Predicting Adaptively Chosen Observables in Quantum Systems

Jerry Huang,[1,*] Laura Lewis,[1,2,3] Hsin-Yuan Huang,[1,2,4] and John Preskill[1,5]

[1] *Institute for Quantum Information and Matter, Caltech, Pasadena, California, USA*
[2] *Google Quantum AI, Venice, California, USA*
[3] *University of Cambridge, Cambridge, United Kingdom*
[4] *Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
[5] *AWS Center for Quantum Computing, Pasadena, California, USA*

(Received 12 June 2025; accepted 28 January 2026; published 9 March 2026)

Recent advances have demonstrated that $\mathcal{O}(\log M)$ measurements suffice to predict M properties of arbitrarily large quantum many-body systems. However, these remarkable findings assume that the properties to be predicted are chosen independently of the data. This assumption can be violated in practice, where scientists adaptively select properties after looking at previous predictions. This work investigates the adaptive setting for three classes of observables: local, Pauli, and bounded-Frobenius-norm observables. We prove that $\Omega(\sqrt{M})$ samples of an arbitrarily large unknown quantum state are necessary to predict expectation values of M adaptively chosen local and Pauli observables, where the system size scales exponentially and polynomially in M, respectively. We also present computationally efficient algorithms that achieve this information-theoretic lower bound. In contrast, for bounded-Frobenius-norm observables, we devise an algorithm requiring only $\mathcal{O}(\log M)$ samples, independent of system size. These results highlight the potential pitfalls of adaptivity in analyzing data from quantum experiments and provide algorithmic tools to safeguard against erroneous predictions in quantum experiments.

DOI: 10.1103/xhn1-vnp9

I. INTRODUCTION

Estimating properties of quantum systems using data collected from physical experiments is fundamental to the advancement of quantum information science. However, the required resources often scale unfavorably with system size n. For instance, full tomography of an n-qubit state requires a number of samples that grows exponentially in n [1–6]. Recent results have addressed this scaling problem by demonstrating that many properties can be predicted from very few samples, even for very large n [7–11]. These sample-efficient protocols are known as shadow tomography. In particular, the classical shadow formalism [8] constructs a succinct classical description of an unknown state ρ that accurately captures expectation values $\{o_i\}$ of a set of observables $\{O_i\}$,

$$o_i(\rho) = \mathrm{tr}(O_i\rho), \quad 1 \le i \le M, \tag{1}$$

using only $\mathcal{O}(\log M)$ samples [12]. This achieves an exponential improvement in sample complexity compared to direct measurement of each observable, even when $M = \mathrm{poly}(n)$.

However, this efficiency relies on a critical assumption: the observables $\{O_i\}$ are chosen *nonadaptively*. This means that the next observable O_{i+1} cannot be selected based on the prediction outcomes for observables $O_1, O_2, \ldots, O_i$. While classical shadows can be acquired using randomized measurements [11] with no knowledge of the observables, the prediction performance guarantee derived in Ref. [8] assumes that the set of observables $\{O_i\}$ is chosen independently of the measurement outcomes and the predicted values of other observables. In contrast, scientific research is often conducted adaptively: hypotheses are formulated from experimental data, tested using the same data, and then adaptively modified and retested. This process extends to quantum experiments, where one may learn certain properties of a quantum state and, inspired by the results, decide on additional properties to predict using the same data. However, there is currently no rigorous guarantee maintaining the $\mathrm{poly}\log(M)$ sample complexity achieved by the classical shadow formalism when observables are chosen *adaptively*.

This work addresses this gap by exploring the reusability of quantum data. We focus on answering the

*Contact author: jyhuang@alumni.caltech.edu

following central question:

> *Can we still predict M properties of an arbitrarily large quantum system from* poly $\log(M)$ *samples when the properties are chosen adaptively?*

Specifically, we seek to maintain the exponential reduction in sample complexity compared to direct observable measurement (as achieved by classical shadows in the nonadaptive setting) even when the number of properties is $M = \text{poly}(n)$. This necessitates avoiding any polynomial dependence of the sample complexity on the system size n. While existing protocols in shadow tomography [7,13,14] achieve a polynomial reduction in sample complexity compared to direct measurement, with a sample complexity of $\mathcal{O}(n \log^2 n)$ for $M = \text{poly}(n)$, they do not fully address our central question, as we aim for a superpolynomially reduced sample complexity of poly $\log(M)$.

Our work relates closely to adaptive data analysis in classical statistics [15–25], where similar challenges arise in maintaining statistical validity under adaptive hypothesis testing [26–30]. In the classical setting, the best-known sample complexity for predicting M adaptively chosen queries is $\mathcal{O}(\sqrt{M})$ [17,23,24] when dimension dependence is not allowed, or $\mathcal{O}(\sqrt{d} \log M)$ samples where d is the dataset dimension [17,31,32]. Despite significant progress in classical statistics, it is not immediately clear whether these results apply to physically relevant classes of observables in the quantum setting, and whether the classical shadow formalism is compromised by adaptively choosing the observables.

By building on techniques in shadow tomography and adaptive data analysis, we present both positive results and impossibility theorems for three physically relevant classes of observables: local, Pauli, and bounded-Frobenius-norm observables. Estimations of these observables are key subroutines in algorithms for quantum chemistry [33–37], quantum field theory [38], and linear algebra [39], as well as protocols for estimating fidelities [8,40,41] and verifying entanglement [11,42,43]. Our main findings are:

(1) For local and Pauli observables, we prove a lower bound of $\Omega(\sqrt{M})$ samples for any algorithm that accurately predicts adaptively chosen properties. We complement this with computationally efficient algorithms achieving matching upper bounds.
(2) For bounded-Frobenius-norm observables, we devise a (computationally inefficient) algorithm that accurately predicts expectation values with a sample complexity of $\mathcal{O}(\log M)$, independent of system size. This improves upon existing shadow tomography protocols by exploiting the low-dimensional nature of these observables.

The $\Omega(\sqrt{M})$ lower bound demonstrates that it is impossible to achieve the desired poly $\log(M)$ sample complexity even for predicting local and Pauli observables. We note that this surprising result requires the system size to be exponential in M for local observables and polynomial in M for Pauli observables. Hence, this impossibility result only applies to systems containing many more qubits than the number of observables M. For bounded-Frobenius-norm observables, the $\mathcal{O}(\log M)$ sample complexity improves upon the best-known scaling of $\mathcal{O}(n \log^2 M)$ [14].

Our work makes several key contributions. First, we demonstrate that a nontrivial extension of lower bounds in classical adaptive data analysis can yield lower bounds for adaptively chosen local and Pauli observables in quantum systems. Second, by identifying and leveraging the connection between existing classical adaptive data analysis algorithms and classical shadows, we obtain upper bounds for local and Pauli observables in the quantum setting. Finally, we introduce a new (computationally inefficient) algorithm for predicting expectation values of bounded-Frobenius-norm observables that achieves logarithmic sample complexity in the adaptive setting. In the process, we develop new algorithms for quantum threshold search (TS) [14] and shadow tomography which may be of independent interest. In the following sections, we present our results in detail, including lower bounds, matching upper bounds, and our novel algorithm for bounded-Frobenius-norm observables. We conclude with a discussion of the broader implications of our work and potential directions for future research.

II. THE ADAPTIVE MODEL

We introduce the model describing the adaptive interaction, with the full formal description provided in Appendix A. We conceptualize the predictive task as a game between an *analyst* (e.g., a human scientist) and a *mechanism* (e.g., a prediction algorithm), where only the mechanism has access to samples of the unknown quantum state ρ. This framework aligns with standard models in classical adaptive data analysis [15]. We consider the analyst to be classical, restricted to classical memory and processing, while the mechanism has access to quantum memory and has quantum processing capabilities. The analyst-mechanism interaction unfolds over M rounds, where in each round $i \in [M] \triangleq \{1, \ldots, M\}$: (1) the analyst selects an observable O_i to query, and (2) the mechanism performs quantum processing on the samples and responds with an estimate $\hat{o}_i \approx \text{tr}(O_i \rho)$. Crucially, the analyst can observe all prior responses $\hat{o}_1, \ldots, \hat{o}_{i-1}$ and utilize this information to determine the next observable O_i to query. This adaptive decision-making process characterizes the *adaptive* nature of the interaction. A visual representation of this interaction model is presented in Fig. 1.

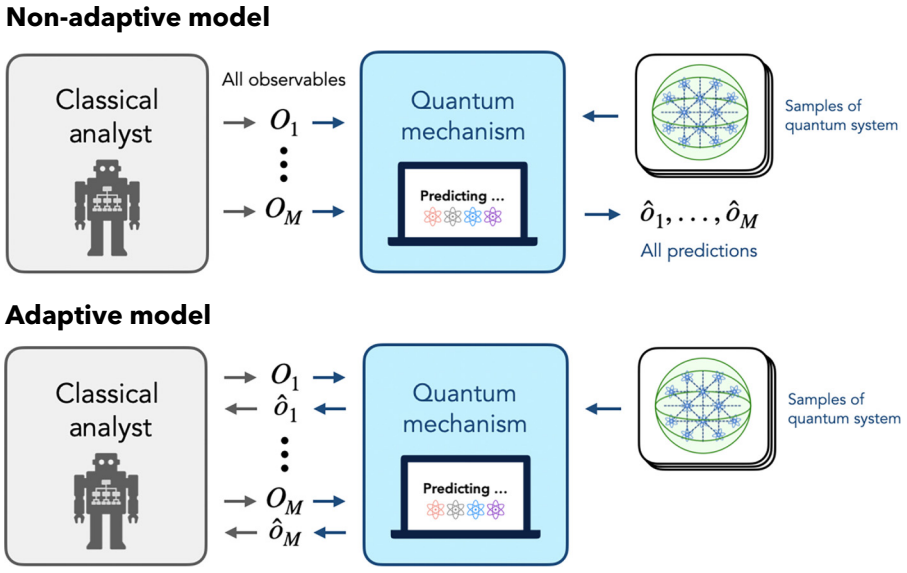

FIG. 1. Nonadaptive vs adaptive model. The analyst queries observables $O_1, \ldots, O_M$ to the mechanism. The mechanism is given access to samples of a quantum state ρ and can conduct quantum processing and measurements in order to output predictions $\hat{o}_i$ for the expectation values $\mathrm{tr}(O_i\rho)$. In the adaptive model (bottom), the analyst can condition on all prior observed predictions $\hat{o}_1, \ldots, \hat{o}_{i-1}$ in order to select the next observable O_i. In contrast, a nonadaptive analyst (top) cannot observe prior predictions before querying observables.

A. Adaptive attack on classical shadows

To illustrate the vulnerabilities introduced by adaptivity, we construct a simple adaptive attack executed by a classical analyst. This attack demonstrates how adversarial behavior can rapidly compromise the original classical shadows protocol [8]. For a review of the classical shadow formalism, we refer readers to Appendix A 2. Our proposed attack is inspired by the linear classification attack from Ref. [15], which in turn draws from Freedman's paradox [44]—a well-known example of standard statistical procedures yielding highly misleading results. Freedman's paradox considers a regression problem where, given access to a large group of variables predicting a response variable, we (1) select a smaller group of variables correlated with the response variable and (2) fit a linear regression model on these variables. Surprisingly, standard statistical tests (e.g., an F-test) report the model as a good fit, even when no correlation exists between the input variables and the response variable.

A similar failure can occur in the quantum setting. Consider N samples of a diagonal density matrix ρ of system size $n = M + 2^M$, where the first M qubits represent variables and the last 2^M qubits correspond to all possible models that can be fit, each conditioned on a unique subset of the M variables. The qubits are correlated such that the spins of the last 2^M qubits are determined by a majority vote of the spins of a subset of the first M qubits.

We examine an analyst's adaptive attack querying only single-qubit observables from the set $\{Z_i : i \in [n]\}$, where Z_i is the Pauli Z observable acting solely on the ith qubit. While the classical shadows protocol can accurately estimate expectation values of exponentially many observables in the number of samples N for nonadaptive queries [8], our adaptive attack causes the protocol to fail with high probability when the number of queries is only $M = \mathcal{O}(N)$, even for single-qubit observables.

The attack mimics Freedman's paradox in a two-step process:

(1) Query Z_i for each of the first M qubits.
(2) Select a subset of qubits whose estimated expectation values are suitably biased toward 1, and query Z_j for the qubit correlated with this subset.

This process yields a similar outcome to Freedman's paradox: the classical shadows estimate for $\mathrm{tr}(Z_j\rho)$ will likely be greater than 0.99, even if the true value is 0. Notably, only the last query is adaptive and is conditioned on the results of the first M nonadaptive queries. We empirically demonstrate this phenomenon in Fig. 2. Because the density matrix is diagonal and we only query Z_i observables, we can simulate this protocol classically. The full description and proof of the attack are provided in Appendix B 1, with details of the numerical experiment in Appendix B 2.

This example underscores the potentially detrimental effect of adaptivity on the accuracy of predictive algorithms: for single-qubit observable queries, a single adaptive query suffices to induce false predictions. Our goal is to prevent such failures and derive rigorous guarantees against false discovery in quantum experiments for specific classes of observables.

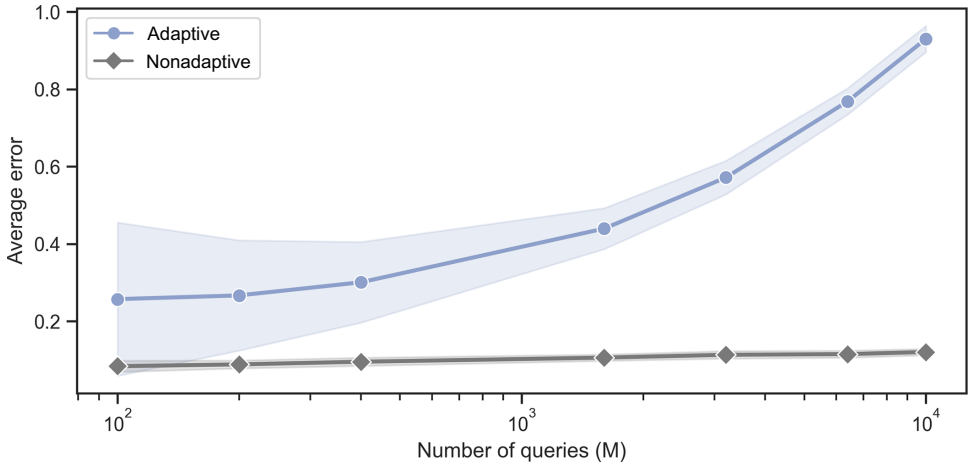

FIG. 2. The dangers of adaptivity. Each point indicates the average error (averaged over 100 independent runs) of the classical shadows protocol after being asked M queries from the analyst for a fixed sample size of $N = 10\,000$. The shaded regions show the standard deviation over the independent runs. We see that after many adaptively chosen queries, the average error of classical shadows is much higher than for nonadaptive queries.

III. MAIN RESULTS

Our primary focus is on the task of predicting properties $\mathrm{tr}(O_i\rho)$, for $i = 1, \ldots, M$, where ρ is an unknown n-qubit quantum state and O_i is an observable. We aim to design a mechanism that can accurately predict these properties to within ϵ error using minimal samples/copies of ρ, even when the properties are chosen *adaptively*. In the following subsections, we address this challenge for three distinct classes of observables: local, Pauli, and bounded-Frobenius-norm observables.

A. Local observables

We define local observables as those acting on a constant number $k = \mathcal{O}(1)$ of qubits. Recall that the adaptive attack described in Sec. II A can cause the classical shadows protocol to fail by querying only $M = \mathcal{O}(N)$ local observables, where N is the number of samples of ρ. Our main result in this section demonstrates that $\mathcal{O}(N^2)$ local observables suffice to cause *any* quantum mechanism to fail, given a sufficiently large system size n. In other words, we prove an information-theoretic sample complexity lower bound of $N = \Omega(\sqrt{M})$ for predicting M adaptively chosen local observables when the system size n is exponential in M. We also show that this lower bound can be saturated by a computationally efficient quantum mechanism. The formal statement is given as follows.

Theorem 1 (Local observables). For predicting expectation values $\mathrm{tr}(O_i\rho)$ of M adaptively chosen local observables $O_1, \ldots, O_M$, any quantum mechanism must use $N = \Omega(\min\{\sqrt{M}, \log n\})$ samples. Moreover, there exists a computationally efficient quantum mechanism using $N = \mathcal{O}(\min\{\sqrt{M}, \log n\})$ samples, running in time $\mathrm{poly}(N, n)$ per observable queried.

To prove the lower bound, we construct a density matrix ρ and an analyst strategy for adaptively querying M local observables, inspired by lower bounds in classical adaptive data analysis [15,32,45] based on interactive fingerprinting codes (IFPCs) [46]. The proof ideas are presented in Sec. V, with detailed proofs in Appendix B 3.

We emphasize that while there exists an $\Omega(\sqrt{M})$ lower bound in classical adaptive data analysis, this only applies for *arbitrary* adaptive queries. In contrast, our lower bound stated here applies when restricted to the class of local observables—a setting with no existing classical result. In fact, our quantum lower bound for local observables yields a new classical lower bound for sparse statistical queries (Corollary B1 in Appendix B 3).

The matching upper bound is achieved by applying classical adaptive data analysis algorithms [17,19] to classical shadow data [8]. We provide two computationally efficient algorithms in Appendixes B 4 and D, both running in time $\mathrm{poly}(N, n)$ per observable. The first algorithm, specific to local observables, achieves a sample complexity of $\mathcal{O}(\sqrt{M})$. The second, more general algorithm applies to any class of observables and measurement primitive, achieving an upper bound of $\mathcal{O}(\sqrt{M}\log M \cdot \|O\|^2_{\mathrm{shadow}})$, where $\|\cdot\|_{\mathrm{shadow}}$ is the shadow norm dependent on the observables and measurement primitive. For local observables, $\|O\|^2_{\mathrm{shadow}}$ is constant, so we achieve nearly the same $\tilde{\mathcal{O}}(\sqrt{M})$ upper bound. We expect this second algorithm to be of independent interest, as it applies to any observable.

B. Pauli observables

For Pauli observables $P \in \{I, X, Y, Z\}^{\otimes n}$, we demonstrate that any quantum mechanism requires at least $\Omega(\sqrt{M})$ samples, assuming a sufficiently large system size. Unlike local observables, the $\Omega(\sqrt{M})$ lower bound

holds when the system size is polynomial in M, which is realizable in many relevant regimes. Combined with the previous section, our results eliminate the hope of replicating the nonadaptive classical shadows scaling in the adaptive setting for arbitrary system sizes.

Theorem 2 (Pauli observables). For predicting expectation values $\mathrm{tr}(P_i\rho)$ of M adaptively chosen Pauli observables $P_1,\ldots,P_M \in \{I,X,Y,Z\}^{\otimes n}$, any quantum mechanism must use $N = \Omega(\min\{\sqrt{M}, n^{1/3}\})$ samples. Moreover, there exists a computationally efficient quantum mechanism using $N = \mathcal{O}(\min\{\sqrt{M}, n\})$ samples, running in time $\mathrm{poly}(N,n)$ per observable queried.

Similarly to the local observables case, the lower bound is obtained by constructing an adversarial choice of density matrix and analyst strategy, leveraging interactive fingerprinting codes [46]. However, the particular construction differs from the case of predicting local observables in order to embed the interaction within a polynomial number of qubits. Proof ideas are presented in Sec. V, with detailed proofs in Appendix C.

Again, we highlight that this result does not follow trivially from classical lower bounds in adaptive data analysis since it is not clear *a priori* whether the adaptive attack using general queries can be mapped to the structured queries we consider. Moreover, our quantum lower bound for Pauli observables implies a new classical lower bound for parity statistical queries (see Corollary C1 in Appendix C 1).

The $\mathcal{O}(\sqrt{M})$ upper bound is achieved by applying a computationally efficient classical adaptive data analysis algorithm [17] to the quantum machine learning (ML) algorithm from [47]. The $\mathcal{O}(n)$ upper bound uses the nonadaptive quantum ML algorithm [47] to predict all 4^n Pauli observables. As our main focus is on the M scaling, we leave open the problem of closing the gap between $\mathcal{O}(n)$ and $\Omega(n^{1/3})$.

C. Bounded-Frobenius-norm observables

We define bounded-Frobenius-norm observables as those observables O satisfying $\mathrm{tr}(O^2) \le B$, where $B = \mathcal{O}(1)$. In this case, we demonstrate that M adaptively chosen observables can be predicted with only $\mathcal{O}(\log(M))$ samples.

Theorem 3 (Bounded-Frobenius-norm observables). For any sequence of adaptively chosen observables $O_1,\ldots,O_M$ satisfying $\mathrm{tr}(O_i^2) = \mathcal{O}(1)$ for all i, there exists a (computationally inefficient) mechanism using $N = \tilde{\mathcal{O}}(\log M/\epsilon^4)$ samples that can predict $\mathrm{tr}(O_1\rho),\ldots,\mathrm{tr}(O_M\rho)$ to ϵ additive error.

Our algorithm builds upon existing shadow tomography protocols [7,14], incorporating two key components: a threshold search subroutine [48] and a prediction subroutine. These components operate synergistically: the prediction subroutine provides estimates for the expectation values, while the threshold search subroutine validates these predictions against true values $\mathrm{tr}(O\rho)$. When predictions deviate significantly from true values, the prediction subroutine refines its model. We implement these components with novel algorithms (detailed in Appendix D) that are applicable to any observables and provide robustness against adaptivity.

Our key insight is that predicting expectations of bounded-Frobenius-norm observables is equivalent to predicting an adaptively chosen observable after "projecting" onto an unknown low-dimensional subspace determined by the unknown state ρ. By learning to predict observables in this subspace, we achieve a system-size-independent $\mathcal{O}(\log M)$ sample complexity, improving upon the best-known adaptive shadow tomography protocol, which has a sample complexity of $\tilde{\mathcal{O}}(n\log^2 M/\epsilon^4)$ [14]. Proof ideas are presented in Sec. V, with full proofs in Appendix E. For low-rank observables, we can further improve the sample complexity.

Theorem 4 (Low-rank observables). For any sequence of adaptively chosen observables $O_1,\ldots,O_M$ with rank at most R, there exists a (computationally inefficient) mechanism using $N = \tilde{\mathcal{O}}(R^2\log M/\epsilon^3)$ samples that can predict $\mathrm{tr}(O_1\rho),\ldots,\mathrm{tr}(O_M\rho)$ to ϵ additive error.

The proof for this theorem is provided in Appendix E.

IV. DISCUSSION

The classical shadow formalism [8] provides an experimentally feasible alternative to full quantum state tomography that allows one to predict many properties of an unknown quantum state from very few samples of the state. However, its rigorous guarantees require that these properties are chosen nonadaptively, whereas adaptivity is a natural component of scientific research. This work makes progress in understanding adaptivity in analyzing data from quantum experiments. We show that obtaining similar sample complexity guarantees as the classical shadow formalism in the adaptive setting is difficult even in highly restricted cases. For local and Pauli observables, we prove that it is not possible in general to prevent an adaptive attack using $\mathcal{O}(\log M)$ samples, where M is the number of queried observables. Moreover, this holds even when the system size is only polynomial in M. However, in the case of bounded-Frobenius-norm observables, we show that it becomes possible to predict many properties with few samples, maintaining the same scaling as in the classical shadow formalism with respect to M. We remark that other works such as Refs. [13,14] have previously suggested a connection between adaptive data analysis and shadow

tomography. Our work provides extensive evidence of this connection, demonstrating that results from adaptive data analysis can be used to prove new sample complexity and impossibility theorems for shadow tomography, and vice versa.

Despite the progress made in this work, many questions remain open. For local and Pauli observables we showed that if the system size is large enough, then no algorithm can prevent an adaptive attack. What are some other families of observables, where adaptive queries can lead to erroneous predictions? While we were able to design a sample-efficient algorithm for predicting adaptively chosen observables with bounded-Frobenius-norm, our algorithm is computationally inefficient. The main bottleneck in our algorithm is a projection onto a low-dimensional subspace. Are there computationally efficient algorithms for predicting bounded-Frobenius-norm observables with a sample complexity of poly $\log(M)$? Or is there a computational lower bound that cannot be surpassed? We remark that there are well-known computational hardness conjectures in classical adaptive data analysis, where any algorithm with a $N = \text{poly}\log(M)$ sample complexity is expected to be computationally inefficient. However, bounded-Frobenius-norm observables are sufficiently structured such that assuming these computational hardness conjectures does not yet rule out the presence of computationally efficient quantum algorithms. We also remark that if one allows system size dependence $N = \text{poly}(n, \log M)$, then the algorithms in Refs. [49,50] achieve a computationally efficient algorithm for low-rank observables.

V. METHODS

In this section, we describe the key ideas behind the proofs of our results. The lower bound proofs for local and Pauli observables are provided in Appendixes B and C. Both proofs leverage interactive fingerprinting codes [46] (see Appendix A 3 for an overview) to establish their guarantees, a technique also employed by Ref. [32] in classical adaptive data analysis. To prove upper bounds for bounded-Frobenius-norm observables, we first introduce a new threshold search algorithm and a shadow tomography algorithm based on classical shadows in Appendix D. Subsequently, in Appendix E, we present our main algorithm, which integrates these two algorithms as the major subroutines, and prove that it achieves the desired $\log M$ upper bound. We give an overview in the following sections.

A. Lower bounds for local and Pauli observables

We establish the lower bounds in Theorems 1 and 2 by reducing them to the guarantees of IFPCs [46]. While the relationship between classical adaptive data analysis [15,32,45] and fingerprinting codes is well-established, its extension to quantum adaptive data analysis remains largely unexplored, particularly when the queried observables are constrained to specific classes. Our primary technical contribution in this section is demonstrating that the requisite IFPC conditions can be satisfied within the quantum interaction model, even when the analyst is limited to querying only local and Pauli observables.

It is worth noting that our techniques can be adapted to derive new classical lower bounds for analogous problems in classical adaptive data analysis (see Corollaries B1 and C1 in Appendixes B 3 and C 1, respectively). This underscores the broader applicability of our approach beyond the quantum realm.

Interactive fingerprinting codes serve as a crucial component in establishing lower bounds for classical adaptive data analysis. Our proofs for Theorems 1 and 2 involve embedding IFPCs into a quantum interaction game with local and Pauli observables. The idea is to reduce our problem to IFPC by designing an analyst such that if the mechanism answers all of the analyst's queries accurately, then it contradicts the definition of an IFPC. To elucidate the underlying principles of IFPCs, we provide a concise overview (for a more comprehensive treatment, see Appendix A 3).

Consider a scenario with d users, some of whom form a colluding group creating illegal content. The primary objective of an IFPC is to embed watermarks into the content with sufficient sophistication to enable tracing of any illegally created material back to its source. IFPCs function within an interactive game involving an adversary $\mathcal{P}$ (representing the colluding group) and a fingerprinting code $\mathcal{F}$. In this setup, $\mathcal{F}$ distributes watermarks to all users, while $\mathcal{P}$ aims to generate illegal content without getting caught. The game unfolds over multiple rounds. In each round, $\mathcal{F}$ examines the illegal content and identifies potential members of the colluding group. A fingerprinting code is deemed *collusion resilient* if it can accurately identify colluding users while avoiding false accusations. The IFPC guarantee asserts the existence of a collusion-resilient interactive fingerprinting code capable of adaptively tracing a colluding group of size N within $\mathcal{O}(N^2)$ rounds.

Let us revisit the interaction model between an analyst and a mechanism, where the mechanism exclusively has access to N copies of the unknown quantum state ρ, and the analyst queries observables for their expectation values. In this scenario, the mechanism strives to respond to the analyst's queries without divulging excessive information about the samples, while the analyst attempts to adaptively extract this information to force the mechanism into generating incorrect answers.

At a high level, the analyst utilizes the IFPC algorithm of $\mathcal{F}$ to force the mechanism to predict inaccurate values. We construct a quantum state where "users" are embedded as computational basis states. The quantum mechanism

aims to hide the computational basis states in the mixture, just as the adversary $\mathcal{P}$ aims to hide colluding users. By enforcing the same restrictions on the mechanism as the adversary $\mathcal{P}$, the IFPC guarantees apply so that there exists a strategy for identifying the computational basis states in $\mathcal{O}(N^2)$ adaptive queries to local and Pauli observables. Then, the analyst can force the mechanism to fail by choosing queries that overfit to the samples. More specifically, the parallel between this model and the IFPC setting can be drawn as follows:

(a) The analyst assumes the role of the fingerprinting code $\mathcal{F}$, selecting queries strategically.
(b) The mechanism corresponds to the adversary $\mathcal{P}$ in the IFPC game.
(c) The colluding users are encoded into the unknown quantum state ρ.
(d) The queried observables correspond to the watermarks in the fingerprinting code.

Provided that the conditions of the IFPC game are maintained (i.e., the adversary $\mathcal{P}$, and by extension the mechanism, only accesses information from nonaccused members of the colluding group), the collusion-resilient guarantees of the IFPC remain valid. Consequently, the mechanism can be compelled to provide inaccurate answers after $\mathcal{O}(N^2)$ queries. This directly aligns with our objective in proving the sample complexity lower bound. Therefore, our task reduces to constructing the density matrix ρ and a strategy for the analyst in a manner that satisfies the conditions of the IFPC game.

The IFPC game requires the following two conditions to hold, which we refer to as the IFPC conditions.

(1) The fingerprinting code $\mathcal{F}$ (i.e., the analyst) must have the capability to query any watermarking assignment from the set $\{-1, 1\}^d$, where each user is assigned either -1 or 1.
(2) The adversary $\mathcal{P}$ (i.e., the mechanism) can only access the watermarks of nonaccused users in the colluding group.

To satisfy these conditions in our quantum setting, we construct ρ as a mixture of $d = 2000N$ computational basis states, each representing a user. The analyst's objective is to recover the computational basis states in the sample. This is achieved by measuring specific qubits in the Z basis, which deterministically assigns each user a value of -1 or 1, effectively serving as the watermark. To fulfill the first IFPC condition, we consider the following. For the local observables case:

(a) The analyst queries observables from the set $\{Z_i : i \in [n]\}$, where Z_i is the Pauli Z observable acting solely on the ith qubit.
(b) The number of qubits scales as 2^d, with each qubit encoding one possible watermarking assignment from $\{-1, 1\}^d$.

For the Pauli observables case:

(a) The analyst queries observables from the set $Z_b \in \{I, Z\}^{\otimes n}$, where $b \in \{0, 1\}^n$ indicates which qubits are acted on by the Pauli-Z operator.
(b) The number of qubits scales as $\mathcal{O}(d)$, which significantly reduced the system size needed to encode all watermarking assignments.

To satisfy the second IFPC condition, we employ distinct strategies for local and Pauli observables. For local observables, we implement randomization in the selection of ρ. This randomization conceals information from the mechanism, ensuring it only accesses data from nonaccused members of the colluding group. For Pauli observables, we incorporate an encryption scheme. This encryption similarly restricts the mechanism's access to information from nonaccused colluding group members.

For both cases, we allocate new qubits to encode watermarking assignments for each of the M rounds. This crucial step ensures that information remains hidden across different rounds of the interaction. Consequently, we introduce an additional factor of M in the number of qubits required for both the local and Pauli observables lower bounds. These strategies collectively guarantee that the mechanism's access to information aligns with the IFPC game's constraints, thereby preserving the integrity of our quantum implementation of the IFPC framework.

With these conditions established, an analyst using the fingerprinting code $\mathcal{F}$ in the IFPC game inherits its guarantees, forcing any mechanism to fail rapidly. The interaction proceeds as follows:

(a) In round $j \in [M]$ of the interaction, the analyst queries observables Z_x [51], where x corresponds to the watermarks sent by $\mathcal{F}$ to the adversary $\mathcal{P}$ in round j of the IFPC game.
(b) Through adaptive querying of Z_x observables, the analyst accumulates information, enabling the recovery of computational basis states in the sample provided to the mechanism.
(c) Upon gathering sufficient information about the sample, the analyst forces the mechanism to produce inaccurate responses. This strategy bears similarity to the adaptive attack in Sec. II A.

The IFPC guarantees ensure that the analyst can force any mechanism to respond inaccurately within $\mathcal{O}(N^2)$ queries. This directly yields our sample complexity lower bound of $N = \Omega(\sqrt{M})$, establishing a fundamental limit on the efficiency of quantum mechanisms. The IFPC conditions

also yield the requisite system size for our lower bounds. For local observables, the required number of qubits is $n = \Omega(M2^{\sqrt{M}})$, which scales exponentially in the number of observables queried M. Meanwhile, for Pauli observables, the system size is much smaller, i.e., $n = \Omega(M^{3/2})$.

B. Adaptive algorithms for classical shadows

We establish rigorous guarantees for new adaptivity-resistant algorithms using the classical shadow formalism [8] for shadow tomography and threshold search, which we black-box apply as subroutines to our main algorithm for predicting bounded-Frobenius-norm observables. While quantum threshold search [14] and Aaronson's shadow tomography [7] may also be substituted to yield system-size-independent upper bounds, we achieve improved scaling by integrating classical shadows [8] with classical adaptive data analysis tools (see Remark E1). Moreover, these algorithms achieve sample complexity bounds that scale better with certain parameters, so we believe them to be of independent interest.

By combining classical shadows and the private multiplicative weights algorithm [31], we derive a shadow tomography algorithm requiring

$$N_{ST} = \tilde{\mathcal{O}}\left(\frac{\sqrt{m}\cdot\log M}{\epsilon^3}\cdot\max_{i\in[M]}\|O_i\|^3_{\text{shadow}}\right) \tag{2}$$

samples. Here, m is the number of bits needed for the classical shadow description and $\|O_i\|_{\text{shadow}}$ is the shadow norm of observable O_i [8]. For $m = \mathcal{O}(n^2)$ (applicable to local and bounded-Frobenius-norm observables), this improves upon the best-known scaling [14] of $\tilde{\mathcal{O}}(n\log^2 M/\epsilon^4)$ by a factor of $\log M/\epsilon$, at the cost of the shadow norm term. Notably, the shadow norm is constant for bounded-Frobenius-norm observables.

We develop a new threshold search algorithm by incorporating the sparse vector algorithm [52] from classical adaptive data analysis and introducing a preprocessing step for classical shadows. We also employ a new variant of classical shadows derived from uniform POVM measurements [53], which offers enhanced concentration properties, leading to improved ϵ scaling.

The main new feature of our threshold search algorithm is its "strong composition" property, enabling sublinear sample complexity growth when adaptively combining multiple instances of threshold search. This contrasts with previous quantum threshold search algorithms, such as Ref. [14], which compose linearly. In those algorithms, each instance runs until an above-threshold event occurs, necessitating fresh copies of ρ for each new instance. Consequently, for a budget of ℓ above-threshold events (termed "mistakes" in Ref. [14]), the sample complexity includes a linear factor of ℓ.

In contrast, classical adaptive data analysis algorithms, particularly the sparse vector algorithm, leverage differential privacy to achieve a *quadratic* improvement in composing multiple instances, introducing only a factor of $\sqrt{\ell}$. By utilizing classical shadows, our threshold search algorithm inherits this property, requiring

$$N_{TS} = \tilde{\mathcal{O}}\left(\frac{\sqrt{\ell}\log M}{\epsilon^2}\cdot\max_i\sqrt{\text{tr}(O_i^2)}\right) \tag{3}$$

samples. This improves over the best-known quantum threshold search algorithm [14], which has a sample complexity of $\mathcal{O}(\ell\log^2 M/\epsilon^2)$. Our approach reduces the sample complexity by a factor of $\sqrt{\ell}\log M$, albeit at the cost of introducing a bound on the Frobenius-norm.

C. Upper bound for bounded-Frobenius-norm observables

We leverage our new shadow tomography and threshold search algorithms to design an improved algorithm for predicting bounded-Frobenius-norm observables. To build intuition, we first consider the case of single-rank observables before extending to bounded-Frobenius-norm observables. We adopt the terminology of "student" and "teacher" from Ref. [14] to describe the two main components of the mechanism. For any query $|\psi\rangle\langle\psi|$, the student predicts $\hat{o}$ for $\langle\psi|\rho|\psi\rangle$, while the teacher verifies this prediction. If $\hat{o}$ is ϵ-close to $\langle\psi|\rho|\psi\rangle$, the teacher outputs "pass"; otherwise, it declares a "mistake" and provides the correct value.

In our algorithm, the student and teacher use the shadow tomography and threshold search algorithms, respectively. This is different from Aaronson's shadow tomography protocol [7], which uses an online learning algorithm [54,55] to implement the student. In Aaronson's protocol, the student learns a full $2^n\times 2^n$ representation of the n-qubit quantum state ρ, which introduces the dimension dependence in its sample complexity. Our approach replaces the student with an algorithm that learns a low-dimensional representation of ρ via iterative dimension expansion. Here, the student only needs to run a low-dimensional instance of shadow tomography to learn this low-dimensional representation, avoiding the dimension dependence. This is sufficient for predicting expectation values of bounded-Frobenius-norm (and low rank) observables and enables learning with few samples.

We make two key observations:

(1) Given any quantum state ρ, there exists a low-dimensional subspace S such that for all pure state $|\psi\rangle$ and $|\psi_S\rangle$ (the projection of $|\psi\rangle$ to S),

$$|\langle\psi|\rho|\psi\rangle - \langle\psi_S|\rho|\psi_S\rangle| \le \epsilon. \tag{4}$$

(2) Given S, there is a low-dimensional quantum state $\hat{\rho}$ such that

$$\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle = \langle\psi_S|\,\rho\,|\psi_S\rangle\,. \tag{5}$$

Note that $\hat{\rho}$ can indeed be constructed to satisfy the required properties of a quantum state (i.e., Hermitian, positive semi-definite, and unit trace).

In light of these two key observations, our algorithm tasks the student with learning (1) the subspace S and (2) an approximation of the low-dimensional quantum state $\hat{\rho}$ (which is used to predict $\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle$). At a high level, the student runs a low-dimensional instance of shadow tomography (using our new adaptive shadow tomography algorithm) in order to learn $\hat{\rho}$ and produce predictions. At the same time, the student maintains a running guess of S. Every time the teacher declares "mistake" on the student's prediction, the student will apply a correction on S. Notably, the teacher only plays an *indirect* role in learning $\hat{\rho}$ (in the sense that a "mistake" declaration by this teacher may not necessarily trigger an update to the representation of $\hat{\rho}$) [56]. This is another key difference with Aaronson's protocol, where the teacher in their algorithm directly facilitates learning the full quantum state representation ρ.

We outline the steps in the algorithm below:

(1) For each received query $|\psi\rangle$, the student projects $|\psi\rangle$ into S and uses the new shadow tomography protocol to predict $\hat{o} = \langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle$.
(2) The student asks the teacher (the new threshold search algorithm) to check if $|\hat{o} - \langle\psi|\,\rho\,|\psi\rangle\,| < \epsilon$.
(3) If the teacher declares a mistake, the teacher will produce the correct expectation value o' using classical shadows.
(4) In the case that the teacher declares a mistake, the student will apply a correction by adding the query $|\psi\rangle$ to S, where the new subspace is now defined as $\mathrm{span}(S \cup \{|\psi\rangle\})$.

The key result to showing that the sample complexity is dimension-independent is the following lemma.

Lemma 1 (Mistake bound). A student algorithm that correctly predicts $\hat{o} = \langle\psi_S|\,\rho\,|\psi_S\rangle$ will make at most $\mathcal{O}(1/\epsilon^2)$ mistakes.

Intuitively, if a mistake is declared on $|\psi\rangle$, we show that the component of $|\psi\rangle$ orthogonal to S has expectation value $\Omega(\epsilon^2)$ with respect to ρ. However, there can exist at most $\mathcal{O}(1/\epsilon^2)$ such states, implying the lemma. The full proof is presented in Appendix E 1.

Using this lemma, one can show that since the number of mistakes is not very large, the threshold search algorithm will not need to make too many corrections (i.e., the budget parameter ℓ is not too large) and can thus be implemented sample-efficiently. Together with the sample complexity bounds in Eqs. (5.1) and (5.2) this yields that

$$N = \tilde{\mathcal{O}}\left(\frac{\log M}{\epsilon^3}\right) \tag{6}$$

samples are sufficient to predict single-rank observables.

To extend to bounded-Frobenius-norm and low-rank observables, we only need to make the following modifications: (1) define a processing step that "projects" the queried observable O to S, and (2) when a mistake is made, add every eigenstate of O with a sufficiently large eigenvalue (in magnitude) to S. In the case of observables with Frobenius norm $\mathrm{Tr}(O^2) \leq B$, the number of mistakes is $\mathcal{O}(B^2/\epsilon^4)$, yielding the corresponding increase in the sample complexity stated in Theorem 3. For low-rank observables, the number of mistakes is only $\mathcal{O}(R^2/\epsilon^2)$, in which case we maintain the same ϵ dependence as for single-rank observables.

ACKNOWLEDGMENTS

The authors thank Yu Tong for valuable and inspiring discussions. The authors also thank Ruohan Shen, Haimeng Zhao, Mehdi Soleimanifar, Tai-Hsuan Yang, Yiyi Cai, Charles Cao, Nadine Meister, and Chris Pattison for insightful comments and feedback. J.H. is supported by a Caltech Summer Undergraduate Fellowship and a Saul and Joan Cogen Memorial SURF fund. L.L. is supported by a Mellon Mays Undergraduate Fellowship and a Marshall Scholarship. H.H. was supported by a Google PhD fellowship and a MediaTek Research Young Scholarship. H.H. acknowledges the visiting associate position at the Massachusetts Institute of Technology. J.P. acknowledges support from the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (DE-SC0025535, DE-SC0025572), the U.S. Department of Energy Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator, and the National Science Foundation (PHY-2317110). This work was done (in part) while a subset of the authors visited the Simons Institute for the Theory of Computing. The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

H.H. and J.P. conceived the project. J.H. led the development of the mathematical proofs with contributions from L.L. and H.H. All authors contributed to the writing of the manuscript.

DATA AVAILABILITY

The data that support the findings of this article are openly available [57].

APPENDIX A: PRELIMINARIES

In Appendix A 1, we formally define the interaction models used in classical adaptive data analysis (Appendix A 1 a) and our setting of quantum adaptive data analysis (Appendix A 1 b).

In Appendix A 2, we review several previous nonadaptive methods for predicting expectation values with favorable sample complexity. Specifically, in Appendix A 2 a, we review the classical shadow formalism [8]. Classical shadows are present in many of our algorithms to achieve sample complexity upper bounds, so it is helpful to understand this formalism. We also review an algorithm [47] for predicting expectation values of nonadaptively chosen Pauli observables in Appendix A 2 b, which we adapt in one of our sample complexity upper bounds. In Appendix A 2 c, we also describe a different version of classical shadows in which the shadow is created via independent measurements of the uniform POVM [53] rather than the random measurement scheme in Ref. [8].

In addition, we review tools that will be useful throughout our proofs. In Appendix A 3, we describe fingerprinting codes [58], which are utilized heavily in our proofs of the lower bounds. Another tool used in our lower bounds is private-key cryptography, which we review in Appendix A 4.

Throughout the Appendixes, we use the notation $[m]$ for some $m \in \mathbb{N}$ to denote $[m] \triangleq \{1, \ldots, m\}$. We also assume that the spectral norm of the observables we want to predict is bounded by one: $\|O\|_\infty \leq 1$.

1. Interaction models

a. Statistical query model

We first introduce a classical model common in the adaptive data analysis literature, which we will refer to as the adaptive statistical query model [15]. In the next section, we will establish an analog in the quantum setting. Suppose there is an unknown distribution $\mathcal{D}$ over a data universe $\mathcal{X}$ with dimension n. The goal is to design an algorithm to answer *statistical queries* [59,60], which are specified by a function $q : \mathcal{X} \to [0, 1]$. The true value of the query q is denoted as

$$q(\mathcal{D}) \triangleq \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]. \tag{A1}$$

We would like to estimate the true expectation value $q(\mathcal{D})$ as accurately as possible, but we typically do not have direct access to $\mathcal{D}$ and are instead granted access to a sample $\mathcal{S} = (x_1, \ldots, x_N) \sim \mathcal{D}^N$, i.e., each $x_i \in \mathcal{S}$ is sampled i.i.d. from $\mathcal{D}$. The most straightforward way to estimate $q(\mathcal{D})$ is to compute the empirical mean $q(\mathcal{S}) = 1/N \sum_{x_i\in\mathcal{S}} q(x_i)$. Note that statistical queries must satisfy

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}^N}[q(\mathcal{S})] = q(\mathcal{D}). \tag{A2}$$

However, an algorithm simply outputting $q(\mathcal{S})$ may fail after only a few queries if the queries are selected adaptively (e.g., see the feature selection attack in Ref. [16] and Lecture 3 of Ref. [61]). In general, we will need to design a *mechanism* that outputs a response based on $\mathcal{S}$ to protect against adaptivity.

Formally, the interaction can be described as a game between an analyst $\mathcal{A}$ and mechanism $\mathcal{M}$. In general, the analyst and mechanism do not have access to the exact true distribution $\mathcal{D}$, and only the mechanism has access to the sample $\mathcal{S}$. $\mathcal{A}$ and $\mathcal{M}$ interact over k rounds, where in each round j, (1) $\mathcal{A}$ asks query q^j and (2) $\mathcal{M}$ uses $\mathcal{S}$ to generate a response $a^j \in [0, 1]$. Here, the analyst's access to the sample $\mathcal{S}$ in the form of answers to statistical queries is the same as that of statistical query learning [59]. However, in this case, $\mathcal{A}$ and $\mathcal{M}$ can depend on prior history. Specifically, the analyst can choose queries *adaptively* based on the transcript of prior interactions $\mathcal{T}^j = (q^1, a^1, \ldots, q^{j-1}, a^{j-1})$. This is in contrast to the typical setting in which the queries q^j are specified beforehand. Again, the goal is to design a mechanism that can answer these adaptive queries correctly, where we characterize its accuracy as follows.

Definition A1. A mechanism $\mathcal{M}$ is (ϵ, δ)-*accurate* for k queries $q^1, \ldots, q^k$ if for every analyst $\mathcal{A}$ and distribution $\mathcal{D}$,

$$\Pr\left[\max_{j\in[k]} |q^j(\mathcal{D}) - a^j| \leq \epsilon\right] \geq 1 - \delta, \tag{A3}$$

where the probability is over the randomness of $\mathcal{M}$, $\mathcal{A}$, and the dataset $S \sim \mathcal{D}^N$. Here, a^j denotes the mechanism $\mathcal{M}$'s response to the query q^j.

b. Quantum interaction model

With this motivation from classical adaptive data analysis, we now discuss the interaction model in the quantum setting. Consider a game between a classical analyst $\mathcal{A}$ and a quantum mechanism $\mathcal{M}$. The interaction can be described as follows. Suppose ρ is an n-qubit unknown quantum state and that $\mathcal{M}$ has access to N copies $\rho^{\otimes N}$. First, $\mathcal{M}$ applies an initial unitary U_0 to $\rho^{\otimes N}$ and a set of ancilla qubits,

$$\mathcal{W}^0(\rho^{\otimes N}) = U_0(\rho^{\otimes N} \otimes |0^k\rangle\langle 0^k|)U_0^\dagger, \tag{A4}$$

where $\mathcal{W}^j(\rho^{\otimes N})$ denotes the mechanism's quantum state at round j given access to the samples $\rho^{\otimes N}$ stored in quantum memory. This executes any preprocessing on the input data. After this initial preprocessing, the analyst and the mechanism will alternate turns. The analyst adaptively generates a classical query q^j at round j (e.g., a classical description of an observable O). Sometimes, we will also denote the query as o^j to align with the

ALGORITHM 1. Quantum interaction model.

Let ρ be an n-qubit density matrix and M be the number of properties to be tested.
Mechanism obtains samples $\rho^{\otimes N}$.
for $j = 1$ *to* $M(N)$ **do**
  Let $\mathcal{T}^j$ be the transcript of prior interactions $\mathcal{T}^j = (q^1, a^1, ..., q^{j-1}, a^{j-1})$.
  Analyst $\mathcal{A}(\mathcal{T}^j)$ generates query q^j corresponding to a property to be tested such that $q^j(\rho) \in [-1, 1]$.
  Mechanism $\mathcal{M}(\mathcal{T}^j, q^j, \rho^{\otimes N})$ generates response $a^j \in [-1, 1]$.
end

notation of Ref. [8]. Moreover, we assume throughout the work that the observable O corresponding to this query has $\|O\|_\infty \leq 1$, an assumption that is also present in Ref. [8]. This query can be viewed as an additional input in the mechanism's quantum state. Then, the mechanism responds to the query by performing a unitary on its current state, measuring some of the qubits, and returning the classical output a^j to the analyst. This interaction is described at a high level in Algorithm 1. More formally, let $\mathcal{T}^j$ be the transcript of prior interactions up to round j, i.e., $\mathcal{T}^j = (q^1, a^1, \ldots, q^{j-1}, a^{j-1})$. For each round $j = 1, \ldots, M$,

(1) $\mathcal{A}$ generates query q^j based on $\mathcal{T}^j$ and gives it to $\mathcal{M}$.
(2) $\mathcal{M}$ takes q^j as input and applies a unitary U_j to its current quantum state $\mathcal{W}^{j-1}$ before performing a partial measurement $\Pi_j = |a^j\rangle\langle a^j| \otimes I$. Here, the mechanism only measures some subset of the qubits so that the measurement operator is identity on the remaining unmeasured qubits. The classical outcome a^j from this measurement is observed and given to $\mathcal{A}$ with probability

$$\mathrm{tr}(\Pi_j (U_j (\mathcal{W}^{j-1} \otimes |q^j\rangle\langle q^j|) U_j^\dagger)). \tag{A5}$$

The mechanism's post-measurement state is

$$\mathcal{W}^j(\rho^{\otimes N}) \triangleq \frac{\Pi_j U_j (\mathcal{W}^{j-1} \otimes |q^j\rangle\langle q^j|) U_j^\dagger \Pi_j^\dagger}{\mathrm{tr}(\Pi_j (U_j (\mathcal{W}^{j-1} \otimes |q^j\rangle\langle q^j|) U_j^\dagger))}. \tag{A6}$$

We now discuss what it means for a mechanism to be robust against any adaptive analyst (or for an analyst to force the mechanism to fail). We first note that for any quantum mechanism $\mathcal{M}$ and classical analyst $\mathcal{A}$, the interaction can be encoded in the following state:

$$\sigma \triangleq \sum_{\substack{q^1,\ldots,q^M \\ a^1,\ldots,a^M}} p_{\mathcal{A}}(q_1|\mathcal{T}^1) \cdots p_{\mathcal{A}}(q_M|\mathcal{T}^M) |\mathcal{T}^{M+1}\rangle\langle\mathcal{T}^{M+1}|, \tag{A7}$$

where $p_{\mathcal{A}}(q_j|\mathcal{T}^j)$ denotes the probability that the analyst $\mathcal{A}$ selects query q^j conditioned on the past history of interactions. We can define an observable O_ρ dependent on ρ that encodes a mechanism's error in a given interaction, i.e.,

$$\mathrm{tr}(O_\rho |\mathcal{T}^{M+1}\rangle\langle\mathcal{T}^{M+1}|) = \max_{j\in[M]} |q_j(\rho) - a^j| \tag{A8}$$

is the mechanism's maximum error under transcript $\mathcal{T}^{M+1}$. So the expected error of the mechanism is $\mathrm{tr}(O_\rho \sigma)$. Denote

$$P^{\mathcal{A}}_{\mathcal{T}^{M+1}} \triangleq \left(\prod_{j=1}^{M} p_{\mathcal{A}}(q_j|\mathcal{T}^j) \right) \tag{A9}$$

as the probability of transcript $\mathcal{T}^{M+1}$ occurring. We say that a mechanism can always answer M adaptive queries to ϵ error with probability $1 - \delta$ if for any analyst $\mathcal{A}$,

$$\sum_{\mathcal{T}:\mathrm{tr}(O_\rho|\mathcal{T}\rangle\langle\mathcal{T}|)\leq\epsilon} P^{\mathcal{A}}_{\mathcal{T}} \geq 1 - \delta. \tag{A10}$$

2. Review of previous nonadaptive methods

In this Appendix, we provide a brief overview of previous work [8,47,53] that can be used to predict expectation values of nonadaptively chosen observables. In particular, we review the classical shadow formalism (which can predict expectation values of, e.g., local observables and observables with bounded Frobenius norm) in Appendix A 2 a, an algorithm for predicting expectation values of Pauli observables in Appendix A 2 b, and a version of the classical shadow formalism using measurements of the uniform POVM in Appendix A 2 c. As we discuss later, all of these methods implicitly assume that the observables are chosen nonadaptively. We review them here because we build on them in our new algorithms.

a. Classical shadow formalism

In this Appendix, we briefly review the classical shadow formalism with an emphasis on aspects that are important in our proofs. For a more detailed presentation, we direct the reader to Ref. [8]. We note that there exist several other techniques for constructing efficient classical representations of quantum systems to predict properties [9,62–66], but we focus on the classical shadow formalism in this work.

Recall that the goal of the classical shadow formalism is to construct a succinct classical representation of an

unknown n-qubit quantum state ρ that suffices to predict properties $\mathrm{tr}(O_i\rho)$ for observables $O_1, \ldots, O_M$. One can construct this classical representation, called the *classical shadow*, by applying a unitary transformation $\rho \mapsto U\rho U^\dagger$, measuring all qubits in the computational basis to obtain an outcome $|\hat{b}\rangle$ for $\hat{b} \in \{0,1\}^n$, and repeating this process several times. Here, U is a unitary selected at random from a fixed ensemble of unitaries $\mathcal{U}$. Different choices of $\mathcal{U}$ lead to different guarantees, as we discuss later. After measuring in the computational basis, we can apply the inverse of U to the outcome state to obtain $U^\dagger|\hat{b}\rangle\langle\hat{b}|U$. In expectation (over the random choice of unitary and the measurement), this quantity contains can be viewed as a quantum channel applied to ρ,

$$\mathcal{M}(\rho) = \mathbb{E}\left[U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right]. \tag{A11}$$

Thus, applying the inverse of this quantum channel to our classical data $U^\dagger|\hat{b}\rangle\langle\hat{b}|U$ gives us an approximation of ρ,

$$\hat{\rho} = \mathcal{M}^{-1}\left(U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right). \tag{A12}$$

The classical shadow representation is then a collection of such $\hat{\rho}$ after repeating this measurement procedure several times,

$$\mathsf{S}(\rho; T) \triangleq \{\hat{\rho}_1, \ldots, \hat{\rho}_T\}. \tag{A13}$$

We can then predict properties $\mathrm{tr}(O_i\rho)$ using the classical shadow via median-of-means [67,68]. Specifically, one can split the classical shadow into K equally sized parts, take the empirical mean of each of these parts to obtain $\hat{\rho}_{(k)}$, compute the property $\mathrm{tr}(O_i\hat{\rho}_{(k)})$ for each of these empirical means, and then take median $\mathrm{median}(\mathrm{tr}(O_i\hat{\rho}_{(1)}), \ldots, \mathrm{tr}(O_i\hat{\rho}_{(K)}))$ of the results. For this simple procedure, the classical shadow formalism achieves the following rigorous guarantee for predicting expectation values.

Theorem A1 (Theorem 1 in Ref. [8]). Let $O_1, \ldots, O_M$ be Hermitian $2^n \times 2^n$ matrices, and let $\epsilon, \delta \in [0,1]$. Then,

$$N = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^2} \max_{1\le i\le M} \left\| O_i - \frac{\mathrm{tr}(O_i)}{2^n}\mathbb{I}\right\|^2_{\mathrm{shadow}}\right) \tag{A14}$$

copies of an unknown quantum state ρ suffice to predict $\hat{o}_i$ such that

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \le \epsilon \tag{A15}$$

for all $1 \le i \le M$, with probability at least $1-\delta$.

Here, $\|\cdot\|_{\mathrm{shadow}}$ denotes the shadow norm, which depends on the ensemble of unitary transformations used to create the classical shadow. Explicitly, the shadow norm is given by

$$\|O\|_{\mathrm{shadow}} = \max_{\sigma:\mathrm{state}} \times \left(\mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\, U\sigma U^\dagger\,|b\rangle\, \langle b|\, U\mathcal{M}^{-1}(O)U^\dagger\,|b\rangle^2\right)^{1/2}. \tag{A16}$$

There are two examples of unitary ensembles emphasized in Ref. [8] that we also focus on here. These are the cases of tensor products of random single-qubit Pauli gates and random n-qubit Clifford circuits. Random Pauli measurements allow us to predict properties for local observables while random Clifford measurements allow us to predict properties for observables with bounded Frobenius norm $\mathrm{tr}(O^2)$. In what follows, we provide sample complexity upper and lower bounds for predicting properties with respect to local observables and observables with bounded Frobenius norm.

Moreover, in the local observable case, there is a particularly simple form for the inverse quantum channel $\mathcal{M}^{-1}$, which we make use of in the later sections. In particular, when predicting local observables, the random unitary $U = U_1 \otimes \cdots \otimes U_n$ can be written as a tensor product of random single-qubit Pauli gates. Then, the inverse quantum channel is given by

$$\hat{\rho} = \bigotimes_{j=1}^{n}\left(3U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j - \mathbb{I}\right), \tag{A17}$$

where $\hat{b}_1, \ldots, \hat{b}_n \in \{0,1\}$. Note that for local observables, the classical shadow of a quantum state can be stored efficiently in classical memory because

$$U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |+i\rangle, |-i\rangle\}. \tag{A18}$$

This is clear because each of the U_j are Pauli gates. Thus, we can represent each classical snapshot $\hat{\rho}_i$ using $\mathcal{O}(n)$ bits ($\mathcal{O}(1)$ bits to store which of the 6 Pauli eigenstates $U_j^\dagger|\hat{b}_j\rangle\langle\hat{b}_j|U_j$ is for each of the n qubits). The entire classical shadow $\mathsf{S}(\rho; T) = \{\hat{\rho}_1, \ldots, \hat{\rho}_T\}$ can then be stored in $\mathcal{O}(nT)$ bits.

b. Nonadaptively chosen Pauli observables

In this Appendix, we give an overview of the algorithm from Ref. [47]. We utilize this in Appendix C 2 to prove the sample complexity upper bound for predicting expectation values of adaptively chosen Pauli observables. For a more detailed presentation, we direct the reader to Ref. [47].

Given copies of an unknown n-qubit quantum state, the goal of the algorithm from Ref. [47] is to predict

expectation values $\mathrm{tr}(P_i\rho)$, of M different (nonadaptively chosen) n-qubit Pauli observables $P_1,\ldots,P_M$, i.e., $P_i \in \{I,X,Y,Z\}^{\otimes n}$. For this task, Ref. [47] achieves the following rigorous guarantee.

Theorem A2 (Theorem 4 in Ref. [47]). Let $\epsilon,\delta > 0$. Let $P_1,\ldots,P_M \in \{I,X,Y,Z\}^{\otimes n}$ be M n-qubit Pauli observables. Then,

$$N = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^4}\right) \tag{A19}$$

copies of an unknown quantum state ρ suffice to predict $\hat{p}_i$ such that

$$|\hat{p}_i - \mathrm{tr}(P_i\rho)| \leq \epsilon \tag{A20}$$

for all $1 \leq i \leq M$, with probability at least $1-\delta$.

The algorithm that achieves this sample complexity upper bound estimates $\mathrm{tr}(P\rho)$ in two steps. First, the algorithm estimates the magnitude $|\mathrm{tr}(P\rho)|^2$. Second, the algorithm determines $\mathrm{sign}(\mathrm{tr}(P\rho))$.

In the first step, one performs N_1 rounds of Bell measurements on $\rho\otimes\rho$. In this way, one can obtain classical measurement data $S^{(t)} = \{S_k^{(t)}\}_{k\in[n]}$, where $t \in [N_1]$ corresponds to the t'th round of measurement and $k \in [n]$ corresponds to the kth qubit. Then, the algorithm estimates the magnitude $|\mathrm{tr}(P\rho)|^2$ via the following classical processing. Let $P = \sigma_1 \otimes \cdots \otimes \sigma_n$. We can compute

$$q_P(S^{(t)}) = \prod_{k=1}^{n} \mathrm{tr}\left((\sigma_k\otimes\sigma_k)S_k^{(t)}\right) \in \{\pm 1\} \tag{A21}$$

for each $t \in [N_1]$, where we suggestively denote this quantity by q_P, similar to our query notation from Appendix A 1. Huang *et al.* [47] shows that the expectation of q_P is the value we wish to estimate

$$\mathbb{E}\left[q_P(S^{(t)})\right] = |\mathrm{tr}(P\rho)|^2. \tag{A22}$$

Thus, one can estimate this expectation value with the empirical mean

$$q_P(\{S^{(t)}\}_{t\in[N_1]}) = \frac{1}{N_1}\sum_{t=1}^{N_1} q_P(S^{(t)}). \tag{A23}$$

Then, Ref. [47] shows that if the Pauli observables $P_1,\ldots,P_M$ are selected nonadaptively, then

$$2N_1 = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^4}\right) \tag{A24}$$

copies of the unknown quantum state ρ suffice to estimate $|\mathrm{tr}(P_i\rho)|^2$ with error at most ϵ^2 for all $1 \leq i \leq M$.

In the second step to determine $\mathrm{sign}(\mathrm{tr}(P\rho))$, Ref. [47] performs a coherent measurement on N_2 copies of ρ. This measurement measures the eigenvalues of the Pauli operator P on each copy of ρ and takes the majority vote, yielding an estimate for the sign of $\mathrm{tr}(P\rho)$. For predicting M Pauli observables $P_1,\ldots,P_M$, this requires

$$N_2 = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^2}\right) \tag{A25}$$

copies of ρ in quantum memory. Moreover, the argument in Ref. [47] allows one to reuse these N_2 copies of ρ stored in quantum memory for each queried Pauli observable P_i. This is due to the quantum union bound [69] (a generalization of the gentle measurement lemma [70,71]), since performing this coherent measurement does not disturb the copies of the quantum state much so that one can perform the coherent measurement repeatedly, for each observable P_i queried. Notice that this in fact makes this step of the algorithm resistant to adaptivity.

c. Classical shadows from uniform POVM measurements

In this Appendix, we discuss a different version of the classical shadow formalism introduced in Appendix A 2 a in which the shadows are generated from independent measurements in the uniform POVM. This version was first introduced in Ref. [53]. These classical shadows from uniform POVM measurements exhibit stronger concentration properties than the canonical classical shadow formalism [8] using Clifford and Pauli measurements. This property will be important for our proof of Theorem D4 in Appendix D 2.

The classical shadow is obtained by performing the following measurement procedure: apply a Haar random unitary to the state $\rho \mapsto U\rho U^\dagger$ and then perform a computational-basis measurement. Upon receiving the n-bit measurement outcome $|\hat{b}\rangle \in \{0,1\}^n$, we can store an (inefficient) classical description of $U^\dagger|b\rangle\hat{\langle}b|U$ in classical memory. Applying an inverse quantum channel, as in the canonical classical shadow formalism [8], gives us an approximation of ρ,

$$\hat{\rho} = \mathcal{M}^{-1}\left(U^\dagger\left|\hat{b}\right\rangle\left\langle\hat{b}\right|U\right), \tag{A26}$$

where the inverted quantum channel is $\mathcal{M}^{-1}(X) = (2^n+1)X - \mathbb{I}$. Note that $\mathbb{E}[\hat{\rho}] = \rho$, where the expectation is over the random choice of unitary and the randomness in the measurement results. The classical shadow representation is then a collection of such $\hat{\rho}$ after repeating this measurement procedure several times,

$$\mathsf{S}(\rho;N) = \{\hat{\rho}_1,\ldots,\hat{\rho}_N\}. \tag{A27}$$

The procedure above is equivalent to repeatedly measuring the unknown state ρ with the uniform POVM $\{d\,|v\rangle\langle v|\,\mathrm{d}v\}$, where $\mathrm{d}v$ is the unique unitarily invariant measure on the complex unit sphere induced by the Haar measure, and applying the same inverted quantum channel to the measurement results. This is similar to the tomography methods in Ref. [72]. This is why we refer to this version of classical shadow as using uniform POVM measurements.

We can then predict properties $\mathrm{tr}(O\rho)$ using the classical shadow by simply taking the empirical mean. Specifically, we can produce an estimate $\hat{o}$ of $\mathrm{tr}(O\rho)$ as follows:

$$\hat{o} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}(O\hat{\rho}_i). \tag{A28}$$

Because $\mathbb{E}[\hat{\rho}_i] = \rho$, this estimate clearly reproduces $\mathrm{tr}(O\rho)$ in expectation. In what follows, we consider the case where $N = 1$, i.e., we take $\hat{\rho}$ in Eq. (A26) to be our classical shadow. The analysis easily generalizes to arbitrary N, but we will need the $N = 1$ case in Appendix D 2.

We now prove the key property that this classical shadows estimate for the expectation value of an observable exhibits exponential concentration.

Proposition A1. Fix an observable O and let $\hat{o} = \mathrm{tr}(O\hat{\rho})$, where $\hat{\rho}$ is an estimate of ρ as in Eq. (A26). Then for any $0 \le \tau \le 1$,

$$\Pr\big[|\hat{o} - \mathbb{E}[\hat{o}]| \ge \tau\big] \le 2\exp\left(-\frac{\tau^2}{16B + 4\sqrt{B}\tau}\right). \tag{A29}$$

In order to prove the proposition, we require the following two well-known results. First, recall a subexponential formulation of the scalar Bernstein inequality. This is a well-known result in probability (see, e.g., Lemma 5.4 in Ref. [73]).

Theorem A3 (Scalar Bernstein inequality). Let $X_1, \ldots, X_m$ be independent random variables with zero mean and variance σ_i^2 such that for all integers $k \ge 2$, there exists an $R > 0$ such that

$$\mathbb{E}|X_i|^k \le k!R^{k-2}\frac{\sigma_i^2}{2} \quad \text{for all} \quad 1 \le i \le m. \tag{A30}$$

Set $\sigma^2 = \sum_{i=1}^m \sigma_i^2$. Then, for all $\tau > 0$

$$\Pr\left[\left|\sum_{i=1}^{m} X_i\right| \ge \tau\right] \le 2\exp\left(-\frac{\tau^2/2}{\sigma^2 + R\tau}\right). \tag{A31}$$

Second, we need the following result about the moments of the uniform POVM, which follows from arguments in representation theory (see, e.g., Refs. [72,74,75]).

Theorem A4 (Moments of the uniform POVM). Fix $k \in \mathbb{N}$ and let $P_{\mathrm{Sym}^{(k)}}$ denote the projector onto the totally symmetric subspace $\mathrm{Sym}^{(k)} \subset (\mathbb{C}^d)^{\otimes k}$. Then,

$$\int_{\mathbb{S}^{d-1}} (|v\rangle\langle v|)^{\otimes k}\mathrm{d}v = \binom{d+k-1}{k}^{-1} P_{\mathrm{Sym}^{(k)}}.$$

One can show this by using the unitary invariance of $\mathrm{d}v$ to show that the left-hand side commutes with $U^{\otimes k}$ for all $d \times d$ unitaries combined with Schur's lemma.

We can use Theorem A4 to prove bounds on the moments of a random variable based on the classical shadows estimate, which will then allow us to apply Theorem A3 to obtain exponential concentration.

Lemma A1. Fix density matrix ρ of dimension d and observable O such that $\mathrm{tr}(O^2) \le B$. For $\hat{\rho}$ given in Eq. (A26), define the random variable $\bar{X} = \mathrm{tr}(O\hat{\rho}) - \mathrm{tr}(O\rho)$. Then, for $k \ge 2$,

$$\mathbb{E}[|\bar{X}|^k] \le k!(4B)^{k/2}.$$

Proof. Set

$$E = O - \frac{\mathrm{tr}(O) + \mathrm{tr}(O\rho)}{d+1}\mathbb{I}. \tag{A32}$$

Then, we can reformulate $\bar{X}$ as

$$\bar{X} = (d+1)\,\langle v|\,E\,|v\rangle \quad \text{with prob } d\,\langle v|\,\rho\,|v\rangle\,. \tag{A33}$$

We first focus on bounding even moments, i.e., k even. Using Theorem A4 to compute the expectation value,

$$\mathbb{E}[|\bar{X}|^k] = \mathbb{E}[\bar{X}^k] = \int_{\mathbb{S}^{d-1}} d\,\langle v|\,\rho\,|v\rangle\,((d+1)\,\langle v|\,E\,|v\rangle)^k\mathrm{d}v \tag{A34}$$

$$= d(d+1)^k\mathrm{tr}\left(\left(\int_{\mathbb{S}^{d-1}}(|v\rangle\langle v|)^{\otimes k+1}\right)\rho\otimes E^{\otimes k}\right) \tag{A35}$$

$$= d(d+1)^k\binom{d+k}{k+1}^{-1}\mathrm{tr}(P_{\mathrm{Sym}^{(k+1)}}\rho\otimes E^{\otimes k}). \tag{A36}$$

We bound the prefactor and the trace term separately. The prefactor can be bounded as

$$d(d+1)^k\binom{d+k}{k+1}^{-1} = \frac{(k+1)!d(d+1)^k}{(d+k)\cdots(d+1)d} \le (k+1)!$$

The trace evaluates all possible contractions of the tensor $\rho\otimes E\otimes\cdots\otimes E$, yielding the bound

$$\mathrm{tr}(P_{\mathrm{Sym}^{(k+1)}}\rho\otimes E^{\otimes k}) \le \max\{\mathrm{tr}(E), \|E\|_F\}^k. \tag{A37}$$

This inequality holds by using the definition of the projector onto the symmetric subspace, submultiplicativity of the

trace, and ρ having unit trace. We compute this explicitly for a few examples to convince the reader. Recall that, following the presentation in Ref. [76], the projection onto the symmetric subspace can be written as

$$P_{\mathrm{Sym}^{(k)}} = \frac{1}{k!}\sum_{\pi\in S_k} V_d(\pi), \tag{A38}$$

where

$$V_d(\pi) = \sum_{i_1,\ldots,i_k\in[d]^k} \left|i_{\pi^{-1}(1)},\ldots,i_{\pi^{-1}(k)}\right\rangle\langle i_1,\ldots,i_k| \tag{A39}$$

for $\pi \in S_k$ with S_k the symmetric group over k elements. Then, the inequality in Eq. (A37) is clear for $k = 1$ since $S_2 = \{(1)(2), (1\ 2)\}$,

$$\mathrm{tr}(P_{\mathrm{Sym}^{(2)}}\rho\otimes E) = \frac{1}{2}(\mathrm{tr}(\rho)\mathrm{tr}(E) + \mathrm{tr}(E\rho)) \le \mathrm{tr}(E). \tag{A40}$$

Similarly, we can also show this for $k = 2$, where $S_3 = \{(1)(2)(3), (1)(2\ 3), (3)(1\ 2), (1\ 2\ 3), (1\ 3\ 2), (2)(1\ 3)\}$,

$$\mathrm{tr}(P_{\mathrm{Sym}^{(3)}}\rho\otimes E\otimes E) \tag{A41}$$

$$= \frac{1}{6}(\mathrm{tr}(\rho)\mathrm{tr}(E)\mathrm{tr}(E) + \mathrm{tr}(\rho)\mathrm{tr}(E^2) + \mathrm{tr}(E)\mathrm{tr}(E\rho) + \mathrm{tr}(E\rho E) + \mathrm{tr}(E^2\rho) + \mathrm{tr}(E)\mathrm{tr}(E\rho)) \tag{A42}$$

$$= \frac{1}{6}(\mathrm{tr}(E)^2 + \mathrm{tr}(E^2) + 2\mathrm{tr}(E)\mathrm{tr}(E\rho) + 2\mathrm{tr}(E^2\rho)) \tag{A43}$$

$$\le \frac{1}{6}(\mathrm{tr}(E)^2 + \mathrm{tr}(E^2) + 2\mathrm{tr}(E)^2 + 2\mathrm{tr}(E^2)). \tag{A44}$$

Here, if $\mathrm{tr}(E)^2 \le \mathrm{tr}(E^2) = \|E\|_2^2$, then the above inequality is bounded by $\mathrm{tr}(E^2) = \|E\|_2^2$. If $\mathrm{tr}(E^2) \le \mathrm{tr}(E)^2$, then the above inequality is bounded by $\mathrm{tr}(E)^2$. Thus, we have that

$$\mathrm{tr}(P_{\mathrm{Sym}^{(3)}}\rho\otimes E\otimes E) \le \max(\mathrm{tr}(E), \|E\|_2^2)^2. \tag{A45}$$

Similar calculations hold for general k, so it is clear that Eq. (A37) holds.

We can further bound this quantity,

$$\mathrm{tr}(P_{\mathrm{Sym}^{(k+1)}}\rho\otimes E^{\otimes k}) \le \max\{\mathrm{tr}(E), \|E\|_F\}^k \le B^{k/2}. \tag{A46}$$

This follows from the following computations:

$$\mathrm{tr}(E) = \mathrm{tr}(O) - \frac{d}{d+1}(\mathrm{tr}(O) + \mathrm{tr}(O\rho)) \le \frac{\mathrm{tr}(O)}{d+1} - \frac{d}{d+1}\mathrm{tr}(O\rho) \le \frac{d}{d+1} + 1 \le 2. \tag{A47}$$

In the first inequality, we use the assumption that $O \le \mathbb{I}$ in PSD ordering. In the second inequality, we use that the spectral norm of O is bounded by one,

$$\|E\|_F^2 = \mathrm{tr}(E^2) \tag{A48}$$

$$= \mathrm{tr}\left(\left(O - \frac{\mathrm{tr}(O) + \mathrm{tr}(O\rho)}{d+1}\mathbb{I}\right)^2\right) \tag{A49}$$

$$= \mathrm{tr}(O^2) - 2\frac{\mathrm{tr}(O) + \mathrm{tr}(O\rho)}{d+1}\mathrm{tr}(O) + \left(\frac{\mathrm{tr}(O) + \mathrm{tr}(O\rho)}{d+1}\right)^2\mathrm{tr}(\mathbb{I}^2) \tag{A50}$$

$$= \mathrm{tr}(O^2) - \frac{d+2}{(d+1)^2}\mathrm{tr}(O)^2 - \frac{2}{(d+1)^2}\mathrm{tr}(O)\mathrm{tr}(O\rho) + \frac{d}{(d+1)^2}\mathrm{tr}(O\rho)^2 \tag{A51}$$

$$\le \mathrm{tr}(O^2) - \frac{d+2}{(d+1)^2}\mathrm{tr}(O)^2 - \frac{2}{(d+1)^2}\mathrm{tr}(O)\mathrm{tr}(O\rho) + \frac{d}{(d+1)^2}\mathrm{tr}(O)^2 \tag{A52}$$

$$= \mathrm{tr}(O^2) - \frac{2}{(d+1)^2}(\mathrm{tr}(O)^2 + \mathrm{tr}(O)\mathrm{tr}(O\rho)) \tag{A53}$$

$$\le \mathrm{tr}(O^2) - \frac{2}{(d+1)^2}\mathrm{tr}(O)\mathrm{tr}(O\rho) \tag{A54}$$

$$\le \mathrm{tr}(O^2) - \frac{2}{(d+1)^2}\mathrm{tr}(O\rho)^2 \tag{A55}$$

$$\le B, \tag{A56}$$

where we use the submultiplicative property of the trace several times. In summary,

$$\mathbb{E}[|\bar{X}|^k] \le (k+1)!B^{k/2} \quad \text{for } k = 2, 4, 6, \ldots. \tag{A57}$$

In order to bound the odd moments, we use the following trick, which allows us to convert odd moments into even moments. The function $x \mapsto x^{k/(k+1)}$ is concave on the positive reals. Jensen's inequality then implies

$$\mathbb{E}[|\bar{X}|^k] \le (\mathbb{E}[|\bar{X}|^{k+1}])^{\frac{k}{k+1}}. \tag{A58}$$

Using Eq. (A57),

$$\mathbb{E}[|\bar{X}|^k] \le \left((k+2)!B^{(k+1)/2}\right)^{\frac{k}{k+1}} \le (k+2)!B^{k/2} \tag{A59}$$

for $k \ge 3$ odd. The final claim follows from the observation:

$$(k+1)! \le (k+2)! \le 2^k k! \tag{A60}$$

■

With this result, we are now ready to prove Proposition A1, which follows by an application of Theorem A3.

Proof of Proposition A1. Let $\bar{X} = \mathrm{tr}(O\hat{\rho}) - \mathrm{tr}(O\rho)$, where $\hat{\rho}$ is a classical shadow as in Eq. (A26). Then, since $\mathbb{E}[\hat{\rho}] = \rho$, then we clearly have that $\bar{X}$ has zero mean and $\bar{X} = \hat{o} - \mathbb{E}[\hat{o}]$ when $\hat{o} = \mathrm{tr}(O\hat{\rho})$. We want to bound $\Pr\left[|\bar{X}| \geq \tau\right]$. By Lemma A1, we found that

$$\mathbb{E}[|\bar{X}|^k] \leq k!(4B)^{k/2}. \tag{A61}$$

Note that the variance of $\bar{X}$ is $8B$ by taking the second moment bound in the above inequality so that choosing $R = 2\sqrt{B}$, $\bar{X}$ satisfies the conditions for Theorem A3. Applying this, we have

$$\begin{aligned}\Pr\left[|\hat{o} - \mathbb{E}[\hat{o}]| \geq \tau\right] &= \Pr\left[|\bar{X}| \geq \tau\right] \\ &\leq 2\exp\left(-\frac{\tau^2/2}{8B + 2\sqrt{B}\tau}\right) \\ &= 2\exp\left(-\frac{\tau^2}{16B + 4\sqrt{B}\tau}\right),\end{aligned} \tag{A62}$$

as claimed. ■

3. Fingerprinting codes

Our sample complexity lower bounds for the local and Pauli observable cases rely heavily on fingerprinting codes and their guarantees.

Fingerprinting codes were originally introduced by Ref. [58] within the context of watermarking digital content to prevent piracy. We first describe the intuition behind these constructions and then give a formal definition. Suppose we are distributing content to a group of d users, but there is a size N colluding group $\mathcal{S}$ outputting illegal copies. Note here that we are suggestively using the same notation as previously introduced for the adaptive data analysis setting (i.e., where N was the number of points in the mechanism's dataset $\mathcal{S}$). Boneh and Shaw [58] show that by assigning unique "watermarks" to the content, one can trace illegal copies back to the user in $\mathcal{S}$ that produced it. Importantly, the illegal copy must be consistent with one of the watermarks given to members in $\mathcal{S}$. Fingerprinting codes are a generalization of these watermarks that can identify some user in S responsible even when the users in $\mathcal{S}$ collude to construct an illegal copy. Reference [77] gave optimal constructions for fingerprinting codes. Moreover, fingerprinting codes are central to sample complexity lower bounds in the differential privacy literature [78,79]. Interactive fingerprinting codes [46] extend this idea even further, being able to adaptively identify each of the users in $\mathcal{S}$ one by one, where constructions are given in Refs. [80,81]. Interactive fingerprinting codes were used in Ref. [32] to prove lower bounds for general adaptive data analysis problems, where the idea is to adaptively choose queries (watermarks) such that the analyst can identify every point in the dataset $\mathcal{S}$. Note that if the analyst knows $\mathcal{S}$, he then can easily overfit, e.g., by choosing a query q such that $q(x) = 1$ if $x \in \mathcal{S}$ and 0 otherwise.

With this intuition, we may now describe interactive fingerprinting codes more formally. Interactive fingerprinting codes operate according to a game between an adversary $\mathcal{P}$ and fingerprinting code $\mathcal{F}$. This game has M rounds and is presented in Algorithm 2. Again, here we are suggestively using the same notation as before (i.e., where M was the number of queries made by the analyst). During round j, the fingerprinting code $\mathcal{F}$ broadcasts to every user $i \in [d]$ a code $c_i^j \in \{0,1\}$. The colluding group $\mathcal{S} \subset [d]$ of size N produces a code a^j (for their illegal content), with the requirement that it must satisfy *consistency*: there exists a user i such that $a^j = c_i^j$. At the end of each round, the fingerprinting code $\mathcal{F}$ accuses a set of users and prevents them from seeing any future broadcasts from $\mathcal{F}$. Thus, in each round, the fingerprinting code $\mathcal{F}$ iteratively identifies users in the colluding group $\mathcal{S}$.

We give a security definition for an interactive fingerprinting code based on this game $\mathsf{IFPC}_{N,d,M}(\mathcal{P},\mathcal{F})$. Informally, we want both completeness and soundness properties: the fingerprinting code should both be able to identify the colluding users $\mathcal{S}$ and not make many false accusations. To formalize this, we define the following quantities. Let

$$\theta^j \triangleq |\{1 \leq k \leq j \mid \nexists i \in [d], a^k = c_i^k\}| \tag{A63}$$

ALGORITHM 2. $\mathsf{IFPC}_{N,d,M}(\mathcal{P},\mathcal{F})$ [32].

$\mathcal{P}$ selects colluding users $\mathcal{S}^1 \subseteq [d]$ (unknown to $\mathcal{F}$)
for $j = 1$ *to* $M(N)$ **do**
  $\mathcal{F}$ outputs column vector $c^j \in \{0,1\}^d$.
  Let $c^j_{S^j} \in \{0,1\}^{|S^j|}$ be the restriction of c^j to the coordinates of S^j. This is given to $\mathcal{P}$.
  $\mathcal{P}$ generates $a^j \in \{0,1\}$. This is given to $\mathcal{F}$.
  $\mathcal{F}$ accuses a set of users $I^j \subseteq [d]$. Let $S^{j+1} = S^j \setminus I^j$.
end

be the number of rounds $1,\dots,j$ such that the output of the adversary $\mathcal{P}$ is not consistent. Also, let

$$\psi^j \triangleq \left| \left(\bigcup_{1\le k\le j} I^k \right) \setminus S \right| \tag{A64}$$

be the number of users in $I^1,\dots,I^j$ that are falsely accused by $\mathcal{F}$. We use these quantities to define a collusion-resilient interactive fingerprinting code.

Definition A2 (Collusion-resilient interactive fingerprinting code). Let N be the size of the colluding group. An algorithm $\mathcal{F}$ is said to be an *N-collusion-resilient interactive fingerprinting code* of length M for d users with failure probability ϵ if for every adversary $\mathcal{P}$,

$$\Pr_{\mathsf{IFPC}[\mathcal{P},\mathcal{F}]}[\theta^M = 0 \vee \psi^M > d/2000] \le \epsilon. \tag{A65}$$

This says that a good (collusion-resilient) fingerprinting code should be able to force the adversary $\mathcal{P}$ to be inconsistent (and hence identify the colluding users $\mathcal{S}$) while not making many false accusations. Here, we allow the fingerprinting code to falsely accuse a small constant fraction $1/2000$ of the total number of users d.

Our lower bounds (and the classical adaptive data analysis lower bounds) require the existence of interactive fingerprinting codes. We recall a construction of interactive fingerprinting codes from Ref. [32].

Theorem A5 (Theorem 2.2. in Ref. [32]). For every $1 \le N \le d$, there is an N-collusion-resilient interactive fingerprinting code of length $\mathcal{O}(N^2)$ for d users with failure probability $\epsilon \le \mathcal{O}(1/d)$ and falsely accuses at most a constant fraction of the users.

Note that in Ref. [32], they state and prove a more general theorem, but this is all that we require. We consider the special case of their theorem in which we tolerate falsely accusing a small constant fraction of users, i.e., see Remark 2.3 in Ref. [32].

4. Cryptography

In this Appendix, we review some basic cryptography, which we use in the proof of Proposition C1. We first define a private-key encryption scheme.

Definition A3 (Private-key encryption scheme [82]). Let λ be the security parameter. Let $\mathcal{K}$ be the key space, $\mathcal{M}$ be the message space, and $\mathcal{C}$ be ciphertext space. A private-key encryption scheme is composed of three algorithms $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$, defined as follows:

(a) $\mathsf{Gen}(1^\lambda)$: On input the security parameter 1^λ, the key-generation algorithm outputs a secret key $\mathsf{sk} \in \mathcal{K}$ at random.

(b) $\mathsf{Enc}_{\mathsf{sk}}(m)$: On input the secret key $\mathsf{sk} \in \mathcal{K}$ and the plaintext $m \in \mathcal{M}$, the encryption algorithm outputs a ciphertext $c \in \mathcal{C}$.

(c) $\mathsf{Dec}_{\mathsf{sk}}(c)$: On input the secret key $\mathsf{sk} \in \mathcal{K}$ and the ciphertext $c \in \mathcal{C}$, the decryption algorithm outputs a plaintext $m \in \mathcal{M}$.

A private-key encryption scheme is correct if

$$\mathsf{Dec}_{\mathsf{sk}}(\mathsf{Enc}_{\mathsf{sk}}(m)) = m. \tag{A66}$$

In addition to correctness, we want our encryption schemes to be secure. In other words, from the encrypted ciphertext, it should be difficult to recover the corresponding plaintext. The definition of security we need in this setting is perfect secrecy.

Definition A4 (Perfect secrecy [82]). An encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is *perfectly secret* if for every pair of messages $m_0, m_1 \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$,

$$\Pr_{\mathsf{sk}\sim\mathsf{Gen}(1^\lambda)}[\mathsf{Enc}_{\mathsf{sk}}(m_0) = c] = \Pr_{\mathsf{sk}\sim\mathsf{Gen}(1^\lambda)}[\mathsf{Enc}_{\mathsf{sk}}(m_1) = c]. \tag{A67}$$

In words, this means that for any pair of messages m_0, m_1, it is impossible to distinguish the ciphertext of m_0 from the ciphertext of m_1. In fact, one can also devise a simple perfectly secret private-key encryption scheme called the one-time pad [83].

Definition A5 (One-time pad). Let λ be the security parameter. Let $\mathcal{K} = \mathcal{M} = \mathcal{C} = \{0,1\}^\lambda$ be the key, message, and ciphertext spaces. The one-time pad is defined by $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ as follows.

(a) $\mathsf{Gen}(1^\lambda)$: Output a uniformly random bitstring $\mathsf{sk} \in \{0,1\}^\lambda$.

(b) $\mathsf{Enc}_{\mathsf{sk}}(m)$: Encrypt $m \in \{0,1\}^\lambda$ as $c = m \oplus \mathsf{sk}$.

(c) $\mathsf{Dec}_{\mathsf{sk}}(c)$: Decrypt $c \in \{0,1\}^\lambda$ as $m = c \oplus \mathsf{sk}$.

The perfect secrecy of the one-time pad was first proven in Ref. [84]. We will not repeat the proof here.

APPENDIX B: LOCAL OBSERVABLES

We consider the task of predicting properties $\mathrm{tr}(O\rho)$, where ρ is the unknown quantum state that we are given copies of and O is a local observable with spectral norm bounded by 1. Here, we consider a local observable to be an observable acting on a constant number of qubits. In this Appendix, we provide a proof of Theorem 1, which characterizes the sample complexity for this problem. We restate the theorem below.

Theorem B1 (Local observables, detailed restatement of Theorem 1). There exists an analyst $\mathcal{A}$ and a density matrix ρ on $n = \Omega(M 2^{2000\sqrt{M}})$ qubits such that

any $[0.99, 1-(1/2000\sqrt{M})]$-accurate quantum mechanism $\mathcal{M}$ estimating expectation values of M adaptively chosen single-qubit observables requires at least

$$N = \Omega(\sqrt{M}) \tag{B1}$$

samples of the quantum state ρ. Meanwhile, for any $\epsilon, \delta \in (0, 1/2)$ such that $\epsilon\delta \geq 4e^{-M/\log^2 M \log\log^4 M}$, there exists an (ϵ, δ)-accurate quantum mechanism $\mathcal{M}$ that can estimate expectation values of M adaptively chosen k-local observables using

$$N = \mathcal{O}\left(\frac{\min\{3^k\sqrt{M\log(1/\epsilon\delta)}, k\log(n/\delta)\}}{\epsilon^2}\right) \tag{B2}$$

samples of the unknown n-qubit quantum state ρ. Moreover, the mechanism is computationally efficient and runs in time $\mathrm{poly}(N, n, \log(1/\delta))$ per observable.

In Appendix B 1, we begin with a warmup showing how adaptivity can cause overfitting for the classical shadows protocol [8]. Then, in Appendix B 3, we prove the sample complexity lower bound, which states that there does not exist a quantum mechanism that can accurately estimate M adaptively chosen local observables given poly $\log(M)$ copies of ρ. Finally, in Appendix B 4, we give a quantum algorithm achieving a matching sample complexity upper bound for answering adaptively chosen local observables.

In both Appendixes B 1 and B 3, we focus on the setting where we are given N copies of a diagonal n-qubit density matrix ρ. This diagonal density matrix can be viewed as the classical probability distribution $\mathcal{D}$ in the statistical query model. In addition, the analyst is constrained to only querying local observables $\{Z_i : i \in [n]\}$, where Z_i is the Pauli Z observable that acts only on the ith qubit. In this notation, we implicitly have the identity matrix on the remaining qubits. For each query Z_i, the mechanism aims to estimate $\mathrm{tr}(Z_i\rho)$. We show that for any mechanism, there always exists a quantum state ρ (of this diagonal form) and an analyst (even with this restricted power of only querying Z_i observables) such that the mechanism fails to estimate the expectation value of the analyst's adversarially chosen observables.

In this setting, one way of viewing the problem is as a classical problem involving n coins, whose joint distribution is dictated by the diagonal elements of ρ. The query $q(x)$ is the outcome of flipping one of the coins in sample x, and the goal would be to predict the true biases $q(\mathcal{D})$ of the queried coins.

1. Adaptive attack

First, we demonstrate the dangers of adaptivity by constructing an attack querying only single-qubit observables that is sufficient to cause the classical shadows protocol to fail after $\mathcal{O}(N)$ queries, where N is the number of samples of the unknown quantum state ρ the mechanism is given. We devise the analyst's adversarial strategy and a density matrix ρ such that classical shadows fails.

Consider a density matrix ρ on $n = M + 2^M$ qubits, where recall M is the number of queries the analyst makes. Here, the qubits are correlated such that the outcomes of the last 2^M qubits are completely determined by the first M. Denote the power set as $\mathcal{P}([M]) \triangleq \{I_1, \ldots, I_{2^M}\}$, i.e., $I_1 = \emptyset, I_2 = \{1\}, \ldots, I_{2^M} = [M]$. We will simulate a classical probability distribution by defining ρ to be the diagonal matrix

$$\rho \triangleq \sum_{Q\in\{0,1\}^n} p(Q)\,|Q\rangle\langle Q|\,Q, \tag{B3}$$

where $Q = (Q_1, \ldots, Q_n) \in \{0, 1\}^n$ so that $|Q\rangle$ is a computational basis states. Here, we use subscripts to denote labels for the different qubits. To construct an adversarial density matrix, we effectively encode the classical distribution from the feature selection attack of Refs. [16,61] into the density matrix. Thus, we define the probability distribution $p(Q)$ as

$$p(Q) = p(Q_1, \ldots, Q_n) \triangleq \begin{cases} \frac{1}{2^M} & \text{if } \mathrm{MAJ}(\{Q_i : i \in I_j\}) = Q_{M+j}, \forall j \in [2^M], \\ 0 & \text{otherwise,} \end{cases} \tag{B4}$$

where $\mathrm{MAJ}(\{Q_i \mid i \in I_j\})$ denotes the majority element of the set $\{Q_i : i \in I_j\}$. In other words,

$$\mathrm{MAJ}(\{Q_i : i \in I_j\}) = \begin{cases} 1 & \text{if } \sum_{i\in I_j} Q_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{B5}$$

As a technical detail, the rule for Q_{M+j} corresponding to $I = \emptyset$ can simply be set ahead arbitrarily, e.g., set to always be 0. We can view the distribution $p(Q)$ of the qubits labeled by $Q = (Q_1, \ldots, Q_n)$ as follows. The first M qubits labeled by $Q_1, \ldots, Q_M$ are uniformly distributed over $\{0, 1\}^M$ while the last 2^M qubits $Q_{M+1}, \ldots, Q_{M+2^M}$ are uniquely determined by the first M qubits via a majority voting rule.

For this density matrix, we want to show that there exists an adversarial analyst that can cause the classical shadows protocol with random Pauli measurements to fail after $\mathcal{O}(N)$ queries. This means that the classical shadows formalism for predicting local observables fails to protect against adaptive queries. Recall that classical shadows for predicting local observables are produced by (i) repeatedly measuring each qubit of ρ independently in a random Pauli basis and receiving output bitstring $|\hat{b}\rangle \in \{0, 1\}^n$ and then (ii) computing N classical snapshots of

ρ of the form $\hat{\rho}_k = \bigotimes_{j=1}^{n}(3U_j^{(k)\dagger}\left|\hat{b}_j^{(k)}\right\rangle\!\left\langle\hat{b}_j^{(k)}\right|U_j^{(k)} - \mathbb{I})$, for $k = 1,\ldots,N$, where $U_j^{(k)}$ is the Pauli matrix with respect to which the jth qubit in the kth copy of ρ was measured. Then, given a queried observable O from the analyst, the classical shadow mechanism estimates $\mathrm{tr}(O\rho)$ by returning $a(O) \triangleq 1/N \sum_{k=1}^{N} \mathrm{tr}(O\hat{\rho}_k)$, i.e., the empirical mean of the expectation values on each of the classical snapshots [85]. We refer to Appendix A 2 a and Ref. [8] for a more in-depth presentation of the classical shadow formalism. In our restricted setting, the analyst only queries from the set of single-qubit observables $\{Z_i : i \in [n]\}$.

Finally, we describe the analyst's behavior, which causes classical shadows to fail. The analyst performs the following procedure:

(1) Query $Z_1,\ldots,Z_M$ and receive $a(Z_i)$ for $i \in [M]$.
(2) Let $I \triangleq \left\{i \in [M] : a(Z_i) \geq \frac{9}{\sqrt{N}}\right\}$. Then query Z_{M+j^*} where $I_{j^*} = I$.

Here, because we indexed the power set $\mathcal{P}([M])$ as $\mathcal{P}([M]) = \{I_1,\ldots,I_{2^M}\}$ and $I \subseteq [M]$, there clearly must exist some $j^* \in [2^M]$ such that $I_{j^*} = I$. We claim that the mechanism's answer $a(Z_{M+j^*})$ to the query Z_{M+j^*} in Step 2 of the above procedure does not estimate the true expectation value $\mathrm{tr}(Z_{M+j^*}\rho)$ well for ρ defined in Eq. (B4).

The intuition for the attack is that the analyst looks for Z_i whose sample outcomes $a(Z_i)$ are especially biased toward the positive direction. We could also have chosen to have the analyst search for bias in the negative direction; as long as the sample outcomes are all biased in some direction, it does not matter which one. Each such Z_i will skew the empirical mean of every Z_{M+j} where $i \in I_j$ in the positive direction (since each such Z_{M+j} is correlated with Z_i in our construction). Thus, Z_{M+j} with $I_j = I$ will contain contributions from every such positively biased Z_i, which will amplify its empirical bias toward the positive direction.

Formally, we have the following guarantee, which states that the analyst overfits after linearly many queries.

Theorem B2 (Adaptive attack). There is a constant c such that if $M \geq c\max(N, \log(1/\delta))$ and $N \geq c\log(1/\delta)$, with probability $1-\delta$,

$$|a(Z_{M+j}) - \mathrm{tr}(Z_{M+j}\rho)| \geq 0.99. \tag{B6}$$

Proof. This can essentially be reduced to the feature selection attack of Refs. [16,61] (Lecture 3).

We first describe the random variable $\hat{z}_k \triangleq \mathrm{tr}(Z_i\hat{\rho}_k)$, which is the expectation value of Z_i with respect to the kth classical snapshot $\hat{\rho}_k$. First, let us consider the case when the random Pauli measurement on the ith qubit is chosen to be Z. This occurs with probability $1/3$. Measuring the ith qubit of the kth copy of ρ in the Z basis yields the outcome $\left|\hat{b}_i^{(k)}\right\rangle$, which we denote as $\left|\hat{b}_i^{(k)}\right\rangle = \left|Q_i^k\right\rangle$ to distinguish it from other measurement results, in line with the notation in Eq. (B3). Measuring the other qubits $j \neq i$ of the kth copy of ρ in an independent random Pauli bases yield outcomes $\left|\hat{b}_j^{(k)}\right\rangle$. Computing $\mathrm{tr}(Z_i\hat{\rho}_k)$ explicitly in this case, we have

$$\begin{aligned}
\hat{z}_k &= \mathrm{tr}(Z_i\hat{\rho}_k) && \text{(B7)}\\
&= \mathrm{tr}\left((I\otimes\cdots\otimes I\otimes Z\otimes I\otimes\cdots\otimes I)\left(\bigotimes_{j=1}^{n} 3U_j^{(k)\dagger}\left|\hat{b}_j^{(k)}\right\rangle\!\left\langle\hat{b}_j^{(k)}\right|U_j^{(k)} - \mathbb{I}\right)\right) && \text{(B8)}\\
&= \mathrm{tr}(3Z(U_i^{(k)\dagger}\left|Q_i^k\right\rangle\!\left\langle Q_i^k\right|U_i^{(k)} - \mathbb{I}))\cdot\prod_{j\neq i}\mathrm{tr}(3U_j^{(k)\dagger}\left|\hat{b}_j^{(k)}\right\rangle\!\left\langle\hat{b}_j^{(k)}\right|U_j^{(k)} - \mathbb{I}) && \text{(B9)}\\
&= \mathrm{tr}(3ZZ\left|Q_i^k\right\rangle\!\left\langle Q_i^k\right|Z - \mathbb{I}) && \text{(B10)}\\
&= \begin{cases} 3 & Q_i^k = 0,\\ -3 & Q_i^k = 1, \end{cases} && \text{(B11)}
\end{aligned}$$

where in the fourth line, we used our assumption that in this case that the random Pauli measurement on the ith qubit is chosen to be Z. Meanwhile, a similar computation shows that if the random Pauli measurement on the ith qubit is X or Y, the expectation value $\hat{z}_k$ evaluates to 3 or -3 with equal probability.

In summary, with probability $1/3$, the ith qubit is measured in the computational basis to be Q_i^k so that $\hat{z}_k$ is -3

or 3 if Q_i^k is 1 or 0, respectively. With probability $2/3$, $\hat{z}_k$ is 3 or -3 with equal probability.

Thus, notice that for all i, $a(Z_i) = 1/N \sum_{k=1}^{N} \mathrm{tr}(Z_i \hat{\rho}_k)$ is a (rescaled) binomial distribution taking values between $[-3, 3]$, where the expected value is 0 and the standard deviation is $3/\sqrt{N}$. This is because Q_i^k takes values 1 or 0 with equal probability, so $\mathrm{tr}(Z_i \hat{\rho}_k)$ is 3 or -3 with equal probability by the above analysis.

We can use this to show that given M large enough, then $|I| = \Omega(M)$. Recall that a binomial random variable deviates from its mean by a constant factor of its standard deviation with constant probability, so

$$\Pr[i \in I] = \Pr\left[a(Z_i) \geq 9/\sqrt{N}\right] = \Omega(1). \tag{B12}$$

Denote this probability as p. This implies that $\mathbb{E}[|I|] = \Omega(M)$. Let X_i be the random variable that is 1 when $i \in I$ and 0 otherwise, and let $X = \sum_{i=1}^{M} X_i = |I|$. By the Chernoff bound,

$$\Pr[X \leq pM - t] \leq \exp\left(-2t^2/M\right). \tag{B13}$$

By setting $t = pM/2$ and rearranging, we get that given $M \geq c \log(1/\delta)$ for some constant c, then $|I| = \Omega(M)$ with probability at least $1 - (\delta/2)$.

Assuming $|I| = \Omega(M)$, we will next show that with M large enough and for a state Q chosen uniformly at random from the sample, the odds of the adaptively chosen qubit Q_{M+j} being 1 is large. Denote Q_{M+j} to be the qubit corresponding to I, where by our majority voting rule in Eq. (B4), $Q_{M+j} = 1$ if and only if $\sum_{i \in I} Q_i > 0$. Moreover, recall $i \in I$ if and only if $a(Z_i) = 1/N \sum_{k=1}^{N} \mathrm{tr}(Z_i \hat{\rho}_k) \geq 9/\sqrt{N}$, where above we analyzed the random variable $\hat{z}_k$ in two cases: (i) with probability $1/3$ the ith qubit was measured in the Z basis, in which case $\hat{z}_k = \mathrm{tr}(Z_i \hat{\rho}_k)$ is 3 if $Q_i^k = 1$ and else -3, and (ii) with probability $2/3$ the ith qubit was measured in the X or Y basis, in which case $\hat{z}_k$ is 3 or -3 with equal probability. Let $T = \{|Q^1\rangle, \ldots, |Q^N\rangle\}$ denote the computational basis state realizations of the N given copies of the quantum state ρ. Thus, for every $i \in I$ and set T

$$\mathop{\mathbb{E}}_{Q \sim T}[\mathbb{1}(Q_i = 1)] \geq \frac{1}{3} \mathop{\mathbb{E}}_{Q \sim T}[\mathbb{1}(\mathrm{tr}(Z_i \hat{\rho}) = 3)] \tag{B14}$$

$$= \frac{1}{3N} \sum_{k=1}^{N} \mathbb{1}(\mathrm{tr}(Z_i \hat{\rho}_k) = 3) \tag{B15}$$

$$\geq \frac{1}{9} a(Z_i) \tag{B16}$$

$$\geq \frac{1}{\sqrt{N}}, \tag{B17}$$

where the first inequality follows by only considering case (i) for the random variable $\hat{z}_k$ above and discarding case (ii). Thus,

$$\mathop{\mathbb{E}}_{Q \sim T}\left[\sum_{i \in |I|} \mathbb{1}(Q_i = 1)\right] \geq \frac{|I|}{\sqrt{N}}. \tag{B18}$$

This implies that $Q_{M+j} = 1$ unless $\sum_{i \in |I|} \mathbb{1}(Q_i = 1)$ differs from its expectation by more than $|I|/\sqrt{N}$. Using another Chernoff bound,

$$\Pr_{Q \sim T}[Q_{M+j} \neq 1]$$

$$= \Pr\left[\sum_{i \in I} \mathbb{1}(Q_i = 1) \leq \mathbb{E}\left[\sum_{i \in I} \mathbb{1}(Q_i = 1)\right] - \frac{|I|}{\sqrt{N}}\right] \tag{B19}$$

$$\leq \exp\left(-\frac{2|I|^2}{N \cdot |I|}\right) \tag{B20}$$

$$= \exp\left(\frac{-2|I|}{N}\right). \tag{B21}$$

If $|I| \geq N \ln(400)/2$, then $\Pr_{Q \sim S}[Q_{M+j} \neq 1] \leq 1/400$. This follows if $M \geq c \cdot N$, for c large enough. We will now bound the difference the classical shadows output $a(Q_{M+j})$ and the empirical mean of measuring the $M+j$th qubit of the samples in the Z basis $q(Z_{M+j}) = 1/N \sum_{Q \in S}[\mathbb{1}(Q_{M+j} = 1)] - 1/N \sum_{Q \in S}[\mathbb{1}(Q_{M+j} \neq 1)]$. Recall that with probability $1/3$ the expectation value $\mathrm{tr}(Z_{M+j} \hat{\rho}_k)$ falls under case (i), where it matches the output of measuring the $(M+j)$th qubit in the Z basis, and with probability $2/3$ it falls under case (ii), where it randomly outputs 3 or -3 with equal probability. This is captured in the random variable $X_k = \mathrm{tr}(Z_{M+j} \hat{\rho}_k) - \left\langle Q_{M+j}^k \middle| Z_{M+j} \middle| Q_{M+j}^k \right\rangle$, which is defined as follows. If $\left\langle Q_{M+j}^k \middle| Z_{M+j} \middle| Q_{M+j}^k \right\rangle = 1$, X_k is 2 with probability $2/3$ and -4 with probability $1/3$. This is because if $\left\langle Q_{M+j}^k \middle| Z_{M+j} \middle| Q_{M+j}^k \right\rangle = 1$, then $Q_{M+j}^k = 0$ so that by case (i) with probability $1/3$, $\mathrm{tr}(Z_{M+j} \hat{\rho}_k) = 3$. Hence, $X_k = 2$ with probability $1/3$. By case (ii), with probability $2/3$, $\mathrm{tr}(Z_{M+j} \hat{\rho}_k)$ is 3 or -3 with equal probability so that $X_k = 2$ with probability $1/3$ and $X_k = -4$ with probability $1/3$. In total, we have that $X_k = 2$ with probability $2/3$ and $X_k = -4$ with probability $1/3$. If $\left\langle Q_{M+j}^k \middle| Z_{M+j} \middle| Q_{M+j}^k \right\rangle = -1$, then by a similar argument, X_k is -2 with probability $2/3$ and 4 with probability $1/3$. By applying Hoeffding's

inequality,

$$\begin{aligned}&\Pr\left[|a(Q_{M+j}) - q(Z_{M+j})| \geq \frac{1}{200}\right]\\&= \Pr\left[|\sum_{k=1}^{N} \mathrm{tr}(Z_{M+j}\hat{\rho}) - \langle Q^k| Z_{M+j} |Q^k\rangle| \geq \frac{N}{200}\right]\\&\leq 2\exp\left\{-\frac{N}{18(200)^2}\right\}.\end{aligned}$$

Taking $N = c\log(1/\delta)$ with c being a large enough constant yields $\Pr\left[|a(Q_{M+j}) - q(Z_{M+j})| \geq \frac{1}{200}\right] \leq \delta/2$. Taking union bound over this event and the event that $|I| = \Omega(M)$, we get that with probability at least $1-\delta$,

$$a(Z_{M+j}) \geq q(Z_{M+j}) - \frac{1}{200} \tag{B22}$$

$$= \frac{1}{N}\sum_{Q\in S}[\mathbb{1}(Q_{M+j} = 1)] - \frac{1}{N}\sum_{Q\in S}[\mathbb{1}(Q_{M+j} \neq 1)] - \frac{1}{200} \tag{B23}$$

$$= \Pr_{Q\sim S}[Q_{M+j} = 1] - \Pr_{Q\sim S}[Q_{M+j} \neq 1] - \frac{1}{200} \tag{B24}$$

$$\geq \frac{199}{200} - \frac{1}{200} \tag{B25}$$

$$\geq 0.99. \tag{B26}$$

Also, since $Q^1 \cdots Q^M$ are uniformly distributed over $\{0,1\}^M$, $\mathrm{tr}(Z_{M+j}\rho) = 0$. This yields the final result. ■

2. Numerical experiment

We consider the same setting as the adaptive attack. For our numerical experiments, we fix the number of samples $N = 10\,000$ and vary the number of queries M that are asked. The quantum state is determined by the majority rule in Eq. (B5) and is uniform among pure states that satisfy the rule, as in Eq. (B4). Recall that one can view this setting as a classical problem involving n coins, whose joint distribution is dictated by the diagonal elements of ρ. Querying Z_i observables gives the outcome of flipping one of the coins. The goal is then to predict the true biases of the queried coins. We simulate this equivalent classical problem in our numerical experiments.

We simulate the same adaptive attack, where $Z_1, \ldots, Z_M$ are queried followed by an adaptive query Z_{M+j^*}. The adaptive error is taken as the error of the adaptively chosen query. We run this for 100 random instances for $M = \{100, 200, 400, 800, 1600, 3200, 6400, 10\,000\}$ and compute the mean error and standard deviation for each M. We also compare this to a nonadaptive querying rule running on the same instances, where the queries are fixed ahead of time. Specifically, we query $Z_1, \ldots, Z_M$ and then Z_{M+j} where Q_{M+j} is conditioned on three-quarters of the first M qubits. The nonadaptive error is computed by taking the maximum error over these queries. Our numerical results in Fig. 2 show that the adaptive error quickly becomes significantly higher than the nonadaptive error as M grows. This confirms our theoretical result from Theorem B2 and emphasizes how false discovery can occur in the experiments. Our implementation for the numerical experiments is available at https://github.com/jrrhuang/quantum-adaptive-data-analysis [57].

3. Lower bound

In this Appendix, we prove a sample complexity lower bound, which demonstrates that in general, we cannot design a mechanism that can accurately estimate M adaptively chosen local observables given poly $\log(M)$ copies of the unknown quantum state ρ. In particular, we show that it is impossible to obtain the logarithmic dependence on the number of queries M present in the classical shadow formalism for the nonadaptive setting [8]. We prove the following lower bound.

Proposition B1 (Local observables, lower bound). There exists an analyst $\mathcal{A}$ and density matrix ρ with system size $n = \Omega(M2^{2000\sqrt{M}})$ such that any $[0.99, 1 - (1/2000\sqrt{M})]$-accurate quantum mechanism $\mathcal{M}$ estimating expectation values of M adaptively chosen single-qubit observables requires at least $N = \Omega(\sqrt{M})$ samples of the quantum state ρ.

While the lower bound on the system size may seem large, our upper bound presented in the next section shows that an exponential scaling in system size is in fact necessary for such a lower bound to hold.

Similarly to the adaptive attack in Appendix B 1, we consider the setting in which the analyst queries single-qubit observables $\{Z_i\}$ and the mechanism has access to N copies of a diagonal density matrix ρ. The idea behind the proof of this lower bound builds on lower bounds found in the classical adaptive statistical query setting [15,32,45]. In particular, Ref. [32] showed an $\Omega(\sqrt{M})$ sample complexity lower bound for the adaptive statistical query setting. We provide a similar construction for an analyst interacting with a quantum mechanism, which achieves this lower bound even when the analyst only queries single-qubit observables. The main tool for the proof will be *interactive fingerprinting codes* [46], which we reviewed in Sec. A 3. The argument is similar to the proof of a sample complexity lower bound for the adaptive statistical query setting [32].

The idea is to reduce our problem to IFPC by designing an analyst $\mathcal{A}$ such that if the mechanism $\mathcal{M}$ answers all

of the analyst's queries accurately, then we contradict the definition of $\mathcal{F}$ being a collusion-resilient fingerprinting code. In particular, we design an attack on the mechanism that simulates the IFPC game (Algorithm 2), where the mechanism $\mathcal{M}$ plays the role of the adversary $\mathcal{P}$ in the IFPC game. We argue that in our attack, the mechanism $\mathcal{M}$ indeed adheres to the same conditions as the adversary $\mathcal{P}$ (i.e., only having access to information from the nonaccused colluding users) so that the guarantees of the interactive fingerprinting codes hold. These guarantees state that the fingerprinting code $\mathcal{F}$ can force the adversary $\mathcal{P}$ to be inconsistent. Then, because $\mathcal{M}$ is the same as the adversary $\mathcal{P}$, this implies that $\mathcal{M}$ also answers one of the analyst's queries incorrectly.

Proof of Proposition B1. We first define the setup of the problem. Throughout this proof, we use $\mathcal{A}$ for the analyst, $\mathcal{M}$ for the mechanism, $\mathcal{F}$ for the fingerprinting code, and $\mathcal{P}$ for the adversary in the IFPC game. Recall that the mechanism is given copies of an unknown quantum state ρ, with which it is asked to predict $\mathrm{tr}(O_i\rho)$ for an (adaptively chosen) query O_i from the analyst. We first precisely define an adversarial choice of ρ. Subsequently, we will present an algorithm for $\mathcal{A}$ that thwarts any mechanism, even when only querying single-qubit observables.

Let d be the number of users in the setting of fingerprinting codes. Let M be the number of observables that the analyst queries.

Defining the density matrix: We define the density matrix ρ on n qubits, where we specify n shortly. In fact, we define an ensemble of density matrices $\{\rho_\ell\}$, where the density matrix ρ given to the mechanism is chosen from the ensemble uniformly at random. Each density matrix ρ_ℓ in the ensemble is diagonal so that it corresponds to a classical distribution. We effectively encode the adversarial choice of distribution from Ref. [32] into this density matrix with technical changes to ensure the guarantees hold in our setting. Moreover, note that we use a diagonal density matrix because the guarantees from interactive fingerprinting codes hold for classical adversaries (i.e., classical mechanisms). We later remark that quantum mechanisms can be reduced to classical mechanisms.

Let $n = \lceil \log d \rceil + M2^d$. We index the first $\lceil \log d \rceil$ qubits by an integer in $\{1, \ldots, \lceil \log d \rceil\}$, and we index the remaining qubits by tuples (k, j) such that $k \in [2^d], j \in [M]$. This reflects the underlying structure that the last qubits are M groups of 2^d qubits. In this way, j indexes the group of qubits and k indexes the specific qubit within the jth group. We describe the intuition for this structure later.

Similarly to the adaptive attack (Appendix B 1), since the density matrices ρ_ℓ in the ensemble we construct is diagonal, we can write

$$\rho_\ell \triangleq \sum_{Q \in \{0,1\}^n} p_\ell(Q)\, |Q\rangle\langle Q|\,, \tag{B27}$$

where $|Q\rangle$ is a computational basis state so that $Q \in \{0,1\}^n$. Note that $\ell \in \{1, \ldots, (2^d!)^M\}$, so our ensemble consists of $(2^d!)^M$ density matrices. The reason for this will become clear later. We can index the entries of the bitstring Q as previously described:

$$Q = Q_1 \cdots Q_{\lceil \log d \rceil} Q_{(1,1)} \cdots Q_{(2^d,1)} \cdots Q_{(1,M)} \cdots Q_{(2^d,M)} \in \{0,1\}^n. \tag{B28}$$

The state $|Q\rangle$ can be viewed as follows. The first $\lceil \log d \rceil$ qubits index the user from the setting of fingerprinting codes, where recall there are d users. Explicitly, each integer in $\{1, \ldots, d\}$ can be represented by in binary using $\lceil \log d \rceil$ bits. The bitstring $Q_1 \cdots Q_{\lceil \log d \rceil} \in \{0,1\}^{\lceil \log d \rceil}$ encodes this binary representation of the user's index $i \in \{1, \ldots, d\}$ [86]. There are also M groups of 2^d qubits. Namely, for $j \in [M]$, these are the groups of qubits corresponding to $Q_{(1,j)} \cdots Q_{(2^d,j)}$. Each group corresponds to a round j of the interaction between the analyst and mechanism, where at round j, the analyst will query a qubit in group j.

We want to define the distribution $p_\ell(Q)$ in a specific way such that the first $\lceil \log d \rceil$ qubits determine the state of the remaining ones. Let $\sigma_{j,\ell} : \{0,1\}^d \to [2^d]$ be a random assignment of $q \in \{0,1\}^d$ to some index $k_j \in [2^d]$ at round j. Here, we suggestively denote a bitstring in $\{0,1\}^d$ by q, where, as hinted before, the analyst will query one of the 2^d qubits in a given group j. When we consider a randomly chosen ρ from the ensemble $\{\rho_\ell\}$, we will often drop the ℓ subscript on $\sigma_{j,\ell}$ and simply write σ_j. Each density matrix ρ_ℓ corresponds to a random assignment σ_ℓ defined by each of the assignments $\sigma_{1,\ell}, \ldots, \sigma_{M,\ell}$. Namely,

$$\sigma_\ell(q^1, \ldots, q^{2^d}) = (\sigma_{1,\ell}(q^1), \ldots, \sigma_{1,\ell}(q^{2^d}), \ldots, \sigma_{M,\ell}(q^1), \ldots, \sigma_{M,\ell}(q^{2^d})). \tag{B29}$$

Thus, $\ell \in \{1, \ldots, (2^d!)^M\}$ because this is the number of random permutations of $[M2^d]$. Define a set of binary strings

$$R_\ell \triangleq \{Q \in \{0,1\}^n : Q_{(\sigma_{j,\ell}(q),j)} = q_i \quad \forall q \in \{0,1\}^d, \forall j \in [M], \text{ where } Q_1 \cdots Q_{\lceil \log d \rceil} = i \in [d]\}. \tag{B30}$$

This defines precisely how the first $\lceil \log d \rceil$ qubits determine the remaining ones. Namely, if the first $\lceil \log d \rceil$ bits are a binary representation of the user's index $i \in [d]$, then the other bits in the bitstring $Q \in \{0,1\}^n$ must adhere to the rule that the bit indexed by $(\sigma_{j,\ell}(q), j)$ is equal to q_i for all $q \in \{0,1\}^d$ and for all rounds $j \in [M]$. With this, we can

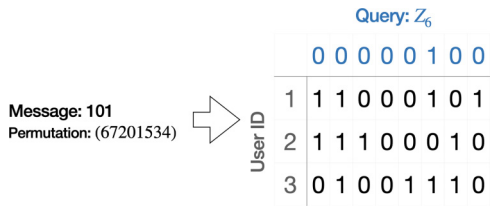

FIG. 3. Example construction for local observables. We consider an example where the query (message) is $q^j = 101$ and the randomly generated permutation is $\sigma_j = (6, 7, 2, 0, 1, 5, 3, 4)$, where we denote each bitstring q by its decimal representation. The table shows the quantum state, with each row being a computational basis state $|Q\rangle$ corresponding to a user (which is encoded in the first $\lceil \log d \rceil$ qubits of $|Q\rangle$). The permutation is encoded into the quantum state as shown in the table, e.g., the first column is $q = 110$ (corresponding to 6). The message is $q^j = 101$ (corresponding to 5) is on the sixth entry of the permutation, so the analyst will query the single-qubit observable Z_6.

define $p_\ell(Q)$ as follows:

$$p_\ell(Q) \triangleq \begin{cases} \frac{1}{d} & \text{if } Q \in R_\ell, \\ 0 & \text{otherwise.} \end{cases} \tag{B31}$$

In this way, we have a uniform distribution over all users, and the remaining $M2^d$ qubits must satisfy the rule specified in the definition of R_ℓ [Eq. (B30)]. An example is illustrated in Fig. 3.

We have completed the definition of this ensemble $\{\rho_\ell\}$ of density matrices. Recall that the density matrix ρ given to the mechanism is chosen from this ensemble uniformly at random.

Defining the analyst's behavior: With this, we can define the analyst's adversarial behavior. Similarly to the adaptive attack (Appendix B 1), the analyst can cause the mechanism to answer queries inaccurately by querying only single-qubit observables. In particular, the analyst will query single-qubit Z observables.

Recall that in our setup of the density matrix, there are M groups of 2^d qubits. Each group corresponds to a round j of the interaction between the analyst and mechanism, where at round j, the analyst will query the observable $Z_{(k,j)}$ for some $k \in [2^d]$. Intuitively, at round j, the analyst queries a qubit chosen from $(1, j), \ldots, (2^d, j)$ uniformly at random and independently for each round. We describe this explicitly as follows. Let $q^j \in \{0,1\}^d$ be the query specified by the analyst in round j. Let $\sigma_j : \{0,1\}^d \to [2^d]$ be a random assignment of the query q^j to some index $k_j \in [2^d]$ for round j. Then, the analyst will query the observable $Z_{(\sigma_j(q^j),j)}$ at round j. Note that this randomization via σ_j is necessary to ensure that the guarantees of the fingerprinting code hold, and this plays the same role as encryption does in Ref. [32]. We justify this more formally later in the proof.

Now, we are ready to state the analyst's full attack. This is given in Algorithm 3.

Here, we use T^j to denote the set of users accused by the fingerprinting code $\mathcal{F}$ up to and including round j. We also denote $S^j = \mathcal{S} \setminus T^j$, where $\mathcal{S}$ is the set of colluding users. Thus, S^j is the set of colluding users that have not yet been accused.

Reduction to IFPC: We first show that our proposed attack follows the same interaction model as the IFPC game. Crucially, we claim that from the mechanism's perspective, querying the observable $Z_{(\sigma_j(q^j),j)}$ that corresponds to q^j (as in Algorithm 3) is effectively the same as querying a random observable $Z_{(i,j)}$ for which the observed $Q_{(i,j)}$ is the same as c^j_{Sj} (as required in Algorithm 2). We also claim that this modification is enough to cause the game to operate in the IFPC model.

To show this, we construct a simulated attack, given in Algorithm 4. The simulated attack is merely fictitious, as it relies on the analyst knowing the set of colluding users $\mathcal{S}$. The analyst first simulates sampling from the density matrix ρ that would be given to the mechanism in the real attack (Algorithm 3) and gives this simulated data to the mechanism. In this way, because the analyst knows $\mathcal{S}$, the dataset that is given to the mechanism in the simulated attack (Algorithm 4) only contains information about the colluding users. Explicitly, the analyst

ALGORITHM 3. Real attack (local version).

Let the number of users be $d = 2000N$.
Sample some n-qubit quantum state ρ from the ensemble $\{\rho_\ell\}$.
Give $\rho^{\otimes N}$ to the mechanism $\mathcal{M}$.
Initialize an N-collusion resilient fingerprinting code $\mathcal{F}$ for d users and length $M = \mathcal{O}(N^2)$.
Let $T^0 = \emptyset$.
for $j = 1$ *to* M **do**
  Let $c^j \in \{0,1\}^d$ be chosen by $\mathcal{F}$.
  Analyst $\mathcal{A}$ chooses query $q^j \in \{0,1\}^d$ such that $q^j_i = c^j_i$ if $i \notin T^{j-1}$ and 0 otherwise.
  $\mathcal{A}$ queries $Z_{(\sigma_j(q^j),j)}$ to $\mathcal{M}$, where $\sigma_j(q^j)$ denotes the index corresponding to q^j.
  $\mathcal{M}$ outputs response a^j, rounds a^j to $\bar{a}^j \in \{0,1\}$, and gives $\bar{a}^j$ to $\mathcal{F}$.
  $\mathcal{F}$ accuses $I^j \subseteq [d]$.
  Set $T^j \leftarrow T^{j-1} \cup I^j$.
end

ALGORITHM 4. Simulated attack (local version).

Let the number of users be $d = 2000N$ and let $\mathcal{U}$ be the uniform distribution over $[d]$.
Choose colluding users $\mathcal{S} \sim \mathcal{U}^N$ and give $\mathcal{S}$ to the analyst $\mathcal{A}$.
$\mathcal{A}$ simulates sampling from the density matrix ρ by generating $\mathcal{S}' = \{|Q\rangle\}$ corresponding to $\mathcal{S}$.
The analyst's procedure for generating $\mathcal{S}'$ is described further in the text.
Give $\mathcal{S}'$ to $\mathcal{M}$.
Initialize N-collusion resilient fingerprinting code $\mathcal{F}$ for d users and length $M = \mathcal{O}(N^2)$.
Let $S^1 = \mathcal{S}$.
for $j = 1$ *to* M **do**
  Let $c^j \in \{0,1\}^d$ be chosen by $\mathcal{F}$. Give $c^j_{S^j}$ to $\mathcal{A}$.
  $\mathcal{A}$ chooses a query at random from among $\{Z_{(\sigma_j(q),j)} : q_{S^j} = c^j_{S^j}\}$ and gives it to $\mathcal{M}$.
  $\mathcal{M}$ outputs response a^j. Round a^j to $\bar{a}^j \in \{0,1\}$, and give $\bar{a}^j$ to $\mathcal{F}$.
  $\mathcal{F}$ accuses $I^j \subseteq [d]$. Set $S^{j+1} \leftarrow S^j \setminus I^j$.
end

creates computational basis states $|Q\rangle$, where $Q_1 \cdots Q_{\lceil \log d \rceil}$ is the binary representation of some colluding user's index $i \in \mathcal{S}$ and $Q_{\sigma_j(q),j} = q_i$ for all $q \in \{0,1\}^d, j \in [M]$. The random assignments σ_j are generated in the same manner as described previously (namely, the same as in the real attack (Algorithm 3). Moreover, the simulated interaction with the sampled states was constructed with only knowledge of $c^j_{S^j}$ and contain no information about c^j outside of S^j. Thus, the simulated attack indeed operates in the correct IFPC interaction model required for fingerprinting code guarantees to hold.

To show that our proposed attack (Algorithm 3) follows the same interaction model as the IFPC game, we claim that the mechanism cannot distinguish between our proposed attack and the simulated attack (Algorithm 4), which as described must be in the IFPC interaction model. We can show that any mechanism cannot distinguish between the simulated and real attacks by noting the following:

(1) The distribution over the assignments σ_j are the same.
(2) For any given round j, the distribution of the queried qubit is the same.

The first point is true because the assignments were randomized uniformly before the first round in both the real and simulated attacks.

The second point is true because of the following argument. For the real attack (Algorithm 3), the assignments σ_j are randomized before the analyst selects q^j (only the analyst knows this random seed). Thus, from the mechanism's perspective, the queried qubit is effectively random, just as in the simulated attack. Note also that we could not argue this if we only had one copy of $\{Q_{(\cdot,1)}\}$ (or if we did not randomize each of the M groups independently). If this were the case, the mechanism could learn whether they were receiving the same queries between rounds, so the indices would not truly be random.

Thus, the mechanism's views are identical in both attacks. Since the distribution over the dataset and indices presented to the mechanism $\mathcal{M}$ are the same, the distribution over answers from $\mathcal{M}$ are also the same. This implies that the probabilities of inconsistency are the same, i.e.,

$$\Pr_{\text{Real}}[\exists j \in [M] : \forall i \in [d], \bar{a}^j \neq c^j_i] = \Pr_{\text{Sim}}[\exists j \in [M] : \forall i \in [d], \bar{a}^j \neq c^j_i]. \tag{B32}$$

By the argument in Ref. [32] for the algorithm for interactive fingerprinting codes, the distribution over challenges c^j are also the same for the real and simulated attacks. Thus, we have

$$\Pr_{\text{Real}}[|T^M \setminus \mathcal{S}| \leq n/2000] = \Pr_{\text{Sim}}[|T^M \setminus \mathcal{S}| \leq n/2000]. \tag{B33}$$

This completes the first part of the proof.

Conclusion: To finish the proof, we need to show that the analyst can force the mechanism to be inaccurate after $\mathcal{O}(N^2)$ adaptive queries. To do so, our argument follows that of the attack sketch (without cryptography) from Lecture 19 of Ref. [61] with a some modification. Throughout the M rounds of interaction between the analyst and mechanism, the guarantees of the fingerprinting code state that with probability $1 - 1/n$, the following hold:

(1) The number of false accusations does not exceed $n/2000$, i.e., $|T^j| \leq N + n/2000 = n/1000$.
(2) The mechanism $\mathcal{M}$ was forced to be inconsistent, in which case either $c^j_i = 1$ for all i or $c^j = -1$ for all i.

During these rounds, either

$$\mathbb{E}_{\mathcal{D}}[Z^j] \geq 1 - 2\frac{|T^j|}{n} \geq 1 - \frac{1}{500} \quad \text{or} \quad \mathbb{E}_{\mathcal{D}}[Z^j] \leq -1, \tag{B34}$$

where Z^j denotes the query that was asked in round j. Suppose that the mechanism answers queries with error at most

$\alpha = 0.99$. Then, in the first case, the answer $a^j > 0$, and in the second case, $a^j < 0$. In both cases, the rounded answers $\bar{a}^j$ are consistent. However, the guarantees of the fingerprinting code state that the probability that $\bar{a}^j$ is consistent on every round is at most $1 - 1/n$. Thus, the rest of the time, the mechanism was forced to answer the query with error $\alpha > 0.99$. ■

We emphasize some observations we made throughout the proof.

Remark B1. Note that the randomization of the qubits in our proposed attack (Algorithm 3) plays an equivalent role to the encryption in Ref. [32]. Namely it ensures the secrecy requirement needed for the IFPC model for queries outside of $\mathcal{S}$.

Remark B2. The guarantee in this proof is for classical mechanisms $\mathcal{M}$. However, since any quantum mechanism can be simulated classically and the arguments of the proof hold without restriction to the time complexity of the mechanism, the guarantee holds for any quantum mechanism as well.

We also point out that the same lower bounds hold for the following analogous classical adaptive data analysis problem. Suppose the distribution $\mathcal{D}$ is over an n-dimensional data universe $\mathcal{X} \subseteq \{0,1\}^n$, and the analyst is constrained to using only *k-sparse statistical queries* $q : \mathcal{X} \mapsto [0,1]$, which we define as functions depending on at most k bits of the input, k being a constant. Given a dataset $\mathcal{S} = (x_1, \ldots, x_N) \sim \mathcal{D}^N$, the goal of the classical mechanism is to accurately estimate $q(\mathcal{D}) = \mathbb{E}_{x\sim\mathcal{D}}[q(x)]$ for M adaptively chosen queries q. Since our proof for the local observables lower bound involves only classical constructions, we may recycle the proof to obtain new classical lower bounds for this sparse queries problem: considering an analyst querying 1-sparse queries of the form $q_b(x) = x_b$, the query q_b corresponds to the observable Z_b and the classical distribution $\mathcal{D}$ can be mapped onto a density matrix.

Corollary B1 (Sparse queries, lower bound). There exists an analyst $\mathcal{A}$ and classical distribution $\mathcal{D}$ acting on $\{0,1\}^n$, where $n = \Omega(M 2^{2000\sqrt{M}})$ such that any $[0.99, 1 - (1/2000\sqrt{M})]$-accurate classical mechanism $\mathcal{M}$ answering M adaptively chosen 1-sparse statistical queries requires at least $N = \Omega(\sqrt{M})$ from $\mathcal{D}$.

4. Upper bound

In this Appendix, we prove a sample complexity upper bound for answering adaptive queries specified by local observables. Recall that in our setting, the analyst's queries are specified by a local observable O_i, and the mechanism wishes to estimate expectation values $\mathrm{tr}(O_i\rho)$ accurately given copies of the unknown quantum state ρ. We design a mechanism that can successfully estimate these expectation values, even when the observable O_i is chosen adaptively. In the case where O_i are chosen nonadaptively, the median of means [67,68] protocol used in Ref. [8] is sufficient. However, we will need to use a different algorithm for when O_i are chosen adaptively. Moreover, while the proof of our sample complexity lower bound (Appendix B 3) only required the analyst to query single-qubit observables, our algorithm works for an analyst querying any local observables. This upper bound matches our lower bound from Proposition B1 exactly in M scaling. In particular, we prove the following proposition.

Proposition B2 (Local observables, upper bound). For every $\epsilon, \delta \in (0, 1/2)$ such that $\epsilon\delta \geq 4e^{-M/\log^2 M \log\log^4 M}$, there exists an (ϵ, δ)-accurate mechanism $\mathcal{M}$ that can estimate expectation values of M adaptively chosen k-local observables with

$$N = \mathcal{O}\left(\frac{\min\{3^k\sqrt{M\log(1/\epsilon\delta)}, k\log(n/\delta)\}}{\epsilon^2}\right) \tag{B35}$$

samples of the unknown n-qubit quantum state ρ. The mechanism runs in time $\mathrm{poly}(N, n, \log(1/\delta))$ per observable.

The idea of the proof is to find the classical shadow (see Appendix A 2 a for an overview) of the unknown quantum state ρ with respect to which we want to estimate expectation values. Then, we can apply algorithms from classical adaptive data analysis [17] to obtain our guarantees. Here, the classical shadow formalism allows us to export our quantum problem to the classical setting, where we can leverage algorithms from the classical literature. Moreover, in the case of local observables, classical shadows allow us to estimate expectation values of queried observables (computationally) efficiently.

We will require the following classical algorithm from adaptive data analysis [17].

Theorem B3 (Corollary 1.3 in Ref. [23]). For every $\epsilon, \delta \in (0, 1/2)$ such that $\epsilon\delta \geq 4e^{-M/\log^2 M \log\log^4 M}$, there is an (ϵ, δ)-accurate mechanism $\mathcal{M}$ for M adaptively chosen (scaled) statistical queries $q : \mathcal{X} \mapsto [-C, C]$ that uses

$$N = \mathcal{O}\left(\frac{C\sqrt{M\log(1/\epsilon\delta)}}{\epsilon^2}\right) \tag{B36}$$

samples from $\mathcal{X}$. The mechanism runs in time $\mathrm{poly}(N, n, \log(1/\delta))$ per observable.

Here, we use Corollary 1.3 in Ref. [23] applied to C/N-sensitive queries. We use this theorem to prove Proposition B2.

Proof of Proposition B2. To derive the upper bound for local observables, we apply the algorithm of Theorem B3 on classical shadows data obtained from random Pauli measurements. Recall from Appendix A 2 a that classical shadows obtained via random Pauli measurements allow us to predict expectation values of local observables well, as desired.

First, we claim that the setting of the classical shadow problem aligns with the classical statistical query model from Appendix A 1 a [87]. Recall that in the classical statistical query model, we wish to design an algorithm (the mechanism) to answer statistical queries $q : \mathcal{X} \to [-C, C]$, whose true value is given by

$$q(\mathcal{D}) = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)], \tag{B37}$$

where $\mathcal{D}$ is an unknown distribution over the data universe $\mathcal{X}$. The mechanism has access to a sample $\mathcal{S} = \{x_1, \ldots, x_N\} \sim \mathcal{D}^N$, which it can use in order to estimate $q(\mathcal{D})$. Moreover, statistical queries must satisfy

$$\mathop{\mathbb{E}}_{\mathcal{S}\sim\mathcal{D}^N}[q(\mathcal{S})] = q(\mathcal{D}), \tag{B38}$$

where $q(\mathcal{S}) = 1/N \sum_{x_i\in\mathcal{S}} q(x_i)$.

The setting of the classical shadow formalism for predicting local observables is indeed analogous to the statistical query model as follows. Let the data universe $\mathcal{X}$ be defined as $\{0,1\}^m$ for $m = \mathcal{O}(n)$, where n is the system size. One classical snapshot of the n-qubit unknown quantum state ρ from random Pauli measurements can be stored in m bits, as justified in Appendix A 2 a. Let $\mathcal{D}$ be the distribution from which the classical shadow data are sampled from. Namely, $\mathcal{D}$ is the distribution over the random choice of Pauli gates and the probabilistic outcomes from measuring ρ. Recall from Appendix A 2 a that we can write

$$\rho = \mathbb{E}\left[\mathcal{M}^{-1}\left(U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right)\right] = \mathop{\mathbb{E}}_{\hat{\rho}\sim\mathcal{D}}[\hat{\rho}], \tag{B39}$$

where U is a random tensor product of n single-qubit Pauli gates, $\hat{b} \in \{0,1\}^n$ is the result of measuring $U\rho U^\dagger$ in the computational basis, and $\mathcal{M}$ is a quantum channel. Here, we denote

$$\hat{\rho} = \mathcal{M}^{-1}\left(U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right). \tag{B40}$$

Then, our sample $\mathcal{S} \sim \mathcal{D}^N$ is the classical shadow $\mathcal{S} = \mathsf{S}(\rho; N) = \{\hat{\rho}_1, \ldots, \hat{\rho}_N\}$ from repeating the random measurement procedure N times to obtain N classical snapshots. The queries are functions $o : \mathcal{X} \to [-C, C]$ (we justify later why the outputs falls in $[-C, C]$) corresponding to an observable O, which are evaluated on a classical snapshot $\hat{\rho} \in \mathcal{X}$ as $o(\hat{\rho}) = \mathrm{tr}(O\hat{\rho})$. The true value of the query is given by

$$o(\mathcal{D}) = \mathop{\mathbb{E}}_{\hat{\rho}\sim\mathcal{D}}[o(\hat{\rho})] = \mathop{\mathbb{E}}_{\hat{\rho}\sim\mathcal{D}}[\mathrm{tr}(O\hat{\rho})] = \mathrm{tr}(O\rho), \tag{B41}$$

where the last equality follows by Eq. (B39). This aligns well with the classical statistical query model, since we wish to estimate $o(\mathcal{D}) = \mathrm{tr}(O\rho)$. Moreover, these functions are indeed statistical queries, as we have

$$\mathop{\mathbb{E}}_{\mathcal{S}\sim\mathcal{D}^N}[o(\mathcal{S})] = \frac{1}{N}\sum_{j=1}^{N}\mathop{\mathbb{E}}_{\mathcal{S}\sim\mathcal{D}^N}[\mathrm{tr}(O\hat{\rho}_j)] = o(\mathcal{D}), \tag{B42}$$

where $o(\mathcal{S}) = 1/N \sum_{j=1}^N o(\hat{\rho}_j)$. To complete this analogy with the classical statistical query model, it remains to show that $o(\hat{\rho}) \in [-C, C]$ for some constant C and for $\hat{\rho} \in \mathcal{X}$ when O is a local observable. Let $\hat{\rho}$ be a classical snapshot as in Eq. (B40). In the case of predicting local observables, the random unitary U utilized to construct this classical snapshot is a random tensor product of n single-qubit Pauli gates $U = U_1 \otimes \cdots \otimes U_n$. Moreover, from Appendix A 2 a, the quantum channel $\mathcal{M}$ has a specific form

$$\hat{\rho} = \bigotimes_{j=1}^{n}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j - \mathbb{I}\right) \quad \text{where } \hat{b}_1, \ldots, \hat{b}_n \in \{0,1\}. \tag{B43}$$

We want to show that $|o(\hat{\rho})| = |\mathrm{tr}(O\hat{\rho})|$ is bounded by a constant for a local observable O. Let $O = \tilde{O} \otimes \mathbb{I}^{\otimes(n-k)}$ be a k-local observable, where we assume without loss of generality that O acts nontrivially only on the first k qubits. Then, we can compute the expectation value $\mathrm{tr}(O\hat{\rho})$ by plugging in Eq. (A17),

$$\mathrm{tr}(O\hat{\rho}) = \mathrm{tr}\left(O\bigotimes_{j=1}^{n}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j - \mathbb{I}\right)\right) \tag{B44}$$

$$= \mathrm{tr}\left(\left(\tilde{O}\bigotimes_{j=1}^{k}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j - \mathbb{I}\right)\right)\otimes\left(\bigotimes_{j=k+1}^{n}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j - \mathbb{I}\right)\right)\right) \tag{B45}$$

$$= \mathrm{tr}\left(\left(\tilde{O}\bigotimes_{j=1}^{k}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j-\mathbb{I}\right)\right)\right)\mathrm{tr}\left(\left(\bigotimes_{j=k+1}^{n}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j-\mathbb{I}\right)\right)\right) \tag{B46}$$

$$= \mathrm{tr}\left(\tilde{O}\bigotimes_{j=1}^{k}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j-\mathbb{I}\right)\right). \tag{B47}$$

In the first line, we used Eq. (A17). In the second line, we used that $O = \tilde{O}\otimes\mathbb{I}^{\otimes(n-k)}$. In the third line, we used the multiplicative properties of the trace. Finally, in the last line, we used that $\mathrm{tr}(\hat{\rho}) = 1$, where we can view this the second term in the third line as a classical snapshot over $n-k$ qubits. Now, by properties of the trace norm, we also have

$$|\mathrm{tr}(O\hat{\rho})| \le \left\|\tilde{O}\bigotimes_{j=1}^{k}\left(3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j-\mathbb{I}\right)\right\|_1 \tag{B48}$$

$$\le \left\|\tilde{O}\right\|_\infty\cdot\prod_{j=1}^{k}\left\|3U_j^\dagger\left|\hat{b}_j\right\rangle\!\left\langle\hat{b}_j\right|U_j-\mathbb{I}\right\|_1 \tag{B49}$$

$$\le 3^k, \tag{B50}$$

where in the last line we used the assumption that the spectral norm of $\tilde{O}$ is bounded by 1. Thus, we have that $|o(\hat{\rho})| \le C = 3^k$, where this is a constant because $k = \mathcal{O}(1)$ by definition of a local observable.

With this, we have now shown that the classical shadow framework fits into the classical statistical query setting. Thus, we can apply any classical algorithm for answering adaptive statistical queries to derive our rigorous guarantees. Applying Theorem B3, we see that in order to estimate expectation values of M adaptively chosen k-local observables, we can use

$$N = \mathcal{O}\left(\frac{3^k\sqrt{M\log(1/\epsilon\delta)}}{\epsilon^2}\right) \tag{B51}$$

samples of the unknown n-qubit quantum state ρ. This gives us the first term in the minimum in the statement of Proposition B2.

To obtain the second term of $k\log n$, we note that any observable O can be decomposed in the Pauli basis as $O = \sum_{P\in\{I,X,Y,Z\}^{\otimes n}}\alpha_P P$ for some coefficients α_P. For a k-local observable, there are only $3^k\cdot\binom{n}{k} = \mathcal{O}(n^k)$ possible different Pauli terms P in this decomposition. One can predict expectation values of all of these observables with respect to the unknown state ρ using the guarantees of the classical shadow formalism, which requires

$$N = \mathcal{O}\left(\frac{k\log(n/\delta)}{\epsilon^2}\right). \tag{B52}$$

This follows from Theorem A1, where the shadow norm is constant for local observables (Proposition S3 in Ref. [8]). Because these $3^k\cdot\binom{n}{k}$ observables are fixed beforehand, we can use these guarantees. Expectation values of any (adaptively chosen) local observable can be expressed in terms of expectations of these $\mathcal{O}(n^k)$ observables, so we can predict any (adaptively chosen) local observable using this classical shadow data. Taking the minimum of these two expressions yields the sample complexity claim.

Finally, the time complexity follows because, as discussed previously, each classical snapshot $\hat{\rho}$ can be stored efficiently in classical memory using $\mathcal{O}(n)$ bits. ■

APPENDIX C: PAULI OBSERVABLES

In this Appendix, we shift our attention to focus on predicting properties $\mathrm{tr}(O\rho)$, where ρ is the unknown n-qubit quantum state that we are given copies of and $O\in\{I,X,Y,Z\}^{\otimes n}$ is a Pauli observable. This type of observable was not explicitly considered in the classical shadow formalism [8]. However, Ref. [47] designed a quantum machine learning model uses $N = \mathcal{O}(\log(M)/\epsilon^4)$ copies of ρ to predict expectation values of M n-qubit Pauli observables. We analyze how the rigorous guarantees change when the Pauli observables are allowed to be chosen adaptively. In this Appendix, we prove Theorem 2, which provides sample complexity upper and lower bounds for this task. We restate the theorem below.

Theorem C1 (Pauli observables, detailed restatement of Theorem 2). There exists an analyst $\mathcal{A}$ and a density matrix ρ on $n = \Omega(M^{3/2})$ qubits such that any $[0.99, 1-(1/2000\sqrt{M})]$-accurate quantum mechanism $\mathcal{M}$ estimating expectation values of M adaptively chosen Pauli observables requires at least

$$N = \Omega(\sqrt{M}) \tag{C1}$$

samples of the quantum state ρ. Meanwhile, for any $\epsilon,\delta\in(0,1/2)$ such that $\epsilon\delta \ge 4e^{-M/\log^2 M\log\log^4 M}$, there exists an (ϵ,δ)-accurate quantum mechanism $\mathcal{M}$ that can estimate expectation values of M adaptively chosen Pauli observables using

$$N = \mathcal{O}\left(\frac{\min\{\sqrt{M\log(1/\epsilon\delta)}, n\}}{\epsilon^4}\right) \tag{C2}$$

samples of the unknown n-qubit quantum state ρ. Moreover, the mechanism is computationally efficient and runs in time $\mathrm{poly}(N, n, \log(1/\delta))$ per observable.

The techniques used in this section are similar to Appendix B. In Appendix C 1, we prove the sample complexity lower bound, which states that in general, there does not exist a quantum mechanism that can accurately estimate M adaptively chosen Pauli observables with $\mathrm{poly}\log(M)$ copies of ρ. In Appendix C 2, we present a quantum algorithm achieving a matching sample complexity upper bound for answering adaptively chosen Pauli observables.

1. Lower bound

In this Appendix, we prove the sample complexity lower bound, which states that in general, there does not exist a quantum mechanism that can accurately estimate M adaptively chosen Pauli observables with $\mathrm{poly}\log(M)$ copies of ρ. This gives a stark contrast to the algorithm in Ref. [47], which can accomplish this task using $\mathcal{O}(\log M/\epsilon^4)$ copies of ρ. We prove the following lower bound.

Proposition C1 (Pauli observables, lower bound). There exists an analyst $\mathcal{A}$ and a density matrix ρ on $n = \Omega(M^{3/2})$ qubits such that any $[0.99, 1-(1/2000\sqrt{M})]$-accurate quantum mechanism $\mathcal{M}$ estimating expectation values of M adaptively chosen Pauli observables requires at least

$$N = \Omega(\sqrt{M}) \tag{C3}$$

samples of the quantum state ρ.

This proof is similar to the proof of Proposition B1, which gave a sample complexity lower bound for the task of predicting local observables. Again, this proof relies on interactive fingerprinting codes [46] (see Appendix A 3) and is similar to lower bounds in classical adaptive data analysis [15,32,45]. Moreover, our construction achieves this lower bound even when the analyst can only query Pauli observables $\{I, Z\}^{\otimes n}$ and when the mechanism is given N copies of a diagonal n-qubit density matrix ρ.

Our proof will require some cryptographic concepts, which we review in Appendix A 4.

The idea is again to reduce our problem to IFPC by designing an analyst $\mathcal{A}$ such that if the mechanism $\mathcal{M}$ answers all of the analyst's queries accurately, then we contradict the definition of $\mathcal{F}$ being a collusion-resilient fingerprinting code. We argue that in our attack, the mechanism $\mathcal{M}$ adheres to the same conditions as the adversary $\mathcal{P}$ in the IFPC game (i.e., only having access to information from the nonaccused colluding users) so that the guarantees of the interactive fingerprinting codes hold. These guarantees state that the fingerprinting code $\mathcal{F}$ can force the adversary $\mathcal{P}$ to be inconsistent, which also implies that $\mathcal{M}$ also answers one of the analyst's queries incorrectly.

In this case, to show that $\mathcal{M}$ only has access to information from the nonaccused colluding users, we use encryption, similarly to Ref. [32]. Namely, we will encrypt all information given to the mechanism via a private-key encryption scheme (reviewed in Appendix A 4) and encode the secret keys needed to decrypt in the samples given to the mechanism. In this way, the mechanism can only decrypt the information corresponding to the colluding users, and all other information is hidden by the security of the private-key encryption scheme.

Proof of Proposition C1. We follow the same proof structure as the proof of Proposition B1. Throughout the proof, we use $\mathcal{A}$ for the analyst, $\mathcal{M}$ for the mechanism, $\mathcal{F}$ for the fingerprinting code, and $\mathcal{P}$ for the adversary in the IFPC game. Let d be the number of users in the setting of fingerprinting codes. Let M be the number of observables that the analyst queries. Also, throughout the proof, we use the notation $x_{i:j}$, where $x = (x_1, \ldots, x_m)$ is a vector, to denote $x_{i:j} \triangleq (x_i, \ldots, x_j)$, where $j \geq i$.

Recall that the mechanism is given copies of an unknown quantum state ρ, with which it is asked to predict $\mathrm{tr}(O_i\rho)$ for an (adaptively chosen) query O_i from the analyst. We first define an encryption scheme that we will use. We then precisely define the adversarial choice of ρ based on this encryption scheme. Subsequently, we will present an algorithm for $\mathcal{A}$ that thwarts any mechanism, even when only querying Pauli observables in $\{I, Z\}^{\otimes n}$.

Encryption scheme: The encryption scheme is a version of the one-time pad (Definition A5). Namely, we use a variation of the one-time pad with $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ defined as follows. Let $\lambda = d$ be the security parameter. Let the key and message spaces be $\{0,1\}^d$ and let the ciphertext space be $\{0,1\}^{2d}$.

(a) $\mathsf{Gen}(1^d)$: Output a uniformly random bitstring $\mathsf{sk} \in \{0,1\}^d$.

(b) $\mathsf{Enc}_{\mathsf{sk}}(m)$: Encrypt $m \in \{0,1\}^d$ as $c = c_1 \cdots c_{2d} \in \{0,1\}^{2d}$, where

$$c_i = \begin{cases} 10 & \text{if } m_i \oplus \mathsf{sk}_i = 1, \\ 01 & \text{if } m_i \oplus \mathsf{sk}_i = 0. \end{cases} \tag{C4}$$

(c) $\mathsf{Dec}_{\mathsf{sk}}(c)$: On input $c \in \{0,1\}^{2d}$, parse c as $c = c_{1:2} \cdots c_{2d-1:2d}$, i.e., consider consecutive pairs of bits in c. Map c to $c' = c'_1 \cdots c'_d$, where

$$c'_i = \begin{cases} 1 & \text{if } c_i = 10, \\ 0 & \text{if } c_i = 01. \end{cases} \tag{C5}$$

Decrypt c as $m = c' \oplus \mathsf{sk} \in \{0,1\}^d$.

Note that this is indeed the same as the one-time pad, but with an extra step that maps the output of encryption to a bitstring of length $2d$ and correspondingly maps back for decryption.

Defining the density matrix: We define the density matrix ρ on n qubits, where we specify n shortly. This density matrix is diagonal so that it corresponds to a classical distribution. As in the proof of Proposition B1, we encode an adversarial choice of classical distribution into this density matrix that takes inspiration from Ref. [32]. Moreover, note that we use a diagonal density matrix because the guarantees from interactive fingerprinting codes hold for classical adversaries (i.e., classical mechanisms). We later remark that quantum mechanisms can be reduced to classical mechanisms.

Let $n = \lceil \log d \rceil + 2Md$. We index the first $\lceil \log d \rceil$ qubits by an integer in $\{1, \ldots, \lceil \log d \rceil\}$, and we index the remaining qubits by tuples (k, j) such that $k \in [2d], j \in [M]$. This reflects the underlying structure that the last qubits are M groups of $2d$ qubits. In this way, j indexes the group of qubits and k indexes the specific qubit within the j th group. We describe the intuition for this structure later.

Since ρ is a diagonal density matrix, we can write

$$\rho \triangleq \sum_{Q \in \{0,1\}^n} p(Q) \, |Q\rangle\langle Q| \,, \tag{C6}$$

where $|Q\rangle$ is a computational basis state so that $Q \in \{0,1\}^n$. We can index the entries of the bitstring Q as previously described:

$$Q = Q_1 \cdots Q_{\lceil \log d \rceil} Q_{(1,1)} \cdots Q_{(2d,1)} \cdots Q_{(1,M)} \cdots Q_{(2d,M)} \in \{0,1\}^n. \tag{C7}$$

The state $|Q\rangle$ can be viewed as follows. The first $\lceil \log d \rceil$ qubits index the user from the setting of fingerprinting codes, where recall there are d users. Explicitly, each integer in $\{1, \ldots, d\}$ can be represented by in binary using $\lceil \log d \rceil$ bits. The bitstring $Q_1 \cdots Q_{\lceil \log d \rceil} \in \{0,1\}^{\lceil \log d \rceil}$ encodes this binary representation of the user's index $i \in \{1, \ldots, d\}$. There are also M groups of $2d$ qubits. Namely, for $j \in [M]$, these are the groups of qubits corresponding to $Q_{(1,j)} \cdots Q_{(2d,j)}$. Each group corresponds to a round j of the interaction between the analyst and mechanism. Moreover, each pair of qubits corresponding to $Q_{(2i-1,j)}$ and $Q_{(2i,j)}$ represent user i in round j.

We want to define the distribution $p(Q)$ in a specific way such that the first $\lceil \log d \rceil$ qubits determine the state of the remaining ones and such that it encodes the secret keys for our encryption scheme. Let $\mathsf{sk}^j \leftarrow \mathsf{Gen}(1^d)$ be the secret key for round $j \in [M]$. We generate a new secret key for each round of interaction between the analyst and mechanism. We use sk^j_i to denote the ith bit of the secret key sk^j for round j, i.e., the one-bit secret key for the ith user in round j. We define the following sets of binary strings:

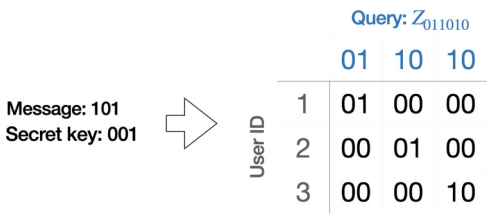

FIG. 4. Example construction for Pauli observables. We consider an example where the query (message) is $q^j = 101$ and the randomly generated secret key is 001. The table shows the quantum state, with each row being a computational basis state $|Q\rangle$ corresponding to a user (which is encoded in the first $\lceil \log d \rceil$ qubits of $|Q\rangle$). The secret key is encoded into the quantum state as shown in the table. We group qubits into pairs in each for clarity, which are defined by the encoding rule for the density matrix. By the encryption rule, the analyst will query the Pauli observable Z_{b^j} where $b^j = 011010$. One can verify that evaluating Z_{b^j} on each state recovers the original message.

$$R_0 \triangleq \{Q \in \{0,1\}^n : Q_{(2i-1,j)} Q_{(2i,j)} = 10, \ \forall j \in [M], \text{ where } Q_1 \cdots Q_{\lceil \log d \rceil} = i \in [d] \text{ and } \mathsf{sk}^j_i = 0\}, \tag{C8}$$

$$R_1 \triangleq \{Q \in \{0,1\}^n : Q_{(2i-1,j)} Q_{(2i,j)} = 10, \ \forall j \in [M], \text{ where } Q_1 \cdots Q_{\lceil \log d \rceil} = i \in [d] \text{ and } \mathsf{sk}^j_i = 1\}, \tag{C9}$$

$$R_2 \triangleq \{Q \in \{0,1\}^n : Q_{(2i-1,j)} Q_{(2i,j)} = 00, \ \forall i \in [d], \ \forall j \in [M]\}, \tag{C10}$$

$$R \triangleq R_0 \cup R_1 \cup R_2. \tag{C11}$$

With this, we can define $p(Q)$ as follows:

$$p(Q) \triangleq \begin{cases} \frac{1}{d} & \text{if } Q \in R, \\ 0 & \text{otherwise.} \end{cases} \tag{C12}$$

This defines precisely how the secret key is encoded into the density matrix. We have completed the definition of this density matrix $\rho = \sum_{Q \in \{0,1\}^n} p(Q) \, |Q\rangle\langle Q|$. An example is illustrated in Fig. 4.

Defining the analyst's behavior: With this, we can define the analyst's adversarial behavior. We will show that the analyst can cause the mechanism to answer queries inaccurately when querying Pauli observables in $\{I, Z\}^{\otimes n}$. We use Z_b for $b \in \{0,1\}^n$ to denote the observable $Z_b = O_1 \otimes \cdots \otimes O_n$, where $O_i = Z$ if $b_i = 1$ and $O_i = \mathbb{I}$ otherwise. Moreover, we use the notation $b^j \in \{0,1\}^n$ to denote the choice of query in the jth round. We denote $b^j_{(i,j)}$ to denote the entry of the vector b^j corresponding to the qubit indexed by (i,j). We also use the notation $b^j_{(i:i',j)}$ to denote the entries of the vector b^j corresponding to the qubits indexed by $(i,j), (i+1,j), \ldots, (i',j)$ for $i' \geq i$ and $i, i' \in [2d]$.

ALGORITHM 5. Real attack (Pauli version).

Let the number of users be $d = 2000N$.
Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be the encryption scheme defined in the text.
Generate secret keys $\mathsf{sk}^j \leftarrow \mathsf{Gen}(1^d)$ for each round $j \in [M]$.
Give $\rho^{\otimes N}$ to the mechanism $\mathcal{M}$ for ρ defined as in the text.
Initialize N-collusion resilient fingerprinting code $\mathcal{F}$ for n users and length $M = \mathcal{O}(N^2)$.
Let $T^0 = \emptyset$.
for $j = 1$ *to* M **do**
  Let $c^j \in \{0,1\}^d$ be chosen by $\mathcal{F}$.
  Define $q^j \in \{0,1\}^d$ such that $q_i^j = c_i^j$ if $i \notin T^{j-1}$ and 0 otherwise.
  Analyst $\mathcal{A}$ chooses query Z_{b^j} where $b^j_{(1:2d,j)} = \mathsf{Enc}(\mathsf{sk}^j, q^j)$ and b^j is 0 for the other entries.
  $\mathcal{A}$ queries Z_{b^j} to $\mathcal{M}$.
  $\mathcal{M}$ outputs response a^j, rounds a^j to $\bar{a}^j \in \{0,1\}$, and gives $\bar{a}^j$ to $\mathcal{F}$.
  $\mathcal{F}$ accuses $I^j \subseteq [d]$. Set $T^j \leftarrow T^{j-1} \cup I^j$.
end

Again, note that the guarantees from interactive fingerprinting codes will only give us guarantees for classical mechanisms. Thus, we argue that these types of queries also have a classical analog. Namely, measuring the n-qubit state ρ is equivalent to flipping n biased coins and obtaining a sample outcome $x \in \{0,1\}^n$. Then, querying the observable Z_b is the same as querying the parity over a selected group of coins indexed by b, i.e., $p(b,x) = b \cdot x \bmod 2$.

The analyst's full attack is given in Algorithm 5. Here, we use T^j to denote the set of users accused by the fingerprinting code $\mathcal{F}$ up to and including round j. We also denote $S^j = \mathcal{S} \setminus T^j$, where $\mathcal{S}$ is the set of colluding users. Thus, S^j is the set of colluding users that have not yet been accused.

Note that this attack is almost the same as Algorithm 3 from the local observables sample complexity lower bound. The only difference is that the queries that the analyst $\mathcal{A}$ gives to the mechanism $\mathcal{M}$ are encrypted. In particular, in round j, given a query $q^j \in \{0,1\}^d$, the analyst instead gives the query Z_{b^j} to the mechanism, where $b^j_{(1:2d,j)} = \mathsf{Enc}_{\mathsf{sk}^j}(q^j)$ and b^j is 0 for the remaining entries. This is so that the mechanism only has access to information from the set of colluding users S^j that have not yet been accused, as we will justify shortly. This encryption replaces the random permutations from the proof of Proposition B1.

We also justify that for an input query q^j, encrypting it to obtain b^j, and querying Z_{b^j} as in Algorithm 5 still results in the same query q^j. Let $Q^{(i)} \in \{0,1\}^n$ be the bitstring from measuring ρ in the computational basis, where $Q_1^{(i)} \cdots Q_{\lceil \log d \rceil}^{(i)} = i$ and let $z^j_{(2i-1:2i,j)} = Q^{(i)}_{(2i-1,j)} Q_{(2i,j)}$. Let $(\mathsf{Gen}^1, \mathsf{Enc}^1, \mathsf{Dec}^1)$ be the encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ defined above but for security parameter $\lambda = 1$. In other words, we use $(\mathsf{Gen}^1, \mathsf{Enc}^1, \mathsf{Dec}^1)$ to denote the special case of $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ when the key space and message space are $\{0,1\}$ and the ciphertext space is $\{0,1\}^2$. Let $\mathsf{sk}^j \leftarrow \mathsf{Gen}(1^d)$. Note that because sk^j is generated uniformly at random in Gen, then each bit sk^j_i of sk^j is also a key that could be generated from Gen^1. Then, notice that by definition of the encryption scheme and the set R [defined in Eq. (C11)], we have $z^j_{(2i-1:2i,j)} = \mathsf{Enc}^1_{\mathsf{sk}^j_i}(1)$. Moreover, we can also write

$$\mathsf{Dec}^1_{\mathsf{sk}^j_i}(c) = \mathsf{Enc}^1_{\mathsf{sk}^j_i}(1) \cdot c \bmod 2, \tag{C13}$$

which can be easily checked by definition of the encryption scheme. Recall that querying the observable Z_b is the same as querying the parity $p(b,x) = b \cdot x \bmod 2$, where x is a sample outcome from measuring ρ. In this case, then by definition of b^j in Algorithm 5, we have

$$b^j_{(2i-1:2i,j)} \cdot z^j_{(2i-1:2i,j)} = \mathsf{Dec}_{\mathsf{sk}^j_i}(\mathsf{Enc}_{\mathsf{sk}^j_i}(q^j_i)) = q^j_i. \tag{C14}$$

In this way, we see that querying Z_{b^j} is the same as querying the unencrypted q^j. Thus, the encryption here is only hiding information but not obscuring what we wish to query.

Reduction to IFPC: We now show that our proposed attack follows the same interaction model as the IFPC game. Crucially, we claim that the mechanism can only access information from the set of colluding users S^j that have not been accused, i.e., the information $c^j_{S^j}$ (as required in Algorithm 1).

To show this, we construct a simulated attack, given in Algorithm 6. The simulated attack is merely fictitious, as it relies on the analyst knowing the set of colluding users $\mathcal{S}$. In particular, for $i \notin \mathcal{S}$, the analyst's simulated attack will set $q^j_i = 0$, which is not possible in the real attack because the analyst does not know $\mathcal{S}$. The analyst first simulates sampling from the density matrix ρ that would be given to the mechanism in the real attack (Algorithm 5) and gives this simulated data to the mechanism. In this way, because the analyst knows $\mathcal{S}$, the dataset that is given to the mechanism in the simulated attack (Algorithm 4) only contains

ALGORITHM 6. Simulated attack (Pauli version).

Let the number of users be $d = 2000N$ and let $\mathcal{U}$ be the uniform distribution over $[d]$.
Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be the encryption scheme defined in the text.
Choose colluding users $\mathcal{S} \sim \mathcal{U}^N$ and give $\mathcal{S}$ to the analyst $\mathcal{A}$.
$\mathcal{A}$ generates secret keys $\mathsf{sk}^j \leftarrow \mathsf{Gen}(1^d)$ for each round $j \in [M]$.
$\mathcal{A}$ simulates sampling from the density matrix ρ by generating $\mathcal{S}' = \{|Q\rangle\}$ corresponding to $\mathcal{S}$.
Give $\mathcal{S}'$ to $\mathcal{M}$.
Initialize N-collusion resilient fingerprinting code $\mathcal{F}$ for n users and length $M = \mathcal{O}(N^2)$.
Let $S^1 = \mathcal{S}$.
for $j = 1$ *to* M **do**
  Let $c^j \in \{0,1\}^d$ be chosen by $\mathcal{F}$. Give $c^j_{S^j}$ to $\mathcal{A}$.
  Define $q^j \in \{0,1\}^d$ such that $q^j_i = c^j_i$ if $i \in S^j$ and 0 otherwise.
  Analyst $\mathcal{A}$ chooses query Z_{b^j} where $b^j_{(1:2d,j)} = \mathsf{Enc}(\mathsf{sk}^j, q^j)$ and b^j is 0 for the other entries.
  $\mathcal{A}$ queries Z_{b^j} to $\mathcal{M}$.
  $\mathcal{M}$ outputs response a^j. Round a^j to $\bar{a}^j \in \{0,1\}$, and give $\bar{a}^j$ to $\mathcal{F}$.
  $\mathcal{F}$ accuses $I^j \subseteq [d]$. Set $S^{j+1} \leftarrow S^j \setminus I^j$.
end

information about the colluding users. Explicitly, the analyst creates computational basis states $|Q\rangle$, where $Q \in R$ [defined in Eq. (C11)], where i is restricted to be in $\mathcal{S}$ instead. Moreover, the simulated interaction with the sampled states was constructed with only knowledge of $c^j_{S^j}$ and contains no information about c^j outside of S^j. Thus, the simulated attack indeed operates in the correct IFPC interaction model required for fingerprinting code guarantees to hold.

To show that our proposed attack (Algorithm 5) follows the same interaction model as the IFPC game, we claim that the mechanism cannot distinguish between our proposed attack and the simulated attack (Algorithm 6), which as described must be in the IFPC interaction model. We can show that any mechanism cannot distinguish between the simulated and real attacks by noting the following:

(1) The distribution over the measurement outcomes for qubits indexed by (k,j) for $k \in [2d], j \in [M]$ is the same.
(2) For a given round j, the distribution of the queries b^j is the same.

The first point is true because both the users (by definition of ρ) and secret keys are chosen uniformly at random in both the simulated and real attacks. Together, these completely determine the rest of the $2dM$ qubits, so their distributions are the same.

The second point is true because of the following argument. Notice that the mechanism only sees the secret keys for $i \in \mathcal{S}$, i.e., it observes $z^j_{(2i-1:2i,j)} = \mathsf{Enc}^1(\mathsf{sk}^j_i, 1)$ only for $i \in \mathcal{S}$. This is clear by construction. Moreover, notice that the only difference between the real and simulated attacks with respect to the query q^j is that q^j (and thus b^j) is only different outside of $\mathcal{S}$. Thus, although the mechanism can know the secret keys for $i \in \mathcal{S}$, b^j is in fact the same for $i \in \mathcal{S}$, so the distribution is the same in this case. On the other hand, for $i \notin \mathcal{S}$, the queries q^j may be different. However, recall that b^j is the encryption of q^j, where our encryption scheme is defined to be a variant of the one-time pad. Thus, by the perfect secrecy of the one-time pad, then the distribution over b^j are identical in this case as well.

Thus, the mechanism's views are identical in both attacks. This implies that the probabilities of inconsistency are the same, i.e.,

$$\Pr_{\text{Real}}[\exists j \in [M] : \forall i \in [d], \bar{a}^j \neq c^j_i] = \Pr_{\text{Sim}}[\exists j \in [M] : \forall i \in [d], \bar{a}^j \neq c^j_i]. \tag{C15}$$

The rest of the proof is the same as the proof of Proposition B1. ■

Just as in the local observables case, there is also an analogous classical adaptive data analysis problem for which we may recycle our proof in order to obtain a new lower bound. Here, the distribution $\mathcal{D}$ is over an n-dimensional data universe $\mathcal{X} \subseteq \{0,1\}^n$, and the analyst queries from the family of *parity queries*, which are defined the same way as noted previously in this Appendix: $q_b(x) = p(b,x) = b \cdot x$, where $b \in \{0,1\}^n$ and $x \in \mathcal{X}$. Our proof for the Pauli observables lower bound relies only on classical constructions, so we may recycle our proof: the parity query q_b corresponds to the observable Z_b and the classical distribution $\mathcal{D}$ can be mapped onto a density matrix.

Corollary C1 (Parity queries, lower bound). There exists an analyst $\mathcal{A}$ and classical distribution $\mathcal{D}$ acting on $\{0,1\}^n$, where $n = \Omega(M^{3/2})$ such that any $[0.99, 1 - (1/2000\sqrt{M})]$-accurate classical mechanism $\mathcal{M}$ answering M adaptively chosen parity queries requires at least $N = \Omega(\sqrt{M})$ from $\mathcal{D}$.

2. Upper bound

In this Appendix, we prove a sample complexity upper bound for predicting expectation values $\mathrm{tr}(P_i\rho)$ for $i \in [M]$ of adaptively chosen n-qubit Pauli observables $P_i \in \{I, X, Y, Z\}^{\otimes n}$ with respect to an unknown quantum state ρ. We design a mechanism that can successfully estimate these expectation values, even when the observable P_i is chosen adaptively. In the case where P_i are chosen nonadaptively, one can use the algorithm in Ref. [47]. However, we will need to use a different algorithm for when P_i are chosen adaptively. Moreover, while the proof of our sample complexity lower bound (Appendix C 1) only required the analyst to query observables in $\{I, Z\}^{\otimes n}$, our algorithm works for an analyst querying any Pauli observables. This upper bound matches our lower bound from Proposition C1 exactly in M scaling. In particular, we prove the following proposition.

Proposition C2 (Pauli observables, upper bound). For every $\epsilon, \delta \in (0, 1/2)$ such that $\epsilon\delta \geq 4e^{-M/\log^2 M \log\log^4 M}$, there is exists an (ϵ, δ)-accurate mechanism $\mathcal{M}$ that can estimate expectation values of M adaptively chosen Pauli observables using

$$N = \mathcal{O}\left(\frac{\min\{\sqrt{M\log(1/\epsilon\delta)}, n\log(1/\delta)\}}{\epsilon^4}\right) \tag{C16}$$

samples of the unknown quantum state ρ. The algorithm runs in time $\mathrm{poly}(N, n, \log(1/\delta))$ per query.

The idea of the proof is similar to the proof of Proposition B2. Namely, we can replace a classical subroutine in the algorithm in Ref. [47] with an algorithm from classical adaptive data analysis [17] (see Theorem B3). This subroutine is the only part of the algorithm from Ref. [47] that is susceptible to adaptivity, so by utilizing the algorithm in Ref. [17] instead, we can protect against this. We refer to Appendix A 2 b for a review of the algorithm from Ref. [47].

Proof of Proposition C2. Recall from Appendix A 2 b that the algorithm from Ref. [47] estimates $\mathrm{tr}(P_i\rho)$ in two steps: (1) estimate the magnitude $|\mathrm{tr}(P_i\rho)|^2$ and (2) determine $\mathrm{sign}(\mathrm{tr}(P_i\rho))$.

As discussed in Appendix A 2 b, the first step is completed by estimating expectations of

$$q_P(S^{(t)}) = \prod_{k=1}^{n} \mathrm{tr}\left((\sigma_k \otimes \sigma_k) S_k^{(t)}\right) \in \{\pm 1\}, \tag{C17}$$

where $S^{(t)} = \{S_k^{(t)}\}_{k\in[n]}$ are classical measurement results with $t \in [N_1]$ the t'th round of measurement and $P = \sigma_1 \otimes \cdots \sigma_n$ [Eq. (A21) in Appendix A 2 b]. Huang *et al.* [47] show that the expectation value of q_P is exactly the magnitude $|\mathrm{tr}(P\rho)|^2$ and thus estimates this expectation to within ϵ^2 error using the empirical mean of q_P. However, taking the empirical mean is not sufficient in the adaptive case, so we will need to use a different algorithm, similarly to the proof of Proposition B2.

Our key observation is to notice that q_P is a statistical query with values in $\{-1, 1\}$. Thus, we can apply the algorithm from Theorem B3 with $C = 1$ to estimate the expectation of each q_{P_i}, i.e., the magnitudes $|\mathrm{tr}(P_i\rho)|^2$. Applying Theorem B3, we can estimate the magnitudes $|\mathrm{tr}(P_i\rho)|^2$ for M adaptively chosen Pauli observables P_i to within ϵ^2 error using

$$2N_1 = \mathcal{O}\left(\frac{\sqrt{M\log(1/\epsilon\delta)}}{\epsilon^4}\right) \tag{C18}$$

copies of the unknown quantum state ρ.

Note that this still holds despite the subsequent post-processing of the estimation, i.e., taking the square root and estimating $\mathrm{sign}(\mathrm{tr}(P\rho))$, since none of them reveal to the analyst any new information about the dataset of measurement outcomes $\{S_k^{(t)}\}$ (where $\mathrm{sign}(\mathrm{tr}(P\rho))$ is estimated using a separate set of samples). Thus, effectively, the analyst only receives information about the dataset $\{S_k^{(t)}\}$ from the output of the Theorem B3 algorithm. Hence, N_1 samples as defined above are indeed sufficient to predict magnitudes even for adaptively chosen observables.

As remarked in Appendix A 2 b, the second step of determining $\mathrm{sign}(\mathrm{tr}(P\rho))$ is already resistant to adaptivity since it reuses copies of ρ via an argument involving the quantum union bound [69]. Thus, we can use the same

$$N_2 = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^2}\right) \tag{C19}$$

as in the original algorithm [47] for determining $\mathrm{sign}(\mathrm{tr}(P_i\rho))$ for $1 \leq i \leq M$. Putting everything together, the total sample complexity is

$$N = 2N_1 + N_2 = \mathcal{O}\left(\frac{\sqrt{M\log(1/\epsilon\delta)}}{\epsilon^4}\right). \tag{C20}$$

To obtain the second term in the minimum in the statement of Proposition C2, we can also protect against adaptivity by simply using the original nonadaptive algorithm [47] to predict expectation values of all 4^n Pauli observables. By Theorem A2, this requires

$$N = \mathcal{O}\left(\frac{n\log(1/\delta)}{\epsilon^4}\right). \tag{C21}$$

Because this guarantees that the algorithm will accurately estimate expectation values of all possible Pauli observables, it does not matter if an observable is chosen adaptively. Taking the minimum of the two expressions (C20) and (C21) gives the claim.

Finally, the computational complexity follows because the dataset of measurement outcomes $S^{(t)}$ can be stored efficiently in classical memory using $2n$ bits. This is clear because, as described in Appendix A 2 b, $S^{(t)}$ is obtained by measuring $\rho \otimes \rho$ in the Bell basis. ■

APPENDIX D: ADAPTIVE ALGORITHMS USING CLASSICAL SHADOWS

In this Appendix, we introduce a series of algorithms for answering adaptive queries using classical shadows [8]. Namely, when observables are chosen adaptively, we provide new algorithms for shadow tomography [7] in Appendix D 1 and quantum threshold search (originally called secret acceptor in Ref. [88]) in Appendix D 2. We will use these algorithms as components of our algorithm for predicting adaptively chosen bounded-Frobenius-norm properties in Appendix E. However, we also believe them to be sufficiently interesting algorithms in their own right, as they apply to arbitrary observables and their nonadaptive versions are often used as subroutines in common quantum algorithms. Throughout this Appendix, we assume that we have access to a classical shadow dataset $\mathcal{S}(\rho; N) = \{\hat{\rho}_1, \ldots, \hat{\rho}_N\}$, where ρ is the unknown quantum state that we have samples of. We refer the reader to Appendix A 2 a for a review of the classical shadow formalism.

1. Adaptive classical shadow tomography

We again consider receiving adaptively chosen observables $O_1, \ldots, O_M$ and are tasked with estimating $o_i = \mathrm{tr}(O_i\rho)$ for all $1 \le i \le M$ given samples of the unknown quantum state ρ. Previously, in Appendix B 4, we saw that if $o_i(\hat{\rho}) = \mathrm{tr}(O_i\hat{\rho})$ is bounded in some range $[-C, C]$ where C is a constant and $\hat{\rho}$ is a classical snapshot of ρ, then we can use classical adaptive data analysis algorithms [17] that apply to statistical queries $q : \mathcal{X} \mapsto [-C, C]$ to obtain rigorous sample complexity guarantees. However, in general, it is not guaranteed that the estimator $o_i(\hat{\rho})$ will have bounded range. For instance, in the case of bounded-Frobenius-norm observables, o_i has an exponentially large range, and we are only given that its variance is bounded by a constant B (Lemma 1 and Proposition 1 in Ref. [89]). For classical shadows predicting arbitrary observables, we need to consider algorithms that apply to more general estimators. We consider two algorithms for this task, one of which is computationally efficient. These two algorithms have different sample complexity guarantees, so we present both, which may be useful in different practical scenarios. Both of our algorithms utilize differentially private algorithms from Ref. [19].

a. Computationally efficient mechanism

First, we present a computationally efficient algorithm for predicting adaptively chosen properties with sample complexity $\tilde{O}(\sqrt{M})$. The algorithm is simply a modification of the original median of means [67,68] protocol used in Ref. [8], where the only difference is that it performs a truncation step and then estimates the median in a differentially private manner. The algorithm can be outlined as follows:

(1) Break the set of classical shadows into K batches of size N, and compute

$$\{\hat{o}_i^{(1)}, \ldots, \hat{o}_i^{(K)}\} \text{ where } \hat{o}_i^{(k)} = \frac{1}{N} \sum_{j=N(k-1)+1}^{Nk} \mathrm{tr}(O_i\hat{\rho}_j). \tag{D1}$$

(2) Let $B = \max_{i=1,\ldots,M} \|O_i - (\mathrm{tr}(O_i)/2^n)\mathbb{I}\|_{\mathrm{shadow}}^2$. Truncate each $\hat{o}_i^{(k)}$ to lie in the range $[-2\sqrt{B/N} - 1, 2\sqrt{B/N} + 1]$, i.e., collect

$$\{\tilde{o}_i^{(1)}, \ldots, \tilde{o}_i^{(K)}\} \quad \text{where } \tilde{o}_i^{(k)} = \begin{cases} -2\sqrt{B/N} - 1 & \text{if } \hat{o}_i^{(k)} < -2\sqrt{B/N} - 1, \\ 2\sqrt{B/N} + 1 & \text{if } \hat{o}_i^{(k)} > 2\sqrt{B/N} + 1, \\ \hat{o}_i^{(k)} & \text{otherwise.} \end{cases} \tag{D2}$$

(3) Use any differentially private algorithm for computing the approximate median of $\{\tilde{o}_i^{(1)}, \ldots, \tilde{o}_i^{(K)}\}$. For example, this can be done via an application of the exponential mechanism [90] (e.g., Refs. [19,91]). Then, return the result $\hat{o}_i(N, K)$.

Our sample complexity guarantee then follows directly from Theorem 1.3 in Ref. [19] and the remarks following it. This sample complexity upper bound is looser than that of Theorem B3 that we used in Appendix B 4. In particular, it has an additional $\log M$ factor. However, it applies more

generally and allows us to bound the sample complexity in terms of the variance of the estimator. In the setting of classical shadows, the variance is bounded by the shadow norm (Lemma 1 in Ref. [8]). Thus, we have the following result.

Theorem D1 (Computationally efficient adaptive classical shadows). For a fixed measurement primitive $\mathcal{U}$, a sequence of adaptively chosen $2^n \times 2^n$ Hermitian matrices $O_1, \ldots, O_M$, and accuracy parameters $\epsilon, \delta \in [0,1]$, set

$$K = \mathcal{O}\left(\sqrt{M \log(1/\delta)} \log(M/\delta\epsilon)\right) \quad \text{and}$$
$$N = \mathcal{O}\left(\frac{1}{\epsilon^2} \max_{i=1,\ldots,M} \|O_i - \frac{\mathrm{tr}(O_i)}{2^n}\mathbb{I}\|^2_{\mathrm{shadow}}\right), \tag{D3}$$

where $\|\cdot\|_{\mathrm{shadow}}$ is the shadow norm. Then, there exists an efficient algorithm using NK independent classical shadows that can accurately estimate all the properties,

$$|\hat{o}_i(N,K) - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \text{for all } 1 \leq i \leq M, \tag{D4}$$

with probability at least $1-\delta$. The running time is $\mathrm{poly}(N, m, \log(1/\epsilon))$, where m is the number of bits required to store a classical shadow.

Here, a measurement primitive is an ensemble of unitaries that is tomographically complete [8].

b. Private multiplicative weights

A simple modification of the algorithm from Appendix D 1 a can yield the desired $\mathcal{O}(\log(M))$ sample complexity scaling for predicting adaptively queried observables. However, we exchange this favorable sample complexity at the cost of computational complexity, as this new algorithm is computationally inefficient.

In particular, to achieve the $\mathcal{O}(\log M)$ sample complexity, one only needs to modify Step 3 in the algorithm from Appendix D 1 a. Specifically, the problem of finding the approximate median in Step 3 can be reduced to answering a sequence of statistical queries via another algorithm (e.g., Refs. [19,92]). Then, one can simply apply the private multiplicative weights algorithm [31] to answer these statistical queries, which gives an alternative method for solving Step 3. This same approach gives a computationally inefficient algorithm in Ref. [19], so our results in the classical shadow setting follow directly from Theorem 5.4 in Ref. [19].

Theorem D2 (Computationally inefficient adaptive classical shadows). For a fixed measurement primitive $\mathcal{U}$, a sequence of adaptively chosen $2^n \times 2^n$ Hermitian matrices $O_1, \ldots, O_M$, and accuracy parameters $\epsilon, \delta \in [0,1]$, there exists an algorithm that uses

$$N = \mathcal{O}\left(\frac{\sqrt{m \log(1/\delta)} \cdot \log(M/\epsilon\delta)}{\epsilon^3} \cdot \max_{i=1,\ldots,M} \|O_i - \frac{\mathrm{tr}(O_i)}{2^n}\mathbb{I}\|^3_{\mathrm{shadow}}\right) \tag{D5}$$

samples to accurately estimate all the properties,

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \text{for all } 1 \leq i \leq M, \tag{D6}$$

with probability at least $1-\delta$. Here, m is the number of bits needed to store a classical shadow.

In particular, when the measurement primitive $\mathcal{U}$ consists of random n-qubit Clifford circuits and tensor products of random single-qubit Clifford circuits, then $m = \mathcal{O}(n^2)$ and $m = \mathcal{O}(n)$, respectively. Note that if $m = \mathcal{O}(n^2)$, this improves upon the shadow tomography protocol of Ref. [14] by a factor of $\log M/\epsilon$ at the cost of the shadow norm term. We will use this algorithm as one of the two main components for our algorithm in Appendix E.

2. Threshold search with classical shadows

In this Appendix, we consider a problem called *threshold search* and provide a new algorithm solving this problem. Threshold search was first introduced by Ref. [88] where it was known as "secret acceptor." In this problem, we receive a sequence of adaptively chosen queries $(O_1, \theta_1), \ldots, (O_M, \theta_M)$ in an online fashion. Here, O_i are the observables and $\theta_i \in [0,1]$ are *thresholds*. Given a budget ℓ, the task is to correctly verify whether $o_i = \mathrm{tr}(O_i\rho)$ exceeds the threshold θ_i. Moreover, we only need to do this until the number of times the threshold is exceeded goes past the budget ℓ (though if this never happens then we must correctly provide answers for all M queries). We introduce a new threshold search algorithm based on the classical shadows formalism, which will be used as another main component of the algorithm in Appendix E.

As in the previous sections, we will apply a classical adaptive data analysis algorithm (the sparse vector algorithm [52,93]) to estimators on classical shadows, though with an additional preprocessing step. For a given classical shadow dataset $\mathcal{S}(\rho; N)$, the algorithm first truncates each $o_i(\hat{\rho})$ to the range $[-T, T]$, where $T = O(\log(1/\epsilon))$. Then, it defines a statistical query estimator of $\mathrm{tr}(O_i\rho)$ as the sample mean of these truncated values and runs the sparse vector algorithm [52] using this estimator. We do the truncation step because the sparse vector algorithm works only for statistical (or, more generally, low-sensitivity) queries. Explicitly, the steps of the algorithm are as follows:

(1) Gather classical shadows data $\mathcal{S}(\rho;N)=\{\hat{\rho}_1,\ldots,\hat{\rho}_N\}$. Let

$$s_i(\hat{\rho})=\mathrm{sign}(\mathrm{tr}(O_i\hat{\rho}))\cdot\min\{|\mathrm{tr}(O_i\hat{\rho})|,T\} \tag{D7}$$

denote the truncation of $o_i(\hat{\rho})$ to the interval $[-T,T]$.
(2) Run the sparse vector algorithm over the statistical queries

$$s_i(\mathcal{S})=\frac{1}{N}\sum_{j=1}^{N}s_i(\hat{\rho}_{(j)}). \tag{D8}$$

It is known that the sparse vector algorithm has the following guarantees, which can be seen by, e.g., applying the transfer theorem in Ref. [17] to Theorem 5.2 in Ref. [19].

Theorem D3 ([52,93]). Given parameters $\epsilon,\delta\in[0,1]$ and a budget $\ell\in\mathbb{N}$, there is an algorithm using

$$K=\mathcal{O}\left(\frac{T\sqrt{\ell}\log(M)\log(1/\epsilon\delta)^{3/2}}{\epsilon^2}\right) \tag{D9}$$

samples that respond to an adaptively chosen sequence of query/threshold pairs $(q_1,\theta_1),\ldots,(q_M,\theta_M)$, where $q_i:\mathcal{X}\mapsto[-T,T]$. With probability at least $1-\delta$, the algorithm correctly does the following for all $i\in[M]$:

(1) If $q_i(\mathcal{D})>\theta_i$, output "No."
(2) If $q_i(\mathcal{D})\le\theta_i-\epsilon$, output "Yes."
(3) If the number of "No"s exceed ℓ, terminate the algorithm.

Using this, we can obtain the following guarantees for our threshold search algorithm described above.

Theorem D4 (Threshold search). Given parameters $\epsilon,\delta\in[0,1]$ and a budget parameter $\ell\in\mathbb{N}$, there is an algorithm using

$$N=\mathcal{O}\left(\frac{\sqrt{B\ell}\log(B)\log(M)\log(1/\epsilon\delta)^{5/2}}{\epsilon^2}\right)\times\left[B=\max_{i=1,\ldots,M}\mathrm{tr}(O_i^2)\right] \tag{D10}$$

samples that responds to an adaptively chosen sequence of observable/threshold pairs $(O_1,\theta_1),\ldots,(O_M,\theta_M)$. With probability at least $1-\delta$, the algorithm correctly does the following for all $i\in[M]$:

(1) If $\mathrm{tr}(O_i\rho)>\theta_i$, always output "No."
(2) If $\mathrm{tr}(O_i\rho)\le\theta_i-\epsilon$, always output "Yes."
(3) If the number of "No"s exceed ℓ, terminate the algorithm.

Previously, the quantum threshold search algorithm [14] was used to solve this problem, with the best-known sample complexity as $\mathcal{O}(\ell\log^2 M/\epsilon^2)$. Our algorithm improves the sample complexity of quantum threshold search by a factor of $\sqrt{\ell}\log M$ in exchange for a term that scales with the bound on the Frobenius norm. In order to obtain this guarantee, we require that the classical shadows data $\mathcal{S}(\rho;N)$ gathered are obtained via the uniform POVM procedure described in Appendix A 2 c rather than the canonical classical shadow procedure [8]. This is due to its stronger concentration properties, which allow us to obtain a better sample complexity guarantee. If one were to use canonical classical shadows [8], a similar proof can obtain a sample complexity guarantee of $N=\tilde{\mathcal{O}}(B\sqrt{\ell}\log(M)/\epsilon^3)$. Importantly, using classical shadows with uniform POVM allows us to reduce the scaling with $1/\epsilon$ from $1/\epsilon^3$ to $1/\epsilon^2$.

Proof of Theorem D4. We first show that the bias of each $s(\hat{\rho}_{(j)})$ defined in Eq. (D7) is small (which implies the bias of $s_i(\mathcal{S})$ is small). We write $o=\mathrm{tr}(O\hat{\rho})$ and $s=s(\hat{\rho})$ for brevity. Letting $\mu=\mathbb{E}[o]\in[-1,1]$, we have that the bias $|\mathbb{E}[o]-\mathbb{E}[s]|$ satisfies the following:

$$\begin{aligned}
&|\mathbb{E}[o]-\mathbb{E}[s]| && \text{(D11)}\\
&=|\mathbb{E}[(o-T)\mathbb{I}(o>T)-(o+T)\mathbb{I}(o<-T)]| && \text{(D12)}\\
&\le|\mathbb{E}[(o-\mu-(T-1))\mathbb{I}(o-\mu>T-1)-(o-\mu+T-1)\mathbb{I}(o-\mu<-T+1)]| && \text{(D13)}\\
&\le|\mathbb{E}[(o-\mu)\mathbb{I}(o-\mu>T-1)]|+|\mathbb{E}[(o-\mu)\mathbb{I}(o-\mu<-T+1)]|+(T-1)\Pr[|o-\mu|>T-1] && \text{(D14)}\\
&\le\sqrt{\mathbb{E}(o-\mu)^2\Pr[o-\mu>T-1]}+\sqrt{\mathbb{E}(o-\mu^2)\Pr[o-\mu<-T+1]}+(T-1)\Pr[|o-\mu|>T-1] && \text{(D15)}\\
&\le\sqrt{8B}\cdot\sqrt{2\Pr[|o-\mu|>T-1]}+(T-1)\Pr[|o-\mu|>T-1] && \text{(D16)}\\
&\le 4\sqrt{B}\cdot\sqrt{2\exp\left(-\frac{(T-1)^2}{16B+4\sqrt{B}(T-1)}\right)}+(T-1)\cdot 2\exp\left(-\frac{(T-1)^2}{16B+4\sqrt{B}(T-1)}\right), && \text{(D17)}
\end{aligned}$$

where Eq. (D14) follows from triangle inequality, Eq. (D15) from the Cauchy-Schwarz inequality, Eq. (D16) from the Cauchy-Schwarz inequality and the second moment bound in Lemma A1, and Eq. (D17) from Proposition A1. Setting $T = 1 + \sqrt{CB}$, it can be verified that if $C \geq 80^2(\log B + 4)^2$, then

$$T - 1 \leq \exp\left(\frac{(T-1)^2}{32B + 8\sqrt{B}(T-1)}\right). \tag{D18}$$

Moreover, if $C \geq (40^2 \log(12/\epsilon))^2$, then

$$2\exp\left(\frac{(T-1)^2}{32B + 8\sqrt{B}(T-1)}\right) \leq \epsilon/6. \tag{D19}$$

Combining these two, we obtain that for $\epsilon < 1$, setting $T = 1 + 40\sqrt{B}\log(48/\epsilon)(\log B + 4)$ yields

$$(T-1)\cdot 2\exp\left(-\frac{(T-1)^2}{16B + 4\sqrt{B}(T-1)}\right) \leq \epsilon/6. \tag{D20}$$

A similar computation shows that if $C \geq 40^2 \log(48B/\epsilon)^2$,

$$4\sqrt{B}\cdot\sqrt{2\exp\left(-\frac{(T-1)^2}{16B + 4\sqrt{B}(T-1)}\right)} \leq \epsilon/6. \tag{D21}$$

Putting everything together, choosing $T = 1 + 40\sqrt{B}\log(48/\epsilon)(\log B + 4)$ yields

$$|\mathbb{E}[o] - \mathbb{E}[s]| \leq \epsilon/3. \tag{D22}$$

This tells us that the truncation of the expectation value to lie in the interval $[-T, T]$ as in Step 1 of the algorithm is still close to the true expectation value $\mathrm{tr}(O_i\rho)$.

In Step 2 of the algorithm, we run the sparse vector algorithm from Theorem D3 over the dataset $\tilde{\mathcal{S}}$ and queries s_i. Notice that Theorem D3 can be seen as the classical analog of quantum threshold search, where in quantum threshold search $q_i(\mathcal{D}) = \mathrm{tr}(O_i\rho)$. For our algorithm, we instead define $q_i(\mathcal{D})$ to be the truncated estimator $s_i : \mathcal{X} \mapsto [-T, T]$, where $\mathcal{X}$ is the space of classical shadows. Moreover, for a given query (O_i, θ_i), we instantiate the sparse vector algorithm with parameters $\epsilon/3, \delta$ and pass in query/threshold pairs $(s_i, \theta_i - \epsilon/3)$. Then, the guarantees of the sparse vector algorithm imply those of Theorem D4:

(1) Suppose $o_i > \theta_i$. Since $o_i = o_i - s_i + s_i \leq \epsilon/3 + s_i$, then $s_i > \theta_i - \epsilon/3 \implies$ output "No."
(2) Suppose $o_i \leq \theta_i - \epsilon$. Since $o_i \geq s_i - \epsilon/3$, then $s_i \leq \theta_i - 2\epsilon/3 \implies$ output "Yes."

The final sample complexity follows from Theorem D3. ■

a. Extension to verifying epsilon closeness

We include an extension of the threshold search algorithm to the problem of verifying whether θ_i is close to $\mathrm{tr}(O_i\rho)$. This variant will be used for our problem of predicting expectation values of observables in Appendix E.

Corollary D1. Consider the same inputs as Theorem D4. There is an algorithm using

$$N = \mathcal{O}\left(\frac{B\sqrt{\ell}\log(B)\log(M)\log(1/\epsilon\delta)^{5/2}}{\epsilon^2}\right) \times \left[B = \max_{i=1,\ldots,M}\mathrm{tr}(O_i^2)\right] \tag{D23}$$

samples and with probability at least $1 - \delta$ correctly implements the following for up to ℓ mistakes:

(1) If $|\mathrm{tr}(O_i\rho) - \theta_i| > \epsilon$, always declare a mistake.
(2) If $|\mathrm{tr}(O_i\rho) - \theta_i| \leq \frac{3}{4}\epsilon$, always pass.
(3) If a mistake was declared, supply a value μ_i' satisfying $|\mathrm{tr}(O_i\rho) - \mu_i'| \leq \epsilon/4$.

Proof. The proof for implementing the verification step (i.e., conditions 1 and 2) are the same as in Ref. [14]. For each query (O_i, θ_i), we can simply give "simulated inputs" $(O_i, \theta_i + \epsilon), (\mathbb{I} - O_i, 1 - \theta_i + \epsilon)$ to the algorithm from Theorem D4 with parameters $\epsilon/4$, $\delta/2$, $2M$, and ℓ. With probability at least $1 - \delta/2$, we get the correct outputs for the threshold search algorithm. We let the algorithm operate as follows.

(1) If threshold search declares a mistake on either one of the queries, i.e.,

$$\mathrm{tr}(O_i\rho) > \theta_i + \epsilon - \frac{1}{4}\epsilon \quad \text{or} \quad \mathrm{tr}(\mathbb{I} - O_i\rho) > 1 - \theta_i + \epsilon - \frac{1}{4}\epsilon, \tag{D24}$$

then the algorithm will correctly declare a mistake for query (O_i, θ_i).
(2) If threshold search passed on both queries, i.e.,

$$\mathrm{tr}(O_i\rho) \leq \theta_i + \epsilon \quad \text{or} \quad \mathrm{tr}(\mathbb{I} - O_i\rho) \leq 1 - \theta_i + \epsilon, \tag{D25}$$

then the algorithm will correctly output "pass."

It can be verified that these give the correct outputs in conditions 1 and 2. The sample complexity for implementing this is given in Theorem D4.

In order to provide the correction μ_i' when mistakes are made, we can simply run the Theorem D1 algorithm with

accuracy parameters $\epsilon/4, \delta$ in parallel whenever a mistake is made (can be run on a separate batch or on all the samples). This requires

$$N = \mathcal{O}\left(\frac{B\sqrt{\ell \log(1/\delta)}\log(\ell/\delta\epsilon)}{\epsilon^2}\right) \quad [B = \max_{i=1,\ldots,M} \mathrm{tr}(O_i^2)] \tag{D26}$$

samples. Combining this with the sample complexity in Theorem D4 yields the final result. ■

APPENDIX E: BOUNDED-FROBENIUS-NORM OBSERVABLES

We consider predicting expectation values of M adaptively chosen observables O_i with bounded-Frobenius-norm, i.e., $\max_i \mathrm{tr}(O_i^2) \leq B$. We also consider the subcase of predicting low-rank observables. For this case, we show that we can design an algorithm for the mechanism that achieves a sample complexity upper bound that scales as $\mathcal{O}(\log(M))$. To help simplify the presentation, we first prove a sample-efficient algorithm for predicting adaptively chosen single-rank observables. Then we extend the ideas to answering adaptively chosen bounded-Frobenius-norm observables and low-rank observables.

1. Single rank observables

We can obtain the following rigorous sample complexity guarantee for predicting adaptively chosen single-rank observables.

Theorem E1 (Single rank observables). There exists an algorithm that can answer M adaptively chosen single rank projectors $\{|\psi^i\rangle\langle\psi^i|\}$ to ϵ error with probability at least $1-\delta$ using

$$N = \mathcal{O}\left(\frac{\log(M)\log(1/\epsilon\delta)^{3/2}\log(1/\epsilon)}{\epsilon^3}\right) \tag{E1}$$

samples of ρ.

In particular, we designed the algorithm in Algorithm 7 for achieving this sample complexity upper bound. In Algorithm 7, the tomography protocol $\mathcal{P}$ can be any protocol that implements shadow tomography, for example, the original shadow tomography protocol [7] or the algorithms from Appendix D 1. Similarly, the TS protocol can be any protocol which solves the threshold search problem (reviewed in Appendix D 2), e.g., the quantum threshold search algorithm [14] or the algorithms from Appendix D 2. Depending on which algorithm is used to implement the tomography protocol or threshold search, we can obtain different sample complexity guarantees. By using the new algorithms in Appendix D, we obtain the sample complexity in Theorem E1.

Our algorithm follows a similar framework to that in Ref. [14], which is based on Aaronson's shadow tomography protocol [7]. Specifically, we implement a student-teacher type interaction: the student responds to a queried observable O by constructing a guess $\hat{o}_i$ for its expectation value $o_i = \mathrm{tr}(O_i\rho)$ and then passing the guess to the teacher. The teacher either "passes," indicating that the guess is close to the true expectation value, or he declares a "mistake" and supplies a different value o_i' that is closer to the true expectation value.

In Algorithm 7, the teacher is implemented using a threshold search algorithm. Then, we can obtain the following guarantees for the student algorithm.

Theorem E2. Suppose the following teacher properties are satisfied for all $i \in [M]$:

(a) If $|\mathrm{tr}(O_i\rho) - \hat{o}_i| > \epsilon$, the teacher declares a mistake.
(b) If $|\mathrm{tr}(O_i\rho) - \hat{o}_i| \leq \frac{3}{4}\epsilon$, the teacher always passes.
(c) If the teacher declares a mistake, the supplied value o_i' satisfies $|\mathrm{tr}(O_i\rho) - o_i'| \leq \frac{1}{4}\epsilon$.

There is an algorithm for the student using $\tilde{\mathcal{O}}(\log M/\epsilon^3)$ copies of ρ that provides guesses $\hat{o}_i$ for $\mathrm{tr}(O_i\rho)$ while

ALGORITHM 7. Mechanism for predicting single-rank observables.

Set $S \leftarrow \{\}$ and initialize the tomography protocol $\mathcal{P}$.
Start an instance of threshold search (TS) with parameters ϵ, δ.
for $i = 1$ *to* M **do**
  Receive rank one observable $O = |\psi^i\rangle\langle\psi^i| = U_i|0^n\rangle$ and construct estimate $\hat{o}_i \approx \langle\psi_S^i|\rho|\psi_S^i\rangle$ using $\mathcal{P}$, where $\psi_S^i = \mathrm{Proj}_S(\psi^i)$.
  Feed $\hat{o}_i$ to TS.
  if *TS declares a mistake* **then**
    Set $S \leftarrow S \cup \{|\psi^i\rangle\}$.
    Receive new answer o_i' from TS and return o_i'.
    Update $\mathcal{P}$ and restart TS.
  else
    Return $\hat{o}_i$.
  end
end

making at most $\mathcal{O}(1/\epsilon^2)$ mistakes, with probability at least $1-\delta$.

The algorithm for the student is essentially to project $|\psi\rangle$ to the space spanned by S and guess the expectation value for the projected state. For a given observable $O = |\psi\rangle\langle\psi|$, let $|\psi\rangle = |\psi_S\rangle + |\psi_\perp\rangle$, where $|\psi_S\rangle$ is the projection of $|\psi\rangle$ onto the space spanned by S and $|\psi_\perp\rangle$ is the component orthogonal to the space spanned by S. Given $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ is spanned by an orthonormal basis $\phi_1, \ldots, \phi_k$ at the current round, the student will estimate

$$s = \langle\psi_S|\,\rho\,|\psi_S\rangle = \sum_{\substack{i\in[k]\\ j\in[k]}} \alpha_i\bar{\alpha}_j \left\langle\phi_j\right|\rho\left|\phi_i\right\rangle \tag{E2}$$

where $|\psi_S\rangle = \sum_{i=1}^k \alpha_i |\phi_i\rangle$. We show that if the student can estimate s_i for all observables $O_1, \ldots, O_M$ to a suitable precision, then he will make at most $\mathcal{O}(1/\epsilon^2)$ mistakes.

Lemma E1. Suppose ρ is a density matrix, i.e., $\mathrm{tr}(\rho) = 1$ and $\rho \succeq 0$. If $|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_k\rangle$ are orthonormal states and $\langle\phi_i|\,\rho\,|\phi_i\rangle > \epsilon\ \forall i$, then $k \le \frac{1}{\epsilon}$.

Proof. For any orthonormal basis $\{|\phi_1\rangle, \ldots, |\phi_{2^n}\rangle\}$, we have $\sum_{i=1}^{2^n} \mathrm{tr}(|\phi_i\rangle\langle\phi_i|\,\rho) = \mathrm{tr}(\rho) = 1$. The result follows immediately. ■

Lemma E2. Suppose the student has access to an oracle that can always produce an estimate $\hat{o}$ such that $|\langle\psi_S|\,\rho\,|\psi_S\rangle - \hat{o}| < \frac{3}{8}\epsilon$. Then, the oracle makes at most $\frac{256}{9}\frac{1}{\epsilon^2}$ mistakes.

Proof. Suppose the teacher declares a mistake on observable $|\psi\rangle\langle\psi|$, where $|\psi\rangle$ has unit norm, i.e., $|\langle\psi|\,\rho\,|\psi\rangle - \hat{o}| \ge \frac{3}{4}\epsilon$. By the oracle guarantees,

$$\begin{aligned}|\langle\psi|\,\rho\,|\psi\rangle - \hat{o}| &\le |\langle\psi|\,\rho\,|\psi\rangle - \langle\psi_S|\,\rho\,|\psi_S\rangle| \\ &\quad + |\langle\psi_S|\,\rho\,|\psi_S\rangle - \hat{o}| \le |\langle\psi|\,\rho\,|\psi\rangle \\ &\quad - \langle\psi_S|\,\rho\,|\psi_S\rangle| + \frac{3}{8}\epsilon. \end{aligned}\tag{E3}$$

Thus, $|\langle\psi|\,\rho\,|\psi\rangle - \langle\psi_S|\,\rho\,|\psi_S\rangle| \ge \frac{3}{8}\epsilon$. Let $\{|\phi_i\rangle\}_{i\in[k]}$ span S and $|\phi_\perp\rangle$ be the normalized state corresponding to $|\psi_\perp\rangle$, so we can write the decompositions $|\psi_S\rangle = \sum_{i=1}^k \alpha_i |\phi_i\rangle$ and $|\psi_\perp\rangle = \alpha_\perp |\phi_\perp\rangle$. We then have

$$\begin{aligned}&|\langle\psi|\,\rho\,|\psi\rangle - \langle\psi_S|\,\rho\,|\psi_S\rangle| \\ &\quad= |\langle\psi_\perp|\,\rho\,|\psi_\perp\rangle + \langle\psi_S|\,\rho\,|\psi_\perp\rangle + \langle\psi_\perp|\,\rho\,|\psi_S\rangle| &\text{(E4)}\\ &\quad= |\alpha_\perp\bar{\alpha}_\perp\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle + \sum_{i=1}^k[\alpha_\perp\bar{\alpha}_i\langle\phi_i|\,\rho\,|\phi_\perp\rangle \\ &\qquad+ \alpha_i\bar{\alpha}_\perp\langle\phi_\perp|\,\rho\,|\phi_i\rangle]| &\text{(E5)}\\ &\quad\le |\alpha_\perp|^2\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle + 2\sum_i |\alpha_\perp||\alpha_i||\langle\phi_\perp|\,\rho\,|\phi_i\rangle|, &\text{(E6)}\end{aligned}$$

where the first and second inequality follows from triangle inequality, respectively. We can bound $|\langle\phi_\perp|\,\rho\,|\phi_i\rangle|$ as follows:

$$\begin{aligned}|\langle\phi_\perp|\,\rho\,|\phi_i\rangle| &= |\langle\phi_\perp|\,(\rho^{\frac{1}{2}})^\dagger\rho^{\frac{1}{2}}\,|\phi_i\rangle| &\text{(E7)}\\ &= |\left\langle\rho^{\frac{1}{2}}\phi_\perp\right|\left.\rho^{\frac{1}{2}}\phi_i\right\rangle| &\text{(E8)}\\ &\le \sqrt{\left\langle\rho^{\frac{1}{2}}\phi_\perp\right|\left.\rho^{\frac{1}{2}}\phi_\perp\right\rangle\left\langle\rho^{\frac{1}{2}}\phi_i\right|\left.\rho^{\frac{1}{2}}\phi_i\right\rangle} &\text{(E9)}\\ &= \sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle\langle\phi_i|\,\rho\,|\phi_i\rangle}, &\text{(E10)}\end{aligned}$$

where the inequality follows by Cauchy-Schwarz inequality. Thus, plugging into the previous equation, we have

$$\begin{aligned}|\langle\psi|\,\rho\,|\psi\rangle - \langle\psi_S|\,\rho\,|\psi_S\rangle| &\le |\alpha_\perp|^2\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle + 2\sum_i |\alpha_\perp||\alpha_i|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle\langle\phi_i|\,\rho\,|\phi_i\rangle} \\ &= |\alpha_\perp|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle}(|\alpha_\perp|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle} + 2\sum_i |\alpha_i|\sqrt{\langle\phi_i|\,\rho\,|\phi_i\rangle}) \\ &\le |\alpha_\perp|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle}\sqrt{(|\alpha_\perp|^2 + 2\sum_i |\alpha_i|^2)(\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle + 2\sum_i \langle\phi_i|\,\rho\,|\phi_i\rangle)} \\ &\le 2|\alpha_\perp|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle},\end{aligned}$$

where the second to last inequality follows by Cauchy-Schwarz inequality, and the last inequality follows since $\mathrm{tr}(\rho) = 1$ and $\sum_{i=1}^{k} |\alpha_i|^2 + |\alpha_\perp|^2 = 1$ (i.e., $|\psi|$ has unit norm). Since $2|\alpha_\perp|\sqrt{\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle} > \frac{3}{8}\epsilon$, then $\langle\phi_\perp|\,\rho\,|\phi_\perp\rangle > (9\epsilon^2/256|\alpha_\perp|^2)$. This implies that every time we make a mistake, we add a new component $|\phi_i\rangle$ orthonormal to the current S with expectation value at least $(9\epsilon^2/256|\alpha_\perp|^2)$. By Lemma E1 and the fact that $|\alpha_\perp| \leq 1$, this can happen at most $\frac{256}{9}(1/\epsilon^2)$ times. ■

a. Estimating $\langle\psi_S|\,\rho\,|\psi_S\rangle$

We now incorporate the tomography protocol from Theorem D2 and show that it can be used to predict $\langle\psi_S|\,\rho\,|\psi_S\rangle$ efficiently.

Lemma E3.

$$N_{ST} = \mathcal{O}\left(\frac{\sqrt{\log(1/\delta)}\cdot\log(M/\epsilon\delta)\log(1/\epsilon)}{\epsilon^3}\right) \tag{E11}$$

samples suffice to estimate $s_i = \mathrm{tr}(|\psi_S^i\rangle\langle\psi_S^i|\,\rho)$ within ϵ error for all $1 \leq i \leq M$ with probability at least $1-\delta$.

Proof. Suppose the size of the set S is at most k. The problem from the tomography protocol's point of view can be framed as follows: at each round, receive a low-dimensional query $|\psi_S\rangle$, which is $|\psi\rangle$ projected onto the low-dimensional space spanned by S. The goal is to estimate $\langle\psi_S|\,\rho\,|\psi_S\rangle$. We show that the full tomography problem involving the $2^n \times 2^n$ quantum state ρ is equivalent to a separate tomography problem in this low-dimensional space. Specifically, we claim that there is a quantum state $\hat{\rho}$ of dimension $2^{\lceil\log(k+1)\rceil}$ such that $\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle = \langle\psi_S|\,\rho\,|\psi_S\rangle$. If this is the case, then as long as we express $|\psi_S\rangle$ in a basis of S, then we can simply run our tomography algorithm on the separate problem with the low-dimensional representation of $|\psi_S\rangle$ as input.

Recall that the set S is constructed iteratively in Algorithm 7, where each time the teacher declares a mistake, a state is added to S. Let S_i be S after the ith mistake, and suppose that, in hindsight, there exists a fixed orthonormal basis $\Phi_i = \{|\phi_1\rangle, \ldots, |\phi_i\rangle\} = \Phi_{i-1} \cup \{|\phi_i\rangle\}$ of S_i for all $1 \leq i \leq k$. Such a basis can clearly be constructed via Gram-Schmidt orthonormalization.

We can use this basis to show the claim that there exists some $\hat{\rho}$ such that $\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle = \langle\psi_S|\,\rho\,|\psi_S\rangle$. To see that the claim is true, consider the $k \times k$ submatrix representing ρ in the basis Φ_k, i.e., $\hat{\rho}_S = \{\langle\phi_i|\,\rho\,|\phi_j\rangle\}_{1\leq i,j\leq k}$. Note that $\langle\psi_S|\,\hat{\rho}_S\,|\psi_S\rangle = \langle\psi_S|\,\rho\,|\psi_S\rangle$ since $|\psi_S\rangle \in \mathrm{span}(S)$. Finally, we only need to "pad" $\hat{\rho}_S$ with additional dimensions to make it a valid quantum state (i.e., unit trace, positive semi-definite, and Hermitian). Specifically, we can obtain $\hat{\rho}$ of dimension $2^{\lceil\log(k+1)\rceil}$ by adding additional dimensions where the additional diagonal elements are non-negative to make $\hat{\rho}$ trace 1. The off-diagonal elements can be set to 0, which would make $\hat{\rho}$ positive semi-definite and Hermitian since $\hat{\rho}_S$ is positive semi-definite and Hermitian. Moreover, $\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle = \langle\psi_S|\,\hat{\rho}_S\,|\psi_S\rangle$.

Thus, we can now consider the problem of predicting $\langle\psi_S|\,\hat{\rho}\,|\psi_S\rangle$, with $\hat{\rho}$ being a $2^{\lceil\log(k+1)\rceil}$ dimensional quantum state. This can be solved using the adaptive tomography algorithm from Theorem D2. In this case, the number of bits needed to represent a classical shadow obtained with Clifford measurements for this hypothetical state is $m = \mathcal{O}(\lceil\log(k+1)\rceil^2)$. We can outline the algorithm as follows:

(1) Obtain a classical shadow dataset of size

$$N_{ST} = \mathcal{O}\left(\frac{\sqrt{\log(1/\delta)}\cdot\log(M/\epsilon\delta)\log(1/\epsilon)}{\epsilon^3}\right) \tag{E12}$$

using Clifford measurements.

(2) Prepare an instance of the tomography algorithm from Theorem D2 initialized with dimension $m = \mathcal{O}(\lceil\log(k+1)\rceil^2)$. This also requires N_{ST} samples.

(3) At any round i, express $|\psi_{S_i}\rangle$ in i dimensions and feed it as input to the tomography algorithm. This ensures that in hindsight every query $|\psi_{S_i}\rangle$ was expressed in the subspace $\mathrm{span}(S_k)$ so that we have a proper reduction to the low-dimensional problem specified previously. Then return the output of the tomography algorithm.

By the mistake bound of Lemma E2, $k = \mathcal{O}(1/\epsilon^2)$. Combining this with Theorem D2 yields the result of Lemma E3. ■

With this, we can provide a simple proof of Theorem E2

Proof of Theorem E2. Combining Lemma E3 with Lemma E2 yields Theorem E2. ■

Finally, we can prove our guarantee for single-rank observables by combining all of our results.

Proof. The teacher can be implemented directly by our threshold search algorithm in Theorem D4. Combined with the mistake bound in Theorem E2 yields a sample complexity of

$$N_{TS} = \mathcal{O}\left(\frac{\log(M)\log(1/\epsilon\delta)^{5/2}}{\epsilon^3}\right) \tag{E13}$$

for implementing threshold search. Summing this with the sample complexity for implementing the tomography

algorithm

$$N_{ST} = \mathcal{O}\left(\frac{\sqrt{\log(1/\delta)} \cdot \log(M/\epsilon\delta)\log(1/\epsilon)}{\epsilon^3}\right) \quad \text{(E14)}$$

yields the desired guarantee,

$$N = N_{ST} + N_{TS} = \mathcal{O}\left(\frac{\log(M)\log(1/\epsilon\delta)^{5/2}\log(1/\epsilon)}{\epsilon^3}\right). \quad \text{(E15)}$$

■

2. Extension to bounded-Frobenius-norm observables

Having solved the single rank observable case, we now describe how to extend the algorithm to predicting bounded-Frobenius-norm observables. We obtain the following guarantees.

Theorem E3. For a sequence of adaptively chosen observables $O_1, \ldots, O_M$ satisfying $\mathrm{tr}(O_i^2) \le B$ for all i and accuracy parameters $\epsilon, \delta \in [0, 1]$, there exists an algorithm that uses

$$N = \mathcal{O}\left(\frac{B^2 \log(M)\log(1/\epsilon\delta)^{5/2}\log(B/\epsilon)}{\epsilon^4}\right) \quad \text{(E16)}$$

samples of ρ to accurately predict every property,

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \le \epsilon \quad \text{for all } 1 \le i \le M, \quad \text{(E17)}$$

with probability at least $1 - \delta$.

Thus, we are able to predict expectation values of adaptively chosen bounded-Frobenius-norm observables with the desired scaling of $\log(M)$, independent of system size. When compared to protocols for adaptive shadow tomography [14], we remove the system size dependence in exchange for an extra factor of $1/\epsilon$.

To achieve this sample complexity upper bound, we modify the algorithm from Appendix E 1. The only modifications are (1) we define a processing step that "projects" the queried observable O to S and (2) when a mistake is made, every eigenstate of O with a sufficiently large (in magnitude) eigenvalue is added to S. In more detail, the algorithm is given in Algorithm 8 and we outline it as follows.

For a given observable O, consider its eigenvalue decomposition $O = \sum_{i=1}^{2^n} w_i |\psi^i\rangle\langle\psi^i|$, where $\sum_{i=1}^{2^n} |w_i|^2 \le B$. Let $\hat{O} = \sum_{i:|w_i|>\epsilon/2} |w_i| |\psi^i\rangle\langle\psi^i|$ be its truncated eigenvalue decomposition, i.e., we include only eigenstates whose eigenvalues have magnitude greater than $\epsilon/2$ and throw out the rest. Note that there are at most $4B/\epsilon^2$ such eigenstates. At every round, the algorithm projects every eigenstate of the truncated observable $\hat{O}$ to span(S) and constructs $\tilde{O} = P_S \hat{O} P_S = \sum_{i:|w_i|>\epsilon/2} w_i |\psi_S^i\rangle\langle\psi_S^i|$, where P_S is the projection operator to S and ψ_S^i is the projection of eigenstate ψ^i to S. Then it uses the tomography protocol from Theorem D2 with $m = 4B/\epsilon^6$, accuracy parameter $\epsilon/2$, and failure probability $\delta/2$ to generate prediction $\tilde{o} \approx \mathrm{tr}(\tilde{O}\rho)$. If a mistake is made, every eigenstate in $\hat{O}$ is added to S. The teacher is implemented by the threshold search algorithm in Theorem D4 with accuracy parameter ϵ and failure probability $\delta/2$.

With this sketch of the algorithm, we can prove Theorem E3, relying on some results from the previous section.

Proof of Theorem E3. The probability that either threshold search or the tomography protocol fails is at most δ by union bound. Thus, for the rest of the proof, we assume that both protocols correctly give outputs across all M rounds. Hence, every $\hat{o}$ received from the shadow tomography protocol satisfies $|\hat{o} - \mathrm{tr}(\tilde{O}\rho)| \le \frac{3}{16}\epsilon$. Now suppose QTS declares a mistake, so we have $|\hat{o} - \mathrm{tr}(\hat{O}\rho)| > 3\epsilon/8$. Using triangle inequality,

$$\frac{3}{8}\epsilon < |\hat{o} - \mathrm{tr}(\hat{O}\rho)| \le |\hat{o} - \mathrm{tr}(\tilde{O}\rho)| + |\mathrm{tr}(\tilde{O}\rho) - \mathrm{tr}(\hat{O}\rho)| \le \frac{3}{16}\epsilon + |\mathrm{tr}(\tilde{O}\rho) - \mathrm{tr}(\hat{O}\rho)|. \quad \text{(E18)}$$

Thus, $|\mathrm{tr}(\tilde{O}\rho) - \mathrm{tr}(\hat{O}\rho)| > 3\epsilon/16$. We claim that there is at least one eigenstate $|\psi\rangle$ in the truncated observable $\hat{O}$ such that

$$|\langle\psi_S|\rho|\psi_S\rangle - \langle\psi|\rho|\psi\rangle| > \frac{3\epsilon^2}{32B}. \quad \text{(E19)}$$

We can see this by noting that if this is not true for any eigenstate of $\hat{O}$, then this gives us a contradiction,

$$\begin{aligned}
&\left|\mathrm{tr}(\tilde{O}\rho) - \mathrm{tr}(\hat{O}\rho)\right| \\
&= \left|\sum_{i:|w_i|>\epsilon/2} w_i(\langle\psi_S^i|\rho|\psi_S^i\rangle - \langle\psi^i|\rho|\psi^i\rangle)\right| && \text{(E20)}\\
&\le \sum_{i:|w_i|>\epsilon/2} |w_i| \left|\langle\psi_S^i|\rho|\psi_S^i\rangle - \langle\psi^i|\rho|\psi^i\rangle\right| && \text{(E21)}\\
&\le \frac{3\epsilon^2}{32B} \sum_{i:|w_i|>\epsilon/2} |w_i| && \text{(E22)}\\
&\le \frac{3}{16}\epsilon, && \text{(E23)}
\end{aligned}$$

where the first inequality is by triangle inequality, the second is by our assumption, and the third by Cauchy-Schwarz inequality and the Frobenius-norm bound.

ALGORITHM 8. Mechanism for bounded-Frobenius-norm observables.

Set $S \leftarrow \{\}$ and initialize tomography protocol $\mathcal{P}$ with accuracy parameter $\frac{3}{16}\epsilon$ and failure probability $\delta/2$.
Start an instance of threshold search (TS) with parameters $\epsilon/2, \delta/2$.
for $i = 1$ *to* M **do**
  Receive truncated observable $\hat{O}_i = \sum_{j:|w_j|>\epsilon/2} w_j \left|\psi^j\right\rangle\!\left\langle\psi^j\right|$.
  Project every eigenstate of $\hat{O}_i$ to span(S) and construct $\tilde{O}_i = \sum_{j:|w_j|>\epsilon/2} w_j \left|\psi_S^j\right\rangle\!\left\langle\psi_S^j\right|$, where $\psi_S^j = \mathrm{Proj}_S(\psi^j)$ is the projection of the eigenstate ψ^j to S.
  Receive prediction $\hat{o}_i \approx \mathrm{tr}(\tilde{O}_i\rho)$ from $\mathcal{P}$. Feed $\hat{o}_i$ to TS.
  if *TS declares a mistake* **then**
    Set $S \leftarrow S \cup \{|\psi^j\rangle : |w_j| > \epsilon/2\}$.
    Receive new answer o_i' from TS and return $\hat{o}_i = o_i'$.
    Update $\mathcal{P}$ and restart TS.
  else
    Return $\hat{o}_i$.
  end
end

Since a mistake was declared, then every eigenstate in $\hat{O}$ gets added to S. We added at least one eigenstate such that $|\langle\psi_S|\rho|\psi_S\rangle - \langle\psi|\rho|\psi\rangle| > 3\epsilon^2/32B$, so by the mistake bound in Lemma E2, a mistake can happen at most $\mathcal{O}(B^2/\epsilon^4)$ times. To obtain the m parameter in the tomography protocol from Theorem D2, note that every time a mistake is made we add at most $4B/\epsilon^2$ states to S. Combining this with the mistake bound, $|S| \le \mathcal{O}(B^3/\epsilon^6)$, so we set $m = \lceil \log(\mathcal{O}(B^3/\epsilon^6) + 1)\rceil^2 = \mathcal{O}(\log(B/\epsilon)^2)$. Moreover, we emphasize that the Frobenius bound still holds for each $\tilde{O} = P_S\hat{O}P_S$: using the fact that $P_S^2 = P_S$ and $I - P_S$ is a projector,

$$\begin{aligned}\mathrm{tr}(A^2) - \mathrm{tr}((P_SA)^2) &= \mathrm{tr}(A^2) - \mathrm{tr}(AP_SA)\\ &= \mathrm{tr}(A(I-P_S)A) = \mathrm{tr}(((I-P_S)A)^2) \ge 0. \end{aligned} \tag{E24}$$

Thus, $\|P_SA\|_F \le \|A\|_F$. It can similarly be shown that $\|AP_S\|_F \le \|A\|_F$. In sum,

$$\|P_S\hat{O}P_S\|_F \le \|\hat{O}\|_F \le \|O\|_F \le B. \tag{E25}$$

Putting these together, the sample complexity of the tomography protocol predicting the projected observables $\tilde{O}_1, \ldots, \tilde{O}_M$ is

$$N_{ST} = \mathcal{O}\left(\frac{B^{3/2}\sqrt{\log(1/\delta)}\cdot\log(M/\epsilon\delta)\log(B/\epsilon)}{\epsilon^3}\right), \tag{E26}$$

by Theorem D2. The sample complexity of the threshold search algorithm verifying the output of the tomography protocol is

$$N_{TS} = \mathcal{O}\left(\frac{B^2\log(M)\log(1/\epsilon\delta)^{5/2}}{\epsilon^4}\right), \tag{E27}$$

by Theorem D4. Combining these two bounds yields the final result, which guarantees that for every $\hat{o}_i$ returned by the mechanism, $|\hat{o}_i - \mathrm{tr}(\hat{O}_i\rho)| \le \epsilon/2$ for all $1 \le i \le M$ with probability at least $1-\delta$. Finally, this implies that the total error is bounded by ϵ,

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \le |\hat{o}_i - \mathrm{tr}(\hat{O}_i\rho)| + |\mathrm{tr}(\hat{O}_i\rho) - \mathrm{tr}(O_i\rho)| \tag{E28}$$

$$\le \epsilon/2 + \sum_{j:|w_j|<\epsilon/2} |w_j|\left\langle\psi^j\right|\rho\left|\psi^j\right\rangle \tag{E29}$$

$$\le \frac{\epsilon}{2} + \frac{\epsilon}{2}\sum_{j:|w_j|<\epsilon/2}\left\langle\psi^j\right|\rho\left|\psi^j\right\rangle \tag{E30}$$

$$\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \tag{E31}$$

$$= \epsilon. \tag{E32}$$

Thus, the total sample complexity is given by

$$N = N_{ST} + N_{TS} = \mathcal{O}\left(\frac{B^2\log(M)\log(1/\epsilon\delta)^{5/2}\log(B/\epsilon)}{\epsilon^4}\right). \tag{E33}$$

■

3. Extension to low-rank observables

We also consider the subcase of predicting expectation values of low-rank observables. In this case, we can again achieve a sample complexity upper bound scaling logarithmically in the number of observables M, independent of system size. Moreover, we obtain a strict improvement over the best-known sample complexity for shadow tomography, given in Ref. [14]. This can be done by directly applying Algorithm 8, with the slight modification that it is not necessary to do the truncation step in this case.

Theorem E4. For a sequence of adaptively chosen observables $O_1, \ldots, O_M$ with rank at most R and accuracy parameters $\epsilon, \delta \in [0, 1]$, there exists an algorithm that uses

$$N = \mathcal{O}\left(\frac{R^2 \log(M) \log(1/\epsilon\delta)^{5/2} \log(R/\epsilon)}{\epsilon^3}\right) \tag{E34}$$

samples of ρ to accurately predict every property,

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \text{for all } 1 \leq i \leq M, \tag{E35}$$

with probability at least $1 - \delta$.

Proof. The analysis is the exact same as that for Theorem E3, except that in Eq. (E19), the guarantee is instead there is at least one eigenstate $|\psi\rangle$ in the observable $\hat{O}$ such that

$$|\langle\psi_S|\rho|\psi_S\rangle - \langle\psi|\rho|\psi\rangle| > \frac{3\epsilon}{16R}, \tag{E36}$$

which follows because $\sum_i |w_i| \leq R$. In this case, the new mistake bound is $\mathcal{O}(R^2/\epsilon^2)$. The new sample complexity of the threshold search algorithm is

$$N_{TS} = \mathcal{O}\left(\frac{R^2 \log(M) \log(1/\epsilon\delta)^{5/2}}{\epsilon^3}\right), \tag{E37}$$

by Theorem D2. Moreover, whenever a mistake is made, we add at most R eigenstates to S, so $|S| \leq \mathcal{O}(R^3/\epsilon^2)$. Moreover, the rank bound is preserved for $\tilde{O}$, i.e., $\mathrm{rank}(\tilde{O}) = \mathrm{rank}(P_S\hat{O}P_S) \leq R$, due to the following property: for any matrices (with dimensions properly aligned) A and B, $\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$. So the sample complexity of the tomography protocol is

$$N_{ST} = \mathcal{O}\left(\frac{R^{3/2}\sqrt{\log(1/\delta)} \log(M/\epsilon\delta) \log(R/\epsilon)}{\epsilon^3}\right), \tag{E38}$$

by Theorem D4. The result follows from combining these bounds:

$$N = N_{ST} + N_{TS} = \mathcal{O}\left(\frac{R^2 \log(M) \log(1/\epsilon\delta)^{5/2} \log(R/\epsilon)}{\epsilon^3}\right). \tag{E39}$$

■

Remark E1. In order to implement the threshold search and tomography components, one could have also used existing algorithms used in shadow tomography, e.g., the quantum threshold search algorithm [14] and Aaronson's shadow tomography algorithm [7,14]. By tracing through the same arguments, we would have also achieved a system-size-independent upper bound. However, we found that the scaling was worse, where the sample complexity was $\tilde{\mathcal{O}}(B^2 \log^2(M)/\epsilon^6)$ for bounded-Frobenius-norm observables and $\tilde{\mathcal{O}}(R^2 \log^2(M)/\epsilon^4)$ for low-rank observables.

[1] Konrad Banaszek, Marcus Cramer, and David Gross, Focus on quantum tomography, New J. Phys. **15**, 125020 (2013).

[2] Robin Blume-Kohout, Optimal, reliable estimation of quantum states, New J. Phys. **12**, 043034 (2010).

[3] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert, Quantum state tomography via compressed sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[4] Zdenek Hradil., Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).

[5] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu, Sample-optimal tomography of quantum states, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[6] Ryan O'Donnell and John Wright, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2016), p. 899.

[7] Scott Aaronson, in *STOC* (Association for Computing Machinery, New York, NY, USA, 2018), p. 325.

[8] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[9] Marco Paini and Amir Kalev, An approximate description of quantum states, arXiv:1910.10543.

[10] Jordan Cotler, Soonwon Choi, Alexander Lukin, Hrant Gharibyan, Tarun Grover, M. Eric Tai, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, Adam M. Kaufman *et al.*, Quantum virtual cooling, Phys. Rev. X **9**, 031013 (2019).

[11] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

[12] The $\mathcal{O}(\log M)$ sample complexity applies when the observables have a shadow norm independent of n, which is true for many classes of physically relevant observables.

[13] Scott Aaronson and Guy N. Rothblum, in *STOC* (Association for Computing Machinery, New York, NY, USA, 2019), p. 322.

[14] Costin Bădescu and Ryan O'Donnell, Improved quantum data analysis, in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (2021), pp. 1398–1411.

[15] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth, in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2015), p. 117.

[16] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth, Generalization in adaptive data analysis and holdout reuse, Adv. Neural Inf. Process. Syst. (2015).

[17] Raef Bassily, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman, Algorithmic stability for adaptive data analysis, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016), pp. 1046–1059.

[18] Daniel Russo and James Zou, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, edited by Arthur Gretton and Christian C. Robert (PMLR, Cadiz, Spain, 2016), Vol. 51, p. 1232.

[19] Vitaly Feldman and Thomas Steinke, Generalization for adaptively chosen estimators via stable median, in *Conference on learning theory* (PMLR, 2017), pp. 728–757.

[20] Vitaly Feldman and Thomas Steinke, Calibrating noise to variance in adaptive data analysis, in *Conference On Learning Theory* (PMLR, 2018), pp. 535–544.

[21] Christopher Jung, Katrina Ligett, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Moshe Shenfeld, A new analysis of differential privacy's generalization guarantees, in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (2021), pp. 9-9.

[22] Arun Ganesh and Jiazheng Zhao, Privately answering counting queries with generalized Gaussian mechanisms, in *2nd Symposium on Foundations of Responsible Computing* (LIPIcs, 2021), Vol. 192, pp. 1:1–1:18.

[23] Yuval Dagan and Gil Kur, A bounded-noise mechanism for differential privacy, in *Proceedings of the 35th Conference on Learning Theory* (PMLR, 2022), Vol. 178, pp. 625–661.

[24] Badih Ghazi, Ravi Kumar, and Pasin Manurangsi, On avoiding the union bound when answering multiple differentially private queries, in *Proceedings of the 34th Conference on Learning Theory* (PMLR, 2021), Vol. 134, pp. 2133–2146.

[25] Guy Blanc, Subsampling suffices for adaptive data analysis, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (2023), pp. 999–1012.

[26] John P. A. Ioannidis, Contradicted and initially stronger effects in highly cited clinical research, Jama **294**, 218 (2005).

[27] John P. A. Ioannidis, Why most published research findings are false, PLoS Med. **2**, e124 (2005).

[28] Florian Prinz, Thomas Schlange, and Khusru Asadullah, Believe it or not: How much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov. **10**, 712 (2011).

[29] C. Glenn Begley and Lee M. Ellis, Raise standards for preclinical cancer research, Nature **483**, 531 (2012).

[30] Andrew Gelman and Eric Loken, in *The Best Writing on Mathematics*, edited by M. Pitici (Sigma Xi, The Scientific Research Society, Research Triangle Park, NC, USA, 2016), p. 305.

[31] Moritz Hardt and Guy N. Rothblum, in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science* (IEEE, New York, NY, USA, 2010), p. 61.

[32] Thomas Steinke and Jonathan Ullman, Interactive fingerprinting codes and the hardness of preventing false discovery, in *Proceedings of the 28th Conference on Learning Theory* (PMLR, 2015), Vol. 40, pp. 1588–1628.

[33] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[34] Ophelia Crawford, Barnaby van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[35] William J. Huggins, Jarrod McClean, Nicholas Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, npj Quantum Inf. **7**, 23 (2021).

[36] Artur F. Izmaylov, Tzu-Ching Yen, Robert A. Lang, and Vladyslav Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, J. Chem. Theory Comput. **16**, 190 (2019).

[37] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[38] Christian Kokail, Christine Maier, Rick van Bijnen, Tiff Brydges, Manoj K. Joshi, Petar Jurcevic, Christine A. Muschik, Pietro Silvi, Rainer Blatt, Christian F. Roos *et al.*, Self-verifying variational quantum simulation of lattice models, Nature **569**, 355 (2019).

[39] Hsin-Yuan Huang, Kishor Bharti, and Patrick Rebentrost, Near-term quantum algorithms for linear systems of equations with regression loss functions, New J. Phys. **23**, 113021 (2021).

[40] Steven T. Flammia and Yi-Kai Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[41] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[42] Otfried Gühne and Géza Tóth, Entanglement detection, Phys. Rep. **474**, 1 (2009).

[43] Andreas Elben, Richard Kueng, Hsin-Yuan R. Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, and Benoît Vermersch, Mixed-state entanglement from local randomized measurements, Phys. Rev. Lett. **125**, 200501 (2020).

[44] David A. Freedman and David A. Freedman, A note on screening regression equations, Am. Stat. **37**, 152 (1983).

[45] Moritz Hardt and Jonathan Ullman, in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science* (IEEE, New York, NY, USA, 2014), p. 454.

[46] Amos Fiat and Tamir Tassa, Dynamic traitor tracing, J. Cryptol. **14**, 211 (2001).

[47] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Information-theoretic bounds on quantum advantage in machine learning, Phys. Rev. Lett. **126**, 190505 (2021).

[48] Threshold search, introduced as "secret acceptor" in Ref. [88], aims to identify whether $\mathrm{tr}(O_j\rho) > \theta_j - \epsilon$ for some j or $\mathrm{tr}(O_i\rho) \leq \theta_i$ for all i, given ρ, observables $O_1, \ldots, O_M$, and thresholds $\theta_1, \ldots, \theta_M$.

[49] Fernando GSL Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu, in *ICALP* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2019).
[50] Joran van Apeldoorn, Andrá S. Gilyén, Sander Gribling, and Ronald de Wolf, Quantum SDP-solvers: Better upper and lower bounds, Quantum **4**, 230 (2020).
[51] For local observables, $x \in [n]$, while for Pauli observables, $x \in \{0,1\}^n$.
[52] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil Vadhan, in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, STOC '09* (Association for Computing Machinery, New York, NY, USA, 2009), p. 381.
[53] Richard Kueng and Hsin-Yuan Huang, Shadow tomography with independent uniform POVM measurements (unpublished, private communication).
[54] Online learning refers to a setting where one is tasked to answer a series of (potentially adversarial) requests in a sequential fashion while minimizing error/loss. This is similar to our setting, where we must predict expectation values for a sequence of observables.
[55] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak, Online learning of quantum states, Adv. Neural Inf. Process. Syst. **31**, 8962 (2018).
[56] To be precise, we are running a (low-dimensional) shadow tomography algorithm as a subroutine within our overall shadow tomography algorithm for bounded Frobenius-norm observables. The low-dimensional subroutine itself contains a teacher and student iteratively interacting to learn $\hat{\rho}$, and mistake declarations by this second teacher directly cause the student to update $\hat{\rho}$.
[57] Jerry Huang, Laura Lewis, Hsin-Yuan Huang, and John Preskill, Figure data for predicting adaptively chosen observables in quantum systems, Zenodo, 2026, https://doi.org/10.5281/zenodo.18451308.
[58] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, IEEE Trans. Inf. Theory **44**, 1897 (1998).
[59] Michael Kearns, Efficient noise-tolerant learning from statistical queries, J. ACM (JACM) **45**, 983 (1998).
[60] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S. Vempala, and Ying Xiao, Statistical algorithms and a lower bound for detecting planted cliques, J. ACM (JACM) **64**, 1 (2017).
[61] Aaron Roth and Adam Smith, The algorithmic foundations of adaptive data analysis, Lecture notes (2017), https://adaptivedataanalysis.com.
[62] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).
[63] Jordan S. Cotler and Frank Wilczek, Quantum overlapping tomography, Phys. Rev. Lett. **124**, 100401 (2020).
[64] Tim J. Evans, Robin Harper, and Steven T. Flammia, Scalable Bayesian Hamiltonian learning, arXiv:1912.07636.
[65] Hsin-Yuan Huang, Richard Kueng, and John Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).
[66] Dax Enshan Koh and Sabee Grewal, Classical shadows with noise, Quantum **6**, 776 (2022).
[67] Arkadij Semenovič Nemirovskij and David Borisovich Yudin, Problem complexity and method efficiency in optimization (Wiley-Interscience, New York, 1983).
[68] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theor. Comput. Sci. **43**, 169 (1986).
[69] Scott Aaronson, in *CCC* (IEEE, New York, NY, USA, 2006), p. 13.
[70] Scott Aaronson, in *Proceedings. 19th IEEE Annual Conference on Computational Complexity, 2004* (IEEE, New York, NY, USA, 2004), p. 320.
[71] Mark M. Wilde, *Quantum Information Theory* (Cambridge University, Cambridge, 2013), 2nd ed.
[72] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A. Tropp, Fast state tomography with optimal error bounds, J. Phys. A: Math. Theor. **53**, 204001 (2020).
[73] Alessandro Rinaldo, 36-710: Advanced Statistical Theory II, Lecture notes, Carnegie Mellon University (2019), https://www.stat.cmu.edu/∼arinaldo/Teaching/36710/F19/.
[74] Andrew J. Scott, Tight informationally complete quantum measurements, J. Phys. A: Math. Gen. **39**, 13507 (2006).
[75] David Gross, Felix Krahmer, and Richard Kueng, A partial derandomization of phaselift using spherical designs, J. Fourier Anal. Appl. **21**, 229 (2015).
[76] Antonio Anna Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, Quantum **8**, 1340 (2024).
[77] Gábor Tardos, Optimal probabilistic fingerprint codes, J. ACM (JACM) **55**, 1 (2008).
[78] Mark Bun, Jonathan Ullman, and Salil Vadhan, in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2014), p. 1.
[79] Jonathan Ullman, in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2013), p. 361.
[80] Tamir Tassa, Low bandwidth dynamic traitor tracing schemes, J. Cryptol. **18**, 167 (2005).
[81] Thijs Laarhoven, Jeroen Doumen, Peter Roelse, Boris Škorić, and Benne de Weger, Dynamic Tardos traitor tracing schemes, IEEE Trans. Inf. Theory **59**, 4230 (2013).
[82] Jonathan Katz and Yehuda Lindell, *Introduction to Modern Cryptography: Principles and Protocols* (Chapman and Hall/CRC, Boca Raton, FL, USA, 2007).
[83] Frank Miller, *Telegraphic Code to Insure Privacy and Secrecy in the Transmission of Telegrams* (CM Cornwell, New York, NY, USA, 1882).
[84] Claude E. Shannon, Communication theory of secrecy systems, Bell Syst. Tech. J. **28**, 656 (1949).
[85] We remark that for local observables, the empirical mean can be used because the classical shadows output has a range bound in this case. However, the attack still works even if median of means is used.
[86] These $\lceil \log d \rceil$ qubits can actually be removed, but we include them for simplicity. They also do not affect the overall system size much.
[87] Note that in Appendix A 1 a, we consider statistical queries $q : \mathcal{X} \to [0,1]$, but this setup is the same for scaled statistical queries $q : \mathcal{X} \to [-C, C]$.

[88] Scott Aaronson, The complexity of quantum states and transformations: From quantum money to black holes, arXiv:1607.05256.

[89] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill, Provably efficient machine learning for quantum many-body problems, Science **377**, eabk3333 (2022).

[90] Frank McSherry and Kunal Talwar, in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (IEEE, New York, NY, USA, 2007), p. 94.

[91] Adam Smith, in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC '11* (Association for Computing Machinery, New York, NY, USA, 2011), p. 813.

[92] Vitaly Feldman, Dealing with range anxiety in mean estimation via statistical queries, in *Proceedings of the 28th International Conference on Algorithmic Learning Theory* (PMLR, 2017), Vol. 76, pp. 629–640.

[93] Cynthia Dwork and Aaron Roth, The algorithmic foundations of differential privacy, Found. Trends Theor. Comput. Sci. **9**, 211 (2014).