SIAM J. COMPUT.
Vol. 47, No. 5, pp. 1888–1938

FINGERPRINTING CODES AND THE PRICE OF APPROXIMATE DIFFERENTIAL PRIVACY*

MARK BUN†, JONATHAN ULLMAN†, AND SALIL VADHAN†

Abstract. We show new information-theoretic lower bounds on the sample complexity of (ε, δ)-differentially private algorithms that accurately answer large sets of counting queries. A counting query on a database $D \in (\{0,1\}^d)^n$ has the form "What fraction of the individual records in the database satisfy the property q?" We show that in order to answer an arbitrary set $\mathcal{Q}$ of $\gg d/\alpha^2$ counting queries on D to within error $\pm\alpha$ it is necessary that $n \geq \tilde{\Omega}(\sqrt{d}\log|\mathcal{Q}|/\alpha^2\varepsilon)$. This bound is optimal up to polylogarithmic factors, as demonstrated by the private multiplicative weights algorithm (Hardt and Rothblum, FOCS'10). In particular, our lower bound is the first to show that the sample complexity required for accuracy and (ε, δ)-differential privacy is asymptotically larger than what is required merely for accuracy, which is $O(\log|\mathcal{Q}|/\alpha^2)$. In addition, we show that our lower bound holds for the specific case of k-way marginal queries (where $|\mathcal{Q}| = 2^k\binom{d}{k}$) when α is not too small compared to d (e.g., when α is any fixed constant). Our results rely on the existence of short *fingerprinting codes* (Boneh and Shaw, CRYPTO'95; Tardos, STOC'03), which we show are closely connected to the sample complexity of differentially private data release. We also give a new method for combining certain types of sample-complexity lower bounds into stronger lower bounds.

Key words. differential privacy, fingerprinting codes, privacy attacks

AMS subject classification. 68Q17

DOI. 10.1137/15M1033587

1. Introduction.

Consider a database $D \in \mathcal{X}^n$, in which each of the n rows corresponds to an individual's record, and each record is an element of some data universe $\mathcal{X}$ (e.g., $\mathcal{X} = \{0,1\}^d$, corresponding to d binary attributes per record). The goal of privacy-preserving data analysis is to enable rich statistical analyses on such a database while protecting the privacy of the individuals. It is especially desirable to achieve (ε, δ)-*differential privacy* [21, 20], which (for suitable choices of ε and δ) guarantees that no individual's data have a significant influence on the information released about the database. A natural way to measure the trade-off between these two goals is via *sample complexity*—the minimum number of records n such that there exists a (possibly computationally unbounded) algorithm that achieves both differential privacy and statistical accuracy.

Some of the most basic statistics are *counting queries*, which are queries of the form "What fraction of individual records in D satisfy some property q?" In particular, we would like to design an algorithm that takes as input a database D and, for some family of counting queries $\mathcal{Q}$, outputs an approximate answer to each of the queries in $\mathcal{Q}$ that is accurate to within, say, $\pm.01$. Suppose we are given a bound on the number of queries $|\mathcal{Q}|$ and the dimensionality of the database records d, but

*Received by the editors August 3, 2015; accepted for publication (in revised form) March 16, 2017; published electronically October 30, 2018. A preliminary version of this work appeared in the Symposium on the Theory of Computing 2014.

http://www.siam.org/journals/sicomp/47-5/M103358.html

Funding: This work was supported by NSF grant CNS-1237235. The first author was supported by an NDSEG Fellowship. The third author was supported by a gift from Google, and a Simons Investigator Award.

†School of Engineering and Applied Sciences & Center for Research on Computation and Society, Harvard University, Cambridge, MA 02138 (mbun@seas.harvard.edu, jullman@ccs.neu.edu, salil@seas.harvard.edu).

otherwise allow the family $\mathcal{Q}$ to be arbitrary. What is the sample complexity required to achieve (ε, δ)-differential privacy and statistical accuracy for $\mathcal{Q}$?

Of course, if we drop the requirement of privacy, then we could achieve perfect accuracy when D contains any number of records. However, in many interesting settings the database D consists of random samples from some larger population, and an analyst is actually interested in answering the queries on the population. Thus, even without a privacy constraint, D would need to contain enough records to ensure that (with high probability) for every query $q \in \mathcal{Q}$, the answer to q on D is close to the answer to q on the whole population, say within $\pm.01$. To achieve this form of *statistical accuracy*, it is well known that it is necessary and sufficient for D to contain $\Theta(\log |\mathcal{Q}|)$ samples.[1] In this work we consider whether there is an additional "price of differential privacy" if we require both statistical accuracy and (ε, δ)-differential privacy (for, say, $\varepsilon = O(1)$, $\delta = o(1/n)$). This benchmark has often been used to evaluate the utility of differentially private algorithms, beginning with the seminal work of Dinur and Nissim [16].

Some of the earliest work in differential privacy [16, 26, 8, 21] gave an algorithm—the so-called *Laplace mechanism*—whose sample complexity is $\widetilde{\Theta}(|\mathcal{Q}|^{1/2})$, and thus incurs a large price of differential privacy. Fortunately, a remarkable result of Blum, Ligett, and Roth [9] showed that the dependence on $|\mathcal{Q}|$ can be improved exponentially to $O(d \log |\mathcal{Q}|)$ where d is the dimensionality of the data. Their work was improved on in several important aspects [23, 27, 44, 35, 32, 34]. The current best upper bound on the sample complexity is $O(\sqrt{d} \log |\mathcal{Q}|)$, which is obtained via the private multiplicative weights mechanism of Hardt and Rothblum [35].

These results show that the price of privacy is small for datasets with few attributes, but may be large for high-dimensional datasets. For example, if we simply want to estimate the mean of each of the d attributes without a privacy guarantee, then $\Theta(\log d)$ samples are necessary and sufficient to get statistical accuracy. However, the best known (ε, δ)-differentially private algorithm requires $\Omega(\sqrt{d})$ samples—an exponential gap. In the special case of *pure* $(\varepsilon, 0)$-differential privacy, a lower bound of $\Omega(d)$ is known [36]. However, for the general case of *approximate* (ε, δ)-differential privacy the best known lower bound is $\Omega(\log d)$ [16]. More generally, there are no known lower bounds that separate the sample complexity of (ε, δ)-differential privacy from the sample complexity required for statistical accuracy alone.

In this work we close this gap almost completely, and show that there is indeed a "price of approximate differential privacy" for high-dimensional datasets.

THEOREM 1.1 (informal). *Any algorithm that takes as input a database $D \in (\{0,1\}^d)^n$, satisfies approximate differential privacy, and estimates the mean of each of the d attributes to within error $\pm 1/3$ requires $n \geq \tilde{\Omega}(\sqrt{d})$ samples.*

We establish this lower bound using a combinatorial object called a *fingerprinting code*, which was originally introduced by Boneh and Shaw [12] for the problem of watermarking copyrighted content. Specifically, we use Tardos' construction of optimal fingerprinting codes [49]. The use of "secure content distribution schemes" to prove lower bounds for differential privacy originates with the work of Dwork et al. [23], who used "traitor-tracing schemes," which are a cryptographic analogue of information-theoretic fingerprinting codes, to prove computational hardness results for differential privacy. Extending this connection, Ullman [51] used fingerprinting codes to construct

[1] For a specific family of queries $\mathcal{Q}$, the necessary and sufficient number of samples is proportional to the *Vapnik–Chervonenkis (VC) dimension* of $\mathcal{Q}$, which can be as large as $\log |\mathcal{Q}|$.

a novel traitor-tracing scheme and obtain a strong computational hardness result for differential privacy.[2] Here we show that a *direct* use of fingerprinting codes yields information-theoretic lower bounds on sample complexity.

Using the additional structure of Tardos' fingerprinting code, we are able to prove *statistical minimax lower bounds* for inferring the marginals of a product distribution from samples while guaranteeing differential privacy for the sample. Specifically, suppose the database $D \in (\{0,1\}^d)^n$ consists of n independent samples from a product distribution over $\{0,1\}^d$ such that the ith coordinate of each sample is set to 1 with probability p_i, for some unknown $p = (p_1, \ldots, p_d) \in [0,1]^d$. We show that if there exists a differentially private algorithm that takes such a database as input, satisfies approximate differential privacy, and outputs $\hat{p}$ such that $\|\hat{p} - p\|_\infty \leq 1/3$, then $n \geq \tilde{\Omega}(\sqrt{d})$. Statistical minimax bounds of this type for differentially private inference problems were first studied by Duchi, Jordan, and Wainwright [18], who proved minimax bounds for algorithms that satisfy the stronger constraint of *local pure* $(\varepsilon, 0)$*-differential privacy*.

Next, we consider the sample complexity of answering an arbitrary set $\mathcal{Q}$ of counting queries to within error $\pm\alpha$. As above, if we assume the database contains samples from a population, and require only that the answers to queries on the sampled database and the population are close, to within $\pm\alpha$, then $\Theta(\log|\mathcal{Q}|/\alpha^2)$ samples are necessary and sufficient for just statistical accuracy. When $|\mathcal{Q}|$ is large (relative to d and $1/\alpha$), the best sample complexity for differential privacy is again achieved by the private multiplicative weights algorithm, and is $O(\sqrt{d}\log|\mathcal{Q}|/\alpha^2)$. For pure differential privacy, a lower bound of $\Omega(d\log|\mathcal{Q}|/\alpha^2)$ is known [33]. On the other hand, the best known lower bound for approximate differential privacy is $\Omega(\max\{\log|\mathcal{Q}|/\alpha, 1/\alpha^2\})$, which follows from the techniques of [16]. To resolve this gap, we give a *composition theorem* that allows us to obtain a nearly optimal lower bound by combining Theorem 1.1 with (variants of) the existing sample-complexity lower bounds. The result shows that the private multiplicative weights algorithm achieves nearly optimal sample complexity as a function of $|\mathcal{Q}|, d$, and α.

THEOREM 1.2 (informal). *For every sufficiently small $\alpha > 0$, $d \geq 6\log(1/\alpha)$, and $s \geq d/\alpha^2$, there exists a family of queries $\mathcal{Q}$ of size s such that any algorithm that takes as input a database $D \in (\{0,1\}^d)^n$, satisfies approximate differential privacy, and outputs an approximate answer to each query in $\mathcal{Q}$ to within $\pm\alpha$ requires $n \geq \tilde{\Omega}(\sqrt{d}\log|\mathcal{Q}|/\alpha^2)$.*

We remark that the condition that $d \geq 6\log(1/\alpha)$ is both necessary (up to the constant factor) and fairly mild. Necessary because the *noisy histogram algorithm* (see, e.g., [53]) requires $n = O(2^{d/2}\sqrt{\log|\mathcal{Q}|}/\alpha)$ samples, which is better than the conclusion of the lower bound when $d < 2\log(1/\alpha)$. Mild because differential privacy cannot be satisfied for large query sets unless $\alpha \gtrsim 1/\sqrt{n}$, so the condition is no stronger than assuming $n \lesssim 2^{d/3}$, in which case the number of samples is exponential in the dimension. Similarly, the condition $s \geq d/\alpha^2$ is also necessary, since adding independent noise to each query requires only $n \gtrsim |\mathcal{Q}|^{1/2}/\alpha$ samples.

Finally, we consider the sample complexity of the natural and well studied class of *k-way marginal queries*, also known as *k-way conjunction queries* (see, e.g., [2, 40, 31, 50, 14, 25]). A k-way marginal query on a database $D \in (\{0,1\}^d)^n$ is specified by a set $S \subseteq [d]$, $|S| \leq k$, and a pattern $t \in \{0,1\}^{|S|}$ and asks "What fraction of records

[2] In fact, one way to prove Theorem 1.1 is by replacing the one-way functions in [51] with a random oracle, and thereby obtain an information-theoretically secure traitor-tracing scheme.

in D has each attribute j in S set to t_j?" The number of k-way marginal queries on $\{0,1\}^d$ is about $2^k\binom{d}{k}$. For the special case of $k = 1$, the queries simply ask for the mean of each attribute, which was discussed above. We prove that the lower bound of Theorem 1.2, which applies to worst-case queries, also holds for the special case of k-way marginal queries when α is not too small.

THEOREM 1.3 (informal). *Any algorithm that takes a database $D \in (\{0,1\}^d)^n$, satisfies approximate differential privacy, and outputs an approximate answer to each of the k-way marginal queries to within $\pm\alpha$ (for α smaller than some universal constant and larger than an inverse polynomial in d) requires $n \geq \tilde{\Omega}(k\sqrt{d}/\alpha^2)$.*

We remark that, since the number of k-way marginal queries is about $2^k\binom{d}{k}$, the sample-complexity lower bound in Theoem 1.3 essentially matches that of Theorem 1.2. The two theorems are incomparable, since Theorem 1.2 applies even when α is exponentially small in d, but only applies for a worst-case family of queries.

1.1. Our techniques. We now describe the main technical ingredients used to prove these results. For concreteness, we will describe the main ideas for the case of k-way marginal queries.

Fingerprinting codes. Fingerprinting codes, introduced by Boneh and Shaw [12], were originally designed to address the problem of watermarking copyrighted content. Roughly speaking, a (fully-collusion-resilient) fingerprinting code is a way of generating codewords for n users in such a way that any codeword can be uniquely traced back to a user. Each legitimate copy of a piece of digital content has such a codeword hidden in it, and thus any illegal copy can be traced back to the user who copied it. Moreover, even if an arbitrary subset of the users collude to produce a copy of the content, then under a certain *marking assumption*, the codeword appearing in the copy can still be traced back to one of the users who contributed to it. The standard marking assumption is that if every colluder has the same bit b in the jth bit of their codeword, then the jth bit of the "combined" codeword in the copy they produce must be also b. We refer the reader to the original paper of Boneh and Shaw [12] for the motivation behind the marking assumption and an explanation of how fingerprinting codes can be used to watermark digital content.

We show that the existence of short fingerprinting codes implies sample-complexity lower bounds for 1-way marginal queries. Recall that a 1-way marginal query q_j is specified by an integer $j \in [d]$ and asks simply "What fraction of records in D have a 1 in the jth bit?" Suppose a coalition of users takes their codewords and builds a database $D \in (\{0,1\}^d)^n$ where each record contains one of their codewords, and d is the length of the codewords. Consider the 1-way marginal query $q_j(D)$. If every user in S has a bit b in the jth bit of their codeword, then $q_j(D) = b$. Thus, if an algorithm answers 1-way marginal queries on D with nontrivial accuracy, its output can be used to obtain a combined codeword that satisfies the marking assumption. By the tracing property of fingerprinting codes, we can use the combined codeword to identify one of the users in the database. However, if we can identify one of the users from the answers, then the algorithm is not differentially private.

This argument can be formalized to show that if there is a fingerprinting code for n users with codewords of length d, then the sample complexity of answering 1-way marginals must be at least n. The nearly optimal construction of fingerprinting codes due to Tardos [49], gives fingerprinting codes with codewords of length $d = \tilde{O}(n^2)$, which implies a lower bound of $n \geq \tilde{\Omega}(\sqrt{d})$ on the sample complexity required to answer 1-way marginals queries.

Composition of sample-complexity lower bounds. Suppose we want to prove a lower bound of $\tilde{\Omega}(k\sqrt{d})$ for answering k-way marginals up to accuracy $\pm.01$ (a special case of Theorem 1.3). Given our lower bound of $\tilde{\Omega}(\sqrt{d})$ for 1-way marginals, and the known lower bound of $\Omega(k)$ for answering k-way marginals implicit in [16, 43], a natural approach is to somehow compose the two lower bounds to obtain a nearly optimal lower bound of $\tilde{\Omega}(k\sqrt{d})$. Our composition technique uses the idea of the $\Omega(k)$ lower bound from [16, 43] to show that if we can answer k-way marginal queries on a large database D with n rows, then we can obtain the answers to the 1-way marginal queries on a "subdatabase" of roughly n/k rows. Our lower bound for 1-way marginals tell us that $n/k = \tilde{\Omega}(\sqrt{d})$, so we deduce $n = \tilde{\Omega}(k\sqrt{d})$.

Actually, this reduction only gives accurate answers to *most* of the 1-way marginals on the subdatabase, so we need an extension of our lower bound for 1-way marginals to differentially private algorithms that are allowed to answer a small fraction of the queries with arbitrarily large error. Proving a sample-complexity lower bound for this problem requires a "robust" fingerprinting code whose tracing algorithm can trace codewords that have errors introduced into a small fraction of the bits. We show how to construct such a robust fingerprinting code of length $d = \tilde{O}(n^2)$, and thus obtain the desired lower bound. Fingerprinting codes satisfying a weaker notion of robustness were introduced by Boneh, Kiayias, and Montgomery [10] and Boneh and Naor [11].[3]

Theorems 1.2 and 1.3 are proven by using this composition technique repeatedly to combine our lower bound for 1-way marginals with (variants of) several known lower bounds that capture the optimal dependence on $\log|\mathcal{Q}|$ and $1/\alpha^2$.

Are fingerprinting codes necessary to prove differential privacy lower bounds? The connection between fingerprinting codes and differential privacy lower bounds extends to arbitrary families $\mathcal{Q}$ of counting queries. We introduce the notion of a generalized fingerprinting code with respect to $\mathcal{Q}$, where each codeword corresponds to a data universe element $x \in \mathcal{X}$ and the bits of the codeword are given by $q(x)$ for each $q \in \mathcal{Q}$, but is the same as an ordinary fingerprinting code otherwise. The existence of a generalized fingerprinting code with respect to $\mathcal{Q}$, for n users, implies a sample-complexity lower bound of n for privately releasing answers to $\mathcal{Q}$. We also show a partial converse to the above result, which states that some sort of "fingerprinting-code-like object" is necessary to prove sample-complexity lower bounds for answering counting queries under differential privacy. This object has similar semantics to a generalized fingerprinting code, however, the marking assumption required for tracing is slightly stronger and the probability that tracing succeeds can be significantly smaller than what is required by the standard definition of fingerprinting codes. Our partial converse parallels the result of Dwork et al. [23] that shows computational hardness results for differential privacy imply a "traitor-tracing-like object." We leave it as an open question to pin down precisely the relationship between fingerprinting codes and information-theoretic lower bounds in differential privacy (and also between traitor-tracing schemes and computational hardness results for differential privacy).

1.2. Other related work.

1.2.1. Previous work. We have mostly focused on the sample complexity as a function of the number of queries, the number of attributes d, and the accuracy parameter α. There have been several works focused on the sample complexity as

[3] In the fingerprinting codes of [11, 10] the adversary is allowed to *erase* a large fraction of the coordinates of the combined codeword, and must reveal which coordinates are erased.

a function of the specific family $\mathcal{Q}$ of queries. For $(\varepsilon, 0)$-differential privacy, Hardt and Talwar [36] showed how to approximately characterize the sample complexity of a family $\mathcal{Q}$ when the accuracy parameter α is sufficiently small. Nikolov, Talwar, and Zhang [42] extended their results to gives an approximate characterization for (ε, δ)-differential privacy and for the full range of accuracy parameters. Specifically, [42] gives an (ε, δ)-differentially private algorithm that answers any family of queries $\mathcal{Q}$ on $\{0,1\}^d$ with error α using a number of samples that is optimal up to a factor of $\mathrm{poly}(d, \log|\mathcal{Q}|)$ that is independent of α. Thus, their algorithm has sample complexity that depends optimally on α. However, their characterization may be loose by a factor of $\mathrm{poly}(d, \log|\mathcal{Q}|)$. In fact, when α is a constant, the lower bound on the sample complexity given by their characterization is always $O(1)$, whereas their algorithm requires $\mathrm{poly}(d, \log|\mathcal{Q}|)$ samples to give nontrivially accurate answers. In contrast, our lower bounds are tight to within $\mathrm{poly}(\log d, \log\log|\mathcal{Q}|, \log(1/\alpha))$ factors, and thus give meaningful lower bounds even when α is constant, but apply only to certain families of queries.

There have been attempts to prove optimal sample-complexity lower bounds for k-way marginals. In particular, when k is a constant, Kasiviswanathan et al. [40] and De [15] prove a lower bound of $\min\{|\mathcal{Q}|^{1/2}/\alpha, 1/\alpha^2\}$ on the sample complexity. Note that when α is a constant, these lower bounds are $O(1)$.

There have also been attempts to explicitly and precisely determine the sample complexity of even simpler query families than k-way conjunctions, such as point functions and threshold functions [5, 6, 7, 13]. These works show that these families can have sample complexity lower than $\tilde{O}(\sqrt{d}\log|\mathcal{Q}|/\alpha^2)$.

In addition to the general computational hardness results referenced above, there are several results that show stronger hardness results for restricted types of efficient algorithms [52, 31, 24].

1.2.2. Subsequent work. Subsequent to our work, Steinke and Ullman [47] refined our use of fingerprinting codes to prove a lower bound of $\Omega(\sqrt{d\log(1/\delta)}/\varepsilon)$ on the number of samples required to release the mean of each of the d attributes under (ε, δ)-differential privacy when $\delta \ll 1/n$. This lower bound is optimal up to constant factors, and improves on Theorem 1.1 by a factor of roughly $\sqrt{\log(1/\delta)} \cdot \log d$. They also improve and simplify our analysis of robust fingerprinting codes.

Our fingerprinting code technique has also been used to prove lower bounds for other types of differentially private data analyses. Namely, Dwork et al. [29] prove lower bounds for differentially private principal component analysis and Bassily, Smith, and Thakurta [4] prove lower bounds for differentially private empirical risk minimization. In order to establish lower bounds for privately releasing threshold functions, Bun et al. [13] construct a fingerprinting-code-like object that yields a lower bound for the problem of releasing a value between the minimum and maximum of a dataset.

Dwork et al. [28] observe that the privacy attack implicit in our negative results is closely related to the influential attacks that were employed by Homer et al. [38] (and further studied in [46]) to violate privacy of public genetic datasets. Using this connection, they show how to make Homer et al.'s attack robust to very general models of noise and how to make the attack work without detailed knowledge of the population the dataset represents.

A pair of works [37, 48] show that fingerprinting codes and the related traitor-tracing schemes imply both information-theoretic lower bounds and computational hardness results for the "false discovery" problem in adaptive data analysis.

Specifically, they show lower bounds for answering an online sequence of adaptively chosen counting queries where the database is a sample from some unknown distribution and the answers must be accurate with respect to that distribution. These works [37, 48] effectively reverse a connection established in [19, 3], which used differentially private algorithms to obtain positive results for this problem.

Our technique for composing lower bounds in differential privacy has also found applications outside of privacy. Specifically, Liberty et al. [41] used this technique to prove nearly optimal lower bounds on the space required to "sketch" a database while approximately preserving answers to k-way marginal queries (called "frequent itemset queries" in their work).

2. Preliminaries.

2.1. Differential privacy.

We define a *database* $D \in \mathcal{X}^n$ to be an ordered tuple of n rows $(x_1, \ldots, x_n) \in \mathcal{X}$ chosen from a *data universe* $\mathcal{X}$. We say that two databases $D, D' \in \mathcal{X}^n$ are *adjacent* if they differ only by a single row, and we denote this by $D \sim D'$. In particular, we can replace the ith row of a database D with some fixed "junk" element of $\mathcal{X}$ to obtain another database $D_{-i} \sim D$. We emphasize that if D is a database of size n, then D_{-i} is also a database of size n.

DEFINITION 2.1 (differential privacy [21]). *Let $\mathcal{A} : \mathcal{X}^n \to \mathcal{R}$ be a randomized algorithm (where n is a varying parameter). $\mathcal{A}$ is* (ε, δ)-differentially private *if for every two adjacent databases $D \sim D'$ and every subset $S \subseteq \mathcal{R}$,*

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in S] + \delta.$$

LEMMA 2.2. *Let $\mathcal{A} : \mathcal{X}^n \to \mathcal{R}$ be a randomized algorithm such that for every $D \in \mathcal{X}^n$, every $i, j \in [n]$, and every subset $S \subseteq \mathcal{R}$,*

$$\Pr[\mathcal{A}(D_{-i}) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(D_{-j}) \in S] + \delta.$$

Let $\perp$ denote the fixed junk element of $\mathcal{X}$. Then $\mathcal{A}' : \mathcal{X}^{n-1} \to \mathcal{R}$ defined by $\mathcal{A}'(x_1, \ldots, x_{n-1}) = \mathcal{A}(x_1, \ldots, x_{n-1}, \perp)$ is $(2\varepsilon, (e^{\varepsilon} + 1)\delta)$-differentially private.

Proof. Let $D = (x_1, \ldots, x_{n-1})$ and $D' = (x_1, \ldots, x'_i, \ldots, x_{n-1})$ be adjacent databases. Then for any $S \subseteq \mathcal{R}$, we have

$$\begin{aligned}
\Pr[\mathcal{A}'(D) \in S] &= \Pr[\mathcal{A}(x_1, \ldots, x_{n-1}, \perp) \in S] \\
&\leq e^{\varepsilon} \Pr[\mathcal{A}(x_1, \ldots, x_{i-1}, \perp, x_{i+1}, \ldots, x_{n-1}, \perp) \in S] + \delta \\
&\leq e^{2\varepsilon} \Pr[\mathcal{A}(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_{n-1}, \perp) \in S] + (e^{\varepsilon} + 1)\delta \\
&= e^{2\varepsilon} \Pr[\mathcal{A}'(D') \in S] + (e^{\varepsilon} + 1)\delta.
\end{aligned}$$

∎

2.2. Counting queries and accuracy.

In this paper we study algorithms that answer counting queries. A counting query on $\mathcal{X}$ is defined by a predicate $q : \mathcal{X} \to \{0, 1\}$. Abusing notation, we define the evaluation of the query q on a database $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ to be its average value over the rows,

$$q(D) = \frac{1}{n} \sum_{i=1}^{n} q(x_i).$$

DEFINITION 2.3 (accuracy for counting queries). *Let $\mathcal{Q}$ be a set of counting queries on $\mathcal{X}$ and $\alpha, \beta \in [0, 1]$ be parameters. For a database $D \in \mathcal{X}^n$, a sequence of answers*

$a = (a_q)_{q\in\mathcal{Q}} \in \mathbb{R}^{|\mathcal{Q}|}$ is (α,β)-accurate for $\mathcal{Q}$ *if* $|q(D) - a_q| \le \alpha$ *for at least a* $1-\beta$ *fraction of queries* $q \in \mathcal{Q}$.

Let $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ *be a randomized algorithm.* $\mathcal{A}$ *is* (α,β)-accurate for $\mathcal{Q}$ *if for every* $D \in \mathcal{X}^n$,

$$\Pr\left[\mathcal{A}(D) \text{ is } (\alpha,\beta)\text{-accurate for } \mathcal{Q}\right] \ge 2/3.$$

When $\beta = 0$ *we may simply write that* a *or* $\mathcal{A}$ *is* α-accurate for $\mathcal{Q}$.

In the definition of accuracy, we have assumed that $\mathcal{A}$ outputs a sequence of $|\mathcal{Q}|$ real-valued answers, with a_q representing the answer to q. Since we are not concerned with the running time of the algorithm, this assumption is without loss of generality.[4]

An important example of a collection of counting queries is the set of *k-way marginals*. For all of our results it will be sufficient to consider only the set of *monotone k-way marginals*.

DEFINITION 2.4 (monotone k-way marginals). *A (monotone) k-way marginal q_S over $\{0,1\}^d$ is specified by a subset $S \subseteq [d]$ of size $|S| \le k$. It takes the value $q_S(x) = 1$ if and only if $x_i = 1$ for every index $i \in S$. The collection of all (monotone) k-way marginals is denoted by $\mathcal{M}_{k,d}$.*

2.3. Sample complexity. In this work we prove lower bounds on the sample complexity required to simultaneously achieve differential privacy and accuracy.

DEFINITION 2.5 (sample complexity). *Let $\mathcal{Q}$ be a set of counting queries on $\mathcal{X}$ and let $\alpha,\beta > 0$ be parameters, and let ε,δ be functions of n. We say that $(\mathcal{Q},\mathcal{X})$ has sample complexity n^* for (α,β)-accuracy and (ε,δ)-differential privacy if n^* is the least $n \in \mathbb{N}$ such that there exists an (ε,δ)-differentially private algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ that is (α,β)-accurate for $\mathcal{Q}$.*

We will focus on the case where $\varepsilon = O(1)$ and $\delta = o(1/n)$. This setting of the parameters is essentially the most permissive for which (ε,δ)-differential privacy is still a meaningful privacy definition. However, pinning down the exact dependence on ε and δ is still of interest. Regarding ε, this can be done via the following standard lemma, which allows us to take $\varepsilon = 1$ without loss of generality.

LEMMA 2.6. *For every set of counting queries $\mathcal{Q}$, universe $\mathcal{X}$, $\alpha,\beta \in [0,1], \varepsilon \le 1$, $(\mathcal{Q},\mathcal{X})$ has sample complexity n^* for (α,β)-accuracy and $(1,o(1/n))$-differential privacy if and only if it has sample complexity $\Theta(n^*/\varepsilon)$ for (α,β)-accuracy and $(\varepsilon,o(1/n))$-differential privacy.*

One direction ($O(n^*/\varepsilon)$ samples are sufficient) is the "secrecy-of-the-sample lemma," which appeared implicitly in [39]. The other direction ($\Omega(n^*/\varepsilon)$ samples are necessary) appears to be folklore.

The next lemma allows us to generically translate sample-complexity lower bounds for constant accuracy into lower bounds that depend on the error parameter α. For some sets of queries, such as 1-way marginals, the dependence we get on α is tight. However, as we will see in section 5, we can obtain lower bounds with an even stronger dependence on α for specific sets of queries.

[4] In certain settings, $\mathcal{A}$ is allowed to output a "summary" $z \in \mathcal{R}$ for some range $\mathcal{R}$. In this case, we would also require that there exists an "evaluator" $\mathcal{E} : \mathcal{R} \times \mathcal{Q} \to \mathbb{R}$ that takes a summary and a query and returns an answer $\mathcal{E}(z,q) = a_q$ that approximates $q(D)$. The extra generality is used to allow $\mathcal{A}$ to run in less time than the number of queries it is answering. However, since we do not bound the running time of $\mathcal{A}$ we can convert any such sanitizer to one that outputs a sequence of $|\mathcal{Q}|$ real-valued answers simply by running the evaluator for every $q \in \mathcal{Q}$.

LEMMA 2.7. *Let $\mathcal{Q}$ be a set of counting queries on $\mathcal{X}$ and let $\beta, \varepsilon, \delta > 0$. Suppose $(\mathcal{Q}, \mathcal{X})$ has sample complexity n^* for (α_0, β)-accuracy and (ε, δ)-differential privacy, where $\alpha_0 \in (0, 1)$ is a constant. Then $(\mathcal{Q}, \mathcal{X})$ has sample complexity $\Omega(n^*/\alpha)$ for (α, β, γ)-accuracy and (ε, δ)-differential privacy.*

Proof. Let $\mathcal{A} : \mathcal{X}^m \to \mathbb{R}^{|\mathcal{Q}|}$ be an (ε, δ)-differentially private and (α, β)-accurate mechanism for releasing answers to $\mathcal{Q}$. We will use $\mathcal{A}$ to construct a mechanism $\mathcal{A}' : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ achieving constant accuracy α_0 on databases of size $n = \lceil m\alpha/\alpha_0 \rceil$. To do so, fix a (publicly known) element $x_0 \in \mathcal{X}$. On input a database $D' \in \mathcal{X}^n$, the mechanism $\mathcal{A}'$ "pads" D' by appending $m - n$ copies of x_0, producing a database D. It then runs $\mathcal{A}$ on D, obtaining answers $(a_q)_{q \in \mathcal{Q}}$. Finally, it releases answers $(a'_q)_{q \in \mathcal{Q}}$, where $a'_q = \frac{1}{n}(ma_q - (m-n)q(x_0))$.

The mechanism $\mathcal{A}'$ inherits (ε, δ)-differential privacy from $\mathcal{A}$, since changing one row of D' changes one row of the padded database D'. Now we argue accuracy. Suppose a_q is an answer such that $|a_q - q(D)| \le \alpha$. Note that by construction, $q(D) = \frac{1}{m}(nq(D') + (m-n)q(x_0))$, and hence $q(D') = \frac{1}{n}(mq(D) - (m-n)q(x_0))$. Thus we have

$$\begin{aligned} |a'_q - q(D')| &= \frac{1}{n}|ma_q - (m-n)q(x_0) - (mq(D) - (m-n)q(x_0))| \\ &= \frac{m}{n}|a_q - q(D)| \\ &\le \frac{m}{n} \cdot \alpha. \end{aligned}$$

Taking $n = \lceil m\alpha/\alpha_0 \rceil$ makes this quantity at most α_0, completing the proof. ∎

For context, we can restate some prior results on differentially private counting query release in our sample-complexity terminology.

THEOREM 2.8 (combination of [16, 26, 8, 21, 9, 35, 32]). *For every set of counting queries $\mathcal{Q}$ on $\mathcal{X}$ and every $\alpha > 0$, $(\mathcal{Q}, \mathcal{X})$ has sample complexity at most*

$$\min\left\{\tilde{O}\left(\frac{\sqrt{|\mathcal{Q}|}}{\alpha}\right), \tilde{O}\left(\frac{\sqrt{|\mathcal{X}|\log|\mathcal{Q}|}}{\alpha}\right), \tilde{O}\left(\frac{\sqrt{\log|\mathcal{X}|}\log|\mathcal{Q}|}{\alpha^2}\right)\right\}$$

for $(\alpha, 0)$-accuracy and $(1, o(1/n))$-differential privacy.

We are mostly interested in a setting of parameters where α is not too small (e.g., constant) and $\log|\mathcal{X}| \ll |\mathcal{Q}| \le \mathrm{poly}(|\mathcal{X}|)$. In this regime the best known sample complexity will be achieved by the final expression, corresponding to the private multiplicative weights algorithm [35] using the analysis of [32]. In light of Lemma 2.6, it is without loss of generality that we have stated these upper bounds for $\varepsilon = 1$.

The next theorem shows that, when the data universe is not too small, the private multiplicative weights algorithm is nearly-optimal as a function of $|\mathcal{Q}|$ and $1/\alpha$ when each parameter is considered individually.

THEOREM 2.9 (combination of [16, 43]). *For every $s \in \mathbb{N}$, and $\alpha \in (0, 1/4)$, there exists a set of s counting queries $\mathcal{Q}$ on a data universe $\mathcal{X}$ of size $\max\{\log s, O(1/\alpha^2)\}$ such that $(\mathcal{Q}, \mathcal{X})$ has sample complexity at least*

$$\max\left\{\Omega\left(\frac{\log|\mathcal{Q}|}{\alpha}\right), \Omega\left(\frac{1}{\alpha^2}\right)\right\}$$

for $(\alpha, 0)$-accuracy and $(1, o(1/n))$-differential privacy.

2.4. Reidentifiable distributions. All of our eventual lower bounds will take the form of a "reidentification" attack, in which we possess data from a large number of individuals, and identify one such individual who was included in the database. In this attack, we choose a distribution on databases and give an adversary (1) a database D drawn from that distribution and (2) either $\mathcal{A}(D)$ or $\mathcal{A}(D_{-i})$ for some row i, where $\mathcal{A}$ is an alleged sanitizer. The adversary's goal is to identify a row of D that was given to the sanitizer. We say that the distribution is reidentifiable if there is an adversary who can identify such a row with sufficiently high confidence whenever $\mathcal{A}$ outputs accurate answers. If the adversary can do so, it means that there must be a pair of adjacent databases $D \sim D_{-i}$ such that the adversary can distinguish $\mathcal{A}(D)$ from $\mathcal{A}(D_{-i})$, which means $\mathcal{A}$ cannot be differentially private.

DEFINITION 2.10 (reidentifiable distribution). *For a data universe $\mathcal{X}$ and $n \in \mathbb{N}$, let $\mathcal{D}$ be a distribution on n-row databases $D \in \mathcal{X}^n$. Let $\mathcal{Q}$ be a family of counting queries on $\mathcal{X}$ and let $\gamma, \xi, \alpha, \beta \in [0,1]$ be parameters. The distribution $\mathcal{D}$ is* (γ, ξ)-reidentifiable from (α, β)-accurate answers to $\mathcal{Q}$ *if there exists a (possibly randomized) adversary $\mathcal{B} : \mathcal{X}^n \times \mathbb{R}^{|\mathcal{Q}|} \to [n] \cup \{\perp\}$ such that for every randomized algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$, the following both hold:*

1. $\Pr_{D \leftarrow_R \mathcal{D}} [(\mathcal{B}(D, \mathcal{A}(D)) = \perp) \wedge (\mathcal{A}(D)$ *is (α, β)-accurate for $\mathcal{Q}$*$)] \leq \gamma$.
2. *For every $i \in [n]$,* $\Pr_{D \leftarrow_R \mathcal{D}} [\mathcal{B}(D, \mathcal{A}(D_{-i})) = i] \leq \xi$.

Here the probability is taken over the choice of D and i as well as the coins of $\mathcal{A}$ and $\mathcal{B}$. We allow $\mathcal{D}$ and $\mathcal{B}$ to share a common state.

Note that, when row i is not in the dataset, then it would be an error for $\mathcal{B}$ to declare that row i is in the dataset, and condition 2 requires that the probability of this error occurring is at most ξ.

The common state between $\mathcal{D}$ and $\mathcal{B}$ should be thought of as auxiliary information about the realization of D that may help $\mathcal{B}$ identify a user i. Formally, we could model this shared state by having $\mathcal{D}$ output an additional string *aux* that is given to $\mathcal{B}$ but not to $\mathcal{A}$. However, we make the shared state implicit to reduce notational clutter. The need for this shared state will become apparent when we use fingerprinting codes to construct reidentifiable distributions; in the context of fingerprinting codes, the shared state represents auxiliary information about a codebook that helps the *Trace* algorithm accuse a guilty pirate.

If $\mathcal{A}$ is an (α, β)-accurate algorithm, then its output $\mathcal{A}(D)$ will be (α, β)-accurate with probability at least $2/3$. Therefore, if $\gamma < 2/3$, we can conclude that

$$\Pr [\mathcal{B}(D, \mathcal{A}(D)) \in [n]] \geq 1 - \gamma - 1/3 = \Omega(1).$$

In particular, there exists some $i^* \in [n]$ for which

$$\Pr [\mathcal{B}(D, \mathcal{A}(D)) = i^*] \geq \Omega(1/n).$$

However, if $\xi = o(1/n)$, then $\Pr [\mathcal{B}(D, \mathcal{A}(D_{-i^*})) = i^*] \leq \xi = o(1/n)$. Thus, for this choice of γ and ξ we will obtain a contradiction to (ε, δ)-differential privacy of the postprocessed algorithm $\mathcal{B}(D, \mathcal{A}(\cdot))$ for any $\varepsilon = O(1)$ and $\delta = o(1/n)$. Note that this conclusion holds even if $\mathcal{D}$ and $\mathcal{B}$ share a common state. We summarize this argument with the following lemma.

LEMMA 2.11. *Let $\mathcal{Q}$ be a family of counting queries on $\mathcal{X}$, $n \in \mathbb{N}$, and $\xi \in [0,1]$. Suppose there exists a distribution on n-row databases $D \in \mathcal{X}^n$ that is (γ, ξ)-reidentifiable from (α, β)-accurate answers to $\mathcal{Q}$. Then there is no (ε, δ)-differentially*

private algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ that is (α, β)-accurate for $\mathcal{Q}$ for any ε, δ such that $e^{-\varepsilon}(1 - \gamma - 1/3)/n - \delta \geq \xi$.

In particular, if there exists a distribution that is $(\gamma, o(1/n))$-reidentifiable from (α, β)-accurate answers to $\mathcal{Q}$ for $\gamma = 1/3$, then no algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ that is (α, β)-accurate for $\mathcal{Q}$ can satisfy $(O(1), o(1/n))$-differential privacy.

3. Lower bounds via fingerprinting codes. In this section we prove that there exists a simple family of d queries that requires $n \geq \tilde{\Omega}(\sqrt{d})$ samples for both accuracy and privacy. Specifically, we prove that for the family of 1-way marginals on d bits, sample complexity $\tilde{\Omega}(\sqrt{d})$ is required to produce differentially private answers that are accurate even just to within $\pm 1/3$. In contrast, without a privacy guarantee, $\Theta(\log d)$ samples from the population are necessary and sufficient to ensure that the answers to these queries on the database and the population are approximately the same. The best previous lower bound for (ε, δ)-differential privacy is also $O(\log d)$, which follows from the techniques of [16, 43].

In section 3.1 we give the relevant background on fingerprinting codes and in section 3.2 we prove our lower bounds for 1-way marginals.

3.1. Fingerprinting codes. Fingerprinting codes were introduced by Boneh and Shaw [12] to address the problem of watermarking digital content. A fingerprinting code is a pair of randomized algorithms $(Gen, Trace)$. The code generator Gen outputs a *codebook* $C \in \{0,1\}^{n \times d}$. Each row c_i of C is the *codeword* of user i. For a subset of users $S \subseteq [n]$, we use $C_S \in \{0,1\}^{|S| \times d}$ to denote the set of codewords of users in S. The parameter d is called the *length* of the fingerprinting code.

The security property of fingerprinting codes asserts that any codeword can be "traced" to a user $i \in [n]$. Moreover, we require that the fingerprinting code is "fully-collusion-resilient"—even if any "coalition" of users $S \subseteq [n]$ gets together and "combines" their codewords in any way that respects certain constraints known as a *marking assumption*, then the combined codeword c' can be traced to a user $i \in S$. That is, there is a tracing algorithm *Trace* that takes as inputs the codebook and combined codeword c' and outputs either a user $i \in [n]$ or $\perp$, and we require that if c' satisfies the constraints, then $Trace(C, c') \in S$ with high probability. Moreover, *Trace* should accuse an innocent user, i.e., $Trace(C, c') \in [n] \setminus S$, with very low probability. Analogous to the definition of reidentifiable distributions (Definition 2.10), we allow *Gen* and *Trace* to share a common state.[5] When designing fingerprinting codes, one tries to make the marking assumption on the combined codeword as weak as possible.

The basic marking assumption is that each bit of the combined word c' must match the corresponding bit for some user in S. Formally, for a codebook $C \in \{0,1\}^{n \times d}$, and a coalition $S \subseteq [n]$, we define the set of *feasible codewords for C_S* to be

$$F(C_S) = \left\{ c' \in \{0,1\}^d \mid \forall j \in [d], \exists i \in S, c'_j = c_{ij} \right\}.$$

Observe that the combined codeword is only constrained on coordinates j where all users in S agree on the jth bit.

We are now ready to formally define a fingerprinting code.

DEFINITION 3.1 (fingerprinting codes). *For any $n, d \in \mathbb{N}$, $\xi \in (0, 1]$, a pair of algorithms $(Gen, Trace)$ is an (n, d)-*fingerprinting code with security ξ *if Gen outputs a codebook $C \in \{0,1\}^{n \times d}$ and for every (possibly randomized) adversary $\mathcal{A}_{FP}$, and every coalition $S \subseteq [n]$, if we set $c' \leftarrow_R \mathcal{A}_{FP}(C_S)$, then*

[5] As in Definition 2.10, we could model this by having *Gen* output an additional string *aux* that is given to *Trace*. However, we make the shared state implicit to reduce notational clutter.

1. $\Pr[c' \in F(C_S) \wedge \mathit{Trace}(C, c') = \perp] \leq \xi$,
2. $\Pr[\mathit{Trace}(C, c') \in [n] \setminus S] \leq \xi$,

where the probability is taken over the coins of Gen, Trace, and $\mathcal{A}_{FP}$*. The algorithms Gen and Trace may share a common state.*

We remark that our proof of Theorem 3.5, showing how to construct reidentifiable distributions from a fingerprinting code, will only require collusion resilience against coalitions S of size $|S| \geq n - 1$. Our choice to state Definition 3.1 using resilience against arbitrary coalitions is more consistent with the literature on fingerprinting codes.

Tardos [49] constructed a family of fingerprinting codes with a nearly optimal number of users n for a given length d.

THEOREM 3.2 (see [49]). *For every* $d \in \mathbb{N}$*, and* $\xi \in [0, 1]$*, there exists an* (n, d)*-fingerprinting code with security* ξ *for*

$$n = n(d, \xi) = \tilde{\Omega}(\sqrt{d/\log(1/\xi)}).$$

As we will see in the next subsection, fingerprinting codes satisfying Definition 3.1 will imply lower bounds on the sample complexity for releasing 1-way marginals with $(\alpha, 0)$-accuracy (accuracy for every query). In order to prove sample-complexity lower bounds for (α, β)-accuracy with $\beta > 0$, we will need fingerprinting codes satisfying a stronger security property. Specifically, we will expand the feasible set $F(C_S)$ to include all codewords that satisfy most feasibility constraints, and require that even codewords in this expanded set can be traced. Formally, for any $\beta \in [0, 1]$, we define

$$F_\beta(C_S) = \left\{ c' \in \{0,1\}^d \mid \Pr_{j \leftarrow_R [d]} \left[\exists i \in S, c'_j = c_{ij} \right] \geq 1 - \beta \right\}.$$

Observe that $F_0(C_S) = F(C_S)$.

DEFINITION 3.3 (error-robust fingerprinting codes). *For any* $n, d \in \mathbb{N}$*,* $\xi, \beta \in [0, 1]$*, a pair of algorithms* $(\mathit{Gen}, \mathit{Trace})$ *is an* (n, d)*-fingerprinting code with security* ξ *robust to a* β *fraction of errors if Gen outputs a codebook* $C \in \{0,1\}^{n \times d}$ *and for every (possibly randomized) adversary* $\mathcal{A}_{FP}$*, and every coalition* $S \subseteq [n]$*, if we set* $c' \leftarrow_R \mathcal{A}_{FP}(C_S)$*, then*

1. $\Pr[c' \in F_\beta(C_S) \wedge \mathit{Trace}(C, c') = \perp] \leq \xi$,
2. $\Pr[\mathit{Trace}(C, c') \in [n] \setminus S] \leq \xi$,

where the probability is taken over the coins of Gen, Trace, and $\mathcal{A}_{FP}$*. The algorithms Gen and Trace may share a common state.*

In section 6 we show how to construct error-robust fingerprinting codes with a nearly optimal number of users that are tolerant to a constant fraction of errors.

THEOREM 3.4. *For every* $d \in \mathbb{N}$*, and* $\xi \in (0, 1]$*, there exists an* (n, d)*-fingerprinting code with security* ξ *robust to a* $1/75$ *fraction of errors for*

$$n = n(d, \xi) = \tilde{\Omega}(\sqrt{d/\log(1/\xi)}).$$

Boneh and Naor [11] introduced a different notion of fingerprinting codes robust to adversarial "erasures." In their definition, the adversary is allowed to output a string in $\{0, 1, ?\}^d$, and in order to trace they require that the fraction of ? symbols is bounded away from 1 and that any non-? symbols respect the basic feasibility constraint. For this definition, constructions with nearly optimal length $d = \tilde{O}(n^2)$, robust to a $1 - o(1)$ fraction of erasures are known [10]. In contrast, our codes are robust to adversarial

"errors." Robustness to a β fraction of errors can be seen to imply robustness to nearly a 2β fraction of erasures but the converse is false. Thus for corresponding levels of robustness our definition is strictly more stringent. Unfortunately we don't currently know how to design a code tolerant to a $1/2 - o(1)$ fraction of errors, so our Theorem 3.4 does not subsume prior results on robust fingerprinting codes.

3.2. Lower bounds for 1-way marginals. We are now ready to state and prove the main result of this section, namely, that there is a distribution on databases $D \in (\{0,1\}^d)^n$ for $n = \tilde{\Omega}(\sqrt{d})$, that is reidentifiable from accurate answers to 1-way marginals.

THEOREM 3.5. *For every $n, d \in \mathbb{N}$, and $\xi \in [0,1]$ if there exists an (n,d)-fingerprinting code with security ξ, robust to a β fraction of errors, then there exists a distribution on n-row databases $D \in (\{0,1\}^d)^n$ that is (ξ,ξ)-reidentifiable from $(1/3,\beta)$-accurate answers to $\mathcal{M}_{1,d}$.*

In particular, if $\xi = o(1/n)$, then there is no algorithm $\mathcal{A} : (\{0,1\}^d)^n \to \mathbb{R}^{|\mathcal{M}_{1,d}|}$ that is $(O(1), o(1/n))$-differentially private and $(1/3,\beta)$-accurate for $\mathcal{M}_{1,d}$.

By combining Theorem 3.5 with Theorem 3.2 we obtain a sample-complexity lower bound for 1-way marginals, and thereby establish Theorem 1.1 in the introduction.

COROLLARY 3.6. *For every $d \in \mathbb{N}$, the family of 1-way marginals on $\{0,1\}^d$ has sample complexity at least $\tilde{\Omega}(\sqrt{d})$ for $(1/3, 1/75)$-accuracy and $(O(1), o(1/n))$-differential privacy.*

Proof of Theorem 3.5. Let $(Gen, Trace)$ be the promised fingerprinting code. We define the reidentifiable distribution $\mathcal{D}$ to simply be the output distribution of the code generator, Gen. And we define the privacy adversary $\mathcal{B}$ to take the answers $a = \mathcal{A}(D) \in [0,1]^{|\mathcal{M}_{1,d}|}$, obtain $\overline{a} \in \{0,1\}^{|\mathcal{M}_{1,d}|}$ by rounding each entry of a to $\{0,1\}$, run the tracing algorithm $Trace$ on the rounded answers $\overline{a}$, and return its output. The shared state of $\mathcal{D}$ and $\mathcal{B}$ will be the shared state of Gen and $Trace$.

Now we will verify that $\mathcal{D}$ is (ξ,ξ)-reidentifiable. First, suppose that $\mathcal{A}(D)$ outputs answers $a = (a_{q_j})_{j \in [d]}$ that are $(1/3,\beta)$-accurate for 1-way marginals. That is, there is a set $G \subseteq [d]$ such that $|G| \geq (1-\beta)d$ and for every $j \in G$, the answer a_{q_j} estimates the fraction of rows having a 1 in column j to within $1/3$. Let $\overline{a}_{q_j}$ be a_{q_j} rounded to the nearest value in $\{0,1\}$. Let j be a column in G. If column j has all 1's, then $a_{q_j} \geq 2/3$, and $\overline{a}_{q_j} = 1$. Similarly, if column j has all 0's, then $a_{q_j} \leq 1/3$, and $\overline{a}_{q_j} = 0$. Therefore, we have

$$a \text{ is } (1/3,\beta)\text{-accurate} \implies \overline{a} \in F_\beta(D). \tag{1}$$

By security of the fingerprinting code (Definition 3.3), we have

$$\Pr\left[\overline{a} \in F_\beta(D) \wedge Trace(D,\overline{a}) = \perp\right] \leq \xi. \tag{2}$$

Combining (1) and (2) implies that

$$\Pr\left[\mathcal{A}(D) \text{ is } (1/3,\beta)\text{-accurate} \wedge Trace(D,\overline{a}) = \perp\right] \leq \xi.$$

But the event $Trace(D,\overline{a}) = \perp$ is exactly the same as $\mathcal{B}(D,\mathcal{A}(D)) = \perp$, and thus we have established the first condition necessary for $\mathcal{D}$ to be (ξ,ξ)-reidentifiable.

The second condition for reidentifiability follows directly from the soundness of the fingerprinting code, which asserts that for every adversary $\mathcal{A}_{FP}$, in particular for $\mathcal{A}$, it holds that

$$\Pr\left[Trace(D, \mathcal{A}_{FP}(D_{-i})) = i\right] \leq \xi.$$

This completes the proof. ∎

Remark 3.7. Corollary 3.6 implies a lower bound of $\tilde{\Omega}(\sqrt{d})$ for any family $\mathcal{Q}$ on a data universe $\mathcal{X}$ in which we can "embed" the 1-way marginals on $\{0,1\}^d$ in the sense that there exists $q_1, \ldots, q_d \in \mathcal{Q}$ such that for every string $x \in \{0,1\}^d$ there is an $x' \in \{0,1\}^d$ such that $(q_1(x'), \ldots, q_d(x')) = x$. (The maximum such d is actually the VC dimension of $\mathcal{X}$ when we view each element $x \in \mathcal{X}$ as defining a mapping $q \mapsto q(x)$. See Definition 5.1.)

Our proof technique does not directly yield a lower bound with any meaningful dependence on the accuracy α. Since the privacy adversary $\mathcal{B}$ simply runs the tracing algorithm on the rounded answers it is given, it is not able to leverage subconstant accuracy to gain an advantage in reidentification. However, Lemma 2.7 lets us generically translate our lower bound for constant accuracy into a lower bound depending linearly on $1/\alpha$. For 1-way marginals, we get an essentially tight sample-complexity lower bound of $\tilde{\Omega}(\sqrt{d}/\alpha)$ for (α, β)-accuracy.

COROLLARY 3.8. *For every $d \in \mathbb{N}$, the family of 1-way marginals on $\{0,1\}^d$ has sample complexity at least $\tilde{\Omega}(\sqrt{d}/\alpha)$ for $(\alpha, 1/75)$-accuracy and $(O(1), o(1/n))$-differential privacy.*

3.2.1. Minimax lower bounds for statistical inference. Using the additional structure of Tardos' fingerprinting code, and our robust fingerprinting codes, we can prove minimax lower bounds for an "inference version" of the problem computing the 1-way marginals of a product distribution.

For any $d \in \mathbb{N}$, and any *marginals* $p = (p_1, \ldots, p_d) \in [0,1]^d$, let $\mathcal{D}_p$ denote the product distribution over strings $x \in \{0,1\}^d$, where each coordinate x_i is an independent draw from a Bernoulli random variable with mean p_i (i.e., x_i is set to 1 with probability p_i and set to 0 otherwise). We use $\mathcal{D}_p^{\otimes n}$ to denote n independent draws from $\mathcal{D}_p$. We say that a vector $q \in [0,1]^d$ is (α, β)-accurate for p if

$$\Pr_{i \leftarrow_R [d]} [|q_i - p_i| \leq \alpha] \geq 1 - \beta.$$

We can now formally define the problem of inferring the marginals p as follows.

DEFINITION 3.9. *Let $\alpha, \beta \in [0,1]$ be parameters. An algorithm $\mathcal{A} : (\{0,1\}^d)^n \to \mathbb{R}^d$ (α, β)-accurately infers the marginals of a product distribution if for every vector of marginals $p \in [0,1]^d$,*

$$\Pr_{D \leftarrow_R \mathcal{D}_p^{\otimes n}, \mathcal{A}\text{'s coins}} [\mathcal{A}(D) \text{ is } (\alpha, \beta)\text{-accurate for } p] \geq 2/3.$$

Our lower bound can thus be stated as follows.

THEOREM 3.10. *Suppose there is a function $n = n(d)$ such that for every $d \in \mathbb{N}$, there exists an algorithm $\mathcal{A} : (\{0,1\}^d)^n \to \mathbb{R}^d$ that satisfies $(O(1), o(1/n))$-differential privacy and $(1/3, 1/75)$-accurately infers the marginals of a product distribution. Then $n = \tilde{\Omega}(\sqrt{d})$.*

Proof Sketch. The proof has the same general structure that we used to prove Theorem 3.5. Here, we describe additional observations about the structure of the fingerprinting codes used in that proof (see section 6 for a description of Tardos' fingerprinting code) that allow it to carry over to the inference version of computing 1-way marginals.

First, in Tardos' (nonrobust) fingerprinting code, the codebook D is chosen by first sampling marginals $p \in [0,1]^d$ from an appropriate distribution and then sampling D from $\mathcal{D}_p^{\otimes n}$. The robust fingerprinting codes we construct in section 6 also have

this property.[6] Thus the instances used to prove Theorem 3.5 indeed consist of independent samples from a product distribution, which is what the inference problem assumes.

Next, recall that the proof of Theorem 3.5 shows that any string that is (α, β)-accurate for the 1-way marginals of D can be traced successfully. It is moreover the case that any string that is (α, β)-accurate for the marginals p can also be traced successfully. This is because the rows of D are sampled independently from $\mathcal{D}_p$, so accuracy for the 1-way marginals of D and accuracy for p coincide with high probability, at least when $n = \omega(\log d)$.

CLAIM 3.11. *Let $p \in [0,1]^d$ and let $D \leftarrow_R \mathcal{D}_p^{\otimes n}$. Let $a \in [0,1]^d$ denote the exact 1-way marginals of D. Then for every $\alpha, \eta > 0$, and $n = \Omega(\log(d/\eta)/\alpha^2)$, we have $\|a - p\|_\infty \leq \alpha$ with probability at least $1 - \eta$ over the choice of D.*

We remark that Steinke and Ullman [47] showed that accuracy with respect to the marginals p actually suffices to trace regardless of the value of n.

These two observations suffice to show that, when n is too small, a differentially private algorithm cannot be accurate for p with high probability over the choices of both p and D. Thus, for every differentially private algorithm, there exists some p such that the algorithm is not accurate with high probability over the choice of D, which means that the algorithm does not accurately infer the marginals of an arbitrary product distribution. ∎

3.3. Fingerprinting codes for general query families. In this section, we generalize the connection between fingerprinting codes and sample-complexity lower bounds for arbitrary sets of queries. We show that a generalized fingerprinting code with respect to any family of counting queries $\mathcal{Q}$ yields a sample-complexity lower bound for $\mathcal{Q}$, which is analogous to our lower bound for 1-way marginals (Theorem 3.5). We then argue that some type of fingerprinting code is necessary to prove any sample-complexity lower bound by exhibiting a tight connection between such lower bounds and a weak variant of our generalized fingerprinting codes.

We begin by defining our generalization of fingerprinting codes. Fix a finite data universe $\mathcal{X}$ and a set of counting queries $\mathcal{Q}$ over $\mathcal{X}$. A generalized fingerprinting code with respect to the family $\mathcal{Q}$ consists of a pair of randomized algorithms (*Gen*, *Trace*). The code generation algorithm *Gen* produces a codebook $C \in \mathcal{X}^n$. Each row c_i of C is the codeword corresponding to user i. A coalition $S \subseteq [n]$ of pirates receives the subset $C_S = \{c_i : i \in S\}$ of codewords, and produces an *answer vector* $a \in [0,1]^{|\mathcal{Q}|}$. We replace the traditional marking condition on the pirates with the generalized constraint that they output a *feasible answer vector*. A natural way to define feasibility for answer vectors is to require a condition similar to (α, β)-accuracy, i.e., an answer vector a is feasible if $|a_q - q(C_S)| \leq \alpha$ for all but a β fraction of queries $q \in \mathcal{Q}$. We thus define a generalized set of feasible answer vectors by

$$F_{\alpha,\beta}(C_S) = \left\{ a \in [0,1]^{|\mathcal{Q}|} \mid \Pr_{q \leftarrow_R \mathcal{Q}} [|a_q - q(C_S)| \leq \alpha] \geq 1 - \beta \right\}.$$

[6] To generate a codebook D' for our robust fingerprinting code, we sample a codebook D from Tardos' fingerprinting code and then insert additional columns of all 1's or all 0's to D in random locations. Equivalently, we can obtain a codebook D' by appending 1's and 0's in random locations of p to obtain a vector p' and then sampling D' from $\mathcal{D}_{p'}^{\otimes n}$.

When $\alpha = 1 - 1/n$, the generalized set of feasible answer vectors captures the traditional marking assumption by rounding each entry of a feasible answer vector to 0 or 1.[7]

DEFINITION 3.12. *A pair of algorithms* $(Gen, Trace)$ *is an* $(n, \mathcal{Q})$*-fingerprinting code for* (α, β)*-accuracy with security* (γ, ξ) *if Gen outputs a codebook* $C \in \mathcal{X}^n$ *and for every (possibly randomized) adversary* $\mathcal{A}_{FP}$*, and every coalition* $S \subseteq [n]$ *with* $|S| \geq n - 1$*, if we set* $a \leftarrow_R \mathcal{A}_{FP}(C_S)$*, then*

1. $\Pr[a \in F_{\alpha,\beta}(C_S) \wedge Trace(C, a) = \perp] \leq \gamma$,
2. $\Pr[Trace(C, a) \in [n] \setminus S] \leq \xi$,

where the probability is taken over the coins of Gen, Trace, and $\mathcal{A}_{FP}$*. The algorithms Gen and Trace may share a common state.*

The security properties of Definition 3.12 differ from those of an ordinary fingerprinting code in two ways so as to enable a clean statement of a composition theorem for generalized fingerprinting codes (Theorem 4.6). First, we use two separate security parameters γ, ξ for the different types of tracing errors, as in the definition of reidentifiable distributions. Second, security only needs to hold for coalitions of size $n - 1$ or n. However, this condition implies security for coalitions of arbitrary size with an increased false accusation probability of $n\xi$.

As in Theorem 3.5, the existence of a generalized $(n, \mathcal{Q})$-fingerprinting code implies a sample-complexity lower bound of n for privately releasing answers to $\mathcal{Q}$, with essentially the same proof.

THEOREM 3.13. *For every* $n \in \mathbb{N}$ *and* $\gamma, \xi \in [0, 1)$*, if there exists an* $(n, \mathcal{Q})$*-fingerprinting code for* (α, β)*-accuracy with security* (γ, ξ)*, then there exists a distribution on* n*-row databases* $D \in \mathcal{X}^n$ *that is* (γ, ξ)*-reidentifiable from* (α, β)*-accurate answers to* $\mathcal{Q}$*.*

In particular, if $\gamma \leq 1/3$ *and* $\xi = o(1/n)$*, then there is no algorithm* $\mathcal{A} : \mathcal{X}^n \to [0, 1]^{|\mathcal{Q}|}$ *that is* $(O(1), o(1/n))$*-differentially private and* (α, β)*-accurate for* $\mathcal{Q}$*.*

We now turn to investigate whether a converse to Theorem 3.13 holds. We show that a sample-complexity lower bound for a family of queries $\mathcal{Q}$ is essentially equivalent to the existence of a weak type of fingerprinting code, where the tracing procedure depends on the family $\mathcal{Q}$ and the tracing error probabilities satisfy certain affine constraints. It remains an interesting open question to determine the precise relationship between privacy lower bounds and our notion of generalized fingerprinting codes.

DEFINITION 3.14. *A pair of algorithms* $(Gen, Trace)$ *is an* $(n, \mathcal{Q})$*-weak fingerprinting code for* (α, β)*-accuracy with security* (ε, δ) *if Gen outputs a codebook* $C \in \mathcal{X}^n$ *and for every (possibly randomized) adversary* $\mathcal{A}_{FP}$ *that outputs a feasible answer vector with probability* $2/3$*, and every coalition* $S \subseteq [n]$ *with* $|S| \geq n - 1$*, if we set* $a \leftarrow_R \mathcal{A}_{FP}(C_S)$*, then*

$$\Pr[Trace(C, a) \neq \perp] > e^{\varepsilon} n \cdot \Pr[Trace(C, a) \in [n] \setminus S] + \delta,$$

where the probabilities are taken over the coins of Gen, Trace, and $\mathcal{A}_{FP}$*. The algorithms Gen and Trace may share a common state.*

[7] An equivalent way to view a codebook is as a set of n codewords $C \in (\{0, 1\}^{|\mathcal{Q}|})^n$, where each user's codeword is $c_i = (q(x))_{q \in \mathcal{Q}}$ for some $x \in \mathcal{X}$. Notice that the case where $\mathcal{Q}$ is the class of 1-way marginals places no constraints on the structure of a codeword, i.e., a codeword can be any binary string. With this viewpoint, the goal of the pirates is to output an answer vector $a \in [0, 1]^{|\mathcal{Q}|}$ with $|a_q - \frac{1}{|S|} \sum_{i \in S} (c_i)_q| \leq \alpha$ for all but a β fraction of the queries $q \in \mathcal{Q}$.

That is, we require the false accusation probability $\Pr[\mathit{Trace}(C,a) \in [n] \setminus S]$ to be much smaller than the total probability of accusing any user. Note that a tracing algorithm that accuses a random user with probability p will falsely accuse a user with probability p/n when $|S| = n-1$; however, this does not satisfy Definition 3.14 because we require the gap between the two probabilities to be at least a factor of $e^{\varepsilon}n$.

Observe that taking $\xi < (1-\delta)/2e^{\varepsilon}n$ in Definition 3.12 yields an $(n,\mathcal{Q})$-weak fingerprinting code with security (ε,δ). However, Definition 3.14 is weaker than Definition 3.12 in a few important ways. First, security only holds against pirates with a failure probability of at most $1/3$. Second, while Definition 3.12 requires completeness error $\Pr[\mathit{Trace}(C,a) = \perp] < \xi$, a weak fingerprinting code allows $\Pr[\mathit{Trace}(C,a) = \perp] = 1 - o(1)$ as long as $\Pr[\mathit{Trace}(C,a) \in [n] \setminus S]$ is sufficiently small.

The following theorem shows that the existence of an $(n,\mathcal{Q})$-weak fingerprinting code is essentially equivalent to a sample-complexity lower bound of n against $\mathcal{Q}$.

THEOREM 3.15. *For every $n \in \mathbb{N}$, if there exists an $(n,\mathcal{Q})$-weak fingerprinting code for (α,β)-accuracy with security (ε,δ), then there exists a distribution on n-row databases $D \in \mathcal{X}^n$ such that no $(\varepsilon/2, \delta/(2e^{\varepsilon/2}n))$-differentially private algorithm $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ outputs (α,β)-accurate answers to $\mathcal{Q}$.*

Conversely, let $\varepsilon \le 3$ and suppose there is no (ε,δ)-differentially private $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ that gives (α,β)-accurate answers to $\mathcal{Q}$ with probability at least $1/2$. Then there exists an $(m = \lceil n/\varepsilon \rceil, \mathcal{Q})$-weak fingerprinting code for $(\alpha - \alpha', \beta)$-accuracy with security $(\varepsilon/6, \delta/(e^{\varepsilon/3} + e^{5\varepsilon/6}))$ for $\alpha' = \tilde{O}(\sqrt{\varepsilon\, VC(\mathcal{Q})/n})$.

Proof. The forward direction follows the ideas of Lemma 2.11 and Theorem 3.5. Suppose for the sake of contradiction that there exists an (ε',δ')-differentially private $\mathcal{A} : \mathcal{X}^n \to \mathbb{R}^{|\mathcal{Q}|}$ that is (α,β)-accurate for $\mathcal{Q}$. Define a pirate strategy $\mathcal{A}_{FP}$ for coalitions of size $|S| \ge n-1$ by running $\mathcal{A}$ on its input C_S (possibly padded to size n by a junk row). Since $\mathcal{A}$ is (α,β)-accurate, with probability at least $2/3$ it produces an answer vector a such that $|a - q(C_S)| \le \alpha$ for all but a β fraction of $q \in \mathcal{Q}$. Hence, $\mathcal{A}_{FP}$ outputs a feasible answer vector with probability $2/3$. Define

$$p = \Pr_{\substack{C \leftarrow_R Gen \\ \mathrm{coins}(\mathcal{A}_{FP}),\mathrm{coins}(\mathit{Trace})}} [\mathit{Trace}(C, \mathcal{A}_{FP}(C)) \neq \perp].$$

Then there exists an i^* such that $\Pr[\mathit{Trace}(C, \mathcal{A}_{FP}(C)) = i^*] \ge p/n$. By differential privacy,

$$\Pr[\mathit{Trace}(C, \mathcal{A}_{FP}(C_{-i^*})) = i^*] \ge e^{-\varepsilon'} \cdot \left(\frac{p}{n} - \delta'\right).$$

On the other hand, by the security of the weak fingerprinting code and differential privacy,

$$\begin{aligned} e^{\varepsilon} \cdot n \cdot \Pr[\mathit{Trace}(C, \mathcal{A}_{FP}(C_{-i^*}) = i^*] &< \Pr[\mathit{Trace}(C, \mathcal{A}_{FP}(C_{-i^*}) \neq \perp] - \delta \\ &\le e^{\varepsilon'} p + \delta' - \delta. \end{aligned}$$

This yields a contradiction whenever $\varepsilon' \le \varepsilon/2$ and $\delta' \le \delta/(1 + e^{\varepsilon/2}n)$.

We now show the converse direction, i.e., that the high sample-complexity of $(\mathcal{Q},\mathcal{X})$ implies the existence of a weak fingerprinting code. We begin with a technical lemma which shows that the high sample complexity of $\mathcal{Q}$ also rules out mechanisms that satisfy only a one-sided constraint on the probability of any event under the replacement of one row.

LEMMA 3.16. *Let $\varepsilon \leq 1/2$. Let $\mathcal{A}$ be an (α, β)-accurate algorithm for $\mathcal{Q}$ on databases $D \in \mathcal{X}^m$. Suppose we have that for all databases $D \in \mathcal{X}^m$, all $i \in [m]$, and all measurable $T \subseteq Range(\mathcal{A})$ that*

$$\Pr_{\substack{j \leftarrow_R [m] \\ coins(\mathcal{A})}} [\mathcal{A}(D_{-j}) \in T] \leq e^{\varepsilon} \Pr_{coins(\mathcal{A})} [\mathcal{A}(D_{-i}) \in T] + \delta.$$

Let $d = VC(\mathcal{Q})$ be the VC dimension of $\mathcal{Q}$ and let

$$\alpha' = \left(\frac{8}{m} \cdot \left(\ln 24 + d \cdot \ln \left(\frac{2em}{d} \right) \right) \right)^{1/2} + \frac{\varepsilon}{m}.$$

Then there exists a $(6\varepsilon, (e^{2\varepsilon} + e^{5\varepsilon})\delta)$-differentially private algorithm $\mathcal{B}$ on databases of size $n = \lceil m/\varepsilon \rceil$ that gives $(\alpha + \alpha', \beta)$-accurate answers to $\mathcal{Q}$ on any database $D' \in \mathcal{X}^n$ with probability at least $1/2$.

Proof. On input to a database $D' \in \mathcal{X}^n$, consider the algorithm $\mathcal{B}'$ that samples a random subset D consisting of m rows from D' (without replacement) and returns $\mathcal{A}(D)$. Then by our hypothesis on $\mathcal{A}$, for every $i \in [n]$ and every measurable $T \subseteq \text{Range}(\mathcal{B}) = \text{Range}(\mathcal{A})$ we have

$$\Pr_{\substack{j \leftarrow_{\text{R}} [n] \\ \text{coins}(\mathcal{B}')}} \left[\mathcal{B}'(D'_{-j}) \in T \right] \leq e^{\varepsilon} \Pr_{\text{coins}(\mathcal{B}')} \left[\mathcal{B}'(D'_{-i}) \in T \right] + \delta.$$

On the other hand, a "secrecy-of-the-sample" argument [39] enables us to obtain the reverse inequality. For a row $k \in [n]$, consider the following two experiments:

Experiment 1: Sample a random subset D of m rows from D'_{-k}.

Experiment 2: Sample $j \leftarrow_{\text{R}} [n]$, and then sample a random subset D of m rows from D'_{-j}.

Any database D sampleable under Experiment 1 appears with probability $1/\binom{n}{m}$, but appears with probability at least

$$\frac{n-m}{n} \cdot \frac{1}{\binom{n}{m}} \geq (1-\varepsilon) \cdot \frac{1}{\binom{n}{m}}$$

under Experiment 2. Therefore,

$$\Pr_{\substack{j \leftarrow_{\text{R}} [n] \\ \text{coins}(\mathcal{B})}} \left[\mathcal{B}(D'_{-j}) \in T \right] \geq e^{-2\varepsilon} \Pr_{\text{coins}(\mathcal{B})} \left[\mathcal{B}(D'_{-k}) \in T \right].$$

Combining the two inequalities shows that for every database $D' \in \mathcal{X}^n$ and every $i, k \in [n]$,

$$\Pr_{\text{coins}(\mathcal{B}')} \left[\mathcal{B}'(D'_{-k}) \in T \right] \leq e^{3\varepsilon} \Pr_{\text{coins}(\mathcal{B}')} \left[\mathcal{B}'(D'_{-i}) \in T \right] + e^{2\varepsilon}\delta.$$

By Lemma 2.2, we have the algorithm $\mathcal{B}(D'_1, \ldots, D'_{n-1}) = \mathcal{B}'(D'_1, \ldots, D'_{n-1}, \perp)$ is $(6\varepsilon, (e^{2\varepsilon} + e^{5\varepsilon})\delta)$-differentially private.

Finally, uniform convergence of the sampling error of $\mathcal{B}'$ implies that it remains an accurate algorithm, and hence so is $\mathcal{B}$. In particular, when D is a random sample of m rows from D' and d is the VC dimension of $\mathcal{Q}$, we have [1]

$$\Pr[\exists q \in \mathcal{Q} : |q(D) - q(D')| > \alpha'] \leq 4 \cdot \left(\frac{2em}{d} \right)^d \cdot \exp\left(-\frac{(\alpha')^2 m}{8} \right).$$

Taking α' as in the theorem statement makes the total failure probability of $\mathcal{B}$ at most $1/2$. ∎

Now we proceed to complete the proof of Theorem 3.15. Suppose $(\mathcal{Q}, \mathcal{X})$ has sample complexity greater than n for $(\alpha + \alpha', \beta)$-accuracy (with failure probability $1/2$) and $(6\varepsilon, (e^{2\varepsilon} + e^{5\varepsilon})\delta)$-differential privacy. By Lemma 3.16, for every (α, β)-accurate mechanism $\mathcal{A}$ for $\mathcal{Q}$ there exists a database $D \in \mathcal{X}^m$ with $m = \lfloor n\varepsilon \rfloor$, a set T, and an index i such that

$$\Pr_{\substack{j \leftarrow_R [m] \\ \text{coins}(\mathcal{A})}} [\mathcal{A}(D_{-j}) \in T] > e^{\varepsilon} \Pr_{\text{coins}(\mathcal{A})} [\mathcal{A}(D_{-i}) \in T] + \delta. \tag{3}$$

We now argue that it is without loss of generality to restrict our attention to mechanisms $\mathcal{A}$ whose range is the finite set $I_m^{|\mathcal{Q}|} = \{0, \frac{1}{2m}, \frac{1}{m}, \ldots, 1 - \frac{1}{2m}, 1\}^{|\mathcal{Q}|}$. To see this, note that the exact answer to any counting query q on a database $D \in \mathcal{X}^m$ is in the set $\{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1 - \frac{1}{m}, 1\}$. Therefore, if an answer $a \in [0,1]$ satisfies $|a - q(D)| \leq \alpha$, then the value

$$\bar{a} = \frac{1}{2m} \cdot (\lceil (a - \alpha)m \rceil + \lfloor (a + \alpha)m \rfloor)$$

is a point in I_m that also satisfies $|\bar{a} - q(D)| \leq \alpha$. Thus, we will henceforth assume that the mechanism's output lies in this finite range.

We now apply the min-max theorem from game theory (or, equivalently, linear programming duality), to exhibit a fixed distribution on (D, T, i) for which inequality (3) holds. Specifically, consider a two-player zero-sum game in which Player 1 chooses a triple (D, T, i), where $D \in \mathcal{X}^m$, $T \subseteq I_m^{|\mathcal{Q}|}$, and $i \in [m]$, and Player 2 chooses a randomized function $\mathcal{A} : \mathcal{X}^m \to I_m^{|\mathcal{Q}|}$ that is (α, β)-accurate for $\mathcal{Q}$. Let the payoff to Player 1 be

$$\Pr_{j \leftarrow_R [m]} [\mathcal{A}(D_{-j}) \in T] - e^{\varepsilon} \mathbb{I}(\mathcal{A}(D_{-i}) \in T).$$

By inequality (3), the value of this game is greater than δ. So by the min-max theorem there exists a mixed strategy for Player 1 that achieves a payoff greater than δ against any mixed strategy for Player 2. (Note that we can apply the min-max theorem because we have assumed that the mechanism's output lies in a finite range.) That is, there exists a distribution $\mathcal{D}$ over triples (D, T, i) such that for any randomized algorithm $\mathcal{A} : \mathcal{X}^m \to I_m^{|\mathcal{Q}|}$ that takes any D to a feasible vector in $F_{\alpha,\beta}(D)$ with probability at least $2/3$,

$$\Pr_{\substack{j \leftarrow_R [m] \\ \text{coins}(\mathcal{A}) \\ (D,T,i) \leftarrow_R \mathcal{D}}} [\mathcal{A}(D_{-j}) \in T] > e^{\varepsilon} \cdot \Pr_{\substack{\text{coins}(\mathcal{A}) \\ (D,T,i) \leftarrow_R \mathcal{D}}} [\mathcal{A}(D_{-i}) \in T] + \delta. \tag{4}$$

Now consider the following code: *Gen* samples a database D, a set T, and an index i according to the promised distribution $\mathcal{D}$. The codebook C is $(D_{\pi(1)}, \ldots, D_{\pi(m)})$, where $\pi : [m] \to [m]$ is a random permutation. On input of an answer vector a, the algorithm *Trace* checks whether $a \in T$. If it is, then *Trace* outputs $\pi(i)$ and, otherwise, outputs $\perp$.

To analyze the security of this code, fix a coalition S of $m-1$ users using a pirate strategy $\mathcal{A}_{FP}$. Because the codebook is a random permutation of the rows of D, it is equivalent to analyzing the original database D and a random coalition of $m-1$ users. Thus the part of the codebook C_S given to the pirates is a random set of $m-1$ rows from D, i.e., D_{-j} for a random $j \in [m]$ with the junk row at index j removed. The condition that $\mathcal{A}_{FP}$ outputs a feasible answer vector is equivalent to $a = \mathcal{A}_{FP}(C_S)$ being an (α, β)-accurate answer vector. Therefore, letting $\mathcal{A} : \mathcal{X}^m \to I_m^{|\mathcal{Q}|}$ be the algorithm that runs $\mathcal{A}_{FP}$ on its input with the junk row removed, we have

$$\Pr_{Gen,Trace,\mathcal{A}_{FP}}[Trace(C,a)\neq\perp] = \Pr_{\substack{\text{coins}(\mathcal{A}_{FP})\\(D,T,i)\leftarrow_R\mathcal{D},\pi}}[\mathcal{A}_{FP}(C_S)\in T] = \Pr_{\substack{j\leftarrow_R[m],\text{coins}(\mathcal{A})\\(D,T,i)\leftarrow_R\mathcal{D}}}[\mathcal{A}(D_{-j})\in T].$$

However, the probability that *Trace* outputs the user j not in the coalition is

$$\Pr_{Gen,Trace,\mathcal{A}_{FP}}[Trace(C,a)=i] = \Pr_{\substack{j\leftarrow_R[m],\text{coins}(\mathcal{A}_{FP})\\(D,T,i)\leftarrow_R\mathcal{D},\pi}}[Trace(C,a)=i\wedge j=i] = \frac{1}{m}\cdot\Pr_{\text{coins}(\mathcal{A}),(D,T,i)\leftarrow_R\mathcal{D}}[\mathcal{A}(D_{-i})\in T],$$

because the events $\{j=i\}$ and $\{Trace(C,a)=i\}$ are independent. Thus by (4),

$$\Pr[Trace(a)\neq\perp] > e^{\varepsilon}m\cdot\Pr[Trace(a)\in[m]\setminus S]+\delta,$$

where both probabilities are taken over the coins of *Gen*, *Trace*, and $\mathcal{A}_{FP}$. ∎

4. A composition theorem for sample complexity. In this section we state and prove a composition theorem for sample-complexity lower bounds. At a high level the composition theorem starts with two pairs, $(\mathcal{Q},\mathcal{X})$ and $(\mathcal{Q}',\mathcal{X}')$, for which we know sample-complexity lower bounds of n and n', respectively, and attempts to prove a sample-complexity lower bound of $n\cdot n'$ for a related family of queries on a related data universe.

Specifically, our sample-complexity lower bound will apply to the "product" of $\mathcal{Q}$ and $\mathcal{Q}'$, defined on $\mathcal{X}\times\mathcal{X}'$. We define the product $\mathcal{Q}\wedge\mathcal{Q}'$ to be

$$\mathcal{Q}\wedge\mathcal{Q}'=\{q\wedge q' : (x,x')\mapsto q(x)\wedge q'(x') \mid q\in\mathcal{Q}, q\in\mathcal{Q}'\}.$$

Since q,q' are boolean valued, their conjunction can also be written $q(x)q'(x')$.

We now begin to describe how we can prove a sample-complexity lower bound for $\mathcal{Q}\wedge\mathcal{Q}'$. First, we describe a certain product operation on databases. Let $D\in\mathcal{X}^n$, $D=(x_1,\ldots,x_n)$, be a database. Let $D'_1,\ldots,D'_n\in(\mathcal{X}')^{n'}$, where $D'_i=(x'_{i1},\ldots,x'_{in'})$ be n databases. We define the product database $D^*=D\times(D'_1,\ldots,D'_n)\in(\mathcal{X}\times\mathcal{X}')^{n\cdot n'}$ as follows: For every $i=1,\ldots,n, j=1,\ldots,n'$, let the (i,j)th row of D^* be $x^*_{(i,j)}=(x_i,x'_{ij})$. Note that we index the rows of D^* by (i,j). We will sometimes refer to $D'_1,\ldots,D'_n$ as the subdatabases of D^*.

The key property of these databases is that we can use a query $q\wedge q'\in\mathcal{Q}\wedge\mathcal{Q}'$ to compute a "subset sum" of the vector $s_{q'}=(q'(D'_1),\ldots,q'(D'_n))$ consisting of the answers to q' on each of the n subdatabases. That is, for every $q\in\mathcal{Q}$ and $q'\in\mathcal{Q}'$,

$$(q\wedge q')(D^*)=\frac{1}{n\cdot n'}\sum_{i=1}^{n}\sum_{j=1}^{n'}(q\wedge q')(x^*_{(i,j)})=\frac{1}{n}\sum_{i=1}^{n}q(x_i)q'(D'_i). \tag{5}$$

Thus, every approximate answer $a_{q\wedge q'}$ to a query $q\wedge q'$ places a subset-sum constraint on the vector $s_{q'}$. (Namely, $a_{q\wedge q'}\approx\frac{1}{n}\sum_{i=1}^{n}q(x_i)q'(D'_i)$.) If the database D and family $\mathcal{Q}$ are chosen appropriately, and the answers are sufficiently accurate, then we will be able to reconstruct a good approximation to $s_{q'}$. Indeed, this sort of "reconstruction attack" is the core of many lower bounds for differential privacy, starting with the work of Dinur and Nissim [16]. The setting they consider is essentially the

special case of what we have just described, where $D'_1, \ldots, D'_n$ are each just a single bit ($\mathcal{X}' = \{0,1\}$, and $\mathcal{Q}'$ contains only the identity query). In section 5 we will discuss choices of D and $\mathcal{Q}$ that allow for this reconstruction.

We now state the formal notion of reconstruction attack that we want D and $\mathcal{Q}$ to satisfy.

DEFINITION 4.1 (reconstruction attacks). *Let $\mathcal{Q}$ be a family of counting queries over a data universe $\mathcal{X}$. Let $n \in \mathbb{N}$ and $\alpha', \alpha, \beta \in [0,1]$ be parameters. Let $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ be a database. Suppose there is an adversary $\mathcal{B}_D : \mathbb{R}^{|\mathcal{Q}|} \to [0,1]^n$ with the following property: For every vector $s \in [0,1]^n$ and every sequence $a = (a_q)_{q \in \mathcal{Q}} \in \mathbb{R}^{|\mathcal{Q}|}$ such that*

$$\left| a_q - \frac{1}{n} \sum_{i=1}^{n} q(x_i) s_i \right| < \alpha$$

for at least a $1 - \beta$ fraction of queries $q \in \mathcal{Q}$, $\mathcal{B}_D(a)$ outputs a vector $t \in [0,1]^n$ such that

$$\frac{1}{n} \sum_{i=1}^{n} |t_i - s_i| \leq \alpha'.$$

Then we say that $D \in \mathcal{X}^n$ enables an α'-reconstruction attack from (α, β)-accurate answers to $\mathcal{Q}$.

A reconstruction attack itself implies a sample-complexity lower bound, as in [16]. However, we show how to obtain stronger sample-complexity lower bounds from the reconstruction attack by applying it to a product database D^* to obtain accurate answers to queries on its subdatabases. For each query $q' \in \mathcal{Q}'$, we run the adversary promised by the reconstruction attack on the approximate answers given to queries of the form $(q \wedge q') \in \mathcal{Q} \wedge \{q'\}$. As discussed above, answers to these queries will approximate subset sums of the vector $s_{q'} = (q'(D'_1), \ldots, q'(D'_n))$. When the reconstruction attack is given these approximate answers, it returns a vector $t_{q'} = (t_{q',1}, \ldots, t_{q',n})$ such that $t_{q',i} \approx s_{q',i} = q'(D'_i)$ on average over i. Running the reconstruction attack for every query q' gives us a collection $t = (t_{q',i})_{q' \in \mathcal{Q}', i \in [n]}$, where $t_{q',i} \approx q'(D'_i)$ on average over both q' and i. By an application of Markov's inequality, for most of the subdatabases D'_i, we have that $t_{q',i} \approx q'(D'_i)$ on average over the choice of $q' \in \mathcal{Q}'$. For each i such that this guarantee holds, another application of Markov's inequality shows that for most queries $q' \in \mathcal{Q}'$ we have $t_{q',i} \approx q'(D'_i)$, which is our definition of (α, β)-accuracy (later enabling us to apply a reidentification adversary for $\mathcal{Q}'$).

The algorithm we have described for obtaining accurate answers on the subdatabases is formalized in Figure 1.

We are now in a position to state the main lemma that enables our composition technique. The lemma says that if we are given accurate answers to $\mathcal{Q} \wedge \mathcal{Q}'$ on D^*

Let $a = (a_{q \wedge q'})_{q \in \mathcal{Q}, q' \in \mathcal{Q}'}$ be an answer vector.
Let $\mathcal{B}_D : \mathbb{R}^{|\mathcal{Q}|} \to [0,1]^n$ be a reconstruction attack.
For each $q' \in \mathcal{Q}'$
  Let $(t_{q',1}, \ldots, t_{q',n}) = \mathcal{B}_D((a_{q \wedge q'})_{q \in \mathcal{Q}})$
Output $(t_{q',i})_{q' \in \mathcal{Q}', i \in [n]}$.

FIG. 1. *The reconstruction $\mathcal{R}^*_D(a)$.*

and the database $D \in \mathcal{X}^n$ enables a reconstruction attack from accurate answers to $\mathcal{Q}$, then we can obtain accurate answers to $\mathcal{Q}'$ on most of the subdatabases $D'_1, \ldots, D'_n \in (\mathcal{X}')^{n'}$.

LEMMA 4.2. *Let $D \in \mathcal{X}^n$ and $D'_1, \ldots, D'_n \in (\mathcal{X}')^{n'}$ be databases and let $D^* \in (\mathcal{X} \times \mathcal{X}')^{n \cdot n'}$ be as above. Let $a = (a_{q \wedge q'})_{q \in \mathcal{Q}, q' \in \mathcal{Q}'} \in \mathbb{R}^{|\mathcal{Q} \wedge \mathcal{Q}'|}$. Let $\alpha', \alpha, \beta \in [0,1]$ be parameters. Suppose that for some parameter $c > 1$, the database D enables an α'-reconstruction attack from $(\alpha, c\beta)$-accurate answers to $\mathcal{Q}$. Then if $(t_{q',i})_{q' \in \mathcal{Q}', i \in [n]} = \mathcal{R}^*_D(a)$ (Figure 1),*

$$\begin{aligned} &a \text{ is } (\alpha, \beta)\text{-accurate for } \mathcal{Q} \wedge \mathcal{Q}' \text{ on } D^* \\ &\implies \Pr_{i \leftarrow_R [n]} \left[(t_{q',i})_{q' \in \mathcal{Q}'} \text{ is } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}' \text{ on } D_i \right] \geq 5/6. \end{aligned}$$

The additional bookkeeping in the proof is to handle the case where a is only accurate for most queries. In this case the reconstruction attack may fail completely for certain queries $q' \in \mathcal{Q}'$ and we need to account for this additional source of error.

Proof of Lemma 4.2. Assume the answer vector $a = (a_{q \wedge q'})_{q \in \mathcal{Q}, q' \in \mathcal{Q}'}$ is (α, β)-accurate for $\mathcal{Q} \wedge \mathcal{Q}'$ on $D^* = D \times (D'_1, \ldots, D'_n)$. By assumption, D enables a reconstruction attack $\mathcal{B}_D$ that succeeds in reconstructing an approximation to $s_{q'} = (q'(D'_1), \ldots, q'(D'_n))$ when given $(\alpha, c\beta)$-accurate answers for the family of queries $\mathcal{Q} \wedge \{q'\}$. Consider the set of q' on which the reconstruction attack succeeds, i.e.,

$$\mathcal{Q}'_{good} = \{q' \mid (a_{q \wedge q'})_{q \in \mathcal{Q}} \text{ is } (\alpha, c\beta)\text{-accurate for } \mathcal{Q} \wedge \{q'\}\}.$$

Since a is (α, β)-accurate, an application of Markov's inequality shows that

$$\Pr\left[q' \in \mathcal{Q}'_{good}\right] \geq 1 - 1/c.$$

Thus, $|\mathcal{Q}'_{good}| \geq (1 - 1/c)|\mathcal{Q}'|$.

Recall that, by (5), we can interpret answers to $\mathcal{Q} \wedge \mathcal{Q}'$ as subset sums of answers to the subdatabases, so for every $q' \in \mathcal{Q}'_{good}$,

$$\left| a_{q \wedge q'} - \frac{1}{n} \sum_{i=1}^{n} q(x_i) q'(D'_i) \right| < \alpha$$

for at least a $1 - c\beta$ fraction of queries $q \wedge q' \in \mathcal{Q} \wedge \{q'\}$. Since D enables a reconstruction attack from $(\alpha, c\beta)$-accurate answers to $\mathcal{Q}$, by Definition 4.1, $\mathcal{B}_D((a_{q \wedge q'})_{q \in \mathcal{Q}})$ recovers a vector $t_{q'} \in [0,1]^n$ such that

$$\frac{1}{n} \sum_{i=1}^{n} |t_{q',i} - q'(D'_i)| < \alpha'.$$

Since this holds for every $q' \in \mathcal{Q}'_{good}$, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{q' \leftarrow_R \mathcal{Q}'_{good}, i \leftarrow_R [n]} \left[|t_{q',i} - q'(D'_i)| \right] \leq \alpha' \\ &\implies \Pr_{i \leftarrow_R [n]} \left[\mathop{\mathbb{E}}_{q' \in \mathcal{Q}'_{good}} \left[|t_{q',i} - q'(D'_i)| \right] \leq 6\alpha' \right] \geq 5/6 \\ &\implies \Pr_{i \leftarrow_R [n]} \left[|t_{q',i} - q'(D'_i)| \leq 6c\alpha' \text{ for at least a } 1 - 1/c \text{ fraction of } q' \in \mathcal{Q}'_{good} \right] \geq 5/6 \\ &\implies \Pr_{i \leftarrow_R [n]} \left[|t_{q',i} - q'(D'_i)| \leq 6c\alpha' \text{ for at least a } 1 - 2/c \text{ fraction of } q' \in \mathcal{Q}' \right] \geq 5/6. \end{aligned}$$

The first two implications are Markov's inequality, and the final implication is because $|\mathcal{Q}'_{good}| \geq (1 - 1/c)|\mathcal{Q}'|$. The statement inside the final probability is precisely that $(t_{q',i})_{q' \in \mathcal{Q}'}$ is $(6c\alpha', 2/c)$-accurate for $\mathcal{Q}'$ on D'_i. This completes the proof of the lemma. ∎

We now explain how the main lemma allows us to prove a composition theorem for sample-complexity lower bounds. We start with a query family $\mathcal{Q}$ on a database $D \in \mathcal{X}^n$ that enables a reconstruction attack, and a distribution $\mathcal{D}'$ over databases in $(\mathcal{X}')^{n'}$ that is reidentifiable from answers to a family $\mathcal{Q}'$. We show how to combine these objects to form a reidentifiable distribution $\mathcal{D}^*$ for queries $\mathcal{Q} \wedge \mathcal{Q}'$ over $(\mathcal{X} \times \mathcal{X}')^{n \cdot n'}$, yielding a sample-complexity lower bound of $n \cdot n'$.

A sample from $\mathcal{D}^*$ consists of $D^* = D \times (D'_1, \ldots, D'_n)$, where each subdatabase D'_i is an independent sample from from $\mathcal{D}'$. The main lemma above shows that if there is an algorithm $\mathcal{A}$ that is accurate for $\mathcal{Q} \wedge \mathcal{Q}'$ on D^*, then an adversary can reconstruct accurate answers to $\mathcal{Q}'$ on most of the subdatabases $D'_1, \ldots, D'_n$. Since these subdatabases are drawn from a reidentifiable distribution, the adversary can then reidentify a member of one of the subdatabases D'_i. Since the identified member of D'_i is also a member of D^*, we will have a reidentification attack against D^* as well.

We are now ready to formalize our composition theorem.

THEOREM 4.3. *Let $\mathcal{Q}$ be a family of counting queries on $\mathcal{X}$, and let $\mathcal{Q}'$ be a family of counting queries on $\mathcal{X}'$. Let $\gamma, \xi, \alpha', \alpha, \beta \in [0,1]$ be parameters. Assume that for some parameters $c > 1$, $\gamma, \xi, \alpha', \alpha, \beta \in [0,1]$, the following both hold:*

1. *There exists a database $D \in \mathcal{X}^n$ that enables an α'-reconstruction attack from $(\alpha, c\beta)$-accurate answers to $\mathcal{Q}$.*
2. *There is a distribution $\mathcal{D}'$ on databases $D \in (\mathcal{X}')^{n'}$ that is (γ, ξ)-reidentifiable from $(6c\alpha', 2/c)$-accurate answers to $\mathcal{Q}'$.*

Then there is a distribution on databases $D^* \in (\mathcal{X} \times \mathcal{X}')^{n \cdot n'}$ that is $(\gamma + 1/6, \xi)$-reidentifiable from (α, β)-accurate answers to $\mathcal{Q} \wedge \mathcal{Q}'$.

Proof. Let $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ be the database that enables a reconstruction attack (Definition 4.1). Let $\mathcal{D}'$ be the promised reidentifiable distribution on databases $D \in (\mathcal{X}')^{n'}$ and $\mathcal{B}' : (\mathcal{X}')^{n'} \times \mathbb{R}^{|\mathcal{Q}'|} \to [n'] \cup \{\perp\}$ be the promised adversary (Definition 2.10).

In Figure 2, we define a distribution $\mathcal{D}^*$ on databases $D' \in (\mathcal{X} \times \mathcal{X}')^{n \cdot n'}$. In Figure 3, we define an adversary $\mathcal{B}^* : (\mathcal{X} \times \mathcal{X}')^{n \cdot n'} \times \mathbb{R}^{|\mathcal{Q} \wedge \mathcal{Q}'|}$ for a reidentification attack. The shared state of $\mathcal{D}^*$ and $\mathcal{B}^*$ will be the shared state of $\mathcal{D}'$ and $\mathcal{B}'$. The next two claims show that $\mathcal{D}^*$ satisfies the two properties necessary to be a $(\gamma + 1/6, \xi)$-reidentifiable distribution (Definition 2.10).

CLAIM 4.4. $\Pr_{\substack{D^* \leftarrow_R \mathcal{D}^* \\ coins(\mathcal{A}), coins(\mathcal{B}^*)}} \left[\begin{smallmatrix} (\mathcal{B}^*(D^*, \mathcal{A}(D^*)) = \perp) \\ \wedge\ (\mathcal{A}(D^*) \text{ is } (\alpha, \beta)\text{-accurate for } \mathcal{Q} \wedge \mathcal{Q}') \end{smallmatrix} \right] \leq \gamma + 1/6.$

Let $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ be a database that enables reconstruction.
Let $\mathcal{D}'$ on $(\mathcal{X}')^{n'}$ be a reidentifiable distribution.
For $i = 1, \ldots, n$, choose $D'_i \leftarrow_R \mathcal{D}'$ (independently)
Output $D^* = D \times (D'_1, \ldots, D'_n) \in (\mathcal{X} \times \mathcal{X}')^{n \cdot n'}$

FIG. 2. *The new distribution $\mathcal{D}^*$.*

Let $D^* = D \times (D'_1, \ldots, D'_n)$.
Run $\mathcal{R}^*_D(\mathcal{A}(D^*))$ (Figure 1) to reconstruct a set of approximate answers $(t_{q',i})_{q' \in \mathcal{Q}', i \in [n]}$.
Choose a random $i \leftarrow_{\mathrm{R}} [n]$.
Output $\mathcal{B}'(D'_i, (t_{q',i})_{q' \in \mathcal{Q}'})$.

FIG. 3. *The privacy adversary* $\mathcal{B}^*(D^*, \mathcal{A}(D^*))$.

Proof of Claim 4.4. Assume that $\mathcal{A}(D^*)$ is (α, β)-accurate for $\mathcal{Q} \wedge \mathcal{Q}'$. By Lemma 4.2, we have

$$\Pr_{\substack{i \leftarrow_{\mathrm{R}} [n] \\ \mathrm{coins}(\mathcal{A}), \mathrm{coins}(\mathcal{B}^*)}} \left[\begin{array}{c} (\mathcal{A}(D^*) \text{ is } (\alpha, \beta)\text{-accurate for } \mathcal{Q} \wedge \mathcal{Q}') \\ \wedge ((t_{q',i})_{q' \in \mathcal{Q}'} \text{ is not } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}' \text{ on } D_i) \end{array} \right] \leq 1/6. \tag{6}$$

By construction of $\mathcal{B}^*$,

(7)

$$\begin{aligned} &\Pr_{D^* \leftarrow_{\mathrm{R}} \mathcal{D}^*} [(\mathcal{B}^*(D^*, \mathcal{A}(D^*)) = \perp) \wedge (\mathcal{A}(D^*) \text{ is } (\alpha, \beta)\text{-accurate for } \mathcal{Q} \wedge \mathcal{Q}')] \\ &\quad = \Pr_{\substack{D^* \leftarrow_{\mathrm{R}} \mathcal{D}^* \\ i \leftarrow_{\mathrm{R}} [n]}} [(\mathcal{B}'(D'_i, (t_{q',i})_{q' \in \mathcal{Q}'}) = \perp) \wedge (\mathcal{A}(D^*) \text{ is } (\alpha, \beta)\text{-accurate for } \mathcal{Q} \wedge \mathcal{Q}')] \\ &\quad \leq \Pr_{\substack{D^* \leftarrow_{\mathrm{R}} \mathcal{D}^* \\ i \leftarrow_{\mathrm{R}} [n]}} [(\mathcal{B}'(D'_i, (t_{q',i})_{q' \in \mathcal{Q}'}) = \perp) \wedge ((t_{q',i}) \text{ is } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}')] + \frac{1}{6} \end{aligned}$$

where the last inequality is by (6). Thus, it suffices to prove that

$$\Pr_{\substack{D^* \leftarrow_{\mathrm{R}} \mathcal{D}^* \\ i \leftarrow_{\mathrm{R}} [n]}} [(\mathcal{B}'(D'_i, (t_{q',i})_{q' \in \mathcal{Q}'}) = \perp) \wedge ((t_{q',i}) \text{ is } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}')] \leq \gamma. \tag{8}$$

We prove this inequality by giving a reduction to the reidentifiability of $\mathcal{D}'$. Consider the following sanitizer $\mathcal{A}'$: On input $D' \leftarrow_{\mathrm{R}} \mathcal{D}'$, $\mathcal{A}'$ first chooses a random index $i^* \leftarrow_{\mathrm{R}} [n]$. Next, it samples $D'_1, \ldots, D'_{i^*-1}, D'_{i^*+1}, \ldots, D'_n \leftarrow_{\mathrm{R}} \mathcal{D}'$ independently, and sets $D'_{i^*} = D'$. Finally, it runs $\mathcal{A}$ on $D^* = D \times (D'_1, \ldots, D'_n)$ and then runs the reconstruction attack $\mathcal{R}^*$ to recover answers $(t_{q',i})_{q' \in \mathcal{Q}', i \in [n]}$ and outputs $(t_{q',i^*})_{q' \in \mathcal{Q}'}$.

Notice that since $D'_1, \ldots, D'_n$ are all independently and identically distributed (i.i.d.) samples from $\mathcal{D}'$, their joint distribution is independent of the choice of i^*. Specifically, in the view of $\mathcal{B}^*$, we could have chosen i^* after seeing its output on D^*. Therefore, the following random variables are identically distributed:

1. $(t_{q',i})_{q' \in \mathcal{Q}'}$, where $(t_{q',i})_{q' \in \mathcal{Q}', i \in [n]}$ is the output of $\mathcal{R}^*_D(\mathcal{A}(D^*))$ on $D^* \leftarrow_{\mathrm{R}} \mathcal{D}^*$, and $i \leftarrow_{\mathrm{R}} [n]$.
2. $\mathcal{A}'(D')$, where $D' \leftarrow_{\mathrm{R}} \mathcal{D}'$.

Thus we have

$$\begin{aligned} &\Pr_{\substack{D^* \leftarrow_{\mathrm{R}} \mathcal{D}^* \\ i \leftarrow_{\mathrm{R}} [n]}} [(\mathcal{B}'(D'_i, (t_{q',i})_{q' \in \mathcal{Q}'}) = \perp) \wedge ((t_{q',i}) \text{ is } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}')] \\ &\quad = \Pr_{D' \leftarrow_{\mathrm{R}} \mathcal{D}'} [(\mathcal{B}'(D', \mathcal{A}'(D')) = \perp) \wedge (\mathcal{A}'(D') \text{ is } (6c\alpha', 2/c)\text{-accurate for } \mathcal{Q}')] \leq \gamma, \end{aligned}$$

where the last inequality follows because $\mathcal{D}'$ is (γ, ξ)-reidentifiable from $(6c\alpha', 2/c)$-accurate answers to $\mathcal{Q}'$. Thus we have established (8). Combining (7) and (8) completes the proof of the claim. ∎

The next claim follows directly from the definition of $\mathcal{B}^*$ and the fact that $\mathcal{D}'$ is (γ, ξ)-reidentifiable.

CLAIM 4.5. *For every* $(i,j) \in [n] \times [n']$, $\Pr_{D \leftarrow_R \mathcal{D}^*}\left[\mathcal{B}^*(D, \mathcal{A}(D_{-(i,j)})) = (i,j)\right] \leq \xi$.

Combining Claims 4.4 and 4.5 suffices to prove that $\mathcal{D}^*$ is $(\gamma + 1/6, \xi)$-reidentifiable from (α, β)-accurate answers to $\mathcal{Q} \wedge \mathcal{Q}'$, completing the proof of the theorem. ∎

The proof of Theorem 4.3 also yields a composition theorem for generalized fingerprinting codes. Specifically, Theorem 4.6 below shows how to combine a reconstruction attack for a query family $\mathcal{Q}$ on a database $D \in \mathcal{X}^n$ with a $(n', \mathcal{Q}')$-generalized fingerprinting code to obtain an $(n \cdot n', \mathcal{Q} \wedge \mathcal{Q}')$-generalized fingerprinting code.

THEOREM 4.6. *Let $\mathcal{Q}$ be a family of counting queries on $\mathcal{X}$, and let $\mathcal{Q}'$ be a family of counting queries on $\mathcal{X}'$. Let $\gamma, \xi, \alpha', \alpha, \beta \in [0,1]$ be parameters. Assume that for some parameters $c > 1$, $\gamma, \xi, \alpha', \alpha, \beta \in [0,1]$, the following both hold:*

1. *There exists a database $D \in \mathcal{X}^n$ that enables an α'-reconstruction attack from $(\alpha, c\beta)$-accurate answers to $\mathcal{Q}$.*
2. *There exists an $(n', \mathcal{Q}')$-generalized fingerprinting code for $(6c\alpha', 2/c)$-accuracy with security (γ, ξ).*

Then there is an $(n \cdot n', \mathcal{Q} \wedge \mathcal{Q}')$-generalized fingerprinting code for (α, β)-accuracy with security $(\gamma + 1/6, \xi)$.

5. Applications of the composition theorem. In this section we show how to use our composition theorem (section 4) to combine our new lower bounds for 1-way marginal queries from section 3 with (variants of) known lower bounds from the literature to obtain our main results. In section 5.1 we prove a lower bound for k-way marginal queries when α is not too small (at least inverse polynomial in d), thereby proving Theorem 1.2 in the introduction. Then in section 5.2 we obtain a similar lower bound for arbitrary counting queries that allows α to take a wider range of parameters.

5.1. Lower bounds for k-way marginals. In this section, we carry out the composition of sample-complexity lower bounds for k-way marginals as described in the introduction (Theorem 1.2). Recall that we obtain our new $\tilde{\Omega}(k\sqrt{d}/\alpha^2)$ lower bound by combining three lower bounds:

1. Our reidentification-based $\tilde{\Omega}(\sqrt{d})$ lower bound for 1-way marginals (section 3.2).
2. A known reconstruction-based lower bound of $\Omega(k)$ for k-way marginals.
3. A known reconstruction-based lower bound of $\Omega(1/\alpha^2)$ for k-way marginals.

The lower bound of $\Omega(k)$ for k-way marginals is a special case of a lower bound of $\Omega(VC(\mathcal{Q}))$ due to [43] and based on [16], where $VC(\mathcal{Q})$ is the VC *dimension* of $\mathcal{Q}$. The lower bound of $\Omega(1/\alpha^2)$ for k-way marginals is due to [40, 15].

To apply our composition theorem, we need to formulate these reconstruction attacks in the language of Definition 4.1. In particular, we observe that these reconstruction attacks readily generalize to allow us to reconstruct fractional vectors $s \in [0,1]^n$, instead of just boolean vectors as in [16, 43].

5.1.1. The $\Omega(k)$ lower bound. First we state and prove that the linear dependence on k is necessary.

DEFINITION 5.1 (VC dimension of counting queries). *Let $\mathcal{Q}$ be a collection of counting queries over a data universe $\mathcal{X}$. We say a set $\{x_1, \ldots, x_k\} \subseteq \mathcal{X}$ is* shattered *by $\mathcal{Q}$ if for every string $v \in \{0,1\}^k$, there exists a query $q \in \mathcal{Q}$ such that*

Input: Queries $\mathcal{Q}$ and $(a_q)_{q\in\mathcal{Q}}$ that are $(\alpha, 0)$-accurate for s.
Find any $t \in [0,1]^n$ such that

$$\left| a_q - \frac{1}{n}\sum_{i=1}^{n} q(x_i)t_i \right| \le \alpha \quad \forall q \in \mathcal{Q}.$$

Output: t.

FIG. 4. *The reconstruction adversary* $\mathcal{B}(D, a)$.

$(q(x_1), \ldots, q(x_k)) = (v_1, \ldots, v_k)$. *The* VC dimension *of* $\mathcal{Q}$ *denoted* $VC(\mathcal{Q})$ *is the cardinality of the largest subset of* $\mathcal{X}$ *that is shattered by* $\mathcal{Q}$.

FACT 5.2. *The set of* k*-way conjunctions* $\mathcal{M}_{k,d}$ *over any data universe* $\{0,1\}^d$ *with* $d \ge k$ *has VC dimension* $VC(\mathcal{M}_{k,d}) \ge k$.[8]

Proof. For each $i = 1, \ldots, k$, let $x_i = (1, 1, \ldots, 0, \ldots, 1)$, where the zero is at the ith index. We will show that $\{x_1, \ldots, x_k\}$ is shattered by $\mathcal{M}_{k,d}$. For a string $v \in \{0,1\}^k$, let the query $q_v(x)$ take the conjunction of the bits of x at indices set to 0 in v. Then $q_v(x_i) = 1$ iff $v_i = 1$, so $(q_v(x_1), \ldots, q_v(x_k)) = (v_1, \ldots, v_k)$. ∎

LEMMA 5.3 (variant of [16, 43]). *Let* $\mathcal{Q}$ *be a collection of counting queries over a data universe* $\mathcal{X}$ *and let* $n = VC(\mathcal{Q})$. *Then there is a database* $D \in \mathcal{X}^n$ *which enables a* 4α*-reconstruction attack from* $(\alpha, 0)$*-accurate answers to* $\mathcal{Q}$.

Proof. Let $\{x_1, \ldots, x_n\}$ be shattered by $\mathcal{Q}$, and consider the database $D = (x_1, \ldots, x_n)$. Let $s \in [0,1]^n$ be an arbitrary string to be reconstructed and let $a = (a_q)_{q\in\mathcal{Q}}$ be $(\alpha, 0)$-accurate answers. That is, for every $q \in \mathcal{Q}$

$$\left| a_q - \frac{1}{n}\sum_{i=1}^{n} q(x_i)s_i \right| \le \alpha.$$

Consider the brute-force reconstruction attack $\mathcal{B}$ defined in Figure 4. Notice that, since a is $(\alpha, 0)$-accurate, $\mathcal{B}$ always finds a suitable vector t. Namely, the original database s satisfies the constraints. We will show that the reconstructed vector t satisfies

$$\frac{1}{n}\sum_{i=1}^{n} |t_i - s_i| \le 4\alpha.$$

Let T be the set of coordinates on which $t_i > s_i$ and let S be the set of coordinates where $s_i > t_i$. Note that

$$\sum_{i=1}^{n} |t_i - s_i| = \sum_{i\in T}(t_i - s_i) + \sum_{i\in S}(s_i - t_i).$$

We will show that absolute values of the sums over T and S are each at most 2α. Since $\{x_1, \ldots, x_n\}$ is shattered by $\mathcal{Q}$, there is a query $q \in \mathcal{Q}$ such that $q(x_i) = 1$ iff

[8] More precisely, $VC(\mathcal{M}_{k,d}) \ge k \log_2(\lfloor d/k \rfloor)$, but we use the simpler bound $VC(\mathcal{M}_{k,d}) \ge k$ to simplify calculation, since our ultimate lower bounds are already suboptimal by polylog(d) factors for other reasons.

$i \in T$. Therefore, by the definitions of t and $(\alpha, 0)$-accuracy,

$$\left| a_q - \frac{1}{n}\sum_{i=1}^{n} q(x_i)t_i \right| = \left| a_q - \frac{1}{n}\sum_{i\in T} t_i \right| \le \alpha \quad \text{and} \quad \left| a_q - \frac{1}{n}\sum_{i\in T} s_i \right| \le \alpha,$$

so by the triangle inequality, $\frac{1}{n}\sum_{i\in T}(t_i - s_i) \le 2\alpha$. An identical argument shows that $\frac{1}{n}\sum_{i\in S}(s_i - t_i) \le 2\alpha$, proving that t is an accurate reconstruction. ∎

5.1.2. The $\Omega(1/\alpha^2)$ lower bound for k-way marginals. We can now state in our terminology the lower bound of De from [15] (building on [40]) showing that the inverse-quadratic dependence on α is necessary.

THEOREM 5.4 (restatement of [15]). *Let k be any constant, $d \ge k$ be any integer, and let $\alpha \ge 1/d^{.499k}$ be a sufficiently small parameter*[9] *(i.e., bounded by an absolute constant). There exists a constant $\beta = \beta(k) > 0$ such that for every $\alpha' > 0$, there exists a database $D \in (\{0,1\}^d)^n$ with $n = \Omega_{\alpha',k}(1/\alpha^2)$ such that D enables an α'-reconstruction attack from (α, β)-accurate answers to the k-way marginals $\mathcal{M}_{k,d}$.*

Although the above theorem is a simple extension of De's lower bound, we sketch a proof for completeness, and refer the interested reader to [15] for a more detailed analysis.

Proof Sketch. The reconstruction attack uses the "ℓ_1-minimization" algorithm, which is shown in Figure 5. To prove that the reconstruction attack succeeds, we will show that there exists a database $D = (x_1, \ldots, x_n) \in \{0,1\}^{n\times d}$ such that for any $s \in [0,1]^n$, if a satisfies

$$\Pr_{q\in\mathcal{M}_{k,d}}\left[\left| a_q - \frac{1}{n}\sum_{i=1}^{n} q(x_i)s_i \right| \le \alpha\right] \ge 1 - \beta,$$

(i.e., a has (α, β)-accurate answers) then $\mathcal{B}_{\mathcal{M}_{k,d}}(D, a)$ returns a vector t such that $\|t - s\|_1 \le \alpha' \cdot n$. Henceforth we refer to such an a simply as (α, β)-accurate for $\mathcal{M}_{k,d}$ on (D, s), as a shorthand. The above guarantee must hold for suitable choices of n, β, and α' to satisfy the theorem.

We will argue that the reconstruction succeeds in two steps. First, we show that reconstruction succeeds if D is a nice. Second, we show that there exists a nice D that has the dimensions promised by the theorem.

Input: Queries $\mathcal{Q}$, $D = (x_1, \ldots, x_n) \in \{0,1\}^{n\times d}$ and $a = (a_q)_{q\in\mathcal{Q}}$.
Let $t \in [0,1]^n$ be

$$\arg\min_{t\in[0,1]^n} \sum_{q\in\mathcal{Q}} \left| a_q - \frac{1}{n}\sum_{i=1}^{n} q(x_i)t_i \right|$$

Output: t.

FIG. 5. *The reconstruction adversary $\mathcal{B}_{\mathcal{Q}}(D, a)$.*

[9] The constant .499 was chosen for simplicity, and can be replaced with any constant strictly smaller than .5.

To explain what we mean by a nice database D, for any $D = (x_1, \ldots, x_n) \in \{0,1\}^{n \times d}$ and family of queries $\mathcal{Q}$ on $\{0,1\}^d$, we define the matrix $M = M_{D,\mathcal{Q}} \in \{0,1\}^{n \times |\mathcal{Q}|}$ as $M(i,q) = q(x_i)$.

De analyzes this reconstruction attack in terms of certain properties of the matrix M. Before stating the conclusion, we will need to define the notion of a Euclidean section. Informally, a matrix M is a Euclidean section if its rowspace[10] contains only vectors that are "spread out."

DEFINITION 5.5 (Euclidean section). *A matrix $M \in \{0,1\}^{n \times m}$ is a δ-*Euclidean section *if for every vector a in the rowspace of M we have $\sqrt{m} \cdot \|a\|_2 \geq \|a\|_1 \geq \delta\sqrt{m} \cdot \|a\|_2$.*

LEMMA 5.6 (see [15]). *Let D be a database and $\mathcal{Q}$ be a set of queries such that $M_{D,\mathcal{Q}} \in \{0,1\}^{n \times |\mathcal{Q}|}$ is a δ-Euclidean section and the least singular value of $M_{D,\mathcal{Q}}$ is σ. Let $s \in [0,1]^n$ be arbitrary. There exists $\beta = \beta(\delta) > 0$ such that if a's are (α, β)-accurate answers for $\mathcal{Q}$ on (D, s), and $t = \mathcal{B}_{\mathcal{Q}}(D, a)$, then t satisfies*

$$\|s - t\|_1 \leq \gamma n$$

for $\gamma = O(\alpha\sqrt{n|\mathcal{Q}|}/\sigma)$. The constant hidden in the $O(\cdot)$ notation depends only on δ.

Thus, it suffices to find database D such that the matrix $M_{D,\mathcal{M}_{k,d}}$ is a Euclidean section (for some fixed constant $\delta > 0$) and has no "small" singular values. A result of Rudelson [45] (strengthening that of Kasiviswanathan et al. [40]) guarantees that such a database exists.

LEMMA 5.7 (see [45]). *Let $k \in \mathbb{N}$ be any constant. Let $d, n \in \mathbb{N}$ be such that $d^k \geq n \log n$. Let $D \in \{0,1\}^{n \times d}$ be a uniform random matrix. Then with probability at least $9/10$, the matrix $M_{D,\mathcal{M}_{k,d}}$ defined above has a least singular value of at least $\sigma = \Omega(d^{k/2})$ (where the hidden constant in the $\Omega(\cdot)$ may depend on k) and is a δ-Euclidean section for some constant $\delta > 0$ that depends only on k.*[11]

In particular, there exists a database $D \in \{0,1\}^{n \times d}$ such that the Hadamard product M satisfies the two properties above.

Using the above lemma, we can now complete the proof. Fix any constant $k \in \mathbb{N}$. Let α, d, n be any parameters such that $d \geq k$, $\alpha \geq 1/d^{.499k}$, and $d^k \geq n \log n$. The precise value of n will be determined later. Let $D \in \{0,1\}^{n \times d}$ be the database promised by Lemma 5.7. Let $\beta = \beta(k) > 0$ be a parameter to be chosen later. Let $\alpha' > 0$ be the desired accuracy of the reconstruction attack.

Now fix any $s \in [0,1]^n$ and let $a \in [0,1]^{|\mathcal{M}_{k,d}|}$ be (α, β)-accurate answers to $\mathcal{M}_{k,d}$ on (D, s). Now, if we let $t = \mathcal{B}_{\mathcal{M}_{k,d}}(D, a)$, by Lemma 5.6, provided that β is smaller than some constant that depends only on δ, which in turn depends only on k, we will have $\|s - t\|_1 \leq \gamma \cdot n$ for

$$\gamma = O\left(\frac{\alpha\sqrt{n|\mathcal{Q}|}}{\sigma}\right) = O\left(\frac{\alpha\sqrt{n}(d/k)^{k/2}}{d^{k/2}}\right) = O(\alpha\sqrt{n}).$$

[10] For a matrix M with rows $M_1, \ldots, M_n$, the rowspace of M is $\{a = \sum_{i=1}^n c_i M_i \mid c_1, \ldots, c_n \in \mathbb{R}\}$.

[11] Rudelson actually proves these statements about a related matrix $M_{D,\mathcal{Q}}$, where $\mathcal{Q} \subseteq \mathcal{M}_{k,d}$. Since, for the $\mathcal{Q}$ he considers, $|\mathcal{Q}| \geq |\mathcal{M}_{k,d}|/(2k)^k$, these statements can easily be seen to hold for the matrix $M_{D,\mathcal{M}_{k,d}}$ itself. Specifically, adding this many more columns to the matrix $M_{D,\mathcal{Q}}$ cannot decrease its least singular value (since $M_{D,\mathcal{Q}}$ already has more columns than rows), and can only decrease the Euclidean section parameter δ by a factor of at most $(2k)^k$.

Note that by Lemma 5.6, the hidden constant in the $O(\cdot)$ notation depends only on the parameter δ such that $M_{D,\mathcal{M}_{k,d}}$ is a δ-Euclidean section. By Lemma 5.7, the parameter δ depends only on k. Thus $\gamma = O(\alpha\sqrt{n})$, where the hidden constant depends only on k. Now, we can choose $n = \Omega(1/\alpha^2)$ such that $\gamma \leq \alpha'$. The hidden constant in the $\Omega(\cdot)$ will depend only on k and α', as required by the theorem. Note that, since we have assumed $\alpha \geq 1/d^{.499k}$, we have $n \log n = \tilde{O}(d^{.998k})$, and so we can define $n = \Omega_{k,\alpha'}(1/\alpha^2)$ while ensuring that $d^k \geq n \log n$. Similarly, we required that β is smaller than some constant that depends only on δ, which in turn depends only on k. Thus, we can set $\beta = \beta(k) > 0$ to be some sufficiently small constant depending only on k, as required by the theorem. This completes our sketch of the proof. ∎

5.1.3. Putting together the lower bound. Now we show how to combine the various attacks to prove Theorem 1.2 in the introduction. We obtain our lower bound by applying two rounds of composition. In the first round, we compose the reconstruction attack of Theorem 5.4 described above with the reidentifiable distribution for 1-way marginals. We then take the resulting reidentifiable distribution and apply a second round of composition using the reconstruction attack based on the VC dimension of k-way marginals.

We remark that it is necessary to apply the two rounds of composition in this order. In particular, we cannot prove Theorem 1.3 by composing first with the VC-dimension-based reconstruction attack. Our composition theorem requires a reidentifiable distribution from (α, β)-accurate answers for $\beta > 0$, whereas the reconstruction attack described in Lemma 5.3 requires $(\alpha, 0)$-accurate answers, and the reconstruction can fail if some queries have error much larger than α. The resulting reidentifiable distribution obtained from composing with this reconstruction attack will also require $(\alpha, 0)$-accurate answers, and thus cannot be composed further.

This limitation of Lemma 5.3 is inherent, because a sample-complexity upper bound of $\tilde{O}(\sqrt{d}/\alpha^2)$ can be achieved for answering any family of queries $\mathcal{Q}$ with (α, β)-accuracy (for any constant $\beta > 0$). Notice that this sample complexity is independent of $VC(\mathcal{Q})$.

We can now formally state and prove our sample-complexity lower bound for k-way marginals, thereby establishing Theorem 1.3 in the introduction.

THEOREM 5.8. *For every constant $\ell \in \mathbb{N}$, every $k, d \in \mathbb{N}$, $\ell + 2 \leq k \leq d$, and every sufficiently small (i.e., bounded by an absolute constant) $\alpha \geq 1/d^{.499\ell}$, there is an*

$$n = n(k, d, \alpha) = \tilde{\Omega}\left(\frac{k\sqrt{d}}{\alpha^2}\right)$$

such that there exists a distribution on n-row databases $D \in (\{0,1\}^d)^n$ that is $(1/2, o(1/n))$-reidentifiable from $(\alpha, 0)$-accurate answers to the k-way marginals $\mathcal{M}_{k,d}$.

Proof. We begin with the following two attacks:

1. By combining Theorems 3.5 and 3.4, there exists a distribution on databases $D' \in (\{0,1\}^{d/3})^{n_d}$ that is $(\gamma = 1/6, \xi = o(1/n_d n_\alpha n_k))$-reidentifiable from $(6c\alpha' = 1/3, 2/c = 1/75)$ accurate answers to the 1-way marginals $\mathcal{M}_{1,d/3}$ for $n_d = \tilde{\Omega}(\sqrt{d}/\log(n_d n_\alpha n_k))$. Here n_α and n_k are set below (the subscript corresponds to the primary parameter that each of the n's will depend on).
2. By Theorem 5.4 (with $\alpha' = 1/2700$ and $k = \ell$), there is a constant $\beta > 0$ such that for any $7200\alpha/\beta \geq 1/d^{.499\ell}$ there exists a database $D \in (\{0,1\}^{d/3})^{n_\alpha}$ for $n_\alpha = \tilde{\Omega}(1/\alpha^2)$ that enables a $(1/2700)$-reconstruction attack from $(7200\alpha/\beta, \beta)$-accurate answers to $\mathcal{M}_{\ell,d/3}$.

Applying Theorem 4.3 (with parameter $c = 150$), we obtain item 1′ below. We then bring in another reconstruction attack for the composition theorem.

1′. There exists a probability distribution on databases in $(\{0,1\}^{2d/3})^{n_d n_\alpha}$ that is $(1/3, o(1/n_d n_\alpha n_k))$-reidentifiable from $(6c'\alpha' = 7200\alpha/\beta, 2/c' = \beta/150)$-accurate answers to $\mathcal{M}_{\ell,d/3} \wedge \mathcal{M}_{1,d/3} \subset \mathcal{M}_{\ell+1,2d/3}$ (by applying Theorem 4.3 to 1 and 2 above).

2′. By Lemma 5.3 and Fact 5.2, there exists a database $D \in (\{0,1\}^{d/3})^{n_k}$ for $n_k = k - \ell - 1$, that enables an $(\alpha' = 4\alpha)$-reconstruction attack from $(\alpha, 0)$-accurate answers to the $(k - \ell - 1)$-way marginals $\mathcal{M}_{k-\ell-1,d/3}$. Note that $(k - \ell - 1) \geq 1$, since we have assumed $k \geq \ell + 2$.

We can then apply Theorem 4.3 to 1′ and 2′ (with parameter $c' = 300/\beta$). Thereby we obtain a distribution $\mathcal{D}$ on databases $D \in (\{0,1\}^{d/3} \times \{0,1\}^{d/3} \times \{0,1\}^{d/3})^{n_d n_\alpha n_k}$ that is $(1/2, \xi)$-reidentifiable from $(\alpha, 0)$-accurate answers to $\mathcal{M}_{k-\ell-1,d/3} \wedge \mathcal{M}_{\ell,d/3} \wedge \mathcal{M}_{1,d/3} \subset \mathcal{M}_{k,d}$.

To complete the theorem, first note that $(\alpha, 0)$-accurate answers to $\mathcal{M}_{k,d}$ imply $(\alpha, 0)$-accurate answers to any subset of $\mathcal{M}_{k,d}$. So our lower bound for the subset $\mathcal{M}_{k-\ell-1,d/3} \wedge \mathcal{M}_{\ell,d/3} \wedge \mathcal{M}_{1,d/3}$ is sufficient to obtain the desired lower bound. Finally, note that

$$n = n_d n_\alpha n_k = \tilde{\Omega}\left(\frac{k\sqrt{d}}{\alpha^2}\right),$$

as desired. This completes the proof. ∎

Using the composition Theorem 4.6 in place of Theorem 4.3, we obtain a version of Theorem 5.8 in the language of generalized fingerprinting codes.

THEOREM 5.9. *For every constant $\ell \in \mathbb{N}$, every $k, d \in \mathbb{N}$, $\ell + 2 \leq k \leq d$, and every sufficiently small (i.e., bounded by an absolute constant) $\alpha \geq 1/d^{.499\ell}$, there is an*

$$n = n(k, d, \alpha) = \tilde{\Omega}\left(\frac{k\sqrt{d}}{\alpha^2}\right)$$

such that there exists an $(n, \mathcal{M}_{k,d})$-generalized fingerprinting code that achieves security $(1/2, o(1/n))$ for $(\alpha, 0)$-accuracy.

5.1.4. A tight lower bound for 2-way marginals. Theorem 5.8 does not give any nontrivial lower bound for 2-way marginals. Intuitively, the problem is that the proof uses two rounds of composition, and thus if we try to instantiate the proof for 2-way marginals, one of the three lower bounds being composed will have to be trivial (i.e., will be a lower bound for 0-way marginals). However, a simple modification of the proof yields a tight lower bound for 2-way marginals that holds even for (α, β)-accuracy.

THEOREM 5.10. *For every $k, d \in \mathbb{N}$, and every sufficiently small (i.e., bounded by an absolute constant) $\alpha \geq 1/d^{.499}$, there is a constant $\beta > 0$ and an*

$$n = n(d, \alpha) = \tilde{\Omega}\big(\sqrt{d}/\alpha^2\big)$$

such that there exists a distribution on n-row databases $D \in (\{0,1\}^d)^n$ that is $(1/2, o(1/n))$-reidentifiable from (α, β)-accurate answers to the 2-way marginals $\mathcal{M}_{2,d}$.

Proof. We begin with the following two attacks:

1. By combining Theorems 3.5 and 3.4, there exists a distribution on databases $D' \in (\{0,1\}^{d/2})^{n_d}$ that is $(\gamma = 1/6, \xi = o(1/n_d n_\alpha))$-reidentifiable from

($6c\alpha' = 1/3, 2/c = 1/75$) accurate answers to the 1-way marginals $\mathcal{M}_{1,d/2}$ for $n_d = \tilde{\Omega}(\sqrt{d}/\log(n_d n_\alpha))$. n_α is set below.

2. By Theorem 5.4 (with $\alpha' = 1/2700$ and $k = 1$), there is a constant $\beta > 0$ such that for any $2700\alpha/\beta \geq 1/d^{.499}$ there exists a database $D \in (\{0,1\}^{d/2})^{n_\alpha}$ for $n_\alpha = \tilde{\Omega}(1/\alpha^2)$ that enables a (1/2700)-reconstruction attack from $(2700\alpha, 600\beta)$-accurate answers to $\mathcal{M}_{1,d/2}$.

Applying Theorem 4.3 (with parameter $c = 150$), we obtain the following: There exists a distribution on databases in $(\{0,1\}^d)^{n_d n_\alpha}$ that is $(1/3, o(1/n_d n_\alpha))$-reidentifiable from $(\alpha, 4\beta)$-accurate answers to $\mathcal{M}_{1,d/2} \wedge \mathcal{M}_{1,d/2} \subset \mathcal{M}_{2,d}$.

To complete the theorem, note that $\mathcal{M}_{1,d/2} \wedge \mathcal{M}_{1,d/2}$ contains exactly 1/4 of all the queries in $\mathcal{M}_{2,d}$, so (α, β)-accurate answers to $\mathcal{M}_{2,d}$ contain $(\alpha, 4\beta)$-accurate answers to the subset $\mathcal{M}_{1,d/2} \wedge \mathcal{M}_{1,d/2}$. So our lower bound for the subset $\mathcal{M}_{1,d/2} \wedge \mathcal{M}_{1,d/2}$ is sufficient to obtain the desired lower bound. Finally, note that

$$n = n_d n_\alpha = \tilde{\Omega}\big(\sqrt{d}/\alpha^2\big),$$

as desired. This completes the proof. ∎

5.2. Lower bounds for arbitrary queries. Using our composition theorem, we can also prove a nearly optimal sample-complexity lower bound as a function of $|\mathcal{Q}|, d$, and α and establish Theorem 1.3 in the introduction.

As was the case in the previous section, the main result of this section will follow from three lower bounds: the $\tilde{\Omega}(\sqrt{d})$ lower bound for 1-way marginals and the $\Omega(VC(\mathcal{Q}))$ bound that we have already discussed, a lower bound of $\Omega(1/\alpha^2)$ for worst-case queries, which is a simple variant of the seminal reconstruction attack of Dinur and Nissim [16], and related attacks such as [22, 30]. Although we already proved a $\Omega(1/\alpha^2)$ lower bound for the simpler family of k-way marginals in the previous section, the lower bound in this section will hold for a much wider range of α than what is known for k-way marginals (roughly $\alpha \geq 2^{-d}$ for arbitrary queries, whereas for k-way marginals we require $\alpha \geq 1/d^\ell$ for some constant ℓ).

5.2.1. The $\Omega(1/\alpha^2)$ lower bound for arbitrary queries. Roughly, the results of [16] can be interpreted in our framework as showing that there is an $\Omega(1/\alpha^2)$-row database that enables a 1/100-reconstruction attack from $(\alpha, 0)$-accurate answers to some family of queries $\mathcal{Q}$, but only when the vector to be reconstructed is boolean. That is, the attack reconstructs a bit vector accurately provided that every query in $\mathcal{Q}$ is answered correctly. Dwork, McSherry, and Talwar, [22] and Dwork and Yekhanin [30] generalized this attack to only require (α, β)-accuracy for some constant $\beta > 0$, and we will make use of this extension (although we do not require computational efficiency, which was a focus of those works). Finally, we need an extension to the case of fractional vectors $s \in [0,1]^n$, instead of boolean vectors $s \in \{0,1\}^n$.

The extension is fairly simple and the proof follows the same outline as the original reconstruction attack from [16]. We are given accurate answers to queries in $\mathcal{Q}$, which we interpret as approximate subset sums of the vector $s \in [0,1]^n$ that we wish to reconstruct. The reconstruction attack will output any vector t from a discretization $\{0, 1/m, \ldots, (m-1)/m, 1\}^n$ of the unit interval that is "consistent" with these subset sums. The main lemma we need is an "elimination lemma" that says that if $\|t - s\|_1$ is sufficiently large, then for a random subset $T \subseteq [n]$,

$$\frac{1}{n}\left|\sum_{i \in T}(t_i - s_i)\right| > 3\alpha$$

with suitably large constant probability. For $m = 1$ this lemma can be established via combinatorial arguments, whereas for the $m > 1$ case we establish it via the Berry–Esséen theorem. The lemma is used to argue that for every t that is sufficiently far from s, a large fraction of the subset-sum queries will witness the fact that t is far from s, and ensure that t is not chosen as the output.

First we state and prove the lemma that we just described, and then we will verify that it indeed leads to a reconstruction attack.

LEMMA 5.11. *Let $\kappa > 0$ be a constant, let $\alpha > 0$ be a parameter with $\alpha \leq \kappa^2/240$, and let $n = 1/576\kappa^2\alpha^2$. Then for every $r \in [-1,1]^n$ such that $\frac{1}{n}\sum_{i=1}^n |r_i| > \kappa$, and a randomly chosen $q \subseteq [n]$,*

$$\Pr_{q \subseteq [n]}\left[\left|\frac{1}{n}\sum_{i \in q} r_i\right| > 3\alpha\right] \geq \frac{3}{5}.$$

Proof of Lemma 5.11. Let r be as in the statement of the lemma. Define a random variable

$$Q_i = \begin{cases} r_i/2 & \text{if } i \in q, \\ -r_i/2 & \text{if } i \notin q. \end{cases}$$

By construction, we have

$$\frac{1}{n}\sum_{i \in q} r_i = \frac{1}{n}\sum_{i=1}^n \left(Q_i + \frac{r_i}{2}\right).$$

Thus,

$$\left|\frac{1}{n}\sum_{i \in q} r_i\right| \leq 3\alpha \Longleftrightarrow \sum_{i=1}^n Q_i \in \left[-3\alpha n - \frac{1}{2}\sum_{i=1}^n r_i, 3\alpha n - \frac{1}{2}\sum_{i=1}^n r_i\right].$$

The condition on the right-hand side says that $\sum_i Q_i$ is in some interval of width $6\alpha n$. Since the random variables Q_i are independent, as q is a randomly chosen subset, we will use the Berry–Esséen Theorem (Theorem 5.13) to conclude that this sum does not fall in any interval of this width too often. Establishing the next claim suffices to prove Lemma 5.11.

CLAIM 5.12. *For any interval $I \subseteq \mathbb{R}$ of width $6\alpha n$,*

$$\Pr\left[\sum_i Q_i \notin I\right] \geq \frac{3}{5}.$$

Proof of Claim 5.12. We use the Berry–Esséen theorem to prove this.

THEOREM 5.13 (Berry–Esséen theorem). *Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{E}[X_i] = 0$, $\sum_i \mathbb{E}\left[X_i^2\right] = \sigma^2$, and $\sum_i \mathbb{E}\left[|X_i|^3\right] = \gamma$. Let $X = (X_1 + \cdots + X_n)/\sigma$ and let Y be a normal random variable with mean 0 and variance 1. Then,*

$$\sup_{z, z' \in \mathbb{R}} |\Pr[X \in [z, z']] - \Pr[Y \in [z, z']]| \leq \frac{2\gamma}{\sigma^3}.$$

In order to apply Theorem 5.13 with $X_i = Q_i$, we need to analyze the moments of the random variables Q_i. The following bounds can be verified from the definition of Q_i and the assumption that $\|r\|_1 \geq \kappa n$.

1. $\mathbb{E}[Q_i] = 0$.
2. $\sigma^2 = \sum_i \mathbb{E}\left[Q_i^2\right] \geq \kappa^2 n/4$.
3. $\gamma = \sum_i \mathbb{E}\left[|Q_i|^3\right] \leq \frac{n}{8}$.

Thus, by Theorem 5.13 we have

$$\sup_{z,z' \in \mathbb{R}} \left| \Pr\left[\frac{Q_1 + \cdots + Q_n}{\sigma} \in [z, z']\right] - \Pr[Y \in [z, z']] \right| \leq \frac{2\gamma}{\sigma^3} \leq \frac{2}{\kappa^3 \sqrt{n}} \leq \frac{1}{5},$$

where the final inequality holds because $n = 1/576\kappa^2\alpha^2 \geq 100/\kappa^6$. It can be verified that for a standard normal random variable Y, and every interval $I \subset \mathbb{R}$ of width $1/2$, it holds that $\Pr[Y \notin I] \geq 4/5$. Thus, for every such interval I,

$$\Pr\left[\frac{Q_1 + \cdots + Q_n}{\sigma} \notin I\right] \geq \frac{4}{5} - \frac{1}{5}$$
$$\implies \Pr[Q_1 + \cdots + Q_n \notin \sigma I] \geq \frac{3}{5},$$

where σI is an interval of width $\sigma/2$. Thus we have obtained that $\sum_i Q_i$ falls outside of any interval of width $\sigma/2$ with probability at least $3/5$. In order to establish the claim, we simply observe that

$$\frac{\sigma}{2} \geq \frac{\kappa\sqrt{n}}{4} \geq 6\alpha n$$

when $n = 1/576\kappa^2\alpha^2$. Thus, the probability of falling outside an interval of width $6\alpha n$ is only larger than the probability of falling outside an interval of width $\sigma/2$. ∎

Establishing Claim 5.12 completes the proof of Lemma 5.11. ∎

THEOREM 5.14. *Let $\alpha' \in (0, 1]$ be a constant, let $\alpha > 0$ be a parameter with $\alpha \leq (\alpha')^2/960$, and let $n = 1/144(\alpha')^2\alpha^2$. For any data universe $\mathcal{X} = \{x_1, \ldots, x_n\}$ of size n, there is a set of counting queries $\mathcal{Q}$ over $\mathcal{X}$ of size at most $O(n \log(1/\alpha))$ such that the database $D = (x_1, \ldots, x_n)$ enables an α'-reconstruction attack from $(\alpha, 1/3)$-accurate answers to $\mathcal{Q}$.*

Proof. First we will give a reconstruction algorithm $\mathcal{B}$ for an arbitrary family of queries. We will then show that for a random set of queries $\mathcal{Q}$ of the appropriate size, the reconstruction attack succeeds for every $s \in [0, 1]^n$ with nonzero probability, which implies that there exists a set of queries satisfying the conclusion of the theorem. We will use the shorthand

$$\langle q, s \rangle = \frac{1}{n} \sum_{i=1}^{n} q(x_i) s_i$$

for vectors $s \in [0, 1]^n$.

In order to show that the reconstruction attack $\mathcal{B}$ from Figure 6 succeeds, we must show that $\frac{1}{n} \sum_{i=1}^{n} |t_i - s_i| \leq \alpha'$. Let $s \in [0, 1]^n$, and let

$$s' \in \{0, 1/m, \ldots, (m-1)/m, 1\}^n$$

be the vector obtained by rounding each entry of s to the nearest $1/m$. Then

$$\frac{1}{n} \sum_{i=1}^{n} |s'_i - s_i| \leq \frac{\alpha}{2} \leq \frac{\alpha'}{2},$$

Input: Queries $\mathcal{Q}$ and $(a_q)_{q\in\mathcal{Q}}$ that are $(\alpha, 1/3)$-accurate for s.
Let $m = \lceil \frac{1}{\alpha} \rceil$
Find any $t \in \{0, 1/m, \ldots, (m-1)/m, 1\}^n$ such that

$$\Pr_{q \leftarrow_R \mathcal{Q}} [|\langle q, t\rangle - a_q| < 2\alpha] > \frac{5}{6}.$$

Output: t.

FIG. 6. *The reconstruction adversary* $\mathcal{B}$.

so it is enough to show that the reconstruction attack outputs a vector close to s'. Observe that the vector s' itself satisfies

$$|\langle q, s'\rangle - a_q| \le |\langle q, s\rangle - a_q| + |\langle q, s' - s\rangle| \le 2\alpha$$

for any subset sum query q, so the reconstruction attack always finds some vector t. To show that the reconstruction is successful, fix any $t \in \{0, 1/m, \ldots, (m-1)/m, 1\}^n$ such that $\frac{1}{n}\sum_{i=1}^n |t_i - s'_i| > \frac{\alpha'}{2}$. If we write $r = s' - t \in \{-1, \ldots, -1/m, 0, 1/m, \ldots, 1\}^n$, then $\frac{1}{n}\sum_{i=1}^n |r_i| > \frac{\alpha'}{2}$ and $\langle q, r\rangle = \langle q, t\rangle - \langle q, s'\rangle$. In order to show that no t that is far from s' can be output by $\mathcal{B}$, we will show that for any $r \in \{-1, \ldots, -1/m, 0, 1/m, \ldots, 1\}$ with $\frac{1}{n}\sum_{i=1}^n |r| > \frac{\alpha'}{2}$,

$$\Pr_{q \leftarrow_R \mathcal{Q}} [|\langle q, r\rangle| > 3\alpha] \ge \frac{1}{2}.$$

To prove this, we first observe by Lemma 5.11 (setting $\kappa = \frac{1}{2}\alpha'$) that for a randomly chosen query q defined on $\mathcal{X}$,

$$\Pr_{q} [|\langle q, r\rangle| > 3\alpha] \ge \frac{3}{5}.$$

The lemma applies because $\langle q, r\rangle = \frac{1}{n}\sum_{i=1}^n q(x_i) r_i$ is a random subset-sum of the entries of r.

Next, we apply a concentration bound to show that if the set $\mathcal{Q}$ of queries is a sufficiently large random set, then for every vector r the fraction of queries for which $|\langle q, r\rangle|$ is large will be close to the expected number, which we have just established is at least $3|\mathcal{Q}|/5$. We use the following version of the Chernoff bound.

THEOREM 5.15 (Chernoff bound). *Let $X_1, \ldots, X_N$ be a sequence of independent random variables taking values in $[0, 1]$. If $X = \sum_{i=1}^N X_i$ and $\mu = \mathbb{E}[X]$, then*

$$\Pr[X \le \mu - \varepsilon] \le e^{-2\varepsilon^2/N}.$$

Consider a set of randomly chosen queries $\mathcal{Q}$. By the above, we have that for every $r \in \{-1, \ldots, -1/m, 0, 1/m, \ldots, 1\}^n$ such that $\frac{1}{n}\sum_{i=1}^n |r| > \frac{\alpha'}{2}$,

$$\mathop{\mathbb{E}}_{\mathcal{Q}} [|\{q \in \mathcal{Q} \mid |\langle q, r\rangle| > 3\alpha\}|] \ge \frac{3|\mathcal{Q}|}{5}.$$

Since the queries are chosen independently, by the Chernoff bound we have

$$\Pr_{\mathcal{Q}}\left[|\{q \in \mathcal{Q} \mid |\langle q, r\rangle| > 3\alpha\}| \leq \frac{|\mathcal{Q}|}{2}\right] \leq e^{-|\mathcal{Q}|/50}.$$

Thus, we can choose $|\mathcal{Q}| = O(n \log m)$ to obtain

$$\Pr_{\mathcal{Q}}\left[\begin{array}{c}\exists r \in \{-1, \ldots, -1/m, 0, 1/m, \ldots, 1\}^n, \\ \frac{1}{n}\sum_{i=1}^n |r_i| > \frac{\alpha'}{2}, \quad |\{q \in \mathcal{Q} \mid |\langle q, y\rangle| > 3\alpha\}| \leq \frac{|\mathcal{Q}|}{2}\end{array}\right] < (2m+1)^n e^{-|\mathcal{Q}|/50} \leq \frac{1}{2}.$$

Thus, we have established that there exists a family of queries $\mathcal{Q}$ such that for every s, t such that $\frac{1}{n}\sum_{i=1}^n |t_i - s_i| > \alpha'$,

$$\Pr_{q \leftarrow_{\mathrm{R}} \mathcal{Q}}[|\langle q, s\rangle - \langle q, t\rangle| > 3\alpha] \geq \frac{1}{2}.$$

Moreover, by $(\alpha, 1/3)$-accuracy, we have

$$\Pr_{q \leftarrow_{\mathrm{R}} \mathcal{Q}}[|a_q - \langle q, s\rangle| > \alpha] \leq \frac{1}{3}.$$

Applying a triangle inequality, we can conclude

$$\Pr_{q \leftarrow_{\mathrm{R}} \mathcal{Q}}[|a_q - \langle q, t\rangle| > 2\alpha] \geq \frac{1}{2} - \frac{1}{3} \geq \frac{1}{6},$$

which implies that t cannot be the output of $\mathcal{B}$. This completes the proof. ∎

5.2.2. Putting together the lower bound. Now we show how to combine the various attacks to prove Theorem 1.2 in the introduction. We obtain our lower bound by applying two rounds of composition. In the first round, we compose the reconstruction attack described above with the reidentifiable distribution for 1-way marginals. We then take the resulting reidentifiable distribution and apply a second round of composition using the reconstruction attack for query families of high VC-dimension.

Just like our lower bound for k-way marginal queries, we remark that it is necessary to apply the two rounds of composition in this order. See section 5.1.3 for a discussion of this issue.

THEOREM 5.16. *For all $d \in \mathbb{N}$, all sufficiently small (i.e., bounded by an absolute constant) $\alpha > 2^{-d/6}$, and all $h \leq 2^{d/3}$, there exists a family of queries $\mathcal{Q}$ of size $O(hd \log(1/\alpha)/\alpha^2)$ and an*

$$n = n(h, d, \alpha) = \tilde{\Omega}\left(\frac{\sqrt{d} \log h}{\alpha^2}\right)$$

such that there exists a distribution on n-row databases $D \in (\{0,1\}^d)^n$ that is $(1/2, o(1/n))$-reidentifiable from $(\alpha, 0)$-accurate answers to $\mathcal{Q}$.

Proof. We begin with the following two attacks:

1. By Theorems 3.5 and 3.4, there exists a distribution on databases in $(\{0,1\}^{d/3})^m$ that is $(1/6, o(1/m\ell \log h))$-reidentifiable from $(1/3, 1/75)$ accurate answers to $\mathcal{M}_{1,d/3}$ for $m = \tilde{\Omega}(\sqrt{d}/\log(m\ell \log h))$. Here ℓ and h are parameters we set below.

2. For some $\ell = \Omega(1/\alpha^2)$, by Theorem 5.14, there exists a database $D \in (\{0,1\}^{d/3})^{\ell}$ that enables an α'-reconstruction attack from $(6c'\alpha, 1/3)$-accurate answers to some $\mathcal{Q}_{rec}$ of size $O((\log(1/\alpha))/\ \alpha^2)$. Here α' is a constant with $6c\alpha' = 1/3$ for a composition parameter c set below, and c' is a constant composition parameter set when we apply the second round of composition.

Applying Theorem 4.3 (with parameter $c = 150$), we obtain item 1′ below. We then bring in another reconstruction attack for the composition theorem.

1′. There exists a probability distribution on databases in $(\{0,1\}^{2d/3})^{m\ell}$ that is $(1/3, o(1/m\ell \log h))$-reidentifiable from $(6c'\alpha, 1/450)$-accurate answers to $\mathcal{Q}_{rec} \wedge \mathcal{M}_{1,d/3}$ (by applying Theorem 4.3 to 1 and 2 above).

2′. By Lemma 5.3, there exists a database $D \in (\{0,1\}^{d/3})^{\log h}$ that enables a (4α)-reconstruction attack from $(\alpha, 0)$-accurate answers to some $\mathcal{Q}_{vc}$ of size h. (In particular, the family of queries can be all $(\log h)$-way marginals on the first $\log h$ bits of the data universe items.)

We can then apply Theorem 4.3 to 1′ and 2′ (with parameter $c' = 900$). Thereby we obtain a distribution $\mathcal{D}$ on databases $D \in (\{0,1\}^{d/3} \times \{0,1\}^{d/3} \times \{0,1\}^{d/3})^{m\ell \log h}$ that is $(1/2, \xi)$-reidentifiable from $(\alpha, 0)$-accurate answers to $\mathcal{Q} = \mathcal{Q}_{vc} \wedge \mathcal{Q}_{rec} \wedge \mathcal{M}_{1,d/3}$.

To complete the theorem we first set

$$n = m\ell \log h = \tilde{\Omega}(\sqrt{d} \log h/\alpha^2),$$

and then observe that

$$|\mathcal{Q}_{vc} \wedge \mathcal{Q}_{rec} \wedge \mathcal{M}_{1,d/3}| = h \cdot O(\ell \log(1/\alpha)/\alpha^2) \cdot d/3 = O(hd \log(1/\alpha)/\alpha^2).$$

This completes the proof. ∎

Again, Theorem 5.16 has a corresponding statement in terms of generalized fingerprinting codes.

THEOREM 5.17. *For all $d \in \mathbb{N}$, all sufficiently small (i.e., bounded by an absolute constant) $\alpha > 2^{-d/6}$, and all $h \leq 2^{d/3}$, there exists a family of queries $\mathcal{Q}$ of size $O(hd \log(1/\alpha)/\alpha^2)$ and an*

$$n = n(h, d, \alpha) = \tilde{\Omega}\left(\frac{\sqrt{d} \log h}{\alpha^2}\right)$$

such that there exists an $(n, \mathcal{Q})$-generalized fingerprinting code with security $(1/2, o(1/n))$ for $(\alpha, 0)$-accuracy.

6. Constructing error-robust fingerprinting codes. In this section, we show how to construct fingerprinting codes that are robust to a constant fraction of errors, which will establish Theorem 3.4. Our codes are based on the fingerprinting code of Tardos [49], which has a nearly optimal number of users, but is not robust to any constant fraction of errors. The number of users in our code is only a constant factor smaller than that of Tardos, and thus our codes also have a nearly optimal number of users.

To motivate our approach, it is useful to see why the Tardos code (and all other fingerprinting codes we are aware of) are not robust to a constant fraction of errors. The reason is that the the only way to introduce an error is to put a 0 in a column containing only 1's or vice versa (recall that the set of codewords, $C \in \{0,1\}^{n \times d}$, can be viewed as an $n \times d$ matrix). We call such columns "marked columns." Thus, if

the adversary is allowed to introduce $\geq m$ errors, where m is the number of marked columns, then he can simply ignore the codewords and output either the all-0 or all-1 codeword, which cannot be traced. Thus, in order to tolerate a β fraction of errors, it is necessary that $m \geq \beta d$, where d is the length of the codeword, and this is not satisfied by any construction we know of (when $\beta > 0$ is a constant). However, Tardos' construction can be shown to remain secure if the adversary is allowed to introduce βm errors, rather than βd errors, for some constant $\beta > 0$. We demonstrate this formally in section 6.2. In addition, we show how to take a fingerprinting code that tolerates βm errors and modify it so that it can tolerate about $\beta d/3$ errors. This reduction is formalized in section 6.1. Combining these two results will give us a robust fingerprinting code.

We remark that prior work [11, 10] has shown how to construct fingerprinting codes satisfying a weaker robustness property. Specifically, their codes allow the adversary to introduce a special "?" symbol in a large fraction of coordinates, but still require that any coordinate that is not a "?" satisfies the feasibility constraint.

Before proceeding with the construction and analysis, we restate some terminology and notation from section 3. Recall that a fingerprinting code is a pair of algorithms (*Gen*, *Trace*), where *Gen* specifies a distribution over codebooks $C \in \{0,1\}^{n \times d}$ consisting of n codewords $(c_1, \ldots, c_n)$, and $\mathit{Trace}(C, c')$ either outputs the identity $i \in [n]$ of an accused user or outputs $\perp$. Recall that *Gen* and *Trace* share a common state. For a coalition $S \subseteq [n]$, we write $C_S \in \{0,1\}^{|S| \times d}$ to denote the subset of codewords belonging to users in S.

Every codebook C, coalition S, and robustness parameter $\beta \in [0,1]$ defines a feasible set of combined codewords,

$$F_\beta(C_S) = \left\{ c' \in \{0,1\}^d \mid \Pr_{j \leftarrow_R [d]} \left[\exists i \in S, c'_j = c_{ij} \right] \geq 1 - \beta \right\}.$$

We now recall the definition of an error-robust fingerprinting code from section 3.1.

DEFINITION 6.1 (error-robust fingerprinting codes (restatement of Definition 3.3)). *For any $n, d \in \mathbb{N}$, $\xi, \beta \in [0,1]$, a pair of algorithms (Gen, Trace) is an (n,d)-fingerprinting code with security ξ robust to a β fraction of errors if Gen outputs a codebook $C \in \{0,1\}^{n \times d}$ and for every (possibly randomized) adversary $\mathcal{A}_{FP}$, and every coalition $S \subseteq [n]$, if we set $c' \leftarrow_R \mathcal{A}_{FP}(C_S)$, then*

1. $\Pr\left[(\mathit{Trace}(C, c') = \perp) \wedge (c' \in F_\beta(C_S))\right] \leq \xi$,
2. $\Pr\left[\mathit{Trace}(C, c') \in [n] \setminus S\right] \leq \xi$,

where the probability is taken over the coins of Gen, Trace, and $\mathcal{A}_{FP}$. The algorithms Gen and Trace may share a common state.

The main result of this section is a construction of fingerprinting codes satisfying Definition 6.1

THEOREM 6.2 (restated from section 3.1). *For every $n \in \mathbb{N}$ and $\xi \in (0,1]$, there exists an (n,d)-fingerprinting code with security ξ robust to a $1/75$ fraction of errors for*

$$d = d(n, \xi) = \tilde{O}(n^2 \log(1/\xi)).$$

Equivalently, for every $d \in \mathbb{N}$, and $\xi \in (0,1]$, there exists an (n,d)-fingerprinting code with security ξ robust to a $1/75$ fraction of errors for

$$n = n(d, \xi) = \tilde{\Omega}(\sqrt{d/\log(1/\xi)}).$$

We remark that we have made no attempt to optimize the fraction of errors to which our code is robust. We leave it as an interesting open problem to construct a robust fingerprinting code for a nearly optimal number of users that is robust to a fraction of errors arbitrarily close to 1/2.

6.1. From weak error robustness to strong error robustness. A key step in our construction is a reduction from constructing error-robust fingerprinting codes to constructing a weaker object, which we call a weakly robust fingerprinting code. The difference between a weakly robust fingerprinting code and an error-robust fingerprinting code of the previous section is that we now demand that only a β fraction of the *marked* positions can have errors, rather than a β fraction of all positions.

In order to formally define weakly robust fingerprinting codes, we introduce some terminology. If $C \in \{0,1\}^{n \times d}$ is a codebook, then for $b \in \{0,1\}$, we say that position $j \in [d]$ is *b-marked in C* if $c_{ij} = b$ for every $i \in [n]$. That is, j is b-marked if every user has the symbol b in the jth position of their codeword. The set $F_\beta(C)$ consists of all codewords c' such that for a $1-\beta$ fraction of positions j, either j is not marked, or j is b-marked and $c'_j = b$. Notice that this constraint is vacuous if fewer than a β fraction of positions are marked.

For a weakly robust fingerprinting code, we will define a more constrained feasible set. Intuitively, a codeword c' is feasible if for a $1-\beta$ fraction of positions that are marked, c'_j is set appropriately. Note that this condition is meaningful even when the fraction of marked positions is much smaller than β. More formally, we define

$$WF_\beta(C_S) = \left\{ c' \in \{0,1\}^d \mid \Pr_{j \leftarrow_R [d]} \left[c'_j = b \mid j \text{ is } b\text{-marked in } C_S \text{ for some } b \in \{0,1\} \right] \geq 1 - \beta \right\}.$$

DEFINITION 6.3 (weakly robust fingerprinting codes). *For any $n, d \in \mathbb{N}$ and $\xi, \beta \in [0,1]$, a pair of algorithms $(Gen, Trace)$ is an (n,d)-*weakly robust fingerprinting code with security ξ weakly robust to a β fraction of errors *if $(Gen, Trace)$ satisfy the conditions of a robust fingerprinting code (for the same parameters) with WF_β in place of F_β.*

The next theorem states that if we have an (n,d)-fingerprinting code that is weakly robust to a β fraction of errors and satisfies a mild technical condition, then we obtain an $(n, O(d))$-fingerprinting code that is robust to an $\Omega(\beta)$ fraction of errors with a similar level of security.

LEMMA 6.4. *For any $n, d \in \mathbb{N}$, $\xi, \beta \in [0,1]$, and $m \in \mathbb{N}$, suppose there is a pair of algorithms $(Gen, Trace)$ which*

1. *are an (n,d)-fingerprinting code with security ξ weakly robust to a β fraction of errors, and*
2. *with probability at least $1-\xi$ over $C \leftarrow_R Gen$, produce C that has at least m 0-marked columns and m 1-marked columns.*

Then there is a pair of algorithms $(Gen', Trace')$ that are an (n, d')-fingerprinting code with security ξ' robust to a $\beta/3$ fraction of errors, where

$$d' = 5d \quad \text{and} \quad \xi' = \xi + 2\exp\left(-\Omega(\beta m^2/d)\right).$$

Proof. The reduction is given in Figure 7. Recall that Gen' and $Trace'$ may share a state, so π and the shared state of Gen and $Trace$ is known to $Trace'$.

Fix a coalition $S \subseteq [n]$. Let $\mathcal{A}'_{FP}$ be an adversary. Sample $C' \leftarrow_R Gen'$ and let $c' = \mathcal{A}'_{FP}(C')$. We will show that the reduction is successful by proving that if $c' \in$

Gen′:
Choose $C \leftarrow_{\mathrm{R}} Gen$, $C \in \{0,1\}^{n \times d}$
Append $2d$ 0-marked columns and $2d$ 1-marked columns to C
Apply a random permutation π to the columns of the augmented codebook
Let the new codebook be $C' \in \{0,1\}^{n \times d'}$ for $d' = 5d$
(We refer to the columns from C as *real* and to the additional columns as *fake*)
Output C'

Trace′(C', c'):
Obtain C by applying π^{-1} to the columns of C' and removing the fake columns
Obtain c by applying π^{-1} to c' and removing the symbols corresponding to fake columns
Output $i \leftarrow_{\mathrm{R}} Trace(C, c)$

FIG. 7. *Reducing robustness to weak robustness.*

$F_{\beta/3}(C')$, then the modified string $c \in WF_\beta(C)$ with probability $1-\exp(-\Omega(\beta m^2/d))$. The reason is that an adversary who is given (a subset of the rows of) C' cannot distinguish real columns that are marked from fake columns. Therefore, the fraction of errors in the real marked columns should be close to the fraction of errors that are either real and marked or fake. Since the total fraction of errors in the entire codebook is at most $\beta/3$, we know that the fraction of errors in real marked columns is not much larger than $\beta/3$. Thus the fraction of errors in the real marked columns will be at most β with high probability. We formalize this argument in the following claim.

CLAIM 6.5.

$$\Pr_{\pi}\left[(c' \in F_{\beta/3}(C')) \wedge (c \in WF_\beta(C))\right] \leq 2\exp(-\Omega(\beta m^2/d)).$$

Proof of Claim 6.5. Our analysis will handle 0-marked and 1-marked columns separately. Assume that $c' \in F_{\beta/3}(C')$ and that the adversary has introduced $k \leq \beta d'/3$ errors to 0-marked columns. Let $m_0 \geq m$ be the number of 0-marked columns. Let R_0 be a random variable denoting the number of columns that are both real and 0-marked in which the adversary introduces an error. Since real 0-marked columns are indistinguishable from fake 0-marked columns, R_0 has a hypergeometric distribution on k draws from a population of size $N = m_0 + 2d$ with m_0 successes. In other words, we can think of an urn with N balls, m_0 of which are labeled "real" and $2d$ of which are labeled "fake." We draw k balls without replacement, and R_0 is the number that are labeled real. This distribution has $\mathbb{E}[R_0] = km_0/N = km_0/(m_0 + 2d)$. Moreover, as shown in [17, section 7.1]), it satisfies the concentration inequality

$$\Pr[|R_0 - \mathbb{E}[R_0]| > t] \leq \exp\left(\frac{-2(N-1)t^2}{(N-k)(k-1)}\right) \leq \exp(-\Omega(t^2/k))$$

since $k \leq 5N/6$. Thus

$$\begin{aligned}\Pr[R_0 > \beta m_0] &\le \Pr[|R_0 - \mathbb{E}[R_0]| > \beta m_0 - \mathbb{E}[R_0]] \\ &\le \exp\left(-\Omega\left(\frac{(\beta m_0 - k m_0/N)^2}{k^2}\right)\right) \\ &\le \exp\left(-\Omega\left(\frac{(\beta m_0)^2(1-d'/6d)^2}{(\beta d'/3)^2}\right)\right) \\ &\le \exp\left(-\Omega\left(\frac{\beta m_0^2}{d}\right)\right)\end{aligned}$$

for any choice of k. An identical argument bounds the probability that the number of errors in real 1-marked columns is more than βm_1. Therefore, the probability that more than a β fraction of marked columns have errors is at most $2\exp(-\Omega(\beta m^2/d))$. ∎

Now define an adversary $\mathcal{A}_{FP}$ that takes C_S as input, simulates Gen' by appending marked columns to C_S and applying a random permutation π, and then applies $\mathcal{A}'_{FP}$ to the resulting codebook C'_S. Then it takes $\mathcal{A}'_{FP}(C'_S)$, applies π^{-1}, removes the fake columns, and outputs the result. Notice that *Trace*′ applies *Trace* to a codebook and codeword generated by exactly the same procedure. If we assume that $\mathcal{A}'_{FP}(C'_S)$ is feasible with parameter $\beta/3$, then by the analysis above, with probability at least $1 - \xi - \exp(-\Omega(\beta m^2/d))$, $\mathcal{A}_{FP}(C_S)$ is weakly feasible with parameter β. Thus,

$$\begin{aligned}&\Pr_{C' \leftarrow_R Gen'}\left[(Trace'(C', \mathcal{A}'_{FP}(C_S)) = \perp) \wedge (\mathcal{A}'_{FP}(C_S) \in F_{\beta/3}(C_S))\right] \\ &\quad \le \Pr_{C \leftarrow_R Gen}\left[(Trace(C, \mathcal{A}_{FP}(C_S)) = \perp \wedge (\mathcal{A}_{FP}(C_S) \in WF_\beta(C_S))\right] + 2e^{-\Omega(\beta m^2/d)} \\ &\quad \le \xi + 2\exp(-\Omega(\beta m^2/d)),\end{aligned}$$

where the first inequality is by Claim 6.5 and the second inequality is by ξ-security of *Trace*.

Since *Trace* does not accuse a user outside of S (except with probability at most ξ) regardless of whether or not that adversary's codeword is feasible, it is immediate that *Trace*′ also does not accuse a user outside of S (except with probability at most ξ). ∎

6.2. Weak robustness of Tardos' fingerprinting code. In this section we show that Tardos' fingerprinting code is weakly robust to a β fraction of errors for $\beta \ge 1/25$. Specifically we prove the following.

LEMMA 6.6. *For every $n \in \mathbb{N}$ and $\xi \in (0,1]$, there exists an (n,d)-fingerprinting code with security ξ weakly robust to a $1/25$ fraction of errors for*

$$d = d(n,\xi) = \tilde{O}(n^2 \log(1/\xi)).$$

Equivalently, for every $d \in \mathbb{N}$, and $\xi \in (0,1]$, there exists an (n,d)-fingerprinting code with security ξ weakly robust to a $1/25$ fraction of errors for

$$n = n(d,\xi) = \tilde{\Omega}(\sqrt{d/\log(1/\xi)}).$$

Tardos' fingerprinting code is described in Figure 8. Note that the shared state of *Gen* and *Trace* will include $p_1, \ldots, p_d$.

Tardos' proof that no user is falsely accused (except with probability ξ) holds for every adversary, regardless of whether or not the adversary's output is feasible, therefore it holds without modification even when we allow the adversary to introduce errors. So we will state the following lemma from [49, section 3] without proof.

Gen:
Let $d = 100n^2 \log(n/\xi)$ be the length of the code.
Let $t = 1/300n$ be a parameter and let t' be such that $\sin^2 t' = t$.
For $j = 1, \ldots, d$:
Choose $r_j \leftarrow_R [t', \pi/2 - t']$ and let $p_j = \sin^2 r_j$. Note that $p_j \in [t, 1-t]$.
For each $i = 1, \ldots, n$, set $C_{ij} = 1$ with probability p_j, independently.
Output C.

Trace(C, c'):
Let $Z = 20n \log(n/\xi)$ be a parameter.
For each $j = 1, \ldots, d$, let $q_j = \sqrt{(1-p_j)/p_j}$.
For each $j = 1, \ldots, d$, and each $i = 1, \ldots, n$, let

$$U_{ij} = \begin{cases} q_j & \text{if } C_{ij} = 1, \\ -1/q_j & \text{if } C_{ij} = 0. \end{cases}$$

For each $i = 1, \ldots, n$:
Let

$$S_i(c') = \sum_{j=1}^{d} c'_j U_{ij}.$$

If $S_i(c') \geq Z/2$, output i.
If $S_i(c') < Z/2$ for every $i = 1, \ldots, n$, output $\perp$.

FIG. 8. *The Tardos fingerprinting code* [49].

LEMMA 6.7 (restated from [49]). *Let* $(Gen, Trace)$ *be the fingerprinting code defined in Algorithm* 8. *Then for every adversary* $\mathcal{A}_{FP}$, *and every* $S \subseteq [n]$,

$$\Pr\left[Trace(C, \mathcal{A}_{FP}(C_S)) \in [n] \setminus S\right] \leq \xi,$$

where the probability is taken over the choice of $C \leftarrow_R Gen$ *and the coins of* $\mathcal{A}_{FP}$.

Most of the remainder of this section is devoted to proving that any adversary who introduces errors into at most a 1/25 fraction of the marked columns can be traced successfully.

LEMMA 6.8. *Let* $(Gen, Trace)$ *be the fingerprinting code defined in Algorithm 8. Then for every adversary* $\mathcal{A}_{FP}$, *and every* $S \subseteq [n]$,

$$\Pr\left[(Trace(C, \mathcal{A}_{FP}(C_S)) = \perp) \wedge (\mathcal{A}_{FP}(C_S) \in WF_{1/25}(C_S))\right] \leq \xi,$$

where the probability is taken over the choice of $C \leftarrow_R Gen$ *and the coins of* $\mathcal{A}_{FP}$.

Before giving the proof, we briefly give a high-level roadmap. Recall that in the construction there is a "score" function $S_i(c')$ that is computed for each user, and *Trace* will output some user whose score is larger than the threshold $Z/2$, if such a user exists. Tardos shows that the sum of the scores over all users is at least $nZ/2$, which demonstrates that there exists a user whose score is above the threshold. His argument works by balancing two contributions to the score: (1) the contribution

from 1-marked columns j, which will always be positive due to the fact that $c'_j = 1$, and (2) the potentially negative contribution from columns that are not 1-marked. Conceptually, he shows that the contribution from the 1-marked columns is larger in expectation than the negative contribution from the other columns, so the expected score is significantly above the threshold. He then applies a Chernoff-type bound to show that the score will be above the threshold with high probability. When the adversary is allowed to introduce errors so that there may be some 1-marked columns j such that $c'_j = 0$, these errors will contribute negatively to the score. The new ingredient in our argument is essentially to bound the negative contribution from these errors. We are able to get a sufficiently good bound to tolerate errors in 1/25 of the coordinates. We expect that a tighter analysis and more careful tuning of the parameters can improve the fraction of errors that can be tolerated.

Proof of Lemma 6.8. We will write $S = [n]$. Doing so is without loss of generality as users outside of S are irrelevant. We will use $\beta = 1/25$ to denote the allowable fraction of errors. Fix an adversary $\mathcal{B}$. Sample $C \leftarrow_{\mathrm{R}} \mathit{Gen}$ and let $c' = \mathcal{B}(C)$. Assume $c' \in \mathit{WF}_\beta(C)$. In order to prove that some user is traced, we will bound the quantity

$$S(c') = \sum_{i=1}^{n} S_i(c') = \sum_{j=1}^{d} c'_j \left(x_j q_j - \frac{n - x_j}{q_j} \right),$$

where $x_j = \sum_{i=1}^{n} C_{ij}$ is defined to be the number of codewords c_i such that $c_{ij} = 1$. Our goal is to show that this quantity is at least $nZ/2$ with high probability. If we can do so, then there must exist a user $i \in [n]$ such that $S_i(c') \geq Z/2$, in which case $\mathit{Trace}(C, c') \neq \perp$.

We may decompose an output c' of $\mathcal{B}(C)$ into a the sum of a codeword $\tilde{c} \in \mathit{WF}_0(C)$ with no errors, and a string $\overline{c}$ that captures errors introduced into at most a β fraction of the marked coordinates. Each codeword c has a unique such decomposition if we assume the following constraints on $\overline{c}$.

1. If j is unmarked, then $\overline{c}_j = 0$.
2. If j is 0-marked, then $\overline{c}_j \in \{0, 1\}$.
3. If j is 1-marked, then $\overline{c}_j \in \{-1, 0\}$.
4. The number of nonzero coordinates of $\overline{c}$ is at most βm, where m is the number of marked columns of c.

We call a $\overline{c}$ satisfying the above constraints *valid*. By the linearity of $S(\cdot)$, we can write

$$S(c') = S(\tilde{c}) + S(\overline{c}).$$

Tardos' analysis of the error-free case proves that $S(\tilde{c})$ is large. In our language, he proves the following.

Claim 6.9 (restated from [49]). *For every adversary $\mathcal{B}$, if $C \leftarrow_R \mathit{Gen}$, $c' \leftarrow_R \mathcal{B}(C)$, and $c' = \tilde{c} + \overline{c}$ as above, then*

$$\Pr\left[(S(\tilde{c}) < nZ) \wedge (\tilde{c} \in \mathit{WF}_0(C))\right] \leq \xi^{\sqrt{n}/4}.$$

Although $S(\overline{c})$ will be negative, and thus $S(c') \leq S(\tilde{c})$, we will show that $S(\overline{c})$ is not too negative. That is, introducing errors into a β fraction of the marked columns in c' cannot reduce $S(c')$ by too much.

We will now establish the following claim.

Claim 6.10. *For any adversary $\mathcal{B}$, if $C \leftarrow_R \mathit{Gen}$, $c' \leftarrow_R \mathcal{B}(C)$, and $c' = \tilde{c} + \overline{c}$ as above, then*

$$\Pr\left[(S(\overline{c}) < -nZ/2) \wedge (\overline{c} \textit{ is valid})\right] \leq \xi/2.$$

Proof of Claim 6.10. We start by making an observation about the distribution of $S(\overline{c}) = S(\overline{c})|_{C,\overline{c}}$, which denotes $S(\overline{c})$ when we condition on a fixed choice of a codebook C and a valid choice of $\overline{c}$. Because the nonzero coordinates of $\overline{c}$ are only in marked columns of C (those in which $x_j = 0$ or $x_j = n$), the distribution of

$$S(\overline{c})|_{C,\overline{c}} = \sum_{j=1}^{d} \overline{c}_j \left(x_j q_j - \frac{n - x_j}{q_j} \right)$$

depends only on the number of nonzero coordinates of $\overline{c}$, and not on their location. To see that this is the case, consider a 0-marked coordinate j on which $\overline{c}_j = 1$. The contribution of j to $S(\overline{c})$ is exactly $-n/q_j$. Similarly, for a 1-marked coordinate j on which $\overline{c}_j = -1$, the contribution of j to $S(\overline{c})$ is exactly $-nq_j$. Thus we can write

$$\begin{aligned} S(\overline{c}) &= \sum_{j=1}^{d} \overline{c}_j \left(x_j q_j - \frac{n - x_j}{q_j} \right) \\ &= -\left(\sum_{j \in [d] : j \text{ is 0-marked and } \overline{c}_j = 1} n/q_j + \sum_{j \in [d] : j \text{ is 1-marked and } \overline{c}_j = -1} nq_j \right). \end{aligned} \tag{9}$$

Each term in the first sum (resp., second sum) is a random variable that depends only on the distribution of q_j conditioned on the the jth column being 0-marked (resp., 1-marked). Recall that q_j is determined by p_j. Moreover, conditioned on a fixed C, the p_j's are independent. To see this, let C_j denote the jth column of the codebook C. Recall that each column C_j is generated independently using p_j, and the p_j's themselves are chosen independently. Letting f_X denote the density function of a random variable X, this means that the joint density

(Bayes' rule)

$$\begin{aligned} f_{p_1,\ldots,p_d}(x_1, \ldots, x_d \mid C_1, \ldots, C_d) &= \frac{\Pr[C_1, \ldots, C_d \mid x_1, \ldots, x_d] f_{p_1,\ldots,p_d}(x_1, \ldots, x_d)}{\Pr[C_1, \ldots, C_d]} \\ &= \frac{\Pr[C_1 \mid x_1] f_{p_1}(x_1)}{\Pr[C_1]} \cdot \ldots \cdot \frac{\Pr[C_d \mid x_d] f_{p_d}(x_d)}{\Pr[C_d]} \\ &= f_{p_1}(x_1 \mid C_1) \cdot \cdots \cdot f_{p_d}(x_d \mid C_d). \end{aligned}$$

This shows that the conditional random variables $p_j|_{C_j}$ are independent. Moreover, since $\overline{c}$ only depends on the codebook C and coins of the adversary $\mathcal{B}$, the p_j's are still independent when we also condition on $\overline{c}$. In fact, the following holds.

Claim 6.11. *Conditioned on any fixed choice of C and $\overline{c}$, the following distributions are all identical, independent, and nonnegative:* (1) *$(n/q_j \mid j$ is 0-marked) for $j \in [d]$, and* (2) *$(nq_j \mid j$ is 1-marked).*

Proof of Claim 6.11. By the discussion above, we know that these random variables are independent. To see that they are identicially distributed, note that the distribution p_j used to generate the jth column of C is symmetric about $1/2$. Therefore, the probability that column j is 0-marked when its entries are sampled according to p_j is the same as the probability that j is 1-marked when its entries are sampled according to $1 - p_j$. Applying Bayes' rule, again using the fact that p_j and $1 - p_j$ have the same distribution, we see that the random variables $(p_j \mid j$ is 0-marked) and $(1 - p_j \mid j$ is 1-marked) are identically distributed. The claim follows since $q_j = \sqrt{(1 - p_j)/p_j}$. ∎

In light of this fact, we can see that the conditional random variable $S(\overline{c})|_{C,\overline{c}}$ is a sum of i.i.d. random variables and the number of these variables in the sum is exactly the number of marked columns j on which $\overline{c}_j$ is nonzero. For any $t \in \mathbb{N}$ and any non-negative random variable Q, the sum of $t+1$ independent draws from Q stochastically dominates[12] the sum of t independent draws from Q. Recall that $S(\overline{c})$ will be negative and we want its magnitude not to be too large. Equivalently, we want the positive sum in (9) not to be too large. Therefore, the "worst-case" for the sum (9) is when $\overline{c}$ has the largest possible number of nonzero coordinates. Recall that the number of nonzero coordinates of $\overline{c}$ is exactly the number of errors introduced by the adversary. Thus, the worst-case adversary $\mathcal{B}^*$ is the one that chooses a random set of exactly βm marked columns and for the chosen columns j that are 0-marked, sets $\overline{c}_j = 1$ and for those that are 1-marked, sets $\overline{c}_j = -1$. In summary, it suffices to consider only the single adversary $\mathcal{B}^*(C)$ that constructs a feasible $\tilde{c}$ and introduces errors in a random set of βm of the marked coordinates in C.

Now we proceed to analyzing $\mathcal{B}^*$. We follow Tardos' approach to analyzing S. A key step in his analysis is to show that the optimal adversary (for the error-free case) chooses the jth coordinate of c' based only on the jth column of C. In our case, the optimal adversary $\mathcal{B}^*$ introduces errors in a random set of exactly βm marked columns, which does not satisfy this independence condition. So instead, we will analyze an adversary $\hat{\mathcal{B}}^*$ that introduces an error in each marked column independently with probability β. This adversary may fail to introduce errors in exactly βm random columns, and thus it is not immediately sufficient to bound $\Pr[S(\overline{c}) < -nZ/2]$ for $c' \leftarrow_{\mathrm{R}} \hat{\mathcal{B}}^*(C)$. However, a standard analysis of the binomial distribution shows that this adversary introduces errors in exactly βm marked columns with probability at least

$$1/2\sqrt{m} \geq 1/2\sqrt{d} = 1/\mathrm{poly}(n, \log(1/\xi)),$$

and conditioned on having βm errors, those errors occur on a uniformly random set of marked columns. Thus, if we can show that

$$\Pr_{c' \leftarrow_{\mathrm{R}} \hat{\mathcal{B}}^*(C)}[S(\overline{c}) < -nZ/2] < \xi^{\sqrt{n}/4},$$

we must also have

$$\Pr_{c' \leftarrow_{\mathrm{R}} \mathcal{B}^*(C)}[S(\overline{c}) < -nZ/2] \leq \mathrm{poly}(n, \log(1/\xi)) \cdot \xi^{\sqrt{n}/4} \leq \xi/2,$$

provided $n, 1/\xi$ are sufficiently large.

For the remainder of the proof, we will show that indeed $\Pr[S(\overline{c}) < -nZ/2] < \xi^{\sqrt{n}/4}$ for $c' \leftarrow_{\mathrm{R}} \mathcal{B}^*(C)$. We do so by bounding the quantity $\mathbb{E}_{\overline{p},C}\left[e^{-\alpha S}\right]$ for a suitable $\alpha > 0$ that we will choose later, and then by applying Markov's inequality. Note that the expectation is taken over both the parameters $\overline{p} = (p_1, \ldots, p_d)$ and the randomness of the adversary.

[12] For random variables X and Y over $\mathbb{R}$, X *stochastically dominates* Y if for every $z \in \mathbb{R}$, $\Pr[X \geq r] \geq \Pr[Y \geq r]$.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\overline{p},C}\left[e^{-\alpha S}\right] &= \sum_C \mathop{\mathbb{E}}_{\overline{p}}\left[e^{-\alpha S}\prod_{j=1}^{d} p_j^{x_j}(1-p_j)^{n-x_j}\right] \\
&= \sum_C \mathop{\mathbb{E}}_{\overline{p}}\left[\prod_{j=1}^{d} p_j^{x_j}(1-p_j)^{n-x_j}e^{-\alpha\overline{c}_j\left(x_jq_j-\frac{n-x_j}{q_j}\right)}\right] \\
&= \sum_C \prod_{j=1}^{d}\mathop{\mathbb{E}}_{p}\left[p^{x_j}(1-p)^{n-x_j}e^{-\alpha\overline{c}_j\left(x_jq_j-\frac{n-x_j}{q_j}\right)}\right].
\end{aligned}$$

The first two equalities are by definition. The third equality follows from observing that for fixed C, each term in the product depends only on the (independent) choice of p_j and the adversary's choice of $\overline{c}_j$, and are thus independent by our choice of adversary $\tilde{\mathcal{B}}^*$. This step is the sole reason why it was helpful to consider an adversarial strategy that treats columns independently. Now we want to interchange the sum and product to obtain a product of identical terms, so we can analyze the contribution of an individual term to the product.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\overline{p},C}\left[e^{-\alpha S}\right] &= \sum_C \prod_{j=1}^{d}\mathop{\mathbb{E}}_{p}\left[p^{x_j}(1-p)^{n-x_j}e^{-\alpha\overline{c}_j\left(x_jq_j-\frac{n-x_j}{q_j}\right)}\right] \\
\text{(independence of } \overline{c}_j\text{'s)} \qquad &= \left(\sum_{x=0}^{n}\binom{n}{x}\mathop{\mathbb{E}}_{p}\left[p^x(1-p)^{n-x}e^{-\alpha\overline{c}\left(xq-\frac{n-x}{q}\right)}\right]\right)^d \\
&= \left(\sum_{x=0}^{n}\binom{n}{x}A_x\right)^d,
\end{aligned}$$

where

$$A_x = \begin{cases} (1-\beta)\,\mathbb{E}_p\left[(1-p)^n\right] + \beta\,\mathbb{E}_p\left[(1-p)^n e^{\alpha n/q}\right] & \text{if } x = 0, \\ \mathbb{E}_p\left[p^x(1-p)^{n-x}\right] & \text{if } 1 \le x \le n-1, \\ (1-\beta)\,\mathbb{E}_p\left[p^n\right] + \beta\,\mathbb{E}_p\left[p^n e^{\alpha nq}\right] & \text{if } x = n. \end{cases}$$

First, observe that, since the distribution of p is symmetric about $1/2$, $A_0 = A_n$. Second, if we let

$$B_x = \mathop{\mathbb{E}}_{p}\left[p^x(1-p)^{n-x}\right]$$

for every $x = 0, 1, \ldots, n$, then we have

$$\begin{aligned}
\sum_{x=0}^{n}\binom{n}{x}A_x &= \left(\sum_{x=0}^{n}\binom{n}{x}B_x\right) + 2(A_n - B_n) \\
&= 1 + 2(A_n - B_n).
\end{aligned}$$

In order to obtain a strong enough bound, we need to show that $A_n - B_n = O(\beta\alpha)$. We can calculate

$$\begin{aligned}
A_n - B_n &= (1-\beta)\mathop{\mathbb{E}}_{p}\left[p^n\right] + \beta\mathop{\mathbb{E}}_{p}\left[p^n e^{\alpha nq}\right] - \mathop{\mathbb{E}}_{p}\left[p^n\right] \\
&= \beta\mathop{\mathbb{E}}_{p}\left[p^n e^{\alpha nq}\right] - \beta\mathop{\mathbb{E}}_{p}\left[p^n\right].
\end{aligned}$$

Now we apply the approximation $e^u \leq 1 + 2u$, which holds for $0 \leq u \leq 1$. To do so, we choose $\alpha = \sqrt{t}/n$. Since $q = \sqrt{(1-p)/p}$ and $p \geq t$, we have $\alpha n q \leq 1$ for this choice of α. Thus we have

$$\begin{aligned} A_n - B_n &= \beta \mathop{\mathbb{E}}_{p} [p^n e^{\alpha n q}] - \beta \mathop{\mathbb{E}}_{p} [p^n] \\ &\leq \beta \mathop{\mathbb{E}}_{p} [p^n (1 + 2\alpha n q)] - \beta \mathop{\mathbb{E}}_{p} [p^n] \\ &= 2\beta\alpha \mathop{\mathbb{E}}_{p} [p^n n q] . \end{aligned}$$

Now, to show that $A_n - B_n = O(\beta\alpha)$, we simply want to show that $\mathbb{E}_p [p^n n q] = O(1)$, which we do by direct calculation:

$$\begin{aligned} \mathop{\mathbb{E}}_{p} \left[p^n n \sqrt{\frac{1-p}{p}} \right] &= n \int_{t'}^{\pi/2 - t'} \frac{\sin^{2n} r \sqrt{\frac{1-\sin^2 r}{\sin^2 r}}}{\pi/2 - 2t'} dr = \frac{\sin^{2n}(\pi/2 - t') - \sin^{2n}(t')}{\pi - 4t'} \\ &= \frac{(1-t)^n - t^n}{\pi - 4t'} = \frac{(1 - 1/300n)^n - (1/300n)^n}{\pi - 4t'} \leq \frac{1}{\pi}. \end{aligned}$$

The final inequality holds as long as n is larger than some absolute constant. (To see that this is the case, recall that $t' = \arcsin(\sqrt{t}) = \arcsin(\sqrt{1/300n}) = \Theta(1/\sqrt{n})$, whereas $(1 - 1/300n)^n = 1 - \Omega(1)$.) So we have established

$$A_n - B_n \leq \frac{2\beta\alpha}{\pi}.$$

Plugging this fact into the analysis above, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\overline{p},C} \left[e^{-\alpha S} \right] &= \left(\sum_{x=0}^{n} \binom{n}{x} A_x \right)^d \\ &= (1 + 2(A_n - B_n))^d \\ &\leq \left(1 + \frac{4\beta\alpha}{\pi} \right)^d \leq e^{4\beta\alpha d/\pi}. \end{aligned}$$

Now all that remains is to apply Markov's inequality to bound this quantity by $\xi^{\sqrt{n}/4}$:

$$\begin{aligned} \Pr[S < -nZ/2] &= \Pr[-\alpha S > \alpha n Z/2] \\ &= \Pr\left[e^{-\alpha S} > e^{\alpha n Z/2} \right] \leq \frac{\mathbb{E}\left[e^{-\alpha S} \right]}{e^{\alpha n Z/2}} \leq \frac{e^{4\beta\alpha d/\pi}}{e^{\alpha n Z/2}} \\ &= e^{4\beta\alpha d/\pi - \alpha n Z/2}. \end{aligned}$$

To get the desired upper bound, it is sufficient to show

$$\frac{\alpha n Z}{2} - \frac{4\beta\alpha d}{\pi} \geq \frac{\sqrt{n}\log(1/\xi)}{4}.$$

We calculate

$$\begin{aligned} \frac{\alpha n Z}{2} - \frac{4\beta\alpha d}{\pi} &= 10\sqrt{t} n \log(n/\xi) - \frac{400\beta}{\pi}\sqrt{t} n \log(n/\xi) \\ &= \left(10 - \frac{400\beta}{\pi} \right) \left(\sqrt{t} n \log(n/\xi) \right) \\ &\geq \left(10 - \frac{400\beta}{\pi} \right) \frac{\sqrt{n}\log(n/\xi)}{18} \\ &\geq \frac{\sqrt{n}\log(1/\xi)}{4}, \end{aligned}$$

where the last inequality holds when $\beta < 1/25$. This is sufficient to complete the proof of Claim 6.10. ∎

Combining Claims 6.9 and 6.10 yields Lemma 6.8 as follows. If $S(c') < nZ/2$, then either $S(\tilde{c}) < nZ$ or $S(\bar{c}) < nZ/2$. Moreover, if $c' \in WF_{1/25}(C)$, we must have both $\tilde{c} \in WF_0(C)$ and a valid $\bar{c}$. A union bound thereby gives us Lemma 6.8. ∎

Lemma 6.7 and 6.8 are sufficient to imply Lemma 6.6, that Tardos' fingerprinting code is weakly robust. In order to apply our reduction from full robustness to weak robustness (Lemma 6.4), we need to also establish that with high probability there are many marked columns in the matrix $C \leftarrow_{\mathrm{R}} Gen$ for Tardos' fingerprinting code.

LEMMA 6.12. *With probability at least* $1 - \xi$ *over the choice of* $C \leftarrow_R Gen$, *it holds that the number of* 0*-marked columns* m_0 *and the number of* 1*-marked columns* m_1 *are both larger than* $m = 5n^{3/2}\log(n/\xi)$.

Proof of Lemma 6.12. To estimate the number of marked columns, define for each $j = 1, \ldots, d$ an indicator random variable D_j for whether column j is 0-marked. The D_j's are i.i.d., and have expectation at least

$$\mathbb{E}\left[D_j | p_j < 1/n\right] \Pr[p_j < 1/n] > \left(1 - \frac{1}{n}\right)^n \Pr[r_j < \arcsin(1/\sqrt{n})] \geq \frac{1}{6\sqrt{n}}.$$

Let $D = \sum_{j=1}^{d} D_j$ be the total number of 0-marked columns. Then $\mathbb{E}[D] \geq 10n\sqrt{n}\log(n/\xi)$, so by the additive Chernoff bound (Theorem 5.15),

$$\Pr[D < 5n\sqrt{n}\log(n/\xi)] < \exp\left(\frac{-2(5n\sqrt{n}\log(n/\xi))^2}{d}\right) < \xi/2.$$

A similar argument holds for 1-marked columns. Thus letting $m = 5n\sqrt{n}\log(n/\xi)$, the codebook C has at least m 0-marked columns and m 1-marked columns with probability at least $1 - \xi$. Now observe that

$$\exp(-\Omega(\beta m^2/d)) < \exp(-\Omega(\beta n \log(n/\xi))) < \xi$$

for n larger than some absolute constant. ∎

Combining Lemma 6.4 (reduction from robustness to weak robustness), Lemma 6.6 (weak robustness of Tardos' code), and Lemma 6.12 (Tardos' code has many marked columns), suffices to prove Theorem 6.2.

Appendix A. Lower bounds on fingerprinting codes via privacy. By the contrapositive of Theorem 3.5, upper bounds on the sample complexity of answering 1-way marginals with differential privacy imply a lower bound on the length d of any fingerprinting code with a given number of users n. As pointed out to us by Adam Smith, this yields a particularly simple, self-contained proof of Tardos' [49] optimal lower bound on the length of fingerprinting codes. Specifically, using the well known Gaussian mechanism for achieving differential privacy, we can design a simple adversary $\mathcal{A}_{FP}$ that violates the security of any traitor tracing scheme with length $d = o(n^2)$.

THEOREM A.1. *There is a function* $n = n(d) = \tilde{O}(\sqrt{d})$ *such that for every* d, *there is no* (n, d)*-fingerprinting code with security* $\xi < 1/6en$.

Proof. Before diving into the proof, we will state the following elementary fact about Gaussian random variables. The fact simply says that a Gaussian random

variable with suitable variance is "close" to a shifted version of itself in a particular sense. This same fact is used to show that adding Gaussian noise of suitable variance provides differential privacy.

FACT A.2. *Let $c, c' \in \mathbb{R}^d$ satisfy $\|c - c'\|_2 \leq \sqrt{d}/n$, $\delta > 0$ be a parameter, and let $\sigma^2 = 2d\ln(1/\delta)/n^2$. Let $z \in \mathbb{R}^d$ be a random vector where each coordinate is an independent draw from a Gaussian distribution with mean 0 and variance σ^2. Then for any (measurable) set $T \subseteq \mathbb{R}^d$,*

$$\Pr_z[c + z \in T] \geq (1/e)\Pr_z[c' + z \in T] - \delta.$$

Now we proceed with the proof. Fix any choice of d. Assume towards a contradiction that there is an (n, d)-fingerprinting code $(Gen, Trace)$ with security $\xi < 1/6en$ for $n = \lceil \sqrt{18d\ln(6en)\ln(3d/2)} \rceil$. Observe that $n = n(d) = \tilde{O}(\sqrt{d})$ as promised in the theorem.

Let $\mathcal{A}_{FP}(C_S)$ be the following adversary. Define the vector $\bar{c} \in [0,1]^d$ as

$$\bar{c} = \frac{1}{n}\sum_{i \in S} c_i.$$

Now, let $z \in \mathbb{R}^d$ be a d-dimensional Gaussian where every coordinate is independent with mean 0 and variance $\sigma^2 = 2d\ln(1/\delta)/n^2$ for $\delta = 1/6en$. Finally, let c' be $\hat{c}$ with each coordinate rounded to $\{0, 1\}$, and output the pirated codeword c'.

First we claim that $\mathcal{A}_{FP}$ outputs feasible codewords with at least constant probability.

CLAIM A.3. *For every S such that $|S| \geq n - 1$, and every codebook $C = (c_{ij}) \in \{0,1\}^{n \times d}$,*

$$\Pr_{c' \leftarrow_R \mathcal{A}_{FP}(C_S)}[c' \in F(C_S)] \geq 2/3.$$

Proof of Claim A.3. By a standard tail bound for the Gaussian, we have

$$\Pr\left[\forall j,\ |z_j| < \sigma\sqrt{\ln(3d/2)}\right] \geq 2/3.$$

Thus, by our choice of σ and $n \geq \sqrt{18d\ln(1/\delta)\ln(3d/2)}$ we have $\Pr[\forall j,\ |z_j| < 1/3] \geq 2/3$. Now the claim follows easily. Specifically, if $c_{ij} = 1$ for every $i \in S$, then $(1/n)\sum_{i \in S} c_{ij} \geq 1 - 1/n$, so $\hat{c}_j > 2/3 - 1/n$ and $c'_j = 1$. A similar argument applies if $c_{ij} = 0$ for every $i \in S$. ∎

Now it remains to show that $\mathcal{A}_{FP}$ cannot be traced successfully. By assumption $(Gen, Trace)$ has security $\xi < 1/6en < 1/3$. Then we have in particular

$$\Pr_{\substack{C \leftarrow_R Gen \\ c' \leftarrow_R \mathcal{A}_{FP}(C)}}[c' \in F(C) \wedge Trace(C, c') = \perp] < \xi.$$

Combining with Claim A.3 we have

$$\Pr_{\substack{C \leftarrow_R Gen \\ c' \leftarrow_R \mathcal{A}_{FP}(C)}}[Trace(C, c') \in [n]] > 1 - 1/3 - \xi > 1/3.$$

Therefore, there exists $i^* \in [n]$ such that

$$\Pr_{\substack{C \leftarrow_R Gen \\ c' \leftarrow_R \mathcal{A}_{FP}(C)}}[Trace(C, c') = i^*] > 1/3n. \tag{10}$$

To complete the proof, it now suffices to show that if $S = [n] \setminus \{i^*\}$, then

$$\Pr_{\substack{C \leftarrow_R Gen \\ c' \leftarrow_R \mathcal{A}_{FP}(C_S)}} [Trace(C, c') = i^*] \geq 1/6en > \xi,$$

which will contradict the security of the fingerprinting code.

To do so, first observe that if

$$\overline{c} = \frac{1}{n} \sum_{i \in [n]} c_i \qquad \text{and} \qquad \overline{c}^S = \frac{1}{n} \sum_{i \in S} c_i,$$

then $\|\overline{c}_j - \overline{c}_j^S\|_2 \leq \sqrt{d}/n$. Now, in case the tracing algorithm is randomized, let $Trace_r$ denote the tracing algorithm when run with its random coins fixed to r. For any string of random coins r, define the set $T_r = \{t \in \mathbb{R}^d \mid Trace_r(C, \text{round}(t)) = i^*\}$. Here, round$(\cdot)$ is the function that rounds each entry of its input to $\{0, 1\}$.[13]

By Fact A.2 (with $\delta = 1/6en > \xi$), for every r,

$$\Pr_z \left[\overline{c}^S + z \in T_r\right] \geq (1/e) \Pr_z [\overline{c} + z \in T_r] - \xi.$$

Applying (10), and averaging over $C \leftarrow_R Gen$ and r, we have

$$\Pr_{\substack{C \leftarrow_R Gen \\ c' \leftarrow_R \mathcal{A}_{FP}(C_S)}} [Trace(C, c') = i^*] \geq (1/e)(1/3n) - 1/6en = 1/6en > \xi,$$

which is the desired contradiction. This completes the proof. ∎

Acknowledgments. We thank Kobbi Nissim for drawing our attention to the question of sample complexity and for many helpful discussions. We thank Adam Smith for suggesting that we use the Gaussian mechanism to provide a new proof of the lower bound on the length of fingerprinting codes. Finally, we thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] M. Anthony and P. L. Bartlett, *Neural Network Learning: Theoretical Foundations*, 1st ed., Cambridge University Press, New York, 2009.

[2] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar, *Privacy, accuracy, and consistency too: a holistic solution to contingency table release*, in PODS, ACM, New York, 2007, pp. 273–282.

[3] R. Bassily, K. Nissim, A. Smith, T. Steinke, U. Stemmer, and J. Ullman, *Algorithmic stability for adaptive data analysis*, in Symposium on Theory of Computing (STOC'16), 2016.

[4] R. Bassily, A. Smith, and A. Thakurta, *Private empirical risk minimization: Efficient algorithms and tight error bounds*, in FOCS, IEEE, Piscataway, NJ, 2014, pp. 464–473.

[5] A. Beimel, S. P. Kasiviswanathan, and K. Nissim, *Bounds on the sample complexity for private learning and private data release*, in TCC, Springer, Berlin, 2010, pp. 437–454.

[6] A. Beimel, K. Nissim, and U. Stemmer, *Characterizing the sample complexity of private learners*, in ITCS, ACM, New York, 2013, pp. 97–110.

[7] A. Beimel, K. Nissim, and U. Stemmer, *Private learning and sanitization: Pure vs. approximate differential privacy*, in APPROX-RANDOM, Springer, Heidelberg, 2013, pp. 363–378.

[8] A. Blum, C. Dwork, F. McSherry, and K. Nissim, *Practical privacy: the SuLQ framework*, in PODS, ACM, New York, 2005, pp. 128–138.

[13] Note, for completeness, that T_r is measurable, since the set of $c' \in \{0,1\}^d$ such that $Trace_r(C, c') = i^*$ is finite (for every fixed n, d) and for every c', $\{t \mid \text{round}(t) = c'\}$ is a hypercube, so T_r is a union of finitely many hypercubes.

[9] A. Blum, K. Ligett, and A. Roth, *A learning theory approach to non-interactive database privacy*, in STOC, ACM, New York, 2008.
[10] D. Boneh, A. Kiayias, and H. W. Montgomery, *Robust fingerprinting codes: a near optimal construction*, in Digital Rights Management Workshop, ACM, New York, 2010, pp. 3–12.
[11] D. Boneh and M. Naor, *Traitor tracing with constant size ciphertext*, in CCS, ACM, New York, 2008, pp. 501–510.
[12] D. Boneh and J. Shaw, *Collusion-secure fingerprinting for digital data*, IEEE Trans. Inform. Theory, 44 (1998), pp. 1897–1905.
[13] M. Bun, K. Nissim, U. Stemmer, and S. P. Vadhan, *Differentially private release and learning of threshold functions*, in FOCS, IEEE, Piscataway, NJ, 2015, pp. 634–649.
[14] K. Chandrasekaran, J. Thaler, J. Ullman, and A. Wan, *Faster private release of marginals on small databases*, ITCS 2014, ACM, New York, 2014, pp. 387–402.
[15] A. De, *Lower bounds in differential privacy*, in TCC, Springer, Heidelberg, 2012, pp. 321–338.
[16] I. Dinur and K. Nissim, *Revealing information while preserving privacy*, in PODS, ACM, New York, 2003, pp. 202–210.
[17] D. P. Dubhashi and S. Sen, *Concentration of measure for randomized algorithms: techniques and applications*, in Handbook of Randomized Computing, vol. 1, Kluwer, Dordrecht, 2001, pp. 35–100.
[18] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, *Local privacy and statistical minimax rates*, in 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, IEEE, Piscataway, NJ, 2013, pp. 429–438.
[19] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. L. Roth, *Preserving statistical validity in adaptive data analysis*, in STOC, ACM, New York, 2015, pp. 117–126.
[20] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, *Our data, ourselves: Privacy via distributed noise generation*, in EUROCRYPT, Springer, Berlin, 2006, pp. 486–503.
[21] C. Dwork, F. McSherry, K. Nissim, and A. Smith, *Calibrating noise to sensitivity in private data analysis*, in TCC, Springer, Berlin, 2006, pp. 265–284.
[22] C. Dwork, F. McSherry, and K. Talwar, *The price of privacy and the limits of LP decoding*, in STOC, ACM, New York, 2007, pp. 85–94.
[23] C. Dwork, M. Naor, O. Reingold, G. N. Rothblum, and S. P. Vadhan, *On the complexity of differentially private data release: Efficient algorithms and hardness results*, in STOC, ACM, New York, 2009, pp. 381–390.
[24] C. Dwork, M. Naor, and S. P. Vadhan, *The privacy of the analyst and the power of the state*, in FOCS, IEEE Computer Society, Los, Alamitos, CA, 2012, pp. 400–409.
[25] C. Dwork, A. Nikolov, and K. Talwar, *Efficient Algorithms for Privately Releasing Marginals via Convex Programming*, manuscript.
[26] C. Dwork and K. Nissim, *Privacy-preserving datamining on vertically partitioned databases*, in CRYPTO, Springer, Berlin, 2004, pp. 528–544.
[27] C. Dwork, G. N. Rothblum, and S. P. Vadhan, *Boosting and differential privacy*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 2010, pp. 51–60.
[28] C. Dwork, A. Smith, T. Steinke, J. Ullman, and S. Vadhan, *Robust traceability from trace amounts*, in FOCS, IEEE, Piscataway, NJ, 2015.
[29] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang, *Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis*, in Symposium on Theory of Computing STOC, ACM, New York, 2014, pp. 11–20.
[30] C. Dwork and S. Yekhanin, *New efficient attacks on statistical disclosure control mechanisms*, in CRYPTO, Springer, Berlin, 2008, pp. 469–480.
[31] A. Gupta, M. Hardt, A. Roth, and J. Ullman, *Privately releasing conjunctions and the statistical query barrier*, in STOC, ACM, New York, 2011, pp. 803–812.
[32] A. Gupta, A. Roth, and J. Ullman, *Iterative constructions and private data release*, in TCC, Springer, Heidelberg, 2012, pp. 339–356.
[33] M. Hardt, *A Study in Privacy and Fairness in Sensitive Data Analysis*, PhD thesis, Princeton University, Princeton, NJ, 2011.
[34] M. Hardt, K. Ligett, and F. McSherry, *A simple and practical algorithm for differentially private data release*, in NIPS, Curran Associates, Red Hook, NY, 2012, pp. 2339–2347.
[35] M. Hardt and G. N. Rothblum, *A multiplicative weights mechanism for privacy-preserving data analysis*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 2010, pp. 61–70.
[36] M. Hardt and K. Talwar, *On the geometry of differential privacy*, in STOC, ACM, New York, 2010, pp. 705–714.
[37] M. Hardt and J. Ullman, *Preventing false discovery in interactive data analysis is hard*, in FOCS, IEEE, Piscataway, NJ, 2014.

[38] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig, *Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays*, PLoS Genetics, 4 (2008), e1000167.

[39] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, *What can we learn privately?*, SIAM J. Comput., 40 (2011), pp. 793–826.

[40] S. P. Kasiviswanathan, M. Rudelson, A. Smith, and J. Ullman, *The price of privately releasing contingency tables and the spectra of random matrices with correlated rows*, in STOC, ACM, New York, 2010, pp. 775–784.

[41] E. Liberty, M. Mitzenmacher, J. Thaler, and J. Ullman, *Space lower bounds for itemset frequency sketches*, in PODS 2016, ACM, New York, 2016, pp. 441–454.

[42] A. Nikolov, K. Talwar, and L. Zhang, *The geometry of differential privacy: The sparse and approximate cases*, in STOC, ACM, New York, 2013, pp. 351–360.

[43] A. Roth, *Differential privacy and the fat-shattering dimension of linear queries*, in APPROX-RANDOM, Springer, Berlin, 2010, pp. 683–695.

[44] A. Roth and T. Roughgarden, *Interactive privacy via the median mechanism*, in STOC, ACM, New York, 2010, pp. 765–774.

[45] M. Rudelson, *Row products of random matrices*, Adv. Math., 231 (2012), pp. 3199–3231.

[46] S. Sankararaman, G. Obozinski, M. I. Jordan, and E. Halperin, *Genomic privacy and limits of individual detection in a pool*, Nature Genetics, 41 (2009), pp. 965–967.

[47] T. Steinke and J. Ullman, *Between pure and approximate differential privacy*, J. Privacy Confidentiality, 7 (2017), Article 2.

[48] T. Steinke and J. Ullman, *Interactive fingerprinting codes and the hardness of preventing false discovery*, in Proceedings of the 28th Conference on Learning Theory (COLT'15), 2015, pp. 1588–1628.

[49] G. Tardos, *Optimal probabilistic fingerprint codes*, J. ACM, 55 (2008), 10.

[50] J. Thaler, J. Ullman, and S. P. Vadhan, *Faster algorithms for privately releasing marginals*, in ICALP, Springer, Heidelberg, 2012, pp. 810–821.

[51] J. Ullman, *Answering $n^{2+o(1)}$ counting queries with differential privacy is hard*, in STOC, ACM, New York, 2013, pp. 361–370.

[52] J. Ullman and S. P. Vadhan, *PCPs and the hardness of generating private synthetic data*, in TCC, Springer, Berlin, 2011, pp. 400–416.

[53] S. Vadhan, *The Complexity of Differential Privacy*, http://privacytools.seas.harvard.edu/publications/complexity-differential-privacy (2016).