.

Theorem 3.1 establishes lower bounds on the (additive) error of frequency estimation in the single-message differentially-private shuffled model.

Theorem 3.1 (Lower bound for single-message differentially private frequency estimation)**.** *There is a sufficiently small constant $c > 0$ such that the following holds: Suppose $n, B \in \mathbb{N}$ with $n \geqslant 1/c$, and $0 < \delta < c/n$. Any (ε, δ)-differentially private n-user single-message shuffled model protocol that is $(\alpha, 1/4)$-accurate satisfies:*

$$\alpha \geqslant \begin{cases} \Omega\left(\dfrac{\log B}{n \log\log B}\right) & \text{for} \quad \frac{\log B}{c \log\log B} \leqslant n \leqslant (\log^2 B)(\log\log B), \qquad \text{(“Small-sample”)} \quad (6) \\ \Omega\left(\dfrac{1}{n^{3/4}\sqrt[4]{\log n}}\right) & \text{for} \quad (\log^2 B)(\log\log B) \leqslant n \leqslant \frac{B^2}{\log B}, \qquad \text{(“Intermediate-sample”)} \quad (7) \\ \Omega\left(\dfrac{\sqrt{B}}{n\sqrt{\log B}}\right) & \text{for} \quad n > \frac{B^2}{\log B}. \qquad \text{(“Large-sample”)} \quad (8) \end{cases}$$

Note that the lower bound on the additive error α is divided into 3 cases, which we call the *small-sample regime* (6), the *intermediate-sample regime* (7), and the *large-sample regime* (8). While the division into separate regimes makes our bounds more technical to state, we point out that this seems necessary in light of the very different protocols that achieve near-optimality in the various regimes (as discussed in Section 1.2 and Appendix B). Moreover, the bound for the low-sample regime of Theorem 3.1 is established in Lemma 3.11, while the bounds for the intermediate-sample and large-sample regimes of Theorem 3.1 are established in Corollary 3.13 and Lemma 3.19, respectively. We note that the proof of the intermediate-sample regime (7) is the most technically involved and constitutes the bulk of the proof of Theorem 3.1.

Furthermore, we observe that the lower bounds (6), (7), and (8) also hold, up to constant factors, for the *expected error* $\mathbb{E}_R \left[\max_{j \in [B]} \left| P(X)_j - \frac{1}{n} \sum_{i=1}^{n} (e_{x_i})_j \right| \right]$ of P on a dataset X. This follows as an immediate consequence of Theorem 3.1 and Markov's inequality.

In the course of proving Theorem 3.1 in the small-sample regime (i.e., (6)), we shall see that the constants can be chosen in such a way so as to establish Corollary 3.2 below (in particular, Corollary 3.2 follows from Lemma 3.9):

Corollary 3.2 (Lower bound for constant-error frequency estimation)**.** *Let c be the constant of Theorem 3.1. If P is a $(1,\delta)$-differentially private protocol for frequency estimation in the shuffled model with $\delta < c/n$ which is $(1/10, 1/10)$-accurate, then $n \geqslant \Omega\left(\frac{\log B}{\log\log B}\right)$.*

Corollary 3.2 improves upon Corollary 32 of [CSU+19], both in the lower bound on the error (which was $\Omega(\log^{1/17} B)$ in [CSU+19]) and on the dependence on δ (which was $\delta < O(n^{-8})$ in [CSU+19]).

The primary component of the proof of Theorem 3.1 is a lower bound on the additive error of $(\varepsilon_L, \delta_L)$-*locally* differentially private protocols $P = (R, A)$, when both $\varepsilon_L \gg 1$ (the *low-privacy* setting) and $\delta_L > 0$ simultaneously hold (see Lemma 3.5). In particular, we prove the following:

Theorem 3.3 (Lower bound for locally differentially private frequency estimation)**.** *There is a sufficiently small constant $c > 0$ such that the following holds. Suppose $n, B \in \mathbb{N}$ with $n \geqslant 1/c$, and that $\varepsilon_L, \delta_L > 0$ with $\delta_L < c\min\{1/(n\log n), \exp(-\varepsilon_L)\}$. Any $(\varepsilon_L, \delta_L)$-locally differentially private protocol that is $(\alpha, 1/4)$-accurate satisfies:*

$$\alpha \geqslant \begin{cases} \Omega\left(\frac{\ln B}{n\varepsilon_L}\right) & \text{for} \quad n \geqslant \frac{\ln B}{c\varepsilon_L}, & \text{("Small-sample")} \quad (9) \\ \tilde{\Omega}\left(\frac{1}{\sqrt{n}\cdot\exp(\varepsilon_L/4)}\right) & \text{for} \quad n \geqslant (\ln B)\exp(\varepsilon_L/2) \\ & \text{and } \frac{2}{3}\cdot\ln(n) \leqslant \varepsilon_L + \ln(1+\varepsilon_L) + \frac{1}{c} \leqslant 2\ln(B), & \text{("Intermediate-sample")} \quad (10) \\ \tilde{\Omega}\left(\frac{1}{n^{2/3}}\right) & \text{for} \quad \ln^{3/2}(B) \leqslant n \leqslant B^3 \text{ and } \varepsilon_L \leqslant \frac{2}{3}\cdot\ln(n), & \text{("Intermediate-sample")} \quad (11) \\ \tilde{\Omega}\left(\frac{\sqrt{B}}{n}\right) & \text{for} \quad n \geqslant B^2 \text{ and } \varepsilon_L \leqslant 2\ln(B), & \text{("Large-sample")} \quad (12) \\ \tilde{\Omega}\left(\frac{B}{n}\right) & \text{for} \quad n \geqslant B^3 \text{ and } \varepsilon_L \leqslant 2\ln(B). & \text{("Large-sample")} \quad (13) \end{cases}$$

Again, the lower bound is divided into cases—the bound for low-sample regime of Theorem 3.3 (namely, (9)) is established in Lemma 3.11, while the bounds for the intermediate-sample (namely, (10) and (11)) and large-sample (namely, (12) and (13)) regimes are established in Lemma 3.12 and Lemma 3.18, respectively.

It turns out that Theorem 3.1 is tight in each of the three regimes (small-sample, intermediate-sample, and large-sample), up to polylogarithmic factors in B and n, as shown by Theorem 3.4:

Theorem 3.4 (Upper bound for single-message shuffled DP frequency estimation)**.** *Fix $B, n \in \mathbb{N}$, $\delta = n^{-O(1)}$, and $\varepsilon \leqslant 1$ that satisfies $\varepsilon = \omega(\ln^2(n)/\min\{\sqrt{B}, \sqrt{n}\})$. For $n \in \mathbb{N}$, there is a shuffled model protocol $P = (R, S, A)$ so that for any $X = (x_1, \ldots, x_n) \in [B]^n$, the frequency estimates $P(X) \in [0,1]^B$ produced by P satisfy*

$$\mathbb{E}\left[\max_{j\in[B]}\left|P(X)_j - \frac{1}{n}\sum_{i=1}^{n}(e_{x_i})_j\right|\right] \leqslant \begin{cases} O\left(\frac{\log B}{n}\right) & \text{for} \quad n \leqslant \frac{\varepsilon^2\log^2 B}{\log^3\log B}, & (14) \\ O\left(\frac{\ln^{3/4}(n)\sqrt{\log B}}{n^{3/4}\sqrt{\varepsilon}}\right) & \text{for} \quad \frac{\varepsilon^2\log^2 B}{\log^3\log B} \leqslant n \leqslant B^2, & (15) \\ O\left(\frac{\sqrt{B\ln(n)\ln(B)}}{n\varepsilon}\right) & \text{for} \quad n > B^2. & (16) \end{cases}$$

The proof of Theorem 3.4 follows by combining existing protocols for locally differentially private frequency estimation with the privacy amplification result of [BBGN19c]. For completeness, we provide the proof in Appendix A.

The remainder of this section is organized as follows. In Section 3.1 we collect some tools that will be used in the proofs of our error lower bounds. In Section 3.2 we establish Theorem 3.1 in the small-sample regime (i.e., (6)). In Sections 3.3 and 3.4 we establish Theorem 3.1 in the intermediate and large-sample regimes (i.e., (7) and (8). Finally, in Section 3.5 we show how similar techniques used to prove Theorem 3.1 lead to a tight lower bound on the selection problem (Theorem 3.22).

Remark 3.1. Before proceeding with the proof of Theorem 3.3 (and thus Theorem 3.1), we briefly explain why the approach of [CSU+19], which establishes a weak variant of Theorem 3.3, cannot obtain the tight bounds that we are able to achieve here. Recall that this approach used:

(i) in a black-box manner, known lower bounds of Bassily and Smith [BS15] and Duchi et al. [DJW18] on the error of "pure" $(\varepsilon_L, 0)$-locally differentially private frequency estimation protocols, together with

(ii) a result of Bun et al. [BNS18] stating that by modifying an $(\varepsilon_L, \delta_L)$-locally differentially private protocol, one can produce an $(8\varepsilon_L, 0)$-locally differentially private protocol without significant loss in accuracy.

It seems to be quite challenging to get tight bounds in the single-message shuffled model using this two-step technique. This is because when $\varepsilon_L \approx \ln n$, the error lower bounds for $(\varepsilon_L, 0)$-differentially private frequency estimation in the local model decay as $\exp(-a\varepsilon_L)$ for some constant a. Suppose that for some constant $C \geqslant 1$, one could show that by modifying any $(\varepsilon_L, \delta_L)$-locally differentially private protocol one could obtain a $(C\varepsilon_L, 0)$-locally differentially private protocol without a large loss in accuracy (for instance, Bun et al. [BNS18] achieves $C = 8$.) Then the resulting error lower bound for shuffled-model protocols would decay as $\exp(-aC \ln n) = n^{-aC}$. This bound will necessarily be off by a polynomial in n unless we can determine the optimal constant C. The proof for $C = 8$ [BNS18, CSU+19] is already quite involved, and in order for this approach to guarantee tight bounds in the single-message setup, we would need to achieve $C = 1$, i.e., turn any $(\varepsilon_L, \delta_L)$-locally differentially private protocol into one with $\delta_L = 0$ and essentially no increase in ε_L whatsoever.

3.1 Preliminaries for Lower Bounds

In this section we collect some useful definitions and lemmas. Throughout this section, we will use the following notational convention:

Definition 3.2 (Notation $p_{x,\mathcal{S}}$)**.** For a fixed local randomizer $R : \mathcal{X} \to \mathcal{Z}$ (which will be clear from the context), and for $x \in \mathcal{X}, \mathcal{S} \subset \mathcal{Z}, z \in \mathcal{Z}$, we will write $p_{x,\mathcal{S}} := \mathbb{P}_R[R(x) \in \mathcal{S}]$ and $p_{x,z} := \mathbb{P}_R[R(x) = z]$, where the probability is over the randomness of R.

Moreover, we will additionally write P_x to denote the distribution on $\mathcal{Z}$ given by $R(x)$. In particular, the density of P_x at $z \in \mathcal{Z}$ is $p_{x,z}$.

We say that a local randomizer $R : \mathcal{X} \to \mathcal{Z}$ is (ε, δ)-differentially private in the n-user shuffled model if the composed protocol $(x_1, \ldots, x_n) \mapsto S(R(x_1), \ldots, R(x_n))$ is (ε, δ)-differentially private. Lemma 3.5 establishes that a protocol R that is (ε, δ)-differentially private in the shuffled model is in fact $(\varepsilon + \ln n, \delta)$-differentially private in the *local* model of differential privacy, which means that the function $x \mapsto R(x)$ is itself $(\varepsilon + \ln n, \delta)$-differentially private.

Lemma 3.5 (Theorem 6.2, [CSU+19])**.** *Suppose $\mathcal{X}, \mathcal{Z}$ are finite sets. If $R : \mathcal{X} \to \mathcal{Z}$ is (ε, δ)-differentially private in the n-user single-message shuffled model, then R is $(\varepsilon + \ln n, \delta)$-locally differentially private.*

(That is, for all $x, y \in \mathcal{X}$, and for all $\mathcal{S} \subset \mathcal{Z}$, we have

$$p_{y,\mathcal{S}} \leqslant p_{x,\mathcal{S}} \cdot e^{\varepsilon} n + \delta.$$

Recall $p_{y,\mathcal{S}} = \mathbb{P}[R(y) \in \mathcal{S}], p_{x,\mathcal{S}} = \mathbb{P}[R(x) \in \mathcal{S}]$ per Definition 3.2.)

As discussed in Section 1, to prove Theorem 3.1 (as well as Theorem 3.22), we use similar ideas to those in the the results of [DJW18, BS15] to *directly* derive a lower bound on the error of locally private frequency estimation in the low and approximate privacy setting (i.e., for $(\varepsilon_L, \delta_L)$-locally differentially private protocols with $\varepsilon_L \approx \ln n$ and $\delta_L > 0$). By Lemma 3.5, doing so suffices to derive a lower bound for frequency estimation in the single-message shuffled model. Our lower bounds for local-model protocols, on their own, may be of independent interest. The locally private frequency estimation lower bounds of [DJW18, BS15], as well as our proof, rely on Fano's inequality, which we recall as Lemma 3.6 below.

For random variables X, Y distributed on a finite set $\mathcal{X}$, let $I(X; Y)$ denote the mutual information between X, Y. We refer the reader to [CT91] for more background on basic information theory.

Lemma 3.6 (Fano's inequality)**.** *Suppose Z, Z' are jointly distributed random variables on a finite set $\mathcal{Z}$. Then*

$$\mathbb{P}[Z = Z'] \leqslant \frac{I(Z; Z') + 1}{\log |\mathcal{Z}|}.$$

Additionally, it will be useful to phrase some of our arguments in terms of the hockey stick divergence between distributions:

Definition 3.3 (Hockey stick divergence)**.** Suppose D, F are probability distributions on a space $\mathcal{X}$ that are absolutely continuous with respect to some measure G on $\mathcal{X}$; let the densities of D, F with respect to G be given by d, f. For any $\rho \geqslant 1$, the *hockey stick divergence of order ρ* between D, F is defined as:

$$\mathscr{D}_\rho(D||F) := \int_{\mathcal{X}} [d(x) - \rho \cdot f(x)]_+ \, dG(x),$$

where $[a]_+ = \max\{a, 0\}$ for $a \in \mathbb{R}$.

The *total variation distance* $\Delta(D, F)$ between two distributions D, F on a set $\mathcal{X}$ is defined as

$$\sup_{\mathcal{S} \subseteq \mathcal{X}} |D(\mathcal{S}) - F(\mathcal{S})| \ .$$

Note that for $\rho = 1$ the hockey stick divergence of order ρ is the total variation distance, i.e., $\mathscr{D}_1(D||F) = \mathscr{D}_1(F||D) = \Delta(D, F)$. The following fact is well-known:

Fact 3.7 (Characterization of hockey stick divergence)**.** *Using the notation of Definition 3.3, we have:*

$$\mathscr{D}_\rho(D||F) = \sup_{\mathcal{S} \in \mathcal{X}} \left(D(\mathcal{S}) - \rho \cdot F(\mathcal{S})\right).$$

For a boolean function $f : \{0,1\}^B \to \mathbb{R}$, the *Fourier transform* of f is given by the function $\hat{f}(S) := \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)} \left[f(x) \cdot (-1)^{\sum_{j=1}^B x_j \cdot \mathbb{1}[j \in S]} \right]$, where $S \subseteq [B]$ is any subset. The *Fourier weight at degree 1* of such a function is defined by $\mathbf{W}^1[f] := \sum_{j \in [B]} \hat{f}(\{j\})^2$. We refer the reader to [O'D14] for further background on the Fourier analysis of boolean functions.

3.2 Small-Sample Regime

In this section we establish Theorem 3.1 in the case that $n \leqslant \log^2 B$ (i.e., we prove (6)). As we noted following Lemma 3.5, we will prove a slightly more general statement, allowing R to be any $(\varepsilon + \ln n, \delta)$-*locally* differentially private randomizer for some $\varepsilon > 0$. Similar results are known [DJW18, BS15]; however, the work of [BS15] only applies to the case that R is $(\varepsilon_L, \delta_L)$-locally differentially private with $\varepsilon_L = O(1)$, and [DJW18] only consider $(\varepsilon_L, 0)$-locally differentially private protocols. Moreover, their dependence on the privacy parameter ε_L is not tight: in particular, for the "small-sample regime" of $n \leqslant O(\log^2 B)$ that we consider in this section, the bounds of [DJW18] decay as $e^{-2\varepsilon_L}$, whereas we will be able to derive bounds scaling as $1/\varepsilon_L$. We will then apply this bound with $\varepsilon_L = \varepsilon + \ln n$ being the privacy parameter of the locally differentially private protocol furnished by Lemma 3.5.

The proof of the error lower bound relies on the following Lemma 3.8, which bounds the mutual information between a uniformly random index $V \in [B]$, and $R(V)$. It improves upon analogous results in [DJW18, BS15], for which the dependence on ε_L is $(e^{\varepsilon_L} - 1)^2$, when ε_L is large.

Lemma 3.8 (Mutual information upper bound for small-sample regime)**.** *Fix $n \in \mathbb{N}$. Let R be an $(\varepsilon_L, \delta_L)$-differentially private local randomizer (in the sense of Definition 2.1). Let $V \sim [B]$ be chosen uniformly at random. Then*

$$I(V; R(V)) \leqslant 2\delta_L \cdot \log B + 1 + \varepsilon_L \log e.$$

Proof. For $v \in \mathcal{X}$, following Definition 3.2, let P_v denote the distribution of $R(v)$, i.e., the distribution of the output of the local randomizer when it gets an input of v. Let $\bar{P} := \frac{1}{B}\sum_{v\in[B]} P_v$. Notice that the distribution of $R(V)$, where $V \sim [B]$ is uniform, is P_v. It follows that

$$I(V; R(V)) = \frac{1}{B}\sum_{v\in[B]}\sum_{z\in\mathcal{Z}} \mathbb{P}[R(v)=z]\cdot\log\left(\frac{\mathbb{P}_R[R(v)=z]}{\mathbb{P}_{V\sim[B],R}[R(V)=z]}\right) = \frac{1}{B}\sum_{v\in[B]} \mathrm{KL}(P_v||\bar{P}). \tag{17}$$

We now upper bound $\mathrm{KL}(P_v||\bar{P})$ for each $v \in [B]$. We first claim that for any $v_0 \in [B]$,

$$p_{Z\sim R(v_0)}\left[\log\left(\frac{p_{v_0,Z}}{\frac{1}{B}\sum_{j\in[B]} p_{j,Z}}\right) > 1+\varepsilon_L\log e\right] \leqslant 2\delta_L. \tag{18}$$

To see that (18) holds, let $\mathcal{S} := \left\{z\in\mathcal{Z} : \frac{p_{v_0,z}}{\frac{1}{B}\sum_{j\in[B]} p_{j,Z}} > 2e^{\varepsilon_L}\right\}$. If (18) does not hold, then $p_{v_0,\mathcal{S}} > 2\delta_L$ and $p_{v_0,\mathcal{S}} > (2e^{\varepsilon_L})\cdot\frac{1}{B}\sum_{j\in[B]} p_{j,\mathcal{S}}$. On the other hand, we have from (ε_L,δ_L)-differential privacy of R that

$$p_{v_0,\mathcal{S}} \leqslant \left(\frac{1}{B}\sum_{j\in[B]} p_{j,\mathcal{S}}\right)\cdot e^{\varepsilon_L} + \delta_L.$$

The above equation is a contradiction in light of the fact that for positive real numbers a, b, $a + b \leqslant \max\{2a, 2b\}$. Notice that for any $z \in \mathcal{Z}, v \in [B]$, it is the case that $\log\left(\frac{p_{v,z}}{\frac{1}{B}\sum_{j\in[B]} p_{j,z}}\right) \leqslant \log B$. It follows that (18) implies that

$$\mathrm{KL}(P_{v_0}||\bar{P}) \leqslant 2\delta_L\cdot\log B + 1 + \varepsilon_L\log e.$$

The statement of Lemma 3.8 follows from the above equation and (17). □

Lemma 3.9, together with Lemma 3.5, establishes Corollary 3.2: in particular, by Lemma 3.5, any single-message shuffled-model (ε,δ)-differentially private protocol P yields a local-model $(\varepsilon+\ln n,\delta)$-differentially private protocol with the same accuracy. Thus we may set $\varepsilon_L = \ln n + \varepsilon$ in Lemma 3.9, so that $n \geqslant \Omega(\log(B)/\varepsilon_L) = \Omega(\log(B)/\log n)$ becomes $n \geqslant \Omega(\log(B)/\log\log(B)))$. Lemma 3.9 is also used in the proof of (6) of Theorem 3.1. The proof is by a standard application of Fano's inequality [DJW18, BS15].

Lemma 3.9 (Sample-complexity lower bound for constant-error frequency estimation)**.** *Suppose $\delta_L < 1/(4n), 0 < \varepsilon_L < \log(B)/20$, and $P = (R, A)$ is a local-model protocol that satisfies (ε_L,δ_L)-local differential privacy and $(1/3, 1/2)$-accuracy. Then $n > \frac{\log B}{20\varepsilon_L}$.*

Proof. Suppose for the purpose of contradiction that $n \leqslant \frac{\log B}{20\varepsilon_L}$.

Let D be the distribution on $(\{0,1\}^B)^n$ that is uniform over all tuples $(e_v, e_v, \ldots, e_v)$, for $v \in [B]$ (recall that e_v is the unit vector for component v, i.e., the vector with a 1 in the vth component).

Consider any sample $X = (x_1, \ldots, x_n)$ in the support of D, so that $x_1 = \cdots = x_n = e_v$ for some $v \in [B]$. If the error $\max_{j\in[B]}\left|P(X)_j - \frac{1}{n}\sum_{i=1}^n (e_{x_i})_j\right|$ is strictly less than $1/2$, then the function

$$f(\hat{x}_1, \ldots, \hat{x}_j) := \arg\max_{j\in[B]} \hat{x}_j$$

will satisfy $f(\hat{x}_1, \ldots, \hat{x}_j) = v$. It follows that

$$\begin{aligned}
&\mathbb{P}_{X\sim D,R}\left[\max_{j\in[B]}\left|\hat{x}_j - \frac{1}{n}\sum_{i=1}^{n}(e_{x_i})_j\right| < 1/2\right] \\
&\leqslant \mathbb{P}_{X\sim D,R}\left[f(A(R(x_1),\ldots,R(x_n))) = V\right] \\
&\leqslant \frac{I(f(A(R(x_1),\ldots,R(x_n)));V)+1}{\log B} && (19)\\
&\leqslant \frac{I((R(x_1),\ldots,R(x_n));V)+1}{\log B} \\
&\leqslant \frac{n\cdot I(R(V);V)+1}{\log B} && (20)\\
&\leqslant \frac{n\cdot(2\delta_L\log B + 1 + \varepsilon_L\log e)+1}{\log B} && (21)\\
&\leqslant 2\delta_L n + \frac{5n\varepsilon_L}{\log B} && (22)\\
&< 1/2 && (23)
\end{aligned}$$

where (19) follows by Fano's inequality and the random variable V is so that $x_1 = \cdots = x_n = e_V$ and V is uniform over $[B]$). Moreover, (21) follows from Lemma 3.8 (and the fact that V is uniform over $[B]$), (20) follows from the chain rule for mutual information, and (22), (23) follow from our assumptions on $n, B, \delta_L, \varepsilon_L$. We now arrive at the desired contradiction to the $(1/3, 1/2)$-accuracy of P. □

The next lemma is an adaptation to the local model of a standard result [SU16, Fact 2.3], stating that the optimal error of a differentially private frequency estimation protocol decays at most inverse linearly in the number of users n.

Lemma 3.10 (Inverse-linear dependence of error on n)**.** *Suppose $P = (R, A)$ is an $(\varepsilon_L, \delta_L)$-locally differentially private algorithm (Definition 2.3) for n-user frequency estimation on $[B]$ that satisfies (α, β)-accuracy.*

Let $n \geqslant n' \geqslant \lfloor 2\alpha n/c \rfloor$ for any $c \leqslant 2\alpha n$. Then there is an $(\varepsilon_L, \delta_L)$-differentially private protocol $P' = (R', A')$ for n'-user frequency estimation on $[B]$ that satisfies (c, β)-accuracy.

Proof. The algorithm P' is given as follows: we have $R' = R$. The analyzer A', on input $(z_1, \ldots, z_{n'}) \in \mathcal{Z}^{n'}$, generates $n - n'$ i.i.d. copies of $R(e_1)$, which we denote by $z_{n'+1}, \ldots, z_n$. (Recall $e_1 = (1, 0, \ldots, 0)$.) Then A' computes the vector $v := A(z_1, \ldots, z_n) \in [0,1]^B$, and outputs v', where $v'_j = \frac{n}{n'} \cdot v_j$ for $j > 1$, and $v'_1 = \frac{n}{n'} \cdot \left(v_1 - \frac{n-n'}{n}\right)$.

To see that P' satisfies (c, β) accuracy, let's fix any input dataset $X = (x_1, \ldots, x_{n'})$. Let $x_j = 1$ for $j > n'$, and set $X' = (x_1, \ldots, x_{n'}, x_{n'+1}, \ldots, x_n)$. The (α, β)-accuracy of P gives that with probability at least $1 - \beta$, $\max_{j\in[B]} \left|P(X')_j - \frac{1}{n}\sum_{i=1}^{n}(x_i)_j\right| \leqslant \alpha$. In such an event, we have that

$$\max_{j\in[B]}\left|\frac{n'v'_j}{n} - \frac{1}{n}\sum_{i=1}^{n'}(x_i)_j\right| \leqslant \alpha.$$

Multiplying the above by n/n' and noting that $n/n' \leqslant c/\alpha$ gives that P' satisfies (c, β)-accuracy. □

A technique similar to the one used in Lemma 3.10 can be used to show that the dependence of the error on ε must be $\Omega(1/\varepsilon)$ in the central model [SU16, Fact 2.3]. However, doing so requires each user's input to be duplicated a total of $\Theta(1/\varepsilon)$ times, and it is not clear how to implement such a transformation in the local model. (In the *multi-message shuffled model*, though, such a transformation can be done and one would recover the $\Omega(1/\varepsilon)$ lower bound.)

Finally we may establish (6); for ease of the reader we state it as a separate lemma:

Lemma 3.11 (Proof of Theorems 3.1 and 3.3 in small-sample regime; i.e., (6) & (9))**.** *There is a sufficiently small positive constant c so that the following holds. Suppose $n, B \in \mathbb{N}$ and $\varepsilon_L, \delta_L \geqslant 0$ with $n \geqslant \log B/(\varepsilon_L c)$, $0 < \delta_L < c/n$, and $0 \leqslant \varepsilon_L \leqslant \log B$. Then there is no protocol for n-user frequency estimation on $[B]$ that satisfies $(\varepsilon_L, \delta_L)$-local differential privacy and $\left(\frac{c \log B}{n\varepsilon_L}, 1/2\right)$-accuracy.*

For $n \geqslant \frac{\log B}{c \log\log B}$, $\delta < c/n, \varepsilon \leqslant \log n$, there is no protocol for n-user frequency estimation on $[B]$ that satisfies (ε, δ)-differential privacy in the single-message shuffled model and $\left(\frac{c \log B}{n \log\log B}, 1/2\right)$-accuracy.

Proof. Suppose that the statement of the lemma did not hold for some protocol P. By Lemma 3.10 with $\alpha = \frac{c \log B}{n\varepsilon_L}$ and $n' = \left\lfloor \frac{8c \log B}{\varepsilon_L} \right\rfloor$ there is an $(\varepsilon_L, \delta_L)$-locally differentially private protocol $P' = (R', A')$ for n'-user frequency estimation that is $(1/4, 1/2)$-accurate. As long as $c < 1/160$ we have a contradiction by Lemma 3.9.

The second statement of the lemma follows by applying Lemma 3.5 and taking $\varepsilon_L = \varepsilon + \ln n$. □

3.3 Intermediate-Sample and Large-Sample Regimes

In this section we prove Theorem 3.1 in the intermediate and large-sample regimes (i.e., (7) and (8)), which is the most technical part of the proof of Theorem 3.1. As we did in the small-sample regime, we in fact prove a more general statement giving a lower bound on the accuracy of all *locally differentially private* protocols in the low and approximate-privacy setting:

Lemma 3.12 (Proof of Theorem 3.3 in the intermediate-sample regime; i.e., (10) & (11))**.** *There is a sufficiently small positive constant c such that the following holds. Suppose $\ln B > 1/c^{56}$,*

$$\frac{1}{c} + \max\left\{\frac{\ln n}{3}, \frac{\varepsilon_L + \ln(1+\varepsilon_L)}{2}\right\} \leqslant \ln B \leqslant \min\left\{n^{2/3}, \frac{n}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}\right\}, \tag{24}$$

and

$$\delta_L \leqslant \min\left\{e^{-\varepsilon_L}, \frac{1}{2\sqrt[4]{n^2(\ln^2 B)\exp(\varepsilon_L)(1+\varepsilon_L)}}, \frac{1}{2(\ln^{1/3} B)n^{2/3}}\right\}. \tag{25}$$

Then there is no protocol for n-user frequency estimation that is $(\varepsilon_L, \delta_L)$-locally differentially private and $(\alpha, 1/4)$-accurate for

$$\alpha = c \cdot \min\left\{\frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}}, \frac{\ln^{1/7} B}{n^{2/3}}\right\}.$$

Remark 3.2. The term $\ln^{1/7} B$ in the definition of α above can be replaced by $\ln^{\zeta} B$ for any constant $\zeta < 1/6$. Moreover, the requirement that $\ln B$ is greater than the (very large) constant $1/c^{56}$ can easily be reduced to $1/c^6$ (with no change in c) by replacing this term $\ln^{1/7} B$ with 1.

By Lemma 3.5 the following is a corollary of Lemma 3.12, establishing Theorem 3.1 in the intermediate-sample regime:

Corollary 3.13 (Proof of Theorem 3.1 in intermediate-sample regime; i.e., (7))**.** *For a sufficiently small positive constant $c < 1$, if $\log B \geqslant 1/c$,*

$$\frac{1}{c} \cdot \log^2 B \cdot \log\log B \leqslant n \leqslant \frac{cB^2}{\log B}, \tag{26}$$

$\delta \leqslant c/n$ and $\varepsilon \leqslant 1$, there is no protocol for n-user frequency estimation in the single-message shuffled model that is (ε, δ)-differentially private and $\left(\frac{c}{n^{3/4}\sqrt[4]{\log n}}, 1/4\right)$-accurate.

Remark 3.3. Notice that the bounds on n in the inequality (7) do not involve c, unlike those in (26). To ensure that (7) holds for $\frac{\log^2 B}{c \log\log B} \geqslant n \geqslant (\log^2 B)(\log\log B)$, note that (6) holds for *all* $n \geqslant \frac{\log B}{c \log\log B}$ (Lemma 3.11) and increase the constant in the $\Omega(\cdot)$ in (7) by a factor of at most $1/c$. To ensure that (7) holds for $\frac{cB^2}{\log B} \leqslant n \leqslant \frac{B^2}{\log B}$, we use the reduction to locally differentially private protocols given by Lemma 3.5, and then note that a locally differentially private protocol which is (α, β)-accurate for n' users implies a locally differentially private protocol with the same privacy parameters and which is $(\alpha n'/n, \beta)$-accurate for $n \leqslant n'$ users by simulating the presence of $n' - n$ fake users who hold a fixed and known item. (This latter reduction also requires increasing the constant in the $\Omega(\cdot)$ in (7) by a factor of at most $1/c$.)

Finally we note that similar reductions hold for the proof of Theorem 3.3 using Lemma 3.12 as well.

Proof of Corollary 3.13. By Lemma 3.5, it suffices to show that there is no protocol for n-user frequency estimation in the local model that is $(1 + \ln n, \delta)$-differentially private and $\left(\frac{c}{n^{3/4}\sqrt[4]{\log n}}, 1/4\right)$-accurate. We now apply Lemma 3.12 with $\varepsilon_L = 1 + \ln n$ and $\delta_L = \delta$. The left-hand side of (24) holds (though perhaps with a different constant c than the one used here) since $n \leqslant \frac{cB^2}{\log B}$ and $c < 1$, and the right-hand side of (24) holds since $cn \geqslant \log^2 B \cdot \log\log B$ (as long as c is sufficiently small). Moreover, (25) holds as long as

$$\delta \leqslant \frac{1}{2 \cdot (en)^{3/4} \cdot (1 + \ln(en))^{1/4} \cdot (\log^{1/2} B)},$$

which is guaranteed by $\delta \leqslant c/n$ and $cn \geqslant \log^2 B \log\log B$ for sufficiently small c. As $\frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}} < n^{-2/3}$ for our choice of ε_L, Lemma 3.12 now yields the desired result. □

The hard distribution used to prove Theorem 3.1 in the small-sample regime set each user's data $X_i \in [B]$ to be equal to some fixed $V \in [B]$ (Lemma 3.9). At a high level, to prove Lemma 3.12, we must adapt this argument to allow us to gain a more fine-grained control over the accuracy of protocols. We do so using the same distribution as in previous works [DJW18, BS15]: in particular, each user's X_i is now only equal to V with some small probability (which is roughly the target accuracy α) and otherwise is uniformly random. Formally, we make the following definition: For each $v \in [B]$ and $\gamma \in (0, 1)$, define a distribution of $X \in [B]$, denoted by $X \sim D_{v,\gamma}$, as

$$X = \begin{cases} v & \text{w.p. } \gamma \\ \mathrm{Unif}([B]) & \text{w.p. } 1 - \gamma, \end{cases}$$

where $\mathrm{Unif}([B])$ denotes the uniform distribution on $[B]$. Let $\bar{D}_\gamma$ denote the joint distribution of (V, X), where $V \sim \mathrm{Unif}([B])$ and $X \sim D_{V,\gamma}$. (Note that the marginal distribution of X under $\bar{D}_\gamma$ is the mixture distribution $D_{V,\gamma} = \frac{1}{B}\sum_{v\in[B]} D_{v,\gamma}$, which is just the uniform distribution on $[B]$.) Analogously to Lemma 3.8, we wish to derive an upper bound on $I(V; R(X))$ when $(V, X) \sim \bar{D}_\gamma$. It is known that if R is $(\varepsilon_L, \delta_L)$-differentially private and $\varepsilon_L = O(1)$, then $I(V; R(X)) \leqslant O(\gamma^2\varepsilon_L^2 + \tilde{O}(\delta_L/(\varepsilon_L\gamma)))$ [BS15], and that for *any* $\varepsilon_L \geqslant 0$, if R is $(\varepsilon_L, 0)$-differentially private, then $I(V; R(X)) \leqslant O(\gamma^2(e^{\varepsilon_L} - 1)^2)$ [DJW18].

Remark 3.4. Suppose we attempt to prove Lemma 3.12 following this strategy, at least when $\frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}} < n^{-2/3}$, which is the regime we encounter for single-message shuffled-model protocols. To do so, it is natural to try to improve the upper bound of Duchi et al. [DJW18] of $I(V; R(X)) \leqslant O(\gamma^2 \exp(2\varepsilon_L))$ to $I(V; R(X)) \leqslant \tilde{O}(\gamma^2 \exp(\varepsilon_L/2))$, which turns out to be sufficient to establish Lemma 3.12. However, this is actually *false*, as can be seen by the local randomizer $R_{\mathrm{RR}} : [B] \to [B]$ of B-randomized response [War65] (see also Appendix A). In particular, suppose we take $\varepsilon_L = (\ln n) + O(1)$, $n > 10B$, and $\gamma \ll \exp(-\varepsilon_L/2) = \Theta(\sqrt{1/n})$; it is in fact necessary to treat these settings of the parameters to prove (7). For these parameters it is easy to check that $I(V; R_{\mathrm{RR}}(X)) = \Theta(\gamma \log B) \gg \gamma^2 \cdot \exp(\varepsilon_L/2)$. Thus it may seem that one cannot derive tight bounds by upper bounding $I(V; R(X))$ when $(V, X) \sim \bar{D}_\gamma$.

It is, however, possible to salvage the technique outlined in Remark 3.4: the crucial observation is that the best-possible additive error of any single-message shuffled-model protocol where each user uses R_{RR} is $\tilde{\Theta}(\sqrt{B}/n)$.

When $\varepsilon = O(1) + \ln n$ (as we will have when applying Lemma 3.5), it is the case that $\sqrt{B}/n > \frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}}$ when $n < \tilde{O}(B^2)$. Therefore, there is still hope to prove a lower bound of $\tilde{\Omega}\left(\frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}}\right)$ on the additive error when $\varepsilon = O(1) + \ln n$ and $n \leqslant \tilde{O}(B^2)$ if we *additionally assume* that the additive error of any local-model protocol using R is bounded above by $\frac{1}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}}$. This is indeed what we manage to do in Lemma 3.14 below:

Lemma 3.14 (Mutual information upper bound for intermediate-sample regime)**.** *There is a sufficiently large positive constant C such that the following holds. Suppose $n, \alpha, \beta, \gamma, \delta, \varepsilon \geqslant 0$, $(V, X) \sim \bar{D}_\gamma$ and $R : [B] \to \mathcal{Z}$ is an $(\alpha, 1/4)$-accurate local randomizer with $C \max\{1/n, 1/\sqrt{nB}\} \leqslant \alpha \leqslant \gamma$, and $C\alpha^2 n \leqslant 1$ which is $(\varepsilon_L, \delta_L)$-differentially private for n-user frequency estimation in the local model with $\delta_L \leqslant \min\left\{\frac{\gamma}{\log B}, e^{-\varepsilon_L}\right\}$. Then*

$$I(V; R(X)) \leqslant C \cdot \left(\gamma^2 \alpha^2 n e^{\varepsilon_L} \cdot (1 + \varepsilon_L) + \gamma \alpha^2 n + \gamma^2\right). \tag{27}$$

Typically the term $\gamma^2\alpha^2 n e^{\varepsilon_L}(1+\varepsilon_L)$ is the dominating one on the right-hand side of (27). In particular, in the application of Lemma 3.14 to establish (7), we will have $\gamma = \tilde{\Theta}(n^{-3/4})$, $\alpha = \tilde{\Theta}(n^{-3/4})$ and $\varepsilon_L = \ln(n) + O(1)$, so that $\gamma^2\alpha^2 n e^{\varepsilon_L}(1+\varepsilon_L) = \tilde{\Theta}(1/n)$, whereas $\gamma\alpha^2 n = \tilde{\Theta}(n^{-5/4})$ and $\gamma^2 = \tilde{\Theta}(n^{-3/2})$.

Remark 3.5. The statement of Lemma 3.14 still holds if R is only assumed to be (α, β)-accurate for any constant $\beta < 1/2$.

We postpone the proof of Lemma 3.14 for now and assuming it, prove Lemma 3.12.

Proof of Lemma 3.12. Let $a = 200$ and $c < 1$ be a sufficiently small positive constant, to be specified later. Let $\alpha = \min\left\{\frac{c}{\sqrt[4]{n^2 \exp(\varepsilon_L)(1+\varepsilon_L)}}, \frac{c \ln^{1/7} B}{n^{2/3}}\right\}$ be the desired error lower bound. Set $\gamma := \min\left\{\alpha \cdot a\sqrt{\ln B}/c, 1/3\right\}$. We make the following observations about γ:

1. $\gamma < 1/2$ is clear from definition of γ.
2. $\gamma^2 \cdot nB \geqslant a^2 \ln B$. This is clear if $\gamma \geqslant 1/3$ by choosing c small enough (recall $\ln B > 1/c^3$). Otherwise, note that $\gamma^2 \cdot nB \geqslant \min\left\{\frac{a^2 B \ln B}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}, \frac{a^2 B \ln B}{n^{1/3}}\right\} \geqslant a^2 \ln B$ since $\max\left\{e^{\varepsilon_L}(1+\varepsilon_L), n^{2/3}\right\} \leqslant B^2$.
3. $\gamma n \geqslant a \ln B$. Again this is clear if $\gamma \geqslant 1/3$. Otherwise, $\gamma n \geqslant \min\left\{\frac{a\sqrt{n \ln B}}{\sqrt[4]{\exp(\varepsilon_L)(1+\varepsilon_L)}}, a\sqrt{\ln B} n^{1/3}\right\} \geqslant a \ln B$ since $\min\left\{n^{2/3}, \frac{n}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}\right\} \geqslant \ln B$.

Suppose that $P = (R, A)$ is a single-message shuffled model protocol which is $(\alpha, 1/4)$-accurate and $(\varepsilon_L, \delta_L)$-differentially private where ε_L, δ_L satisfy (24) and (25). Now suppose $V \sim [B]$ uniformly and $X_1, \ldots, X_n \sim D_{V,\gamma}$ are independent (conditioned on V).

Fix an arbitrary $v \in [B]$, and let us momentarily condition on the event that $V = v$. Consider the conditional distribution of $X_1, \ldots, X_n \sim D_{v,\gamma}$. For any $u \neq v$, we have, by the Chernoff bound, in the case that $\gamma/3 \leqslant 1/B$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(e_{X_i})_u \geqslant 1/B + \gamma/3\right] \leqslant \exp\left(-\frac{(\gamma B)^2 \cdot n/B}{27}\right) = \exp(-\gamma^2 Bn/27) \leqslant \exp(-a^2 \ln B/27). \tag{28}$$

In the case that $\gamma/3 > 1/B$, again by the Chernoff bound, we have

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(e_{X_i})_u \geqslant 1/B + \gamma/3\right] \leqslant \exp\left(-\frac{(\gamma B) \cdot n/B}{9}\right) = \exp(-\gamma n/9) = \exp(-a\sqrt{\ln B} n^{1/4}/9) \leqslant \exp(-a \ln B/9). \tag{29}$$

Next, note that by definition of the distribution $D_{V,\gamma}$, we have that for each $1 \leqslant i \leqslant n$, $\mathbb{E}[(e_{X_i})_v] = \frac{1-\gamma}{B} + \gamma$. It then follows by the Chernoff bound that

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(e_{X_i})_v \leqslant (1-\gamma)/B + 2\gamma/3\right] &\leqslant \exp(-a^2(1-\gamma)\ln B/72) + \exp(-\gamma n/72)\\ &\leqslant \exp(-a^2\ln B/144) + \exp(-a\ln B/72). \end{aligned} \tag{30}$$

Since $1/B + \gamma/3 < (1-\gamma)/B + 2\gamma/3$ and $P = (R, A)$ is $(\gamma/3, 1/3)$-accurate (as $\gamma/3 \geqslant \alpha$), it follows by a union bound over all $u \in [B]$ in (28), (29) and (30) that with probability at least

$$1 - 1/3 - \exp(-a/9) - 2\exp(-a\ln B/72), \tag{31}$$

we have that

$$\arg\max_{u\in[B]} P((X_1,\ldots,X_n)) = \arg\max_{u\in[B]} A(R(X_1),\ldots,R(X_n))_u = v.$$

Moreover, by our choice of $a = 200$, we ensure that the probability in (31) is strictly greater than $1/4$. For such a, using the fact that $v \in [B]$ is arbitrary, we have shown that

$$\mathbb{P}\left[\arg\max_{u\in[B]}\{A(R(X_1),\ldots,R(X_n))_u\} = V\right] > 1/4. \tag{32}$$

Now we will apply Lemma 3.14 to derive an upper bound on the probability in the above equation. First we check that the conditions of Lemma 3.14 are met. By (24) and $\ln B \geqslant 1/c^3$ we have that

$$(\alpha n)^2 \geqslant \min\left\{c^2 n^{2/3}, \frac{c^2 n}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}\right\} \geqslant c^2\ln B \geqslant 1/c, \tag{33}$$

so by choosing c small enough, we can guarantee that $\alpha \geqslant C/n$, where C is the constant of Lemma 3.14. Similarly, by (24), we have that

$$\alpha^2 nB \geqslant \min\left\{c^2 B/n^{1/3}, \frac{c^2 B}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}\right\} \geqslant c^2\exp(1/c), \tag{34}$$

and again by choosing c small enough, we can guarantee that $\alpha \geqslant C/\sqrt{nB}$, where C is the constant of Lemma 3.14.

The choice of γ ensures that $\gamma \geqslant \alpha$, and $C\alpha^2 n = \min\left\{\frac{Cc^2}{\sqrt{\exp(\varepsilon_L)(1+\varepsilon_L)}}, \frac{Cc^2\ln^{2/7}B}{n^{1/3}}\right\} \leqslant Cc^2$, which can be made less than 1 by choosing c sufficiently small. Finally, $\delta_L \leqslant \min\left\{e^{-\varepsilon_L}, \frac{\gamma}{2\ln B}\right\} \leqslant \min\left\{e^{-\varepsilon_L}, \frac{\gamma}{\log B}\right\}$ by (25). Therefore, by Fano's inequality and Lemma 3.14, for any function $f : [0,1]^B \to [B]$,

$$\begin{aligned}\mathbb{P}[f(A(R(X_1),\ldots,R(X_n))) = V] &\leqslant \frac{I(V;(R(X_1),\ldots,R(X_n))) + 1}{\ln B}\\ &\leqslant \frac{\sum_{i=1}^{n} I(V;R(X_i)) + 1}{\ln B}\\ &\leqslant \frac{1 + n\cdot\left(C\cdot\left(\gamma^2\alpha^2 n e^{\varepsilon_L}\cdot(1+\varepsilon_L) + \gamma\alpha^2 n + \gamma^2\right)\right)}{\ln B}\\ &\leqslant \frac{1 + nC\cdot\left(\gamma^2\alpha^2 n e^{\varepsilon_L}\cdot(1+\varepsilon_L) + \frac{2a}{c^2}\cdot\gamma\alpha^2 n\right)}{\ln B} \qquad (35)\\ &\leqslant \frac{1 + Cc\ln B}{\ln B}. \qquad (36)\end{aligned}$$

Inequality (35) follows since

$$\frac{\gamma^2}{\gamma\alpha^2 n} \leqslant \frac{\alpha a\sqrt{\ln B}/c}{\alpha^2 n} \leqslant \frac{a}{c}\cdot\frac{\sqrt{\ln B}}{\alpha n} \leqslant \frac{a}{c^2},$$

where we have used inequality (33). Inequality (36) follows since, by choice of α,

$$\max\left\{\gamma^2\alpha^2 n\exp(\varepsilon_L)(1+\varepsilon_L), \frac{a}{c^2}\cdot\gamma\alpha^2 n\right\} \leqslant \max\left\{\frac{c^2a^2(\ln B)n\exp(\varepsilon_L)(1+\varepsilon_L)}{n^2\exp(\varepsilon_L)(1+\varepsilon_L)}, \frac{a^2(\ln^{13/14}B)\cdot n}{cn^2}\right\}$$
$$\leqslant \frac{c\ln B}{n}, \tag{37}$$

where (37) follows from $\ln B > 1/c^{56} > (a^2/c^2)^{14}$ and a choice of $c < 1/a^2$.

As long as $c < 1/(10C)$, the expression in (36) is bounded above by 1/4, which contradicts (32). □

We now prove Lemma 3.14. The proof uses the assumption that the local randomizer R is $(\alpha, 1/4)$-accurate (Definition 3.1) to derive, for each $v \in [B]$, a lower bound on the total variation distance between the distributions of $R(v)$ and $R(V)$, where $V \sim \mathrm{Unif}([B])$. Intuitively, it makes sense that if, for some $v \in [B]$, the distribution of $R(v)$ is close to $R(V)$, then no analyzer can reliably compute how many users hold the item v. However, showing rigorously that this holds for *any* analyzer A is nontrivial, and we state this result as a separate lemma:

Lemma 3.15 (Lower bound on total variation distance between $R(v)$ & $R(V)$)**.** *Suppose $R : [B] \to \mathcal{Z}$ is an $(\alpha/6, \beta)$-accurate local randomizer such that*

$$\max\left\{\frac{3B\log(4/(1-2\beta))}{n}, \sqrt{\frac{3B\log(4/(1-2\beta))}{n}}\right\} \leqslant \frac{\alpha B}{4}. \tag{38}$$

Let the distribution of $R(v)$ be denoted P_v (Definition 3.2) and the distribution of $R(V)$, where $V \sim \mathrm{Unif}([B])$ be denoted Q. Then there is some $C = \Theta\left(\frac{1}{(1-2\beta)^2}\right)$, such that for each $v \in [B]$, $\Delta(P_v, Q) \geqslant 1 - C\alpha^2 n$.

A result similar to Lemma 3.15 was established in [CSS12]; however, their result only establishes a lower bound on $\Delta(P_v, P_u)$ for $u \neq v$, which does not lead to tight bounds on $\Delta(P_v, Q)$, as we need. The proof of Lemma 3.15 is provided in Section 3.4.

Proof of Lemma 3.14. Fix a local randomizer $R : [B] \to \mathcal{Z}$ satisfying the requirements of the lemma statement. Recall (per Definition 3.2) that for $v \in [B]$, we use P_v to denote the distribution of $R(v)$, and that $p_{v,\cdot}$ denotes the density of P_v, so that for each $z \in \mathcal{Z}$, $p_{v,z} = \mathbb{P}_R[R(v) = z]$. For $v \in [B]$, additionally let $P_{v,\gamma}$ denote the distribution of $R(X)$ when $X \sim D_{v,\gamma}$, and let Q denote the distribution of $R(X)$ when $(X, V) \sim \bar{D}_\gamma$. Note that Q is the distribution of $R(X)$ when $X \sim \mathrm{Unif}([B])$, so indeed does not depend on γ. First note (see (17)) that

$$I(V; R(X)) = \frac{1}{B}\sum_{v\in[B]} \mathrm{KL}(P_{v,\gamma}||Q).$$

For each $\mathcal{S} \subset \mathcal{Z}$ and $z \in \mathcal{Z}$, write $q_{\mathcal{S}} := \mathbb{P}[Z \in \mathcal{S}]$ and $q_z := \mathbb{P}[Z = z]$, where $Z \sim Q$. Notice that for each $z \in \mathcal{Z}$, we have $p_{v,z} \leqslant B \cdot q_z$ since $q_{v,z} = \frac{1}{B}\sum_{v\in[B]} p_{v,z}$.

Next, for each $h \geqslant 0$ and $v \in [B]$, set

$$\mathcal{U}_{v,h} := \left\{z \in \mathcal{Z} : \frac{p_{v,z}}{q_z} = h\right\}.$$

Let ρ_v be the (Borel) probability measure on $\mathbb{R}$ given by, for $\mathcal{S} \subset \mathbb{R}$,

$$\rho_v(\mathcal{S}) := \sum_{h\geqslant 0} p_{v,\mathcal{U}_{v,h}}\delta_h(\mathcal{S}),$$

where the sum is well-defined since only finitely many h are such that $\mathcal{U}_{v,h}$ is nonempty. (Here δ_h is the measure such that $\delta_h(\mathcal{S}) = 1$ if $h \in \mathcal{S}$, and otherwise $\delta_h(\mathcal{S}) = 0$.) Since $\sum_{z\in\mathcal{Z}} p_{v,z} = \sum_{h\geqslant 0} p_{v,\mathcal{U}_{v,h}} = 1$, ρ_v is indeed a probability measure.

Next notice that

$$\begin{aligned}
\mathrm{KL}(P_{v,\gamma}||Q) &= \sum_{z\in\mathcal{Z}} (\gamma p_{v,z} + (1-\gamma)q_z)\cdot \log\left(1+\gamma\left(\frac{p_{v,z}}{q_z}-1\right)\right) \\
&\leqslant \sum_{z:p_{v,z}\geqslant q_z} p_{v,z}\left(\frac{1}{p_{v,z}/q_z}+\gamma\left(1-\frac{1}{p_{v,z}/q_z}\right)\right)\cdot\log\left(1+\gamma\cdot\frac{p_{v,z}}{q_z}\right) \\
&+ \sum_{z:p_{v,z}<q_z} (\gamma p_{v,z} + (1-\gamma)q_z)\cdot \log\left(1+\gamma\left(\frac{p_{v,z}}{q_z}-1\right)\right) \\
&= \int_{h=1}^{B}\left(\frac{1}{h}+\gamma\left(1-\frac{1}{h}\right)\right)\log(1+\gamma h)d\rho_v(h) \\
&+ \sum_{z\in\cup_{h<1}\mathcal{U}_{v,h}} (\gamma p_{v,z} + (1-\gamma)q_z)\cdot\log\left(1+\gamma\left(\frac{p_{v,z}}{q_z}-1\right)\right). \qquad (39)
\end{aligned}$$

We begin by working towards an upper bound on the first term in the above expression (39), corresponding to values $h \geqslant 1$. Our first goal is to show that for $h \gg 1$, $p_{v,\mathcal{U}_{v,h}}$ is small for most v. To do so, define, for any $\lambda \geqslant 1$,

$$\mathcal{T}_{v,\lambda} := \left\{z\in\mathcal{Z} : \frac{p_{v,z}}{q_z}\geqslant\lambda\right\} = \bigcup_{h\geqslant\lambda}\mathcal{U}_{v,h}.$$

We next make the following claim:

Claim 3.16. *For any $v\in[B]$, for each $z\in\mathcal{T}_{v,\lambda}$, there are at most B/λ values of $v'\in[B]$ such that $z\in\mathcal{T}_{v',\lambda}$.*

Claim 3.16 is a simple consequence of Markov's inequality on the distribution of the random variable $p_{v,z}$, where $v\sim[B]$ uniformly.

Next, consider any $z\in\mathcal{Z}$ such that there is some $u\in[B]$ with $z\in\mathcal{T}_{u,\lambda}$; let the set of such z be denoted by $\mathcal{W}_\lambda$, i.e., $\mathcal{W}_\lambda = \bigcup_u \mathcal{T}_{u,\lambda}$. By Claim 3.16, for each $z\in\mathcal{W}_\lambda$, there are at most B/λ values $u\in[B]$ such that $z\in\mathcal{T}_{u,\lambda}$. Let the set of such values be denoted by $\mathcal{S}_z\subset[B]$, and construct an ordering of those $u\in\mathcal{S}_z$ so that $p_{u,z}$ are in decreasing order with respect to this ordering. Now, for a fixed λ and for each $u\in[B]$, and $1\leqslant k\leqslant B/\lambda$, construct a subset $\tilde{\mathcal{T}}^{(k)}_{u,\lambda}\subset\mathcal{T}_{u,\lambda}$ such that each $z\in\mathcal{W}_\lambda$ appears in at most one set $\tilde{\mathcal{T}}^{(k)}_{u,\lambda}$ (over all $u\in[B]$), and such a u is the kth element of $\mathcal{S}_z$ with respect to the ordering above (if it exists). It is an immediate consequence of this construction that for each fixed k, the sets $\tilde{\mathcal{T}}^{(k)}_{u,\lambda}$, $u\in[B]$, are pairwise disjoint. Moreover, for each fixed u, the sets $\tilde{\mathcal{T}}^{(k)}_{u,\lambda}$, $1\leqslant k\leqslant B/\lambda$ are pairwise disjoint, and their union is $\mathcal{T}_{u,\lambda}$. It follows that

$$\sum_{k=1}^{B/\lambda}\sum_{u\in[B]} p_{u,\tilde{\mathcal{T}}^{(k)}_{u,\lambda}} = \sum_{z\in\mathcal{Z}}\sum_{k=1}^{B/\lambda}\sum_{u:z\in\tilde{\mathcal{T}}^{(k)}_{u,\lambda}} p_{u,z} = \sum_{z\in\mathcal{Z}}\sum_{u:z\in\mathcal{T}_{u,\lambda}} p_{u,z} = \sum_{u\in[B]} p_{u,\mathcal{T}_{u,\lambda}}.$$

From (ε_L,δ_L)-differential privacy of R we have that

$$\max_{\mathcal{S}\subset\mathcal{Z}}\{p_{u,\mathcal{S}} - e^{\varepsilon_L}\cdot p_{v,\mathcal{S}}\}\leqslant\delta_L. \qquad (40)$$

By Fact 3.7 and the fact that the sets $\tilde{\mathcal{T}}^{(1)}_{u,\lambda},\ldots,\tilde{\mathcal{T}}^{(B/\lambda)}_{u,\lambda}$ are pairwise disjoint for any $u\in[B]$, we have that for all $u,v\in[B]$,

$$\sum_{k=1}^{B/\lambda}\sum_{z\in\tilde{\mathcal{T}}^{(k)}_{u,\lambda}}[p_{u,z} - e^{\varepsilon_L}\cdot p_{v,z}]_+\leqslant\delta_L. \qquad (41)$$

Averaging (40) over $v\in[B]$ gives that $\max_{\mathcal{S}\subset\mathcal{Z}}\{p_{u,\mathcal{S}} - e^{\varepsilon_L}q_{\mathcal{S}}\}\leqslant\delta_L$. Fact 3.7 then gives

$$\sum_{k=1}^{B/\lambda}\sum_{z\in\tilde{\mathcal{T}}^{(k)}_{u,\lambda}}[p_{u,z} - e^{\varepsilon_L}q_z]_+\leqslant\delta_L. \qquad (42)$$

By (41) and (42), we have that, for all $u, v \in [B]$,

$$e^{\varepsilon_L} \cdot \sum_{k=1}^{B/\lambda} \sum_{z \in \tilde{\mathcal{T}}_{u,\lambda}^{(k)}} \min\{p_{v,z}, q_z\} \geqslant p_{u,\mathcal{T}_{u,\lambda}} - 2\delta_L. \tag{43}$$

For each $v \in [B]$ and $1 \leqslant k \leqslant B/\lambda$,

$$1 - \Delta(P_v, Q) \geqslant \sum_{z \in \mathcal{Z}} \min\{p_{v,z}, q_z\} \geqslant \sum_{u \in [B]} \sum_{z \in \tilde{\mathcal{T}}_{u,\lambda}^{(k)}} \min\{p_{v,z}, q_z\}.$$

Averaging over k and using (43), it follows that for any $v \in [B]$,

$$\begin{aligned}
1 - \Delta(P_v, Q) &\geqslant \frac{\lambda}{B} \sum_{k=1}^{B/\lambda} \sum_{u \in [B]} \sum_{z \in \tilde{\mathcal{T}}_{u,\lambda}^{(k)}} \min\{p_{v,z}, q_z\} \\
&\geqslant \frac{\lambda}{B} \sum_{u \in [B]} \frac{p_{u,\mathcal{T}_{u,\lambda}} - 2\delta_L}{e^{\varepsilon_L}} \\
&\geqslant \left(\frac{\lambda}{e^{\varepsilon_L} B} \sum_{u \in [B]} p_{u,\mathcal{T}_{u,\lambda}} \right) - \frac{2\delta_L \lambda}{e^{\varepsilon_L}}.
\end{aligned}$$

By Lemma 3.15 and the $(\alpha, 1/4)$-accuracy of R, together with the fact that

$$\alpha/4 \geqslant \max\left\{ \frac{3\log(4/(1 - 2 \cdot 1/4))}{n}, \sqrt{\frac{3\log(4/(1 - 2 \cdot 1/4))}{nB}} \right\}$$

as long as the constant C is chosen large enough (recall the assumption $\alpha \geqslant C \max\{1/n, 1/\sqrt{nB}\}$), we have that, perhaps by making C even larger, $1 - \Delta(P_v, Q) \leqslant C\alpha^2 n$. In particular, it follows that

$$\frac{1}{B} \sum_{u \in [B]} p_{u,\mathcal{T}_{u,\lambda}} \leqslant \frac{e^{\varepsilon_L}}{\lambda} \cdot \left(C\alpha^2 n + \frac{2\delta_L \lambda}{e^{\varepsilon_L}} \right) = \frac{C\alpha^2 n e^{\varepsilon_L}}{\lambda} + 2\delta_L. \tag{44}$$

Using the inequality $\log(1 + \gamma h) \leqslant \gamma h$, we can now upper bound the first term in (39), when averaged over $v \in [B]$, as follows:

$$\begin{aligned}
&\frac{1}{B} \sum_{v \in [B]} \int_{h=1}^{B} \left(\frac{1}{h} + \gamma \left(1 - \frac{1}{h} \right) \right) \log(1 + \gamma h) d\rho_v(h) \\
&\leqslant \gamma + \frac{1}{B} \sum_{v \in [B]} \int_{h=1}^{2\exp(\varepsilon_L)} \gamma \log(1 + \gamma h) d\rho_v(h) + \frac{1}{B} \sum_{v \in [B]} \int_{h=2\exp(\varepsilon_L)}^{B} \gamma \log(1 + \gamma h) d\rho_v(h) \\
&\leqslant \gamma + \gamma^2 \int_{h=1}^{2\exp(\varepsilon_L)} h \cdot \left(\frac{1}{B} \sum_{v \in [B]} d\rho_v(h) \right) + \gamma \log B \int_{h=2\exp(\varepsilon_L)}^{B} \left(\frac{1}{B} \sum_{v \in [B]} d\rho_v(h) \right).
\end{aligned} \tag{45}$$

(In the integrals above, for an integral of the form $\int_{h=x}^{y}$ we integrate over the *closed* interval $[x, y]$, so that point masses at x and y are included.) Let ρ be the Borel measure on $\mathbb{R}$ defined by $\rho = \frac{1}{B} \sum_{v \in [B]} \rho_v$.

Recall that $(\varepsilon_L, \delta_L)$-differential privacy of R gives that for any $\mathcal{S} \subset \mathcal{Z}$, $p_{v,\mathcal{S}} \leqslant e^{\varepsilon_L} q_{\mathcal{S}} + \delta_L$. Thus, setting $\mathcal{S} = \{z \in \mathcal{Z} : p_{v,z} \geqslant 2e^{\varepsilon_L} q_z\}$ gives $p_{v,\mathcal{S}} \leqslant e^{\varepsilon_L} q_{\mathcal{S}} + \delta_L \leqslant e^{\varepsilon_L} \cdot p_{v,\mathcal{S}}/(2e^{\varepsilon_L}) + \delta_L$, so that $p_{v,\mathcal{S}} \leqslant 2\delta_L$. It follows by averaging this inequality over all $v \in [B]$ that

$$\gamma \log B \int_{h=2\exp(\varepsilon_L)}^{B} d\rho(h) \leqslant \gamma \log B \cdot 2\delta_L. \tag{46}$$

Next, note that (44) gives us that for any $\lambda \geqslant 1$,

$$\int_{h=\lambda}^{2\exp(\varepsilon_L)} d\rho(h) \leqslant \int_{h \geqslant \lambda} d\rho(h) = \frac{1}{B} \sum_{u \in [B]} p_{u,\mathcal{T}_{u,\lambda}} \leqslant \frac{C\alpha^2 n \cdot e^{\varepsilon_L}}{\lambda} + 2\delta_L.$$

It follows that

$$\begin{aligned}
&\int_{h=1}^{2\exp(\varepsilon_L)} h d\rho(h) \\
&\leqslant \sum_{h' \in \{1,2,\ldots,2^{\lceil \log 2\exp(\varepsilon_L) \rceil}\}} \int_{h'/2}^{h'} h d\rho(h) \\
&\leqslant \sum_{h' \in \{1,2,\ldots,2^{\lceil \log 2\exp(\varepsilon_L) \rceil}\}} 2h' \cdot \left(\frac{2C\alpha^2 n e^{\varepsilon_L}}{h'} + 2\delta_L \right) \\
&\leqslant (4 + \varepsilon_L \cdot \log(e)) \cdot \left(\alpha^2 n \cdot 4Ce^{\varepsilon_L}\right) + 32\delta_L e^{\varepsilon_L}. \qquad (47)
\end{aligned}$$

Next we upper bound the terms $h < 1$ in the sum of (39). Again using Lemma 3.15 and the $(\alpha, 1/4)$-accuracy of R, we see that there is a constant C such that for each $v \in [B]$, it holds that $1 - \Delta(P_v, Q) \leqslant C\alpha^2 n$. Hence we have

$$\sum_{z : p_{v,z} < q_z} p_{v,z} \leqslant C\alpha^2 n$$

and

$$\sum_{z : p_{v,z} < q_z} q_z \geqslant 1 - C\alpha^2 n.$$

We next need the following claim:

Claim 3.17. *Let $\tau \in (0,1)$. Suppose $\mathcal{Y}$ is a finite set and for each $z \in \mathcal{Y}$, $p_z, q_z \in [0,1]$ are defined such that $\sum_{z \in \mathcal{Y}} p_z \leqslant \tau$ and $1 - \tau \leqslant \sum_{z \in \mathcal{Y}} q_z \leqslant 1$. Suppose also that $p_z \leqslant q_z$ for all $z \in \mathcal{Y}$. Then for any $\gamma \in (0, 1/2)$,*

$$\sum_{z \in \mathcal{Y}} (\gamma p_z + (1-\gamma) q_z) \cdot \log\left(1 + \gamma\left(\frac{p_z}{q_z} - 1\right)\right) \leqslant -\gamma + 2\gamma\tau + \gamma^2(1+\tau).$$

Proof. Using the fact that $\log(1+x) \leqslant x$ for all $x \geqslant -1$, we have

$$\begin{aligned}
&\sum_{z \in \mathcal{Y}} (\gamma p_z + (1-\gamma) q_z) \cdot \log\left(1 + \gamma\left(\frac{p_z}{q_z} - 1\right)\right) \\
&\leqslant \sum_{z \in \mathcal{Y}} (\gamma p_z + (1-\gamma) q_z) \gamma \cdot \left(\frac{p_z}{q_z} - 1\right) \\
&\leqslant -(1-\tau)(1-\gamma)\gamma + \sum_{z \in \mathcal{Y}} (\gamma p_z + (1-\gamma) q_z) \gamma p_z / q_z \\
&\leqslant -(1-\tau-\gamma)\gamma + \gamma\tau + \sum_{z \in \mathcal{Y}} \gamma^2 p_z^2 / q_z \\
&\leqslant -\gamma + 2\gamma\tau + \gamma^2 + \gamma^2\tau.
\end{aligned}$$

□

Using Claim 3.17 with $\mathcal{Y} = \{z \in \mathcal{Z} : p_{v,z} < q_z\}$, $\tau = C\alpha^2 n \leqslant 1$ gives us that we may upper bound the second term in (39) as follows:

$$\sum_{z \in \cup_{h<1} \mathcal{U}_{v,h}} (\gamma p_{v,z} + (1-\gamma) q_z) \cdot \log\left(1 + \gamma\left(\frac{p_{v,z}}{q_z} - 1\right)\right) \leqslant -\gamma + 2C\gamma\alpha^2 n + 2\gamma^2. \qquad (48)$$

Combining (39), (45), (46), (47), and (48), we obtain

$$\frac{1}{B}\sum_{v\in[B]}\mathrm{KL}(P_{v,\gamma}||Q) \leqslant \left(\gamma+\gamma^2(4+\varepsilon_L\cdot\log(e))\cdot\alpha^2 n\cdot 4Ce^{\varepsilon_L}+\gamma^2\cdot 32e^{\varepsilon_L}\delta_L+2\gamma\delta_L\log B\right)+\left(-\gamma+2C\gamma\alpha^2 n+2\gamma^2\right)$$
$$\leqslant \gamma^2\alpha^2 n e^{\varepsilon_L}\cdot 4C\cdot(4+\varepsilon_L\log(e))+2C\gamma\alpha^2 n+36\gamma^2.$$

The second inequality above uses the facts that $\delta_L \leqslant \gamma/\log B$ and $\delta \leqslant 1/e^{\varepsilon_L}$.

□

Finally we prove Theorems 3.3 and 3.1 in the large-sample regime. The proof is a simple application of Lemma 3.10.

Lemma 3.18 (Proof of Theorem 3.3 in large-sample regime; i.e., (12) & (13))**.** *There is a sufficiently small positive constant c so that the following holds. Suppose $n, B \in \mathbb{N}$ with $n \geqslant 1/c$ and $\varepsilon_L, \delta_L \geqslant 0$ with $0 < \delta_L < c/(n\log n)$, and $0 \leqslant \varepsilon_L \leqslant 2\ln B - \ln\ln B - 1/c$. Then there is no protocol for n-user frequency estimation on $[B]$ that satisfies $(\varepsilon_L, \delta_L)$-local differential privacy and $(\alpha, 1/2)$-accuracy where*

$$\alpha \geqslant \begin{cases} \frac{c\sqrt{B}}{n\log^{1/4} B} & \text{for } n \geqslant B^2, \\ \frac{cB\ln^{1/7} B}{n} & \text{for } n \geqslant B^3. \end{cases}$$

Proof. We first treat the case that $n \geqslant B^2$. Set $n_0 = B^2$ and $\varepsilon_L = 2\ln B - \ln\ln B - 1/c$. Lemma 3.12 establishes that there is no $(\varepsilon_L, \delta_L)$-locally differentially private protocol that satisfies $\left(\frac{c}{\sqrt{n_0 B\sqrt{\log B}}}, 1/4\right)$-accuracy, for a sufficiently small constant c. But $1/\sqrt{n_0 B\sqrt{\log B}} = \frac{\sqrt{B}}{n_0\log^{1/4} B}$ as $n_0 = B^2$.

But by Lemma 3.10, any $(\varepsilon_L, \delta_L)$-locally differentially private protocol with n users, $\delta < c/(n\log n)$ and $\varepsilon_L = 2\ln B - \ln\ln B - 1/c$ which is $\left(\frac{c\sqrt{B}}{4n\log^{1/4} B}, 1/4\right)$-accurate yields an $(\varepsilon_L, \delta_L)$-locally differentially private protocol with n_0 users which is $\left(\frac{c\sqrt{B}}{n_0\log^{1/4} B}, 1/4\right)$-accurate.

The proof for the case $n \geqslant B^3$ is virtually identical, with $n_0 = B^2$ replaced by $n_1 = B^3$, for which the lower bound of Lemma 3.12 states that there is no $(\varepsilon_L, \delta_L)$-locally differentially private protocol that satisfies $\left(\frac{\ln^{1/7} B}{n^{2/3}}, 1/4\right)$-accuracy. □

Lemma 3.19 (Proof of Theorem 3.1 in large-sample regime; i.e., (8))**.** *For a sufficiently small positive constant c so that the following holds. If $\log B > 1/c$, $0 \leqslant \delta < c/n$, and $0 \leqslant \varepsilon \leqslant 1$, then there is no protocol for n-user frequency estimation in the single-message shuffled-model that is (ε, δ)-differentially private and $\left(\frac{c\sqrt{B}}{n\log B}, 1/4\right)$-accurate.*

Proof. Here we cannot use Lemma 3.5 in tandem with Lemma 3.18 since Lemma 3.18 requires a privacy parameter $\varepsilon_L \leqslant 2\ln B$ and the one produced by Lemma 3.5 grows as $\ln n$, which can be arbitrarily large. However, note that it is evident that Lemma 3.10 still applies if the protocol $P = (R, A)$ in the lemma statement is replaced by a single-message shuffled-model protocol $P = (R, S, A)$ (and the the protocol P' guaranteed by the lemma is $P' = (R', S, A')$).

In particular, letting c be the constant of Corollary 3.13, for $n_0 = \frac{cB^2}{\log B}$, Corollary 3.13 guarantees that for $\delta \leqslant 1/n_0, \varepsilon \leqslant 1$, there is no (ε, δ)-differentially private protocol in the single-message shuffled model that is $\left(\frac{c}{n_0^{3/4}\log^{1/4} n_0}, 1/4\right)$-accurate. But $\frac{c}{n_0^{3/4}\log^{1/4} n_0} = \frac{c^{5/4}\sqrt{B}}{n_0(\log^{1/4} n_0)(\log^{1/4} B)}$ by our choice of n_0.

By the modification of Lemma 3.10 mentioned in the previous paragraph, there is a sufficiently small constant $c' > 0$ so that for any $n \geqslant \frac{cB^2}{\log B}$, there is no (ε, δ)-differentially private protocol for frequency estimation in the n-user single-message shuffled model that is $\left(\frac{c'\sqrt{B}}{n\sqrt{\log B}}, 1/4\right)$-accurate. □

3.4 Proof of Lemma 3.15

In this section we prove Lemma 3.15. We first establish a more general statement in Lemma 3.20 below, which shows that if two distributions D, F have total variation distance bounded away from 1, then two distributions which can be obtained as the histograms of mixtures of i.i.d. samples from D, F have small total variation distance (much smaller than $\Delta(D, F)$).

We first recall the notation that we use to denote histograms of distributions. Given a tuple of random variables $(Y_1, \ldots, Y_n)$, $\mathrm{hist}(Y_1, \ldots, Y_n)$ denotes the distribution of the histogram of $(Y_1, \ldots, Y_n)$, i.e., of the function that maps each $z \in \mathcal{Z}$ to $|\{i : Y_i = z\}|$. We will denote histograms as functions $h : \mathcal{Z} \to \mathbb{N}$. If $(z_1, \ldots, z_n) \in \mathcal{Z}^n$ is such that all the z_i are distinct, then its histogram $h = \mathrm{hist}(z_1, \ldots, z_n)$ is a function $h : \mathcal{Z} \to \{0, 1\}$.

Lemma 3.20 (Total variation distance between histograms of mixture distributions)**.** *Suppose D, F are distributions on a finite set $\mathcal{Z}$. Suppose that for $\gamma \leqslant 1/\sqrt{n}$ such that $(1-\gamma)n/2$ is an integer, if $Z_1, \ldots, Z_n \sim D$ and $W_1, \ldots, W_n \sim F$ are iid, then*

$$\begin{aligned} &\Delta(\mathrm{hist}(Z_1, \ldots, Z_{(1-\gamma)n/2}, W_{(1-\gamma)n/2+1}, \ldots, W_n)), \\ &\mathrm{hist}(Z_1, \ldots, Z_{(1+\gamma)n/2}, W_{(1+\gamma)n/2+1}, \ldots, W_n)) \geqslant c. \end{aligned} \tag{49}$$

Then $\Delta(D, F) \geqslant 1 - c'\gamma^2 n$ for $c' = \Theta(1/c^2)$.

Notice that in the statement of Lemma 3.20

Proof of Lemma 3.20. We first introduce some notation. Given a set $\mathcal{Z}$, let $\mathcal{H}_{\mathcal{Z}}$ denote the set of all histograms on $\mathcal{Z}$; notice that elements $h \in \mathcal{H}_{\mathcal{Z}}$ can be thought of as functions $h : \mathcal{Z} \to \mathbb{Z}_{\geqslant 0}$. Given distributions D, F on a set $\mathcal{Z}$, as well as positive integers ν, n with $\nu \leqslant n$, let $R^{\nu,n}_{D,F}$ denote the distribution of the random variable

$$\mathrm{hist}(Z_1, \ldots, Z_\nu, W_{\nu+1}, \ldots, W_n),$$

where $Z_1, \ldots, Z_\nu \sim D$ i.i.d., and $W_{\nu+1}, \ldots, W_n \sim F$ i.i.d. Thus (49) may equivalently be written as

$$\Delta\left(R^{(1-\gamma)n/2,n}_{D,F}, R^{(1+\gamma)n/2,n}_{D,F}\right) \geqslant c. \tag{50}$$

A key tool in the proof of Lemma 3.20 is the data processing inequality (for total variation distance), stated below for convenience:

Lemma 3.21 (Data processing inequality)**.** *Suppose $\mathcal{Z}, \mathcal{Z}'$ are sets, D_0, D_1 are distributions on $\mathcal{Z}$ and $f : \mathcal{Z} \to \mathcal{Z}'$ is a randomized function. Suppose $Z_0 \sim D_0, Z_1 \sim D_1$. Then $\Delta(f(Z_0), f(Z_1)) \leqslant \Delta(Z_0, Z_1)$.*

Using Lemma 3.21 twice, we will reduce to the case in which $|\mathcal{Z}| = 3$, in two stages. For the first stage, consider the set $\mathcal{Z} \sqcup \mathcal{Z} := \{(z, b) : z \in \mathcal{Z}, b \in \{0, 1\}\}$, as well as the function $f : \mathcal{Z} \sqcup \mathcal{Z} \to \mathcal{Z}$, defined by $f((z, b)) = z$. Moreover define distributions D', F' on $\mathcal{Z} \sqcup \mathcal{Z}$, as follows: for all $z \in \mathcal{Z}$, we have

$$\begin{aligned} D'((z, 0)) &= D(z) - \min\{D(z), F(z)\}. \\ F'((z, 0)) &= F(z) - \min\{D(z), F(z)\}. \\ D'((z, 1)) &= \min\{D(z), F(z)\}. \\ F'((z, 1)) &= \min\{D(z), F(z)\}. \end{aligned}$$

It is immediate from the definition of D', F' that $\Delta(D', F') = \Delta(D, F)$. Moreover, it is clear that if Z_0, Z_1 are distributed according to D', F', respectively, then $f(Z_0), f(Z_1)$ are distributed according to D, F, respectively. To describe the effect of applying f to histograms on $\mathcal{Z} \sqcup \mathcal{Z}$, we make the following definition:

Definition 3.4 (Push-forward histogram)**.** Consider sets $\mathcal{Y}, \mathcal{Y}'$ together with a (possibly randomized) function $f : \mathcal{Y} \to \mathcal{Y}'$. For a histogram $h \in \mathcal{H}_{\mathcal{Y}}$, the *push-forward histogram* $f_* h \in \mathcal{H}_{\mathcal{Y}'}$ (i.e., $f_* h : \mathcal{Y}' \to \mathbb{Z}_{\geqslant 0}$) is defined as follows. If h is expressed as $h = \mathrm{hist}(y_1, \ldots, y_n)$ for $y_1, \ldots, y_n \in \mathcal{Y}$, then $f_* h$ is the (possibly random) histogram given by $\mathrm{hist}(f(y_1), \ldots, f(y_n))$.

It follows that if $H \in \mathcal{H}_{\mathcal{Z}\sqcup\mathcal{Z}}$ is a random variable distributed according to $R^{\nu,n}_{D',F'}$, then the push-forward histogram $f_*H : \mathcal{Z} \to \mathbb{Z}_{\geqslant 0}$ (which in this case is given by $f_*H(z) = H((z,0)) + H((z,1))$), is distributed according to $R^{\nu,n}_{D,F}$. Since f_*H is a (deterministic) function of H, it follows from Lemma 3.21 that

$$\Delta\left(R^{(1-\gamma)n/2,n}_{D,F}, R^{(1+\gamma)n/2,n}_{D,F}\right) \leqslant \Delta\left(R^{(1-\gamma)n/2,n}_{D',F'}, R^{(1+\gamma)n/2,n}_{D',F'}\right). \tag{51}$$

Next, we define a randomized function $g : \{-1,0,1\} \to \mathcal{Z}\sqcup\mathcal{Z}$, as follows. First make the following definitions:

$$\rho_0 = \sum_{z\in\mathcal{Z}} D'((z,1)) = \sum_{z\in\mathcal{Z}} F'((z,1)), \quad \rho_1 = \sum_{z\in\mathcal{Z}} F'((z,0)) = \sum_{z\in\mathcal{Z}} D'((z,0)) = 1-\rho_0.$$

Let $g(0)$ be the distribution over $\{(z,1) : z \in \mathcal{Z}\}$ that assigns to the point $(z,1)$ a mass of $\frac{D'((z,1))}{\rho_0} = \frac{F'((z,1))}{\rho_0}$. Let $g(-1)$ be the distribution over $\{(z,0) : z \in \mathcal{Z}\}$ that assigns to the point $(z,0)$ a mass of $\frac{D'((z,0))}{\rho_1}$. Finally let $g(1)$ be the distribution over $\{(z,0) : z \in \mathcal{Z}\}$ that assigns to the point $(z,0)$ a mass of $\frac{F'((z,0))}{\rho_1}$.

Finally define distributions D'', F'' on $\{-1,0,1\}$ as follows:

$$\begin{aligned} D''(-1) = \rho_1, \quad D''(0) = \rho_0, \quad D''(1) = 0 \\ F''(-1) = 0, \quad F''(0) = \rho_0, \quad F''(1) = \rho_1. \end{aligned}$$

From the *definitions* of D'', F'' we see that $\Delta(D'',F'') = \rho_1 = \Delta(D,F)$. Moreover, if Z'_0, Z'_1 are distributed according to D'', F'', respectively, then $g(Z'_0), g(Z'_1)$ are distributed according to D', F'. Next, let $H' \in \mathcal{H}_{\{-1,0,1\}}$ be the random histogram distributed according to $R^{\nu,n}_{D'',F''}$. Then the push-forward histogram $g_*H' \in \mathcal{H}_{\mathcal{Z}\sqcup\mathcal{Z}}$ is distributed according to $R^{\nu,n}_{D',F'}$. It follows from Lemma 3.21 that

$$\Delta\left(R^{(1-\gamma)n/2,n}_{D',F'}, R^{(1+\gamma)n/2,n}_{D',F'}\right) \leqslant \Delta\left(R^{(1-\gamma)n/2,n}_{D'',F''}, R^{(1+\gamma)n/2,n}_{D'',F''}\right). \tag{52}$$

Thus, since $\Delta(D'',F'') = \Delta(D',F') = \Delta(D,F)$, and by (50), (51), and (52) it suffices to show that there is some constant c' such that, assuming $\Delta\left(R^{(1-\gamma)n/2,n}_{D'',F''}, R^{(1+\gamma)n/2,n}_{D'',F''}\right) \geqslant c$, it follows that $1-\rho_0 = \Delta(D'',F'') \geqslant 1-c'\gamma^2 n$.

Let random variables $H^{(1-\gamma)n/2}, H^{(1+\gamma)n/2}$ be distributed according to $R^{(1-\gamma)n/2,n}_{D'',F''}$ and $R^{(1+\gamma)n/2,n}_{D'',F''}$, respectively. Since $H^{(1\pm\gamma)n/2}$ are each histograms of n elements, they are completely determined by their values on -1 and 1. Next note that the distribution of the tuple $(H^{(1-\gamma)n/2}(-1), H^{(1-\gamma)n/2}(1))$ is the distribution of $(Bin((1-\gamma)n/2, 1-\rho_0), Bin((1+\gamma)n/2, 1-\rho_0))$, where the two binomial random variables are independent. Similarly, the distribution of the tuple $(H^{(1+\gamma)n/2}(-1), H^{(1+\gamma)n/2}(1))$ is the distribution of $(Bin((1+\gamma)n/2, 1-\rho_0), Bin((1-\gamma)n/2, 1-\rho_0))$. To upper bound

$$\Delta((H^{(1-\gamma)n/2}(-1), H^{(1-\gamma)n/2}(1)), (H^{(1+\gamma)n/2}(-1), H^{(1+\gamma)n/2}(1)))$$

it therefore suffices to upper bound $\Delta\left(Bin((1-\gamma)n/2, 1-\rho_0), Bin((1+\gamma)n/2, 1-\rho_0)\right)$. To do this, we use [Roo06, Theorem 2, Eq. (15)], which implies that as long as $\gamma \geqslant 1/n$ and $\frac{3\gamma^2 n(1-\rho_0)}{\rho_0} < 1/4$, we have that

$$\Delta\left(Bin((1-\gamma)n/2, 1-\rho_0), Bin((1+\gamma)n/2, 1-\rho_0)\right) \leqslant 2\sqrt{e}\cdot\sqrt{\frac{3\gamma^2 n(1-\rho_0)}{\rho_0}} \tag{53}$$

If $\frac{3\gamma^2 n(1-\rho_0)}{\rho_0} \geqslant 1/4$ then we get $\rho_0 \leqslant 12\gamma^2 n$. Otherwise, (53) together with $\Delta\left(R^{(1-\gamma)n/2,n}_{D'',F''}, R^{(1+\gamma)n/2,n}_{D'',F''}\right) \geqslant c$ gives us that $\frac{c}{2} \leqslant 2\sqrt{e}\cdot\sqrt{3\gamma^2 n/\rho_0}$, meaning that $\rho_0 \leqslant O(\gamma^2 n/c^2)$. In particular, for some constant $c' = \Theta(1/c^2)$, we have shown that $1-\rho_0 \geqslant 1-c'\gamma^2 n$. □

Now we are ready to prove Lemma 3.15 using Lemma 3.20.

Proof of Lemma 3.15. By increasing α by at most a constant factor we may ensure that $(1 \pm \alpha)n/2$ are integers. We define two distributions of inputs, D_1 and D_2. For D_1, exactly $(1-\alpha)n/2$ of the x_i are set to v and the remaining $(1+\alpha)n/2$ of the x_i are drawn uniformly from $[B]$. For D_2, exactly $(1+\alpha)n/2$ of the x_i are set to v and the remaining $(1-\alpha)n/2$ of the x_i are drawn uniformly from $[B]$. Under both D_1 and D_2, the subset of which users are chosen to have their x_i fixed to v is chosen uniformly at random.

By the Chernoff bound, for $\lambda \geqslant 0$, as long as $\alpha \leqslant 1$, we have that

$$\mathbb{P}_{(x_1,\ldots,x_n)\sim D_1}\left[\sum_{i=1}^{n}(e_{x_i})_v \geqslant (1-\alpha)n/2 + (1+\lambda)(1+\alpha)n/(2B)\right] \leqslant \exp(-n\min\{\lambda,\lambda^2\}/(3B)). \quad (54)$$

$$\mathbb{P}_{(x_1,\ldots,x_n)\sim D_2}\left[\sum_{i=1}^{n}(e_{x_i})_v \leqslant (1+\alpha)n/2 + (1-\lambda)(1-\alpha)\cdot n/(2B)\right] \leqslant \exp(-n\min\{\lambda,\lambda^2\}/(3B)). \quad (55)$$

Choose λ so that $\exp(-n\min\{\lambda,\lambda^2\}/(3B)) = (1-2\beta)/4$, i.e., $\min\{\lambda,\lambda^2\}/B = 3/n \cdot \log(4/(1-2\beta))$. Explicitly, we have:

$$\lambda = \max\left\{\frac{3B\log(4/(1-2\beta))}{n}, \sqrt{\frac{3B\log(4/(1-2\beta))}{n}}\right\}.$$

We now consider two possible cases for λ:

Case 1. In the case that $\lambda = 3B\log(4/(1-2\beta))/n$, we have $\lambda \geqslant 1$. Moreover, since $2\lambda/B = 6\log(4/(1-2\beta))/n \leqslant \alpha/2$ (by (38)), the fact that R is $(\alpha/6,\beta)$-accurate implies that R is $(\frac{1}{3}\cdot(\alpha - 2\lambda/B),\beta)$-accurate. In turn it follows that R is $(\frac{1}{3}\cdot(\alpha-(\lambda+\alpha)/B),\beta)$-accurate.

Case 2. In the case that $\lambda = \sqrt{3B\log(4/(1-2\beta))/n}$, we have $\lambda \leqslant 1$. Since $\lambda/B = \sqrt{3\log(4/(1-2\beta))/(nB)} \leqslant \alpha/4 \leqslant \alpha/2 - \alpha/B$ (by (38)), the fact that R is $(\alpha/6,\beta)$-accurate implies that R is $(\frac{1}{3}\cdot(\alpha-(\lambda+\alpha)/B),\beta)$-accurate.

In both cases, it follows that, by Definition 3.1, there exists some analyzer $A : \mathcal{Z}^n \to [0,1]$ so that, for any dataset $X = (x_1,\ldots,x_n)$, with probability $1-\beta$ over the local randomizers, $\left|A(R(x_1),\ldots,R(x_n)) - \frac{1}{n}\sum_{i=1}^{n}(e_{x_i})_v\right| \leqslant \frac{1}{3}\cdot(\alpha-(\lambda+\alpha)/B)$. (Here we are only using accuracy on the vth coordinate.) Define $f : [0,1] \to \{0,1\}$ by $f(x) = 1$ if $x \geqslant \frac{1}{2} + \frac{(1+\lambda\alpha)n}{2B}$ and $f(x) = 0$ otherwise. Using (54) and (55) it follows that

$$\mathbb{E}_{(x_1,\ldots,x_n)\sim D_2}\left[f(A(R(x_1),\ldots,R(x_n)))\right] - \mathbb{E}_{(x_1,\ldots,x_n)\sim D_1}\left[f(A(R(x_1),\ldots,R(x_n)))\right] \geqslant 1 - 2\beta - 2\exp(-n\lambda/(12B)). \quad (56)$$

For $b \in \{1,2\}$, let $H_b : \mathcal{Z} \to \mathbb{Z}_{\geqslant 0}$ denote the random variable that is the histogram of the $R(x_i)$ when $(x_1,\ldots,x_n)$ are drawn according to D_b, and let $\tilde{D}_b$ be the distribution of H_b. For a histogram $h : \mathcal{Z} \to \mathbb{Z}_{\geqslant 0}$ with $\sum_{z\in\mathcal{Z}} h(z) = n$, define $G(h) \in \mathcal{Z}^n$ to be the random variable obtained by permuting uniformly at random the multiset consisting of $h(z)$ copies of z for each $z \in \mathcal{Z}$. Note that for $b \in \{1,2\}$, the distribution of $G(H_b)$ is exactly the distribution of $(R(x_1),\ldots,R(x_n))$ when $(x_1,\ldots,x_n) \sim D_b$. It follows from (56) and our choice of λ that

$$\mathbb{E}_{H_2\sim\tilde{D}_2}\left[f(A(G(H_2)))\right] - \mathbb{E}_{H_1\sim\tilde{D}_1}\left[f(A(G(H_1)))\right] \geqslant \frac{1-2\beta}{2}.$$

Hence $\Delta(\tilde{D}_1,\tilde{D}_2) \geqslant 1 - 2\beta - 2\exp(-n\lambda/(12B))$.

By definition, $\tilde{D}_1$ is the distribution of $\mathrm{hist}(Z_1,\ldots,Z_{n(1-\alpha)/2},W_{n(1-\alpha)/2+1},\ldots,W_n)$ and $\tilde{D}_2$ is the distribution of $\mathrm{hist}(Z_1,\ldots,Z_{n(1+\alpha)/2},W_{n(1+\alpha)/2+1},\ldots,W_n$ when $Z_1,\ldots,Z_n \sim P_v$ iid, and $W_1,\ldots,W_n \sim Q$ iid. By Lemma 3.20 with $D = P_v, F = Q$, we must have that $\Delta(P_v,Q) \geqslant 1 - c\alpha^2 n$ for some $c = \Theta\left(\frac{1}{(1-2\beta)^2}\right)$. □

3.5 Lower Bounds for Single-Message Selection

In this section we prove Theorem 1.2, stated formally below:

Theorem 3.22 (Nearly tight lower bound for single-message shuffled model selection)**.** *Suppose $n, B \in \mathbb{N}$ and $\varepsilon \leqslant O(1)$ and $\delta \leqslant o(1/(nB))$. Any single-message shuffled model protocol that is (ε,δ)-differentially private and solves the selection problem with n users and with probability at least $4/5$ must have $n \geqslant \Omega(B)$.*

As we did for frequency estimation in the single-message shuffled model, we will prove Theorem 3.22 by appealing to Lemma 3.5 and proving an analogous statement for $(\varepsilon_L, \delta_L)$-differentially private *local-model* protocols where the privacy parameter ε_L is approximately $\ln n$. The proof is similar in structure to that of [Ull18], established a lower bound on n which is tight in the case that $\varepsilon_L = O(1)$ and $\delta_L = 0$.

We begin by defining a distribution under which we shall show selection to be hard.

Definition 3.5 (Distributions $D_{\ell,j}$)**.** Fix $B \in \mathbb{N}$. For $\ell \in \{0,1\}, j \in [B]$, let the distribution $D_{\ell,j}$ be the uniform distribution on the subcube $\{x \in \{0,1\}^B : x_j = \ell\}$.

Let $\bar{D}$ denote the joint distribution of (L, J, X), where $(L, J) \sim \mathrm{Unif}(\{0,1\} \times [B])$ and conditioned on L, J, $X \sim D_{L,J}$.

Definition 3.6 (Distributions $P_{\ell,j}$)**.** Next suppose that for some finite set $\mathcal{Z}$, $R : \{0,1\}^B \to \mathcal{Z}$ is a fixed local randomizer. Let $P_{\ell,j}$ be the distribution of $R(X)$ when $X \sim D_{\ell,j}$. Let Q be the distribution $\frac{1}{2B}\sum_{j\in[B],\ell\in\{0,1\}} P_{\ell,j}$. Note that Q is the distribution of $R(X)$ when $X \sim \mathrm{Unif}(\{0,1\}^B)$.

To prove Theorem 3.22, we first establish Lemma 3.23 below, which applies to any protocol that is differentially private in the *local model* of differential privacy (with a large privacy parameter ε_L):

Lemma 3.23. *For a sufficiently small positive constant c, the following holds. Suppose $R : \{0,1\}^B \to \mathcal{Z}$ is an $(\varepsilon_L, \delta_L)$-(locally) differentially private protocol with $\delta_L \leqslant \frac{c}{n(B+\exp(\varepsilon_L))}$. Moreover suppose that $A : \mathcal{Z}^n \to \{0,1\} \times [B]$ is a function so that, if $L \sim \{0,1\}, J \sim [B]$ are uniform and independent, then*

$$\mathbb{P}_{L,J,X_1,\ldots,X_n\sim(D_{L,J}|L,J)}\left[A(R(X_1),\ldots,R(X_n)) = (L,J)\right] \geqslant \frac{1}{3}. \tag{57}$$

Then

$$n \geqslant \frac{cB\log B}{1+\varepsilon_L}.$$

Theorem 3.22 is a straightforward consequence of Lemma 3.23 and Lemma 3.5 (see [Ull18]). We provide the proof for completeness.

Proof of Theorem 3.22. Let $c_0 \in (0,1)$ be a sufficiently small positive constant to be specified later. Suppose for the purpose of contradiction that $P_S = (R, S, A)$ is an (ε, δ)-differentially private single-message shuffled model protocol that solves the selection problem with $n < c_0 B$ users and probability at least $4/5$. By Lemma 3.5, $P_L := (R, A)$ is an $(\varepsilon + \ln n, \delta)$-locally differentially private protocol that solves the selection problem with n users and probability at least $4/5$.

It follows by a Chernoff bound and a union bound that if $J \sim \mathrm{Unif}([B])$ and $X_1, \ldots, X_n \sim D_{1,J}|J$, then

$$\mathbb{P}_{J,X_1,\ldots,X_n\sim D_{1,J}|J}\left[A(R(X_1),\ldots,R(X_n)) = J\right] \geqslant 3/4 \tag{58}$$

as long as $n \geqslant \Omega(\sqrt{\log B})$. (In particular, we can guarantee that with probability at least $1 - 1/20$, for all $j' \neq J$, $\sum_{i=1}^n (X_i)_{j'} < \sum_{i=1}^n (X_i)_J - n/10 = 9n/10$.)

It follows from (58) that if $L \sim \mathrm{Unif}(\{0,1\})$ is independent of J,

$$\mathbb{P}_{J,X_1,\ldots,X_n\sim D_{L,J}|J}\left[A(R(X_1),\ldots,R(X_n)) = (J,L)\right] \geqslant 3/8.$$

The above equation is a contradiction to Lemma 3.23 in light of the fact that R is $(\varepsilon + \ln n, \delta)$-differentially private, $n < c_0 B < \frac{cB\log B}{1+\varepsilon+\ln n}$, and

$$\delta < \frac{c}{n(B + \exp(\varepsilon + \ln(n)))}.$$

(The above bound on δ can be seen by noting that $\delta < c_0/(nB)$ by assumption and $\exp(\varepsilon+\ln(n)) = O(n) \leqslant O(B)$.) □

The bulk of the proof of Theorem 3.22 is to establish an upper bound on $I((L,J); R(X))$, when $(L, J, X) \sim \bar{D}$. Lemma 3.24 below provides this upper bound.

Lemma 3.24. *Suppose $\varepsilon_L \geqslant 0$, $\delta_L \in (0,1)$, and R is $(\varepsilon_L, \delta_L)$-differentially private. Then we have that*

$$\mathbb{E}_{\ell\sim L, j\sim J}\left[\mathrm{KL}(P_{\ell,j}||Q)\right] \leqslant O\left(\frac{1+\varepsilon_L}{B} + \delta\cdot(B+e^{\varepsilon_L})\right),$$

where $P_{\ell,j}, Q$ are as defined in Definition 3.6.

The proof of Lemma 3.23 from Lemma 3.24 is entirely standard [Ull18]. We provide a proof for completeness.

Proof of Lemma 3.23. Suppose L, J are drawn uniformly from $\{0,1\}\times[B]$, and then $X_1,\dots,X_n \sim D_{L,J}$ are drawn i.i.d. Let $Z_1 = R(X_1),\dots,Z_n = R(x_n)$ denote the resulting random variables after passing $X_1,\dots,X_n$ through the local randomizer R. (In particular, $Z_1,\dots,Z_n$ are drawn i.i.d. according to $P_{L,J}$.) By Fano's inequality, for any deterministic function $f : \mathcal{Z}^n \to \{0,1\}\times[B]$, we have that

$$\begin{aligned}
&\mathbb{P}_{L,J,Z_1,\dots,Z_n}\left[f(Z_1,\dots,Z_n) = (L,J)\right] \\
&\leqslant \frac{1 + I((Z_1,\dots,Z_n);(L,J))}{\log 2B} \\
&\leqslant \frac{1 + n\cdot I(Z_1;(L,J))}{\log 2B} \\
&= \frac{1 + n\cdot \mathrm{KL}((Z_1,L,J)||Z_1\otimes(L,J))}{\log 2B} \\
&= \frac{1 + n\cdot \mathbb{E}_{(\ell,j)\sim\mathrm{Unif}(\{0,1\}\times[B])}[\mathrm{KL}(P_{\ell,j}||Q)]}{\log 2B} \\
&\leqslant \frac{1 + C\cdot\left(\frac{n(1+\varepsilon_L)}{B} + \delta n\cdot(B+e^{\varepsilon_L})\right)}{\log 2B}, \qquad (59)
\end{aligned}$$

where (59) uses Lemma 3.24 and C is a sufficiently large constant. If $n < \frac{cB\log B}{1+\varepsilon_L}$, then using the assumption on δ we may bound (59) above by

$$\frac{1 + C\cdot c\log B + c}{\log B},$$

which is strictly less than $1/3$ for a sufficiently small constant c, thus contradicting (57). □

Finally we prove Lemma 3.24.

Proof of Lemma 3.24. Recall the notation of Definition 3.2: For $x\in\{0,1\}^d$ and $z\in\mathcal{Z}$, we have $p_{x,z} = \mathbb{P}_R[R(x) = z]$, and for $\mathcal{S}\subset\mathcal{Z}$, $p_{x,\mathcal{S}} = \mathbb{P}_R[R(x)\in\mathcal{S}]$. Also set $q_z = \frac{1}{2^B}\sum_{x\in\{0,1\}^B} p_{x,z} = \mathbb{P}_{X\sim U_B,R}[R(X) = z] = p_{Z\sim Q}[Z = z]$ and $q_{\mathcal{S}} = \sum_{z\in\mathcal{S}} q_z$ for $\mathcal{S}\subset\mathcal{Z}$. Notice that

$$\begin{aligned}
&\mathbb{E}_{\ell\sim\mathrm{Unif}(\{0,1\}),j\sim\mathrm{Unif}([B])}\left[\mathrm{KL}(P_{\ell,j}||Q)\right] \\
&= \mathbb{E}_{\ell,j}\left[\sum_{z\in\mathcal{Z}} \mathbb{P}_{Z\sim P_{\ell,j}}[Z=z]\cdot\log\left(\frac{\mathbb{P}_{Z\sim P_{\ell,j}}[Z=z]}{\mathbb{P}_{Z\sim Q}[Z=z]}\right)\right] \\
&= \frac{1}{2B}\sum_{\ell\in\{0,1\},j\in[B]}\sum_{z\in\mathcal{Z}}\left(\frac{1}{2^{B-1}}\sum_{x\in\{0,1\}^B:x_j=\ell} p_{x,z}\right)\cdot\log\left(\frac{\frac{1}{2^{B-1}}\sum_{x\in\{0,1\}^B:x_j=\ell} p_{x,z}}{q_z}\right) \\
&= \sum_{z\in\mathcal{Z}}\frac{1}{2^B}\sum_{y\in\{0,1\}^B} p_{y,z}\cdot\frac{1}{B}\sum_{j\in[B]}\log\left(\frac{\frac{1}{2^{B-1}}\sum_{x:x_j=y_j} p_{x,z}}{q_z}\right) \\
&= \sum_{z\in\mathcal{Z}} q_z\cdot\frac{1}{B}\sum_{j\in[B]}\left(\frac{1}{2^B}\sum_{y\in\{0,1\}^B}\frac{p_{y,z}}{q_z}\cdot\log\left(\frac{1}{2^{B-1}}\sum_{x:x_j=y_j}\frac{p_{x,z}}{q_z}\right)\right). \qquad (60)
\end{aligned}$$

For each $z \in \mathcal{Z}$, define a function $f_z : \{0,1\}^B \to \mathbb{R}_{\geqslant 0}$ by $f_z(x) := \frac{p_{x,z}}{q_z}$. Thus, for any $j \in [B]$, $\mathbb{E}_{x\sim \mathrm{Unif}(\{0,1\}^B)}[f_z(x)] = \frac{\mathbb{E}_{x\sim D_{0,j}}[f_z(x)]+\mathbb{E}_{x\sim D_{1,j}}[f_z(x)]}{2} = 1$. We may now upper bound (60) by

$$\mathbb{E}_{\ell\sim\mathrm{Unif}(\{0,1\}),j\sim\mathrm{Unif}([B])}\left[\mathrm{KL}(P_{\ell,j}||Q)\right] \leqslant \sum_{z\in\mathcal{Z}} q_z \cdot \frac{1}{2B}\sum_{j,\ell}\mathbb{E}_{y\sim D_{\ell,j}}[f_z(y)] \cdot \log\left(\mathbb{E}_{y\sim D_{\ell,j}}[f_z(y)]\right). \tag{61}$$

For each $z \in \mathcal{Z}$, $x \in \{0,1\}^B$, set

$$g_z(x) = \begin{cases} f_z(x) & : f_z(x) \leqslant 2e^{\varepsilon_L} \\ 0 & : f_z(x) > 2e^{\varepsilon_L}, \end{cases}$$

and $h_z(x) := f_z(x) - g_z(x)$.

We next note the following basic fact.

Fact 3.25. *Suppose $g_0, g_1, h_0, h_1 \geqslant 0$ are real numbers such that $\frac{g_0+g_1+h_0+h_1}{2} = 1$. Then*

$$g_0\log(g_0+h_0) + g_1\log(g_1+h_1) \leqslant \frac{1}{2}(g_1-g_0)^2 + \frac{1}{2}(g_1-g_0)(h_1-h_0).$$

Proof of Fact 3.25. Let $c \in [0,1]$ be such that $g_0 + h_0 = 1 - c$ and $g_1 + h_1 = 1 + c$. Then using the fact that $\log(1+x) \leqslant x$ for all $x \geqslant -1$,

$$\begin{aligned} &g_0\log(g_0+h_0) + g_1\log(g_1+h_1) \\ &= g_0\log(1-c) + g_1\log(1+c) \\ &\leqslant -g_0c + g_1c \\ &= \frac{(g_1-g_0)+(h_1-h_0)}{2}\cdot(g_1-g_0), \end{aligned}$$

which leads to the desired claim. □

Recall that for a boolean function $f : \{0,1\}^B \to \mathbb{R}$ we have $\hat{f}(\{j\}) = \frac{1}{2}\left(\mathbb{E}_{x\sim D_{j,0}}[f(x)] - \mathbb{E}_{x\sim D_{j,1}}[f(x)]\right)$ for each $j \in [B]$. Using Fact 3.25 in (61) with $g_0 = \mathbb{E}_{x\sim D_{0,j}}[g_z(x)]$, $g_1 = \mathbb{E}_{x\sim D_{1,j}}[g_z(x)]$, $h_0 = \mathbb{E}_{x\sim D_{0,j}}[h_z(x)]$, and $h_1 = \mathbb{E}_{x\sim D_{1,j}}[h_z(x)]$ for each $z \in \mathcal{Z}$, $j \in [B]$, we obtain

$$\begin{aligned} &\mathbb{E}_{\ell\sim\mathrm{Unif}(\{0,1\}),j\sim\mathrm{Unif}([B])}\left[\mathrm{KL}(P_{\ell,j}||Q)\right] \\ &\leqslant \sum_{z\in\mathcal{Z}} q_z\cdot\frac{1}{B}\sum_{j\in[B]}\hat{g}_z(\{j\})^2 + \hat{g}_z(\{j\})\hat{h}_z(\{j\}) + \sum_{z\in\mathcal{Z}} q_z\cdot\frac{1}{2B}\sum_{j\in[B],\ell\in\{0,1\}}\mathbb{E}_{x\sim D_{\ell,j}}[h_z(x)]\cdot\log\left(\mathbb{E}_{x\sim D_{\ell,j}}[f_z(x)]\right) \\ &\leqslant \sum_{z\in\mathcal{Z}}\frac{q_z}{B}\mathbf{W}^1[g_z] + \sum_{z\in\mathcal{Z}}\frac{q_z}{B}\sum_{j\in[B]}\hat{g}_z(\{j\})\hat{h}_z(\{j\}) + \sum_{z\in\mathcal{Z}}\frac{q_z}{2}\sum_{j,\ell}\mathbb{E}_{x\sim D_{\ell,j}}[h_z(x)] \qquad (62) \\ &= \sum_{z\in\mathcal{Z}}\frac{q_z}{B}\mathbf{W}^1[g_z] + \sum_{z\in\mathcal{Z}}\frac{q_z}{B}\sum_{j\in[B]}\hat{g}_z(\{j\})\hat{h}_z(\{j\}) + \sum_{z\in\mathcal{Z}}\frac{Bq_z}{2}\cdot\mathbb{E}_{x\sim\mathrm{Unif}(\{0,1\}^B)}[h_z(x)]. \qquad (63) \end{aligned}$$

where the (62) uses the fact that $f_z(x) = \frac{p_{x,z}}{q_z} \leqslant 2^B$ for any $x \in \{0,1\}^B, z \in \mathcal{Z}$.

Next, notice that for an arbitrary non-negative-valued boolean function $f : \{0,1\}^B \to \mathbb{R}_{\geqslant 0}$, for and $j \in [B]$ we have $\hat{f}(\{j\}) \leqslant \mathbb{E}_{x\sim\mathrm{Unif}(\{0,1\}^B)}[f(x)]$. Also using the fact that $g_z(x) \leqslant 2e^{\varepsilon_L}$ for each $z \in \mathcal{Z}, x \in \{0,1\}^B$, we see that

$$\sum_{z\in\mathcal{Z}}\frac{q_z}{B}\sum_{j\in[B]}\hat{g}_z(\{j\})\hat{h}_z(\{j\}) \leqslant \sum_{z\in\mathcal{Z}} 2e^{\varepsilon_L}q_z\cdot\mathbb{E}_{x\sim\mathrm{Unif}(\{0,1\}^B)}[h_z(x)]. \tag{64}$$

Next we derive an upper bound on $\sum_{z\in\mathcal{Z}} q_z\cdot\mathbb{E}_{x\sim\mathrm{Unif}(\{0,1\}^B)}[h_z(x)]$. Here we will use the (ε_L,δ_L)-differential privacy of R; intuitively, the differential privacy of R constrains the ratio $p_{x,z}/q_z$ to be small for most $x \in \{0,1\}^B, z \in \mathcal{Z}$, except with probability δ_L.

For each $x \in \{0,1\}^B$, set $\mathcal{T}_x := \left\{ z \in \mathcal{Z} : \frac{p_{x,z}}{q_z} > 2e^{\varepsilon_L} \right\}$. Note that $h_z(x) > 0$ if and only if $z \in \mathcal{T}_x$. Then

$$\begin{aligned}
&\sum_{z \in \mathcal{Z}} q_z \cdot \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[h_z(x)] \\
&= \sum_{z \in \mathcal{Z}} q_z \cdot \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)} \left[\mathbb{1}[z \in \mathcal{T}_x] \cdot h_z(x) \right] \\
&= \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)} \left[\sum_{z \in \mathcal{T}_x} q_z \cdot h_z(x) \right] \\
&= \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)} \left[\sum_{z \in \mathcal{T}_x} p_{x,z} \right] \\
&= \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)} \left[p_{x,\mathcal{T}_x} \right], \qquad (65)
\end{aligned}$$

where we have used that for $z \in \mathcal{T}_x$, $h_z(x) = f_z(x) = p_{x,z}/q_z$. But since R is $(\varepsilon_L, \delta_L)$-differentially private, we have that $p_{x,\mathcal{T}_x} \leqslant e^{\varepsilon_L} \cdot p_{y,\mathcal{T}_x} + \delta_L$ for any $x, y \in \{0,1\}^B$. Averaging over all y, we obtain $p_{x,\mathcal{T}_x} \leqslant e^{\varepsilon_L} \cdot q_{\mathcal{T}_x} + \delta_L \leqslant e^{\varepsilon_L} \cdot \frac{p_{x,\mathcal{T}_x}}{2e^{\varepsilon_L}} + \delta_L$, so $p_{x,\mathcal{T}_x} \leqslant 2\delta_L$. Since this holds for all x, it follows by (63), (64) and (65) that

$$\begin{aligned}
\mathbb{E}_{\ell \sim \mathrm{Unif}(\{0,1\}), j \sim \mathrm{Unif}([B])} \left[\mathrm{KL}(P_{\ell,j} || Q) \right] &\leqslant \left(\sum_{z \in \mathcal{Z}} \frac{q_z}{B} \mathbf{W}^1[g_z] \right) + (2e^{\varepsilon_L} + B/2) \cdot \sum_{z \in \mathcal{Z}} q_z \cdot \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[h_z(x)] \\
&\leqslant \left(\sum_{z \in \mathcal{Z}} \frac{q_z}{B} \mathbf{W}^1[g_z] \right) + (2e^{\varepsilon_L} + B/2) \cdot 2\delta_L. \qquad (66)
\end{aligned}$$

By definition of g_z we have that $\hat{g}_z(\varnothing) = \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[g_z(x)] \leqslant \mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[f_z(x)] = 1$. For each $z \in \mathcal{Z}$, define a function $g'_z : \{0,1\}^B \to \mathbb{R}_{\geqslant 0}$, by $g'_z(x) = g_z(x) + (1 - \hat{g}_z(\varnothing))$. Certainly $\mathbf{W}^1[g_z] = \mathbf{W}^1[g'_z]$, $0 \leqslant g'_z(x) \leqslant 1 + 2e^{\varepsilon_L}$ for all x, and $\mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[g'_z(x)] = 1$. Now we apply the level-1 inequality, stated below for convenience.

Theorem 3.26 (Level-1 Inequality, [O'D14], Section 5.4)**.** *Suppose $f : \{0,1\}^B \to \mathbb{R}_{\geqslant 0}$ is a non-negative-valued boolean function with $0 \leqslant f(x) \leqslant L$ for all $x \in \{0,1\}^B$. Suppose also that $\mathbb{E}_{x \sim \mathrm{Unif}(\{0,1\}^B)}[f(x)] = 1$. Then $\mathbf{W}^1[f] \leqslant 6 \ln(L)$.*

Using Theorem 3.26, for each $z \in \mathcal{Z}$, with $f = g'_z, L = 1 + 2e^{\varepsilon_L}$, we get that $\mathbf{W}^1[g'_z] \leqslant 6 \ln(1 + 2e^{\varepsilon_L}) \leqslant 6 \ln(3e^{\varepsilon_L})$. From (66) it follows that

$$\mathbb{E}_{\ell \sim \mathrm{Unif}(\{0,1\}), j \sim \mathrm{Unif}([B])} \left[\mathrm{KL}(P_{\ell,j} || Q) \right] \leqslant \frac{6 \ln(3e^{\varepsilon_L})}{B} + (2e^{\varepsilon_L} + B/2) \cdot 2\delta_L,$$

as desired. □

4 Multi-Message Protocols for Frequency Estimation

In this section, we present new algorithms for private frequency estimation in the shuffled model that significantly improve on what can be achieved in the local model of differential privacy. By our previous lower bounds, such protocols must necessarily use multiple messages. Our results are summarized in Table 2, which focuses on communication requirements of users, the size of the additive error on query answers, and the time required to answer a query (after creating a data structure based on the shuffled dataset). To our best knowledge, the only previously known upper bounds for these problems in the shuffled model (going beyond local differential privacy) followed via a reduction to private aggregation [CSU+19]. Using the currently best protocol for private aggregation in the shuffled model [BBGN19a, GPV19] yields the result stated in the first row of Table 2. Since the time to answer a frequency query differs by a factor of $\tilde{\Theta}(n)$ between our public and private coin protocols, we include query time bounds in the table. We start by stating the formal guarantees on our *private-coin* multi-message protocol.

Problem	Messages per user	Message size in bits	Error	Query time
Frequency estimation (private randomness) [CSU+19, BBGN19a, GPV19]	B	$\log B$	$\sqrt{\log(B)\log\frac{1}{\delta}}/\varepsilon$ (expected error)	1
Frequency estimation (private randomness) Section 4.1	$\frac{\log(1/\varepsilon\delta)}{\varepsilon^2}$	$\log n \log B$	$\log B + \frac{\sqrt{\log(B)\log(1/(\varepsilon\delta))}}{\varepsilon}$	$\frac{n\log\left(\frac{1}{\varepsilon\delta}\right)\log n\log B}{\varepsilon^2}$
Frequency estimation (public randomness) Section 4.2	$\frac{\log^3(B)\log(\log(B)/\delta)}{\varepsilon^2}$ B^η	$\log n + \log\log B$ $\log B$	$\frac{\log^{3/2}(B)\sqrt{\log(\log B/\delta)}}{\varepsilon}$ $\sqrt{\log(B)\log\frac{1}{\delta}}/\varepsilon$	$\log B$ 1

Table 2: Overview of bounds on frequency estimation in the shuffled model with multiple messages. Each user is assumed to hold $k = 1$ value from $[B]$, and $\eta > 0$ is a constant. The query time stated is the additional time to answer a query, assuming a preprocessing of the output of the shuffler that takes time linear in its length. Note that frequencies and counts are not normalized, i.e., they are integers in $\{0, \ldots, n\}$. For simplicity of presentation in this table, constant factors are suppressed, the bounds are stated for error probability $\beta = B^{-O(1)}$, and the following are assumed: n is bounded above by B, and $\delta < 1/\log B$.

Theorem 4.1 (Frequency estimation via private-coin multi-message shuffling)**.** *Let n and B be positive integers and $\varepsilon > 0$ and $\delta \in (0,1)$ be real numbers. Then there exists a private-coin (ε,δ)-differentially private algorithm in the shuffled model for frequency estimation on n users and domain size B with error $O\left(\log B + \frac{\sqrt{\log(B)\log(1/(\varepsilon\delta))}}{\varepsilon}\right)$ and with $O\left(\frac{\log(1/\varepsilon\delta)}{\varepsilon^2}\right)$ messages per user, where each message consists of $O(\log n \log B)$ bits. Moreover, any frequency query can be answered in time $O\left(\frac{n\log\left(\frac{1}{\varepsilon\delta}\right)\log n\log B}{\varepsilon^2}\right)$.*

Theorem 4.1 is proved in Section 4.1. We next state the formal guarantees of our *public-coin* multi-message protocols, whose main advantage compared to the private-coin protocol is that it has polylogarithmic query time.

Theorem 4.2 (Frequency estimation via public-coin multi-message shuffling)**.** *Let n and B be positive integers and $\varepsilon > 0$ and $\delta \in (0,1)$ be real numbers. Then there exists a public-coin (ε,δ)-differentially private algorithm in the shuffled model for frequency estimation on n users and domain size B with error $O\left(\frac{\log^{3/2}(B)\sqrt{\log(\log B/\delta)}}{\varepsilon}\right)$ and with $O\left(\frac{\log^3(B)\log(\log(B)/\delta)}{\varepsilon^2}\right)$ messages per user, where each message consists of $O(\log n + \log\log B)$ bits. Moreover, any frequency query can be answered in time $O(\log B)$.*

The public-coin protocol in Theorem 4.2 and its analysis are presented in Section 4.2.

Prior work [BS15, BNST17, BNS18] has focused on the case of computing heavy hitters when each user holds only a single element. While we focus primarily on this case, we will also consider the application of frequency estimation to the task of computing range counting queries (given in Section 5 below), where we apply our protocols for a frequency oracle as a black box and need to deal with the case in which a user can hold $k > 1$ inputs. Thus, in the rest of this section, we state our results for more general values of k (although we do not attempt to optimize our algorithms for large values of k, as k will be at most $\mathrm{poly}\log(n)$ in our application to range counting queries). Moreover, our results for $k \geqslant 1$ can be interpreted as establishing bounds for privately computing sparse families of counting queries (see Appendix B).

For clarity, we point out that the privacy of our protocols holds for every setting of the public random string. In other words, public randomness is assumed to be known by the analyzer, and affects error but not privacy.

4.1 Private-Coin Protocol

In this section, we give a private-coin protocol (i.e., one where the only source of randomness is the private coin source at each party) for frequency estimation with polylogarithmic error and polylogarithmic bits of communication

per user. In the case of local DP, private-coin protocols were recently obtained by Acharya et al. in [ASZ19, AS19]. These works made use of the Hadamard response for the local randomizers instead of previous techniques developed in the local model and which relied on public randomness. The Hadamard response was also used in [CKS19, CKS18, NXY+16] for similar applications, namely, private frequency estimation.

Overview. For any power of two $B \in \mathbb{N}$, let $H_B \in \{-1,1\}^{B \times B}$ denote the $B \times B$ Hadamard matrix and for $j \in [B-1]$, set $\mathcal{H}_{B,j} := \{j' \in [B] \mid H_{j+1,j'} = 1\}$[13]. By orthogonality of the rows of H_B, we have that $|\mathcal{H}_{B,j}| = B/2$ for any $j \in [B-1]$ and for all $j \neq j'$, it is the case that $|\mathcal{H}_{B,j} \cap \mathcal{H}_{B,j'}| = B/4$. For any $\tau \in \mathbb{N}$, we denote the τ-wise Cartesian product of $\mathcal{H}_{B,j}$ by $\mathcal{H}^{\tau}_{B,j} \subset [B]^{\tau}$. In the *Hadamard response* [ASZ19], a user whose data consists of an index $j \in [B]$ sends to the server a random index $j' \in [B]$ that is, with probability $\frac{e^{\varepsilon}}{1+e^{\varepsilon}}$, chosen uniformly at random from the "Hadamard codeword" $\mathcal{H}_{B,j}$ and, with probability $\frac{1}{1+e^{\varepsilon}}$, chosen uniformly from $[B] \backslash \mathcal{H}_{B,j}$.

In the shuffled model, much less randomization is needed to protect a user's privacy than in the local model of differential privacy, where the Hadamard response was previously applied. In particular, we can allow the users to send more information about their data to the server, along with some "blanket noise"[14] which helps to hide the true value of *any* one individual's input. Our adaptation of the Hadamard response to the multi-message shuffled model for computing frequency estimates (in the case where each user holds up to k elements) proceeds as follows (see Algorithm 1 for the detailed pseudo-code). Suppose the n users possess data $\mathcal{S}_1, \ldots, \mathcal{S}_n \subset [B]$ such that $|\mathcal{S}_i| \leqslant k$ — equivalently, they possess $x_1, \ldots, x_n \in \{0,1\}^B$, such that for each $i \in [n]$, $\|x_i\|_1 \leqslant k$ (the nonzero indices of x_i are the elements of $\mathcal{S}_i$). Given x_i, the local randomizer R^{Had} *augments* its input by adding $k - \|x_i\|_1$ arbitrary elements from the set $\{B+1, \ldots, 2B-1\}$ (recall that $k < B$). (Later, the analyzer will simply ignore the augmented input in $\{B+1, \ldots, 2B-1\}$ from the individual randomizers. The purpose of the augmentation is to guarantee that all sets $\mathcal{S}_i$ will have cardinality exactly k, which facilitates the privacy analysis.) Let the augmented input be denoted $\tilde{x}_i$, so that $\tilde{x}_i \in \{0,1\}^{2B-1}$ and $\|\tilde{x}_i\|_1 = k$. For each index j at which $(\tilde{x}_i)_j \neq 0$, the local randomizer chooses τ indices $a_{j,1}, \ldots, a_{j,\tau}$ in $\mathcal{H}_{2B,j}$ uniformly and independently, and sends each tuple $(a_{j,1}, \ldots, a_{j,\tau})$ to the shuffler. It also generates ρ tuples $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})$ where each of $\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau}$ is uniform over $[2B]$, and sends these to the shuffler as well; these latter tuples constitute "blanket noise" added to guarantee differential privacy.

Given the output of the shuffler, the analyzer A^{Had} determines estimates $\hat{x}_j$ for the frequencies of each $j \in [B]$ by counting the number of messages $(a_1, \ldots, a_\tau) \in [2B]^{\tau}$ which belong to $\mathcal{H}^{\tau}_{2B,j}$. The rationale is that each user i such that $j \in \mathcal{S}_i$ will have sent such a message in $\mathcal{H}^{\tau}_{2B,j}$. As the analyzer could have picked up some of the blanket noise in this count, as well as tuples sent by users holding some $j' \neq j$, since $\mathcal{H}^{\tau}_{2B,j} \cap \mathcal{H}^{\tau}_{2B,j'} \neq \varnothing$, it then corrects this count (Algorithm 1, Line 19) to obtain an unbiased estimate $\hat{x}_j$ of the frequency of j.

Analysis. The next theorem summarizes the privacy, accuracy and efficiency properties of Algorithm 1 for general values of k.

Theorem 4.3. *There is a sufficiently large positive absolute constant ζ such that the following holds. Suppose $n, B, k \in \mathbb{N}$ with $k < B$, and $0 \leqslant \varepsilon, \delta, \beta \leqslant 1$. Consider the shuffled-model protocol $P^{\mathrm{Had}} = (R^{\mathrm{Had}}, S, A^{\mathrm{Had}})$ with $\tau = \log n$ and $\rho = \frac{36k^2}{\varepsilon^2}\left(\ln \frac{ek}{\varepsilon\delta}\right)$. Then P^{Had} is a (ε, δ)-differentially private protocol (Definition 2.2) with $O\left(\frac{k^2 \log(k/\varepsilon\delta)}{\varepsilon^2}\right)$ messages per user, each consisting of $O(\log n \log B)$ bits, such that for inputs $x_1, \ldots, x_n \in \{0,1\}^B$ satisfying $\|x_i\|_1 \leqslant k$, the estimates $\hat{x}_j$ produced by the output of $P^{\mathrm{Had}}(n, B, \tau, \rho, k)$ satisfy*

$$\mathbb{P}\left[\forall j \in [B] \ : \ \left|\hat{x}_j - \sum_{i=1}^{n} x_{i,j}\right| \leqslant O\left(\log(B/\beta) + \frac{k\sqrt{\log(B/\beta)\log(k/\varepsilon\delta)}}{\varepsilon}\right)\right] \geqslant 1 - \beta. \tag{67}$$

Moreover, any frequency query can be answered in time $O\left(n \log n \log B \left(\frac{k^2 \log(k/\varepsilon\delta)}{\varepsilon^2}\right)\right)$.

[13] Since the first row of the Hadamard matrix H_B is all 1's, we cannot use the first row in our frequency estimation protocols. This is the reason for the subscript of $j+1$ in the definition of $\mathcal{H}_{B,j}$.

[14] This uses the expression of Balle et al. [BBGN19c]

Algorithm 1: Local randomizer and analyzer for frequency estimation via Hadamard response

1 $R^{\mathrm{Had}}(n, B, \tau, \rho, k)$**:**
  Input: Set $\mathcal{S} \subset [B]$ specifying i's input set;
  Parameters $n, B, \tau, \rho, k \in \mathbb{N}$
  Output: A multiset $\mathcal{T} \subset \{0,1\}^{\log 2B \cdot \tau}$
2 **for** $j = B+1, B+2, \ldots, 2B-1$ **do**
    `// Augmentation step`
3 **if** $|\mathcal{S}| < k$ **then**
4 $\mathcal{S} \leftarrow \mathcal{S} \cup \{j\}$
5 **for** $j \in \mathcal{S}$ **do**
6 Choose $a_{j,1}, \ldots, a_{j,\tau} \in \mathcal{H}_{2B,j}$ uniformly and independently at random
7 **for** $g = 1, 2, \ldots, \rho$ **do**
8 Choose $\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau} \in [2B]$ uniformly and independently at random
9 **return** $\mathcal{T} := \bigcup_{j \in \mathcal{S}} \{(a_{j,1}, \ldots, a_{j,\tau})\} \cup \bigcup_{1 \leqslant g \leqslant \tilde{\rho}} \{(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})\}$ `// Each element of` $\mathcal{T}$ `is viewed as an element of` $(\{0,1\}^{\log 2B})^{\tau}$`, by associating each element of` $[2B]$ `with its binary representation.`
10 $A^{\mathrm{Had}}(n, B, \tau, \rho, k)$**:**
  Input: Multiset $\{y_1, \ldots, y_m\}$ consisting of outputs of local randomizers, $y_i \in (\{0,1\}^{\log 2B})^{\tau}$;
  Parameters $n, B, \tau, \rho, k \in \mathbb{N}$
  Output: A vector $\hat{x} \in \mathbb{R}^B$ containing estimates of the frequency of each $j \in [B]$
11 **for** $j \in [B]$ **do**
12 Let $\hat{x}_j \leftarrow 0$
13 **for** $j \in [B]$ **do**
14 **for** $i \in [m]$ **do**
15 Write $y_i \in (\{0,1\}^{\log 2B})^{\tau}$ as $y_i := (a_{i,1}, \ldots, a_{i,\tau})$, with $a_{i,1}, \ldots, a_{i,\tau} \in \{0,1\}^{\log 2B}$
16 **if** $\{a_{i,1}, \ldots, a_{i,\tau}\} \subset \mathcal{H}_{2B,j}$ **then**
17 $\hat{x}_j \leftarrow \hat{x}_j + 1$
18 **for** $j \in [B]$ **do**
19 $\hat{x}_j \leftarrow \frac{1}{1-2^{-\tau}} \cdot (\hat{x}_j - (\rho + k) n 2^{-\tau})$ `// De-biasing step`
20 **return** $\hat{x}$

Before we prove Theorem 4.3, instantiating Theorem 4.3 with $k = 1$ directly implies Theorem 4.1.

Theorem 4.3 is a direct consequence of Lemmas 4.4, 4.5, and 4.6, which establish the privacy, accuracy, and efficiency guarantees, respectively, of protocol P^{Had}. The remainder of this section presents and proves the aforementioned lemmas. We begin with Lemma 4.4, which establishes DP guarantees of P^{Had}.

Lemma 4.4 (Privacy of P^{Had})**.** *Fix $n, B \in \mathbb{N}$ with B a power of 2. Let $\tau = \log n$, $\varepsilon \leqslant 1$, and $\rho = \frac{36 \ln 1/\delta}{\varepsilon^2}$. Then the algorithm $S \circ R^{\mathrm{Had}}(n, B, \tau, \rho, k)$ is $(k\varepsilon, \delta \exp(k\varepsilon)/\varepsilon)$-differentially private.*

Proof. For convenience let $P := S \circ R^{\mathrm{Had}}(n, B, \tau, \rho, k)$ be the protocol whose (ε, δ)-differential privacy we wish to establish. With slight abuse of notation, we will assume that P operates on the augmented inputs $(\tilde{x}_1, \ldots, \tilde{x}_n)$ (see Algorithm 1, Line 4). In particular, for inputs $(x_1, \ldots, x_n)$ that lead to augmented inputs $(\tilde{x}_1, \ldots, \tilde{x}_n)$, we will let $P(\tilde{x}_1, \ldots, \tilde{x}_n)$ be the output of P when given as inputs $x_1, \ldots, x_n$. Let $\mathcal{Y}$ be the set of multisets consisting of elements of $\{0,1\}^{\log 2B \times \tau}$; notice that the output of P lies in $\mathcal{Y}$.

By symmetry, it suffices to show that for any augmented inputs of the form $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_{n-1}, \tilde{x}_n)$ and $\tilde{X}' =$

$(\tilde{x}_1, \ldots, \tilde{x}_{n-1}, \tilde{x}'_n)$, and for any subset $\mathcal{U} \subset \mathcal{Y}$, we have that

$$\mathbb{P}[P(\tilde{x}_1, \ldots, \tilde{x}_n) \in \mathcal{U}] \leqslant e^{\varepsilon} \cdot \mathbb{P}[P(\tilde{x}_1, \ldots, \tilde{x}_{n-1}, \tilde{x}'_n) \in \mathcal{U}] + \delta. \tag{68}$$

We first establish (68) for the special case that $\tilde{x}_n, \tilde{x}'_n$ differ by 1 on two indices, say j, j', while having the same ℓ_1 norm: in particular, we have $|(\tilde{x}_n)_j - (\tilde{x}'_n)_j| = 1$ and $|(\tilde{x}_n)_{j'} - (\tilde{x}'_n)_{j'}| = 1$. By symmetry, without loss of generality we may assume that $j = 1, j' = 2$ and that $(\tilde{x}_n)_j - (\tilde{x}'_n)_j = 1$ while $(\tilde{x}'_n)_{j'} - (\tilde{x}_n)_{j'} = 1$. To establish (68) in this case, we will in fact prove a stronger statement: for inputs $(\tilde{x}_1, \ldots, \tilde{x}_n)$, define the *view* of an adversary, denoted by $\mathrm{View}_P(\tilde{x}_1, \ldots, \tilde{x}_n)$, as the tuple consisting of the following components:

- For each $i \in [n-1]$, the set $\hat{\mathcal{S}}_i := \bigcup_{j:(\tilde{x}_i)_j=1}\{(a_{j,1}, \ldots, a_{j,\tau})\}$ of tuples output by user i corresponding to her true input $\tilde{x}_i$.
- The set $\hat{\mathcal{S}}_n := \bigcup_{j:j\notin\{1,2\},(\tilde{x}_n)_j=1}\{(a_{j,1}, \ldots, a_{j,\tau})\}$ of tuples output by user n corresponding to her true (augmented) input $\tilde{x}_n$, except (if applicable) the string that would be output if $(\tilde{x}_n)_1 = 1$ or $(\tilde{x}_n)_2 = 1$.
- The multiset $\{y_1, \ldots, y_m\}$ consisting of the outputs of the n users of the protocol P.

It then suffices to show the following:

$$\mathbb{P}_{V\sim\mathrm{View}_P(\tilde{x}_1,\ldots,\tilde{x}_n)}\left[\frac{\mathbb{P}[\mathrm{View}_P(\tilde{x}_1, \ldots, \tilde{x}_{n-1}, \tilde{x}_n) = V]}{\mathbb{P}[\mathrm{View}_P(\tilde{x}_1, \ldots, \tilde{x}_{n-1}, \tilde{x}'_n) = V]} \geqslant e^{\varepsilon}\right] \leqslant \delta. \tag{69}$$

(See [BBGN19c, Theorem 3.1] for a similar argument.)

Notice that each of the elements $y_1, \ldots, y_m$ in the output of the protocol P consists of a tuple $(a_1, \ldots, a_\tau)$, where each $a_1, \ldots, a_\tau \in [2B]$. Now we will define a joint distribution (denoted by $\mathcal{D}$) of random variables $(W_{a_1,\ldots,a_\tau})_{a_1,\ldots,a_\tau\in[2B]}, Q, Q'$, where, for each $(a_1, \ldots, a_\tau) \in [2B]^\tau$, $W_{a_1,\ldots,a_\tau} \in \mathbb{Z}_{\geqslant 0}$, and $Q, Q' \in [2B]^\tau$, as follows. For each tuple $(a_1, \ldots, a_\tau) \in [2B]^\tau$, we let $W_{a_1,\ldots,a_\tau}$ be jointly distributed from a multinomial distribution over $[2B]^\tau$ with ρn trials. For each $(a_1, \ldots, a_\tau) \in [2B]^\tau$, let $\hat{W}_{a_1,\ldots,a_\tau}$ be the random variable representing the number of tuples $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})$ generated on Line 8 of Algorithm 1 satisfying $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau}) = (a_1, \ldots, a_\tau)$. Notice that the joint distribution of all $W_{a_1,\ldots,a_\tau}$ is the same as the joint distribution of $\hat{W}_{a_1,\ldots,a_\tau}$, for $(a_1, \ldots, a_\tau) \in [2B]^\tau$. Intuitively, $W_{a_1,\ldots,a_\tau}$ represents the blanket noise added by the outputs $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})$ in Line 8 of Algorithm 1. Also let $Q, Q' \in [2B]^\tau$ be random variables that are distributed uniformly over $\mathcal{H}^\tau_{2B,1}, \mathcal{H}^\tau_{2B,2}$, respectively. Then since the tuples $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})$ are distributed independently of the tuples $(a_{j,1}, \ldots, a_{j,\tau})$ $(j \in \mathcal{S}_i)$, (69) is equivalent to

$$\mathbb{P}_{w_{a_1,\ldots,a_\tau},q,q'\sim\mathcal{D}}\left[\frac{\mathbb{P}_{W_{a_1,\ldots,a_\tau},Q,Q'\sim\mathcal{D}}\left[\forall(a_1,\ldots,a_\tau)\in[2B]^\tau : W_{a_1,\ldots,a_\tau} + \mathbb{1}[Q=(a_1,\ldots,a_\tau)] = w_{a_1,\ldots,a_\tau} + \mathbb{1}[q=(a_1,\ldots,a_\tau)]\right]}{\mathbb{P}_{W_{a_1,\ldots,a_\tau},Q,Q'\sim\mathcal{D}}\left[\forall(a_1,\ldots,a_\tau)\in[2B]^\tau : W_{a_1,\ldots,a_\tau} + \mathbb{1}[Q'=(a_1,\ldots,a_\tau)] = w_{a_1,\ldots,a_\tau} + \mathbb{1}[q=(a_1,\ldots,a_\tau)]\right]} \geqslant e^{\varepsilon}\right] \leqslant \delta. \tag{70}$$

Set $\tilde{w}_{a_1,\ldots,a_\tau} := w_{a_1,\ldots,a_\tau} + \mathbb{1}[q = (a_1, \ldots, a_\tau)]$. By the definition of $\mathcal{D}$ we have

$$\begin{aligned}
&\mathbb{P}_{W_{a_1,\ldots,a_\tau},Q,Q'\sim\mathcal{D}}\left[\forall(a_1,\ldots,a_\tau)\in[2B]^\tau : W_{a_1,\ldots,a_\tau} + \mathbb{1}[Q=(a_1,\ldots,a_\tau)] = \tilde{w}_{a_1,\ldots,a_\tau}\right] \\
&= \mathbb{E}_{Q\sim\mathcal{D}}\left[(2B)^{-\tau\rho n}\cdot\binom{(2B)^\tau}{\{\tilde{w}_{a_1,\ldots,a_\tau} - \mathbb{1}[Q=(a_1,\ldots,a_\tau)]\}_{(a_1,\ldots,a_\tau)\in[2B]^\tau}}\right] \\
&= \left(\frac{2}{2B}\right)^\tau \cdot (2B)^{-\tau\rho n}\binom{(2B)^\tau}{\{\tilde{w}_{a_1,\ldots,a_\tau}\}_{(a_1,\ldots,a_\tau)\in[2B]^\tau}} \cdot \sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,1}} \tilde{w}_{a'_1,\ldots,a'_\tau}.
\end{aligned}$$

In the above equation, the notation such as $\binom{(2B)^\tau}{\{\tilde{w}_{a_1,\ldots,a_\tau}\}_{(a_1,\ldots,a_\tau)\in[2B]^\tau}}$ refers to the multinomial coefficient, equal to $\frac{((2B)^\tau)!}{\prod_{a_1,\ldots,a_\tau\in[2B]}\tilde{w}_{a_1,\ldots,a_\tau}!}$. Similarly, for the denominator of the expression in (70),

$$\begin{aligned}
&\mathbb{P}_{W_{a_1,\ldots,a_\tau},Q,Q'\sim\mathcal{D}}\left[\forall(a_1,\ldots,a_\tau)\in[2B]^\tau : W_{a_1,\ldots,a_\tau} + \mathbb{1}[Q'=(a_1,\ldots,a_\tau)] = \tilde{w}_{a_1,\ldots,a_\tau}\right] \\
&= \mathbb{E}_{Q\sim\mathcal{D}}\left[(2B)^{-\tau\rho n}\cdot\binom{(2B)^\tau}{\{\tilde{w}_{a_1,\ldots,a_\tau} - \mathbb{1}[Q'=(a_1,\ldots,a_\tau)]\}_{(a_1,\ldots,a_\tau)\in[2B]^\tau}}\right] \\
&= \left(\frac{2}{2B}\right)^\tau \cdot (2B)^{-\tau\rho n}\binom{(2B)^\tau}{\{\tilde{w}_{a_1,\ldots,a_\tau}\}_{(a_1,\ldots,a_\tau)\in[2B]^\tau}} \cdot \sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,2}} \tilde{w}_{a'_1,\ldots,a'_\tau}.
\end{aligned}$$

Thus, (70) is equivalent to

$$\mathbb{P}_{w_{a_1,\ldots,a_\tau},q,q'\sim\mathcal{D}}\left[\frac{\sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,1}}\tilde{w}_{a'_1,\ldots,a'_\tau}}{\sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,2}}\tilde{w}_{a_1,\ldots,a'_\tau}}\geqslant e^{\varepsilon}\right]\leqslant\delta. \tag{71}$$

Notice that $\sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,1}}\tilde{w}_{a'_1,\ldots,a'_\tau}$ is distributed as $1+\mathrm{Bin}(\rho n,2^{-\tau})$, since $q\in\mathcal{H}^\tau_{2B,1}$ with probability 1 (by definition of $Q\sim\mathcal{D}$), and each of the ρn trials in determining the counts $w_{a_1,\ldots,a_\tau}$ belongs to $\mathcal{H}_{2B,2}$ with probability $2^{-\tau}$. Similarly, $\sum_{a'_1,\ldots,a'_\tau\in\mathcal{H}_{2B,1}}\tilde{w}_{a'_1,\ldots,a'_\tau}$ is distributed as $\mathrm{Bin}(\rho n+1,2^{-\tau})$; notice in particular that q, which is distributed uniformly over $\mathcal{H}^\tau_{2B,1}$, is in $\mathcal{H}^\tau_{2B,2}$ with probability $2^{-\tau}$. By the multiplicative Chernoff bound, we have that, for $\eta\leqslant 1$, it is the case

$$\mathbb{P}_{W\sim\mathrm{Bin}(\rho n,1/n)}\left[|W-\rho|>\rho\eta\right]\leqslant\exp\left(\frac{-\eta^2\rho}{3}\right). \tag{72}$$

As long as we take $\rho=\frac{36\ln(1/\delta)}{\varepsilon^2}$, inequality (72) will be satisfied with $\eta=\varepsilon/6$, which in turn implies inequality (71) since

$$\frac{(1+\varepsilon/6)\rho+1}{\rho(1-\varepsilon/6)}\leqslant\frac{e^{\varepsilon/6}\rho+1}{e^{-\varepsilon/3}\rho}\leqslant\frac{e^{4\varepsilon/6}\rho}{e^{-\varepsilon/3}\rho}\leqslant e^{\varepsilon},$$

where the second inequality above uses $(\rho+1)/\rho\leqslant e^{\varepsilon/2}$ for our choice of ρ.

We have thus established inequality (68) for the case that $\tilde{x}_n,\tilde{x}'_n$ differ by 1 on two indices. For the general case, consider any neighboring datasets $X=(\tilde{x}_1,\ldots,\tilde{x}_n)$ and $X'=(\tilde{x}_1,\ldots,\tilde{x}_{n-1},\tilde{x}'_n)$; we can find a sequence of at most $k-1$ intermediate datasets $(\tilde{x}_1,\ldots,\tilde{x}_{n-1},\tilde{x}^{(\mu)}_n)$, $1\leqslant\mu\leqslant k-1$ such that $\tilde{x}^{(\mu)}_n$ and $\tilde{x}^{(\mu-1)}_n$ differ by 1 on two indices. Applying inequality (68) to each of the k neighboring pairs in this sequence, we see that for any $\mathcal{U}\subset\mathcal{Y}$,

$$\mathbb{P}[P(X)\in\mathcal{U}]\leqslant e^{k\varepsilon}\cdot\mathbb{P}[P(X')\in\mathcal{U}]+\delta\cdot(1+e^{\varepsilon}+\cdots+e^{(k-1)\varepsilon})\leqslant e^{k\varepsilon}\cdot\mathbb{P}[P(X')\in\mathcal{U}]+\delta\cdot\frac{2e^{k\varepsilon}}{\varepsilon},$$

where we have used $\varepsilon\leqslant 1$ in the final inequality above. □

We next prove the accuracy of Algorithm 1.

Lemma 4.5 (Accuracy of P^{Had})**.** *Fix $n,B\in\mathbb{N}$ with B a power of 2. Then with τ,ρ as in Lemma 4.4, the estimate $\hat{x}$ produced in A^{Had} in the course of the shuffled-model protocol $P^{\mathrm{Had}}=(R^{\mathrm{Had}},S,A^{\mathrm{Had}})$ with input $x_1,\ldots,x_n\in\{0,1\}^B$ satisfies*

$$\mathbb{P}\left[\left\|\hat{x}-\sum_{i=1}^n x_i\right\|_\infty\leqslant\sqrt{3\ln(2B/\beta)\cdot\max\{3\ln(2B/\beta),\rho+k\}}\right]\geqslant 1-\beta.$$

Proof. Fix any $j\in[B]$. Let $\zeta_j=\sum_{i=1}^n(x_i)_j$. We will upper bound the probability that $\xi_j:=\hat{x}_j-\zeta_j$ is large. Notice that the distribution of $\hat{x}_j$ is given by

$$\frac{1}{1-2^{-\tau}}\cdot\left(\zeta_j+\mathrm{Bin}(\rho n+kn-\zeta_j,2^{-\tau})-(\rho n+kn)2^{-\tau}\right).$$

This is because each of the ρn tuples $(\tilde{a}_{g,1},\ldots,\tilde{a}_{g,\rho})$ chosen uniformly from $[2B]^\tau$ on Line 8 of Algorithm 1 has probability $2^{-\tau}$ of belonging to $\mathcal{H}^\tau_{2B,j}$, and each of the $kn-\zeta_j$ tuples $(a_{j',1},\ldots,a_{j',\tau})$ (for $j'\in\mathcal{S}_i$, $j'\neq j$, $i\in[n]$) chosen uniformly from $\mathcal{H}^\tau_{2B,j'}$ in Line 6 of Algorithm 1 also has probability $2^{-\tau}$ of belonging to $\mathcal{H}^\tau_{2B,j}$. (Moreover, each of the ζ_j tuples $(a_{j,1},\ldots,a_{j,\tau})$ chosen in Line 6 of the local randomizer always belongs to $\mathcal{H}^\tau_{2B,j}$.)

Therefore, the distribution of ξ_j is given by

$$\frac{1}{1-2^{-\tau}}\cdot\left(\mathrm{Bin}(\rho n+kn-\zeta_j,2^{-\tau})-(\rho n+kn-\zeta_j)2^{-\tau}\right).$$

Using that $\tau = \log n$, we may rewrite the above as

$$\frac{n}{n-1} \cdot (\mathrm{Bin}((\rho + k - \zeta_j/n) \cdot n, 1/n) - (\rho + k - \zeta_j/n)).$$

For any reals $c > 0$ and $0 \leqslant \eta \leqslant 1$, by the Chernoff bound, we have

$$\mathbb{P}_{\xi \sim \mathrm{Bin}(cn,1/n)} [|z - c| > \eta c] \leqslant 2 \exp\left(\frac{-\eta^2 c}{3}\right).$$

Moreover, for $\eta > 1$, we have

$$\mathbb{P}_{\xi \sim \mathrm{Bin}(cn,1/n)} [|z - c| > \eta c] \leqslant 2 \exp\left(\frac{-\eta c}{3}\right).$$

We have $2\exp(-\eta^2 c/3) \leqslant \beta/B$ as long as $\eta \geqslant \sqrt{3\ln(2B/\beta)/c}$. Set $c = \rho + k - \zeta_j/n$, so that $\rho \leqslant c \leqslant \rho + k$. First suppose that $\sqrt{3\ln(2B/\beta)/(\rho + k - \zeta_j/n)} \leqslant 1$. Then we see that

$$\mathbb{P}[|\hat{x}_j - \zeta_j| > \sqrt{3\ln(2B/\beta) \cdot (\rho + k)}] \leqslant \beta/B. \tag{73}$$

In the other case, namely $\rho + k - \zeta_j/n = c < 3\ln(2B/\beta)$, set $\eta = 3\ln(2B/\beta)/c$, and we see that

$$\mathbb{P}[|\hat{x}_j - \zeta_j| > 3\ln(2B/\beta)] \leqslant \beta/B. \tag{74}$$

The combination of (73) and (74) with a union bound over all $j \in [B]$ completes the proof of Lemma 4.5. □

Next we summarize the communication and computation settings of the shuffled model protocol P^{Had}.

Lemma 4.6 (Efficiency of P^{Had})**.** *Let $n, B, \tau, \rho, k \in \mathbb{N}$. Then the protocol $P^{\mathrm{Had}} = (R^{\mathrm{Had}}(n, B, \tau, \rho, k), S, A^{\mathrm{Had}}(n, B, \tau, \rho, k))$ satisfies the following:*

1. *On input $(x_1, \ldots, x_n) \in \{0,1\}^B$, the output of the local randomizers $R^{\mathrm{Had}}(n, B, \tau, \rho, k)$ consists of $n(k + \rho)$ messages of length $\tau \log 2B$ bits.*
2. *The runtime of the analyzer $A^{\mathrm{Had}}(n, B, \tau, \rho, k)$ on input $\{y_1, \ldots, y_m\}$ is at most $O(Bm\tau)$ and its output has space $O(B\log(n(k+\rho)))$ bits. Moreover, if $\tau = \log n$ (i.e., as in Lemma 4.4), and if its input $\{y_1, \ldots, y_m\}$ is the output of the local randomizers on input $x_1, \ldots, x_n$ (so that $m = n(\rho + k)$), there is a modification of the implementation of A^{Had} in Algorithm 1 that, for $\beta \in [0,1]$, completes in time $O((\rho+k)n\log^3 B + B\rho \log B/\beta)$ with probability $1 - \beta$.*
3. *There is a separate modification of $A^{\mathrm{Had}}(n, B, \tau, \rho, k)$ that on input $\{y_1, \ldots, y_m\}$ produces an output data structure $(\mathrm{FO}, \mathcal{A})$ with space $O(m\tau \log B)$ bits, such that a single query $\mathcal{A}(\mathrm{FO}, j)$ of some $j \in [B]$ takes time $O(m\tau \log B)$.*

Proof. The first item is immediate from the definition of R^{Had} in Algorithm 1. For the second item, note first that A^{Had} as written in Algorithm 1 takes time $O(Bm\tau \log B)$: for each message $y_i = (a_{i,1}, \ldots, a_{i,\tau})$, it loops through each $j \in [B]$ to check if each $a_{i,g} \in \mathcal{H}_{2B,j}$ for $1 \leqslant g \leqslant \tau$ (determination of whether $a_{i,g} \in \mathcal{H}_{2B,j}$ takes time $O(\log B)$).

Now suppose that the messages $y_1, \ldots, y_m$ are the union of the multisets output by each of n shufflers on input $x_1, \ldots, x_n$. Notice that for each $j' \in [2B-1]$, the number of messages $(a_{j,1}, \ldots, a_{j,\tau}) \in \mathcal{H}_{2B,j}$ (Line 6 of Algorithm 1) such that $j \neq j'$ and also $(a_{j,1}, \ldots, a_{j,\tau}) \subset \mathcal{H}_{2B,j'}$ is distributed as $Bin(n', 1/n)$ for some $n' \leqslant n$ (recall $\tau = \log n$). Moreover, for each $j' \in [2B-1]$, the number of messages of the form $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau})$ (Line 8 of Algorithm 1) satisfying $(\tilde{a}_{g,1}, \ldots, \tilde{a}_{g,\tau}) \subset \mathcal{H}_{2B,j'}$ is distributed as $Bin(\rho n, 1/n)$. Therefore, by the multiplicative Chernoff bound and a union bound, for any $0 \leqslant \beta \leqslant 1$, the sum, over all $j' \in [B]$, of the number of messages $(a_1, \ldots, a_\tau)$ that belong to $\mathcal{H}_{2B,j'}$ is bounded above by

$$kn + \frac{4B(n + \rho n)(1 + \ln(B/\beta))}{n} = kn + 4B(1+\rho)(1 + \ln(B/\beta)), \tag{75}$$

with probability $1-\beta$. Next, consider any individual message $y_i = (a_{i,1}, \ldots, a_{i,\tau})$ (as on Line 16). Notice that the set of j such that $\{a_{i,1}, \ldots, a_{i,\tau}\} \subset \mathcal{H}_{2B,j}$ can be described as follows: write $j = (j_1, \ldots, j_{\log 2B}) \in \{0,1\}^{2B}$ to denote the binary representation of j, and arrange the $\log 2B$-bit binary representations of each of $a_{i,1}, \ldots, a_{i,\tau}$ to be the rows of a $\tau \times \log 2B$ matrix $A \in \{0,1\}^{\tau \times \log 2B}$. Then $\{a_{i,1}, \ldots, a_{i,\tau}\} \subset \mathcal{H}_{2B,j}$ if and only if $Aj = 0$, where arithmetic is performed over $\mathbb{F}_2$. This follows since for binary representations $i = (i_1, \ldots, i_{\log 2B}) \in \{0,1\}^{\log 2B}$ and $j = (j_1, \ldots, j_{\log 2B}) \in \{0,1\}^{\log 2B}$, the (i,j)-element of H_B is $(-1)^{\sum_{t=1}^{\log 2B} i_t j_t}$. Using Gaussian elimination one can enumerate the set of $j \in \{0,1\}^{\log 2B}$ in the kernel of A in time proportional to the sum of $O(\log^3 B)$ and the number of j in the kernel. Since the sum, over all messages y_i, of the number of such j is bounded above by (75) (with probability $1-\beta$), the total running time of this modification of A^{Had} becomes $O((\rho+k)n\log^3 B + kn + B\rho\log(B/\beta))$.

For the last item of the theorem, the analyzer simply outputs the collection of all tuples $y_i = (a_{i,1}, \ldots, a_{i,\tau})$; to query the frequency of some $j \in [B]$, we simply run the for loop on Line 14 of Algorithm 1, together with the debiasing step of Line 19. □

4.2 Public-Coin Protocol with Small Query Time

In this subsection, we give a public-coin protocol for frequency estimation in the shuffled model with error $\mathrm{poly}\log B$ and communication per user $\mathrm{poly}(\log B, \log n)$ bits. As discussed in Section 1.3, our protocol is based on combining the Count Min data structure [CM05a] with a multi-message version of randomized response [War65]. We start by giving a more detailed overview of the protocol.

Overview. On a high level, the presence of public randomness (which is assumed to be known to the analyzer) allows the parties to jointly sample random seeds for hash functions which they can use to compute and communicate (input-dependent) updates to a probabilistic data structure. The data structure that we will use is Count Min which we recall next. Assume that each of n users holds an input from $[B]$ where $n \ll B$. We hash the universe $[B]$ into s buckets where $s = O(n)$. [15] Then for each user, we increment the bucket to which its input hashes. This ensures that for every element of $[B]$, its hash bucket contains an overestimate of the number of users having that element as input. However, these bucket values are not enough to unambiguously recover the number of users holding any specific element of $[B]$—this is because on average, B/s different elements hash to the same bucket. To overcome this, the Count Min data structure repeats the above idea $\tau = O(\log B)$ times using independent hash functions. Doing so ensures that for each element $j \in [B]$, it is the case that (i) no other element $j' \in [B]$ hashes to the same buckets as j for all τ repetitions, and (ii) for at least one repetition, no element of $[B]$ that is held by a user (except possibly j itself) hashes to the same bucket as j. To make the Count Min data structure differentially private, we use a multi-message version of randomized response [War65]. Specifically, we ensure that sufficient independent noise is added to each bucket of each repetition of the Count Min data structure. This is done by letting each user independently, using its private randomness, increment every bucket with a very small probability. The noise stability of Count Min is used to ensure that the frequency estimates remain accurate after the multi-message noise addition. We further use the property that updates to this data structure can be performed using a logarithmic number of "increments" to entries in the sketch for two purposes: (i) to bound the privacy loss for a single user, and (ii) to obtain a communication-efficient implementation of the protocol. The full description appears in Algorithm 2.

Analysis. We next show the accuracy, efficiency, and privacy guarantees of Algorithm 2 which are summarized in the following theorem.

Theorem 4.7. *There is a sufficiently large positive absolute constant ζ such that the following holds. Suppose $n, B, k \in \mathbb{N}$, and $0 \leqslant \varepsilon, \delta, \beta \leqslant 1$. Consider the shuffled-model protocol $P^{\mathrm{CM}} = (R^{\mathrm{CM}}, S, A^{\mathrm{CM}})$ with $\tau =$*

[15] It is possible to introduce a trade-off here: By increasing s we can improve privacy and accuracy at the cost of requiring more communication. For simplicity we present our results for the case $s = O(n)$, minimizing communication, and discuss larger s at the end of section 4.2

Algorithm 2: Local randomizer, analyzer and query for frequency estimation via Count Min.

$R^{\mathrm{CM}}(n, B, \tau, \gamma, s)$**:**
- **Input:** Subset $\mathcal{S} \subset [B]$ specifying the user's input set
- **Parameters:** $n, B, \tau, s \in \mathbb{N}$ and $\gamma \in [0, 1]$
- **Public Randomness**: A random hash family $\{h_t : [B] \to [s],\ \forall t \in [\tau]\}$
- **Output:** A multiset $\mathcal{T} \subset [\tau] \times [s]$
- **for** $j \in \mathcal{S}$ **do**
  - **for** $t \in [\tau]$ **do**
    - Add the pair $(t, h_t(j))$ to $\mathcal{T}$.
- **for** $t \in [\tau]$ **do**
  - **for** $\ell \in [s]$ **do**
    - Sample $b_{t,\ell}$ from $\mathrm{Ber}(\gamma)$.
    - **if** $b_{t,\ell} = 1$ **then**
      - Add the pair (t, ℓ) to $\mathcal{T}$.
- **return** $\mathcal{T}$.

$A^{\mathrm{CM}}(n, B, \tau, s)$**:**
- **Input:** Multiset $\{y_1, \ldots, y_m\}$ containing outputs of local randomizers
- **Parameters:** $n, B, \tau, s \in \mathbb{N}$
- **Public Randomness**: A random hash family $\{h_t : [B] \to [s],\ \forall t \in [\tau]\}$
- **Output:** A noisy Count Min data structure $C : [\tau] \times [s] \to \mathbb{N}$
- **for** $t \in [\tau]$ **do**
  - **for** $\ell \in [s]$ **do**
    - $C[t, \ell] = 0$.
- **for** $j \in [m]$ **do**
  - $C[y_j] \leftarrow C[y_j] + 1$.
- **return** C

$Q^{\mathrm{CM}}(n, B, \tau, s)$**:**
- **Input:** Element $j \in [B]$
- **Parameters:** $n, B, \tau, s \in \mathbb{N}$
- **Public Randomness**: A random hash family $\{h_t : [B] \to [s],\ \forall t \in [\tau]\}$
- **Output:** A non-negative real number which is an estimate of the frequency of element j
- **return** $\hat{x}_j := \max\{\min\{C[t, h_t[j]] - \gamma n :\ t \in [\tau]\}, 0\}$

$\log(2B/\beta)$*,* $s = 2kn$*, and*

$$\gamma = \frac{1}{n} \cdot \zeta \cdot \max\left\{\log n, \frac{\log^2(B/\beta)k^2 \log(\log(B/\beta)k/\delta)}{\varepsilon^2}\right\}.$$

Then P^{CM} *is* (ε, δ)*-differentially private (Definition 2.2), each user sends* $O(\gamma kn \log(B/\beta))$ *messages consisting of* $O(\log n + \log\log B/\beta)$ *bits each with probability* $1 - \beta$*, and for inputs* $x_1, \ldots, x_n \in \{0, 1\}^B$ *(*$\|x_i\|_1 \leqslant k$*), the estimates* $\hat{x}_j$ *produced by* Q^{CM} *satisfy:*

$$\mathbb{P}\left[\forall j \in [B]\ :\ \left|\hat{x}_j - \sum_{i=1}^{n} x_{i,j}\right| \leqslant O\left(\sqrt{\log\left(\frac{Bn}{\beta}\right) \cdot \left(\log\left(\frac{Bn}{\beta}\right) + \frac{k^2 \log^2(B/\beta) \log((\log B/\beta)k/\delta)}{\varepsilon^2}\right)}\right)\right] \geqslant 1 - \beta. \tag{76}$$

Moreover, any frequency query can be answered in time $O(\log B/\beta)$.

Notice that by decreasing β by at most a constant factor (and thus increasing the error bounds by at most a constant factor), we may ensure that $\tau = \log(2B/\beta)$ in the theorem statement is an integer. Note also that the additive error in (76) is $\tilde{O}(k/\varepsilon)$, where the $\tilde{O}(\cdot)$ hides factors logarithmic in $B, n, k, 1/\delta, 1/\beta$.

Theorem 4.7 with $k = 1$ directly implies Theorem 4.2. The next lemma is used to prove the accuracy of Algorithm 2.

Lemma 4.8 (Accuracy of P^{CM})**.** *Let n, B, and τ be positive integers, and $\gamma \in [0,1]$, $\xi \in [0, \sqrt{\gamma n}]$ be real parameters. Then the estimate $\hat{x}_j$ produced by Q^{CM} on input $j \in [B]$ and as an outcome of the shuffled-model protocol $P^{\mathrm{CM}} = (R^{\mathrm{CM}}, S, A^{\mathrm{CM}})$ with input $x_1, \ldots, x_n \in \{0,1\}^B$ ($\|x_i\|_1 \leqslant k$) satisfies $\hat{x}_j \geqslant \sum_{i=1}^n x_{i,j}$ and*

$$\mathbb{P}\left[\left|\hat{x}_j - \sum_{i=1}^n x_{i,j}\right| \leqslant \xi\sqrt{\gamma n}\right] \geqslant 1 - (kn/s)^\tau - 2^{\log(2s\tau) - \xi^2/3}.$$

Proof. We consider the entries $\{C[t, h_t[j]] \mid t \in [\tau]\}$ of the noisy Count Min data structure. We first consider the error due to the other inputs that are held by the users. Then we consider the error due to the noise blanket. We bound each of these two errors with high probability and then apply a union bound.

First, note that for any element $j \in [B]$, the probability that for every repetition index $t \in [\tau]$, some element $j' \in [B]$ held by one of the users (except possibly j itself) satisfies $h_t(j') = h_t(j)$, is at most $(kn/s)^\tau$. As in the original analysis of Count Min [CM05a], this holds even if the hash functions h_t are sampled from a family of pairwise independent hash functions.

It remains to show that with probability at least $1 - 2^{1+\log(s\tau) - \xi^2/3}$, the absolute value of the deviation of the blanket noise in each of these entries from its expectation $\gamma \cdot n$ is at most $O\left(\sqrt{\gamma n}\right)$. By a union bound over all $s\tau$ pairs of bucket indices and repetition indices, it is enough to show that for each $t \in [\tau]$ and each $\ell \in [s]$, with probability at least $1 - 2^{1-\xi^2/3}$, the absolute value of the blanket noise in $C[t, h_t[j]]$ is at most $\xi\sqrt{\gamma n}$. This follows from the fact that the blanket noise in the entry $C[t, h_t[j]]$ is the sum of n independent $\mathrm{Ber}(\gamma)$ random variables (one contributed by each user). The bound now follows from the multiplicative Chernoff bound.

Finally, by a union bound the overall error is at most $O(\gamma n)$ with probability at least $1 - (kn/s)^\tau - 2^{\Theta(\log(s\tau) - \gamma n)}$. □

By removing the subtraction of γn on the final line of Algorithm 2, we can guarantee that the estimate returned by the Count Min sketch is never less than the true count of an element. This would lead to, however, an expected error of $O(\gamma n)$ as opposed to $O(\sqrt{\gamma n})$ in Lemma 4.8. The next lemma shows the efficiency of Algorithm 2.

Lemma 4.9 (Efficiency of P^{CM})**.** *Let n, B, τ, s be positive integers and $\gamma \in [0,1]$. Then,*

1. *With probability at least $1 - n \cdot 2^{-\Theta(\gamma s\tau)}$, the output of $R^{\mathrm{CM}}(n, B, \tau, \gamma, s)$ on input $\mathcal{S}$ consists of at most $|\mathcal{S}| + O(\gamma s\tau)$ messages each consisting of $\lceil \log_2(\tau) \rceil + \lceil \log_2(s) \rceil$ bits.*
2. *The runtime of the analyzer $A^{\mathrm{CM}}(n, B, \tau, s)$ on input $\{y_1, \ldots, y_m\}$ is $O(\tau s + m)$ and the space of the data structure that it outputs is $O(\tau s \log m)$ bits.*
3. *The runtime of any query $Q^{\mathrm{CM}}(n, B, \tau, s)$ is $O(\tau)$.*

Proof. The second and third parts follow immediately from the operation of Algorithm 2. To prove the first part, note that each user sends $|\mathcal{S}|$ messages corresponding to its inputs along with a number of "blanket noise" terms. This number is a random variable drawn from the binomial distribution $\mathrm{Bin}(\tau s, \gamma)$. Moreover, each of these messages is a pair consisting of a repetition index (belonging to $[\tau]$) and a bucket index (belonging to $[s]$). The proof now follows from the multiplicative Chernoff bound along with a union bound over all n users. □

The next lemma establishes the privacy of Algorithm 2.

Lemma 4.10 (Privacy of P^{CM})**.** *Let n and B be positive integers. Then, for $\gamma n \geqslant \frac{90k^2\tau^2 \ln(2\tau k/\delta)}{\varepsilon^2}$, the algorithm $S \circ R^{\mathrm{CM}}(n, B, \tau, \gamma, s)$ is (ε, δ)-differentially private.*

To prove Lemma 4.10, we need some general tools linking sensitivity of vector-valued functions, smoothness of distributions and approximate differential privacy—these are given next in Section 4.3. The proof of Lemma 4.10 is deferred to Section 4.4. We are now ready to prove Theorem 4.7.

Proof of Theorem 4.7. Privacy is an immediate consequence of Lemma 4.10. To establish accuracy (i.e., (76)), note first that Lemma 4.8 guarantees that for any $j \in [B]$ and any $\xi \in [0, \sqrt{\gamma n}]$, $|\hat{x}_j - \sum_{i=1}^n x_{i,j}| \leqslant \xi\sqrt{\gamma n}$ with probability at least $1 - (kn/s)^\tau - 2^{\log(2s\tau)-\xi^2/3}$. We now choose $\xi = \sqrt{3 \cdot \log\left(\frac{4Bs\tau}{\beta}\right)}$; this ensures that $2^{\log(2s\tau)-\xi^2/3} \leqslant \beta/(2B)$. Moreover, we have that $\xi \leqslant \sqrt{\gamma n}$ by our choice of γ in the theorem statement.

It now follows from a union bound over all $j \in [B]$ that

$$\mathbb{P}\left[\forall j \in [B] \ : \ \left|\hat{x}_j - \sum_{i=1}^n x_{i,j}\right| \leqslant O\left(\sqrt{\log\left(\frac{Bn}{\beta}\right) \cdot \left(\log\left(\frac{Bn}{\beta}\right) + \frac{k^2 \log^2(B/\beta)\log(\log(B/\beta)k/\delta)}{\varepsilon^2}\right)}\right)\right] \geqslant 1-\beta. \tag{77}$$

Here we have used that $k \leqslant B$. □

Improving error and privacy by increasing communication. Theorem 4.7 bounds the error of Algorithm 2 with parameters $s = O(n)$ and $\tau = O(\log B/\beta)$. For constant $\eta > 0$ it is interesting to consider the parameterization $s = O(n(n/\beta)^\eta)$ and $\tau = O(1/\eta)$. By Lemma 4.10 differential privacy can be ensured in this setting with $\gamma n = O(k^2 \log(k/\delta)/\varepsilon^2)$. The randomizer of Algorithm 2 sends a number of blanket messages that is $O(\gamma s)$ in expectation, i.e., $O((n/\beta)^\eta k^2 \log(k/\delta)/\varepsilon^2)$. An argument mirroring the proof of Lemma 4.8 shows that the *pointwise* error of an estimate $\hat{x}_j$ is bounded by $\sqrt{\gamma n \log(1/\beta)} = O(k\sqrt{\log(k/\delta)\log(1/\beta)}/\varepsilon)$ with probability $1 - \beta$. Thus, error as well as communication is independent of the domain size B. To get a bound that is directly comparable to Theorem 4.7, holding for all queries in $[B]$, we may reduce the pointwise error probability β by a factor B and apply a union bound, resulting in communication $O((Bn/\beta)^\eta k^2 \log(k/\delta)/\varepsilon^2)$ and error $O(k\sqrt{\log(k/\delta)\log(B/\beta)}/\varepsilon)$. Query time is $\tau = O(1)$. This strictly improves the results that follow from [CSU+19, BBGN19a, GPV19] (see Table 1). Very recently, Balcer and Cheu [BC19] showed a different trade-off in the case where the number of messages is very large: $B + 1$ messages of size $O(\log B)$ each with error $O(\log(1/\delta)/\varepsilon^2 + \sqrt{\log(1/\delta)\log(n/\beta)}/\varepsilon)$, which is independent of B.

4.3 Useful Tools

If the blanket noise added to each bucket of the Count Min sketch were distributed as i.i.d. Gaussian or Laplacian random variables, the proof of Lemma 4.10 would follow immediately from known results. Due to the discrete and distributed nature of the problem, we are forced to instead use Binomial blanket noise. To prove Lemma 4.10, we will need some general tools linking approximate differential privacy to smoothness of distributions (and in particular the Binomial distribution); these tools are essentially known, but due to the lack of a suitable reference we prove all the prerequisite results.

Definition 4.1 (Sensitivity)**.** The ℓ_1-*sensitivity* (or *sensitivity*, for short) of $f : \mathcal{X}^n \to \mathbb{Z}^m$ is given by:

$$\Delta(f) = \max_{X \sim X'} \|f(X) - f(X')\|_1.$$

It is well-known [DMNS06] that the mechanism given by adding independent Laplacian noise with variance $2\Delta(f)^2/\varepsilon^2$ to each coordinate of $f(X)$ is $(\varepsilon, 0)$-differentially private. Laplace noise, however, is unbounded in both the positive and negative directions, and this causes issues in the shuffled model (roughly speaking, it would require each party to send infinitely many messages). In our setting we will need to ensure that the noise added to each coordinate is bounded, so to achieve differential privacy we will not be able to add Laplacian noise. As a result we will only be able to obtain (ε, δ)-differential privacy for $\delta > 0$. We specify next the types of noise that we will use instead of Laplacian noise.

Definition 4.2 (Smooth distributions)**.** Suppose $\mathcal{D}$ is a distribution supported on $\mathbb{Z}$. For $k \in \mathbb{N}$, $\varepsilon \geqslant 0$ and $\delta \in [0, 1]$, we say that $\mathcal{D}$ is (ε, δ, k)-*smooth* if for all $-k \leqslant k' \leqslant k$,

$$\mathbb{P}_{Y \sim \mathcal{D}}\left[\frac{\mathbb{P}_{Y' \sim \mathcal{D}}[Y' = Y]}{\mathbb{P}_{Y' \sim \mathcal{D}}[Y' = Y + k']} \geqslant e^{|k'|\varepsilon}\right] \leqslant \delta.$$

Definition 4.3 (Incremental functions)**.** Suppose $k \in \mathbb{N}$. We define $f : \mathcal{X}^n \to \mathbb{Z}^m$ to be *k-incremental* if for all neighboring datasets $X \sim X'$, $\|f(X) - f(X')\|_\infty \leqslant k$.

The following lemma formalizes the types of noise we can add to $f(X)$ to obtain such a privacy guarantee. Its proof appears in Appendix C.

Lemma 4.11. *Suppose $f : \mathcal{X}^n \to \mathbb{Z}^m$ is k-incremental (Definition 4.3) and $\Delta(f) = \Delta$. Suppose $\mathcal{D}$ is a distribution supported on $\mathbb{Z}$ that is (ε, δ, k)-smooth. Then the mechanism*

$$X \mapsto f(X) + (Y_1, \ldots, Y_m),$$

where $Y_1, \ldots, Y_m \sim \mathcal{D}$, i.i.d., is (ε', δ')-differentially private, where $\varepsilon' = \varepsilon \cdot \Delta, \delta' = \delta \cdot \Delta$.

In order to prove Lemma 4.10, we will also use the following statement about the smoothness of the binomial distribution (that we will invoke with a small value of the head probability γ). Its proof appears in Appendix C.

Lemma 4.12 (Smoothness of $\mathrm{Bin}(n, \gamma)$)**.** *Let $n \in \mathbb{N}$, $\gamma \in [0, 1/2]$, $0 \leqslant \alpha \leqslant 1$, and $k \leqslant \alpha\gamma n/2$. Then the distribution $\mathrm{Bin}(n, \gamma)$ is (ε, δ, k)-smooth with $\varepsilon = \ln((1+\alpha)/(1-\alpha))$ and $\delta = e^{-\frac{\alpha^2\gamma n}{8}} + e^{-\frac{\alpha^2\gamma n}{8+2\alpha}}$.*

4.4 Privacy Proof

We are now ready to prove Lemma 4.10 using the results on k-incremental functions from the previous section, thereby establishing the privacy of Algorithm 2. An alternative approach to establishing privacy of Algorithm 2 is to first do so for the case $k = 1$ and then apply the advanced composition lemma [DR14a]. However, doing so leads to an error bound that incurs at least an additional $\sqrt{k}$ factor since one has to make ε smaller by a factor of $\sqrt{k}$. In order to prove Lemma 4.10, we could use Theorem 1 of [ASY$^+$18] instead of our Lemma 4.12 but their result would give worse bounds for $k > 1$.

Proof of Lemma 4.10. Fix ε, δ. Notice that $S \circ R^{\mathrm{CM}}(n, B, \tau, \gamma, s)$ can be obtained as a post-processing of the noisy Count Min data structure $C : [\tau] \times [s] \to \mathbb{N}$ in Algorithm 2, so it suffices to show that the algorithm bringing the players' inputs to this Count Min data structure is (ε, δ)-differentially private. Consider first the Count Min data structure $\tilde{C} : [\tau] \times [s] \to \mathbb{N}$ with no noise, so that $\tilde{C}[t, \ell]$ measures the number of inputs x inside some user's set $\mathcal{S}_i$ such that $h_t(x) = \ell$. We next note that the function mapping the users' inputs $(\mathcal{S}_1, \ldots, \mathcal{S}_n)$ to $\tilde{C}$ has sensitivity (in terms of Definition 4.1) at most $k\tau$ and is k-incremental (in terms of Definition 4.3). Moreover, Lemma 4.12 (with $\alpha = \varepsilon/(3\tau k)$) implies that the binomial distribution $\mathrm{Bin}(n, \gamma)$ is $(\varepsilon/(\tau k), \delta/(\tau k), k)$-smooth (in terms of Definition 4.2) as long as $\delta \geqslant 2\tau k e^{-\frac{\varepsilon^2\gamma n}{90\tau^2 k^2}}$ and $k \leqslant \varepsilon\gamma n/(6\tau k)$. In particular, we need

$$\gamma n \geqslant \frac{90\tau^2 k^2 \ln(2\tau k/\delta)}{\varepsilon^2}.$$

By construction in Algorithm 2, $C[t, s] = \tilde{C}[t, s] + \mathrm{Bin}(n, \gamma)$, where the binomial random variables are independent for each t, s. Applying Lemma 4.11, we get that the Count Min data structure is (ε, δ)-differentially private (with respect to Definition 2.2). □

5 Multi-Message Protocols for Range Counting Queries

We recall the definition of range queries. Let $\mathcal{X} = [B]$ and consider a dataset $X = (x_1, \ldots, x_n) \in [B]^n$. Notice that a statistical query may be specified by a vector $w \in \mathbb{R}^B$, and the answer to this statistical query on the dataset X is given by $\langle w, \mathrm{hist}(X)\rangle$. For all queries w we consider, we will in fact have $w \in \{0, 1\}^B$, and thus w specifies a *counting query*. Here $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product; throughout the paper, we slightly abuse notation and allow an inner product to be taken of a row vector and a column vector. A *1-dimensional range query* $[j, j']$, where $1 \leqslant j \leqslant j' \leqslant B$, is a counting query such that $w_j = w_{j+1} = \cdots = w_{j'} = 1$, and all other entries of w are

Problem	Messages per user	Message size in bits	Error	Query time
d-dimensional range counting (public) Theorem 5.10	$\frac{\log^{3d+3}(B)\log\frac{1}{\delta}}{\varepsilon^2}$	$\log n + \log\log B$	$\frac{\log^{2d+3/2}(B)\log\frac{1}{\delta}}{\varepsilon}$	$\log^{d+1} B$
d-dimensional range counting (private) Theorem 5.9	$\frac{\log^{2d}(B)\log\frac{1}{\varepsilon\delta}}{\varepsilon^2}$	$\log(n)\log B$	$\frac{\log^{2d+1/2}(B)\log\frac{1}{\varepsilon\delta}}{\varepsilon}$	$\frac{n\log^{3d+2}(B)\log\frac{1}{\varepsilon\delta}}{\varepsilon^2}$

Table 3: Overview of results on differentially private range counting in the shuffled model. The query time stated is the additional time to answer a query, assuming a preprocessing of the output of the shuffler that takes time linear in its length. Note that frequencies and counts are not normalized, i.e., they are integers in $\{0,\ldots,n\}$. For simplicity, constant factors are suppressed, the bounds are stated for error probability $\beta = B^{-O(1)}$, and the following are assumed: dimension d is a constant, n is bounded above by B, and $\delta < 1/\log B$.

0. For *d-dimensional range queries*, the elements of $[B]$ will map to points on a d-dimensional grid, and a certain subset of vectors $w \in \{0,1\}^B$ represent the d-dimensional range queries. In this section, we use the frequency oracle protocols in Section 4 to derive protocols for computing counting queries with per-user communication $\mathrm{poly}\log(B)$ and additive error $\mathrm{poly}\log(\max\{n, B\})$.

In Section 5.2, we adapt the matrix mechanism of [LHR+10, LM12] to use the frequency oracle protocols of Section 4 as a black-box for computation of counting queries, which include range queries as a special case. In Section 5.3, we instantiate this technique for the special case of 1-dimensional range queries, and in Section 5.4 we consider the case of multi-dimensional range queries. In Section 5.5 we collect the results from Sections 5.2 through 5.4 to formally state our guarantees on range query computation in the shuffled model, as well as the application to M-estimation of the median, as mentioned in the Introduction.

5.1 Frequency Oracle

We now describe a basic data primitive that encapsulates the results in Section 4 and that we will use extensively in this section. Fix positive integers B and $k \leqslant B$ as well as positive real numbers κ and β. For each $v \in [B]$, let $e_v \in \{0,1\}^B$ be the unit vector with $(e_v)_j = 1$ if $j = v$, else $(e_v)_j = 0$. In the *(κ,β,k)-frequency oracle* problem [HKR12, BS15], each user $i \in [n]$ holds a subset $\mathcal{S}_i \subset [B]$ of size at most k. Equivalently, user i holds the sum of the unit vectors e_v corresponding to the elements v of $\mathcal{S}_i$, i.e., the vector $x_i \in \{0,1\}^B$ such that $(x_i)_j = 1$ if and only if $j \in \mathcal{S}_i$. Note that $\|x_i\|_1 \leqslant k$ for all i. At times we will restrict ourselves to the case that $k = 1$; in such cases we will often use x_i to denote the single element $x_i \in [B]$ held by user i, and write $e_{x_i} \in \{0,1\}^B$ for the corresponding unit vector.

The goal is to design a (possibly randomized) data structure FO and a deterministic algorithm $\mathcal{A}$ (*frequency oracle*) that takes as input the data structure FO and an index $j \in [B]$, and outputs in time T an estimate that, with high probability, is within an additive κ from $\sum_{i=1}^n (x_i)_j$. Formally:

Definition 5.1 ((κ,β,k)-frequency oracle)**.** A protocol with inputs $x_1,\ldots,x_n \in \{0,1\}^B$ computes an (κ,β,k)-*frequency oracle* if it outputs a pair $(\mathrm{FO},\mathcal{A})$ such that for all datasets $(x_1,\ldots,x_n)$ with $\|x_i\|_1 \leqslant k$ for $i \in [n]$,

$$\mathbb{P}\left[\forall j \in [B] : \left|\mathcal{A}(\mathrm{FO}, j) - \sum_{i=1}^n (x_i)_j\right| \leqslant \kappa\right] \geqslant 1-\beta.$$

The probability in the above expression is over the randomness in creating the data structure FO.

Note that given such a frequency oracle, one can recover the (2κ)-*heavy hitters*, namely those j such that $\sum_{i=1}^n (x_i)_j \geqslant 2\kappa$, in time $O(T \cdot B)$, by querying $\mathcal{A}(\mathrm{FO},1),\ldots,\mathcal{A}(\mathrm{FO},B)$ (for a more efficient reduction see Appendix D).

5.2 Reduction to Private Frequency Oracle via the Matrix Mechanism

Our protocol for computing range queries is a special case of a more general protocol, which is in turn inspired by the *matrix mechanism* of [LHR+10, LM12]. We begin by introducing this more general protocol and explaining how it allows us to reduce the problem of computing range queries in the shuffled model to that of computing a frequency oracle in the shuffled model.

Finally, for a matrix $M \in \mathbb{R}^B \times \mathbb{R}^B$, define the *sensitivity* of M as follows:

Definition 5.2 (Matrix sensitivity, [LHR+10])**.** For a matrix M, let the *sensitivity* of M, denoted Δ_M, be the maximum ℓ_1 norm of a column of M.

For any column vector $y \in \mathbb{R}^B$, Δ_M measures the maximum ℓ_1 change in My if a single element of y changes by 1. The matrix mechanism, introduced by Li et al. [LHR+10, LM12] in the central model of DP, allows one to release answers to a given set of counting queries in a private manner. It is parametrized by an invertible matrix M, and given input X, releases the following noisy perturbation of $\mathrm{hist}(X)$:

$$\mathrm{hist}(X) + \Delta_M \cdot M^{-1} z, \tag{78}$$

where $z \in \mathbb{R}^B$ is a random vector whose components are distributed i.i.d. according to some distribution calibrated to the privacy parameters ε, δ. The response to a counting query $w \in \mathbb{R}^B$ is then given by $\langle w, \mathrm{hist}(X) + \Delta_M \cdot M^{-1} z \rangle$. The intuition behind the privacy of (78) is as follows: (78) can be obtained as a post-processing of the mechanism $X \mapsto M(\mathrm{hist}(X)) + \Delta_M \cdot z$, namely via multiplication by M^{-1}. If we choose, for instance, each z_i to be an independent Laplacian of variance $2/\varepsilon$, then the algorithm $X \mapsto M(\mathrm{hist}(X)) + \Delta_M \cdot z$ is simply the Laplace mechanism, which is $(\varepsilon, 0)$-differentially private [DMNS06].

In our modification of the matrix mechanism, the parties will send data that allows the analyzer to directly compute the "pre-processed input" $M(\mathrm{hist}(X)) + \Delta_M \cdot z$. Moreover, due to limitations of the shuffled model and to reduce communication, the distribution of the noise z will be different from what has been previously used [LHR+10, LM12]. For our application, we will require M to satisfy the following properties:

(1) For any counting query w corresponding to a d-dimensional range query, wM^{-1} has at most $\mathrm{poly}\log(B)$ nonzero entries, and all of those nonzero entries are bounded in absolute value by some $c > 0$. (Here $w \in \{0,1\}^B$ is viewed as a row vector.)

(2) $\Delta_M \leqslant \mathrm{poly}\log(B)$.

By property (2) above and the fact that all entries of M are in $\{0,1\}$, (approximate) computation of the vector $M(\mathrm{hist}(X))$ can be viewed as an instance of the frequency oracle problem where user $i \in [n]$ holds the $\leqslant \mathrm{poly}\log(B)$ nonzero entries of the vector $M(\mathrm{hist}(x_i))$. This follows since $M(\mathrm{hist}(x_i))$ is the x_ith column of M, $\Delta_M \leqslant \mathrm{poly}\log(B)$, and $\mathrm{hist}(X) = \sum_{i=1}^n \mathrm{hist}(x_i)$. Moreover, suppose there is some choice of local randomizer and analyzer (such as those in Section 4) that approximately solve the frequency oracle problem, i.e., compute an approximation $\hat{y}$ of $M(\mathrm{hist}(X))$ up to an additive error of $\mathrm{poly}\log B$, in a differentially private manner. Since wM^{-1} has at most $\mathrm{poly}\log(B)$ nonzero entries, each of magnitude at most c, it follows that

$$\langle wM^{-1}, \hat{y} \rangle \tag{79}$$

approximates the counting query $\langle w, \mathrm{hist}(X) \rangle$ up to an additive error of $c \cdot \mathrm{poly}\log(B)$.

Algorithm 3: Local randomizer for matrix mechanism

$R^{\mathrm{matrix}}(n, B, M, R^{\mathrm{FO}})$**:**

  Input: $x \in [B]$, parameters $n, B \in \mathbb{N}$, $M \in \{0,1\}^{B \times B}$, $R^{\mathrm{FO}} : \{0,1\}^B \to \mathcal{T}^*$

  Output: Multiset $\mathcal{S} \subset \mathcal{T}$, where $\mathcal{T}$ is the output set of R^{FO}

  Let $\mathcal{A}_x \leftarrow \{j \in [B] : M_{jx} \neq 0\}$

  `// A_x is the set of nonzero entries of the xth column of M`

  return $R^{\mathrm{FO}}(\mathcal{A}_x)$

Perhaps surprisingly, for any constant $d \geqslant 1$, we will be able to find a matrix M that satisfies properties (1) and (2) above for d-dimensional range queries with $c = 1$. This leads to the claimed $\mathrm{poly}\log(B)$ error for computation of d-dimensional range queries, as follows: the local randomizer R^{matrix} (Algorithm 3) is parametrized by integers $n, B \in \mathbb{N}$, a matrix $M \in \{0,1\}^{B \times B}$, and a local randomizer $R^{\mathrm{FO}} : [B] \to \mathcal{T}^*$ that can be used in a shuffled model protocol that computes a frequency oracle. (Here $\mathcal{T}$ is an arbitrary set, and R^{FO} computes a sequence of messages in $\mathcal{T}$.) Given input $x \in [B]$, R^{matrix} returns the output of R^{FO} when given as input the set of nonzero entries of the xth column of M. The corresponding analyzer A^{matrix} (Algorithm 4) is parametrized by integers $n, B \in \mathbb{N}$, a matrix $M \in \{0,1\}^{B \times B}$, and an analyzer A^{FO} for computation of a frequency oracle in the shuffled model. Given a multiset $\mathcal{S}$ consisting of the shuffled messages output by individual randomizers R^{matrix}, it returns (79), namely the inner product of wM^{-1} and the output of A^{FO} when given $\mathcal{S}$ as input. To complete the construction of a protocol

Algorithm 4: Analyzer for matrix mechanism

$A^{\mathrm{matrix}}(n, B, M, A^{\mathrm{FO}})$**:**

Input: Multiset $\mathcal{S} \subset [B]$ consisting of the shuffled reports;
Parameters $\mathcal{W} \subset \{0,1\}^B$ specifying a set of counting queries, $n, B \in \mathbb{N}, M \in \{0,1\}^{B \times B}$, analyzer A^{FO} for frequency oracle computation

Output: Map associating each $w \in \mathcal{W}$ to $f_w \in [0,1]$, specifying an estimate for each counting query w

Let $(\mathrm{FO}, \mathcal{A}) \leftarrow A^{\mathrm{FO}}(\mathcal{S})$

`// Frequency oracle output by` A^{FO} `(see Definition 5.1)`

return *Map associating each* $w \in \mathcal{W}$ *to* $f_w := \sum_{j \in [B] : (wM^{-1})_j \neq 0} (wM^{-1})_j \cdot \mathcal{A}(\mathrm{FO}, j)$

`// Let` $\hat{y} \in \mathbb{R}^B$ `be such that` $\hat{y}_j = \mathcal{A}(\mathrm{FO}, j)$`; then this returns the map associating` $w \in \mathcal{W}$ `to` $\langle wM^{-1}, \hat{y} \rangle$`.`

for range query computation in the shuffled model, it remains to find a matrix M satisfying properties (1) and (2) above. We will do so in Sections 5.3 and 5.4. First we state here the privacy and accuracy guarantees of the shuffled protocol $P^{\mathrm{matrix}} = (R^{\mathrm{matrix}}, S, A^{\mathrm{matrix}})$.

Theorem 5.1 (Privacy of P^{matrix})**.** *Suppose R^{FO} is a local randomizer for computation of an (κ, β, k)-frequency oracle with n users and universe size B, which satisfies (ε, δ)-differential privacy in the shuffled model. Suppose $M \in \{0,1\}^B$ satisfies $\Delta_M \leqslant k$. Then the shuffled protocol $S \circ R^{\mathrm{matrix}}(n, B, M, R^{\mathrm{FO}})$ is (ε, δ)-differentially private.*

Proof. Let $\mathcal{Y}$ be the message space of the randomizer R^{FO}, and $\mathcal{Y}'$ be the set of multisets consisting of elements of $\mathcal{Y}$. Let $P = S \circ R^{\mathrm{matrix}}(n, B, M, R^{\mathrm{FO}})$. Consider neighboring datasets $X = (x_1, \ldots, x_n) \in [B]^n$ and $X' = (x_1, \ldots, x_{n-1}, x_n') \in [B]^n$. We wish to show that for any $\mathcal{T} \subset \mathcal{Y}$,

$$\mathbb{P}[P(X) \in \mathcal{T}] \leqslant e^{\varepsilon} \cdot \mathbb{P}[P(X') \in \mathcal{T}] + \delta. \tag{80}$$

For $i \in [n]$, let $\mathcal{S}_i = \{j \in [B] : M_{j,x_i} \neq 0\}$ and $\mathcal{S}_n' = \{j \in [B] : M_{j,x_n'} \neq 0\}$. Since $\Delta_M \leqslant k$, we have $|\mathcal{S}_i| \leqslant k$ for $i \in [n]$ and $|\mathcal{S}_n'| \leqslant k$. Since the output of R^{matrix} on input x_i is simply $R^{\mathrm{FO}}(\mathcal{S}_i)$,

$$P(X) = S(R^{\mathrm{FO}}(\mathcal{S}_1), \ldots, R^{\mathrm{FO}}(\mathcal{S}_n)), \quad P(X') = S(R^{\mathrm{FO}}(\mathcal{S}_1), \ldots, R^{\mathrm{FO}}(\mathcal{S}_{n-1}), R^{\mathrm{FO}}(\mathcal{S}_n')).$$

Then (80) follows by the fact that $(\mathcal{S}_1, \ldots, \mathcal{S}_n)$ and $(\mathcal{S}_1, \ldots, \mathcal{S}_{n-1}, \mathcal{S}_n')$ are neighboring datasets for the (κ, β, k)-frequency problem and $S \circ R^{\mathrm{FO}}$ is (ε, δ)-differentially private. □

Theorem 5.2 (Accuracy & efficiency of P^{matrix})**.** *Suppose $R^{\mathrm{FO}}, A^{\mathrm{FO}}$ are the local randomizer and analyzer for computation of an (κ, β, k)-frequency oracle with n users and universe size B. Suppose also that $\mathcal{W} \subset \{0,1\}^B$ is a set of counting queries and $M \in \{0,1\}^B$ is such that, for any $w \in \mathcal{W}$, $\|wM^{-1}\|_1 \leqslant a$ and $\Delta_M \leqslant k$. Consider the shuffled model protocol $P^{\mathrm{matrix}} = (R^{\mathrm{matrix}}(n, B, M, R^{\mathrm{FO}}), S, A^{\mathrm{matrix}}(n, B, M, A^{\mathrm{FO}}, \mathcal{W}))$. For any dataset $X = (x_1, \ldots, x_n)$, let the (random) estimates produced by the protocol P^{matrix} on input X be denoted by $f_w \in [0,1]$ $(w \in \mathcal{W})$. Then:*

$$\mathbb{P}\left[\forall w \in \mathcal{W} : |f_w - \langle w, \mathrm{hist}(X) \rangle| \leqslant \kappa \cdot a\right] \geqslant 1 - \beta. \tag{81}$$

Moreover, if the set of nonzero entries of wM^{-1} and their values can be computed in time T, and A^{FO} releases a frequency oracle $(\mathrm{FO}, \mathcal{A})$ which takes time T' to query an index j, then for any $w \in \mathcal{W}$, the estimate f_w can be computed in time $O(T + a \cdot T')$ by A^{matrix}.

Proof. For $i \in [n]$, let $\mathcal{S}_i = \{j \in [B] : M_{j,x_i} \neq 0\}$ be the set of nonzero entries of the x_ith column of M. Denote by $(\mathrm{FO}, \mathcal{A})$ the frequency oracle comprising the output $A^{\mathrm{FO}}(S(R^{\mathrm{FO}}(\mathcal{S}_1), \dots, R^{\mathrm{FO}}(\mathcal{S}_n)))$. Define $\hat{y} \in \mathbb{R}^B$ by $\hat{y}_j = \mathcal{A}(\mathrm{FO}, j)$, for $j \in [B]$. Then the output of P^{matrix}, namely

$$P^{\mathrm{matrix}}(X) = A^{\mathrm{matrix}}(S(R^{\mathrm{matrix}}(x_1), \dots, R^{\mathrm{matrix}}(x_n))),$$

is given by the map associating each $w \in \mathcal{W}$ to $\langle wM^{-1}, \hat{y} \rangle$ (Algorithms 3 and 4).

Since $(\mathrm{FO}, \mathcal{A})$ is an (κ, β, k)-frequency oracle, we have that

$$\mathbb{P}\left[\|\hat{y} - \mathrm{hist}(\mathcal{S}_1, \dots, \mathcal{S}_n)\|_\infty \leqslant \kappa\right] \geqslant 1 - \beta.$$

Notice that the histogram of $\mathcal{S}_i$ is given by the x_ith column of M, which is equal to $M\mathrm{hist}(x_i)$. Thus $\mathrm{hist}(\mathcal{S}_1, \dots, \mathcal{S}_n) = M\mathrm{hist}(x_1, \dots, x_n)$. By Hölder's inequality, it follows that with probability $1 - \beta$, for all $w \in \mathcal{W}$,

$$\left|\langle wM^{-1}, \hat{y} \rangle - \langle wM^{-1}, M\mathrm{hist}(x_1, \dots, x_n) \rangle\right| \leqslant \kappa \cdot \|wM^{-1}\|_1 \leqslant \kappa \cdot a.$$

But $\langle wM^{-1}, M\mathrm{hist}(x_1, \dots, x_n) \rangle = wM^{-1}M\mathrm{hist}(x_1, \dots, x_n) = \langle w, \mathrm{hist}(x_1, \dots, x_n) \rangle$ is the answer to the counting query w. This establishes (81).

The final claim involving efficiency follows directly from Line 3 of Algorithm 4. □

5.3 Single-Dimensional Range Queries

We first present the matrix M discussed in previous section for the case of $d = 1$, i.e., single-dimensional range queries. In this case, the set $\mathcal{X} = [B]$ is simply identified with B consecutive points on a line, and a range query $[j, j']$ is specified by integers $j, j' \in [B]$ with $j \leqslant j'$. We will assume throughout that B is a power of 2. (This assumption is without loss of generality since we can always pad the input domain to be of size a power of 2, with the loss of a constant factor in our accuracy bounds.) We begin by presenting the basic building block in the construction of M, namely that of a *range query tree* $\mathcal{T}_B$ with B leaves and a *chosen set* $\mathcal{C}_B$ of B nodes of $\mathcal{T}_B$:

Definition 5.3 (Range query tree)**.** Suppose $B \in \mathbb{N}$ is a power of 2, $\ell \in \mathbb{N}, \gamma \in (0, 1)$. Define a complete binary tree $\mathcal{T}_B$ of depth $\log B$, where each node stores a single integer-valued random variable:

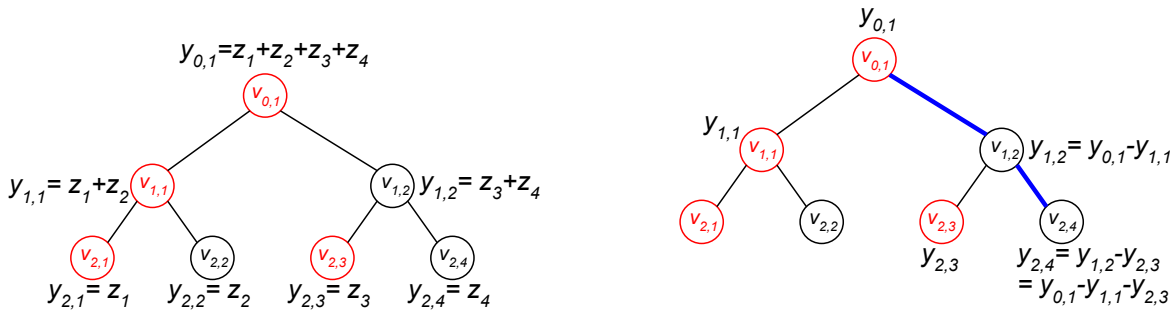
1. For a depth $0 \leqslant t \leqslant \log B$ and an index $1 \leqslant s \leqslant B/2^{\log B - t}$, let $v_{t,s}$ be the sth vertex of the tree at depth t (starting from the left). We will denote the value stored at vertex $v_{t,s}$ by $y_{t,s}$. The values $y_{t,s}$ will always have the property that $y_{t,s} = y_{t+1,2s-1} + y_{t+1,2s}$; i.e., the value stored at $v_{t,s}$ is the sum of the values stored at the two children of $v_{t,s}$.
2. Let $\mathcal{C}_B = \{v_{t,s} : 0 \leqslant t \leqslant \log B, s \equiv 1 \pmod 2\}$. Let the B nodes in $\mathcal{C}_B$ be ordered in the top-to-bottom, left-to-right order. In particular, $v_{0,1}$ comes first, $v_{1,1}$ is second, $v_{1,3}$ is third, $v_{2,1}$ is fourth, and in general: the jth node in this ordering ($1 < j \leqslant B$) is v_{t_j,s_j}, where $t_j = \lceil \log_2 j \rceil$, $s_j = 2(j - 2^{t_j - 1}) - 1$.
3. For $1 \leqslant j \leqslant B$, we will denote $z_j := y_{\log B, j}$ and $y_j = y_{t_j,s_j}$.

See Figure 2 for an illustration of $\mathcal{T}_4$. The next lemma establishes some basic properties of the set $\mathcal{C}_B$:

Lemma 5.3. *Fix d a power of 2. We have the following regarding the set $\mathcal{C}_B$ defined in Definition 5.3:*

1. *$\mathcal{C}_B$ is the union of the the root and set of nodes of $\mathcal{T}_B$ that are the left child of their parent.*
2. *For any node $u \notin \mathcal{C}_B$, there is some $v \in \mathcal{T}_B$ (which is an ancestor of u) so that there is a path from v to u that consists entirely of following the right child of intermediate nodes, starting from v.*

Proof of Lemma 5.3. The first part is immediate from the definition of $\mathcal{C}_B$. For the second part, given u, we walk towards the root, continually going to the parent of the current node. The first time we arrive at a node that is the left child of its parent, we will be at a node in $\mathcal{C}_B$; we let this node be v. □



(a) Range query tree, $B = 4$

(b) Path P constructed to derive (82)

Figure 2: (a) The range query tree $\mathcal{T}_4$. The nodes in $\mathcal{C}_4$ are highlighted in red. The labels $y_{t,s}, z_s$ next to nodes show the values stored at the nodes and the relations between them. Notice that in the case $B = 4$, we have $(t_1, s_1) = (0, 1), (t_2, s_2) = (1, 1), (t_3, s_3) = (2, 1), (t_4, s_4) = (2, 3)$. (b) The path P described in (82) for $j = 4$ is highlighted in blue. For this case ($B = j = 4$) we have $z_4 = y_{0,1} - y_{1,2} - y_{2,3}$.

Next we make two more definitions that will aid in the analysis:

Definition 5.4. For an integer $j \in [B]$, let $v(j)$ denote the number of steps from a node to its parent one must take starting at the leaf $v_{\log B,j}$ of the tree $\mathcal{T}_B$ to get to a node in $\mathcal{C}_B$. Equivalently, $v(j)$ is the 2-adic valuation of j (i.e., the base-2 logarithm of the largest power of 2 dividing j).

Definition 5.5. For a positive integer j, let $c(j)$ be the number of ones in the binary representation of j.

By property (1) of Definition 5.3, the set of all values $y_{t,s}$, for $0 \leqslant t \leqslant \log B$, $1 \leqslant s \leqslant B/2^t$, is entirely determined by the values z_s: in particular, for any $v_{t,s}$, $y_{t,s}$ is the sum of all z_s for which the leaf $v_{\log B,s}$ is a descendant of $v_{t,s}$. Conversely, given the values of $y_{t,s}$ for which $v_{t,s} \in \mathcal{C}_B$ (equivalently, the values y_{t_j,s_j} for $j \in [B]$), the values $z_j = y_{\log B,j}$ are determined as follows:

$$z_j = y_{\log B,j} = y_{\log B - v(j), j/2^{v(j)}} - \sum_{t=1}^{v(j)-1} y_{\log B - v(j)+t, j/2^{v(j)-t}-1}. \tag{82}$$

Graphically, we follow the path P from $v_{\log B,j}$ to the root until we hit a node $v_{t,s}$ in $\mathcal{C}_B$; then z_j is the difference of $y_{t,s}$ and the sum of the variables stored at the left child of each node in the path P. (See Figure 2 for an example.)

It follows from the argument in the previous paragraph that the linear transformation that sends the vector $(z_1, \ldots, z_B)$ to the vector $(y_{t_1,s_1}, \ldots, y_{t_B,s_B})$ is invertible; let $M_B \in \{0,1\}^{B\times B}$ be the matrix representing this linear transformation. By (82), which describes the linear transformation induced by M_B^{-1}, we have that $M_B^{-1} \in \{-1, 0, 1\}^{B\times B}$.

Since each leaf has $1 + \log B$ ancestors (including itself), we immediately obtain:

Lemma 5.4. *The sensitivity of M_B is given by $\Delta_{M_B} = 1 + \log B$.*

Next consider any range query $[j, j']$, so that $1 \leqslant j \leqslant j' \leqslant B$, and let $w \in \mathbb{R}^B$ be the row vector representing this range query (see Section 5.2). In particular all entries of w are 0 apart from $w_j, w_{j+1}, \ldots, w_{j'}$, which are all 1.

Lemma 5.5. *For a vector w representing a range query $[j, j']$, the vector wM_B^{-1} belongs to $\{-1, 0, 1\}^B$, and it has at most $c(j-1) + c(j') \leqslant 2 \log B$ nonzero entries. Moreover, the set of these nonzero entries (and their values) can be computed in time $O(\log B)$.*

Proof of Lemma 5.5. Since M_B is invertible, wM_B^{-1} is the unique vector $\nu \in \mathbb{R}^B$ such that for any values of $\{y_{t,s}\}_{0\leqslant t\leqslant B, s\in[B/2^t]}$ satisfying property (1) of Definition 5.3, we have

$$z_j + z_{j+1} + \cdots + z_{j'} = y_{\log B,j} + \cdots + y_{\log B,j'} = \langle \nu, (y_{t_1,s_1}, \ldots, y_{t_B,s_B})\rangle.$$

Next let $v_{\tilde{t},\tilde{s}}$ be the first node in $\mathcal{C}_B$ that is reached on the leaf-to-root path starting at $v_{\log B,j'}$. Recall from Definition 5.4 that $\tilde{t} = \log B - v(j')$. Consider the path on the tree $\mathcal{T}_B$ from the root $v_{1,1}$ to the node $v_{\tilde{t},\tilde{s}}$. Suppose the right child is taken at $h-1$ vertices of this path; it is not hard to see that $h = c(j')$ (see Definition 5.5). For $1 \leqslant k \leqslant h$, at the kth vertex on this path where the right child is taken, set $v_{t'_k,s'_k}$ to be the left child of the parent vertex (so that $v_{t'_k,s'_k}$ is not on the path). By Lemma 5.3, $v_{t'_k,s'_k} \in \mathcal{C}_B$. Also set $v_{t'_h,s'_h} = v_{\tilde{t},\tilde{s}}$. Then from Definition 5.3 (property (1)) we have

$$z_1 + \cdots + z_{j'} = y_{\log B,1} + \cdots + y_{\log B,j'} = \sum_{k=1}^{h} y_{t'_k,s'_k}. \tag{83}$$

The same computation for $j-1$ replacing j' yields, with $\hat{h} = c(j-1)$,

$$z_1 + \cdots + z_{j-1} = y_{\log B,1} + \cdots + y_{\log B,j-1} = \sum_{k=1}^{\hat{h}} y_{\hat{t}_k,\hat{s}_k}, \tag{84}$$

where the pairs $(\hat{t}_k, \hat{s}_k)$ replace the pairs (t'_k, s'_k). Taking the difference of (83) and (84) yields

$$z_j + \cdots + z_{j'} = \sum_{k=1}^{h} y_{t'_k,s'_k} - \sum_{k=1}^{\hat{h}} y_{\hat{t}_k,\hat{s}_k},$$

i.e., $z_j + \cdots + z_{j'}$ is a linear combination of at most $c(j-1) + c(j')$ elements of $\{y_{t,s} : v_{t,s} \in \mathcal{C}_B\}$, with coefficients in $\{-1, 1\}$. The sets $\{(t'_k, s'_k)\}_{1\leqslant k\leqslant h}$ and $\{(\hat{t}_k, \hat{s}_k)\}_{1\leqslant k\leqslant \hat{h}}$ can be computed in $O(\log B)$ time by walking on the leaf-to-root path starting at $v_{\log B,j'}$ and $v_{\log B,j-1}$, respectively. This establishes Lemma 5.5. □

Lemmas 5.4 and 5.5 establish properties (1) and (2) required of the matrix $M = M_B$ to guarantee $\operatorname{poly}\log(B)$ accuracy and $\operatorname{poly}\log(B)$ communication for private computation of 1-dimensional range queries. In the following section we use M_B to construct a matrix which satisfies the same properties for d-dimensional range queries for any $d \geqslant 1$.

5.4 Multi-Dimensional Range Queries

Fix any $d \geqslant 1$, and suppose the universe $\mathcal{X}$ consists of B_0 buckets in each dimension, i.e., $\mathcal{X} = [B_0]^d$. In this case, a range query $[j_1, j'_1] \times [j_2, j'_2] \times \cdots \times [j_d, j'_d]$ is specified by integers $j_1, j_2, \ldots, j_d, j'_1, j'_2, \ldots, j'_d \in [B_0]$ with $j_i \leqslant j'_i$ for all $i = 1, 2, \ldots, d$.

Throughout this section, we will consider the case that d is a constant (and B_0 is large). Moreover suppose that B_0 is a power of 2 (again, this is without loss of generality since we can pad each dimension to be a power of 2 at the cost of a blowup in $|\mathcal{X}|$ by at most a factor of 2^d). Write $B = |\mathcal{X}| = B_0^d$. Our goal is to define a matrix $M_{B,d}$ which satisfies analogues of Lemmas 5.4 and 5.5 for $w \in \{0,1\}^B$ representing multi-dimensional range queries (when $[B]$ is identified with $[B_0]^d$).

The idea behind the construction of $M_{B,d}$ is to apply the linear transformation M_{B_0} in each dimension, operating on a single-dimensional slice of the input vector $(z_{j_1,\ldots,j_d})_{j_1,\ldots,j_d\in[B_0]}$ (when viewed as a d-dimensional tensor) at a time. Alternatively, $M_{B,d}$ can be viewed combinatorially through the lens of *range trees* [Ben79]: $M_{B,d}$ is a linear transformation that takes the vector $(z_{j_1,\ldots,j_d})$ to a B-dimensional vector whose components are the values stored at the nodes of a range tree defined in a similar manner to the range query tree $\mathcal{T}_B$ for the case $d = 1$. However, we opt to proceed linear algebraically: the matrix $M_{B,d}$ is defined as follows. Fix a vector $z \in \mathbb{R}^B$. We will index the elements of z with d-tuples of integers in $[B_0]$, i.e., we will write $z = (z_{j_1,\ldots,j_d})_{j_1,\ldots,j_d\in[B_0]}$. For $1 \leqslant p \leqslant d$, let $M^{\mathrm{pre}}_{B,p}$ be the linear transformation that applies M_{B_0} to each vector $(z_{j_1,\ldots,j_{p-1},1,j_{p+1},\ldots,j_d}, \ldots, z_{j_1,\ldots,j_{p-1},B_0,j_{p+1},\ldots,j_d})$, where $j_1, \ldots, j_{p-1}, j_{p+1}, \ldots, j_d \in [B_0]$. That is, M_{B_0} is applied to each slice of the vector z, where the slice is being taken along the pth dimension. Then let

$$M_{B,d} := M^{\mathrm{pre}}_{B,d} \circ \cdots \circ M^{\mathrm{pre}}_{B,1}(z). \tag{85}$$

We will also use an alternate characterization of $M_{B,d}$, which we develop next. First identify $\mathbb{R}^B$ with the d-wise tensor product of $\mathbb{R}^{B_0}$, in the following (standard) manner: Let $e_1, \ldots, e_{B_0} \in \mathbb{R}^{B_0}$ be the standard basis vectors in $\mathbb{R}^{B_0}$. Then the collection of all $e_{j_1} \otimes \cdots \otimes e_{j_d}$, where $j_1, \ldots, j_d \in [B_0]$, form a basis for $\mathbb{R}^{B_0} \otimes \cdots \otimes \mathbb{R}^{B_0}$. Under the identification $\mathbb{R}^B \simeq (\mathbb{R}^{B_0})^{\otimes d}$, a vector $z = (z_{j_1,\ldots,j_d})_{j_1,\ldots,j_d \in [B_0]} \in \mathbb{R}^B$ is identified with the following linear combination of these basis vectors:

$$\sum_{j_1,\ldots,j_d \in [B_0]} z_{j_1,\ldots,j_d} \cdot e_{j_1} \otimes \cdots \otimes e_{j_d}.$$

Under this identification, the matrix $M_{B,d}$ corresponds to the following linear transformation of $(\mathbb{R}^{B_0})^{\otimes d}$:

$$M_{B_0} \otimes \cdots \otimes M_{B_0} : (\mathbb{R}^{B_0})^{\otimes d} \to (\mathbb{R}^{B_0})^{\otimes d}.$$

In the following lemmas, we will often abuse notation to allow $M_{B,d}$ to represent both the above linear transformation as well as the matrix in $\mathbb{R}^{B \times B}$ representing this transformation.

Lemma 5.6. *We have that $M_{B,d} \in \{0,1\}^{B \times B}$ and the sensitivity of $M_{B,d} : \mathbb{R}^B \to \mathbb{R}^B$ is bounded by $\Delta_{M_{B,d}} \leqslant (1 + \log B_0)^d$.*

Proof of Lemma 5.6. Notice that the $((j_1, \ldots, j_d), (j'_1, \ldots, j'_d))$ entry of $M_{B,d}$ is given by the following product:

$$\prod_{p=1}^{d} (M_{B_0})_{j_p, j'_p}.$$

Since $M_{B_0} \in \{0,1\}^{B_0 \times B_0}$, it follows immediately that $M_{B,d} \in \{0,1\}^{B \times B}$. Moreover, to upper bound the sensitivity of $M_{B,d}$ note that for any $(j'_1, \ldots, j'_d) \in [B_0]^d$,

$$\sum_{(j_1,\ldots,j_d) \in [B_0]^d} \prod_{p=1}^{d} (M_{B_0})_{j_p, j'_p} = \prod_{p=1}^{d} \left(\sum_{j_p=1}^{B_0} (M_{B_0})_{j_p, j'_p} \right) \leqslant (\Delta_{M_{B_0}})^d \leqslant (1 + \log B_0)^d,$$

where the last inequality above uses Lemma 5.4. □

Lemma 5.7. *For the vector w representing any range query $[j_1, j'_1] \times \cdots \times [j_d, j'_d]$, the vector $wM_{B,d}^{-1}$ belongs to $\{-1, 0, 1\}^B$ and moreover it has at most*

$$\prod_{p=1}^{d} (c(j_p - 1) + c(j'_p)) \leqslant (2 \log B_0)^d = (2 \log(B^{1/d}))^d$$

nonzero entries.

Proof of Lemma 5.7. The inverse $M_{B,d}^{-1}$ of $M_{B,d}$ is given by the d-wise tensor product $M_{B_0}^{-1} \otimes \cdots \otimes M_{B_0}^{-1}$. This can be verified by noting that this tensor product and $M_{B,d}$ multiply (i.e., compose) to the identity:

$$\begin{aligned}
(M_{B_0}^{-1} \otimes \cdots \otimes M_{B_0}^{-1}) \cdot M_{B,d} &= (M_{B_0}^{-1} \otimes \cdots \otimes M_{B_0}^{-1}) \cdot (M_{B_0} \otimes \cdots \otimes M_{B_0}) \\
&= (M_{B_0}^{-1} \cdot M_{B_0}) \otimes \cdots \otimes (M_{B_0}^{-1} \cdot M_{B_0}) \\
&= I_{B_0} \otimes \cdots \otimes I_{B_0} \\
&= I_B.
\end{aligned}$$

Recall that the (row) vector w representing the range query $[j_1, j'_1] \times \cdots \times [j_d, j'_d]$ satisfies, for each $(j''_1, \ldots, j''_d) \in [B_0]^d$, $w_{j''_1,\ldots,j''_d} = 1$ if and only if $j''_p \in [j_p, j'_p]$ for all $1 \leqslant p \leqslant d$, and otherwise $w_{j''_1,\ldots,j''_d} = 0$. Therefore, we may write w as the product of row vectors $w = w_1 \otimes \cdots \otimes w_d$, where for $1 \leqslant p \leqslant d$, w_p is the (row) vector representing

the range query $[j_p, j'_p]$. In particular, for $1 \leqslant j'' \leqslant B_0$, the j''th entry of w_p is 1 if and only if $j'' \in [j_p, j'_p]$. It follows that

$$wM_{B,d}^{-1} = (w_1 \otimes \cdots \otimes w_d)(M_{B_0}^{-1} \otimes \cdots \otimes M_{B_0}^{-1}) = w_1 M_{B_0}^{-1} \otimes \cdots \otimes w_d M_{B_0}^{-1}. \tag{86}$$

By Lemma 5.5, for $1 \leqslant p \leqslant d$, the vector $w_p M_{B_0}^{-1}$ has entries in $\{-1, 0, 1\}$, at most $c(j_p - 1) + c(j'_p)$ of which are nonzero. Since $wM_{B,d}^{-1}$ is the tensor product of these vectors and the set $\{-1, 0, 1\}$ is closed under multiplication, it also has entries in $\{-1, 0, 1\}$, at most $\prod_{p=1}^{d}(c(j_p - 1) + c(j'_p))$ of which are nonzero. □

The following lemma allows us to bound the running time of the local randomizer (Algorithm 3) and analyzer (Algorithm 4):

Lemma 5.8. *Given B, d with $B = B_0^d$, the following can be computed in $O(\log^d B_0)$ time:*

(1) Given indices $(j_1, \ldots, j_d) \in [B_0]^d$, the nonzero indices of $M_{B,d}$ for the column indexed by $(j_1, \ldots, j_d)$.

(2) Given a vector $w \in \mathbb{R}^B$ specifying a range query, the set of nonzero elements of $wM_{B,d}^{-1}$ and their values (which are in $\{-1, 1\}$).

Proof of Lemma 5.8. We first deal with the case $d = 1$, i.e., the matrix $M_{B,1} = M_B$. Given $j, j' \in [B]$, the (j', j)-entry of M_B is 1 if and only if the node $v_{t_{j'}, s_{j'}}$ of the tree $\mathcal{T}_B$ is an ancestor of the leaf $v_{\log B, j}$. Since $t_j = \lceil \log_2 j \rceil, s_j = 2(j - 2^{t_j - 1}) - 1$, whether or not $v_{t_{j'}, s_{j'}}$ is an ancestor of $v_{\log B, j}$ can be determined in $O(\log B)$ time, thus establishing (1) for the case $d = 1$. Notice that the statement of Lemma 5.5 immediately gives (2) for the case $d = 1$.

To deal with the case of general d, notice that $M_{B,d} = (M_{B_0})^{\otimes d}$. Therefore, for a given $(j_1, \ldots, j_d)$ the set

$$\{(j'_1, \ldots, j'_d) : (M_{B,d})_{(j'_1, \ldots, j'_d), (j_1, \ldots, j_d)} = 1\} \tag{87}$$

of nonzero indices in the $(j_1, \ldots, j_d)$-th column of $M_{B,d}$ is equal to the Cartesian product

$$\bigtimes_{1 \leqslant p \leqslant d} \{j'_p : (M_{B_0})_{j'_p, j_p} = 1\}.$$

Since each of the sets $\{j'_p : (M_{B_0})_{j'_p, j_p} = 1\}$ can be computed in time $O(\log B_0)$ (using the case $d = 1$ solved above), and is of size $O(\log B_0)$, the product of these sets (87) can be computed in time $O(\log^d B_0)$, thus completing the proof of item (1) in the lemma.

The proof of item (2) for general d is similar. For $1 \leqslant p \leqslant d$, let w_p be the vector in $\mathbb{R}^{B_0}$ corresponding to the 1-dimensional range query $[j_p, j'_p]$. Then recall from (86) we have that $wM_{B,d}^{-1} = w_1 M_{B_0}^{-1} \otimes \cdots \otimes w_d M_{B_0}^{-1}$. By item (2) for $d = 1$, the nonzero entries of each of $w_p M_{B_0}^{-1}$ (and their values) can be computed in time $O(\log B_0)$; since each of these sets has size $O(\log B_0)$, the set of nonzero entries of $wM_{B,d}^{-1}$, which is the Cartesian product of these sets, as well as the values of these entries, can be computed in time $O(\log^d B_0)$. □

5.5 Guarantees for Differentially Private Range Queries

In this section we state the guarantees of Theorems 5.1 and 5.2 on the privacy and accuracy of the protocol $P^{\text{matrix}} = (R^{\text{matrix}}(n, B, M, R^{\text{FO}}), S, A^{\text{matrix}}(n, B, M, A^{\text{FO}}))$ for range query computation when $M = M_{B,d}$ and the pair $(R^{\text{FO}}, A^{\text{FO}})$ is chosen to be either $(R^{\text{CM}}, A^{\text{CM}})$ (Count Min sketch-based approach; Algorithm 2) or $(R^{\text{Had}}, A^{\text{Had}})$ (Hadamard response-based approach; Algorithm 1).

For the Hadamard response-based frequency oracle, we obtain the following:

Theorem 5.9. *Suppose $B_0, n, d \in \mathbb{N}$, $B = B_0^d$, and $0 \leqslant \varepsilon \leqslant 1$, and $\beta, \delta \geqslant 0$ with $1/\beta \leqslant B^{O(1)}$[16]. Consider the shuffled-model protocol $P^{\text{matrix}} = (R^{\text{matrix}}, S, A^{\text{matrix}})$, where:*

- *$R^{\text{matrix}} = R^{\text{matrix}}(n, B, M_{B,d}, R^{\text{Had}})$ is defined in Algorithm 3;*
- *$A^{\text{matrix}} = A^{\text{matrix}}(n, B, M_{B,d}, A^{\text{Had}})$ is defined in Algorithm 4;*

[16]The assumption that $1/\beta$ is polynomial in B is purely for simplicity and can be removed at the cost of slightly more complicated bounds.

- *and* $R^{\mathrm{Had}} = R^{\mathrm{Had}}(n, B, \log n, \rho, (\log 2B)^d)$ *and* $A^{\mathrm{Had}} = A^{\mathrm{Had}}(n, B, \log n, \rho, (\log 2B)^d)$ *are defined in Algorithm 1, and*

$$\rho = \frac{36(\log 2B)^{2d} \ln(e(\log 2B)^d/(\varepsilon\delta))}{\varepsilon^2}. \tag{88}$$

Then:

- *The protocol* P^{matrix} *is* (ε, δ)*-differentially private in the shuffled model (Definition 2.2).*
- *For any dataset* $X = (x_1, \ldots, x_n) \in ([B_0]^d)^n$*, with probability* $1 - \beta$*, the frequency estimate of* P^{matrix} *for each* d*-dimensional range query has additive error at most* $O(\varepsilon^{-1} d^{1/2} (2 \log B)^{2d+1/2} \cdot \sqrt{\log((\log B)/(\varepsilon\delta))})$*.*
- *The local randomizers send a total of* $O(n \cdot \rho)$ *messages, each of length* $O(\log n \log B)$*. The analyzer can either (a) produce a data structure of size* $O(B \log(n\rho))$ *bits such that a single range query can be answered in time* $O((2 \log B)^d)$*, or (b) produce a data structure of size* $O(n\rho \log n \log B)$ *such that a single range query can be answered in time* $O(n\rho(2 \log B)^d \log n \log B)$*.*

Proof of Theorem 5.9. Lemma 5.6 guarantees that $\Delta_{M_{B,d}} \leqslant (1 + \log B)^d = (\log 2B)^d$. Then by Theorem 5.1, to show (ε, δ)-differential privacy of P^{matrix} it suffices to show (ε, δ)-differential privacy of the shuffled-model protocol $P^{\mathrm{Had}} := (R^{\mathrm{Had}}, S, A^{\mathrm{Had}})$. By Theorem 4.3 with $k = (\log 2B)^d$, this holds with ρ as in (88).

Next we show accuracy of P^{matrix}. Lemma 5.7 guarantees that for any $w \in \{0,1\}^B$ representing a range query, $wM_{B,d}^{-1}$ has at most $(2 \log B)^d$ nonzero entries, all of which are either -1 or 1. Moreover, by Theorem 4.3 with $k = (\log 2B)^d$ and ρ as in (88), for any $1 \geqslant \beta \geqslant 0$, the shuffled model protocol P^{Had} provides a

$$\left(O\left(\log(B/\beta) + \frac{(\log 2B)^d \sqrt{\log(B/\beta) \log((\log 2B)^d/(\varepsilon\delta))}}{\varepsilon} \right), \beta, (\log 2B)^d \right)$$

frequency oracle. By Theorem 5.2 and the assumption that $1/\beta \leqslant B^{O(1)}$, it follows that with probability $1 - \beta$, the frequency estimates of P^{matrix} on each d-dimensional range query have additive error at most

$$\leqslant O\left(\frac{(2 \log B)^{2d+1/2} \cdot \sqrt{d \left(\log((\log B)/(\delta\varepsilon))\right)}}{\varepsilon} \right).$$

This establishes the claim regarding accuracy of P^{matrix}.

To establish the last item (regarding efficiency), notice that the claims regarding communication (the number of messages and message length) follow from Lemma 4.6 with $k = (\log 2B)^d$. Part (a) of the claim regarding efficiency of the analyzer follows from item 2 of Lemma 4.6 and the last sentence in the statement of Theorem 5.2. Part (b) of the claim regarding efficiency of the analyzer follows from item 3 of Lemma 4.6 and the last sentence in the statement of Theorem 5.2. □

Similarly, for the Count Min sketch-based frequency oracle, we obtain

Theorem 5.10. *There is a sufficiently large constant* ζ *such that the following holds. Suppose* $B_0, n, d \in \mathbb{N}$*,* $B = B_0^d \geqslant n$[17]*, and* $0 \leqslant \varepsilon \leqslant 1$*, and* $\beta, \delta \geqslant 0$*. Consider the shuffled-model protocol* $P^{\mathrm{matrix}} = (R^{\mathrm{matrix}}, S, A^{\mathrm{matrix}})$*, where:*

- $R^{\mathrm{matrix}} = R^{\mathrm{matrix}}(n, B, M_{B,d}, R^{\mathrm{CM}})$ *is defined in Algorithm 3;*
- $A^{\mathrm{matrix}} = A^{\mathrm{matrix}}(n, B, M_{B,d}, A^{\mathrm{CM}})$ *is defined in Algorithm 4;*
- *and* $R^{\mathrm{CM}} = R^{\mathrm{CM}}(n, B, \log 2B/\beta, \gamma, 2kn)$ *and* $A^{\mathrm{CM}} = A^{\mathrm{CM}}(n, B, \log 2B/\beta, 2kn)$ *are defined in Algorithm 2, where*

$$\gamma = \frac{1}{n} \cdot \zeta \cdot \frac{\log^2(B/\beta) k^2 \log(\log(B/\beta)k/\delta)}{\varepsilon^2}$$

and $k = (\log 2B^{1/d})^d = (\log 2B_0)^d$*.*

Then:

[17] The assumption $n \leqslant B$ is made to simplify the bounds and can be removed.

- *The protocol P^{matrix} is (ε, δ)-differentially private in the shuffled model (Definition 2.2).*
- *For any dataset $X = (x_1, \ldots, x_n) \in ([B_0]^d)^n$, with probability $1 - \beta$, the frequency estimate of P^{matrix} for each d-dimensional range query has additive error at most*

$$O\left(\frac{(2\log B_0)^{2d}}{\varepsilon} \cdot \sqrt{\log^3(B/\beta)\log\left((\log(B/\beta))(\log 2B_0)^d/\delta\right)}\right).$$

- *With probability at least $1 - \beta$, each local randomizer sends a total of at most*

$$\tilde{m} := O\left(\frac{\log^3(B/\beta)(\log 2B_0)^{3d}\log((\log(B/\beta))(\log 2B_0)^d/\delta)}{\varepsilon^2}\right)$$

messages, each of length $O(\log(\log B/\beta) + \log((\log 2B_0)^d n))$. Moreover, in time $O(n\tilde{m})$, the analyzer produces a data structure of size $O(n\log(B/\beta)(\log 2B_0)^d \log(n\tilde{m}))$ bits, such that a single range query can be answered in time $O((2\log B_0)^d \cdot \log B/\beta)$.

Proof of Theorem 5.10. Lemma 5.6 guarantees that $\Delta_{M_{B,d}} \leqslant (1 + \log B_0)^d = (\log 2B^{1/d})^d$. (Recall our notation that $B = (B_0)^d$.) Then by Theorem 5.1, to show (ε, δ)-differential privacy of P^{matrix} it suffices to show (ε, δ)-differential privacy of the shuffled-model protocol $P^{\mathrm{CM}} := (R^{\mathrm{CM}}, S, A^{\mathrm{CM}})$. For the parameters above this follows from Theorem 4.7.

Next we show accuracy of P^{matrix}. Lemma 5.7 guarantees that for any $w \in \{0,1\}^B$ representing a range query, $wM_{B,d}^{-1}$ has at most $(2\log B_0)^d$ nonzero entries, all of which are either -1 or 1. Moreover, by Theorem 4.7, the shuffled model protocol P^{CM} provides an $(\kappa, \beta, (\log 2B_0)^d)$-frequency oracle with

$$\kappa \leqslant O\left(\frac{(\log 2B_0)^d}{\varepsilon} \cdot \sqrt{\log^3(B/\beta)\log\left((\log(B/\beta))(\log 2B_0)^d/\delta\right)}\right).$$

By Theorem 5.2 with $a = (2\log B_0)^d$, it follows that with probability $1 - \beta$, the frequency estimates of P^{matrix} on each d-dimensional range query have additive error at most

$$O\left(\frac{(2\log B_0)^{2d}}{\varepsilon} \cdot \sqrt{\log^3(B/\beta)\log\left((\log(B/\beta))(\log 2B_0)^d/\delta\right)}\right).$$

This establishes the second item. The final item follows from Lemma 4.9, part (2) of Lemma 5.8, and the final sentence in the statement of Theorem 5.2. □

6 Conclusion and Open Problems

The shuffled model is a promising new privacy framework motivated by the significant interest on anonymous communication. In this paper, we studied the fundamental task of frequency estimation in this setup. In the single-message shuffled model, we established nearly tight bounds on the error for frequency estimation and on the number of users required to solve the selection problem. We also obtained communication-efficient multi-message private-coin protocols with exponentially smaller error for frequency estimation, heavy hitters, range counting, and M-estimation of the median and quantiles (and more generally sparse non-adaptive SQ algorithms). We also gave public-coin protocols with, in addition, small query times. Our work raises several interesting open questions and points to fertile future research directions.

Our $\tilde{\Omega}(B)$ lower bound for selection (Theorem 1.2) holds for single-message protocols even with unbounded communication. We conjecture that a lower bound on the error of $B^{\Omega(1)}$ should hold even for multi-message protocols (with unbounded communication) in the shuffled model, and we leave this as a very interesting open question. Such a lower bound would imply a first separation between the central and (unbounded communication) multi-message shuffled model.

Another interesting question is to obtain a private-coin protocol for frequency estimation with polylogarithmic error, communication per user, and query time; reducing the query time of our current protocol below $\tilde{O}(n)$ seems

challenging. In general, it would also be interesting to reduce the polylogarithmic factors in our guarantees for range counting as that would make them practically useful.

Another interesting direction for future work is to determine whether our efficient protocols for frequency estimation with much less error than what is possible in the local model could lead to more accurate and efficient shuffled-model protocols for fundamental primitives such as clustering [Ste20] and distribution testing [ACFT19], for which current locally differentially private protocols use frequency estimation as a black box.

Finally, a promising future direction is to extend our protocols for sparse non-adaptive SQ algorithms to the case of sparse aggregation. Note that the queries made by sparse non-adaptive SQ algorithms correspond to the special case of sparse aggregation where all non-zero queries are equal to 1. Extending our protocols to the case where the non-zero coordinates can be arbitrary numbers would, e.g., capture sparse stochastic gradient descent (SGD) updates, an important primitive in machine learning. More generally, it would be interesting to study the complexity of various other statistical and learning tasks [Smi11, WZ10, BST14, CMS11, CM08, CSS13] in the shuffled privacy model.

Acknowledgments

We would like to thank James Bell, Albert Cheu, Úlfar Erlingsson, Vitaly Feldman, Adrià Gascón, Peter Kairouz, Pasin Manurangsi, Stefano Mazzocchi, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, Abhradeep Guha Thakurta, Salil Vadhan, and Vinod Vaikuntanathan as well as the anonymous reviewers of a previous version of this paper, for very helpful comments and suggestions. In particular, we would like to thank an anonymous reviewer who pointed out the connection to nonadaptive SQ algorithms in Corollary 1.4.

A Proof of Theorem 3.4

In this section we prove Theorem 3.4. The proof is a simple consequence of the privacy amplification result of [BBGN19c] and known accuracy bounds for locally-differentially private protocols. We first recall the privacy amplification result:

Theorem A.1 (Privacy amplification of single-message shuffling, [BBGN19c], Corollary 5.3.1)**.** *Suppose $R : \mathcal{X} \to \mathcal{Z}$ is an $(\varepsilon_L, 0)$-locally differentially private randomizer with $\varepsilon_L \leqslant \frac{\ln(n/\ln(1/\delta))}{2}$ for some $\delta > 0$. Then the shuffled algorithm $(x_1, \ldots, x_n) \mapsto S(R(x_1), \ldots, R(x_n))$ is (ε, δ)-differentially private with $\varepsilon = O\left(\varepsilon_L \cdot e^{\varepsilon_L} \cdot \sqrt{\ln(1/\delta)/n}\right)$.*

Proof of Theorem 3.4. We first treat the case of $n \leqslant \tilde{O}(B^2)$ (i.e., the cases (14), (15)), where the locally differentially private protocol we use is B-RAPPOR [EPK14, DJW18]. In particular, we consider the protocol $P_{\mathrm{RAPPOR}} = (R_{\mathrm{RAPPOR}}, A_{\mathrm{RAPPOR}})$. For a given privacy parameter $\varepsilon_L \geqslant 1$, the local randomizer $R_{\mathrm{RAPPOR}} : [B] \to \{0,1\}^B$ is defined as follows: for $v \in [B]$, $R_{\mathrm{RAPPOR}}(v) = (Z_1, \ldots, Z_B)$, where each Z_k is an independent bit that equals $(e_v)_k$ with probability $\frac{\exp(\varepsilon_L/2)}{1+\exp(\varepsilon_L/2)}$ and equals $1 - (e_v)_k$ with probability $\frac{1}{1+\exp(\varepsilon_L/2)}$. For later use in the proof, we will also define $R_{\mathrm{RAPPOR}}(\varnothing) = (Z_1, \ldots, Z_B)$, where each $Z_k \sim \mathrm{Ber}\left(\frac{1}{1+\exp(\varepsilon_L/2)}\right)$.

The analyzer $A_{\mathrm{RAPPOR}} : (\{0,1\}^B)^n \to \{0,1\}^B$ is defined as follows: given as input n bit-vectors $(z^1, \ldots, z^n)$, the analyzer outputs the vector $(\hat{x}_1, \ldots, \hat{x}_B) \in [0,1]^n$ of frequency estimates defined by

$$\hat{x}_j = \frac{1}{n} \cdot \left(\left(\sum_{i=1}^{n} z_j^i \right) - \frac{n}{1 + \exp(\varepsilon_L/2)} \right) \cdot \left(\frac{\exp(\varepsilon_L/2) + 1}{\exp(\varepsilon_L/2) - 1} \right). \tag{89}$$

First we prove the accuracy of the local-model protocol $P_{\mathrm{RAPPOR}} = (R_{\mathrm{RAPPOR}}, A_{\mathrm{RAPPOR}})$; it is clear, by symmetry of A_{RAPPOR} that the same accuracy bounds hold when we insert the shuffler S. The expression (89) satisfies the following property: for any dataset $X = (x_1, \ldots, x_n)$, if $Z^i := R_{\mathrm{RAPPOR}}(x_i)$, and it happens that $\sum_{i=1}^n Z_j^i$ equals its expected value (over the randomness in the local randomizers R_{RAPPOR}), then $\hat{x}_j$ is

equal to the true frequency $\frac{1}{n}\sum_{i=1}^{n}(e_{x_i})_j$. It follows that if $\left|\frac{1}{n}\sum_{i=1}^{n} Z_j^i - \mathbb{E}_{R_{\mathrm{RAPPOR}}}\left[\frac{1}{n}\sum_{i=1}^{n} Z_j^i\right]\right| \leqslant \kappa$, then $\left|\hat{x}_j - \frac{1}{n}\sum_{i=1}^{n}(e_{x_i})_j\right| \leqslant \kappa \cdot \left(\frac{\exp(\varepsilon_L/2)+1}{\exp(\varepsilon_L/2)-1}\right) \leqslant O(\kappa)$, where the final inequality follows from $\varepsilon_L \geqslant 1$.

Let $p = \frac{1}{\exp(\varepsilon_L/2)+1}$. The random variable $\left|\sum_{i=1}^{n} Z_j^i - \mathbb{E}_{R_{\mathrm{RAPPOR}}}\left[\sum_{i=1}^{n} Z_j^i\right]\right|$ is stochastically dominated by the random variable $|Y - np|$, where $Y \sim \mathrm{Bin}\,(n,p)$. It follows that if $Y_1, \ldots, Y_B \sim \mathrm{Bin}\,(n,p)$ i.i.d., then

$$\begin{aligned}
&\mathbb{E}\left[\max_{j\in[B]}\left|\sum_{i=1}^{n} Z_j^i - \mathbb{E}_{R_{\mathrm{RAPPOR}}}\left[\sum_{i=1}^{n} Z_j^i\right]\right|\right] \\
&\leqslant \mathbb{E}\left[\max_{j\in[B]}|Y_j - np|\right] \\
&\leqslant \mathbb{E}\left[\left(\max_{j\in[B]}\{Y_j\} - np\right)\right] + E[|Y_1 - np|]. \qquad (90)
\end{aligned}$$

By Jensen's inequality $\mathbb{E}[|Y_1 - np|] \leqslant \sqrt{np}$. By Exercise 2.19 of [BLM12], we have that

$$\mathbb{E}\left[\max_{j\in[B]}\{Y_j\}\right] \leqslant np\exp\left(1 + W\left(\frac{\ln(B) - np}{enp}\right)\right),$$

where $W(\cdot)$ is the Lambert W function.[18] We consider two cases regarding the value of pn:

Case 1. $epn < \ln B$. In this case we use the fact that $W(x) \leqslant \ln(e\cdot x)$ for all $x \geqslant 1$. Then since W is increasing,

$$\begin{aligned}
np\exp\left(1 + W\left(\frac{\ln(B) - np}{enp}\right)\right) &\leqslant np\exp\left(1 + W\left(\ln(B)/(enp)\right)\right) \\
&\leqslant np\exp(1 + \ln(\ln(B)/(np))) \\
&\leqslant np\exp(\ln(e\ln(B)/(np))) \\
&= np\cdot\frac{e\ln(B)}{np} = e\ln B.
\end{aligned}$$

Thus in this case (90) is bounded above by $O(\ln B)$.

Case 2. $epn \geqslant \ln B$. In this case we use the fact that $W\left(\frac{-1}{e} + x\right) \leqslant -1 + 3\sqrt{x}$ for all $x \geqslant 0$. In particular, it follows from this fact that

$$\begin{aligned}
np\exp\left(1 + W\left(\frac{\ln(B)}{enp} - \frac{1}{e}\right)\right) &\leqslant np\exp\left(1 - 1 + 3\sqrt{\frac{\ln(B)}{enp}}\right) \\
&\leqslant np\cdot O\left(\sqrt{\frac{\ln B}{enp}}\right) \\
&= O\left(\sqrt{np\ln(B)}\right),
\end{aligned}$$

where the second inequality uses the fact that $\ln(B)/np = O(1)$. Thus in this case (90) is bounded above by $O(\sqrt{np\ln B})$.

Next we analyze privacy of R_{RAPPOR} in the n-user shuffled model. It is clear that R_{RAPPOR} is $(\varepsilon_L, 0)$-differentially private. In fact, R_{RAPPOR} satisfies the following stronger property: for any $v \in [B]$, and any vector $z \in \{0,1\}^B$, we have that $e^{-\varepsilon_L/2}\mathbb{P}[R_{\mathrm{RAPPOR}}(\varnothing) = z] \leqslant \mathbb{P}[R_{\mathrm{RAPPOR}}(v) = z] \leqslant e^{\varepsilon_L/2}\mathbb{P}[R_{\mathrm{RAPPOR}}(\varnothing) = z]$. Now write $M(x_1, \ldots, x_n) = S(R_{\mathrm{RAPPOR}}(x_1), \ldots, R_{\mathrm{RAPPOR}}(x_n))$. It is not difficult to see by inspecting the proof of Theorem A.1 that the following holds, as long as ε, δ are chosen so that $\frac{\varepsilon_L}{2} \leqslant \frac{\ln(n/\ln(1/\delta))}{2}$ and $\varepsilon =$

[18] The lambert W function $W : [-1/e, \infty) \to \mathbb{R}$ is defined implicitly by $W(x)\exp(W(x)) = x$ and $W(x) \geqslant -1$ for all $x \geqslant -1/e$. It is an increasing function.

$O\left(\varepsilon_L e^{\varepsilon_L/2}\sqrt{\ln(1/\delta)/n}\right)$: For any subset $\mathcal{S} \subset (\{0,1\}^B)^n$, and any dataset $(x_1,\ldots,x_n) \in [B]^n$,

$$\mathbb{P}\left[M(x_1,\ldots,x_{n-1},x_n) \in \mathcal{S}\right] \leqslant e^{\varepsilon}\mathbb{P}\left[M(x_1,\ldots,x_{n-1},\varnothing) \in \mathcal{S}\right] + \delta$$
$$\mathbb{P}\left[M(x_1,\ldots,x_{n-1},\varnothing) \in \mathcal{S}\right] \leqslant e^{\varepsilon}\mathbb{P}\left[M(x_1,\ldots,x_{n-1},x_n) \in \mathcal{S}\right] + \delta.$$

It follows that $(x_1,\ldots,x_n) \mapsto M(x_1,\ldots,x_n)$ is $(2\varepsilon, \delta(1+e^{\varepsilon}))$-differentially private (i.e., in the n-user shuffled model). Thus, by choosing $\varepsilon_L = \ln(n/\ln(1/\delta)) - 2\ln\ln n + 2\ln(\varepsilon) + O(1)$, we obtain the accuracy bounds in (14) and (15); in particular, the accuracy bound in Case 1 corresponds to $n \leqslant \frac{\varepsilon^2 \log^2 B}{\log^3 \log B}$ and the accuracy bound in Case 2 corresponds to $n \geqslant \frac{\varepsilon^2 \log^2 B}{\log^3 \log B}$.

Finally we treat the case $n > \Omega(B^2)$ (i.e., the case (16)). In this case we will use the local randomizer of B-randomized response [War65]. In particular, the local randomizer $R_{\mathrm{RR}} : [B] \to [B]$ is defined as follows: for $u, v \in [B]$,

$$\mathbb{P}[R_{\mathrm{RR}}(v) = u] = \begin{cases} \frac{\exp(\varepsilon_L)}{\exp(\varepsilon_L)+B-1} & : u = v \\ \frac{1}{\exp(\varepsilon_L)+B-1} & : u \neq v. \end{cases}$$

The analyzer $A_{\mathrm{RR}} : [B]^n \to [B]^n$, when given outputs of local randomizers $(z_1,\ldots,z_n) \in [B]^n$, produces frequency estimates $A(z_1,\ldots,z_n) = (\hat{x}_1,\ldots,\hat{x}_B)$, given by

$$\hat{x}_j = \frac{1}{n}\left(\left(\sum_{i=1}^n (e_{z_i})_j\right) - \frac{n}{\exp(\varepsilon_L)+B-1}\right) \cdot \left(\frac{\exp(\varepsilon_L)+B-1}{\exp(\varepsilon_L)-1}\right). \tag{91}$$

First we analyze the accuracy of A_{RR}. The analysis is quite similar to that of A_{RAPPOR}. In particular, first note that (91) satisfies the following property: for any dataset $X = (x_1,\ldots,x_n)$, if $Z^i := R_{\mathrm{RR}}(x_i)$, and it happens that $\sum_{i=1}^n (e_{Z^i})_j$ equals its expected value (over the randomness in the local randomizers R_{RR}), then $\hat{x}_j$ is equal to the true frequency $\frac{1}{n}\sum_{i=1}^n (e_{x_i})_j$. It follows that if $\left|\frac{1}{n}\sum_{i=1}^n (e_{Z^i})_j - \mathbb{E}_{R_{\mathrm{RR}}}\left[\frac{1}{n}\sum_{i=1}^n (e_{Z^i})_j\right]\right| \leqslant \kappa$, then $\left|\hat{x}_j - \frac{1}{n}\sum_{i=1}^n (e_{x_i})_j\right| \leqslant \kappa \cdot \left(\frac{\exp(\varepsilon_L)+B-1}{\exp(\varepsilon_L)-1}\right) \leqslant O\left(\kappa \cdot \left(\frac{\exp(\varepsilon_L)+B}{\exp(\varepsilon_L)}\right)\right)$, where the last inequality follows from $\varepsilon_L \geqslant 1$ (as we will see below).

Let $q = \frac{B-1}{\exp(\varepsilon_L)+B-1}$. The random variable $|\sum_{i=1}^n (e_{Z^i})_j - \mathbb{E}_{R_{\mathrm{RR}}}[\sum_{i=1}^n (e_{Z^i})_j]|$ is stochastically dominated by the random variable $|Y - n(1-q)|$, where $Y \sim \mathrm{Bin}(n, 1-q)$. It follows by binomial concentration that if $Y_1,\ldots,Y_B \sim \mathrm{Bin}(n,1-q)$ i.i.d. and $qn \geqslant \Omega(\ln B)$, then

$$\mathbb{E}_{R_{\mathrm{RR}}}\left[\max_{j\in[B]}\left|\sum_{i=1}^n (e_{Z^i})_j - \mathbb{E}_{R_{\mathrm{RR}}}\left[\sum_{i=1}^n (e_{Z^i})_j\right]\right|\right] \leqslant O(\sqrt{qn\ln(B)}).$$

Thus the expected error of $P_{\mathrm{RR}} = (R_{\mathrm{RR}}, A_{\mathrm{RR}})$ is bounded above by

$$\mathbb{E}_{(\hat{x}_1,\ldots,\hat{x}_B)\sim P_{\mathrm{RR}}}\left[\max_{j\in[B]}\left|\hat{x}_j - \frac{1}{n}\sum_{i=1}^n (e_{x_i})_j k\right|\right] \leqslant O\left(\sqrt{q\ln(B)/n} \cdot \left(\frac{\exp(\varepsilon_L)+B}{\exp(\varepsilon_L)}\right)\right). \tag{92}$$

Next we analyze the privacy of R_{RR} in the n-user shuffled model. To do so, note that R_{RR} is clearly $(\varepsilon_L, 0)$-(locally) differentially private. Thus, by [BBGN19c, Theorem 3.1][19], if we take $\varepsilon_L = \ln(n/\ln(1/\delta)) + 2\ln(\varepsilon) + O(1)$, then by the assumption $\varepsilon \geqslant \omega(\ln^2(n)/\min\{\sqrt{B}, \sqrt{n}\})$, the shuffled-model protocol $(R_{\mathrm{RR}}, S, A_{\mathrm{RR}})$ is (ε,δ)-differentially private in the n-user shuffled model. As long as $\sqrt{n} > B$ (i.e., $n > B^2$, so that $\frac{n\varepsilon^2}{\ln 1/\delta} \geqslant \Omega(B)$), the error in (92) is bounded above by $O(\sqrt{q\ln(B)/n}) = O\left(\frac{1}{n\varepsilon} \cdot \sqrt{B\ln(n)\ln(B)}\right)$. □

[19] In particular, the parameter γ in Theorem 3.1 of [BBGN19c] is set to $\frac{B}{\exp(\varepsilon_L)+B-1}$.

B Low-Communication Simulation of Sparse Non-Adaptive SQ Algorithms

We now discuss an equivalent formulation of our results in terms of non-adaptive statistical query algorithms. A *statistical query* on a set $\mathcal{X}$ is specified by a binary-valued predicate function $q : \mathcal{X} \to \{0, 1\}$, and, for a distribution $\mathcal{D}$ on $\mathcal{X}$, takes the value $q(\mathcal{D}) := \mathbb{E}_{x\sim\mathcal{D}}[q(x)]$. Special cases of statistical queries include *frequency queries*, specified by $q(x) = \mathbb{1}[x = y]$ for some $y \in \mathcal{X}$, and *range queries*, given by $q(x) = \mathbb{1}[x \in R]$, where R is a rectangle in $\mathcal{X}$. For $\tau \in [0, 1]$, a statistical query oracle $\mathrm{SQ}_{\mathcal{D},\tau}$ of *tolerance* τ, takes as input a statistical query q and outputs a value $\mathrm{SQ}_{\mathcal{D},\tau}(q) \in [q(\mathcal{D}) - \tau, q(\mathcal{D}) + \tau]$. A *statistical query (SQ) algorithm of tolerance* τ, $\mathcal{Q}$, may access the distribution $\mathcal{D}$ through a number of queries q to an oracle $\mathrm{SQ}_{\mathcal{D},\tau}$. $\mathcal{Q}$ is called *non-adaptive* if the distribution of its queries is fixed *a priori*, i.e., does not depend on the results of any of these queries. It was observed in the work of Blum et al. [BDNM05] that any statistical query (SQ) algorithm can be simulated by a differentially private protocol (in the central model). The same was shown for locally differentially private protocols by Kasiviswanathan et al. [KLN$^+$08], albeit with worse parameters. In fact, it is known [KLN$^+$08] that (non-adaptive) SQ algorithms are equivalent to (noninteractive) locally-differentially private algorithms, up to a polynomial factor in the tolerance τ. We refer the reader to [KLN$^+$08] for further background on SQ algorithms.

A straightforward corollary of the techniques used to show Theorem 1.3 is that one can efficiently and privately simulate *sparse* non-adaptive statistical query algorithms in the shuffled model. In particular, for $k \in \mathbb{N}$, we say that a non-adaptive SQ algorithm $\mathcal{Q}$ is k-*sparse* if, for each $x \in \mathcal{X}$, with probability 1, there are at most k distinct statistical queries q that $\mathcal{Q}$ makes satisfying $q(x) = 1$. Sparsity is a more stringent condition than having low sensitivity: in particular, if a k-sparse algorithm makes B queries $q_1, \ldots, q_B$, then, letting $\hat{\mathcal{D}}$ be the empirical distribution over a dataset $(x_1, \ldots, x_n)$, the mapping $(x_1, \ldots, x_n) \mapsto (q_1(\hat{\mathcal{D}}), \ldots, q_B(\hat{\mathcal{D}}))$ has ℓ_1 sensitivity $2k/n$. We show in Corollary B.1 below that for any universe $\mathcal{X}$, if $\mathcal{Q}$ is a k-sparse SQ algorithm making at most B queries, then one can privately simulate $\mathcal{Q}$ in the shuffled model, as long as the tolerance τ is roughly a multiplicative factor of k/n times the corresponding error in Theorem 1.3 (and communication blown up by a factor of k^2).

Next we state and prove the formal version of Corollary 1.4:

Corollary B.1. *Fix any set $\mathcal{X}$, and let $\mathcal{Q}$ be a k-sparse non-adaptive SQ algorithm on $\mathcal{X}$ making at most B queries of tolerance τ for a distribution $\mathcal{D}$. Suppose that $\varepsilon, \delta \in (0, 1)$,*

$$n \geqslant \Omega\left(\frac{\log(B/\beta)}{\tau^2} + \frac{k\log(B/\beta)\sqrt{\log\left(\frac{k}{\delta\varepsilon}\right)}}{\varepsilon\tau}\right). \tag{93}$$

Then there is a private-coin (ε, δ)-differentially private algorithm P in the shuffled that receives as input n iid samples $x_1, \ldots, x_n \sim \mathcal{D}$, and produces output that agrees with that of $\mathcal{Q}$ with probability at least $1 - \beta$. Moreover, each user sends in expectation $O\left(\frac{k^2\log(k/(\delta\varepsilon))}{\varepsilon^2}\right)$ messages consisting of $O(\log n \log B)$ bits each.

The sample complexity bound (93) improves upon an analogous result for locally differentially private simulation of $\mathcal{Q}$, for which $n \geqslant \tilde{\Omega}\left(\frac{k}{\tau^2\varepsilon^2}\right)$ samples suffice [ENU20, KLN$^+$08, Theorem 5.7]. Moreover, for small k, it is close to what one gets in the central model, namely that $n \geqslant \tilde{\Omega}\left(\frac{1}{\tau^2} + \frac{\sqrt{k}}{\tau\varepsilon}\right)$ [BDNM05] samples suffice. These observations follow from the fact that the ℓ_2 sensitivity of the collection of queries made by $\mathcal{Q}$ is bounded above by $\sqrt{k}$.

Proof of Corollary B.1. Let us fix a set of B queries $q_1, \ldots, q_B$ made by $\mathcal{Q}$. We will show that with probability $1 - \beta$ over the sample $X := (x_1, \ldots, x_n)$, the algorithm P can output real numbers $P_1(X), \ldots, P_B(X)$ so that for $1 \leqslant j \leqslant B$, $|q_j(\mathcal{D}) - P_j(X)| \leqslant \tau$. Thus, conditioned on $\mathcal{Q}$ making the queries $q_1, \ldots, q_B$, P simulates an SQ oracle of tolerance τ with probability $1 - \beta$. The claimed result regarding the accuracy of P follows by taking expectation over $q_1, \ldots, q_B$.

As long as $n \geqslant C \cdot \frac{\log B}{\tau^2}$ for a sufficiently large constant C, the Chernoff-Hoeffding bound guarantees that with probability $1 - \beta/2$, for $1 \leqslant j \leqslant B$, we have $\left|q_j(\mathcal{D}) - \frac{1}{n}\sum_{i=1}^n q_j(x_i)\right| \leqslant \tau/2$. Define a new universe $\mathcal{X}' := \{(q_1(x), q_2(x), \ldots, q_B(x)) : x \in \mathcal{X}\}$. Since $\mathcal{Q}$ is k-sparse, we have that $\mathcal{X}' \subseteq \{v \in \{0, 1\}^B : \|v\|_1 \leqslant k\}$.

Now let P be the protocol P^{Had} with the parameters from Theorem 4.3, and $P(X)_j$, $1 \leqslant j \leqslant B$, be the values $\hat{x}_j/n$ from Theorem 4.3. Then Theorem 4.3 gives that

$$\mathbb{P}\left[\forall j \in [B] : \left|\frac{1}{n}\sum_{i=1}^{n} q_j(x_i) - P(X)_j\right| \leqslant O\left(\frac{\log(B)k\sqrt{\log(k/(\varepsilon\delta))}}{n}\right)\right] \geqslant 1 - \beta/2.$$

By the choice of n in (93), with probability at least $1 - \beta/2$ over P^{Had}, we have $\left|P(X)_j - \frac{1}{n}\sum_{i=1}^{n} q_j(x_i)\right| \leqslant \tau/2$. Thus, with probability at least $1 - \beta$ over P^{Had} and the sample X, we have $|P(X)_j - q_j(\mathcal{D})| \leqslant \tau$ for all $j \in [B]$, as desired. □

Instead of the Hadamard response-based protocol, we could use the (public coin) count-min sketch based protocol P^{CM} of Theorem 4.7 in the above corollary. This would give the inferior sample complexity bound of

$$n \geqslant \Omega\left(\frac{k\log^{3/2}(B)\sqrt{\log(k(\log B)/\delta)}}{\varepsilon\tau}\right),$$

an would involve each user sending in expectation $O\left(\frac{(\log^3 B)k^3\log(k(\log B)/\delta)}{\varepsilon^2}\right)$ messages consisting of $O(\log k + \log n + \log\log B)$ bits each. (Recall that the advantage of the count-min sketch based protocol was efficient computation of a given statistical query in the data-structural setting when B is prohibitively large to compute all of them.)

Notice that the error and communication bounds in Corollary B.1 degrade polynomially in k; thus, for families $\mathcal{Q}$ which do not have any particular sparsity structure and for which $|\mathcal{Q}| \geqslant n$, the bounds of Corollary B.1 are vacuous. In the central model of differential privacy, much effort has gone into determining the optimal sample complexity of simulating in a differentially private manner an arbitrary (not necessarily sparse) non-adaptive statistical query algorithm. For instance, Nikolov et al. [NTZ13] demonstrated a mechanism that achieves sample complexity nearly equal to an efficiently computable lower bound for any differentially private mechanism releasing a fixed set of statistical queries.[20] The earlier work of Hardt and Rothblum [HR10] showed that there is an (ε, δ)-differentially private mechanism that with high probability simulates a non-adaptive SQ algorithm $\mathcal{Q}$ making B queries as long as $n \geqslant \Omega\left(\frac{\log(B)\sqrt{\log|\mathcal{X}|\log(1/\delta)}}{\varepsilon\tau^2}\right)$.

We leave the question of generalizing Corollary B.1 to the case of non-sparse $\mathcal{Q}$ in a way analogous to [NTZ13, HR10] as an interesting question for future work. In particular, we would hope to maintain polylogarithmic-in-B growth of the tolerance and communication, while settling for a number of samples n that grows as $\frac{1}{\varepsilon\tau^\alpha}$ for some $\alpha > 1$.

C Proofs of Auxiliary Lemmas from Section 4

In this section, we prove Lemmas 4.11 and 4.12.

Lemma 4.11. *Suppose $f : \mathcal{X}^n \to \mathbb{Z}^m$ is k-incremental (Definition 4.3) and $\Delta(f) = \Delta$. Suppose $\mathcal{D}$ is a distribution supported on $\mathbb{Z}$ that is (ε, δ, k)-smooth. Then the mechanism*

$$X \mapsto f(X) + (Y_1, \ldots, Y_m),$$

where $Y_1, \ldots, Y_m \sim \mathcal{D}$, i.i.d., is (ε', δ')-differentially private, where $\varepsilon' = \varepsilon \cdot \Delta, \delta' = \delta \cdot \Delta$.

Proof of Lemma 4.11. Consider neighboring datasets $X = (x_1, \ldots, x_{n-1}, x_n)$ and $X' = (x_1, \ldots, x_{n-1}, x'_n)$. We will show

$$\mathbb{P}_{y_1,\ldots,y_m\sim\mathcal{D}}\left[\frac{\mathbb{P}_{Y_1,\ldots,Y_m\sim\mathcal{D}}[f(X) + (Y_1, \ldots, Y_m) = f(X) + (y_1, \ldots, y_m)]}{\mathbb{P}_{Y_1,\ldots,Y_m\sim\mathcal{D}}[f(X') + (Y_1, \ldots, Y_m) = f(X) + (y_1, \ldots, y_m)]} \geqslant e^{\varepsilon'}\right] \leqslant \delta'. \tag{94}$$

[20] This optimality is with respect to mean squared error over the set of queries.

To see that (94) suffices to prove the Lemma 4.11, fix any subset $\mathcal{S} \subset \mathbb{Z}^m$, and write $P(X) = f(X) + (Y_1, \ldots, Y_m)$ to denote the randomized protocol. Let $\mathcal{T}$ denote the set of $f(X) + (y_1, \ldots, y_m) \in \mathbb{Z}^m$ such that the event in (94) does not hold; then we have $\mathbb{P}[P(X) \notin \mathcal{T}] \leqslant \delta'$. It follows that

$$\begin{aligned}
\mathbb{P}\left[P(X) \in \mathcal{S}\right] &\leqslant \delta' + \sum_{w \in \mathcal{T} \cap \mathcal{S}} \mathbb{P}[P(X) = w] \\
&= \delta' + \sum_{w \in \mathcal{T} \cap \mathcal{S}} \mathbb{P}_{Y_1, \ldots, Y_m \sim \mathcal{D}}[f(X) + (Y_1, \ldots, Y_m) = w] \\
&\leqslant \delta' + \sum_{w \in \mathcal{T} \cap \mathcal{S}} e^{\varepsilon'} \mathbb{P}_{Y_1, \ldots, Y_m \sim \mathcal{D}}[f(X') + (Y_1, \ldots, Y_m) = w] \\
&= \delta' + \sum_{w \in \mathcal{T} \cap \mathcal{S}} e^{\varepsilon'} \cdot \mathbb{P}[P(X') = w] \\
&\leqslant \delta' + e^{\varepsilon'} \mathbb{P}[P(X') \in \mathcal{S}].
\end{aligned}$$

It then suffices to show (94). For $j \in [m]$, let $k_j = f(X)_j - f(X')_j$. Since the sensitivity of f is Δ, we have $\sum_{j=1}^m |k_j| \leqslant \Delta$. It follows that (94) is equivalent to

$$\mathbb{P}_{y_1, \ldots, y_m \sim \mathcal{D}} \left[\prod_{j=1}^m \frac{\mathbb{P}_{Y_j \sim \mathcal{D}}[Y_j = y_j]}{\mathbb{P}_{Y_j \sim \mathcal{D}}[Y_j = y_j + k_j]} \geqslant e^{\varepsilon'} \right] \leqslant \delta'. \tag{95}$$

For (95) to hold it in turn suffices, by a union bound and the fact that at most Δ of the k_j are nonzero, that for each j with $k_j \neq 0$,

$$\mathbb{P}_{y \sim \mathcal{D}} \left[\frac{\mathbb{P}_{Y \sim \mathcal{D}}[Y = y]}{\mathbb{P}_{Y \sim \mathcal{D}}[Y = y + k_j]} \geqslant e^{|k_j| \varepsilon'/\Delta} \right] \leqslant \delta'/\Delta. \tag{96}$$

But (96) follows for $\varepsilon'/\Delta = \varepsilon, \delta'/\Delta = \delta$ since $\mathcal{D}$ is (ε, δ, k)-smooth. This completes the proof. □

Lemma 4.12. *Let $n \in \mathbb{N}$, $\gamma \in [0, 1/2]$, $0 \leqslant \alpha \leqslant 1$, and $k \leqslant \alpha\gamma n/2$. Then the distribution $\mathrm{Bin}(n, \gamma)$ is (ε, δ, k)-smooth with $\varepsilon = \ln((1+\alpha)/(1-\alpha))$ and $\delta = e^{-\frac{\alpha^2 \gamma n}{8}} + e^{-\frac{\alpha^2 \gamma n}{8+2\alpha}}$.*

Proof of Lemma 4.12. Recall that for $Y \sim \mathrm{Bin}(n, \gamma)$ and $0 \leqslant y \leqslant n$, we have $\mathbb{P}[Y = y] = \gamma^y (1-\gamma)^{n-y} \binom{n}{y}$. Thus, we have that, for any $k \geqslant k' \geqslant -k$,

$$\frac{\mathbb{P}_{Y \sim \mathrm{Bin}(n,\gamma)}[Y = y]}{\mathbb{P}_{Y \sim \mathrm{Bin}(n,\gamma)}[Y = y + k']} = \frac{(1-\gamma)^{k'}}{\gamma^{k'}} \cdot \frac{(y+k')!(n-y-k')!}{y!(n-y)!}. \tag{97}$$

We define the interval $\mathcal{E} := [(1-\alpha)\gamma n + k', (1+\alpha)\gamma n - k']$ where α is any positive constant smaller than 1. As long as $k' \leqslant \alpha\gamma n/2$, $\mathcal{E}$ contains the interval $\mathcal{E}' := [(1-\alpha/2)\gamma n, (1+\alpha/2)\gamma n]$. By the multiplicative Chernoff Bound, we have that

$$\mathbb{P}_{y \sim \mathrm{Bin}(n,\gamma)}[y \notin \mathcal{E}] \leqslant e^{-\frac{\alpha^2 \gamma n}{8}} + e^{-\frac{\alpha^2 \gamma n}{8+2\alpha}}. \tag{98}$$

Note that for any $y \in \mathcal{E}$, if $k' \geqslant 0$, it is the case that

$$\frac{(1-\gamma)^{k'}}{\gamma^{k'}} \frac{(y+k')!(n-y-k')!}{y!(n-y)!} = \frac{(1-\gamma)^{k'}}{\gamma^{k'}} \cdot \frac{(y+1)\cdots(y+k')}{(n-y)\cdots(n-y-k'+1)} \leqslant (1+\alpha)^{k'}. \tag{99}$$

For $y \in \mathcal{E}$ and if $k' \leqslant 0$, it is the case that

$$\frac{(1-\gamma)^{k'}}{\gamma^{k'}} \frac{(y+k')!(n-y-k')!}{y!(n-y)!} = \frac{\gamma^{|k'|}}{(1-\gamma)^{|k'|}} \cdot \frac{(n-y+1)\cdots(n-y+|k'|)}{y(y-1)\cdots(y-|k'|)} \leqslant \left(\frac{(1-\gamma+\gamma\alpha)}{(1-\gamma)(1-\alpha)} \right)^{|k'|} \leqslant \left(\frac{1+\alpha}{1-\alpha} \right)^{|k'|}, \tag{100}$$

where the last inequality above uses $\gamma \leqslant 1/2$. We now proceed to show smoothness by conditioning on the event $y \in \mathcal{E}$ as follows:

$$\begin{aligned}
&\mathbb{P}_{y\sim\mathrm{Bin}(n,\gamma)}\left[\frac{\mathbb{P}_{Y\sim\mathrm{Bin}(n,\gamma)}[Y=y]}{\mathbb{P}_{Y\sim Bin(n,\gamma)}[Y=y+k']} \geqslant e^{|k'|\varepsilon}\right] \\
&\leqslant \mathbb{P}_{y\sim\mathrm{Bin}(n,\gamma)}\left[\frac{\mathbb{P}_{Y\sim\mathrm{Bin}(n,\gamma)}[Y=y]}{\mathbb{P}_{Y\sim\mathrm{Bin}(n,\gamma)}[Y=y+k']} \geqslant e^{|k'|\varepsilon} \mid y\in\mathcal{E}\right] + \mathbb{P}[y\notin\mathcal{E}] \\
&\leqslant \mathbb{P}_{y\sim\mathrm{Bin}(n,\gamma)}\left[\frac{\mathbb{P}_{Y\sim\mathrm{Bin}(n,\gamma)}[Y=y]}{\mathbb{P}_{Y\sim\mathrm{Bin}(n,\gamma)}[Y=y+k']} \geqslant e^{|k'|\varepsilon} \mid y\in\mathcal{E}\right] + e^{-\frac{\alpha^2\gamma n}{8}} + e^{-\frac{\alpha^2\gamma n}{8+2\alpha}} \qquad (101)\\
&= e^{-\frac{\alpha^2\gamma n}{8}} + e^{-\frac{\alpha^2\gamma n}{8+2\alpha}}, \qquad (102)
\end{aligned}$$

where (101) follows from (98) and (102) follows from (97), (99), and (100) as well as our choice of ε. □

D Heavy Hitters

Let τ denote the heavy hitter threshold, and assume that τ is large enough so that with high probability the maximum frequency estimation error of Theorem 4.7 is at most $\tau/2$. We wish to return a set of $O(n/\tau)$ elements that include all heavy hitters (it may also return other elements, so in that sense this is an approximate answer). One option is to use Theorem 4.7 directly: Iterate over all elements in $[B]$, compute an estimate of each count, and output the elements whose estimate is larger than $\tau/2$. This gives a runtime of $\tilde{O}(B)$.

Algorithm 2 can be combined with the prefix tree idea of Bassily et al. [BNST17] to reduce the server decoding time (for recovering all heavy hitters and their counts up to additive polylogarithmic factors) from $\tilde{O}(B)$ to $\tilde{O}(n/\tau)$. For completeness we sketch the reduction here. The combined algorithm would use $\lceil\log_2(B)\rceil$ differentially private frequency estimation data structures obtained from Algorithm 2. To make the whole data structure differentially private we decrease the privacy parameters, such that each data structure is $(\varepsilon/\lceil\log_2(B)\rceil, \delta/\lceil\log_2(B)\rceil)$-differentially private. (In turn, this increases the error and the bound on how small τ can be by a polylogarithmic factor in B.)

For each element $x \in [B]$ we would consider the prefixes of the binary representation of x, inserting the length-i prefix in the ith frequency estimation data structure. The decoding procedure iteratively identifies the prefixes of length 1, 2, 3, ... with a true count of at least τ. With high probability this is a subset of the prefixes that have an estimated count of at least $\tau/2$, and there are $O(n/\tau)$ such prefixes at each level. When a superset of these "heavy" prefixes have been determined at level i, we only need to estimate the frequencies of the two length-$(i+1)$ extensions of each considered prefix. This reduces the server decoding time to $\tilde{O}(n/\tau)$ with high probability (while maintaining a polylogarithmic bound on the number of bits of communication per user).

E M-Estimation of Median and Quantiles

We now discuss how to obtain results for M-estimation of the median and quantiles using our result for range counting (from Section 5). For simplicity, let $x_1, x_2, \ldots, x_n \in [0,1]$ be data points held by n users.

We recall that there is no differentially private algorithm for estimating the value of the true median with error $o(1)$, i.e., computing $\tilde{x}$ which is within additive error $o(1)$ of the true median. This is because the true median can be highly sensitive to a single data point, which precludes the possibility of outputting a close approximation of the median without revealing much information about a single user. For instance, consider the case in which $n = 2k+1$ and $x_1 = x_2 = \cdots = x_k = 0$ while $x_{k+1} = x_{k+2} = \cdots = x_n = 1$. It is clear that the median of this set is 0, but changing a single value x_k to 1 would change the median of the set to 1.

To get around the above limitations, we consider a different notion known as *M-estimation* of the median, defined as follows: Consider the function

$$M(y) = \frac{1}{2}\sum_{i=1}^{n} |x_i - y|.$$

Note that the median is a value $\tilde{x}$ that minimizes the quantity $M(\tilde{x})$. The problem of M-estimation seeks to compute a value of y which approximates this quantity, and the error is considered to be the additive error between $M(y)$ and $M(\tilde{x})$. Our results imply a differentially private multi-message protocol for M-estimation that obtains both error $\operatorname{poly}\log n$ and communication per user $\operatorname{poly}\log n$ bits.

Theorem E.1 (Multi-message protocol for M-estimating the median)**.** *Suppose $x_1, x_2, \ldots, x_n \in [0,1]$. Then there is a differentially private multi-message protocol in the shuffled model that M-estimates the median of $x_1, x_2, \ldots, x_n$ with communication $\operatorname{poly}\log n$ bits per user and additive error $\operatorname{poly}\log n$, i.e., outputs $y \in [0,1]$ such that*

$$M(y) \leqslant \min_{\tilde{x}} M(\tilde{x}) + \operatorname{poly}\log n.$$

Proof. We reduce the problem of M-estimation to range counting. First, we divide the interval $[0,1]$ into $B = n$ subintervals $I_1, I_2, \ldots, I_B$, where $I_j = [(j-1)/B, j/B]$. Each user will associate his element x_i with an index $z_i \in [B]$ corresponding to an interval I_{z_i} which contains x_j. Note that if j is the smallest element of $[B]$ such that $|[1,j] \cap \{z_1, z_2, \ldots, z_n\}| \geqslant n/2$, then I_j contains a minimizer of $M(y)$. Thus, we wish to determine this value of j.

Thus, we obtain a protocol as follows: We use our protocol for range counting queries in the shuffled model (see Section 5) to compute the queries $[1,j]$ for $j = 1, 2, \ldots, B$ for the dataset $z_1, z_2, \ldots, z_B$ and compute the first j for which the query $[1,j]$ yields a count of $\geqslant n/2$ (or $j = B$ if no query yields such a count). Then the analyzer outputs j/B as the estimate for the median.

We now determine the error of thee aforementioned protocol. Note that by the guarantees of the range counting protocol, the error for the range counts is $\operatorname{poly}\log B$. This results in a corresponding $\operatorname{poly}\log B$ additive error due to range counting queries for the estimation of $M(\tilde{x})$. Moreover, note that there is an additional error resulting from the discretization. Since each interval is of length $1/B$, the error resulting from discretization is n/B. Hence, the total error is $n/B + \operatorname{poly}\log B = \operatorname{poly}\log n$ for our choice of B. □

Remark E.1. It should be noted that virtually the same argument as above yields a differentially private protocol for M-estimation of *quantiles*. Given a set of points $x_1, x_2, \ldots, x_n \in [0,1]$, we say that y is a k^{th} q-quantile of the dataset if

$$|[0,y) \cap \{x_1, x_2, \ldots, x_n\}| \leqslant \frac{k}{q}$$

and

$$|[0,y] \cap \{x_1, x_2, \ldots, x_n\}| \geqslant \frac{k}{q}.$$

In particular, the median is a special case, namely, the (only) q-quantile for $q = 2$. The above argument applies, except that the function M to be minimized (which is minimized by k^{th} q-quantiles) is given by

$$M(y) = \sum_{i=1}^{n} \left(\left(1 - \frac{k}{q}\right)(y - x_i)_+ + \frac{k}{q}(y - x_i)_- \right),$$

and again, the task is to find a y such that

$$M(y) \leqslant \min_{\tilde{x}} M(\tilde{x}) + \operatorname{poly}\log n.$$

Moreover, in the reduction to range counting queries, one instead determines the smallest value j such that $|[1,j] \cap \{z_1, z_2, \ldots, z_n\}| \geqslant kn/q$ and the rest of the analysis follows verbatim.

References

[ACFT19] Jayadev Acharya, Clément Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In *AISTATS*, pages 2067–2076, 2019.

[App17] Apple Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 2017.

[AS19] Jayadev Acharya and Ziteng Sun. Communication complexity in locally private distribution estimation and heavy hitters. In *ICML*, pages 97:51–60, 2019.

[ASY+18] Naman Agarwal, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Brendan McMahan. cpsgd: Communication-efficient and differentially-private distributed sgd. In *Advances in Neural Information Processing Systems*, pages 7564–7575, 2018.

[ASZ19] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *AISTATS*, pages 1120–1129, 2019.

[BBGN19a] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Differentially private summation with multi-message shuffling. *CoRR*, abs/1906.09116, 2019.

[BBGN19b] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Improved summation from shuffling. *arXiv: 1909.11225*, 2019.

[BBGN19c] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. The privacy blanket of the shuffle model. In *CRYPTO*, pages 638–667, 2019.

[BBGN20] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Private summation in the multi-message shuffle model. *arXiv:2002.00817*, 2020.

[BC19] Victor Balcer and Albert Cheu. Separating local & shuffled differential privacy via histograms, 2019.

[BDNM05] Avrim Blum, Cynthia Dwork, Kobbi Nissim, and Frank McSherry. Practical privacy: the SuLQ framework. In *PODS*, pages 128–138, 2005.

[BEM+17] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnés, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *SOSP*, pages 441–459, 2017.

[Ben79] Jon Louis Bentley. Decomposable searching problems. *IPL*, 8(5):244–251, 1979.

[BLM12] Stephane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: a nonasmpytotic theory of independence*. Clarendon Press, Oxford, 2012.

[BLR08] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *STOC*, pages 609–618, 2008.

[BNS13] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In *APPROX-RANDOM*, pages 363–378, 2013.

[BNS18] Mark Bun, Jelani Nelson, and Uri Stemmer. Heavy hitters and the structure of local privacy. In *PODS*, pages 435–447, 2018.

[BNST17] Raef Bassily, Kobbi Nissim, Uri Stemmer, and Abhradeep Guha Thakurta. Practical locally private heavy hitters. In *NIPS*, pages 2288–2296, 2017.

[BNSV15] Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially private release and learning of threshold functions. In *FOCS*, pages 634–649, 2015.

[BS15] Raef Bassily and Adam Smith. Local, private, efficient protocols for succinct histograms. In *STOC*, pages 127–135, 2015.

[BST14] Raef Bassily, Adam D. Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *FOCS*, pages 464–473, 2014.

[CCFC02] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. In *ICALP*, pages 693–703, 2002.

[CDN15] Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen. *Secure Multiparty Computation*. Cambridge University Press, 2015.

[CH08] Graham Cormode and Marios Hadjieleftheriou. Finding frequent items in data streams. *VLDB*, 1(2):1530–1541, 2008.

[CKS18] Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. Marginal release under local differential privacy. In *SIGMOD*, pages 131–146, 2018.

[CKS19] Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. Answering range queries under local differential privacy. In *Proceedings of International Conference on Management of Data (SIGMOD)*, page 18321834, 2019.

[CM05a] Graham Cormode and Shan Muthukrishnan. An improved data stream summary: The Count-Min sketch and its applications. *Journal of Algorithms*, 55(1):58–75, 2005.

[CM05b] Graham Cormode and Shan Muthukrishnan. What's hot and what's not: tracking most frequent items dynamically. *TODS*, 30(1):249–278, 2005.

[CM08] Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *NIPS*, pages 289–296, 2008.

[CMS11] Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *JMLR*, 12:1069–1109, 2011.

[Cor11] Graham Cormode. Sketch techniques for approximate query processing. *Foundations and Trends in Databases. NOW publishers*, 2011.

[CPS+12] Graham Cormode, Cecilia Procopiuc, Divesh Srivastava, Entong Shen, and Ting Yu. Differentially private spatial decompositions. In *ICDE*, pages 20–31, 2012.

[CSS11] T.-H. Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Trans. Inf. Syst. Secur.*, 14(3):26:1–26:24, 2011.

[CSS12] T-H. Hubert Chan, Elaine Shi, and Dawn Song. Optimal lower bound for differentially private multi-part aggregation. In *European Symposium on Algorithms*, 2012.

[CSS13] Kamalika Chaudhuri, Anand D. Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *JMLR*, 14(1):2905–2943, 2013.

[CSU+19] Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via mixnets. In *EUROCRYPT*, pages 375–403, 2019.

[CT91] Thomas A. Cover and Joy M. Thomas. *Elements of Information Theory*. Wiley, 1991.

[CY20] Graham Cormode and Ke Yi. *Small Summaries for Big Data*. Cambridge University Press, 2020.

[DJW13] John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *FOCS*, pages 429–438, 2013.

[DJW18] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *JASA*, 113(521):182–201, 2018.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006.

[DKY17] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *NIPS*, pages 3571–3580, 2017.

[DL09] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *STOC*, pages 371–380, 2009.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[DNPR10] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *STOC*, pages 715–724, 2010.

[DNR+09] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N Rothblum, and Salil Vadhan. On the complexity of differentially private data release: Efficient algorithms and hardness results. In *STOC*, pages 381–390, 2009.

[DNRR15] Cynthia Dwork, Moni Naor, Omer Reingold, and Guy N Rothblum. Pure differential privacy for rectangle queries via private partitions. In *ASIACRYPT*, pages 735–751, 2015.

[DR14a] Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Now Publishers Inc., 2014.

[DR+14b] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[Dwo06] Cynthia Dwork. Differential privacy. In *ICALP*, pages 1–12, 2006.

[EFM+19] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *SODA*, pages 2468–2479, 2019.

[EFM+20] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Shuang Song, Kunal Talwar, and Abhradeep Thakurta. Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. *arXiv preprint arXiv:2001.03618*, 2020.

[ENU20] Alexander Edmonds, Aleksander Nikolov, and Jonathan Ullman. The power of factorization methods in local and central differential privacy. In *Symposium on the Theory of Computing*, 2020.

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *CCS*, pages 1054–1067, 2014.

[EV03] Cristian Estan and George Varghese. New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice. *TOCS*, 21(3):270–313, 2003.

[GGI+02] Anna C Gilbert, Sudipto Guha, Piotr Indyk, Yannis Kotidis, Sivaramakrishnan Muthukrishnan, and Martin J Strauss. Fast, small-space algorithms for approximate histogram maintenance. In *STOC*, pages 389–398, 2002.

[GGK+20] Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Pure differentially private summation from anonymous messages. In *Information Theoretic Cryptography (ITC)*, 2020.

[GK+01] Michael Greenwald, Sanjeev Khanna, et al. Space-efficient online computation of quantile summaries. *ACM SIGMOD Record*, 30(2):58–66, 2001.

[GMPV19] Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. *arXiv:1909.11073*, 2019.

[GPV19] Badih Ghazi, Rasmus Pagh, and Ameya Velingker. Scalable and differentially private distributed aggregation in the shuffled model. *arXiv:1906.08320*, 2019.

[Gre16] Andy Greenberg. Apple's "differential privacy" is about collecting your data – but not your data. *Wired, June*, 13, 2016.

[HKR12] Justin Hsu, Sanjeev Khanna, and Aaron Roth. Distributed private heavy hitters. In *ICALP*, pages 461–472, 2012.

[HLM12] Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. In *NIPS*, pages 2339–2347, 2012.

[HR10] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *FOCS*, pages 61–70, 2010.

[HRMS10] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the accuracy of differentially private histograms through consistency. *VLDB*, 3(1-2):1021–1032, 2010.

[IKOS06] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography from anonymity. In *FOCS*, pages 239–248, 2006.

[KBR16] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *ICML*, pages 2436–2444, 2016.

[KLL16] Zohar Karnin, Kevin Lang, and Edo Liberty. Optimal quantile approximation in streams. In *FOCS*, pages 71–78, 2016.

[KLN+08] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Rashkodnikova, and Adam Smith. What can we learn privately? In *FOCS*, pages 531–540, 2008.

[KMSZ08] Joe Kilian, André Madeira, Martin J Strauss, and Xuan Zheng. Fast private norm estimation and heavy hitters. In *TCC*, pages 176–193, 2008.

[Lei11] Jing Lei. Differentially private m-estimators. In *NIPS*, pages 361–369, 2011.

[LHR+10] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Milau, and Andrew McGregor. Optimizing linear counting queries under differential privacy. In *PODS*, pages 123–134, 2010.

[LLV07] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE*, pages 106–115, 2007.

[LM12] Chao Li and Gerome Miklau. An adaptive mechanism for accurate query answering under differential privacy. In *VLDB*, volume 5(6), pages 514–525, 2012.

[MG82] Jayadev Misra and David Gries. Finding repeated elements. *Science of Computer Programming*, 2(2):143–152, 1982.

[MN12] S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *STOC*, pages 1285–1292, 2012.

[MP80] J Ian Munro and Mike S Paterson. Selection and sorting with limited storage. *TCS*, 12(3):315–323, 1980.

[MRL98] Gurmeet Singh Manku, Sridhar Rajagopalan, and Bruce G Lindsay. Approximate medians and other quantiles in one pass and with limited memory. *ACM SIGMOD Record*, 27(2):426–435, 1998.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

[NTZ13] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. On the geometry of differential privacy: the sparse and approximate cases. In *STOC*, pages 351–360, 2013.

[NXY+16] Thong Nguyen, Xiaokui Xiao, Yin Yang, Sui Cheung Hui, Hyejin Shin, and Junbum Shin. Collecting and analyzing data from smart device users with local differential privacy. In *arXiv:1606.05053*, 2016.

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[QYL13] Wahbeh Qardaji, Weining Yang, and Ninghui Li. Understanding hierarchical methods for differentially private histograms. *VLDB*, 6(14):1954–1965, 2013.

[Roo06] Bero Roos. Binomial approximation to the Poisson binomial distribution: The Krawtchouk Expansion. *Theory of Probability and its Applications*, 45(2):258–272, 2006.

[Sha14] Stephen Shankland. How Google tricks itself to protect Chrome user privacy. *CNET, October*, 2014.

[Smi11] Adam D. Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *STOC*, pages 813–822, 2011.

[Ste20] Uri Stemmer. Locally private k-means clustering. In *Proceedings of the 2020 Symposium on Discrete Algorithms*, 2020.

[SU16] Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7(2):3–22, 2016.

[SU17] Thomas Steinke and Jonathan Ullman. Tight lower bounds for differentially private selection. In *FOCS*, pages 552–563, 2017.

[Ull18] Jonathan Ullman. Tight lower bounds for locally differentially private selection. In *arXiv:1802.02638*, 2018.

[Vad17] Salil Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer, 2017.

[War65] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *JASA*, 60(309):63–69, 1965.

[WBLJ17] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *USENIX Security*, pages 729–745, 2017.

[WXD+19] Tianhao Wang, Min Xu, Bolin Ding, Jingren Zhou, Ninghui Li, and Somesh Jha. Practical and robust privacy amplification with multi-party differential privacy. *arXiv:1908.11515*, 2019.

[WZ10] Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *JASA*, 105(489):375–389, 2010.

[XWG10] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. Differential privacy via wavelet transforms. *TKDE*, 23(8):1200–1214, 2010.

[YB17] Min Ye and Alexander Barg. Optimal schemes for discrete distribution estimation under local differential privacy. In *ISIT*, pages 759–763, 2017.

[YZ13] Ke Yi and Qin Zhang. Optimal tracking of distributed heavy hitters and quantiles. *Algorithmica*, 65(1):206–223, 2013.