Online Learning of Quantum States

Scott Aaronson
UT Austin [*]
aaronson@cs.utexas.edu

Xinyi Chen
Princeton University
xinyic2015@gmail.com

Elad Hazan
Princeton University and Google Brain
ehazan@cs.princeton.edu

Satyen Kale
Google AI, New York
satyenkale@google.com

Ashwin Nayak
University of Waterloo [†]
ashwin.nayak@uwaterloo.ca

Abstract

Suppose we have many copies of an unknown n-qubit state ρ. We measure some copies of ρ using a known two-outcome measurement E_1, then other copies using a measurement E_2, and so on. At each stage t, we generate a current hypothesis ω_t about the state ρ, using the outcomes of the previous measurements. We show that it is possible to do this in a way that guarantees that $|\mathrm{Tr}(E_i\omega_t) - \mathrm{Tr}(E_i\rho)|$, the error in our prediction for the next measurement, is at least ε at most $\mathrm{O}\big(n/\varepsilon^2\big)$ times. Even in the "non-realizable" setting—where there could be arbitrary noise in the measurement outcomes—we show how to output hypothesis states that incur at most $\mathrm{O}(\sqrt{Tn}\ )$ excess loss over the best possible state on the first T measurements. These results generalize a 2007 theorem by Aaronson on the PAC-learnability of quantum states, to the online and regret-minimization settings. We give three different ways to prove our results—using convex optimization, quantum postselection, and sequential fat-shattering dimension—which have different advantages in terms of parameters and portability.

1 Introduction

State tomography is a fundamental task in quantum computing of great practical and theoretical importance. In a typical scenario, we have access to an apparatus that is capable of producing many copies of a quantum state, and we wish to obtain a description of the state via suitable measurements. Such a description would allow us, for example, to check the accuracy with which the apparatus constructs a specific target state.

How many single-copy measurements are needed to "learn" an unknown n-qubit quantum state ρ? Suppose we wish to reconstruct the full $2^n \times 2^n$ density matrix, even approximately, to within ε in trace distance. If we make no assumptions about ρ, then it is straightforward to show that the number of measurements needed grows exponentially with n. In fact, even when we allow joint measurement of multiple copies of the state, an exponential number of copies of ρ are required (see,

[*] Supported by a Vannevar Bush Faculty Fellowship from the US Department of Defense. Part of this work was done while the author was supported by an NSF Alan T. Waterman Award.

[†] Research supported in part by NSERC Canada.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

e.g., O'Donnell and Wright [2016], Haah et al. [2017]). (A "joint measurement" of two or more states on disjoint sequences of qubits is a *single* measurement of all the qubits together.)

Suppose, on the other hand, that there is some probability distribution $\mathcal{D}$ over possible yes/no measurements, where we identify the measurements with $2^n \times 2^n$ Hermitian matrices E with eigenvalues in $[0, 1]$. Further suppose we are only concerned about learning the state ρ well enough to predict the outcomes of *most* measurements E drawn from $\mathcal{D}$—where "predict" means approximately calculating the probability, $\mathrm{Tr}(E\rho)$, of a "yes" result. Then for how many (known) sample measurements E_i, drawn independently from $\mathcal{D}$, do we need to know the approximate value of $\mathrm{Tr}(E_i\rho)$, before we have enough data to achieve this?

Aaronson [2007] proved that the number of sample measurements needed, m, grows only *linearly* with the number of qubits n. What makes this surprising is that it represents an exponential reduction compared to full quantum state tomography. Furthermore, the prediction strategy is extremely simple. Informally, we merely need to find any "hypothesis state" ω that satisfies $\mathrm{Tr}(E_i\omega) \approx \mathrm{Tr}(E_i\rho)$ for all the sample measurements $E_1, \ldots, E_m$. Then with high probability over the choice of sample measurements, that hypothesis ω necessarily "generalizes", in the sense that $\mathrm{Tr}(E\omega) \approx \mathrm{Tr}(E\rho)$ for most additional E's drawn from $\mathcal{D}$. The learning theorem led to followup work including a full characterization of quantum advice (Aaronson and Drucker [2014]); efficient learning for stabilizer states (Rocchetto [2017]); the "shadow tomography" protocol (Aaronson [2018]); and recently, the first experimental demonstration of quantum state PAC-learning (Rocchetto et al. [2017]).

A major drawback of the learning theorem due to Aaronson is the assumption that the sample measurements are drawn *independently* from $\mathcal{D}$—and moreover, that the same distribution $\mathcal{D}$ governs both the training samples, and the measurements on which the learner's performance is later tested. It has long been understood, in computational learning theory, that these assumptions are often unrealistic: they fail to account for adversarial environments, or environments that change over time. This is precisely the state of affairs in current experimental implementations of quantum information processing. Not all measurements of quantum states may be available or feasible in a specific implementation, *which* measurements are feasible is dictated by Nature, and as we develop more control over the experimental set-up, more sophisticated measurements become available. The task of learning a state prepared in the laboratory thus takes the form of a game, with the theorist on one side, and the experimentalist and Nature on the other: the theorist is repeatedly challenged to predict the behaviour of the state with respect to the next measurement that Nature allows the experimentalist to realize, with the opportunity to refine the hypothesis as more measurement data become available.

It is thus desirable to design learning algorithms that work in the more stringent *online learning* model. Here the learner is presented a sequence of input points, say $x_1, x_2, \ldots$, one at a time. Crucially, there is no assumption whatsoever about the x_t's: the sequence could be chosen adversarially, and even adaptively, which means that the choice of x_t might depend on the learner's behavior on $x_1, \ldots, x_{t-1}$. The learner is trying to learn some unknown function $f(x)$, about which it initially knows only that f belongs to some hypothesis class $\mathcal{H}$—or perhaps not even that; we also consider the scenario where the learner simply tries to compete with the best predictor in $\mathcal{H}$, which might or might not be a good predictor. The learning proceeds as follows: for each t, the learner first guesses a value y_t for $f(x_t)$, and is then told the true value $f(x_t)$, or perhaps only an approximation of this value. Our goal is to design a learning algorithm with the following guarantee: *regardless of the sequence of x_t's, the learner's guess, y_t, will be far from the true value $f(x_t)$ at most k times* (where k, of course, is as small as possible). The x_t's on which the learner errs could be spaced arbitrarily; all we require is that they be bounded in number.

This leads to the following question: can the learning theorem established by Aaronson [2007] be generalized to the online learning setting? In other words: is it true that, given a sequence $E_1, E_2, \ldots$ of yes/no measurements, where each E_t is followed shortly afterward by an approximation of $\mathrm{Tr}(E_t\rho)$, there is a way to anticipate the $\mathrm{Tr}(E_t\rho)$ values by guesses $y_t \in [0, 1]$, in such a way that $|y_t - \mathrm{Tr}(E_t\rho)| > \varepsilon$ at most, say, $\mathrm{O}(n)$ times (where $\varepsilon > 0$ is some constant, and n again is the number of qubits)? The purpose of this paper is to provide an affirmative answer.

Throughout the paper, we consider only two-outcome measurements of an n qubit mixed state ρ, and we specify such a measurement by a $2^n \times 2^n$ Hermitian matrix E with eigenvalues in $[0, 1]$. We say that E "accepts" ρ with probability $\mathrm{Tr}(E\rho)$ and "rejects" ρ with probability $1 - \mathrm{Tr}(E\rho)$. We prove that:

Theorem 1. *Let ρ be an n-qubit mixed state, and let $E_1, E_2, \ldots$ be a sequence of 2-outcome measurements that are revealed to the learner one by one, each followed by a value $b_t \in [0,1]$ such that $|\mathrm{Tr}(E_t\rho) - b_t| \leq \varepsilon/3$. Then there is an explicit strategy for outputting hypothesis states $\omega_1, \omega_2, \ldots$ such that $|\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\rho)| > \varepsilon$ for at most $\mathrm{O}\left(\frac{n}{\varepsilon^2}\right)$ values of t.*

We also prove a theorem for the so-called *regret minimization model* (i.e., the "non-realizable case"), where we make no assumption about the input data arising from an actual quantum state, and our goal is simply to do not much worse than the best hypothesis state that could be found with perfect foresight. In this model, the measurements $E_1, E_2, \ldots$ are presented to a learner one-by-one. In iteration t, after seeing E_t, the learner is challenged to output a hypothesis state ω_t, and then suffers a "loss" equal to $\ell_t(\mathrm{Tr}(E_t\omega_t))$ where ℓ_t is a real function that is revealed to the learner. Important examples of loss functions are L_1 loss, when $\ell_t(z) := |z - b_t|$, and L_2 loss, when $\ell_t(z) := (z - b_t)^2$, where $b_t \in [0,1]$. The number b_t may be an approximation of $\mathrm{Tr}(E_t\rho)$ for some fixed but unknown quantum state ρ, but is allowed to be arbitrary in general. In particular, the pairs (E_t, b_t) may not be consistent with any quantum state. Define the *regret* R_T, after T iterations, to be the amount by which the actual loss of the learner exceeds the loss of the best single hypothesis:

$$R_T := \sum_{t=1}^{T} \ell_t(\mathrm{Tr}(E_t\omega_t)) - \min_{\varphi} \sum_{t=1}^{T} \ell_t(\mathrm{Tr}(E_t\varphi)) \ .$$

The learner's objective is to minimize regret. We show that:

Theorem 2. *Let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements on an n-qubit state presented to the learner, and $\ell_1, \ell_2, \ldots$ be the corresponding loss functions revealed in successive iterations in the regret minimization model. Suppose ℓ_t is convex and L-Lipschitz; in particular, for every $x \in \mathbb{R}$, there is a sub-derivative $\ell'_t(x)$ such that $|\ell'_t(x)| \leq L$. Then there is an explicit learning strategy that guarantees regret $R_T = \mathrm{O}(L\sqrt{Tn}\,)$ for all T. This is so even assuming the measurement E_t and loss function ℓ_t are chosen adaptively, in response to the learner's previous behavior.*

Specifically, the algorithm applies to L_1 loss and L_2 loss, and achieves regret $\mathrm{O}(\sqrt{Tn}\,)$ for both.

The online strategies we present enjoy several advantages over full state tomography, and even over "state certification", in which we wish to test whether a given quantum state is close to a desired state or far from it. Optimal algorithms for state tomography (O'Donnell and Wright [2016], Haah et al. [2017]) or certification (Bădescu et al. [2017]) require joint measurements of an exponential number of copies of the quantum state, and assume the ability to perform noiseless, universal quantum computation. On the other hand, the algorithms implicit in Theorems 1 and 2 involve only single-copy measurements, allow for noisy measurements, and capture ground reality more closely. They produce a hypothesis state that mimics the unknown state with respect to measurements that *can be* performed in a given experimental set-up, and the accuracy of prediction improves as the set of available measurements grows. For example, in the realizable case, i.e., when the data arise from an actual quantum state, the average L_1 loss per iteration is $\mathrm{O}(\sqrt{n/T}\,)$. This tends to zero, as the number of measurements becomes large. Note that L_1 loss may be as large as $1/2$ per iteration in the worst case, but this occurs at most $\mathrm{O}(\sqrt{nT}\,)$ times. Finally, the algorithms have run time exponential in the number of qubits in each iteration, but are entirely *classical*. Exponential run time is unavoidable, as the measurements are presented explicitly as $2^n \times 2^n$ matrices, where n is the number of qubits. If we were required to output the hypothesis states, the length of the output—also exponential in the number of qubits—would again entail exponential run time.

It is natural to wonder whether Theorems 1 and 2 leave any room for improvement. Theorem 1 is asymptotically optimal in its mistake bound of $\mathrm{O}(n/\varepsilon^2)$; this follows from the property that n-qubit quantum states, considered as a hypothesis class, have ε-fat-shattering dimension $\Theta(n/\varepsilon^2)$ (see, for example, Aaronson [2007]). On the other hand, there is room to improve Theorem 2. The bounds of which we are aware are $\Omega(\sqrt{Tn}\,)$ for the L_1 loss (see, e.g., [Arora et al., 2012, Theorem 4.1]) in the non-realizable case and $\Omega(n)$ for the L_2 loss in the realizable case, when the feedback consists of the measurement outcomes. (The latter bound, as well as an $\Omega(\sqrt{Tn}\,)$ bound for L_1 loss in the same setting, come from considering quantum mixed states that consist of n independent classical coins, each of which could land heads with probability either $1/2$ or $1/2 + \varepsilon$. The paramater ε is set to $\sqrt{n/T}$.)

We mention an application of Theorem 1, that appears in simultaneous work. Aaronson [2018] has given an algorithm for the so-called *shadow tomography* problem. Here we have an unknown D-dimensional pure state ρ, as well as known two-outcome measurements $E_1, \ldots, E_m$. Our goal is to approximate $\operatorname{Tr}(E_i \rho)$, for *every* i, to within additive error ε. We would like to do this by measuring $\rho^{\otimes k}$, where k is as small as possible. Surprisingly, Aaronson [2018] showed that this can be achieved with $k = \widetilde{\mathrm{O}}((\log M)^4 (\log D)/\varepsilon^5)$, that is, a number of copies of ρ that is only *polylogarithmic* in both D and M. One component of his algorithm is essentially tantamount to online learning with $\widetilde{\mathrm{O}}(n/\varepsilon^3)$ mistakes—i.e., the learning algorithm we present in Section 4 of this paper. However, by using Theorem 1 from this paper in a black-box manner, we can improve the sample complexity of shadow tomography to $\widetilde{\mathrm{O}}((\log M)^4 (\log D)/\varepsilon^4)$. Details appear in (Aaronson [2018]).

To maximize insight, in this paper we give *three* very different approaches to proving Theorems 1 and 2 (although we do not prove every statement with all three approaches). Our first approach is to adapt techniques from online convex optimization to the setting of density matrices, which in general may be over a complex Hilbert space. This requires extending standard techniques to cope with convexity and Taylor approximations, which are widely used for functions over the real domain, but not over the complex domain. We also give an efficient iterative algorithm to produce predictions. This approach connects our problem to the modern mainstream of online learning algorithms, and achieves the best parameters (as stated in Theorems 1 and 2).

Our second approach is via a postselection-based learning procedure, which starts with the maximally mixed state as a hypothesis and then repeatedly refines it by simulating postselected measurements. This approach builds on earlier work due to Aaronson [2005], specifically the proof of $\mathsf{BQP/qpoly} \subseteq \mathsf{PP/poly}$. The advantage is that it is almost entirely self-contained, requiring no "power tools" from convex optimization or learning theory. On the other hand, the approach does not give optimal parameters, and we do not know how to prove Theorem 2 with it.

Our third approach is via an upper-bound on the so-called *sequential fat-shattering dimension* of quantum states, considered as a hypothesis class (see, e.g., Rakhlin et al. [2015]). In the original quantum PAC-learning theorem by Aaronson, the key step was to upper-bound the so-called *ε-fat-shattering dimension* of quantum states considered as a hypothesis class. Fat-shattering dimension is a real-valued generalization of VC dimension. One can then appeal to known results to get a sample-efficient learning algorithm. For online learning, however, bounding the fat-shattering dimension no longer suffices; one instead needs to consider a possibly-larger quantity called sequential fat-shattering dimension. However, by appealing to a lower bound due to Nayak [1999], Ambainis et al. [2002] for a variant of quantum *random access codes*, we are able to upper-bound the sequential fat-shattering dimension of quantum states. Using known results—in particular, those due to Rakhlin et al. [2015]—this implies the regret bound in Theorem 2, up to a multiplicative factor of $\log^{3/2} T$. The statement that *the hypothesis class of n-qubit states has ε-sequential fat-shattering dimension* $\mathrm{O}(n/\varepsilon^2)$ might be of independent interest: among other things, it implies that *any* online learning algorithm that works given bounded sequential fat-shattering dimension, will work for online learning of quantum states. We also give an alternative proof for the lower bound due to Nayak for quantum random access codes, and extend it to codes that are decoded by what we call *measurement decision trees*. We expect these also to be of independent interest.

1.1 Structure of the paper

We start by describing background and the technical learning setting as well as notations used throughout (Section 2). In Section 3 we give the algorithms and main theorems derived using convexity arguments and online convex optimization. In Section 4 we state the main theorem using a postselection algorithm. In Section 5 we give a sequential fat-shattering dimension bound for quantum states and its implication for online learning of quantum states. Proofs of the theorems and related claims are presented in the appendices.

2 Preliminaries and definitions

We define the trace norm of a matrix M as $\|M\|_{\mathrm{Tr}} := \operatorname{Tr} \sqrt{MM^\dagger}$, where $M^\dagger$ is the adjoint of M. We denote the ith eigenvalue of a Hermitian matrix X by $\lambda_i(X)$, its minimum eigenvalue by $\lambda_{\min}(X)$, and its maximum eigenvalue by $\lambda_{\max}(X)$. We sometimes use the notation $X \bullet Y$ to denote the trace inner-product $\mathrm{Tr}(X^\dagger Y)$ between two complex matrices of the same dimensions. By 'log' we denote the natural logarithm, unless the base is explicitly mentioned.

An n-qubit quantum state ρ is an element of C_n, where C_n is the set of all trace-1 positive semi-definite (PSD) complex matrices of dimension 2^n:

$$C_n = \{M \in \mathbb{C}^{2^n \times 2^n} \ , \ M = M^\dagger \ , \ M \succeq 0 \ , \ \mathrm{Tr}(M) = 1\} \ .$$

Note that C_n is a convex set. A two-outcome measurement of an n-qubit state is defined by a $2^n \times 2^n$ Hermitian matrix E with eigenvalues in $[0, 1]$. The measurement E "accepts" ρ with probability $\mathrm{Tr}(E\rho)$, and "rejects" with probability $1 - \mathrm{Tr}(E\rho)$. For the algorithms we present in this article, we assume that a two-outcome measurement is specified via a classical description of its defining matrix E. In the rest of the article, unless mentioned otherwise, a "measurement" refers to a "two-outcome measurement". We refer the reader to the book by Watrous [2018] for a more thorough introduction to the relevant concepts from quantum information.

Online learning and regret. In online learning of quantum states, we have a sequence of iterations $t = 1, 2, 3, \ldots$ of the following form. First, the learner constructs a state $\omega_t \in C_n$; we say that the learner "predicts" ω_t. It then suffers a "loss" $\ell_t(\mathrm{Tr}(E_t\omega_t))$ that depends on a measurement E_t, both of which are presented by an adversary. Commonly used loss functions are L_2 loss (also called "mean square error"), given by

$$\ell_t(z) := (z - b_t)^2 \ ,$$

and L_1 loss (also called "absolute loss"), given by

$$\ell_t(z) := |z - b_t| \ ,$$

where $b_t \in [0, 1]$. The parameter b_t may be an approximation of $\mathrm{Tr}(E_t\rho)$ for some fixed quantum state ρ not known to the learner, obtained by measuring multiple copies of ρ. However, in general, the parameter is allowed to be arbitrary.

The learner then "observes" feedback from the measurement E_t; the feedback is also provided by the adversary. The simplest feedback is the realization of a binary random variable Y_t such that

$$Y_t = \begin{cases} 1 & \text{with probability } \mathrm{Tr}(E_t\rho) \ , \qquad \text{and} \\ 0 & \text{with probability } 1 - \mathrm{Tr}(E_t\rho) \ . \end{cases}$$

Another common feedback is a number b_t as described above, especially in case that the learner suffers L_1 or L_2 loss.

We would like to design a strategy for updating ω_t based on the loss, measurements, and feedback in all the iterations so far, so that the learner's total loss is minimized in the following sense. We would like that over T iterations (for a number T known in advance), the learner's total loss is not much more than that of the hypothetical strategy of outputting the same quantum state φ at every time step, where φ minimizes the total loss *with perfect hindsight*. Formally this is captured by the notion of *regret* R_T, defined as

$$R_T := \sum_{t=1}^{T} \ell_t(\mathrm{Tr}(E_t\omega_t)) - \min_{\varphi \in C_n} \sum_{t=1}^{T} \ell_t(\mathrm{Tr}(E_t\varphi)) \ .$$

The sequence of measurements E_t can be arbitrary, even adversarial, based on the learner's previous actions. Note that if the loss function is given by a fixed state ρ (as in the case of mean square error), the minimum total loss would be 0. This is called the "realizable" case. However, in general, the loss function presented by the adversary need not be consistent with any quantum state. This is called the "non-realizable" case.

A special case of the online learning setting is called *agnostic learning*; here the measurements E_t are drawn from a fixed and unknown distribution $\mathcal{D}$. The setting is called "agnostic" because we still do not assume that the losses correspond to any actual state ρ (i.e., the setting may be non-realizable).

Online mistake bounds. In some online learning scenarios the quantity of interest is not the mean square error, or some other convex loss, but rather simply the total number of "mistakes" made. For example, we may be interested in the number of iterations in which the predicted probability of acceptance $\mathrm{Tr}(E_t\omega_t)$ is more than ε-far from the actual value $\mathrm{Tr}(E_t\rho)$, where ρ is again a fixed state not known to the learner. More formally, let

$$\ell_t(\mathrm{Tr}(E_t\omega_t)) \coloneqq |\,\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\rho)|$$

be the absolute loss function. Then the goal is to bound the number of iterations in which $\ell_t(\mathrm{Tr}(E_t\omega_t)) > \varepsilon$, regardless of the sequence of measurements E_t presented by the adversary. We assume that in this setting,the adversary provides as feedback an approximation $b_t \in [0, 1]$ that satisfies $|\,\mathrm{Tr}(E_t\rho) - b_t| \le \frac{\varepsilon}{3}$.

3 Online learning of quantum states

In this section, we use techniques from online convex optimization to minimize regret. The same algorithms may be adapted to also minimize the number of mistakes made.

3.1 Regularized Follow-the-Leader

We first follow the template of the Regularized Follow-the-Leader algorithm (RFTL; see, for example, [Hazan, 2015, Chapter 5]). The algorithm below makes use of von Neumann entropy, which relates to the Matrix Exponentiated Gradient algorithm (Tsuda et al. [2005]).

Algorithm 1 RFTL for Quantum Tomography

```
Input: $T$, $\mathcal{K} \coloneqq C_n$, $\eta < \frac{1}{2}$
Set $\omega_1 \coloneqq 2^{-n}\mathbb{I}$.
for $t = 1, \dots, T$ do
    Predict $\omega_t$. Consider the convex and $L$-Lipschitz loss function $\ell_t : \mathbb{R} \to \mathbb{R}$ given by measurement $E_t$ : $\ell_t(\mathrm{Tr}(E_t\varphi))$. Let $\ell'_t(x)$ be a sub-derivative of $\ell_t$ with respect to $x$. Define
        $\nabla_t \coloneqq \ell'_t(\mathrm{Tr}(E_t\omega_t))E_t$ .
    Update decision according to the RFTL rule with von Neumann entropy:
        $\omega_{t+1} \coloneqq \arg\min_{\varphi\in\mathcal{K}} \left\{ \eta \sum_{s=1}^{t} \mathrm{Tr}(\nabla_s\varphi) + \sum_{i=1}^{2^n} \lambda_i(\varphi)\log\lambda_i(\varphi) \right\}$ .    (1)
end for
```

Remark 1: The mathematical program in Eq. (1) is convex, and thus can be solved in polynomial time in the dimension, which is 2^n.

Theorem 3. *Setting $\eta = \sqrt{\frac{(\log 2)n}{2TL^2}}$, the regret of Algorithm 1 is bounded by $2L\sqrt{(2\log 2)Tn}$.*

Remark 2: In the case where the feedback is an independent random variable Y_t, where $Y_t = 0$ with probability $1 - \mathrm{Tr}(E_t\rho)$ and $Y_t = 1$ with probability $\mathrm{Tr}(E_t\rho)$ for a fixed but unknown state ρ, we define ∇_t in Algorithm 1 as $\nabla_t \coloneqq 2(\mathrm{Tr}(E_t\omega_t) - Y_t)E_t$. Then $\mathbb{E}[\nabla_t]$ is the gradient of the L_2 loss function where we receive precise feedback $\mathrm{Tr}(E_t\rho)$ instead of Y_t. It follows from the proof of Theorem 3 that the expected L_2 regret of Algorithm 1, namely

$$\mathbb{E}\left[\sum_{t=1}^{T}(\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\rho))^2\right] ,$$

is bounded by $\mathrm{O}(\sqrt{Tn})$.

The proof of Theorem 3 appears in Appendix B. The proof is along the lines of [Hazan, 2015, Theorem 5.2], except that the loss function does not take a raw state as input, and our domain for optimization is complex. Therefore, the mean value theorem does not hold, which means we need to approximate the Bregman divergence instead of replacing it by a norm as in the original proof. Another subtlety is that convexity needs to be carefully defined with respect to the complex domain.

3.2 Matrix Multiplicative Weights

The Matrix Multiplicative Weights (MMW) algorithm [Arora and Kale, 2016] provides an alternative means of proving Theorem 2. The algorithm follows the template of Algorithm 1 with step 5 replaced by the following update rule:

$$\omega_{t+1} \coloneqq \frac{\exp(-\frac{\eta}{L}\sum_{\tau=1}^{t}\nabla_\tau)}{\mathrm{Tr}(\exp(-\frac{\eta}{L}\sum_{\tau=1}^{t}\nabla_\tau))} \quad . \tag{2}$$

In the notation of Arora and Kale [2016], this algorithm is derived using the loss matrices $M_t = \frac{1}{L}\nabla_t = \frac{1}{L}\ell_t'(\mathrm{Tr}(E_t\omega_t))E_t$. Since $\|E_t\| \leq 1$ and $|\ell_t'(\mathrm{Tr}(E_t\omega_t))| \leq L$, we have $\|M_t\| \leq 1$, as requred in the analysis of the Matrix Multiplicative Weights algorithm. We have the following regret bound for the algorithm (proved in Appendix C):

Theorem 4. *Setting $\eta = \sqrt{\frac{(\log 2)n}{4T}}$, the regret of the algorithm based on the update rule* (2) *is bounded by $2L\sqrt{(\log 2)Tn}$.*

3.3 Proof of Theorem 1

Consider either the RFTL or MMW based online learning algorithm described in the previous subsections, with the 1-Lipschitz convex absolute loss function $\ell_t(x) = |x - b_t|$. We run the algorithm in a sub-sequence of the iterations, using only the measurements presented in those iterations. The subsequence of iterations is determined as follows. Let ω_t denote the hypothesis maintained by the algorithm in iteration t. We run the algorithm in iteration t if $\ell_t(\mathrm{Tr}(E_t\omega_t)) > \frac{2\varepsilon}{3}$. Note that whenever $|\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\rho)| > \varepsilon$, we have $\ell_t(\mathrm{Tr}(E_t\omega_t)) > \frac{2\varepsilon}{3}$, so we update the hypothesis according to the RFTL/MMW rule in that iteration.

As we explain next, the algorithm makes at most $\mathrm{O}(\frac{n}{\varepsilon^2})$ updates regardless of the number of measurements presented (i.e., regardless of the number of iterations), giving the required mistake bound. For the true quantum state ρ, we have $\ell_t(\mathrm{Tr}(E_t\rho)) < \frac{\varepsilon}{3}$ for all t. Thus if the algorithm makes T updates (i.e., we run the algorithm in T of the iterations), the regret bound implies that $\frac{2\varepsilon}{3}T \leq \frac{\varepsilon}{3}T + \mathrm{O}(\sqrt{Tn})$. Simplifying, we get the bound $T = \mathrm{O}(\frac{n}{\varepsilon^2})$, as required.

4 Learning Using Postselection

In this section, we give a direct route to proving a slightly weaker version of Theorem 1: one that does not need the tools of convex optimization, but only tools intrinsic to quantum information.

In the following, by a "register" we mean a designated sequence of qubits. Given a two-outcome measurement E on n-qubits states, we define an operator $\mathcal{M}$ that "postselects" on acceptance by E. (While a measurement results in a random outcome distributed according to the probability of acceptance or rejection, *postselection* is a hypothetical operation that produces an outcome of one's choice with certainty.) Let U be any unitary operation on $n+1$ qubits that maps states of the form $|\psi\rangle|0\rangle$ to $\sqrt{E}\,|\psi\rangle|0\rangle + \sqrt{\mathbb{I}-E}\,|\psi\rangle|1\rangle$. Such a unitary operation always exists (see, e.g., [Watrous, 2018, Theorem 2.42]). Denote the $(n+1)$th qubit by register B. Let $\Pi \coloneqq \mathbb{I} \otimes |0\rangle\langle 0|$ be the orthogonal projection onto states that equal $|0\rangle$ in register B. Then we define the operator $\mathcal{M}$ as

$$\mathcal{M}(\varphi) \coloneqq \frac{1}{\mathrm{Tr}(E\varphi)}\ \mathrm{Tr}_B\big(U^{-1}\Pi U\,(\varphi \otimes |0\rangle\langle 0|)\,U^{-1}\Pi U\big) \quad , \tag{3}$$

if $\mathrm{Tr}(E\varphi) \neq 0$, and $\mathcal{M}(\varphi) \coloneqq 0$ otherwise. Here, Tr_B is the *partial trace* operator over qubit B [Watrous, 2018, Section 1.1]. This operator $\mathcal{M}$ has the effect of mapping the quantum state φ to the (normalized) post-measurement state when we perform the measurement E and get outcome "yes" (i.e., the measurement "accepts"). We emphasize that we use a fresh ancilla qubit initialized to state $|0\rangle$ as register B in every application of the operator $\mathcal{M}$. We say that the postselection succeeds with probability $\mathrm{Tr}(E\varphi)$.

We need a slight variant of a well-known result, which Aaronson called the "Quantum Union Bound" (see, for example, Aaronson [2006, 2016], Wilde [2013]).

Theorem 5 (variant of Quantum Union Bound; Gao [2015])**.** *Suppose we have a sequence of two-outcome measurements $E_1, \ldots, E_k$, such that each E_i accepts a certain mixed state φ with probability at least $1 - \varepsilon$. Consider the corresponding operators $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ that postselect on acceptance by the respective measurements $E_1, E_2, \ldots, E_k$. Let $\widetilde{\varphi}$ denote the state $(\mathcal{M}_k \mathcal{M}_{k-1} \cdots \mathcal{M}_1)(\varphi)$ obtained by applying each of the k postselection operations in succession. Then the probability that all the postselection operations succeed, i.e., the k measurements all accept φ, is at least $1 - 2\sqrt{k\varepsilon}$. Moreover, $\|\widetilde{\varphi} - \varphi\|_{\mathrm{Tr}} \leq 4\sqrt{k\varepsilon}$.*

We may infer the above theorem by applying Theorem 1 from (Gao [2015]) to the state φ augmented with k ancillary qubits $B_1, B_2, \ldots, B_k$ initialized to 0, and considering k orthogonal projection operators $U_i^{-1} \Pi_i U_i$, where the unitary operator U_i and the projection operator Π_i are as defined for the postselection operation $\mathcal{M}_i$ for E_i. The ith projection operator $U_i^{-1} \Pi_i U_i$ acts on the registers holding φ and the ith ancillary qubit B_i.

We prove the main result of this section using suitably defined postselection operators in an online learning algorithm (proof in Appendix D):

Theorem 6. *Let ρ be an unknown n-qubit mixed state, let $E_1, E_2, \ldots$ be a sequence of two-outcome measurements, and let $\varepsilon > 0$. There exists a strategy for outputting hypothesis states $\omega_0, \omega_1, \ldots$, where ω_t depends only on $E_1, \ldots, E_t$ and real numbers $b_1, \ldots, b_t$ in $[0,1]$, such that as long as $|b_t - \mathrm{Tr}(E_t \rho)| \leq \varepsilon/3$ for every t, we have*

$$|\mathrm{Tr}(E_{t+1}\omega_t) - \mathrm{Tr}(E_{t+1}\rho)| > \varepsilon$$

for at most $\mathrm{O}\big(\frac{n}{\varepsilon^3} \log \frac{n}{\varepsilon}\big)$ values of t. Here the E_t's and b_t's can otherwise be chosen adversarially.

5 Learning Using Sequential Fat-Shattering Dimension

In this section, we prove regret bounds using the notion of *sequential fat-shattering dimension*. Let S be a set of functions $f : U \to [0,1]$, and $\varepsilon > 0$. Then, following Rakhlin et al. [2015], let the *ε-sequential fat-shattering dimension* of S, or $\mathrm{sfat}_\varepsilon(S)$, be the largest k for which we can construct a complete binary tree T of depth k, such that

- each internal vertex $v \in T$ has associated with it a point $x_v \in U$ and a real $a_v \in [0,1]$, and
- for each leaf vertex $v \in T$ there exists an $f \in S$ that causes us to reach v if we traverse T from the root such that at any internal node w we traverse the left subtree if $f(x_w) \leq a_w - \varepsilon$ and the right subtree if $f(x_w) \geq a_w + \varepsilon$. If we view the leaf v as a k-bit string, the function f is such that for all ancestors u of v, we have $f(x_u) \leq a_u - \varepsilon$ if $v_i = 0$, and $f(x_u) \geq a_u + \varepsilon$ if $v_i = 1$, when u is at depth $i - 1$ from the root.

An n-qubit state ρ induces a function f on the set of two-outcome measurements E defined as $f(E) \coloneqq \mathrm{Tr}(E\rho)$. With this correspondence in mind, we establish a bound on the sequential fat-shattering dimension of the set of n-qubit quantum states. The bound is based on a generalization of "random access coding" (Nayak [1999], Ambainis et al. [2002]) called "serial encoding". We derive the following bound on the length of serial encoding. Let $\mathrm{H}(x) \coloneqq -x \log_2 x - (1-x)\log_2(1-x)$ be the binary entropy function.

Corollary 7. *Let k and n be positive integers. For each k-bit string $y \coloneqq y_1 \cdots y_k$, let ρ_y be an n-qubit mixed state such that for each $i \in \{1, 2, \ldots, k\}$, there is a two-outcome measurement E' that depends only on i and the prefix $v \coloneqq y_1 y_2 \cdots y_{i-1}$, and has the following properties*

(iii) if $y_i = 0$ then $\mathrm{Tr}(E'\rho_y) \leq a_v - \varepsilon$, and

(iv) if $y_i = 1$ then $\mathrm{Tr}(E'\rho_y) \geq a_v + \varepsilon$,

where $\varepsilon \in (0, 1/2]$ and $a_v \in [0,1]$ is a "pivot point" associated with the prefix v. Then

$$n \quad \geq \quad \left(1 - \mathrm{H}\left(\frac{1-\varepsilon}{2}\right)\right) k \ .$$

In particular, $k = \mathrm{O}\big(n/\varepsilon^2\big)$.

(The proof is presented in Appendix E).

Corollary 7 immediately implies the following theorem:

Theorem 8. *Let U be the set of two-outcome measurements E on an n-qubit state, and let S be the set of all functions $f : U \to [0,1]$ that have the form $f(E) := \operatorname{Tr}(E\rho)$ for some ρ. Then for all $\varepsilon > 0$, we have $\operatorname{sfat}_\varepsilon(S) = \mathrm{O}\big(n/\varepsilon^2\big)$.*

Theorem 8 strengthens an earlier result due to Aaronson [2007], which proved the same upper bound for the "ordinary" (non-sequential) fat-shattering dimension of quantum states considered as a hypothesis class.

Now we may use existing results from the literature, which relate sequential fat-shattering dimension to online learnability. In particular, in the non-realizable case, Rakhlin et al. [2015] recently showed the following:

Theorem 9 (Rakhlin et al. [2015])**.** *Let S be a set of functions $f : U \to [0,1]$ and for every integer $t \geq 1$, let $\ell_t : [0,1] \to \mathbb{R}$ be a convex, L-Lipschitz loss function. Suppose we are sequentially presented elements $x_1, x_2, \ldots \in U$, with each x_t followed by the loss function ℓ_t. Then there exists a learning strategy that lets us output a sequence of hypotheses $f_1, f_2, \ldots \in S$, such that the regret is upper-bounded as:*

$$\sum_{t=1}^{T} \ell_t\left(f_t(x_t)\right) \leq \min_{f\in S} \sum_{t=1}^{T} \ell_t\left(f(x_t)\right) + 2LT \inf_{\alpha} \left\{ 4\alpha + \frac{12}{\sqrt{T}} \int_{\alpha}^{1} \sqrt{\operatorname{sfat}_\beta(S) \log\left(\frac{2\mathrm{e}T}{\beta}\right)} \mathrm{d}\beta \right\}.$$

This follows from Theorem 8 in (Rakhlin et al. [2015]) as in the proof of Proposition 9 in the same article.

Combining Theorem 8 with Theorem 9 gives us the following:

Corollary 10. *Suppose we are presented with a sequence of two-outcome measurements $E_1, E_2, \ldots$ of an n-qubit state, with each E_t followed by a loss function ℓ_t as in Theorem 9. Then there exists a learning strategy that lets us output a sequence of hypothesis states $\omega_1, \omega_2, \ldots$ such that the regret after the first T iterations is upper-bounded as:*

$$\sum_{t=1}^{T} \ell_t\left(\operatorname{Tr}(E_t\omega_t)\right) \leq \min_{\omega\in C_n} \sum_{t=1}^{T} \ell_t\left(\operatorname{Tr}(E_t\omega)\right) + \mathrm{O}\Big(L\sqrt{nT}\log^{3/2} T\Big).$$

Note that the result due to Rakhlin et al. [2015] is non-explicit. In other words, by following this approach, we do not derive any specific online learning algorithm for quantum states that has the stated upper bound on regret; we only prove non-constructively that such an algorithm exists.

We expect that the approach in this section, based on sequential fat-shattering dimension, could also be used to prove a mistake bound for the realizable case, but we leave that to future work.

6 Open Problems

We conclude with some questions arising from this work. The regret bound established in Theorem 2 for L_1 loss is tight. Can we similarly achieve optimal regret for other loss functions of interest, for example for L_2-loss? It would also be interesting to obtain regret bounds in terms of the loss of the best quantum state in hindsight, as opposed to T (the number of iterations), using the techniques in this article. Such a bound has been shown by [Tsuda et al., 2005, Lemma 3.2] for L_2-loss using the Matrix Exponentiated Gradient method.

In what cases can one do online learning of quantum states, not only with few samples, but also with a polynomial amount of computation? What is the tight generalization of our results to measurements with d outcomes? Is it the case, in online learning of quantum states, that *any* algorithm works, so long as it produces hypothesis states that are approximately consistent with all the data seen so far? Note that none of our three proof techniques seem to imply this general conclusion.

References

S. Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1: 1–28, 2005. Earlier version in CCC'2004. quant-ph/0402095.

S. Aaronson. QMA/qpoly is contained in PSPACE/poly: de-Merlinizing quantum protocols. In *Proc. Conference on Computational Complexity*, pages 261–273, 2006. quant-ph/0510230.

S. Aaronson. The learnability of quantum states. *Proc. Roy. Soc. London*, A463(2088):3089–3114, 2007. quant-ph/0608142.

S. Aaronson. The complexity of quantum states and transformations: From quantum money to black holes, February 2016. Lecture Notes for the 28th McGill Invitational Workshop on Computational Complexity, Holetown, Barbados. With guest lectures by A. Bouland and L. Schaeffer. www.scottaaronson.com/barbados-2016.pdf.

S. Aaronson. Shadow tomography of quantum states. In *Proc. ACM STOC*, STOC 2018, pages 325–338, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5559-9. doi: 10.1145/3188745.3188802. arXiv:1711.01053.

S. Aaronson and A. Drucker. A full characterization of quantum advice. *SIAM J. Comput.*, 43(3): 1131–1183, 2014. Earlier version in STOC'2010. arXiv:1004.0377.

A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Quantum dense coding and quantum finite automata. *J. of the ACM*, 49:496–511, 2002. Combination of an earlier version in STOC'1999, pp. 376-383, arXiv:quant-ph/9804043 and (Nayak [1999]).

S. Arora and S. Kale. A combinatorial, primal-dual approach to semidefinite programs. *J. ACM*, 63 (2):12:1–12:35, 2016.

S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

K. M. R. Audenaert and J. Eisert. Continuity bounds on the quantum relative entropy. *Journal of Mathematical Physics*, 46(10):102104, 2005. arXiv:quant-ph/0503218.

C. Bădescu, R. O'Donnell, and J. Wright. Quantum state certification. Technical Report arXiv:1708.06002 [quant-ph], arXiv.org, 2017.

Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

E. A. Carlen and E. H. Lieb. Remainder terms for some quantum entropy inequalities. *Journal of Mathematical Physics*, 55(4), 2014. arXiv:1402.3840.

J. Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92:052331, Nov 2015.

J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, Sept 2017.

E. Hazan. *Introduction to Online Convex Optimization*, volume 2 of *Foundations and Trends in Optimization*. 2015.

A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proc. IEEE FOCS*, pages 369–376, 1999. quant-ph/9904093.

R. O'Donnell and J. Wright. Efficient quantum tomography. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 899–912, New York, NY, USA, 2016. ACM.

A. Rakhlin, K. Sridharan, and A. Tewari. Online learning via sequential complexities. *The Journal of Machine Learning Research*, 16(1):155–186, 2015.

A. Rocchetto. Stabiliser states are efficiently PAC-learnable. arXiv:1705.00345, 2017.

A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino. Experimental learning of quantum states. arXiv:1712.00127, 2017.

K. Tsuda, G. Rätsch, and M. K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *Journal of Machine Learning Research*, 6:995–1018, 2005.

J. Watrous. *Theory of Quantum Information*. Cambridge University Press, May 2018.

M. Wilde. Sequential decoding of a general classical-quantum channel. *Proc. Roy. Soc. London*, A469(2157):20130259, 2013. arXiv:1303.0808.

A Auxiliary Lemmas

The following lemma is from (Tsuda et al. [2005]), given here for completeness.

Lemma 11. *For Hermitian matrices A, B and Hermitian PSD matrix X, if $A \succeq B$, then* $\mathrm{Tr}(AX) \geq \mathrm{Tr}(BX)$.

Proof. Let $C \coloneqq A - B$. By definition, $C \succeq 0$. It suffices to show that $\mathrm{Tr}(CX) \geq 0$. Let $VQV^\dagger$ be the eigen-decomposition of X, and let $C = VPV^\dagger$, where $P \coloneqq V^\dagger CV \succeq 0$. Then $\mathrm{Tr}(CX) = \mathrm{Tr}(VPQV^\dagger) = \mathrm{Tr}(PQ) = \sum_{i=1}^n P_{ii}Q_{ii}$. Since $P \succeq 0$ and all the eigenvalues of X are nonnegative, $P_{ii} \geq 0$, $Q_{ii} \geq 0$. Therefore $\mathrm{Tr}(CX) \geq 0$. □

Lemma 12. *If A, B are Hermitian matrices, then* $\mathrm{Tr}(AB) \in \mathbb{R}$.

Proof. The proof is similar to Lemma 11. Let $VQV^\dagger$ be the eigendecomposition of A. Then Q is a real diagonal matrix. We have $B = VPV^\dagger$, where $P \coloneqq V^\dagger BV$. Note that $P^\dagger = V^\dagger B^\dagger V = P$, so P has a real diagonal. Then $\mathrm{Tr}(AB) = \mathrm{Tr}(VQV^\dagger VPV^\dagger) = \mathrm{Tr}(VQPV^\dagger) = \mathrm{Tr}(QP) = \sum_{i=1}^n Q_{ii}P_{ii}$. Since $Q_{ii}, P_{ii} \in \mathbb{R}$ for all i, $\mathrm{Tr}(AB) \in \mathbb{R}$. □

B Proof of Theorem 3

Proof of Theorem 3. Since ℓ_t is convex, for all $\varphi \in \mathcal{K}$,

$$\ell_t(\mathrm{Tr}(E_t\omega_t)) - \ell_t(\mathrm{Tr}(E_t\varphi)) \leq \ell_t'(\mathrm{Tr}(E_t\omega_t))\,[\mathrm{Tr}(E_t\omega_t) - \mathrm{Tr}(E_t\varphi)] = \nabla_t \bullet (\omega_t - \varphi) \quad .$$

(Recall that '$\bullet$' denotes the trace inner-product between complex matrices of the same dimensions.) Summing over t,

$$\sum_{t=1}^T [\ell_t(\mathrm{Tr}(E_t\omega_t)) - \ell_t(\mathrm{Tr}(E_t\varphi))] \leq \sum_{t=1}^T [\mathrm{Tr}(\nabla_t\omega_t) - \mathrm{Tr}(\nabla_t\varphi)] \quad .$$

Define $g_t(X) = \nabla_t \bullet X$, and $g_0(X) = \frac{1}{\eta}R(X)$, where $R(X)$ is the negative von Neumann Entropy of X (in nats). Denote $D_R^2 \coloneqq \max_{\varphi,\varphi' \in \mathcal{K}}\{R(\varphi) - R(\varphi')\}$. By [Hazan, 2015, Lemma 5.2], for any $\varphi \in \mathcal{K}$, we have

$$\sum_{t=1}^T [g_t(\omega_t) - g_t(\varphi)] \leq \sum_{t=1}^T \nabla_t \bullet (\omega_t - \omega_{t+1}) + \frac{1}{\eta}D_R^2 \quad . \tag{4}$$

Define $\Phi_t(X) = \{\eta \sum_{s=1}^t \nabla_s \bullet X + R(X)\}$, then the convex program in line 5 of Algorithm 1 finds the minimizer of $\Phi_t(X)$ in $\mathcal{K}$. The following claim shows that that the minimizer is always *positive definite* (proof provided later in this section):

Claim 13. *For all $t \in \{1, 2, ..., T\}$, we have $\omega_t \succ 0$.*

For $X \succ 0$, we can write $R(X) = \mathrm{Tr}(X \log X)$, and define

$$\nabla\Phi_t(X) \coloneqq \eta \sum_{s=1}^t \nabla_s + \mathbb{I} + \log X \quad .$$

The definition of $\nabla\Phi_t(X)$ is analogous to the gradient of $\Phi_t(X)$ if the function is defined over real symmetric matrices. Moreover, the following condition, similar to the optimality condition over a real domain, is satisfied (proof provided later in this section).

Claim 14. *For all $t \in \{1, 2, \ldots, T-1\}$,*

$$\nabla\Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) \geq 0 \ . \tag{5}$$

Denote

$$B_{\Phi_t}(\omega_t \| \omega_{t+1}) \coloneqq \Phi_t(\omega_t) - \Phi_t(\omega_{t+1}) - \nabla\Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) \ .$$

Then by the Pinsker inequality (see, for example, Carlen and Lieb [2014] and the references therein),

$$\frac{1}{2}\|\omega_t - \omega_{t+1}\|_{\mathrm{Tr}}^2 \leq \mathrm{Tr}(\omega_t \log \omega_t) - \mathrm{Tr}(\omega_t \log \omega_{t+1}) = B_{\Phi_t}(\omega_t \| \omega_{t+1}) \ .$$

We have

$$\begin{aligned} B_{\Phi_t}(\omega_t \| \omega_{t+1}) &= \Phi_t(\omega_t) - \Phi_t(\omega_{t+1}) - \nabla\Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) \\ &\leq \Phi_t(\omega_t) - \Phi_t(\omega_{t+1}) \\ &= \Phi_{t-1}(\omega_t) - \Phi_{t-1}(\omega_{t+1}) + \eta \nabla_t \bullet (\omega_t - \omega_{t+1}) \\ &\leq \eta \nabla_t \bullet (\omega_t - \omega_{t+1}) \ , \end{aligned} \tag{6}$$

where the first inequality follows from Claim 14, and the second because $\Phi_{t-1}(\omega_t) \leq \Phi_{t-1}(\omega_{t+1})$ (ω_t minimizes $\Phi_{t-1}(X)$). Therefore

$$\frac{1}{2}\|\omega_t - \omega_{t+1}\|_{\mathrm{Tr}}^2 \leq \eta \nabla_t \bullet (\omega_t - \omega_{t+1}) \ . \tag{7}$$

Let $\|M\|_{\mathrm{Tr}}^*$ denote the dual of the trace norm, i.e., the spectral norm of the matrix M. By Generalized Cauchy-Schwartz [Bhatia, 1997, Exercise IV.1.14, page 90],

$$\begin{aligned} \nabla_t \bullet (\omega_t - \omega_{t+1}) &\leq \|\nabla_t\|_{\mathrm{Tr}}^* \ \|\omega_t - \omega_{t+1}\|_{\mathrm{Tr}} \\ &\leq \|\nabla_t\|_{\mathrm{Tr}}^* \sqrt{2\eta \nabla_t \bullet (\omega_t - \omega_{t+1})} \ . \qquad \text{by Eq. (7).} \end{aligned}$$

Rearranging,

$$\nabla_t \bullet (\omega_t - \omega_{t+1}) \leq 2\eta \|\nabla_t\|_{\mathrm{Tr}}^{*2} \leq 2\eta G_R^2 \ ,$$

where G_R is an upper bound on $\|\nabla_t\|_{\mathrm{Tr}}^*$. Combining with Eq. (4), we arrive at the following bound

$$\sum_{t=1}^T \nabla_t \bullet (\omega_t - \varphi) \leq \sum_{t=1}^T \nabla_t \bullet (\omega_t - \omega_{t+1}) + \frac{1}{\eta} D_R^2 \leq 2\eta T G_R^2 + \frac{1}{\eta} D_R^2 \ .$$

Taking $\eta = \frac{D_R}{G_R\sqrt{2T}}$, we get $\sum_{t=1}^T \nabla_t \bullet (\omega_t - \varphi) \leq 2 D_R G_R \sqrt{2T}$. Going back to the regret bound,

$$\sum_{t=1}^T [\ell_t(\mathrm{Tr}(E_t \omega_t)) - \ell_t(\mathrm{Tr}(E_t \varphi))] \leq \sum_{t=1}^T \nabla_t \bullet (\omega_t - \varphi) \leq 2 D_R G_R \sqrt{2T} \ .$$

We proceed to show that $D_R = \sqrt{(\log 2) n}$. Let Δ_{2^n} denote the set of probability distributions over $[2^n]$. By definition,

$$D_R^2 = \max_{\varphi, \varphi' \in \mathcal{K}} \{R(\varphi) - R(\varphi')\} = \max_{\varphi \in \mathcal{K}} -R(\varphi) = \max_{\lambda \in \triangle_{2^n}} \sum_{i=1}^{2^n} \lambda_i \log \frac{1}{\lambda_i} = n \log 2 \ .$$

Since the dual norm of the trace norm is the spectral norm, we have

$$\|\nabla_t\|_{\mathrm{Tr}}^* = \|\ell_t'(\mathrm{Tr}(E_t \omega_t)) E_t\| \leq L \|E_t\| \leq L \ .$$

Therefore $\sum_{t=1}^T [(\ell_t(\mathrm{Tr}(E_t \omega_t)) - \ell_t(\mathrm{Tr}(E_t \varphi))] \leq 2L\sqrt{(2 \log 2) n T}$. □

Proof of Claim 13. Let $P \in \mathcal{K}$ be such that $\lambda_{\min}(P) = 0$. Suppose $P = VQV^\dagger$, where Q is a diagonal matrix with real values on the diagonal. Assume that $Q_{1,1} = \lambda_{\max}(P)$ and $Q_{2^n, 2^n} = \lambda_{\min}(P) = 0$. Let $P' = VQ'V^\dagger$ such that $Q'_{1,1} = Q_{1,1} - \varepsilon$, $Q'_{2^n, 2^n} = \varepsilon$ for $\varepsilon < \lambda_{\max}(P)$, and $Q'_{ii} = Q_{ii}$ for $i \in \{2, 3, ..., 2^n - 1\}$, so $P' \in \mathcal{K}$. We show that there exists $\varepsilon > 0$ such that

$\Phi_t(P') \leq \Phi_t(P)$. Expanding both sides of the inequality, we see that it is equivalent to showing that for some ε,

$$\eta \sum_{s=1}^{t} \nabla_s \bullet (P' - P) \leq \lambda_1(P) \log \lambda_1(P) - \lambda_1(P') \log \lambda_1(P') - \varepsilon \log \varepsilon \ \ .$$

Let $\alpha = \lambda_1(P) = Q_{1,1}$, and $A = \eta \sum_{s=1}^{t} \nabla_s$. The inequality then becomes

$$A \bullet (P' - P) \leq \alpha \log \alpha - (\alpha - \varepsilon) \log(\alpha - \varepsilon) - \varepsilon \log \varepsilon \ \ .$$

Observe that $\|A\| \leq \eta \sum_{s=1}^{t} \|\nabla_s\| = \eta \sum_{s=1}^{t} \|\ell'_s(\mathrm{Tr}(E_s \omega_s)) E_s\| \leq \eta L t$. So by the Generalized Cauchy-Schwartz inequality,

$$A \bullet (P' - P) \leq \eta L t \, \|P' - P\|_{\mathrm{Tr}} \leq 2\varepsilon \eta L t \ \ .$$

Since η, t, α, L are finite and $-\log \varepsilon \to \infty$ as $\varepsilon \to 0$, there exists ε small such that $2\eta L t \leq \log \alpha - \log \varepsilon$. We have

$$\begin{aligned} 2\eta L t \varepsilon &\leq \varepsilon \log \alpha - \varepsilon \log \varepsilon \\ &= \alpha \log \alpha - (\alpha - \varepsilon) \log \alpha - \varepsilon \log \varepsilon \\ &\leq \alpha \log \alpha - (\alpha - \varepsilon) \log(\alpha - \varepsilon) - \varepsilon \log \varepsilon \ \ . \end{aligned}$$

So there exists $\varepsilon > 0$ such that $\Phi_t(P') \leq \Phi_t(P)$. If P has multiple eigenvalues that are 0, we can repeat the proof and show that there exists a PD matrix P' such that $\Phi_t(P') \leq \Phi_t(P)$. Since ω_t is a minimizer of Φ_{t-1} and $\omega_1 \succ 0$, we conclude that $\omega_t \succ 0$ for all t. □

Proof of Claim 14. Suppose $\nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) < 0$. Let $a \in (0,1)$ and $\bar{X} = (1-a)\omega_{t+1} + a\omega_t$, then $\bar{X}$ is a density matrix and is positive definite. Define $\triangle = \bar{X} - \omega_{t+1} = a(\omega_t - \omega_{t+1})$. We have

$$\begin{aligned} \Phi_t(\bar{X}) - \Phi_t(\omega_{t+1}) &= a \nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) + B_{\Phi_t}(\bar{X} \| \omega_{t+1}) \\ &\leq a \nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) + \frac{\mathrm{Tr}(\triangle^2)}{\lambda_{\min}(\omega_{t+1})} \\ &= a \nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) + \frac{a^2 \ \mathrm{Tr}((\omega_t - \omega_{t+1})^2)}{\lambda_{\min}(\omega_{t+1})} \ \ . \end{aligned}$$

The above inequality is due to [Audenaert and Eisert, 2005, Theorem 2]. Dividing by a on both sides, we have

$$\frac{\Phi_t(\bar{X}) - \Phi_t(\omega_{t+1})}{a} \leq \nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) + \frac{a \, \mathrm{Tr}((\omega_t - \omega_{t+1})^2)}{\lambda_{\min}(\omega_{t+1})} \ \ .$$

So we can find a small enough such that the right hand side of the above inequality is negative. However, we would have $\Phi_t(\bar{X}) - \Phi_t(\omega_{t+1}) < 0$, which is a contradiction. So $\nabla \Phi_t(\omega_{t+1}) \bullet (\omega_t - \omega_{t+1}) \geq 0$. □

C Proof of Theorem 4

Proof of Theorem 4. Note that for any density matrix φ, we have $M_t \bullet \varphi = \frac{1}{L} \ell'_t(\mathrm{Tr}(E_t \omega_t)) \, \mathrm{Tr}(E_t \varphi)$. Then, the regret bound for Matrix Multiplicative Weights [Arora and Kale, 2016, Theorem 3.1] implies that for any density matrix φ, we have

$$\sum_{t=1}^{T} \ell'_t(\mathrm{Tr}(E_t \omega_t)) \, \mathrm{Tr}(E_t \omega_t) \leq \sum_{t=1}^{T} \ell'_t(\mathrm{Tr}(E_t \omega_t)) \, \mathrm{Tr}(E_t \varphi) + \eta L T + \frac{L \log(2^n)}{\eta} \ \ .$$

Here, we used the bound $M_t^2 \bullet \omega_t \leq 1$. Next, since ℓ_t is convex, we have

$$\ell'_t(\mathrm{Tr}(E_t \omega_t)) \, \mathrm{Tr}(E_t \omega_t) - \ell'_t(\mathrm{Tr}(E_t \omega_t)) \, \mathrm{Tr}(E_t \varphi) \geq \ell_t(\mathrm{Tr}(E_t \omega_t)) - \ell_t(\mathrm{Tr}(E_t \varphi)) \ \ .$$

Using this bound, and the stated value of η, we get the required regret bound. □

D Proof of Theorem 6

Proof of Theorem 6. Let $\rho^* := \rho^{\otimes k}$ be an amplified version of ρ, over a Hilbert space of dimension $D := 2^{kn}$, for some k to be set later. Throughout, we maintain a classical description of a D-dimensional "amplified hypothesis state" ω_t^*, which we view as being the state of k registers with n qubits each. We ensure that ω_t^* is always symmetric under permuting the k registers. Given ω_t^*, our actual n-qubit hypothesis state ω_t is then obtained by simply tracing out $k-1$ of the registers.

Given an amplified hypothesis state ω^*, let E_t^* be a two-outcome measurement that acts on ω^* as follows: it applies the measurement E_t to each of the k registers separately, and accepts if and only if the fraction of measurements that accept equals b_t, up to an additive error at most $\varepsilon/2$.

Here is the learning strategy. Our initial hypothesis, $\omega_0^* := \mathbb{I}/D$, is the D-dimensional maximally mixed state, corresponding to $\omega_0 := \mathbb{I}/2^n$. (The *maximally mixed state* corresponds to the notion of a uniformly random quantum superposition.) For each $t \geq 1$, we are given descriptions of the measurements $E_1, \ldots, E_t$, as well as real numbers $b_1, \ldots, b_t$ in $[0,1]$, such that $|b_i - \mathrm{Tr}\,(E_i\rho)| \leq \varepsilon/3$ for all $i \in [t]$. We would like to update our old hypothesis ω_{t-1}^* to a new hypothesis ω_t^*, ideally such that the difference $|\mathrm{Tr}\,(E_{t+1}\omega_t) - \mathrm{Tr}\,(E_{t+1}\rho)|$ is small. We do so as follows:

- Given b_t, as well classical descriptions of ω_{t-1}^* and E_t, decide whether $\mathrm{Tr}\big(E_t^*\omega_{t-1}^*\big) \geq 1 - \frac{\varepsilon}{6}$.
- If yes, then set $\omega_t^* := \omega_{t-1}^*$ (i.e., we do not change the hypothesis).
- Otherwise, let ω_t^* be the state obtained by applying E_t^* to ω_{t-1}^* and postselecting on E_t^* accepting. In other words, $\omega_t^* := \mathcal{M}(\omega_{t-1}^*)$, where $\mathcal{M}$ is the operator that postselects on acceptance by E_t^* (as defined above).

We now analyze this strategy. Call t "good" if $\mathrm{Tr}\big(E_t^*\omega_{t-1}^*\big) \geq 1 - \frac{\varepsilon}{6}$, and "bad" otherwise. Below, we show that

(i) there are at most $\mathrm{O}\big(\frac{n}{\varepsilon^3}\log\frac{n}{\varepsilon}\big)$ bad t's, and

(ii) for each good t, we have $|\mathrm{Tr}(E_t\omega_{t-1}) - \mathrm{Tr}(E_t\rho)| \leq \varepsilon$.

We start with claim (i). Suppose there have been ℓ bad t's, call them $t(1), \ldots, t(\ell)$, where $\ell \leq (n/\varepsilon)^{10}$ (we justify this last assumption later, with room to spare). Then there were ℓ events where we postselected on E_t^* accepting ω_{t-1}^*. We conduct a thought experiment, in which the learning strategy maintains a quantum register initially in the maximally mixed state $\mathbb{I}/D$, and applies the postselection operator corresponding to E_t^* to the quantum register whenever t is bad. Let p be the probability that all ℓ of these postselection events succeed. Then by definition,

$$p = \mathrm{Tr}\Big(E_{t(1)}^*\omega_{t(1)-1}^*\Big)\cdots\mathrm{Tr}\Big(E_{t(\ell)}^*\omega_{t(\ell)-1}^*\Big) \leq \Big(1 - \frac{\varepsilon}{6}\Big)^{\ell}.$$

On the other hand, suppose counterfactually that we had started with the "true" hypothesis, $\omega_0^* := \rho^* = \rho^{\otimes k}$. In that case, we would have

$$\begin{aligned}\mathrm{Tr}\Big(E_{t(i)}^*\rho^*\Big) &= \Pr\Big[E_{t(i)} \text{ accepts } \rho \text{ between } \Big(b_{t(i)} - \frac{\varepsilon}{2}\Big)k \text{ and } \Big(b_{t(i)} + \frac{\varepsilon}{2}\Big)k \text{ times}\Big] \\ &\geq 1 - 2\,\mathrm{e}^{-2k(\varepsilon/6)^2}\end{aligned}$$

for all i. Here the second line follows from the condition that $\big|\mathrm{Tr}\big(E_{t(i)}\rho\big) - b_{t(i)}\big| \leq \varepsilon/6$, together with the Hoeffding bound.

We now make the choice $k := \frac{C}{\varepsilon^2}\log\frac{n}{\varepsilon}$, for some constant C large enough that

$$\mathrm{Tr}\Big(E_{t(i)}^*\rho^*\Big) \geq 1 - \frac{\varepsilon^{10}}{400n^{10}}$$

for all i. So by Theorem 5, all ℓ postselection events would succeed with probability at least

$$1 - 2\sqrt{\ell\frac{\varepsilon^{10}}{400n^{10}}} \geq 0.9 \ .$$

We may write the maximally mixed state, $\mathbb{I}/D$, as

$$\frac{1}{D}\rho^* + \left(1 - \frac{1}{D}\right)\xi \ ,$$

for some other mixed state ξ. For this reason, even when we start with initial hypothesis $\omega_0^* = \mathbb{I}/D$, all ℓ postselection events still succeed with probability

$$p \geq \frac{0.9}{D} \ .$$

Combining our upper and lower bounds on p now yields

$$\frac{0.9}{2^{kn}} \leq \left(1 - \frac{\varepsilon}{6}\right)^\ell$$

or

$$\ell = \mathrm{O}\left(\frac{kn}{\varepsilon}\right) = \mathrm{O}\left(\frac{n}{\varepsilon^3}\log\frac{n}{\varepsilon}\right),$$

which incidentally justifies our earlier assumption that $\ell \leq (n/\varepsilon)^{10}$.

It remains only to prove claim (ii). Suppose that

$$\mathrm{Tr}\big(E_t^* \omega_{t-1}^*\big) \geq 1 - \frac{\varepsilon}{6} \ . \tag{8}$$

Imagine measuring k quantum registers prepared in the joint state ω_{t-1}^*, by applying E_t to each register. Since the state ω_{t-1}^* is symmetric under permutation of the k registers, we have that $\mathrm{Tr}(E_t\omega_{t-1})$, the probability that E_t accepts the first register, equals the expected fraction of the k registers that E_t accepts. The bound in Eq. (8) means that, with probability at least $1 - \frac{\varepsilon}{6}$ over the measurement outcomes, the fraction of registers which E_t accepts is within $\pm\varepsilon/2$ of b_t. The k measurement outcomes are not necessarily independent, but the fraction of registers accepted never differs from b_t by more than 1. So by the union bound, we have

$$|\mathrm{Tr}(E_t\omega_{t-1}) - b_t| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{6} = \frac{2\varepsilon}{3} \ .$$

Hence by the triangle inequality,

$$|\mathrm{Tr}(E_t\omega_{t-1}) - \mathrm{Tr}(E_t\rho)| \leq \frac{2\varepsilon}{3} + |b_t - \mathrm{Tr}(E_t\rho)| \leq \varepsilon \ ,$$

as claimed. □

E Proof of Corollary 7

We begin with a bound for a generalization of "random access coding" (Nayak [1999], Ambainis et al. [2002]) or what is also known as the Index function problem in communication complexity. The generalization was called "serial encoding" by Nayak [1999] and arose in the context of quantum finite automata. The serial encoding problem is also called Augmented Index in the literature on streaming algorithms.

The following theorem places a bound on how few qubits serial encoding may use. In other words, it bounds the number of bits we may encode in an n-qubit quantum state when an arbitrary bit out of the n may be recovered well via a two-outcome measurement. The bound holds even when the measurement for recovering y_i may depend adaptively on the previous bits $y_1y_2\cdots y_{i-1}$ of y, *which we need not know*.

Theorem 15 (Nayak [1999])**.** *Let k and n be positive integers. For each k-bit string $y := y_1\cdots y_k$, let ρ_y be an n-qubit mixed state such that for each $i \in \{1, 2, \ldots, k\}$, there is a two-outcome measurement E that depends only on i and the prefix $y_1y_2\cdots y_{i-1}$, and has the following properties*

(i) if $y_i = 0$ then $\mathrm{Tr}(E\rho_y) \leq p$, and

(ii) if $y_i = 1$ then $\mathrm{Tr}(E\rho_y) \geq 1 - p$,

where $p \in [0, 1/2]$ is the error in predicting the bit y_i at vertex v. (We say ρ_y "serially encodes" y.) Then $n \geq (1 - \mathrm{H}(p))k$.

In Appendix F, we present a strengthening of this bound when the bits of y may be only be recovered in an adaptive order that is *a priori* unknown. The stronger bound may be of independent interest.

In the context of online learning, the measurements used in recovering bits from a serial encoding are required to predict the bits with probability bounded away from given "pivot points". Theorem 15 may be specialized to this case as in Corollary 7, which we prove below.

Proof of Corollary 7. This is a consequence of Theorem 15, when combined with the following observation. Given the measurement operator E', parameter ε, and pivot point a_v as in the statement of the corollary, we define a new two-outcome measurement E to be associated with vertex v:

$$E \quad := \quad \begin{cases} \frac{E'}{2a_v} & \text{if } a_v \geq \frac{1}{2} \ , \qquad \text{and} \\ \frac{1}{2(1-a_v)} \left(E' + (1 - 2a_v)\mathbb{I}\right) & \text{if } a_v < \frac{1}{2} \ . \end{cases}$$

The measurement E may be interpreted as producing a fixed outcome 0 or 1 with some probability depending on a_v, and applying the given measurement E' with the remaining probability, so as to translate the pivot point a_v to $1/2$.

We may verify that the operator E satisfies the requirements (i) and (ii) of Theorem 15 with $p := (1 - \varepsilon)/2$. We therefore conclude that $n \geq (1 - \mathrm{H}((1 - \varepsilon)/2)k$. Since $\mathrm{H}(1/2 - \delta) \leq 1 - 2\delta^2$, for $\delta \in [0, 1/2]$, we get $k = \mathrm{O}\big(n/\varepsilon^2\big)$. □

F Lower bound on quantum random access codes

Here we present an alternative proof of the linear lower bound on quantum random access codes Nayak [1999], Ambainis et al. [2002]. It goes via the Matrix Multiplicative Weights algorithm, but gives us a slightly weaker dependence on decoding error. We also present an extension of the original bound to more general codes. These may be of independent interest.

Theorem 16. *Let k and n be positive integers with $k > n$. For all k-bit strings $y = y_1, y_2, \ldots, y_k$, let ρ_y be the n-qubit quantum mixed state that encodes y. Let $p \in [0, 1/2]$ be an error tolerance parameter. Suppose that there exist measurements $E_1, E_2, \ldots, E_k$ such that for all $y \in \{0, 1\}^k$ and all $i \in [k]$, we have $|\operatorname{Tr}(E_i \rho_y) - y_i| \leq p$. Then $n \geq \frac{(1/2-p)^2}{4(\log 2)} k$.*

Proof. Run the MMW algorithm described in Section 3.2 with the absolute loss function $\ell_t(x) := |x - y_t|$ for $t = 1, 2, \ldots, k$ iterations. In iteration t, provide as feedback E_t and the label $y_t \in \{0, 1\}$ defined as follows:

$$y_t = \begin{cases} 0 & \text{if } \operatorname{Tr}(E_t \omega_t) > \frac{1}{2} \\ 1 & \text{if } \operatorname{Tr}(E_t \omega_t) \leq \frac{1}{2} \ . \end{cases}$$

Let $y \in \{0, 1\}^k$ be the bit string formed at the end of the process. Then it is easy to check the following two properties by the construction of the labels: for any $t \in [k]$, we have

1. $\ell_t(\omega_t) = |\operatorname{Tr}(E_t \omega_t) - y_t| \geq 1/2$, and
2. $\ell_t(\rho_y)) = |\operatorname{Tr}(E_t \rho_y) - y_t| \leq p$.

By Theorem 4, the MMW algorithm with absolute loss has a regret bound of $2\sqrt{(\log 2)kn}$. So the above bounds imply that $k/2 \leq pk + 2\sqrt{(\log 2)kn}$, which implies that $n \geq \frac{(1/2-p)^2}{4 \log 2} k$. □

Note that in the above proof, we may allow the measurement in the ith iteration, i.e., the one used to decode the ith bit, to depend on the previous bits $y_1, y_2, \ldots, y_{i-1}$. Thus, the lower bound also applies to serial encoding.

Next we consider encoding of bit-strings y into quantum states ρ_y with a more relaxed notion of decoding. The encoding is such that given the encoding for an unknown string y, *some* bit i_1 of y

can be decoded. Given the value y_{i_1} of of this bit, a new bit i_2 of y can be decoded, and the index i_2 may depend on y_{i_1}. More generally, given a sequence of bits $y_{i_1}y_{i_2}\ldots y_{i_j}$ that may be decoded in this manner, a new bit i_{j+1} of y can be decoded, for any $j \in \{0, 1, \ldots, k-1\}$. Here, the index i_{j+1} and the measurement used to recover the corresponding bit of y may depend on the sequence of bits $y_{i_1}y_{i_2}\ldots y_{i_j}$. We show that *even with this relaxed notion of decoding*, we cannot encode more than a linear number of bits into an n-qubit state.

We first formalize the above generalization of random access encoding. We view a complete binary tree of depth $d \geq 0$ as consisting of vertices $v \in \{0,1\}^{\leq d}$. The root of the tree is labeled by the empty string ϵ and each internal vertex v of the tree has two children $v0, v1$. We specify an adaptive sequence of measurements through a "measurement decision tree". The tree specifies the measurement to be applied next, given a prefix of such measurements along with the corresponding outcomes.

Definition 1. *Let k be a positive integer. A* measurement decision tree *of depth k is a complete binary tree of depth k, each internal vertex v of which is labeled by a triple (S, i, E), where $S \in \{1, \ldots, k\}^l$ is a sequence of length $l := |v|$ of distinct indices, $i \in \{1, \ldots, k\}$ is an index that does not occur in S, and E is a two-outcome measurement. The sequences associated with the children $v0, v1$ of v (if defined) are both equal to (S, i).*

For a k-bit string y, and sequence $S := (i_1, i_2, \ldots, i_l)$ with $0 \leq l \leq k$ and $i_j \in \{1, 2, \ldots, k\}$, let y_S denote the substring $y_{i_1}y_{i_2}\cdots y_{i_l}$.

Theorem 17. *Let k and n be positive integers. For each k-bit string $y := y_1 \cdots y_k$, let ρ_y be an n-qubit mixed state (we say ρ_y "encodes" y). Suppose there exists a measurement decision tree T of depth k such that for each internal vertex v of T and all $y \in \{0,1\}^k$ with $y_S = v$, where (S, i, E) is the triple associated with the vertex v, we have $|\mathrm{Tr}(E\rho_y) - y_i| \leq p_v$, where $p_v \in [0, 1/2]$ is the error in predicting the bit y_i at vertex v. Then $n \geq (1 - \mathrm{H}(p))k$, where H is the binary entropy function, and $p := \frac{1}{k}\sum_{l=1}^{k}\frac{1}{2^l}\sum_{v\in\{0,1\}^l} p_v$ is the average error.*

Proof. Let Y be a uniformly random k-bit string. We define a random permutation Π of $\{1, \ldots, k\}$ correlated with Y that is given by the sequence of measurements in the root to leaf path corresponding to Y. More formally, let $\Pi(1) := i$, where i is the index associated with the root of the measurement decision tree T. For $l \in \{2, \ldots, k\}$, let $\Pi(l) := j$, where j is the index associated with the vertex $Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}$ of the tree T. Let Q be a quantum register such that the joint state of YQ is

$$\frac{1}{2^k}\sum_{y\in\{0,1\}^k} |y\rangle\langle y| \otimes \rho_y \ .$$

The quantum mutual information between Y and Q is bounded as $\mathrm{I}(Y : Q) \leq |Q| = n$. Imagine having performed the first $l-1$ measurements given by the tree T on state Q and having obtained the correct outcomes $Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}$. These outcomes determine the index $\Pi(l)$ of the next bit that may be learned. By the Chain Rule, for any $l \in \{1, \ldots, k-1\}$,

$$\begin{aligned}&\mathrm{I}\big(Y_{\Pi(l)}\cdots Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}\big)\\ &\quad= \mathrm{I}\big(Y_{\Pi(l)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}\big) + \mathrm{I}\big(Y_{\Pi(l+1)}\cdots Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l)}\big)\ .\end{aligned}$$

Let E be the operator associated with the vertex $V := Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}$. By hypothesis, the measurement E predicts the bit $Y_{\Pi(l)}$ with error at most p_V. Using the Fano Inequality, and averaging over the prefix V, we get

$$\mathrm{I}\big(Y_{\Pi(l)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(l-1)}\big) \geq \mathbb{E}_V(1 - \mathrm{H}(p_V))\ .$$

Applying this repeatedly for $l \in \{1, \ldots, k-1\}$, we get

$$\begin{aligned}\mathrm{I}(Y : Q) &= \mathrm{I}\big(Y_{\Pi(1)} : Q\big) + \mathrm{I}\big(Y_{\Pi(2)} : Q \mid Y_{\Pi(1)}\big) + \mathrm{I}\big(Y_{\Pi(3)} : Q|Y_{\Pi(1)}Y_{\Pi(2)}\big)\\ &\qquad + \cdots + \mathrm{I}\big(Y_{\Pi(k)} : Q \mid Y_{\Pi(1)}Y_{\Pi(2)}\cdots Y_{\Pi(k-1)}\big)\\ &\geq \sum_{l=1}^{k}\frac{1}{2^l}\sum_{v\in\{0,1\}^l}(1 - \mathrm{H}(p_v))\\ &\geq (1 - \mathrm{H}(p))k\ ,\end{aligned}$$

by concavity of the binary entropy function, and the definition of p. □