Pure Differentially Private Summation from Anonymous Messages

Badih Ghazi
Google Research, Mountain View, CA, USA
badihghazi@gmail.com

Noah Golowich
Google Research, Mountain View, CA, USA
MIT, Cambridge, MA, USA
nzg@mit.edu

Ravi Kumar
Google Research, Mountain View, CA, USA
ravi.k53@gmail.com

Pasin Manurangsi
Google Research, Mountain View, CA, USA
pasin30055@gmail.com

Rasmus Pagh
Google Research, Mountain View, CA, USA
IT University of Copenhagen, Denmark
Basic Algorithms Research Copenhagen, Denmark
pagh@itu.dk

Ameya Velingker
Google Research, Mountain View, CA, USA
ameyav@google.com

Abstract

The *shuffled* (aka *anonymous*) model has recently generated significant interest as a candidate distributed privacy framework with trust assumptions better than the central model but with achievable error rates smaller than the local model. In this paper, we study *pure* differentially private protocols in the shuffled model for *summation*, a very basic and widely used primitive. Specifically:

- For the binary summation problem where each of n users holds a bit as an input, we give a pure ϵ-differentially private protocol for estimating the number of ones held by the users up to an absolute error of $O_\epsilon(1)$, and where each user sends $O_\epsilon(\log n)$ one-bit messages. This is the first pure protocol in the shuffled model with error $o(\sqrt{n})$ for constant values of ϵ.

  Using our binary summation protocol as a building block, we give a pure ϵ-differentially private protocol that performs summation of real numbers in $[0, 1]$ up to an absolute error of $O_\epsilon(1)$, and where each user sends $O_\epsilon(\log^3 n)$ messages each consisting of $O(\log \log n)$ bits.
- In contrast, we show that for any pure ϵ-differentially private protocol for binary summation in the shuffled model having absolute error $n^{0.5-\Omega(1)}$, the per user communication has to be at least $\Omega_\epsilon(\sqrt{\log n})$ bits. This implies (i) the first separation between the (bounded-communication) multi-message shuffled model and the central model, and (ii) the first separation between pure and approximate differentially private protocols in the shuffled model.

Interestingly, over the course of proving our lower bound, we have to consider (a generalization of) the following question that might be of independent interest: given $\gamma \in (0, 1)$, what is the smallest positive integer m for which there exist two random variables X^0 and X^1 supported on $\{0, \ldots, m\}$ such that (i) the total variation distance between X^0 and X^1 is at least $1 - \gamma$, and (ii) the moment generating functions of X^0 and X^1 are within a constant factor of each other everywhere? We show that the answer to this question is $m = \Theta(\sqrt{\log(1/\gamma)})$.

2012 ACM Subject Classification Security and privacy → Privacy-preserving protocols; Mathematics of computing → Probabilistic algorithms; Theory of computation → Communication complexity

1st Conference on Information-Theoretic Cryptography (ITC 2020).
Editors: Yael Tauman Kalai, Adam D. Smith, and Daniel Wichs; Article No. 15; pp. 15:1–15:23
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Keywords and phrases Pure differential privacy, Shuffled model, Anonymous messages, Summation, Communication bounds

Digital Object Identifier 10.4230/LIPIcs.ITC.2020.15

Related Version A full version of the paper is available at `https://arxiv.org/abs/2002.01919`.

Funding *Noah Golowich*: Supported at MIT by a Fannie & John Hertz Foundation Fellowship, an MIT Akamai Fellowship, and an NSF Graduate Fellowship.

Acknowledgements We are grateful to Borja Balle, Kunal Talwar, and Vitaly Feldman for helpful discussions.

1 Introduction

Since its introduction by Dwork et al. [19, 18], *differential privacy (DP)* has become widely popular as a rigorous mathematical definition of privacy. This has led to practical deployments at companies such as Apple [28, 4], Google [22, 38], and Microsoft [17], and in government agencies such as the United States Census Bureau [2]. The most widely studied setting with DP is the so-called *central* model (denoted $\mathrm{DP}_{\mathrm{central}}$) where an analyzer observes the crude user data but is supposed to release a differentially private data structure. Many accurate private algorithms have been discovered in the central model; however, the model is limited when the analyst is not to be trusted with the user data. To remedy this, the more appealing *local* model of DP (denoted $\mathrm{DP}_{\mathrm{local}}$) [33] (also [41]) requires the messages sent by each user to the analyst to be private. Nevertheless, the local model suffers from large estimation errors that are known to be on the order of $\sqrt{n}$, where n is the number of users, for a variety of problems including summation, the focus of this work [10, 14]. This has motivated the study of the *shuffled* model of DP (denoted $\mathrm{DP}_{\mathrm{shuffled}}$), which is intended as a middle-ground with trust assumptions better than those of the central model and estimation accuracy better than the local model.

While an analogous setup was first introduced in crytpography by Ishai et al. [31] in their work on cryptography from anonymity, the shuffled model was first proposed for privacy-preserving computations by Bittau et al. [11] in their Encode-Shuffle-Analyze architecture. In this setup, each user sends (potentially several) messages to a trusted shuffler, who randomly permutes all incoming messages before passing them to the analyst. We will treat the shuffler as a black box in this work, though we point out that various efficient cryptographic implementations of the shuffler have been considered, including onion routing, mixnets, third-party servers, and secure hardware (see, e.g., the discussions in [31, 11]). The privacy properties of $\mathrm{DP}_{\mathrm{shuffled}}$ were first studied, independently, by Erlingsson et al. [21] and Cheu et al. [15]. Moreover, several recent works have sought to nail down the trade-offs between accuracy, privacy, and communication [15, 8, 27, 6, 25, 26, 7, 5, 20, 9].

Pure- and Approximate-DP

The two most widely used notions of DP are pure-DP [19] and approximate-DP [18], which we recall next. For any parameters $\epsilon \geq 0$ and $\delta \in [0, 1]$, a randomized algorithm P is (ϵ, δ)*-DP* if for every pair datasets X, X' differing on a single user's data, and for every subset $\mathcal{S}$ of transcripts of P, it is the case that

$$\Pr[P(X) \in \mathcal{S}] \leq e^{\epsilon} \cdot \Pr[P(X') \in \mathcal{S}] + \delta, \tag{1}$$

where the probabilities are taken over the randomness in P. The notion of ϵ-DP is the special case where δ is set to 0 in (1); we use the terms *pure-DP* when $\delta = 0$ and *approximate-DP* when $\delta > 0$. While δ is intuitively an upper bound on the probability that an (ϵ, δ)-DP algorithm fails to be ϵ-DP, this failure event can in principle be catastrophic, revealing all the user inputs to the analyst. Pure-DP protocols are thus highly desirable as they guarantee more stringent protections against the leakage of user data. In the central and local settings, several prior works either obtained pure protocols in regimes where approximate protocols were previously known, or proved separations between pure and approximate protocols (e.g., [30, 16, 37, 39, 13]).

Summation

A basic primitive in data analytics and machine learning is the *summation* (aka *aggregation*) of inputs held by different users. Indeed, private summation is a critical building block in the emerging area of *federated learning* [34], where a machine learning model, say a neural network, is to be trained on data held by many users without having the users send their data over to a central analyzer (see [32] for a recent extensive overview). To do so, private variants of Stochastic Gradient Descent have been developed and their privacy/accuracy trade-offs analyzed (e.g., [1]). The gist of these procedures is the private summation of users' gradient updates. Private summation is also closely related to functions in the widely studied class of *counting queries* (e.g., [40, 12, 30, 29, 37]).

Several recent work studied approximate-$\mathrm{DP}_{\mathrm{shuffled}}$ protocols for summation [15, 8, 27, 6, 26, 7, 5]. For binary summation, Cheu et al. [15] show that the standard randomized response is an (ϵ, δ)-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for binary summation and that it incurs an absolute error of only $O(\sqrt{\log n})$ for ϵ constant and δ inverse polynomial in n. For real summation in the single-message shuffled model (denoted $\mathrm{DP}^{1}_{\mathrm{shuffled}}$), where each user sends a single message to the shuffler, Balle et al. [8] show that the tight error for approximate protocols is $\Theta(n^{1/6})$. For real summation in the multi-message shuffled model (denoted $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$), where a user can send more than one message, the state-of-the-art approximate protocol was recently obtained in [26, 7][1] and it incurs error at most $O(1/\epsilon)$ with every user sending $O(1 + \frac{\log(1/\delta)}{\log n})$ messages of $O(\log n)$ bits each.

The aforementioned protocols, along with several other results (including the work on "privacy amplification by shuffling" of Erlingsson et al. [21] and Balle et al. [8]), demonstrate the power of the shuffled model over the local model in terms of privacy, as any $(\epsilon, o(1/n))$-$\mathrm{DP}_{\mathrm{local}}$ summation protocol must incur an error of $\Omega_\epsilon(\sqrt{n})$ [14]. However, all of the protocols proposed so far in the shuffled model only achieve an advantage over the local model when allowed approximation. This leads us to the following basic and perplexing question that is the focus of our work:

▶ **Question 1.** *Are there pure-*$\mathrm{DP}_{\mathrm{shuffled}}$ *protocols that achieve better utility than any* $\mathrm{DP}_{\mathrm{local}}$ *protocol?*

1.1 Main Results

We positively answer the above question for the problem of summation. Namely, we give the first pure-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for binary summation with error depending only on ϵ but independent of n and with logarithmic communication per user.

[1] See also [9] (merger of [7, 6]) together with a novel recursive protocol with *poly*$(\log \log n)$ error where each user sends $O(\log \log n)$ messages.

▶ **Theorem 2 (Pure Binary Summation via Shuffling).** *For every positive real number ϵ, there is a (non-interactive) ϵ-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for binary summation that has expected error $O_\epsilon(1)$ and where each user sends $O_\epsilon(\log n)$ messages each consisting of a single bit.*

We use the protocol in Theorem 2 as a building block in order to also obtain a protocol with constant error and polylogarithmic communication per user for the more general task of real summation where each user input is a real number in $[0, 1]$.

▶ **Theorem 3 (Pure Real Summation via Shuffling).** *For every positive real number ϵ, there is a (non-interactive) ϵ-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for real summation that has expected error $O_\epsilon(1)$ and where each user sends $O_\epsilon(\log^3 n)$ messages each consisting of $O(\log \log n)$ bits.*

In light of Theorem 2, a natural question is if there is a (non-interactive) pure-DP protocol for binary summation with logarithmic (or even constant) error and constant communication per user, as in the approximate case. We show that no such protocol exists, even for very large (polynomial) errors:

▶ **Theorem 4 (Communication Lower Bound).** *In any non-interactive ϵ-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for binary summation with expected error at most $n^{0.5-\Omega(1)}$, the worst-case per user communication must be $\Omega_\epsilon(\sqrt{\log n})$ bits.*

1.2 Implications

Our results described above imply new separations between different types of DP protocols (e.g., $\mathrm{DP}_{\mathrm{central}}$, $\mathrm{DP}_{\mathrm{local}}$, $\mathrm{DP}^{1}_{\mathrm{shuffled}}$, and $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$), and also give the first accurate pure-$\mathrm{DP}_{\mathrm{shuffled}}$ protocol for histograms. We elaborate on these next.

Pure Local vs Shuffled Protocols

In $\mathrm{DP}_{\mathrm{local}}$, the tight accuracy for binary summation is known to be $\Theta(\sqrt{n})$ for approximate protocols [41, 10, 14]. Our Theorems 2 and 3 give the first pure-$\mathrm{DP}_{\mathrm{shuffled}}$ protocols with error $o(\sqrt{n})$ for binary and real summation respectively, and in fact they only incur constant error for both of these problems. Furthermore, Bun et al. [13] gave a generic transformation from any sequentially interactive approximate-$\mathrm{DP}_{\mathrm{local}}$ protocol to a pure-$\mathrm{DP}_{\mathrm{local}}$ protocol with essentially the same accuracy and each user communicates only $O(\log \log n)$ bits. In contrast, our Theorem 4 implies that in any such transformation in the shuffled model (if one exists), the per user communication has to be $\Omega(\sqrt{\log n})$.

Pure vs Approximate Shuffled Protocols

Cheu et al. [15] showed that the standard randomized response [41] is an approximate-DP protocol for binary summation that incurs only logarithmic error (for ϵ constant and δ inverse polynomial in n), and where each user sends a single bit. In contrast, our Theorem 4 implies that the communication cost of any pure-DP protocol for binary summation with logarithmic error (and in fact with error as large as $n^{0.5-\Omega(1)}$) is $\Omega(\sqrt{\log n})$ bits. Put together, these two results imply the first separation between the communication complexity of pure-$\mathrm{DP}_{\mathrm{shuffled}}$ and approximate-$\mathrm{DP}_{\mathrm{shuffled}}$ protocols.

Pure Single-Message vs Multi-Message Shuffled Protocols

As recently shown by [5], any pure-$\mathrm{DP}^{1}_{\mathrm{shuffled}}$ protocol implies a pure-$\mathrm{DP}_{\mathrm{local}}$ protocol with the same accuracy. This implies that any pure-$\mathrm{DP}^{1}_{\mathrm{shuffled}}$ protocol for binary summation must incur error $\Omega_\epsilon(\sqrt{n})$. Our Theorem 2 thus implies a huge separation of $\Theta_\epsilon(\sqrt{n})$ between the errors possible for pure-$\mathrm{DP}^{1}_{\mathrm{shuffled}}$ and pure-$\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$ protocols.

Multi-Message Shuffled vs Central Protocols

It is well-known that the tight error for binary summation in $\mathrm{DP}_{\mathrm{central}}$ is $O(1/\epsilon)$ [19]. Theorem 4 proves that any $\mathrm{DP}_{\mathrm{shuffled}}$ protocol with per user communication $o_\epsilon(\sqrt{\log n})$ bits must incur error $n^{0.5-\Omega(1)}$. It thereby gives the first separation between (bounded-communication) $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$ and $\mathrm{DP}_{\mathrm{central}}$ protocols. Indeed, this is, to the best of our knowledge, the first separation between the accuracy of (bounded-communication) $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$ protocols and those of $\mathrm{DP}_{\mathrm{central}}$ protocols with the same privacy parameters; all previous lower bounds for $\mathrm{DP}_{\mathrm{shuffled}}$ [15, 8, 25] only apply to single-message protocols.

Pure Protocol for Histograms

Our pure binary summation protocol (Theorem 2) implies as a black-box the first pure-DP protocol with polylogarithmic error for computing *histograms* (aka *point functions* or *frequency estimation*), albeit with very large communication (see Appendix A of the full version [24] for more details). It remains a very interesting open question to obtain a communication-efficient and accurate pure-DP protocol for histograms (see Section 5 for more on this and other open questions).

1.3 Overview of Techniques

Binary Summation Protocol

We first explain why all existing summation protocols in the shuffled model with error $o(\sqrt{n})$ are not $O(1)$-DP. First, note that as observed by [5], any pure-$\mathrm{DP}^{1}_{\mathrm{shuffled}}$ protocol implies a pure-$\mathrm{DP}_{\mathrm{local}}$ protocol with the same accuracy and privacy. Combined with the fact that any $O(1)$-$\mathrm{DP}_{\mathrm{local}}$ protocol for summation must have error $\Omega(\sqrt{n})$, this implies the same lower bound for any pure $O(1)$-$\mathrm{DP}^{1}_{\mathrm{shuffled}}$ protocol. In particular, this rules out the binary randomized response [41] that was analyzed in the shuffled model by [15]. It also rules out the protocol implied by shuffling RAPPOR [22], and more generally any protocol obtained by the amplification via shuffling approach of [21, 8]. Moreover, in the multi-message shuffled setup, the state-of-the-art real summation protocols of [26, 7], which rely on the Split-and-Mix procedure [31], only give approximate-DP.

A different $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$ protocol for binary summation can be obtained by instantiating the recent $\mathrm{DP}^{\geq 1}_{\mathrm{shuffled}}$ protocols for computing histograms [25], with a domain size of $B = 2$. On a high-level, the two resulting protocols – one of which is based on the Count Min sketch and the other on the Hadamard response – can be seen as special cases of the following common template: each user (i) samples a number ρ of messages that depend on their input, (ii) independently samples a number η of noise messages, and (iii) sends these $\rho + \eta$ messages to the shuffler. Loosely, the analyzer then outputs the number of messages "consistent with" the queried input. However, it can be seen that any protocol following this template will not be pure-DP, as the supports of the distribution of the count observed at the analyzer can shift by 1 when a single user input is changed. The crucial insight in our pure protocol for binary summation will be to *correlate* the input-dependent messages and the noise messages sampled by each user in steps (i) and (ii) above. By doing so, we not only aim to ensure that the supports are identical but that the two densities are also within a small multiplicative factor on any point. We implement this idea using binary messages by having each user send d bits on both inputs 0 and 1. Specifically, the user will start by flipping a suitably biased coin. If it lands as head, the user will send $(d+1)/2$ zeros and $(d-1)/2$ ones when the input is 0, and vice versa when the input is 1. If the coin lands as tail, the user will sample an integer z from

Algorithm 1 Randomizer for binary summation.

```
procedure BinaryRandomizer$_{\epsilon,n}(x)$
    Let $p, d, s$ be as in Lemma 9 (depending on $\epsilon, n$)
    $a \leftarrow \mathrm{Ber}(p)$
    if $a = 0$ then
        if $x = 0$ then
            return the multiset with $\left(\frac{d-1}{2}\right)$ ones and $\left(\frac{d+1}{2}\right)$ zeros
        else
            return the multiset with $\left(\frac{d+1}{2}\right)$ ones and $\left(\frac{d-1}{2}\right)$ zeros
    else
        $z \leftarrow \mathrm{DLap}_d(d/2, s)$
        return the multiset with $z$ ones and $(d - z)$ zeros
```

Algorithm 2 Analyzer for binary summation.

```
procedure BinaryAnalyzer$_{\epsilon,n}(R)$
    Let $d$ be as in Lemma 9 (depending on $\epsilon, n$)
    return $\frac{nd}{2} + \sum_{y \in R} \left(y - \frac{1}{2}\right)$
```

a truncated discrete Laplace distribution and send z zeros and $d - z$ ones (see Algorithm 1 and Equation (2) for more details). The analyzer (Algorithm 2) then outputs the number of received ones after debiasing. Note that the number of ones received by the analyzer is a random variable taking values between 0 and dn inclusive. To prove that the algorithm is private, we intuitively wish to argue that the noise distribution satisfies the property that its density values on any two adjacent points are within a multiplicative e^ϵ factor. However, the technical challenge stems from the fact that this noise distribution depends on the specific input sequence (and as we discussed above this dependence is necessary!). Instead, we have to analyze the n-fold convolution of the individual responses, and show that the density values of the resulting distribution on any two adjacent points in $\{0, 1, \ldots, dn\}$ are within a multiplicative factor of e^ϵ, for any input sequence. The crux of the proof is to relate the tails of different convolutions of the truncated discrete Laplace distribution (Lemmas 10 and 11). We determine a setting of (i) the mixture probability coefficient (denoted by p in Algorithm 1), (ii) the parameter d, and (iii) the "inverse scaling coefficient" of the truncated discrete Laplace distribution (denoted by s in Algorithm 1), for which the privacy property holds and for which the resulting expected absolute error is $O_\epsilon(1)$.

We point out that the dependence of the error on ϵ that we obtain is $\tilde{O}(1/\epsilon^{3/2})$ for $\epsilon \leq O(1)$ (see Theorem 8). An interesting open question is whether this dependence can be further reduced to $O(1/\epsilon)$, which is the tight error in the central model [19].

Real Summation Protocol

We use our pure private binary summation protocol outlined above as a building block in order to obtain a pure private real summation protocol and prove Theorem 3. We note that Cheu et al. [15] had given a transformation from binary summation to real summation, but their reduction results in a protocol with a very large communication of $\tilde{\Omega}(\sqrt{n})$ bits in order to achieve logarithmic error. We instead give a (different) transformation that results in a protocol with polylogarithmic communication. The high-level idea of our reduction is the following: consider the binary representation of the inputs after rounding them to $O(\log n)$

Algorithm 3 Randomizer for real summation.

```
procedure $\textsc{RealRandomizer}_{(\epsilon_j)_{j\in\mathbb{N}},n}(x)$
    for $j = 1$ to $2\log n$ do
        $x[j] \leftarrow j$th most significant bit of $x$
        $S_j \leftarrow \textsc{BinaryRandomizer}_{\epsilon_j,n}(x[j])$        $S_j$ is a multiset of zeros and ones.
        $R_j \leftarrow \{j\} \times S_j$        $R_j$ is a multiset of tuples $(j, 0)$ and $(j, 1)$.
    return $\bigcup_{j=1}^{2\log n} R_j$
```

Algorithm 4 Analyzer for real summation.

```
procedure $\textsc{RealAnalyzer}_{(\epsilon_j)_{j\in\mathbb{N}},n}(R)$
    for $j = 1$ to $2\log_2 n$ do
        $R_j \leftarrow \{y_1 \mid y \in R \text{ and } y_0 = j\}$        Multiset of bit messages for the $j$th bit.
        $a_j \leftarrow \textsc{BinaryAnalyzer}_{\epsilon_j,n}(R_j)$
    return $\sum_{j=1}^{2\log n} a_j/2^j$
```

bits of precision, then approximate the sum for each bit position independently, and finally combine the estimates into an approximation of the (real-valued) sum of the inputs. Since the bit sum estimates have geometrically decreasing weights, we can afford to increase the error on less significant bits. In terms of privacy, this means that for the jth most significant bit, we run an ϵ_j-DP binary summation protocol where $\epsilon_1, \epsilon_2, \ldots$ is a decreasing sequence. The protocol is illustrated in Algorithms 3 and 4. By carefully choosing the sequence $\epsilon_1, \epsilon_2, \ldots$, we can ensure that the total pure privacy parameter $\sum_j \epsilon_j$ is small, while the total error is a constant times the error for the sum of the most significant bits of the inputs. Intuitively, choosing $\epsilon_1, \epsilon_2, \ldots$ to be a geometrically decreasing sequence (e.g., $\epsilon_j = \frac{0.9^j \cdot \epsilon}{10}$) should suffice for our purposes. However since the communication complexity of our binary summation protocol also depends on the privacy parameter ϵ, such a choice of the sequence would result in *poly*(n) communication complexity. To overcome this, our actual sequence has a "cut-off" so that the ϵ_j's do not go below $\Theta\left(\frac{\epsilon}{\log n}\right)$. This completes the proof overview.

Lower Bound

We next outline the proof of Theorem 4. Without loss of generality, we consider an arbitrary ϵ-DP$_{\text{shuffled}}$ protocol performing binary summation with error $n^{0.5-\Omega(1)}$, and where every user sends m messages each belonging to the domain $\{1, \ldots, k\}$. We wish to lower bound the number of bits of communication per user in this protocol, which is equal to $m \log k$. We denote by $\mathbf{X}^0$ and $\mathbf{X}^1$ the random multisets of messages sent by a user in this protocol under inputs 0 and 1 respectively. Note that $\mathbf{X}^0$ and $\mathbf{X}^1$ are supported on the set $\Delta_{k,m} := \{(z_1, \ldots, z_k) \in \mathbb{Z}^k_{\geq 0} \mid z_1 + \cdots + z_k = m\}$. Here, z_i captures the number of i messages sent by the user for each $i \in \{1, \ldots, k\}$.

Using the pure privacy of the protocol, we can argue that the ratio of the moment generating functions (MGFs) of $\mathbf{X}^0$ and $\mathbf{X}^1$ cannot take a very large or a very small value. Specifically, using the fact that the MGF of a sum of independent random variables is equal to the product of the individual MGFs, we derive a simple yet powerful property that should be satisfied by any ϵ-DP protocol in the shuffled model: the ratio of the MGFs of $\mathbf{X}^0$ and $\mathbf{X}^1$ should always lie in the interval $[e^{-\epsilon}, e^{\epsilon}]$. We will refer to such random variables as having an e^{ϵ}*-bounded MGF ratio* (see Section 4.1 for more details). We remark that while MGFs have been used before in DP by Abadi et al. [1] and subsequent works on Renyi DP (starting from [36]), these usages are in a completely different context compared to ours. In particular, these prior works keep track of the moments in order to bound the privacy parameters under composition of protocols. To the best of our knowledge, MGFs have neither been used in lower bounds for DP nor in the shuffled model before.

Then, using the accuracy of the protocol, we can deduce that the total variation distance between $\mathbf{X}^0$ and $\mathbf{X}^1$ has to be large. We do so by invoking a result from the literature [14, 25] showing that for any binary summation protocol that incurs an absolute error of α, the total variation distance between $\mathbf{X}^0$ and $\mathbf{X}^1$ must be at least $1 - \Theta(\alpha/\sqrt{n})$ (see Theorem 16 for more details). Since $\alpha = n^{0.5-\Omega(1)}$ in our case, we get a lower bound of $1 - n^{-\Omega(1)}$ on the total variation distance between $\mathbf{X}^0$ and $\mathbf{X}^1$.

Equipped with these two ingredients, the task of lower bounding the per user communication cost of the protocol reduces to lower bounding the following quantity:

▶ **Definition 5.** *Given parameters $\epsilon > 0$ and $\gamma \in [0, 1]$, we define $C_{\epsilon,\gamma}$ as the minimum value of $m \log k$ for which there exist two random variables supported on $\Delta_{k,m}$ that are at total variation distance is at least $1 - \gamma$ but that have an e^ϵ-bounded MGF ratio.*

Note that any lower bound on the value of $C_{\epsilon,\gamma}$ can be used to infer a lower bound on the per user communication cost. In order to prove Theorem 4, and given our setting of $\gamma = 1/n^{\Omega(1)}$, it is thus enough for us to show that $C_{\epsilon,\gamma} \geq \Omega_\epsilon(\sqrt{\log(1/\gamma)})$. To prove this bound, it suffices to show that if two random variables $\mathbf{X}^0, \mathbf{X}^1$ have an e^ϵ-bounded MGF ratio, then their total variation distance must be at least $1 - \exp(O_\epsilon(m^2 \log k))$. For each $\mathbf{x} \in \Delta_{k,m}$, we view $\Pr[\mathbf{X}^0 = \mathbf{x}]$ and $\Pr[\mathbf{X}^1 = \mathbf{x}]$ as variables. The e^ϵ-bounded MGF ratio constraints can then be written as infinitely many linear inequalities over these variables. Moreover, the total variation distance between $\mathbf{X}^0$ and $\mathbf{X}^1$ can be written as a maximum of linear combinations of these same variables. We therefore get a linear program with infinitely many constraints, and we would like to show that any solution to it has "cost" (i.e., total variation distance) at least $1 - \exp(O_\epsilon(m^2 \log k))$. We do so by giving a dual solution with cost at most $1 - \exp(O_\epsilon(m^2 \log k))$, which by weak duality implies our desired bound (see Section 4 for more details).

A natural question is if the lower bound $C_{\epsilon,\gamma} \geq \Omega_\epsilon(\sqrt{\log(1/\gamma)})$ outlined above can be improved, as that would immediately lead to an improved communication complexity lower bound. However, we show that the lower bound is tight, even in the special case where $k = 2$. Namely, we give two random variables supported on $\Delta_{2,m}$ with $m = \Theta_\epsilon(\sqrt{\log(1/\gamma)})$ that are at total variation distance at least $1 - \gamma$ but that have an e^ϵ-bounded MGF ratio. Our construction is based on truncations of discrete Gaussian random variables (see Section 4.3 for more details). We note that this limitation only applies to the approach of lower bounding the per user communication complexity via lower bounding $C_{\epsilon,\gamma}$. It remains possible that other approaches might give better lower bounds. For instance, one might be able to proceed by giving a necessary condition for the accuracy of binary summation protocols that is stronger than the total variation distance bound that we used, or a necessary condition for pure privacy that is better than our e^ϵ-bounded MGF ratio property.

1.4 Organization

We start with some notation and background in Section 2. Our protocol for binary summation is presented and analyzed in Section 3. In Section 4, we prove our lower bound (Theorem 4). Any deferred proofs from these sections appear in the full version of the paper [24]. We conclude with some interesting open questions in Section 5. Our full proof for real summation (Theorem 3) and our corollary for histograms are deferred to the full version [24].

2 Preliminaries

Shuffled Model of Privacy

Let n be the number of users and let $\mathcal{X}$ be the domain. For each i in $[n] := \{1, \ldots, n\}$, we denote by x_i the input held by the ith user, and further assume that $x_i \in \mathcal{X}$. In the binary summation case, we have that $\mathcal{X} = \{0, 1\}$ while in the real summation case, we let $\mathcal{X}$ be the set $[0, 1]$ of real numbers. A protocol $P = (R, S, A)$ in the shuffled model consists of three algorithms: (i) the *local randomizer* $R(\cdot)$ whose input is the data of one user and whose output is a sequence of messages, (ii) the *shuffler* $S(\cdot)$ whose input is the concatenation of the outputs of the local randomizers and whose output is a uniform random permutation of its inputs, and (iii) the *analyzer* $A(\cdot)$ whose input is the output of the shuffler and whose output is the output of the protocol. The privacy in the shuffled model is guaranteed with respect to the input to the analyzer, i.e., the output of the shuffler.

▶ **Definition 6** (DP in the shuffled model, [21, 15])**.** *A protocol $P = (R, S, A)$ is (ϵ, δ)-$\mathrm{DP}_{\text{shuffled}}$ if, for any dataset $X = (x_1, \ldots, x_n)$, the algorithm $S(R(x_1), \ldots, R(x_n))$ is (ϵ, δ)-DP. In the special case where $\delta = 0$, we say that the protocol P is ϵ-$\mathrm{DP}_{\text{shuffled}}$.*

Note that the $\mathrm{DP}_{\text{local}}$ model corresponds to the case where S is the identity function.

▶ **Definition 7** (Non-Interactive Protocols)**.** *Let k and m be positive integers. In a* non-interactive *(aka* one-round*) protocol, each of the n users (i.e., randomizers) receives an input b and outputs at most m messages each consisting of $\log k$ bits, according to a certain distribution (depending on b), and using private randomness. We say that such a protocol has a* communication complexity *of $m \log k$.*

It is often convenient to view each message as a number in $[k]$. We use $\mathbf{X}^b \in \mathbb{Z}^k_{\geq 0}$ to denote the random variable whose sth coordinate X^b_s denotes the number of s-messages output by the randomizer on input b. Note that it is always the case that $\sum_{s \in [k]} X^b_s = m$, i.e., $\mathrm{supp}(\mathbf{X}^b) \subseteq \Delta_{k,m} := \{(z_1, \ldots, z_k) \in \mathbb{Z}^k_{\geq 0} \mid z_1 + \cdots + z_k = m\}$.

3 Pure Binary Summation Protocol via Shuffling

In this section we prove Theorem 2, restated formally below.

▶ **Theorem 8.** *For every sufficiently large n and $O(1) \geq \epsilon > 1/n^{2/3}$, there is an ϵ-$\mathrm{DP}_{\text{shuffled}}$ protocol for summation for inputs $x_1, \ldots, x_n \in \{0, 1\}$ where each user sends $O\left(\frac{\log n}{\epsilon}\right)$ one-bit messages to the analyzer and has expected error at most $O\left(\frac{\sqrt{\log 1/\epsilon}}{\epsilon^{3/2}}\right)$.*

We remark that the assumption $\epsilon > \frac{1}{n^{2/3}}$ is made w.l.o.g., because, for $\epsilon \leq \frac{1}{n^{2/3}}$, there is a trivial algorithm that achieves square error of $O(1/\epsilon^{3/2})$: the analyzer just always outputs 0.

Throughout this section we assume that for some absolute constant C, $\epsilon \leq C$, and thus in particular e^ϵ can be bounded above by an absolute constant. (The constant C can be arbitrary.) It is well-known that any ϵ-$\mathrm{DP}_{\text{central}}$ protocol for summation has error $\Omega(1/\epsilon)$ [40]. Thus the error in Theorem 8 is suboptimal by a factor of at most $\tilde{O}(1/\sqrt{\epsilon})$.

The remainder of the section is organized as follows. In Section 3.1, we present the protocol used to prove Theorem 8. In Section 3.2, we prove the accuracy and privacy guarantees of Theorem 8, and in Section 3.3 we outline the proof of a technical lemma needed in the privacy analysis.

3.1 The Protocol

To described the protocol, we will use the truncated version of the discrete Laplace distribution, for which we condition the support to be on $[\mu - w/2, \mu + w/2]$ where $w \geq 1$ is the "width" of the support. We denote such a distribution by $\mathrm{DLap}_w(\mu, s)$. More specifically, its probability mass function satisfies

$$\Pr_{Z \sim \mathrm{DLap}_w(\mu,s)}[Z = z] = \begin{cases} \frac{1}{C_w(\mu,s)} \cdot e^{-|z-\mu|/s} & \text{if } \mu - w/2 \leq z \leq \mu + w/2 \text{ and } z \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Here $C_w(\mu, s) = \sum_{z \in [\mu - w/2, \mu + w/2] \cap \mathbb{Z}} e^{-|z-\mu|/s}$ is simply the normalization factor.

Our randomizer and analyzer are presented in Algorithm 1 and Algorithm 2, respectively. The protocol has 3 parameters: the number of messages d, the "inverse scaling exponent" s, and the "noise probability" p. We always assume that d is a positive odd integer[2]. These parameters will be chosen later (in Lemma 9).

3.2 Privacy Analysis

For $b \in \{0, 1\}$, we write $\mathcal{R}_b$ to denote the distribution[3] on the number of ones output by the randomizer on input b. (This distribution depends on d, s, p but we do not include them in the notation to avoid being cumbersome.) Notice that we can decompose $\mathcal{R}_b$ as a mixture $p \cdot \mathrm{DLap}_d(d/2, s) + (1 - p) \cdot \mathbf{1}(\frac{d-1}{2} + b)$, where we use $\mathbf{1}(\vartheta)$ to denote the distribution that is ϑ with probability 1.

To prove the privacy guarantee of Theorem 8, we first note that we may focus only on the neighboring datasets $(0, \ldots, 0, 0)$ and $(0, \ldots, 0, 1)$; this follows since we may assume (due to symmetry) that more than half of the bits are zero and we can then condition out the results from the 1 bits that they share. (See the proof of Theorem 8 for a formalization of this.) For these datasets, Lemma 9 below bounds the ratio of the probabilities of ending up with a particular union of outputs from these two datasets.

▶ **Lemma 9.** *There is a sufficiently small constant $c_0 \in (0, 1)$ so that the following holds. For any sufficiently large $n \in \mathbb{N}$ and any $c_0 \geq \epsilon > \frac{1}{n^{2/3}}$, let $s = \frac{10}{\epsilon}$, $p = \frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{n(1-e^{-0.1\epsilon})}$, and $d = 4\left\lceil \frac{1000\, e^{100\epsilon}}{(1-e^{-0.1\epsilon})} \cdot \log\left(\frac{n}{1-e^{-0.1\epsilon}}\right)\right\rceil + 3$. Then, for all $t \in \{0, \ldots, dn\}$, we have*

$$\frac{\Pr_{Z_1,\ldots,Z_n \sim \mathcal{R}_0}[Z_1 + \cdots + Z_n = t]}{\Pr_{Z_1,\ldots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \mathcal{R}_1}[Z_1 + \cdots + Z_n = t]} \in [e^{-\epsilon}, e^{\epsilon}]. \tag{3}$$

This means that, for the above selection of parameters, the protocol is ϵ-DP. Using Lemma 9, we prove Theorem 8.

Proof of Theorem 8. We may assume without loss of generality that $\epsilon \leq c_0$, as otherwise we may set ϵ to $\min\{\epsilon, c_0\}$ instead.

We use the local randomizer $\textsc{BinaryRandomizer}_{\epsilon,n}$ of Algorithm 1 and the analyzer $\textsc{BinaryAnalyzer}_{\epsilon,n}$ of Algorithm 2, with the parameters s, d, p given by the expressions in Lemma 9, except with n replaced by $\lceil (n+1)/2 \rceil$. Explicitly, we have $s = \frac{10}{\epsilon}$, $p =$

[2] We only assume that d is odd for convenience, so that $\left(\frac{d-1}{2}\right)$ and $\left(\frac{d+1}{2}\right)$ are integers. Using an even d and replacing these two quantities with $d/2 - 1$ and $d/2 + 1$ also works, provided that the proofs are adjusted appropriately.

[3] This is the distribution of $\mathbf{X}^b$ defined in Section 2.

$\frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{\lceil (n+1)/2 \rceil (1-e^{-0.1\epsilon})}$ and $d = 4\left\lceil \frac{1000\, e^{100\epsilon}}{(1-e^{-0.1\epsilon})} \cdot \log\left(\frac{\lceil (n+1)/2 \rceil}{1-e^{-0.1\epsilon}}\right)\right\rceil + 3$. We prove the accuracy guarantee first, which is a simple consequence of the choices of p, d made in Lemma 9, followed by the privacy guarantee, which uses Lemma 9.

Proof of accuracy

Fix a dataset $X = (x_1, \dots, x_n) \in \{0,1\}^n$. Let $Y \in \mathbb{R}$ be the count released by the analyzer. Moreover, for $1 \le i \le n$, let $Z_1, \dots, Z_n$ be i.i.d. random variables distributed according to $\nu = \mathrm{DLap}_d(d/2, s)$. It is easy to check that $\mathrm{Var}[Z_i] \le 2s^2$. Moreover, let $M \in [n]$ be the number of users for whom the Bernoulli random variable a is equal to 1. In particular, $M \sim \mathrm{Binom}(n, p)$. The expected absolute error is given by

$$\begin{aligned}
\mathbb{E}\left[\left|Y - \sum_{i=1}^{n} x_i\right|\right] &\le \sum_{m=0}^{n} \Pr[M = m] \cdot \left(\frac{m}{2} + \mathbb{E}\left[\left|Z_1 + \dots + Z_m - \frac{md}{2}\right|\right]\right) \\
\text{(by Jensen's inequality)} \quad &\le \mathbb{E}[M/2] + \sum_{m=0}^{n} \Pr[M = m] \cdot \sqrt{\mathbb{E}\left[\left(Z_1 + \dots + Z_m - \frac{md}{2}\right)^2\right]} \\
&\le pn/2 + \sum_{m=0}^{n} \Pr[M = m] \cdot \sqrt{m} \cdot \sqrt{2s^2} \\
&\le pn/2 + \sqrt{2}s \cdot \sqrt{pn}. \qquad (4)
\end{aligned}$$

Since $p = \frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{\lceil (n+1)/2 \rceil (1-e^{-0.1\epsilon})}$, we have $pn/2 + \sqrt{2pn} \cdot s = O\left(\frac{\sqrt{\log 1/\epsilon}}{\epsilon^{3/2}}\right)$. Combined with (4), this gives us the desired upper bound on the expected error of the protocol.

Proof of privacy

Let $X = (x_1, \dots, x_{n-1}, x_n) \in \{0,1\}^n$ and $X' = (x_1, \dots, x_{n-1}, x'_n) \in \{0,1\}^n$ be two neighboring datasets. By symmetry, without loss of generality, we may assume that $x_n = 0$ and at least $n_0 := \lceil n-1 \rceil/2$ of the values of $x_1, \dots, x_{n-1}$ are also 0. By permuting the users, we may also assume without loss of generality that $x_1 = x_2 = \dots = x_{n_0} = 0$. For $1 \le i \le n$, let $Y_i \in [0, d]$ denote the (random) number of 1s output by user i when their input is x_i. Also let $Y'_n \in [0, 1]$ denote the (random) number of 1's output by user n when its input is x'_n. By [8, Lemma A.2], to show that $\frac{\Pr[Y_1 + \dots + Y_{n-1} + Y_n = t]}{\Pr[Y_1 + \dots + Y_{n-1} + Y'_n = t]} \in [e^{-\epsilon}, e^{\epsilon}]$ for all $t \in \mathbb{N}$, it suffices to show that for all $t_0 \in \mathbb{N}$, $\frac{\Pr[Y_1 + \dots + Y_{n_0} + Y_n = t_0]}{\Pr[Y_1 + \dots + Y_{n_0} + Y'_n = t_0]} \in [e^{-\epsilon}, e^{\epsilon}]$. Now the validity of the latter is an immediate consequence of Lemma 9 with the parameter n of Lemma 9 equal to $n_0 + 1$. ◀

From now on, we will use ν and ω_b as abbreviations for $\mathrm{DLap}_d(d/2, s)$ and $\mathbf{1}(\frac{d-1}{2} + b)$ respectively, where d, s are defined as in Lemma 9.

Let us denote by $P_{m,k}$ the probability that m independent random variables from the noise distribution ν sums up to k; more formally, $P_{m,k} := \Pr_{Z_1, \dots, Z_m \sim \nu}[Z_1 + \dots + Z_m = k]$. For convenience, we define $P_{0,0} = 1$ and $P_{0,k} = 0$ for all $k \neq 0$.

As we will see in the proof of Lemma 9 below, expansions of the numerator and denominator of the left hand side of (3) result in similar terms involving $P_{m,k}$, except occasionally with (i) k differing by one or (ii) m differing by 1 and k differing by $\left(\frac{d-1}{2}\right)$ or $\left(\frac{d-3}{2}\right)$. Hence, to bound the ratio between the two, we have to find some relation between $P_{m,k}, P_{m,k-1}, P_{m+1,k+\left(\frac{d-1}{2}\right)}$, and $P_{m+1,k+\left(\frac{d-3}{2}\right)}$. The exact inequality we will use here is stated below and its proof overview is given in Section 3.3.

▶ **Lemma 10.** *For any sufficiently large $n \in \mathbb{N}$, let ϵ, d and s be as in Lemma 9. Then the following hold: For any integers $\frac{10\log(1/(1-e^{-0.1\epsilon}))}{1-e^{-0.1\epsilon}} \leq m \leq n-1$, and $\ell_1, \ell_2 \in \left\{\frac{d-1}{2}, \frac{d-3}{2}\right\}$, if $p \geq \frac{100\, e^{100\epsilon}}{n(1-e^{-0.1\epsilon})}$, then we have*

$$e^{-\epsilon}p(1-e^{-\epsilon/2}) \cdot \left(P_{m+1,k+\ell_1} + \frac{(n-m-1)}{m+1} \cdot P_{m+1,k+\ell_2}\right) + e^{0.2\epsilon} \cdot P_{m,k-1} \geq P_{m,k}.$$

We also need the following lemma, which can be interpreted as an anti-concentration result. Recall that $P_{i,j} = \Pr_{Z_1,\dots,Z_i \sim \nu}[Z_1 + \cdots + Z_i = j]$. For any $a \in \mathbb{N}$, if also $Z'_1, \dots, Z'_a \sim \nu$, then as $\mathbb{E}[Z'_1 + \cdots + Z'_a] = da/2$ and the distribution of $Z'_1 + \cdots + Z'_a$ has sufficient mass at its expectation, one expects that $P_{i+a,j+a} = \Pr[Z_1 + \cdots + Z_i + Z'_1 + \cdots + Z'_a = j + da/2]$ is not too much smaller than $P_{i,j}$. Lemma 11 says that in fact $\Pr[Z_1 + \cdots + Z_i + Z'_1 + \cdots + Z'_a = j + da/2 - d/2]$ is not too much smaller than $P_{i,j}$. Its proof is deferred to the full version [24].

▶ **Lemma 11.** *For any $i, j, a \in \mathbb{N}_0$ such that $a \leq s^2/1000$, we have $P_{i+a,j+a\left(\frac{d-1}{2}\right)} \geq \frac{\sqrt{a}}{40s^3} \cdot P_{i,j}$.*

With Lemmas 10 and 11 ready, we can now prove Lemma 9 as follows.

Proof of Lemma 9. Let $c_0 \in (0,1)$ be some sufficiently small positive constant, to be specified later. We would like to show that, for all $t \in \{0, \dots, dn\}$, the following hold:

$$\Pr_{Z_1,\dots,Z_n \sim \mathcal{R}_0}[Z_1 + \cdots + Z_n = t] \leq e^{\epsilon} \cdot \Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \mathcal{R}_1}[Z_1 + \cdots + Z_n = t], \tag{5}$$

and

$$\Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \mathcal{R}_1}[Z_1 + \cdots + Z_n = t] \leq e^{\epsilon} \cdot \Pr_{Z_1,\dots,Z_n \sim \mathcal{R}_0}[Z_1 + \cdots + Z_n = t]. \tag{6}$$

Due to space constraints, we will only prove (5) here; (6) can be proved in a similar manner. To prove (5), let us first decompose the probability on the left and the right hand sides based on whether Z_n is sampled from the noise distribution ν. This gives

$$\begin{aligned}&\Pr_{Z_1,\dots,Z_n \sim \mathcal{R}_0}[Z_1 + \cdots + Z_n = t]\\ &= p \cdot \Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \nu}[Z_1 + \cdots + Z_n = t]\\ &\quad + (1-p) \cdot \Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0}\left[Z_1 + \cdots + Z_{n-1} = t - \left(\frac{d-1}{2}\right)\right],\end{aligned}$$

and

$$\begin{aligned}&\Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \mathcal{R}_1}[Z_1 + \cdots + Z_n = t]\\ &= p \cdot \Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \nu}[Z_1 + \cdots + Z_n = t]\\ &\quad + (1-p) \cdot \Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0}\left[Z_1 + \cdots + Z_{n-1} = t - \left(\frac{d+1}{2}\right)\right].\end{aligned}$$

Moreover, observe that, by expanding based on the number of variables among $Z_1, \dots, Z_{n-1}$ that uses the noise distribution (i.e., i below), we have

$$\Pr_{Z_1,\dots,Z_{n-1} \sim \mathcal{R}_0, Z_n \sim \nu}[Z_1 + \cdots + Z_n = t]$$

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot P_{i+1, t-(n-1-i)\left(\frac{d-1}{2}\right)},$$

and

$$\begin{aligned}
&\Pr_{Z_1,\ldots,Z_{n-1}\sim\mathcal{R}_0}\left[Z_1 + \cdots + Z_{n-1} = t - \left(\frac{d-1}{2}\right)\right] \\
&= \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
&\Pr_{Z_1,\ldots,Z_{n-1}\sim\mathcal{R}_0}\left[Z_1 + \cdots + Z_{n-1} = t - \left(\frac{d+1}{2}\right)\right] \\
&= \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1}. \\
&= \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot e^{\epsilon/2} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\qquad + \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot (e^{\epsilon} - e^{\epsilon/2}) P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\geq \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot e^{\epsilon/2} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\qquad + \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot (e^{\epsilon/2} - 1) P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\geq \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot e^{\epsilon/2} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\qquad + \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i+1} p^{i+1} (1-p)^{n-2-i} \cdot (e^{\epsilon/2} - 1) P_{i+1, t-(n-i-1)\left(\frac{d-1}{2}\right)-1} \\
&\geq \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} e^{\epsilon/2} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\qquad + \frac{1}{e^{\epsilon}} \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot \frac{p(n-1-i)}{i+1} \cdot (e^{\epsilon/2} - 1) P_{i+1, t-(n-i-1)\left(\frac{d-1}{2}\right)-1}.
\end{aligned}$$

Using the above expressions, we may write the difference between the right hand side and the left hand side of (5) as

$$\begin{aligned}
&e^{\epsilon} \cdot \Pr_{Z_1,\ldots,Z_{n-1}\sim\mathcal{R}_0, Z_n\sim\mathcal{R}_1}[Z_1 + \cdots + Z_n = t] - \Pr_{Z_1,\ldots,Z_n\sim\mathcal{R}_0}[Z_1 + \cdots + Z_n = t] \\
&\geq (e^{\epsilon} - 1) \cdot p \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot P_{i+1, t-(n-1-i)\left(\frac{d-1}{2}\right)} \\
&\quad + (1-p) \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} e^{\epsilon/2} \cdot P_{i, t-(n-i)\left(\frac{d-1}{2}\right)-1} \\
&\quad + (1-p) \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot \frac{p(n-1-i)}{i+1} \cdot (e^{\epsilon/2} - 1) P_{i+1, t-(n-i-1)\left(\frac{d-1}{2}\right)-1}
\end{aligned}$$

$$
\begin{aligned}
&- (1-p) \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \cdot P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \\
&\geq (1-p) \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \Delta_i, \qquad (7)
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_i &:= p(1-e^{-\epsilon/2}) \cdot \left(P_{i+1,t-(n-1-i)\left(\frac{d-1}{2}\right)} + \frac{n-1-i}{i+1} \cdot P_{i+1,t-(n-i-1)\left(\frac{d-1}{2}\right)-1} \right) \\
&\quad + e^{\epsilon/2} \cdot P_{i,t-(n-i)\left(\frac{d-1}{2}\right)-1} - P_{i,t-(n-i)\left(\frac{d-1}{2}\right)},
\end{aligned}
$$

and we have used that $e^{\epsilon} - 1 \geq e^{\epsilon/2} - 1 \geq 1 - e^{-\epsilon/2}$ for $\epsilon \geq 0$.

By Lemma 10 with $m = i, k = t - (n-i)\left(\frac{d-1}{2}\right), \ell_1 = \left(\frac{d-1}{2}\right), \ell_2 = \left(\frac{d-3}{2}\right)$, we have

$$
\Delta_i \geq (e^{0.3\epsilon} - 1) P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \geq 0, \qquad (8)
$$

for all i such that $\frac{10 \log(1/(1-e^{-0.1\epsilon}))}{1-e^{-0.1\epsilon}} \leq i \leq n-1$. Let $i_0 := \frac{10 \log(1/(1-e^{-0.1\epsilon}))}{1-e^{-0.1\epsilon}}$. It remains to lower bound the terms in (7) for $0 \leq i < i_0$. To do so, we will "borrow" the additional mass of $(e^{0.3\epsilon} - 1)P_{i,t-(n-i)\left(\frac{d-1}{2}\right)}$ from the terms with $i \geq i_0$. To show that this borrowing gives sufficient positive mass from the terms $P_{i,t-(n-i)\left(\frac{d-1}{2}\right)}$ with $i \geq i_0$, we will use Lemma 11.

Next, let $i_{\max} \in \{0, 1, \ldots, i_0 - 1\}$ and $i_{\min} \in \{i_0, i_0 + 1, \ldots, 2p(n-1)\}$ be defined so that:

$$
\begin{aligned}
P_{i_{\max},t-(n-i_{\max})\left(\frac{d-1}{2}\right)} &\geq P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \quad \forall i \in \{0, 1, \ldots, i_0 - 1\} \\
P_{i_{\min},t-(n-i_{\min})\left(\frac{d-1}{2}\right)} &\leq P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \quad \forall i \in \{i_0, i_0 + 1, \ldots, 2p(n-1)\}.
\end{aligned}
$$

As $p = \frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{n(1-e^{-0.1\epsilon})}$ and $\epsilon < c_0 \leq 1$, we have that as long as c_0 is sufficiently small,

$$
2p(n-1) \leq \frac{100 e^{100} \log(5/\epsilon)}{0.1\epsilon} \leq \frac{1}{4\epsilon^2} = s^2/1000.
$$

It follows from Lemma 11 with $a = i_{\min} - i_{\max} \leq 2p(n-1)$ that

$$
P_{i_{\min},t-(n-i_{\min})\left(\frac{d-1}{2}\right)} \geq \frac{1}{40s^3} P_{i_{\max},t-(n-i_{\max})\left(\frac{d-1}{2}\right)}.
$$

Let $M \sim \mathrm{Binom}(n-1, p)$ be a binomial random variable. Then, as (8) holds for $n - 1 \geq i \geq i_0$, we have

$$
\begin{aligned}
&\sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \Delta_i \\
&\geq - \sum_{i=0}^{i_0-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \\
&\quad + \sum_{i=i_0}^{2p(n-1)} \binom{n-1}{i} p^i (1-p)^{n-1-i} (e^{0.3\epsilon} - 1) \cdot P_{i,t-(n-i)\left(\frac{d-1}{2}\right)} \\
&\geq -P_{i_{\max},t-(n-i_{\max})\left(\frac{d-1}{2}\right)} \sum_{i=0}^{i_0-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \qquad (9) \\
&\quad + 0.3\epsilon \cdot P_{i_{\min},t-(n-i_{\min})\left(\frac{d-1}{2}\right)} \sum_{i=i_0}^{2p(n-1)} \binom{n-1}{i} p^i (1-p)^{n-1-i}
\end{aligned}
$$

$$\geq P_{i_{\max},t-(n-i_{\max})\left(\frac{d-1}{2}\right)} \cdot \left(\frac{0.3\epsilon}{40s^3} \cdot \Pr[i_0 \leq M \leq 2p(n-1)] - \Pr[M < i_0]\right). \tag{10}$$

By the Chernoff bound, for sufficiently large n and since $pn = n \cdot \frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{n(1-e^{-0.1\epsilon})} \geq 100$, we have

$$\Pr[M > 2p(n-1)] \leq \exp(-p(n-1)/3) \leq \exp(-pn/4) < 1/2.$$

Moreover, since $i_0 = \frac{10 \log(1/(1-e^{-0.1\epsilon}))}{1-e^{-0.1\epsilon}} \leq pn/3 \leq p(n-1)/2$ and $\epsilon \leq 1$ in the current case,

$$\Pr[M < i_0] \leq \exp(-p(n-1)/8) \leq \exp(-pn/10) \leq \exp\left(\frac{10}{1-e^{-0.1\epsilon}}\right) < 1/4.$$

Hence, recalling $s = \frac{10}{\epsilon}$, $d = 4\left\lceil \frac{1000\, e^{100\epsilon}}{(1-e^{-0.1\epsilon})} \cdot \log\left(\frac{n}{1-e^{-0.1\epsilon}}\right)\right\rceil + 3$, $p = \frac{100\, e^{100\epsilon} \log(1/(1-e^{-0.1\epsilon}))}{n(1-e^{-0.1\epsilon})}$, as well as the assumption $\epsilon > 1/n^{2/3}$,

$$\begin{aligned} &\frac{0.3\epsilon}{40s^3} \cdot \Pr[i_0 \leq M \leq 2p(n-1)] - \Pr[M < i_0] \\ &\geq \frac{0.3\epsilon}{160s^3} - \exp(-1/(2\epsilon)) \\ &\geq c\epsilon^4 - \exp(-1/(2\epsilon)), \end{aligned}$$

for some sufficiently small positive absolute constant c. The above quantity is positive as long as $\exp(1/(2\epsilon)) \geq \frac{1}{c\epsilon^4}$, i.e., as long as $\epsilon \leq c'$ for some absolute constant $c' > 0$ (which holds as long as we select $c_0 \leq c'$). From this and (7), we can conclude that (5) holds. ◀

3.3 Proof Overview of Lemma 10

In this subsection, we give an overview of the proof of Lemma 10; we defer the full proof to the full version [24]. Throughout this section, we will use the several additional notation:

- First, we will overload the notation and use $\nu(z)$ to denote the probability mass function of ν at z, i.e., $\nu(z) := \Pr_{Z\sim\nu}[Z = z]$.
- We often represent a sequence of integers $a_1, \dots, a_m$ as a vector $\mathbf{a} = (a_1, \dots, a_m)$. For such a vector, we use $\nu(\mathbf{a})$ as a shorthand for the product $\nu(a_1)\cdots\nu(a_m)$, and $\text{zero}(\mathbf{a})$ as a shorthand for the number of zero coordinates, i.e., $\text{zero}(\mathbf{a}) := |\{i \in [m] \mid a_i = 0\}|$.
- We use $S_{m,k,d}$ to denote the set of all sequences of integers $a_1, \dots, a_m$ between 0 and d (inclusive) whose sum is k; more formally, $S_{m,k,d} := \Delta_{m,k} \cap [0,d]^m$. Since d will be fixed throughout, we omit d and simply write $S_{m,k}$.
- Next, for any $i \in \mathbb{R}$, we use $S^{\text{zero}<i}_{m,k}$ (resp. $S^{\text{zero}\geq i}_{m,k}$) to denote the sets of sequences in $S_{m,k}$ whose number of zero-coordinates is less than (resp., at least) i. More formally, $S^{\text{zero}<i}_{m,k} := \{\mathbf{a} \in S_{m,k} \mid \text{zero}(\mathbf{a}) < i\}$ and $S^{\text{zero}\geq i}_{m,k} := \{\mathbf{a} \in S_{m,k} \mid \text{zero}(\mathbf{a}) \geq i\}$.

To prove Lemma 10, let us observe that we may expand $P_{m,k}$ as

$$P_{m,k} = \sum_{\mathbf{a}\in S_{m,k}} \nu(\mathbf{a}) = \sum_{\mathbf{a}\in S^{\text{zero}<i}_{m,k}} \nu(\mathbf{a}) + \sum_{\mathbf{a}\in S^{\text{zero}\geq i}_{m,k}} \nu(\mathbf{a}),$$

where i will be chosen later in the proof.

We will bound the two terms on the right separately. More specifically, we will show that

$$\sum_{\mathbf{a}\in S^{\text{zero}<i}_{m,k}} \nu(\mathbf{a}) \leq e^{0.5\epsilon} \cdot P_{m,k-1}, \tag{11}$$

and that for $\ell_1, \ell_2 \in \{\frac{d-1}{2}, \frac{d-3}{2}\}$,

$$\sum_{\mathbf{a} \in S_{m,k}^{\text{zero}\geq i}} \nu(\mathbf{a}) \leq p(1 - e^{-0.5\epsilon}) \cdot \left(P_{m+1,k+\ell_1} + \frac{(n-m+1)}{m+1} \cdot P_{m+1,k+\ell_2}\right). \tag{12}$$

Once we have these two inequalities, Lemma 10 immediately follows. The intuition behind the two inequalities is quite simple. For (11), since each sequence $\mathbf{a} \in S_{m,k}^{\text{zero}<i}$ contains few zeros, we should be able to pick a non-zero a_i and decrease it by one and end up with a sequence in $S_{m,k-1}$ instead; since the discrete Laplace distribution's mass (i.e., ν) on a_i and on $a_i - 1$ differs (multiplicatively) by a factor of at most $e^{1/s}$, the mass of the modified sequence also differs from the original sequence by a factor of $e^{1/s}$.

For (12), the intuition is pretty similar. We start with a sequence $\mathbf{a} \in S_{m,k}^{\text{zero}\geq i}$ and we will modify it to end up with a sequence in $S_{m+1,k+\ell}$ where ℓ is either $\left(\frac{d-1}{2}\right)$ or $\left(\frac{d-3}{2}\right)$. The intuition here is that since $\mathbf{a}$ contains many zero coordinates, there are many ways for us to divide ℓ among these zero coordinates and an additional coordinate, which would result naturally in a sequence in $P_{m+1,k+\ell}$.

To turn the intuition into a formal proof, we need to be careful about "double counting" a modified sequence. As an example, for (11), suppose we would like to modify a sequence in $S_{m,k}^{\text{zero}<i}$ to one in $S_{m,k-1}$ by decreasing any non-zero coordinate. Then, it is possible that two sequences $(1, 0, a_3, \dots, a_n)$ and $(0, 1, a_3, \dots, a_n)$ results in the same sequence $(0, 0, a_3, \dots, a_n)$.

In order to avoid such "double counting", we divide our proofs into two parts. First, we show that we may replace $S_{m,k}^{\text{zero}<i}$ (resp. $S_{m,k}^{\text{zero}\geq i}$) with the set of sequences whose first coordinate is non-zero (resp., whose first few coordinates are zeros).

4 Lower Bound for Binary Summation

We next prove our lower bound on the communication complexity of any non-interactive pure-DP$_{\text{shuffled}}$ protocol that can perform bit addition with small error (Theorem 4). To prove our lower bound, first recall the following notion from probability theory.

▶ **Definition 12** (Moment Generating Function). *Let $\mathbf{Y}$ be a $\mathbb{R}^k$-valued random variable for some $k \in \mathbb{N}$. Its* moment generating function (MGF) *is defined as $\mathbf{M}_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}[e^{\langle \mathbf{t}, \mathbf{Y} \rangle}]$.*

Throughout this section, we will be dealing with pairs of random variables whose MGFs are within a certain factor of each other. The following definition will be particularly handy.

▶ **Definition 13** (Bounded MGF ratio). *We say that two $\mathbb{R}^k$-valued random variables $\mathbf{Y}, \mathbf{Y}'$ have e^ϵ-*bounded MGF ratio *if and only if, for all $\mathbf{t} \in \mathbb{R}^k$ we have that $\frac{\mathbf{M}_{\mathbf{Y}}(\mathbf{t})}{\mathbf{M}_{\mathbf{Y}'}(\mathbf{t})} \in [e^{-\epsilon}, e^{\epsilon}]$.*

Furthermore, let $\mathrm{SD}(\mathbf{Y}, \mathbf{Y}')$ denote the *total variation distance* between them.

Our proofs follow the outline from Section 1.3. Specifically, in Section 4.1, we prove that a pure-DP$_{\text{shuffled}}$ protocol implies bounded MGF ratio. Then, in Section 4.2, we give a lower bound on $C_{\epsilon,\gamma}$ from Definition 5, which then implies Theorem 4. Finally, in Section 4.3, we provide an example which shows that our lower bound for the question is tight.

4.1 Pure-DP Implies Bounded MGF Ratio

We start by proving a necessary (but not sufficient) condition on ϵ-DP protocols in terms of the MGFs of $\mathbf{X}^0, \mathbf{X}^1$. A straightforward observation we will use is the following:

▶ **Observation 14.** *Let $\mathbf{Y}, \mathbf{Y}'$ be two random variables with the same support $\mathrm{supp}(\mathbf{Y}) = \mathrm{supp}(\mathbf{Y}') \subseteq \mathbb{R}^k$ such that $\frac{\Pr[\mathbf{Y}=\mathbf{v}]}{\Pr[\mathbf{Y}'=\mathbf{v}]} \in [e^{-\epsilon}, e^{\epsilon}]$. Then, $\mathbf{Y}, \mathbf{Y}'$ satisfies e^ϵ-bounded MGF ratio.*

Furthermore, recall a (well-known) multiplicative property of MGF that, if $\mathbf{Y}, \mathbf{Y}' \in \mathbb{R}^k$ are two independent random variables, then $\mathbf{M}_{\mathbf{Y}+\mathbf{Y}'}(\mathbf{t}) = \mathbf{M}_{\mathbf{Y}}(\mathbf{t}) \cdot \mathbf{M}_{\mathbf{Y}'}(\mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^k$. With this in mind, we can now prove the desired result:

▶ **Lemma 15.** *For any ϵ-DP protocol, $\mathbf{X}^0, \mathbf{X}^1$ must satisfy e^ϵ-bounded MGF ratio.*

Proof. Consider two input vectors $0 \dots 00$ and $0 \dots 01$. Let $\mathbf{Y}^0, \mathbf{Y}^1 \in \mathbb{Z}^k$ denote the views of the shuffled output on the corresponding input vectors, where $\mathbf{Y}^0_j$ denote the number of j's received by the analyzer for the input vector $0 \dots 00$ and $\mathbf{Y}^1_j$ denote the number of j's received by the analyzer for the input vector $0 \dots 01$. Notice that $\mathbf{Y}^0$ is a sum of n i.i.d. copies of $\mathbf{X}^0$ and $\mathbf{Y}^1$ is a sum of $(n-1)$ i.i.d. copies of $\mathbf{X}^0$ and a copy of $\mathbf{X}^1$. Observe also that ϵ-DP implies that $\mathbf{Y}^0, \mathbf{Y}^1$ satisfy the condition in Observation 14. From this, we have

$$[e^{-\epsilon}, e^\epsilon] \ni \frac{\mathbf{M}_{\mathbf{Y}^0}(\mathbf{t})}{\mathbf{M}_{\mathbf{Y}^1}(\mathbf{t})} = \frac{(\mathbf{M}_{\mathbf{X}^0}(\mathbf{t}))^n}{(\mathbf{M}_{\mathbf{X}^0}(\mathbf{t}))^{n-1} \cdot \mathbf{M}_{\mathbf{X}^1}(\mathbf{t})} = \frac{\mathbf{M}_{\mathbf{X}^0}(\mathbf{t})}{\mathbf{M}_{\mathbf{X}^1}(\mathbf{t})}.$$ ◀

4.2 From Bounded MGF Ratio to Communication Lower Bound

We will now use the bounded MGF ratio property to bound the communication complexity of any non-interactive protocol for summation that incurs small error. To do so, let us recall below a known result that any protocol that can perform binary summation to within a small error must have large statistical distance between $\mathbf{X}^0$ and $\mathbf{X}^1$:

▶ **Theorem 16** ([14]). *Any non-interactive protocol that can perform binary summation to within an expected absolute error of α must satisfy* $\mathrm{SD}(\mathbf{X}^0, \mathbf{X}^1) \geq 1 - O\left(\frac{\alpha}{\sqrt{n}}\right)$.

Thanks to Lemma 15 and Theorem 16, to prove our lower bound (Theorem 4), it now suffices to show that, for any $\mathbf{Y}, \mathbf{Y}'$ whose supports lie in $\Delta_{k,m}$ that satisfy e^ϵ-bounded MGF ratio and $\mathrm{SD}(\mathbf{Y}, \mathbf{Y}')$ is large, then $m \log k$ must be large. The main lemma of this subsection, which is a quantitative version of the aforementioned statement, is stated formally below.

▶ **Lemma 17.** *Let $\mathbf{Y}, \mathbf{Y}'$ be two random variables supported on $\Delta_{k,m}$ with e^ϵ-bounded MGF ratio. Then,* $\mathrm{SD}(\mathbf{Y}, \mathbf{Y}') \leq 1 - 2^{-O_\epsilon(m^2 \log k)}$.

It is straightforward to see that Lemma 17, Lemma 15, and Theorem 16 together implies Theorem 4. We devote the rest of this subsection to the proof of Lemma 17.

Dual Approach and Proof of Lemma 17

For notational convenience, we use $p_{\mathbf{y}}$ and $p'_{\mathbf{y}}$ to denote $\Pr[\mathbf{Y} = \mathbf{y}]$ and $\Pr[\mathbf{Y}' = \mathbf{y}]$ respectively.

Before we formalize the proof below, let us first present an informal overview. Recall that $1 - \mathrm{SD}(\mathbf{Y}, \mathbf{Y}')$ is equal to $\sum_{\mathbf{y} \in \Delta_{k,m}} \min\{p_{\mathbf{y}}, p'_{\mathbf{y}}\} = \min_{S \subseteq \Delta_{k,m}} \left\{\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}}\right\}$. Hence, it suffices for us to show that, for every $S \subseteq \Delta_{k,m}$, we have

$$\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}} \geq 2^{-O_\epsilon(m^2 \log k)}. \tag{13}$$

We will give a "dual certificate" for this statement. Notice that since the total probability of each of $\mathbf{Y}, \mathbf{Y}'$ must be one, we have $\sum_{\mathbf{y} \in \Delta_{k,m}} p_{\mathbf{y}} = 1$ and $\sum_{\mathbf{y} \in \Delta_{k,m}} p'_{\mathbf{y}} = 1$. We also have the non-negativity constraints: $p_{\mathbf{y}}, p'_{\mathbf{y}} \geq 0$ for all $\mathbf{y} \in \Delta_{k,m}$. Finally, the e^ϵ-bounded MGF ratio property between $\mathbf{Y}$ and $\mathbf{Y}'$ translates to the following linear inequalities for all $\mathbf{t} \in \mathbb{R}^k$:

$$\sum_{\mathbf{y} \in \Delta_{k,m}} e^{\langle \mathbf{t}, \mathbf{y} \rangle} \cdot p'_{\mathbf{y}} - \sum_{\mathbf{y} \in \Delta_{k,m}} e^{\langle \mathbf{t}, \mathbf{y} \rangle - \epsilon} \cdot p_{\mathbf{y}} \geq 0, \sum_{\mathbf{y} \in \Delta_{k,m}} e^{\langle \mathbf{t}, \mathbf{y} \rangle} \cdot p_{\mathbf{y}} - \sum_{\mathbf{y} \in \Delta_{k,m}} e^{\langle \mathbf{t}, \mathbf{y} \rangle - \epsilon} \cdot p'_{\mathbf{y}} \geq 0. \tag{14}$$

Hence, we have a system of linear inequalities and we would like to certify a linear inequality (13). We may do so by writing (13) as a linear combination of the constraints.

As a wishful thinking, if we could somehow "extract" only the $p_{\mathbf{y}}$ and $p'_{\mathbf{y}}$ terms from (14), we would be done because we would simply have $e^\epsilon \cdot p_{\mathbf{y}} \geq p'_{\mathbf{y}} \geq e^{-\epsilon} \cdot p_{\mathbf{y}}$ which can easily be combined with the total probability and non-negativity constraints to get a good bound on $\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}}$. Of course, such extraction is not possible since, for any value $\mathbf{t}$ we plug into (14), we always get non-zero coefficients for all vectors in $\Delta_{k,m}$, not just $\mathbf{y}$.

With this in mind, our relaxed goal is to select $\mathbf{t} = \tau(\mathbf{y})$ for each $\mathbf{y}$ in such a way that the coefficient of $\mathbf{y}$ from its own inequality (i.e., $\mathbf{t} = \tau(\mathbf{y})$) "dominates" the coefficients of $\mathbf{y}$ from other inequalities (i.e., $\mathbf{t} = \tau(\mathbf{y}')$ for $\mathbf{y}' \neq \mathbf{y}$). A formal version of the statement is proved below. Note that $e^{\beta(\mathbf{y})}$ should be thought of as the "scaling factor" for the inequality for $\mathbf{y}$.

▶ **Lemma 18.** *For any $\epsilon > 0$, there exist mappings $\tau : \Delta_{k,m} \to \mathbb{R}^k$ and $\beta : \Delta_{k,m} \to \mathbb{R}$ such that the following hold for all $\mathbf{y} \in \Delta_{k,m}$:*

$$0 \geq \langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y}) \geq \zeta := -O_\epsilon(m^2 \log k), \tag{15}$$

and

$$e^{\langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y})} \geq 2e^\epsilon \cdot \sum_{\mathbf{y}' \in \Delta_{k,m} \setminus \{\mathbf{y}\}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle + \beta(\mathbf{y}')}. \tag{16}$$

Proof. Let $\rho = \epsilon + 10 \ln(k+1) + 10$. We pick $\tau(\mathbf{y}) = \rho \cdot 2\mathbf{y}$ and $\beta(\mathbf{y}) = \rho \cdot \left(-\|\mathbf{y}\|_2^2 - m^2\right)$. It is obvious to see that (15) holds. To prove (16), let us first observe the following identity:

$$\langle \tau(\mathbf{y}'), \mathbf{y} \rangle + \beta(\mathbf{y}') = \langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y}) - \rho \cdot \|\mathbf{y} - \mathbf{y}'\|_2^2. \tag{17}$$

Thus, we may rewrite the right hand side of (16) as

$$\sum_{\mathbf{y}' \in \Delta_{k,m} \setminus \{\mathbf{y}\}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle + \beta(\mathbf{y}')}$$
$$\overset{(17)}{=} e^{\langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y})} \cdot \left(\sum_{i=1}^{2m^2} e^{-\rho i} \cdot \left| \{ \mathbf{y}' \in \Delta_{k,m} \mid \|\mathbf{y} - \mathbf{y}'\|_2^2 = i \} \right| \right). \tag{18}$$

Furthermore, we have $|\{\mathbf{y}' \in \Delta_{k,m} \mid \|\mathbf{y} - \mathbf{y}'\|_2^2 = i\}| \leq |\{\mathbf{z} \in \mathbb{Z}^k \mid \|\mathbf{z}\|_2^2 = i\}| \leq 2^i \cdot \binom{k+i-1}{i} \leq (2e(k+1))^i$. Plugging this back into (18), we have

$$\sum_{\mathbf{y}' \in \Delta_{k,m} \setminus \{\mathbf{y}\}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle + \beta(\mathbf{y}')} \leq e^{\langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y})} \cdot \left(\sum_{i=1}^{2m^2} \left(e^{-\rho} \cdot 2e(k+1) \right)^i \right)$$
$$(\text{From our choice of } \rho) \leq e^{\langle \tau(\mathbf{y}), \mathbf{y} \rangle + \beta(\mathbf{y})} \cdot \frac{1}{2e^\epsilon}.$$ ◀

With Lemma 18 ready, we can now prove Lemma 17.

Proof of Lemma 17. Let τ, β be as in Lemma 18. Consider any set $S \subseteq \Delta_{k,m}$.

For every $\mathbf{y}' \in S$, $\mathbf{M}_{\mathbf{Y}}(\tau(\mathbf{y}')) \geq e^{-\epsilon} \cdot \mathbf{M}_{\mathbf{Y}'}(\tau(\mathbf{y}'))$ is equivalent to

$$\sum_{\mathbf{y} \in \Delta_{k,m}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} \cdot p_{\mathbf{y}} - \sum_{\mathbf{y} \in \Delta_{k,m}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \cdot p'_{\mathbf{y}} \geq 0. \tag{19}$$

Similarly, for every $\mathbf{y}' \in \Delta_{k,m} \setminus S$, $\mathbf{M}_{\mathbf{Y}'}(\tau(\mathbf{y}')) \geq e^{-\epsilon} \cdot \mathbf{M}_{\mathbf{Y}}(\tau(\mathbf{y}'))$ can be rearranged as

$$\sum_{\mathbf{y} \in \Delta_{k,m}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} \cdot p'_{\mathbf{y}} - \sum_{\mathbf{y} \in \Delta_{k,m}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \cdot p_{\mathbf{y}} \geq 0. \tag{20}$$

By adding (19) for all $\mathbf{y}' \in S$ with (20) for all $\mathbf{y}' \in \Delta_{k,m} \setminus S$, we have

$$\begin{aligned} &\sum_{\mathbf{y} \in \Delta_{k,m}} \left(\sum_{\mathbf{y}' \in S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} - \sum_{\mathbf{y}' \in \Delta_{k,m} \setminus S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \right) p_{\mathbf{y}} \\ &+ \sum_{\mathbf{y} \in \Delta_{k,m}} \left(\sum_{\mathbf{y}' \in \Delta_{k,m} \setminus S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} - \sum_{\mathbf{y}' \in S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \right) p'_{\mathbf{y}} \geq 0. \end{aligned} \tag{21}$$

Now, for all $\mathbf{y} \in S$, we can upper bound the coefficient of $p'_{\mathbf{y}}$ in (21) by

$$\begin{aligned} &\sum_{\mathbf{y}' \in \Delta_{k,m} \setminus S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} - \sum_{\mathbf{y}' \in S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \\ &\leq \sum_{\mathbf{y}' \in \Delta_{k,m} \setminus \{\mathbf{y}\}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} - e^{\left\langle \tau(\mathbf{y}), \mathbf{y} \right\rangle - \beta(\mathbf{y}) - \epsilon} \overset{(16)}{\leq} -0.5 e^{\left\langle \tau(\mathbf{y}), \mathbf{y} \right\rangle - \beta(\mathbf{y}) - \epsilon} \overset{(15)}{\leq} -e^{\zeta - 1 - \epsilon}. \end{aligned}$$

Similarly, for all $\mathbf{y} \in \Delta_{k,m} \setminus S$, the coefficient of $p_{\mathbf{y}}$ in (21) is at most $-e^{\zeta-1-\epsilon}$.

Moreover, for all $\mathbf{y} \in S$, we can upper bound the coefficient in (21) of $p_{\mathbf{y}}$ by

$$\begin{aligned} &\sum_{\mathbf{y}' \in S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} - \sum_{\mathbf{y}' \in \Delta_{k,m} \setminus S} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}') - \epsilon} \\ &\leq \sum_{\mathbf{y}' \in \Delta_{k,m}} e^{\left\langle \tau(\mathbf{y}'), \mathbf{y} \right\rangle - \beta(\mathbf{y}')} \overset{(16)}{\leq} \left(1 + \frac{1}{2e^{\epsilon}}\right) e^{\left\langle \tau(\mathbf{y}), \mathbf{y} \right\rangle - \beta(\mathbf{y})} \overset{(15)}{\leq} 2. \end{aligned}$$

Similarly, for all $\mathbf{y} \in S$, the coefficient of $p'_{\mathbf{y}}$ in (21) is at most 2.

Plugging these back into (21), we have

$$0 \leq 2 \left(\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}} \right) - e^{\zeta - 1 - \epsilon} \left(\sum_{\mathbf{y} \in S} p'_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p_{\mathbf{y}} \right).$$

Using the fact that $\sum_{\mathbf{y} \in \Delta_{k,m}} p_{\mathbf{y}} = \sum_{\mathbf{y} \in \Delta_{k,m}} p'_{\mathbf{y}} = 1$, we can simplify the RHS above to

$$2e^{\zeta - 1 - \epsilon} \leq (2 + e^{\zeta - 1 - \epsilon}) \left(\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}} \right).$$

This means that

$$\left(\sum_{\mathbf{y} \in S} p_{\mathbf{y}} + \sum_{\mathbf{y} \in \Delta_{k,m} \setminus S} p'_{\mathbf{y}} \right) \geq \frac{2e^{\zeta - 1 - \epsilon}}{2 + e^{\zeta - 1 - \epsilon}} \overset{(15)}{\geq} 2^{-O_\epsilon(m^2 \log k)}.$$

This establishes (13) and hence we have $\mathrm{SD}(\mathbf{Y}, \mathbf{Y}') \leq 1 - 2^{-O_\epsilon(m^2 \log k)}$ as desired. ◀

4.3 Limitations of the Lower Bound Approach

In this subsection, we argue that the bound we achieve in Lemma 17 is essentially tight, even for $k = 2$. In other words, our approach of using only bounded MGF ratio property and the total variation distance bound from Theorem 16 cannot give any lower bound better than $O_\epsilon(\sqrt{\log n})$. Specifically, the main lemma of this section is stated below.

▶ **Lemma 19.** *For every $\epsilon > 0$ and $\gamma \in (0, 0.5)$, there exist two random variables $\mathbf{Y}, \mathbf{Y}'$ supported on (subsets of) $\Delta_{2,m}$ for some $m = O_\epsilon(\sqrt{\log(1/\gamma)})$ such that $\mathrm{SD}(\mathbf{Y}, \mathbf{Y}') \geq 1 - \gamma$ and that $\mathbf{Y}, \mathbf{Y}'$ satisfy the e^ϵ-bounded MGF ratio property.*

Similar to when we analyze our binary summation protocol in Section 3, it suffices to prove the following one-dimensional version of the above statement, where the two random variables are from $\{0, 1, \dots, m\}$ rather than $\Delta_{2,m}$.

▶ **Lemma 20.** *For every $\epsilon > 0$ and $\gamma \in (0, 0.5)$, there exist two random variables Y^0 and Y^1 supported on $\{0, \dots, m\}$ for some $m = O_\epsilon(\sqrt{\log(1/\gamma)})$ such that $\mathrm{SD}(Y^0, Y^1) \geq 1 - \gamma$ and that Y^0, Y^1 satisfy e^ϵ-bounded MGF ratio property.*

4.3.1 Discrete Gaussian Distributions

Our construction for Lemma 20 is based on the discrete Gaussian distribution, which can be defined as follows. Let the (one-dimensional) Gaussian function centered at c with parameter s as $\rho_{s,c}(x) := \exp\left(-\frac{\pi(x-c)^2}{s^2}\right)$ for all $x \in \mathbb{R}$. For any countable $A \subseteq \mathbb{R}$, let $\rho_{s,c}(A) := \sum_{x \in A} \rho_{s,c}(x)$. When $\rho_{s,c}(A)$ is finite, the discrete Gaussian distribution over A centered at c with parameter s, denoted by $\mathcal{D}_{A,s,c}$, has $\mathcal{D}_{A,s,c}(x) = \frac{\rho_{s,c}(x)}{\rho_{s,c}(A)}$ for all $x \in A$.

We will use a special case of a well-known property of lattices (cf. [35, 23, 3]). Recall that a one-dimensional lattice is $a\mathbb{Z} := \{at \mid t \in \mathbb{Z}\}$ for some $a \in \mathbb{R}^+$. Informally, the property states that, if for sufficiently large s , "shifting" the discrete Gaussian distribution by c does not change its normalization factor too much. This is formalized below.

▶ **Lemma 21** (e.g. [23, Lemma 2.6]). *For any constants $a, \delta \in \mathbb{R}^+$, there exists a sufficiently large constant $s^* = s^*(a, \delta)$ such that, for any $c \in \mathbb{R}$, we have $\frac{\rho_{s^*,c}(a\mathbb{Z})}{\rho_{s^*,0}(a\mathbb{Z})} \in [e^{-\delta}, 1]$.*

We will also use the following (rather straightforward) observation that, similar to the (continuous) Gaussian distribution, we may choose a sufficiently large truncation point $\ell^* a$ for which the total mass of all points x with $|X - c| > \ell^* a$ is arbitrarily small.

▶ **Observation 22.** *For any constants $a, \delta \in \mathbb{R}^+$, let $s^* = s^*(a, \delta)$ be as in Lemma 21. Then, for any $\lambda > 0$, there exists a sufficiently large positive integer $\ell^* = \ell^*(a, \delta, \lambda)$ such that, for any $c \in \mathbb{R}$, we have $\Pr_{X \sim \mathcal{D}_{a\mathbb{Z},s^*,c}}[|X - c| > \ell^* a] \leq \lambda$.*

4.3.2 Proof Overview of Lemma 20

Distributions of both Y^0, Y^1 will place $\frac{\gamma}{2}$ probability masses at each of 0 and m, and these two points shared by the supports of Y^0 and Y^1. (This ensures that the total variation distance of Y^0 and Y^1 are at least $1 - \gamma$.) In the middle, we then place discrete Gaussian distributions centered at $c = m/2$ for Y^0 and Y^1, with that of Y^0 only supported on even numbers whereas that of Y^1 supported on odd numbers. These discrete Gaussian distributions are truncated so that the supports are within the range of $[c - w, c + w]$ for some parameter w.

The intuition behind the construction is as follows. First, when $|t| \geq O_\epsilon(\sqrt{\log(1/\gamma)})$, it is not hard to see that the MGFs at t are dominated by the terms corresponding to the points

0 or m. Our parameters are selected in such a way that, when this is not the case, it must be that $|t| \ll w$. In this case, we observe that the MGFs of discrete Guassian distributions are simply proportional to normalization terms of other discrete Gaussian distributions, shifted by $O(t)$ (and truncated appropriately). Since $|t| \ll w$, we can then apply Lemma 21 and Observation 22 to get a good bound on these terms. This concludes the main ideas in the proof; due to space constraints, the full proof is deferred to the full version [24].

5 Conclusions and Open Questions

We gave the first pure-DP$_{\text{shuffled}}$ protocols for binary and real summation with constant error. We further prove a communication lower bound for any non-interactive protocols for binary summation. While these have advanced our understanding of pure-DP$_{\text{shuffled}}$ protocols, many questions remain open after this work. Specifically, the immediate open questions are:

- Can we improve the error guarantee in the (binary and real) summation protocols to achieve the asymptotically optimal error $1/\epsilon$, which can be achieved in DP$_{\text{central}}$ [19]?
- What is the optimal per user communication complexity of non-interactive DP$_{\text{shuffled}}$ protocols for binary and real summation? As we have shown, the communication complexity for binary summation lies between $O_\epsilon(\log n)$ and $\Omega_\epsilon(\sqrt{\log n})$. On the other hand, for real summation, the only lower bound is the trivial $\Omega(\log n)$ bound (which holds even without privacy concerns) whereas our upper bound is $O_\epsilon(\log^3 n)$.
- In Appendix A of the full version [24], we show that our binary summation protocol also yields a pure-DP protocol for histograms with error $O_\epsilon(\log B \log n)$ but with $O_\epsilon(B \log n)$ per user communication complexity. The latter is in contrast to the approximate-DP multi-message protocol of [25], which has a per user communication complexity of only $O_\epsilon(poly(\log n, \log B))$ and incurs a similar error. It is thus an interesting open question to come up with (or rule out) a pure-DP protocol with a smaller communication complexity.
- Can we exploit interactivity to break our $\Omega_\epsilon(\sqrt{\log n})$ communication lower bound? Alternately, can we prove any non-trivial lower bound that holds also with interaction?

On a high-level, it would also be interesting to develop tools to help prove guarantees for pure-DP$_{\text{shuffled}}$ protocols. In the case of approximate-DP, there are amplification theorems [21, 8] that can yield an approximate-DP$_{\text{shuffled}}$ protocol from a DP$_{\text{local}}$ protocol. Although this may not be optimal in some cases (as shown by the multi-message protocols in [25, 7, 26]), such theorems can be conveniently applied to a large class of protocols and yield good approximate-DP guarantees. On the other hand, our proofs in this work are specific to our carefully designed protocols. It would be much more convenient if one can give a unifying theorem that proves pure privacy guarantees for any protocol with easily verifiable conditions.

References

1 Martín Abadi, Andy Chu, Ian J. Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *CCS*, pages 308–318, 2016. doi:10.1145/2976749.2978318.

2 John M Abowd. The US Census Bureau adopts differential privacy. In *KDD*, pages 2867–2867, 2018.

3 Shweta Agrawal, Craig Gentry, Shai Halevi, and Amit Sahai. Discrete Gaussian leftover hash lemma over infinite domains. In *ASIACRYPT*, pages 97–116, 2013. doi:10.1007/978-3-642-42033-7_6.

4 Apple Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 2017.

5 Victor Balcer and Albert Cheu. Separating local & shuffled differential privacy via histograms. *arXiv: 1911.06879*, 2019.

6 Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Differentially private summation with multi-message shuffling. *arXiv: 1906.09116*, 2019.

7 Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Improved summation from shuffling. *arXiv: 1909.11225*, 2019.

8 Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. The privacy blanket of the shuffle model. In *CRYPTO*, pages 638–667, 2019. doi:10.1007/978-3-030-26951-7_22.

9 Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Private summation in the multi-message shuffle model. *arXiv: 2002.00817*, 2020.

10 Amos Beimel, Kobbi Nissim, and Eran Omri. Distributed private data analysis: Simultaneously solving how and what. In *CRYPTO*, pages 451–468, 2008.

11 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.

12 Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *STOC*, pages 609–618, 2008.

13 Mark Bun, Jelani Nelson, and Uri Stemmer. Heavy hitters and the structure of local privacy. In *PODS*, pages 435–447, 2018.

14 T.-H. Hubert Chan, Elaine Shi, and Dawn Song. Optimal lower bound for differentially private multi-party aggregation. In *ESA*, pages 277–288, 2012. doi:10.1007/978-3-642-33090-2_25.

15 Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *EUROCRYPT*, pages 375–403, 2019. doi:10.1007/978-3-030-17653-2_13.

16 Anindya De. Lower bounds in differential privacy. In *TCC*, pages 321–338, 2012. doi:10.1007/978-3-642-28914-9_18.

17 Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *NIPS*, pages 3571–3580, 2017.

18 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.

19 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

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:2001.03618*, 2020.

21 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.

22 Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *CCS*, pages 1054–1067, 2014.

23 Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *STOC*, pages 197–206, 2008. doi:10.1145/1374376.1374407.

24 Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Pure differentially private summation from anonymous messages. *arXiv: 2002.01919*, 2020. arXiv:2002.01919.

25 Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. On the power of multiple anonymous messages. Cryptology ePrint Archive, Report 2019/1382, 2019.

26 Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. *arXiv: 1909.11073*, 2019.

27 Badih Ghazi, Rasmus Pagh, and Ameya Velingker. Scalable and differentially private distributed aggregation in the shuffled model. *arXiv: 1906.08320*, 2019.

28 Andy Greenberg. Apple's "differential privacy" is about collecting your data – but not your data. *Wired, June*, 13, 2016.

29 Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *FOCS*, pages 61–70, 2010.

30 Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *STOC*, pages 705–714, 2010.

31 Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography from anonymity. In *FOCS*, pages 239–248, 2006.

32 Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and open problems in federated learning. *arXiv: 1912.04977*, 2019.

33 Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Rashkodnikova, and Adam Smith. What can we learn privately? In *FOCS*, pages 531–540, 2008.

34 Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv: 1610.05492*, 2016.

35 Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. *SICOMP*, 37(1):267–302, 2007. doi:10.1137/S0097539705447360.

36 Ilya Mironov. Rényi differential privacy. In *CSF*, pages 263–275, 2017. doi:10.1109/CSF.2017.11.

37 Aleksandar Nikolov, Kunal Talwar, and Li Zhang. On the geometry of differential privacy: the sparse and approximate cases. In *STOC*, pages 351–360, 2013.

38 Stephen Shankland. How Google tricks itself to protect Chrome user privacy. *CNET, October*, 2014.

39 Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *arXiv: 1501.06095*, 2015.

40 Salil Vadhan. *The Complexity of Differential Privacy*, pages 347–450. Springer International Publishing, 2017.

41 Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *JASA*, 60(309):63–69, 1965.