On the Power of Multiple Anonymous Messages: Frequency Estimation and Selection in the Shuffle Model of Differential Privacy

Badih Ghazi[1], Noah Golowich[2⋆], Ravi Kumar[1], Rasmus Pagh[3⋆⋆], and Ameya Velingker[1]

[1] Google Research
badihghazi@gmail.com, ravi.k53@gmail.com, ameyav@google.com
[2] MIT EECS
nzg@mit.edu
[3] BARC and University of Copenhagen
pagh@di.ku.dk

Abstract. It is well-known that general secure multi-party computation can in principle be applied to implement differentially private mechanisms over distributed data with utility matching the curator (a.k.a. central) model. In this paper we study the power of protocols running on top of a much weaker primitive: A non-interactive anonymous channel, known as the *shuffle* model in the differential privacy literature. Such protocols are implementable in a scalable way using known cryptographic methods and are known to enable non-interactive, differentially private protocols with error much smaller than what is possible in the local model. We study fundamental counting problems in the shuffle model and obtain tight, up to polylogarithmic factors, bounds on the error and communication in several settings.

For the classic 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 ℓ_∞ error in the *single-message* shuffle 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 nearly optimal for single-message protocols.
- Protocols in the *multi-message* shuffle model with $\mathrm{poly}(\log B, \log n)$ bits of communication per user and ℓ_∞ error at most $\mathrm{poly}(\log B, \log n)$, 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.

⋆ This work was done while interning at Google Research. Supported at MIT by a Fannie & John Hertz Foundation Fellowship and an NSF Graduate Fellowship.

⋆⋆ This work was initiated while visiting Google Research. Supported by VILLUM Foundation grant 16582.

For the *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 shuffle model. This significantly improves on the $\Omega(B^{1/17})$ lower bound obtained by Cheu et al. (Eurocrypt 2019).

A key ingredient in our lower bound proofs is a lower bound on the error of *locally*-private frequency estimation in the low-privacy (a.k.a. high ε) regime. For this we develop new tools to improve the results of Duchi et al. (FOCS 2013; JASA 2018) and Bassily & Smith (STOC 2015), whose techniques only gave tight bounds in the high-privacy setting.

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., [76]), *differential privacy (DP)* [46,44] has emerged as the gold standard due to its broad applicability and nice features such as composition and post-processing (see, e.g., [51,93] 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., [18,49,50]). 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 [71] (also [97]), 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 [62,5], Google [55,87], and Microsoft [40], 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.

Shuffle Privacy Model. The aforementioned trade-offs have motivated the study of the *shuffle* 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. [68] on cryptography from anonymity, the shuffle model was first proposed for privacy-preserving protocols by Bittau et al. [16] in their Encode-Shuffle-Analyze architecture. In the shuffle setting, each user sends one or more

messages to the analyzer using an *anonymous* channel that does not reveal where each message comes from. Such 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 for a formal description of the shuffle 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., [68,16]). A comprehensive overview of recent work on anonymous communication can be found on Free Haven's Selected Papers in Anonymity website [57].

The DP properties of the shuffle model were first analytically studied, independently, in the works of Erlingsson et al. [54] and Cheu et al. [29]. Protocols within the shuffle 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 shuffle model with as small error and total communication as possible. An example of the power of the shuffle model was established by Erlingsson et al. [54] and extended by Balle et al. [9], who showed that every locally differentially private algorithm directly yields a single-message protocol in the shuffle model with significantly better privacy. In this paper we study the optimal error achievable for fundamental tasks such as frequency estimation (i.e., histograms) and selection in the shuffle model of differential privacy. We show that in many settings, multi-message protocols can achieve significantly smaller error than single-message protocols, and we introduce such low-error multi-message protocols that have the additional property of having low communication.

The study of differential privacy in the shuffle model can be seen as part of a movement towards an integrated study of differential privacy and cryptographic protocols, i.e., "DP-cryptography" [94].

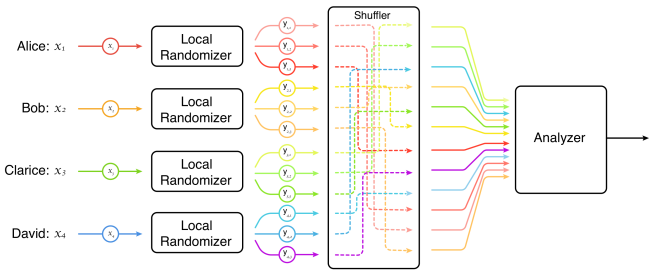

Fig. 1: Computation in the shuffle model consists of local randomization of inputs in the first stage, followed by a shuffle of all outputs of the local randomizers, after which the shuffled output is passed on to an analyzer.

Overview. The remainder of the paper is organized as follows. In Section 2 we review some preliminaries for differential privacy and the shuffle model. In Section 3 we give an overview of our main theorems for the frequency estimation and selection problems, and in Section 4 we overview the proofs of our main results. In Section 5 we discuss applications of our results to problems such as range queries and median estimation. In Section 6 we discuss related work in detail, and we conclude in Section 7. Full proofs of our results as well as the precise statements of some theorems are relegated to the supplementary material; see Section A.

2 Preliminaries

Before stating our main results, we formally introduce the basics of differential privacy and the shuffle model.

Notation. 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. 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., $\mathcal{Y}^* = \bigcup_{n \geqslant 0} \mathcal{Y}^n$. For positive integers n, B, we write $\mathrm{polylog}(n, B)$ to denote the class of functions $f(n, B)$ for which there is a constant C so that for all $n, B \in \mathbb{N}$, $f(n, B) \leqslant C(\log(nB))^C$.

Datasets. 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 $\mathcal{S}$).

Differential Privacy. 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 $X' = (x_1, \ldots, x_{n-1}, x'_n)$, for $x_1, \ldots, x_n, x'_n \in \mathcal{X}$. 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 21 (Differential privacy [46,44]) *A randomized algorithm $P : \mathcal{X}^n \to \mathcal{Z}$ is (ε, δ)-differentially private (DP) 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 1 (Post-processing, e.g., [50]). *If P is (ε, δ)-differentially private, then for every randomized function A, the composed function $A \circ P$ is (ε, δ)-differentially private.*

Shuffle Model. We review the *shuffle model* of differential privacy [16,54,29]. 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 shuffle 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. In the *single-message* shuffle model, we require $m_i = 1$ for each i; in the *multi-message* shuffle model, m_i may be any 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 shuffle and local model is the introduction of the (trusted) shuffler S between the local randomizer and the analyzer. Similar to the local model, in the shuffle 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 22 (Differential privacy in the shuffle model, [54,29]). *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 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 23, is a variant of the shuffle model where the shuffler S is replaced by the identity function.

Definition 23 (Local differential privacy [71]). *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 21. 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 Overview of Results

In this work, we study several basic problems related to *counting* in the shuffle 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 estimation, and range counting and study whether it is possible to obtain (ε, δ)-DP in the shuffle model with accuracy close to what is possible in the central model, while keeping communication low. This section contains an overview of our main results.

The *frequency estimation* problem (also known as computing *histograms*) is at the core of many of the problems we study. In the simplest version, for some positive integer B, each of n users gets an element of the domain $\mathcal{X} := [B]$, and the goal is to estimate the number of users in a dataset X holding element j, namely $\mathrm{hist}(X)_j$, for each query element $j \in [B]$. We study frequency estimation with the ℓ_∞ *error*, meaning that we define the error of a frequency estimation protocol to be the the maximum additive error for the frequency estimate of any coordinate j. In particular, if $\hat{f} \in \mathbb{R}^B$ is a vector of frequency estimates for a dataset X, then the ℓ_∞ error is $\|\mathrm{hist}(X) - \hat{f}\|_\infty = \max_{j \in [B]} |\mathrm{hist}(X)_j - \hat{f}_j|$. Frequency estimation is a fundamental primitive that is used in various data structural, sketching, and streaming applications (see Section 5 for its use in the shuffled protocols for range counting and median estimation as well as Section 6 for a sample of related work on the problem). 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., [93, 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 $\delta < o(1/n)$ [12] (this regime for δ covers all values for δ of interest in the setting of differential privacy).[4]

In the high-level exposition of our results given below, we let n and B be any positive integers. We typically take $\varepsilon > 0$ to be any constant, and $\delta > 0$ to be inverse polynomial in n. This assumption on ε and δ covers a regime of parameters that is relevant in practice. We will make use of tilde notation (e.g., $\tilde{O}$, $\tilde{\Theta}$) to indicate the suppression of multiplicative factors that are polynomial in $\log B$ and $\log n$. Theorem statements which do not make such assumptions and contain full dependence on all parameters may be found in the supplementary material.

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

Theorem 1 (Informal version of Theorems 5 & 7). *Any $(O(1), o(1/n))$-differentially private frequency estimation protocol in the single-message shuffle model has expected ℓ_∞ error $\tilde{\Omega}(\min(\sqrt[4]{n}, \sqrt{B}))$. Moreover, there is a single-message $(O(1), o(1/n))$-differentially private protocol with error $\tilde{O}(\min(\sqrt[4]{n}, \sqrt{B}))$.*

[4] Most of the large-scale deployments of local DP in practice (e.g., [5,55]) have been variants of frequency estimation protocols.

The main contribution of Theorem 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 follows by combining the recent result of Balle et al. [9] (building on the earlier result of Erlingsson et al. [54]) with RAPPOR [55] and B-ary randomized response [97] (see Section 4.1 and Section C for more details).

The precise version of Theorem 1 with polylogarithmic factors (i.e., Theorem 5) 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 2. This improves on a result of Cheu et al. [29, Corollary 32], which gives a lower bound of $n = \omega(\log^{1/17} B)$ for this task.

Multi-Message Protocols for Frequency Estimation. Theorem 1 implies that in the single-message shuffle model, the error has to grow polynomially with $\min(n, B)$, even with unbounded communication (i.e., message length). We next present (non-interactive) *multi-message* protocols in the shuffle model of DP for frequency estimation with only *polylogarithmic* error and communication. One of the protocols is a *public-coin* protocol, meaning that it makes use of a source of public randomness (known to all parties, including the adversary); the other protocol is a *private-coin* protocol, meaning that no such assumption is made. 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.[5]

Theorem 2 (Informal version of Theorems 15 & 16). *There are private-coin and public-coin multi-message $(O(1), 1/n^{O(1)})$-DP protocols in the shuffle model for frequency estimation satisfying the following:*

- *The private-coin protocol has ℓ_∞ error $O(\max\{\log B, \log n\})$, total communication of $O(\log B \log^2 n)$ bits per user, and query time $\tilde{O}(n)$.*
- *The public-coin protocol has ℓ_∞ error $O(\log^{3/2}(B)\sqrt{\log(n\log(B))})$, total communication of $O(\log^4(B)\log^2(n))$ bits per user, and query time $O(\log B)$.*

Combining Theorems 1 and 2 yields the first separation between single-message and multi-message protocols for frequency estimation. Moreover, Theorem 2 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 quantile estimation), discussed in Section 5. Finally, Theorem 2 implies the following consequence for statistical query (SQ) algorithms with respect to a distribution $\mathcal{D}$ on $\mathcal{X}$ (see Section G 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

[5] The analyzers for both protocols in Theorem 2 have pre-processing time $\tilde{O}(n)$ on the output of the shuffler. In the regime $B \gg n$ (which is often of interest), this running time precludes them from computing all frequencies up-front.

	Local		Local + shuffle	Shuffle, single-message	Shuffle, multi-message	Central
Expected max. error	$\tilde{O}(\sqrt{n})$	$\tilde{\Omega}(\sqrt{n})$	$\tilde{O}(\min(\sqrt[4]{n}, \sqrt{B}))$	$\tilde{\Omega}(\min(\sqrt[4]{n}, \sqrt{B}))$	polylog(n, B)	polylog(n, B)
Comm. per user	$\Theta(1)$	any	$O(B)$ (err $\sqrt[4]{n}$) $\log B$ (err $\sqrt{B}$)	any	polylog(n, B)	n.a.
References	[11]	[12]	[97,55,9]	Thms. 7 & 5	Thm. 15	[78,90]

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.

is drawn i.i.d. from $\mathcal{D}$, the algorithm $\mathcal{A}$ can be efficiently simulated in the shuffle model as follows:

Corollary 1 (Informal version of Corollary 4). *For any non-adaptive k-sparse SQ algorithm $\mathcal{A}$ with B queries of tolerance $\tau > 0$ and any $\beta \in (0,1)$, there is a (private-coin) shuffle 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 improves upon the simulation of non-adaptive SQ algorithms in the *local model* [71], 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 shuffle model. We emphasize that the main novelty of Corollary 1 is in the regime that $k^2/\varepsilon^2 \ll B$; in particular, though prior work on low-communication private summation in the shuffle model [29,59,10] implies an algorithm for simulating $\mathcal{A}$ with roughly the same bound on the number of users n as in Corollary 1 and communication $\Omega(B)$, it was unknown whether the communication could be reduced to have logarithmic dependence on B, as in Corollary 1.

Single-Message Bounds for Selection. The techniques that we develop to prove the lower bound in Theorem 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[6], 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

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

index $j^* \in [B]$ such that

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

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 [41,90,92] and the references therein). Our work improves an $\Omega(B^{1/17})$ lower bound on n in the single-message shuffle model due to Cheu et al. [29]. For $\varepsilon = 1$, the exponential mechanism [78] 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 [92].

Theorem 3 (Informal version of Theorem 11). *For any single-message $(O(1), o(1/(nB)))$-DP protocol in the shuffle model that solves the selection problem given in Equation (1), the number n of users should be $\Omega(B)$.*

The lower bound in Theorem 3 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 shuffle model) and that uses the B-randomized response [97,92]. Cheu et al. [29] have previously obtained a multi-message protocol for selection with $O(\sqrt{B})$ users, and combined with this result Theorem 3 yields the first separation between single-message and multi-message protocols for selection.

In subsequent work Chen et al. [28] have extended Theorem 3 to the setting when each user only sends *few* messages; in particular, they show that if each user sends at most m messages in the shuffle model, then the number of users should be $\Omega(B/m)$. Their proof uses generally similar techniques to ours.

4 Proof outlines

4.1 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 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[7] when $\varepsilon_L = \ln(n) + O(1)$. In fact, we prove lower bounds for $(\varepsilon_L, \delta_L)$-locally differentially protocols for a broader range of parameters ε_L, δ_L in Theorem 6; a special case of this result which includes the setting $\varepsilon_L = \ln(n) + O(1)$ relevant for the shuffle model is stated below:

[7] 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 shuffle model.

Theorem 4 (Local DP lower bound; informal version of Theorem 6). *Suppose that* $\varepsilon_L, \delta_L > 0$ *satisfy*

$$\frac{2}{3} \cdot \ln n \leqslant \varepsilon_L + \ln(1+\varepsilon_L) \leqslant \min\left\{2\ln(B) - O(1), 2\ln(n) - 2\ln\ln(B)\right\},$$

and $\delta_L < o\left(\min\left\{\frac{1}{n \ln n}, \exp(-\varepsilon_L)\right\}\right)$. *Then any* $(\varepsilon_L, \delta_L)$*-locally differentially private protocol for frequency estimation on* $[B]$ *must have* ℓ_∞ *error at least* $\tilde{\Omega}\left(\frac{\sqrt{n}}{e^{\varepsilon_L/4}}\right)$, *where the tilde hides factors polynomial in* $\log B, \log n$.

While lower bounds for local DP frequency estimation were previously obtained in the seminal works of Bassily and Smith [12] and Duchi, Jordan and Wainwright [42], two critical reasons make them less useful for our purposes: (i) for $\varepsilon_L = \omega(1)$ (i.e., in the low-privacy regime) they only apply to the case where $\delta_L = 0$ (i.e., pure privacy)[8], and (ii) even for $\delta_L = 0$, their dependence on ε_L is sub-optimal when $\varepsilon_L = \omega(1)$: the results of [42], for instance, imply a lower bound of $\Omega\left(\frac{\sqrt{n \log B}}{e^{\varepsilon_L}}\right)$ on the ℓ_∞ error.[9] By contrast, Theorem 4 covers the *low and approximate privacy* regime; we next discuss its proof.

Let R be an $(\varepsilon_L, \delta_L)$-locally differentially private randomizer. The general approach in the proof of Theorem 4, which was also taken in [12,42], 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 some appropriate choice of $\alpha \in (0,1)$, and is drawn uniformly at random from $[B]$ otherwise, then the mutual information between V and the local randomizer output $R(X)$ satisfies

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

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 [38] 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.

[8] As we discuss in Remark 1, generic reductions [29,20] 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.

[9] If we were to ignore the assumption of $\delta_L = 0$ and try to use this bound for $\varepsilon_L = \ln(n) + O(1)$ to attempt to derive a lower bound in the single-message shuffle model in the context of Theorem 1, we would get a lower bound of $\Omega(\sqrt{\log(B)/n})$ on the ℓ_∞ error, which for $n \gg \log B$ is (much) worse than even the lower bound of $\Omega(\min\{\log B, \log n\})$ from the *central* model.

To prove the lower bound of Theorem 4 using this approach, we choose $\alpha n = \tilde{\Theta}(\sqrt{n}/e^{\varepsilon_L/4})$, and show that

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

For the application to the single-message shuffle model, we will have $\varepsilon_L = \ln(n) + O(1)$ and so $\alpha = \tilde{\Theta}(n^{-3/4})$; 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 [12, 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 $\delta_L = 0$, Duchi et al. [42] generalized this result to the case $\varepsilon_L \geqslant 1$, proving that[10] $I(V; R(X)) \leqslant O(\alpha^2 e^{2\varepsilon_L})$. Even ignoring the constraint $\delta_L = 0$, this bound of [42] is 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 particular, 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 $\delta_L \leqslant n^{-O(1)}$. For instance, it is violated if one takes R to be R_{RR}, the local randomizer of the B-randomized response [97]. 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 $\alpha \ll 1/n^{2/3}$, which is the regime we have to consider in order to prove Theorem 1[11], it holds that $\alpha \gg \alpha^4 n \exp(\ln(n))$, thus contradicting (3). (See also Remark 4 for an explanation of how a slightly different strategy also fails.) The insight derived from this counterexample is 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 must have error at least $\Omega(\sqrt{B})$ when combined with $R_{\mathrm{RR}}(\cdot)$, which is larger than αn, the error that would be ruled out by the subsequent application of Fano's inequality. This leads 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:

[10] This bound is not stated explicitly in [42], though [42, Lemma 7] proves a similar result whose proof can readily be modified appropriately.

[11] i.e., we will take $\alpha n = \tilde{\Theta}(n^{1/4})$, so $\alpha = \tilde{\Theta}(n^{-3/4})$.

(i) Proving a structural property satisfied by the randomizer corresponding to any accurate frequency estimation protocol. Namely, we show in Lemma 10 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 to 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 U be a random variable uniformly distributed on $[B]$. Assume that the total variation distance $\Delta(R(v), R(U))$ 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 histogram of 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 U (with parameter $q := \mathbb{P}[R(U) = 1]$). Then, we use the anti-concentration properties of binomial random variables in order to argue that if $|p - q| = \Delta(R(v), R(U))$ 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 13). The full proof appears in Lemma 10.

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.

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.[12] On a high level, it turns out that the mutual

[12] 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 B.3.

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 not enough inputs $v \in [B]$ such that the mass $\mathbb{P}[R(X) \in \mathcal{Y}_v]$ is large. Intuitively, for v so that $\mathbb{P}[R(X) \in \mathcal{Y}_v]$ is large, the local randomizer "effectively hides" the fact that the uniform input X is v given that X indeed equals v and $R(v) \in \mathcal{Y}_v$.

(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 for all v. By putting these together, we can conclude that event (a) does not occur (see Lemma 10 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 B.3 and Lemma 9.

Having established Theorem 4 giving a lower bound for locally differentially private estimation in the low-privacy regime, Theorem 1 follows in a straightforward manner: the only step is to apply a lemma of Cheu et al. [29] (restated as Lemma 2 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 shuffle model (i.e., we take $\varepsilon_L = \varepsilon + \ln(n)$). Indeed, for $\varepsilon_L = \ln(n) + O(1)$, the error lower bound from Theorem 4 is $\tilde{\Omega}(\sqrt{n}/e^{\varepsilon_L/4}) = \tilde{\Omega}(n^{1/4})$. Finally, we point out that while the above outline assumed that $n \leqslant B^2$, 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 6).

Tightness of Lower Bounds. The lower bounds sketched above are nearly tight. The upper bound of Theorem 1 follows from combining existing results showing that the single-message shuffle model provides privacy amplification of locally differentially private protocols [54,9], with known locally differentially private protocols for frequency estimation [97,55,42,9]. In particular, as recently shown by Balle et al. [9], a pure $(\varepsilon_L, 0)$-differentially private local randomizer yields a

protocol in the shuffle 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.[13] Then:

- When combined with RAPPOR [55,42], we get an upper bound of $\tilde{O}(n^{1/4})$ on the error.
- When combined with the B-randomized response [97,3], we get an error upper bound of $\tilde{O}(\sqrt{B})$.

The full details appear in Section C. Put together, these imply that the minimum in our lower bound in Theorem 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 shuffle model); indeed, a stronger bound in Equation (3) would lead to larger lower bounds in the single-message shuffle 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 single-message shuffle model (Theorem 3). 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 [92] 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 shuffle 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 3 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

[13] Note that we cannot use the earlier amplification by shuffling result of [54], 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 $\ln n$.

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 [92] 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 (we cannot show this for all x as in general $\delta_L > 0$).

(b) Using the *Level-1 inequality* from the analysis of Boolean functions [84] (see Theorem 13 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 B.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 [29] to deduce the desired lower bound in the single-message shuffle model.

4.2 Overview of Multi-Message Protocols

An important consequence of our lower bound in Theorem 1 is that one cannot achieve an error of $\mathrm{polylog}(n, B)$ 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 shuffle 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 [54,9] to improve the privacy guarantees.

Thus, shuffled versions of the B-randomized response [97,3], RAPPOR [55,42,3], the Bassily–Smith protocol [12], TreeHist and Bitstogram [11], and the Hadamard response protocol [3,2], will still incur an error of $\Omega(\min(\sqrt[4]{n}, \sqrt{B}))$.

Moreover, although the single-message protocol of Cheu et al. [29] for binary aggregation (as well as the multi-message protocols given in [60,7,59,8] 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 $\text{polylog}(n, B)$, 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 and communication.

Our protocols. We design a multi-message version of the private-coin Hadamard response of Acharya et al. [3,2] 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 [3,2]). 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 shuffle model, and it has polylogarithmic error and communication per user (see Theorem 15, Lemmas 17, 18, and 19 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 bounded above by $\text{polylog}(n, B)$. This algorithm is based on a multi-message version of randomized response combined in a delicate manner with the Count Min data structure [34] (for more details, see Section D.2). Previous work [12,11] on DP has used Count Sketch [24], which is a close variant of Count Min, to reduce heavy hitter computation to frequency estimation. In contrast, our use of Count Min has the purpose of reducing the amount of communication per user.

5 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 $\tau = \mathrm{polylog}(n, B)$ 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. [11]), our multi-message protocols for frequency estimation yield multi-message protocols for computing the τ-heavy hitters with $\tau = \mathrm{polylog}(n, B)$ and total communication of $\mathrm{polylog}(n, B)$ bits per user (for more details, see Section H).

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?[14] 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. [22] obtained an upper bound of $(\log B)^{O(d)}$ on the error (see Section 6 for more related work). It has also been studied in the local DP model [33]; 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 shuffle 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 F 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 2) with a communication-efficient implementation, in the multi-message shuffle model, of the space-partitioning data structure used in the central model protocol of Chan et al. [22]. 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

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

this into a protocol with small maximum communication in the shuffle model, we develop an approach analogous to the matrix mechanism [74,75]. We argue that the transformation of the aforementioned central model algorithm for range counting into a private protocol in the multi-message shuffle 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. [48] in the central model does not seem to transfer to the shuffle 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 $o(1)$) [93, 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 [73,42] (for more on related work, see Section 6 below). Our private range counting protocol described above yields a multi-message protocol with communication $\mathrm{polylog}(n)$ per user and that M-estimates the median up to error $\mathrm{polylog}(n)$, i.e., outputs a value $y \in [0, 1]$ such that $\sum_i |x_i - y| \leqslant \min_{\tilde{x}} \sum_i |x_i - \tilde{x}| + \mathrm{polylog}(n)$ (see Theorem 23 in Section I). Beyond M-estimation of the median, our work implies private multi-message protocols for estimating *quantiles* with $\mathrm{polylog}(n)$ error and $\mathrm{polylog}(n)$ bits of communication per user (see Section I for more details).

6 Related Work

Shuffle Privacy Model. Following the proposal of the Encode-Shuffle-Analyze architecture by Bittau et al. [16], several recent works have sought to formalize the trade-offs in the shuffle model with respect to standard local and central DP [54,9] as well as devise private schemes in this model for tasks such as secure aggregation [29,9,60,7,59,8]. In particular, for the task of *real* aggregation, Balle et al. [9] showed that in the single-message shuffle model, the optimal error is $\Theta(n^{1/6})$ (which is better than the error in the local model which is known to be $\Theta(n^{1/2})$).[15] By contrast, recent follow-up work gave multi-message protocols for the same task with error and communication of $\mathrm{polylog}(n)$ [60,7,59,8][16]. Our

[15] Although the single-message real summation protocol of Balle et al. [9] 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.

[16] A basic primitive in these protocols is a "split-and-mix" procedure that goes back to the work of Ishai et al. [68].

work is largely motivated by the aforementioned body of works demonstrating the power of the shuffle 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. [96] recently designed an extension of the shuffle 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. [53] studied multi-message ("report fragmenting") protocols for frequency estimation in a practical shuffle 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 2 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) is a fundamental problem that has been extensively studied in numerous computational models including data structures, sketching, streaming, and communication complexity, (in particular, [79,24,56,34,35,31,80,77,63,61,101,70]). Heavy hitters and frequency estimation have also been studied extensively in the standard models of DP, e.g., [97,67,12,11,95,20,2]. The other problems we consider in the shuffle model, namely, range counting, M-estimation of the median, and quantiles, have been well-studied in the literature on data structures and sketching [37] as well as in the context of DP in the central and local models. Dwork and Lei [45] 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 [73] 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 [98], the aforementioned approach exhibits an improved rate of convergence. Furthermore, motivated by risk bounds under privacy constraints, Duchi et al. [42] 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 [100,69] 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 [42] and are quite different from the approach we use for the ℓ_∞ norm in the proof of Theorem 1 (specifically, in the proof of Theorem 6).

We also note that an anti-concentration lemma qualitatively similar to our Lemma 10 was used by Chan et al. [23, 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. [58] 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 [30] and the references therein). Early works on differentially private frequency estimation , e.g., [43,64], apply naturally to range counting, though the approach of summing up frequencies yields large errors for queries with large ranges.

For $d = 1$, Dwork et al. [47] 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. [22] 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 [81] 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 $(\log B + \log(n)^{O(d)})/\varepsilon$ [48], obtained using ideas from [47,22] 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 $\log^* B$ [14,21].

Xiao et al. [99] showed how to obtain private range count queries by using Haar wavelets, while Hay et al. [66] formalized the method of maintaining a hierarchical representation of data; the aforementioned two works were compared and refined by Qardaji et al. [85]. Cormode et al. [33] 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. [74,75] 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. [36].

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. [39]). Kilian et al. [72] 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 shuffle 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 [93] and the references therein), they almost always use multiple rounds, and of-

ten have large overheads compared to the cost of computing $f(x_1, \ldots, x_n)$ in a non-private setting.

7 Conclusions and Open Problems

The shuffle model is a promising new privacy framework motivated by the significant interest in anonymous communication. In this paper, we studied the fundamental task of frequency estimation in this setup. In the single-message shuffle model, we established nearly tight bounds on the error for frequency estimation: while in the local model the error is well-known to be $\tilde{\Theta}(\sqrt{n})$, we proved that the right bound in the single-message model is the minimum of $\tilde{\Theta}(n^{1/4})$ and $\tilde{\Theta}(\sqrt{B})$, which interestingly are achieved by shuffling the widely used RAPPOR and the B-randomized response protocols, respectively. Moreover, we proved a nearly tight lower bound on the number of users required to solve the selection problem in the single-message shuffle model. 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 3) 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 shuffle 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 shuffle 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 shuffle model protocols for fundamental primitives such as clustering [91] and distribution testing [1], 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 [88,98,13,26,25,27] in the shuffle privacy model.

References

1. Acharya, J., Canonne, C., Freitag, C., Tyagi, H.: Test without trust: Optimal locally private distribution testing. In: AISTATS. pp. 2067–2076 (2019)
2. Acharya, J., Sun, Z.: Communication complexity in locally private distribution estimation and heavy hitters. In: ICML. pp. 97:51–60 (2019)
3. Acharya, J., Sun, Z., Zhang, H.: Hadamard response: Estimating distributions privately, efficiently, and with little communication. In: AISTATS. pp. 1120–1129 (2019)
4. Agarwal, N., Suresh, A.T., Yu, F.X.X., Kumar, S., McMahan, B.: cpsgd: Communication-efficient and differentially-private distributed sgd. In: Advances in Neural Information Processing Systems. pp. 7564–7575 (2018)
5. Apple Differential Privacy Team: Learning with privacy at scale. Apple Machine Learning Journal (2017), https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf
6. Balcer, V., Cheu, A.: Separating local & shuffled differential privacy via histograms. In: ITC. pp. 1:1–1:14 (2020)
7. Balle, B., Bell, J., Gascón, A., Nissim, K.: Differentially private summation with multi-message shuffling. CoRR **abs/1906.09116** (2019)
8. Balle, B., Bell, J., Gascón, A., Nissim, K.: Improved summation from shuffling. arXiv: 1909.11225 (2019)
9. Balle, B., Bell, J., Gascón, A., Nissim, K.: The privacy blanket of the shuffle model. In: CRYPTO. pp. 638–667 (2019)
10. Balle, B., Bell, J., Gascón, A., Nissim, K.: Private summation in the multi-message shuffle model. arXiv:2002.00817 (2020)
11. Bassily, R., Nissim, K., Stemmer, U., Thakurta, A.G.: Practical locally private heavy hitters. In: NIPS. pp. 2288–2296 (2017)
12. Bassily, R., Smith, A.: Local, private, efficient protocols for succinct histograms. In: STOC. pp. 127–135 (2015)
13. Bassily, R., Smith, A.D., Thakurta, A.: Private empirical risk minimization: Efficient algorithms and tight error bounds. In: FOCS. pp. 464–473 (2014)
14. Beimel, A., Nissim, K., Stemmer, U.: Private learning and sanitization: Pure vs. approximate differential privacy. In: APPROX-RANDOM. pp. 363–378 (2013)
15. Bentley, J.L.: Decomposable searching problems. IPL **8(5)**, 244–251 (1979)
16. Bittau, A., Erlingsson, Ú., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnés, J., Seefeld, B.: Prochlo: Strong privacy for analytics in the crowd. In: SOSP. pp. 441–459 (2017)
17. Blum, A., Dwork, C., Nissim, K., McSherry, F.: Practical privacy: the SuLQ framework. In: PODS. pp. 128–138 (2005)
18. Blum, A., Ligett, K., Roth, A.: A learning theory approach to non-interactive database privacy. In: STOC. pp. 609–618 (2008)
19. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: a nonasmpytotic theory of independence. Clarendon Press, Oxford (2012)
20. Bun, M., Nelson, J., Stemmer, U.: Heavy hitters and the structure of local privacy. In: PODS. pp. 435–447 (2018)

21. Bun, M., Nissim, K., Stemmer, U., Vadhan, S.: Differentially private release and learning of threshold functions. In: FOCS. pp. 634–649 (2015)
22. Chan, T.H., Shi, E., Song, D.: Private and continual release of statistics. ACM Trans. Inf. Syst. Secur. **14**(3), 26:1–26:24 (2011)
23. Chan, T.H.H., Shi, E., Song, D.: Optimal lower bound for differentially private multi-part aggregation. In: European Symposium on Algorithms (2012)
24. Charikar, M., Chen, K., Farach-Colton, M.: Finding frequent items in data streams. In: ICALP. pp. 693–703 (2002)
25. Chaudhuri, K., Monteleoni, C.: Privacy-preserving logistic regression. In: NIPS. pp. 289–296 (2008)
26. Chaudhuri, K., Monteleoni, C., Sarwate, A.D.: Differentially private empirical risk minimization. JMLR **12**, 1069–1109 (2011)
27. Chaudhuri, K., Sarwate, A.D., Sinha, K.: A near-optimal algorithm for differentially-private principal components. JMLR **14**(1), 2905–2943 (2013)
28. Chen, L., Ghazi, B., Kumar, R., Manurangsi, P.: On distributed differential privacy and counting distinct elements. arXiv:2009.09604 (2020)
29. Cheu, A., Smith, A.D., Ullman, J., Zeber, D., Zhilyaev, M.: Distributed differential privacy via mixnets. In: EUROCRYPT. pp. 375–403 (2019)
30. Cormode, G.: Sketch techniques for approximate query processing. Foundations and Trends in Databases. NOW publishers (2011)
31. Cormode, G., Hadjieleftheriou, M.: Finding frequent items in data streams. VLDB **1**(2), 1530–1541 (2008)
32. Cormode, G., Kulkarni, T., Srivastava, D.: Marginal release under local differential privacy. In: SIGMOD. pp. 131–146 (2018)
33. Cormode, G., Kulkarni, T., Srivastava, D.: Answering range queries under local differential privacy. In: Proceedings of International Conference on Management of Data (SIGMOD). p. 18321834 (2019)
34. Cormode, G., Muthukrishnan, S.: An improved data stream summary: The Count-Min sketch and its applications. Journal of Algorithms **55**(1), 58–75 (2005)
35. Cormode, G., Muthukrishnan, S.: What's hot and what's not: tracking most frequent items dynamically. TODS **30**(1), 249–278 (2005)
36. Cormode, G., Procopiuc, C., Srivastava, D., Shen, E., Yu, T.: Differentially private spatial decompositions. In: ICDE. pp. 20–31 (2012). https://doi.org/10.1109/ICDE.2012.16, http://dx.doi.org/10.1109/ICDE.2012.16
37. Cormode, G., Yi, K.: Small Summaries for Big Data. Cambridge University Press (2020), http://cormode.org/ssbd
38. Cover, T.A., Thomas, J.M.: Elements of Information Theory. Wiley (1991)
39. Cramer, R., Damgård, I.B., Nielsen, J.B.: Secure Multiparty Computation. Cambridge University Press (2015)
40. Ding, B., Kulkarni, J., Yekhanin, S.: Collecting telemetry data privately. In: NIPS. pp. 3571–3580 (2017)
41. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and statistical minimax rates. In: FOCS. pp. 429–438 (2013)
42. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Minimax optimal procedures for locally private estimation. JASA **113**(521), 182–201 (2018)
43. Dwork, C.: Differential privacy. In: ICALP. pp. 1–12 (2006)
44. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: Privacy via distributed noise generation. In: EUROCRYPT. pp. 486–503 (2006)

45. Dwork, C., Lei, J.: Differential privacy and robust statistics. In: STOC. pp. 371–380 (2009)
46. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: TCC. pp. 265–284 (2006)
47. Dwork, C., Naor, M., Pitassi, T., Rothblum, G.N.: Differential privacy under continual observation. In: STOC. pp. 715–724 (2010)
48. Dwork, C., Naor, M., Reingold, O., Rothblum, G.N.: Pure differential privacy for rectangle queries via private partitions. In: ASIACRYPT. pp. 735–751 (2015)
49. Dwork, C., Naor, M., Reingold, O., Rothblum, G.N., Vadhan, S.: On the complexity of differentially private data release: Efficient algorithms and hardness results. In: STOC. pp. 381–390 (2009)
50. Dwork, C., Roth, A.: The Algorithmic Foundations of Differential Privacy. Now Publishers Inc. (2014)
51. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science **9**(3–4), 211–407 (2014)
52. Edmonds, A., Nikolov, A., Ullman, J.: The power of factorization methods in local and central differential privacy. In: Symposium on the Theory of Computing (2020)
53. Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Song, S., Talwar, K., Thakurta, A.: Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. arXiv preprint arXiv:2001.03618 (2020)
54. Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., Thakurta, A.: Amplification by shuffling: From local to central differential privacy via anonymity. In: SODA. pp. 2468–2479 (2019)
55. Erlingsson, Ú., Pihur, V., Korolova, A.: RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In: CCS. pp. 1054–1067 (2014)
56. Estan, C., Varghese, G.: New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice. TOCS **21**(3), 270–313 (2003)
57. Freehaven: Selected papers in anonymity. https://www.freehaven.net/anonbib/
58. Ghazi, B., Golowich, N., Kumar, R., Manurangsi, P., Pagh, R., Velingker, A.: Pure differentially private summation from anonymous messages. In: Information Theoretic Cryptography (ITC) (2020)
59. Ghazi, B., Manurangsi, P., Pagh, R., Velingker, A.: Private aggregation from fewer anonymous messages. arXiv:1909.11073 (2019)
60. Ghazi, B., Pagh, R., Velingker, A.: Scalable and differentially private distributed aggregation in the shuffled model. arXiv:1906.08320 (2019)
61. Gilbert, A.C., Guha, S., Indyk, P., Kotidis, Y., Muthukrishnan, S., Strauss, M.J.: Fast, small-space algorithms for approximate histogram maintenance. In: STOC. pp. 389–398 (2002)
62. Greenberg, A.: Apple's "differential privacy" is about collecting your data – but not your data. Wired, June **13** (2016)
63. Greenwald, M., Khanna, S., et al.: Space-efficient online computation of quantile summaries. ACM SIGMOD Record **30**(2), 58–66 (2001)
64. Hardt, M., Ligett, K., McSherry, F.: A simple and practical algorithm for differentially private data release. In: NIPS. pp. 2339–2347 (2012), http://dl.acm.org/citation.cfm?id=2999325.2999396
65. Hardt, M., Rothblum, G.N.: A multiplicative weights mechanism for privacy-preserving data analysis. In: FOCS. pp. 61–70 (2010)

66. Hay, M., Rastogi, V., Miklau, G., Suciu, D.: Boosting the accuracy of differentially private histograms through consistency. VLDB **3**(1-2), 1021–1032 (2010). https://doi.org/10.14778/1920841.1920970, http://dx.doi.org/10.14778/1920841.1920970
67. Hsu, J., Khanna, S., Roth, A.: Distributed private heavy hitters. In: ICALP. pp. 461–472 (2012)
68. Ishai, Y., Kushilevitz, E., Ostrovsky, R., Sahai, A.: Cryptography from anonymity. In: FOCS. pp. 239–248 (2006)
69. Kairouz, P., Bonawitz, K., Ramage, D.: Discrete distribution estimation under local privacy. In: ICML. pp. 2436–2444 (2016)
70. Karnin, Z., Lang, K., Liberty, E.: Optimal quantile approximation in streams. In: FOCS. pp. 71–78 (2016)
71. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Rashkodnikova, S., Smith, A.: What can we learn privately? In: FOCS. pp. 531–540 (2008)
72. Kilian, J., Madeira, A., Strauss, M.J., Zheng, X.: Fast private norm estimation and heavy hitters. In: TCC. pp. 176–193 (2008)
73. Lei, J.: Differentially private m-estimators. In: NIPS. pp. 361–369 (2011)
74. Li, C., Hay, M., Rastogi, V., Milau, G., McGregor, A.: Optimizing linear counting queries under differential privacy. In: PODS. pp. 123–134 (2010)
75. Li, C., Miklau, G.: An adaptive mechanism for accurate query answering under differential privacy. In: VLDB. vol. 5(6), pp. 514–525 (2012)
76. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: ICDE. pp. 106–115 (2007)
77. Manku, G.S., Rajagopalan, S., Lindsay, B.G.: Approximate medians and other quantiles in one pass and with limited memory. ACM SIGMOD Record **27**(2), 426–435 (1998)
78. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: FOCS. pp. 94–103 (2007)
79. Misra, J., Gries, D.: Finding repeated elements. Science of Computer Programming **2**(2), 143–152 (1982)
80. Munro, J.I., Paterson, M.S.: Selection and sorting with limited storage. TCS **12**(3), 315–323 (1980)
81. Muthukrishnan, S., Nikolov, A.: Optimal private halfspace counting via discrepancy. In: STOC. pp. 1285–1292 (2012)
82. Nguyen, T., Xiao, X., Yang, Y., Hui, S.C., Shin, H., Shin, J.: Collecting and analyzing data from smart device users with local differential privacy. In: arXiv:1606.05053 (2016)
83. Nikolov, A., Talwar, K., Zhang, L.: On the geometry of differential privacy: the sparse and approximate cases. In: STOC. pp. 351–360 (2013)
84. O'Donnell, R.: Analysis of Boolean functions. Cambridge University Press (2014)
85. Qardaji, W., Yang, W., Li, N.: Understanding hierarchical methods for differentially private histograms. VLDB **6**(14), 1954–1965 (2013). https://doi.org/10.14778/2556549.2556576, http://dx.doi.org/10.14778/2556549.2556576
86. Roos, B.: Binomial approximation to the Poisson binomial distribution: The Krawtchouk Expansion. Theory of Probability and its Applications **45(2)**, 258–272 (2006)
87. Shankland, S.: How Google tricks itself to protect Chrome user privacy. CNET, October (2014)
88. Smith, A.D.: Privacy-preserving statistical estimation with optimal convergence rates. In: STOC. pp. 813–822 (2011)

89. Steinke, T., Ullman, J.: Between pure and approximate differential privacy. Journal of Privacy and Confidentiality **7(2)**, 3–22 (2016)
90. Steinke, T., Ullman, J.: Tight lower bounds for differentially private selection. In: FOCS. pp. 552–563 (2017)
91. Stemmer, U.: Locally private k-means clustering. In: Proceedings of the 2020 Symposium on Discrete Algorithms (2020)
92. Ullman, J.: Tight lower bounds for locally differentially private selection. In: arXiv:1802.02638 (2018)
93. Vadhan, S.: The complexity of differential privacy. In: Tutorials on the Foundations of Cryptography, pp. 347–450. Springer (2017)
94. Wagh, S., He, X., Machanavajjhala, A., Mittal, P.: Dp-cryptography: Marrying differential privacy and cryptography in emerging applications. CoRR **abs/2004.08887** (2020), https://arxiv.org/abs/2004.08887, to appear in Communications of the ACM.
95. Wang, T., Blocki, J., Li, N., Jha, S.: Locally differentially private protocols for frequency estimation. In: USENIX Security. pp. 729–745 (2017)
96. Wang, T., Xu, M., Ding, B., Zhou, J., Li, N., Jha, S.: Practical and robust privacy amplification with multi-party differential privacy. arXiv:1908.11515 (2019)
97. Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. JASA **60**(309), 63–69 (1965)
98. Wasserman, L., Zhou, S.: A statistical framework for differential privacy. JASA **105**(489), 375–389 (2010)
99. Xiao, X., Wang, G., Gehrke, J.: Differential privacy via wavelet transforms. TKDE **23**(8), 1200–1214 (2010)
100. Ye, M., Barg, A.: Optimal schemes for discrete distribution estimation under local differential privacy. In: ISIT. pp. 759–763 (2017)
101. Yi, K., Zhang, Q.: Optimal tracking of distributed heavy hitters and quantiles. Algorithmica **65**(1), 206–223 (2013)