Entanglement is Necessary for Optimal Quantum Property Testing

Sebastien Bubeck
Microsoft Research
Redmond, WA
sebubeck@microsoft.com

Sitan Chen
MIT
Cambridge, MA
sitanc@mit.edu

Jerry Li
Microsoft Research
Redmond, WA
jerrl@microsoft.com

***Abstract*—There has been a surge of progress in recent years in developing algorithms for testing and learning quantum states that achieve optimal copy complexity [1], [2], [3], [4], [5], [6]. Unfortunately, they require the use of entangled measurements across many copies of the underlying state and thus remain outside the realm of what is currently experimentally feasible. A natural question is whether one can match the copy complexity of such algorithms using only independent—but possibly adaptively chosen—measurements on individual copies.**

We answer this in the negative for arguably the most basic quantum testing problem: deciding whether a given d-dimensional quantum state is equal to or ϵ-far in trace distance from the maximally mixed state. While it is known how to achieve optimal $O(d/\epsilon^2)$ copy complexity using entangled measurements, we show that with independent measurements, $\Omega(d^{4/3}/\epsilon^2)$ is necessary, even if the measurements are chosen adaptively. This resolves a question posed in [7]. To obtain this lower bound, we develop several new techniques, including a chain-rule style proof of Paninski's lower bound for classical uniformity testing, which may be of independent interest.

***Index Terms*—quantum property testing; quantum tomography; distribution testing**

I. Introduction

This paper considers the problem of *quantum state certification*. Here, we are given N copies of an unknown mixed state $\rho \in \mathbb{C}^{d\times d}$ and a description of a known mixed state σ, and our goal is to make measurements on these copies[1] and use the outcomes of these measurements to distinguish whether $\rho = \sigma$, or if it is ϵ-far from σ in trace norm. An important special case of this is when σ is the maximally mixed state, in which case the problem is known as *quantum mixedness testing*.

This problem is motivated by the need to verify the output of quantum computations. In many applications, a quantum algorithm is designed to prepare some known d-dimensional mixed state σ. However, due to the possibility of noise or device defects, it is unclear whether or not the output state is truly equal to σ. Quantum state certification allows us to verify the correctness of the quantum algorithm. In addition to this more practical motivation, quantum state certification can be seen as the natural non-commutative analogue of *identity testing* of (classical) probability distributions, a well-studied problem in statistics and theoretical computer science.

Recently, [1] demonstrated that $\Theta(d/\epsilon^2)$ copies are necessary and sufficient to solve quantum mixedness testing with good confidence. Subsequently, [6] demonstrated that the same copy complexity suffices for quantum state certification. Note that these copy complexities are sublinear in the number of parameters in ρ, and in particular, are less than the $\Theta(d^2/\epsilon^2)$ copies necessary to learn ρ to ϵ error in trace norm [2], [3].

To achieve these copy complexities, the algorithms in [1], [6] heavily rely on entangled measurements. These powerful measurements allow them to leverage the representation theoretic structure of the underlying problem to dramatically decrease the copy complexity. However, this power comes with some tradeoffs. Entangled measurements require that all N copies of ρ are measured simultaneously. Thus, all N copies of ρ must be kept in quantum memory without any of them de-cohering. Additionally, the positive-operator valued measure (POVM) elements that formally define the quantum measurement must all be of size $d^N \times d^N$; in particular, the size of the POVM elements scales exponentially with N. Both of these issues are problematic for using any of these algorithms in practice [8]. Entangled measurements are also necessary for the only known sample-optimal algorithms for quantum tomography [2], [3], [4].

This leads to the question: can these sample complexities be achieved using weaker forms of measurement? There are two natural classes of such restricted measurements to consider:

- an (unentangled) *nonadaptive measurement* fixes N POVMs ahead of time, measures each copy of ρ using one of these POVMs, then uses the results to make its decision.
- an (unentangled) *adaptive measurement* measures each copy of ρ sequentially, and can potentially choose its next POVM based on the results of the outcomes of the previous experiments.

It is clear that arbitrarily entangled measurements are strictly more general than adaptive measurements, which are in turn strictly more general than nonadaptive ones. However, both nonadaptive and adaptive measurements have the advantage that the quantum memory they require is substantially smaller than what is required for a generic entangled measurement. In particular, only one copy of ρ need be prepared at any given

[1]Formally, a measurement is specified by a *positive operator-valued measure (POVM)*, which is given by a set of positive-definite Hermitian matrices $\{M_x\}$ summing to the identity, and the probability of observing measurement outcome x is equal to $\mathrm{Tr}(\rho M_x)$. See Definition IV.1 for details.

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00070

time, as opposed to the N copies that must simultaneously be created, if we use general entangled measurements.

Separating the power of entangled vs. nonentangled measurements for such quantum learning and testing tasks was posed as an open problem in [7]. In this paper, we demonstrate the first such separations for quantum state certification, and to our knowledge, the first separation between adaptive measurements and entangled measurements without any additional assumptions on the measurements, for any quantum estimation task.

We first show a sharp characterization of the copy complexity of quantum mixedness testing with nonadaptive measurements:

Theorem I.1. *If only unentangled, nonadaptive measurements are used, $\Theta(d^{3/2}/\epsilon^2)$ copies are necessary and sufficient to distinguish whether $\rho \in \mathbb{C}^{d\times d}$ is the maximally mixed state, or if ρ has trace distance at least ϵ from the maximally mixed state, with probability at least $2/3$.*

Second, we show that $\omega(d)$ copies are necessary, even with adaptive measurements. We view this as our main technical contribution. Formally:

Theorem I.2. *If only unentangled, possibly adaptive, measurements are used, $\Omega(d^{4/3}/\epsilon^2)$ copies are necesssary to distinguish whether $\rho \in \mathbb{C}^{d\times d}$ is the maximally mixed state, or has trace distance at least ϵ from the maximally mixed state, with probability at least $2/3$.*

As quantum state certification is a strict generalization of mixedness testing, Theorems I.1 and I.2 also immediately imply separations for that problem as well. Note that the constant $2/3$ in the above theorem statements is arbitrary and can be replaced with any constant greater than $1/2$. We also remark that our lower bounds make no assumptions on the number of outcomes of the POVMs used, which can be infinite (see Definition IV.1).

A. Overview of our techniques

In this section, we give a high-level description of our techniques. We start with the lower bounds.

a) "Lifting" classical lower bounds to quantum ones: Our lower bound instance can be thought of as the natural quantum analogue of Paninski's for (classical) uniformity testing:

Theorem I.3 (Theorem 4, [9])**.** *$\Omega(\sqrt{d}/\epsilon^2)$ samples are necessary to distinguish whether a distribution p over $\{1,\dots,d\}$ is ϵ-far from the uniform distribution in total variation distance, with confidence at least $2/3$.*

At a high level, Paninski demonstrates that it is statistically impossible to distinguish between the distribution $p_0^{\le N}$ of N independent draws from the uniform distribution, and the distribution $p_1^{\le N}$ of N independent draws from a random perturbation of the uniform distribution, where the marginal probability of each element in $\{1,\dots,d\}$ has been randomly perturbed by $\pm\epsilon/d$ (see Example III.7).

The hard instance we consider can be viewed as the natural quantum analogue of Paninski's construction. Roughly speaking, rather than simply perturbing the marginal probabilities of every element in $\{1,\dots,d\}$, which corresponds to randomly perturbing the diagonal elements of the mixed state, we also randomly rotate it (see Construction 1). We note that this hard instance is not novel and has been considered before in similar settings [1], [7], [3]. However, our analysis technique is quite different from previous bounds, especially in the adaptive setting.

The technical crux of Paninski's lower bound is to upper bound the total variation distance between $p_0^{\le N}$ and $p_1^{\le N}$ in terms of the χ^2-divergence between the two. This turns out to have a simple, explicit form, and can be calculated exactly. This works well because, conditioned on the choice of the random perturbation in $p_1^{\le N}$, both of the distributions $p_0^{\le N}$ and $p_1^{\le N}$ have a product structure, as they consist of N independent samples.

This product structure still holds true in the quantum case when we restrict to non-adaptive measurements. This allows us to do a more involved version of Paninski's calculation in the quantum case and thus obtain the lower bound in Theorem I.1.

However, this product structure breaks down completely in the adaptive setting, as now the POVMs, and hence, the measurement outcomes that we observe, for the t-th copy of ρ, can depend heavily on the previous outcomes. As a result, the χ^2-divergence between the analogous quantities to $p_0^{\le N}$ and $p_1^{\le N}$ no longer have a nice, closed form, and it is not clear how to proceed using Paninski's style of argument.

Instead, inspired by the literature on bandit lower bounds [10], [11], we upper bound the total variation distance between $p_0^{\le N}$ and $p_1^{\le N}$ by the KL divergence between these two quantities. The primary advantage of doing so is that the KL divergence satisfies the chain rule. This allows us to partially disentangle how much information that the t-th copy of ρ gives the algorithm, conditioned on the outcomes of the previous experiments.

At present, this chain-rule formulation of Paninski's lower bound seems to be somewhat lossy. Even in the classical case, we need additional calculations tailored to Paninski's instance to recover the $\Omega(\sqrt{d}/\epsilon^2)$ bound for uniformity testing (see the appendix of the full version), without which our approach can only obtain a lower bound of $\Omega(d^{1/3}/\epsilon^2)$ (see Section VI). At a high level, this appears to be why we do not obtain a lower bound of $\Omega(d^{3/2}/\epsilon^2)$ for adaptive measurements. We leave the question of closing this gap as an interesting future direction.

b) "Projecting" quantum upper bounds to classical ones: While the lower bound techniques we employ are motivated by the lower bounds for classical testing, they do not directly use any of those results. In contrast, to obtain our upper bounds, we demonstrate a direct reduction from non-adaptive mixedness testing to classical uniformity testing. The reduction is as follows. First, we choose a random orthogonal measurement basis. Measuring ρ in this basis induces some distribution over $\{1,\dots,d\}$. If ρ is maximally mixed, this distribution is the uniform distribution. Otherwise, if it is far from maximally

mixed, then by similar concentration of measure phenomena as used in the proof of the lower bounds, with high probability this distribution will be quite far from the uniform distribution in L_2 distance. Thus, to distinguish these two cases, we can simply run a classical L_2 uniformity tester [12], [13], [14]. See Appendix II for more details.

c) Concentration of measure over the unitary group: In both our lower bounds and upper bounds, it will crucial to carefully control the deviations of various functions of Haar random unitary matrices. In fact, specializations of quantities we encounter have been extensively studied in the literature on quantum transport in mesoscopic systems, namely the conductance of a chaotic cavity [15], [16], [17], [18], [19], though the tail bounds we need are not captured by these works (see Section IV-C for more details). Instead, we will rely on more general tail bounds [20] that follow from log-Sobolev inequalities on the unitary group $U(d)$.

B. Related Work

The literature on quantum (and classical) testing and learning is vast and we cannot hope to do it justice here; for conciseness we only discuss some of the more relevant works below.

Quantum state certification fits into the general framework of quantum state property testing problems. Here the goal is to infer non-trivial properties of the unknown quantum state, using fewer copies than are necessary to fully learn the state. See [21] for a more complete survey on property testing of quantum states. Broadly speaking, there are two regimes studied here: the asymptotic regime and the non-asymptotic regime.

In the asymptotic regime, the goal is to precisely characterize the exponential convergence of the error as $n \to \infty$ and d, ϵ are held fixed and relatively small. In this setting, quantum state certification is commonly referred to as *quantum state discrimination*. See e.g. [22], [23], [24] and references within. However, this allows for rates which could depend arbitrarily badly on the dimension.

In contrast, we work in the non-asymptotic regime, where the goal is to precisely characterize the rate of convergence as a function of d and ϵ. The closest work to ours is arguably [1] and [6]. The former demonstrated that the copy complexity of quantum mixedness testing is $\Theta(d/\epsilon^2)$, and the latter showed that quantum state certification has the same copy complexity. However, as described previously, the algorithms which achieve these copy complexities heavily rely on entangled measurements.

Another interesting line of work focuses on the case where the measurements are only allowed to be Pauli matrices [25], [26], [27], [28]. Unfortunately, even for pure states, these algorithms require $\Omega(d)$ copies of ρ. We note in particular the paper of [26], which gives a $\Omega(d)$ lower bound for the copy complexity of the problem, even when the Pauli measurements are allowed to be adaptively chosen. However, their techniques do not appear to generalize easily to arbitrary adaptive measurements.

We also mention [29] which gives algorithms for various quantum property testing problems using *local measurements* which non-adaptively operate on each individual qubit. Because this is a more restrictive family of measurements, the sample complexity for these algorithms suffers some polynomial overhead as a function of d.

A related task is that of quantum tomography, where the goal is to recover ρ, typically to good fidelity or low trace norm error. The paper [3] showed that $O(d^2 \log(d/\epsilon)/\epsilon^2)$ copies suffice to obtain ϵ trace error, and that $\Omega(d^2/\epsilon^2)$ copies are necessary. Independently, [2] improved their upper bound to $O(d^2/\epsilon^2)$. These papers, in addition to [4], also discuss the case when ρ is low rank, where $o(d^2)$ copy complexity can be achieved. Notably, all the upper bounds that achieve the tight bound heavily require entanglement. In [3], they demonstrate that $\Omega(d^3/\epsilon^2)$ copies are necessary, if the measurements are nonadaptive. It is a very interesting question to understand the power of adaptive measurements for this problem as well.

Quantum state certification and quantum mixedness testing are the natural quantum analogues of classical identity testing and uniformity testing, respectively, which both fit into the general setting of (classical) distribution testing. There is again a vast literature on this topic; see e.g. [30], [31] for a more extensive treatment of the topic. Besides the papers covered previously and in the surveys, we highlight a line of work on testing with *conditional sampling oracles* [32], [33], [34], [35], [36], [37], a classical model of sampling which also allows for adaptive queries. It would be interesting to see if the techniques we develop here can also be used to obtain stronger lower bounds in this setting.

C. Miscellaneous Notation

We gather here useful notation for the rest of the paper. Let $[d]$ denote the set $\{1, \dots, d\}$. Given a finite set S, we will use $x \sim_u S$ to denote x sampled uniformly at random from S. Given two strings s and t, let $s \circ t$ denote their concatenation. Given $t > 1$ and a sequence $x_1, ..., x_{t-1}$, define $x_{<t} \triangleq (x_1, ..., x_{t-1})$. We will also sometimes refer to this as $x_{\leq t-1}$. Also, let $x_{<1} \triangleq \emptyset$.

Given distributions P, Q, the total variation distance between P and Q is $d_{\mathrm{TV}}(P, Q) \triangleq \frac{1}{2}\|P - Q\|_1$. If P is absolutely continuous with respect to Q, let $\frac{\mathrm{d}P}{\mathrm{d}Q}(\cdot)$ denote the Radon-Nikodym derivative. The KL-divergence between P and Q is $\mathrm{KL}(P\|Q) \triangleq \mathbb{E}_{x\sim Q}[\frac{\mathrm{d}P}{\mathrm{d}Q}(x) \log \frac{\mathrm{d}P}{\mathrm{d}Q}(x)]$. The chi-squared divergence between P and Q is $\chi^2(P\|Q) \triangleq \mathbb{E}_{x\sim Q}[\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(x) - 1\right)^2]$.

Let $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_{HS}$ denote trace, operator, and Hilbert-Schmidt norms respectively. Let $\rho_{\mathsf{mm}} \triangleq \frac{1}{d}\mathbf{I}$ denote the maximally mixed state. Given a matrix M, let $\widehat{M} \triangleq M/\operatorname{Tr}(M)$. Given $\mathbf{A} \in \mathbb{C}^{d\times d}$ and $\pi \in \mathcal{S}_n$ with cycle decomposition $(C_1, ..., C_m)$, let $\langle A \rangle_\pi \triangleq \prod_{i=1}^m \operatorname{Tr}(A^{|C_i|})$.

We will use the following extensively:

Fact I.4 (Integration by parts)**.** *Let $a, b \in \mathbb{R}$. Let Z be a nonnegative random variable satisfying $Z \leq b$ and such that*

for all $x \geq a$, $\mathbb{P}[Z > x] \leq \tau(x)$. Let $f : [0, b] \to \mathbb{R}_{\geq 0}$ be nondecreasing and differentiable. Then

$$\mathbb{E}[f(Z)] \leq f(a)(1 + \tau(a)) + \int_a^b \tau(x) f'(x) \, \mathrm{d}\, x.$$

Finally, throughout this work, we will freely abuse notation and use the same symbols to denote probability distributions, their laws, and their density functions.

II. The Upper Bound

In this section we prove the upper bound of Theorem I.1:

Theorem II.1. *Given $0 < \epsilon < 1$ and $N = O(d^{3/2}/\epsilon^2)$ copies of ρ, there is an algorithm* TestMixed*(ρ, d, ϵ) that makes unentangled measurements and with probability $4/5$ distinguishes whether $\|\rho - \rho_{\mathsf{mm}}\|_1 \geq \epsilon$ or $\rho = \rho_{\mathsf{mm}}$.*

Our mixedness tester is extremely simple: pick a random orthogonal POVM corresponding to a Haar-random basis of $\mathbb{C}^d$, measure $O(d^{3/2}/\epsilon^2)$ copies of ρ with this POVM, and use these measurement outcomes to check whether the distribution over measurement outcomes is too far (in L_2 distance) from uniform, in which case ρ is far from ρ_{mm}. We will need the following result on classical uniformity testing in L_2.

Theorem II.2 ([12], [13], [14])**.** *Given $0 < \epsilon < 1$ and sample access to a distribution q over $[d]$, there is an algorithm* TestUniformityL2*(q, d, ϵ) that uses $N = O(\sqrt{d}/\epsilon^2)$ samples from q and with probability $9/10$ distinguishes whether q is the uniform distribution over $[d]$ or $\epsilon/\sqrt{d}$-far in L_2 distance from the uniform distribution.*

Certainly when $\rho = \rho_{\mathsf{mm}}$, for any orthogonal POVM corresponding to an orthonormal basis of $\mathbb{C}^d$, the induced distribution over measurement outcomes will be the uniform distribution over d elements. The point is that when ρ is ϵ-far from maximally mixed, for a Haar-random orthogonal POVM, the induced distribution over measurement outcomes will be $O(\epsilon/d)$-far in L_2 distance from the uniform distribution over d elements with high probability. So Theorem II.2 would imply an algorithm for testing mixedness which makes $O(d^{3/2}/\epsilon^2)$ unentangled, nonadaptive measurements. Formally, our algorithm is specified in Algorithm 1 below.

Algorithm 1 TestMixed

1: **Input**: $N \triangleq \Theta(d^{3/2}/\epsilon^2)$ copies of unknown state ρ
2: **Output**: NO if $\|\rho - \frac{1}{d} \cdot \mathbf{I}\|_1 \geq \epsilon$, YES if $\rho = \frac{1}{d} \cdot \mathbf{I}$
3: Sample a Haar-random unitary matrix $U \in \mathbb{C}^{d \times d}$.
4: Define the POVM $\{|U_i\rangle \langle U_i|\}_{1 \leq i \leq d}$ and measure with this POVM N times to get N independent samples from the distribution q over measurement outcomes.
5: Run TestUniformityL2$(q, d, \epsilon/\sqrt{d})$ from Theorem II.2
6: If q far from uniform, output NO, else YES.

Proof. As the POVM defined in Step 4 of TestMixed is Haar-random, we may assume $\rho = \mathbf{\Lambda}$ without loss of generality, where $\mathbf{\Lambda}$ is the diagonal matrix whose first $d/2$ diagonal entries are $1/d + \epsilon/d$ and whose last $d/2$ diagonal entries are $1/d - \epsilon/d$. Also define the diagonal matrix $\mathbf{X}'$ whose first $d/2$ diagonal entries are 1 and whose last $d/2$ diagonal entries are -1.

Let q be the distribution over measurement outcomes, and let u be the uniform distribution over $[d]$. Note that for any $i \in [d]$, the marginal probability $q_i = (\mathbf{U}^\dagger \mathbf{\Lambda} \mathbf{U})_{ii}$, so $\|q - u\|_2^2 = -1/d + \sum q_i^2$. By a standard Weingarten calculation (see the full version), $\mathbb{E}[\sum_i q_i^2] = \frac{1}{d+1}(1 + \mathrm{Tr}(\rho^2)) \leq \frac{\epsilon^2}{d(d+1)}$, from which the theorem follows by Markov's and the guarantees of TestUniformityL2 in Theorem II.2. □

III. Lower Bound Strategies

The lower bounds we show in this work are lower bounds on the number of observations needed to distinguish between a simple null hypothesis and a mixture of alternatives. For instance, in the context of classical uniformity testing, the null hypothesis is that the underlying distribution is the uniform distribution over $[d]$, and the mixture of alternatives considered in [9] is that the underlying distribution was drawn from a particular *distribution over distributions* p which are ϵ-far in total variation distance from the uniform distribution (see Example III.7). In our setting, the null hypothesis is that the underlying state is the maximally mixed state ρ_{mm}, and the mixture of alternatives will be a particular *distribution over quantum states* ρ which are ϵ-far in trace distance from ρ_{mm} (see Construction 1).

Note that in order to obtain dimension-dependent lower bounds, as in classical uniformity testing, it is essential that the alternative hypothesis be a mixture. If the task were instead to distinguish whether the underlying state was ρ_{mm} or some *specific* alternative state ρ, then if we make independent measurements in the eigenbasis of ρ, it takes only $O(1/\epsilon^2)$ such measurements to tell apart the two scenarios.

For this reason we will be interested in the following abstraction which contains as special cases both Paninski's lower bound instance for uniformity testing [9] and our lower bound instance for mixedness testing, and which itself is a special case of Le Cam's two-point method [38]. We will do this in a few steps. First, we give a general formalism for what it means to perform possibly adaptive measurements:

Definition III.1 (Adaptive measurements)**.** *Given an underlying space $\mathcal{S}$, a natural number $N \in \mathbb{N}$, and a (possibly infinite) universe $\mathcal{U}$ of measurement outcomes, a* measurement schedule *A using N measurements is any (potentially random) algorithm which outputs $M_1, \ldots, M_N : \mathcal{S} \to \mathcal{U}$, where each M_i is a potentially random function. We say that A is* nonadaptive *if the choice of M_i is independent of the choice of M_j for all $j \neq i$, and we say A is* adaptive *if the choice of M_t depends only on the outcomes of $M_1, \ldots, M_{t-1}$ for all $t \in [N]$.*

To instantiate this for the quantum setting, we let the underlying space $\mathcal{S}$ be the set of mixed states, and we restrict the measurement functions to be (possibly adaptively chosen) POVMs. See Definition IV.2 for a formal definition.

Definition III.2. *A* distribution testing task *is specified by two disjoint sets $\mathcal{S}_0, \mathcal{S}_1$ in $\mathcal{S}$. For any $N \in \mathbb{N}$, and any measurement schedule A, we say that A solves the problem if there exists a (potentially random) post-processing algorithm $f : \mathcal{U}^N \to \{0,1\}$ so that for any $\alpha \in \{0,1\}$, if $D \in \mathcal{S}_\alpha$, then*

$$\mathbb{P}[f\left(M_1(D) \circ \cdots \circ M_N(D)\right) = \alpha] \geq 2/3\,,$$

where $M_1, \ldots, M_N$ are generated by A.

For instance, to instantiate the quantum mixedness testing setting, we let $\mathcal{S}$ be the set of mixed states, we let $\mathcal{S}_0 = \{\rho_{\mathsf{mm}}\}$ be the set containing only ρ_{mm}, the maximally mixed state, and we let $\mathcal{S}_1 = \{\rho : \|\rho - \rho_{\mathsf{mm}}\|_1 > \epsilon\}$. Note that the choice of $2/3$ for the constant is arbitrary and can be replaced (up to constant factors in N) with any constant strictly larger than $1/2$. With this, we can now define our lower bound setup:

Definition III.3 (Lower Bound Setup: Simple Null vs. Mixture of Alternatives)**.** *In the setting of Definition III.2, a* distinguishing task *is specified by a null object $D_0 \in \mathcal{S}_0$, a set of alternate objects $\{D_\zeta\} \subseteq \mathcal{S}_1$ parametrized by ζ, and a distribution $\mathcal{D}$ over ζ.*

For any measurement schedule A which generates measurement functions $M_1, \ldots, M_N$, let $p_0^{\leq N} = p_0^{\leq N}(A)$ and $p_1^{\leq N} = p_1^{\leq N}(A)$ be distributions over strings $x_{\leq N} \in \mathcal{U}^N$, which we call transcripts of length N*. The distribution $p_0^{\leq N}$ corresponds to the distribution of $M_1(D_0) \circ \cdots \circ M_N(D_0)$. The distribution $p_1^{\leq N}$ corresponds to the distribution of of $M_1(D_\zeta) \circ \cdots \circ M_N(D_\zeta)$, where $\zeta \sim \mathcal{D}$.*

The following is a standard result which allows us to relate this back to property testing:

Fact III.4. *Let $\mathcal{S}_0, \mathcal{S}_1$ be a property, let $N \in \mathbb{N}$, and let $\mathcal{A}$ be a class of measurement schedules using N measurements. Suppose that there exists a distinguishing task so that for every $A \in \mathcal{A}$, we have that $d_{\mathrm{TV}}(p_0^{\leq N}(A), p_1^{\leq N}(A)) \leq 1/3$. Then the distribution testing task cannot be solved with N samples by any algorithm in $\mathcal{A}$.*

For the remainder of the paper, we will usually implicitly fix a measurement schedule A, and just write $p_0^{\leq N}$ and $p_1^{\leq N}$. The properties that we assume (e.g. adaptive or nonadaptive) of this algorithm should be clear from context, if it is relevant.

We next define some important quantities which repeatedly arise in our calculations:

Definition III.5. *In the setting of Definition III.3, for any $t \in [N]$, define $p_0^t(\cdot|x_{<t}), p_1^t(\cdot|x_{<t})$ to be the respective conditional laws of the t-th entry, given preceding transcript $x_{<t}$. For any ζ, let $p_1^{\leq N}|\zeta$ be the distribution over transcripts from N independent observations from D_ζ.*

Assume additionally that $p_1^{\leq N}|\zeta$ are absolutely continuous with respect to $p_0^{\leq N}$, for every $\zeta \in \mathrm{supp}(\mathcal{D})$. Then, there will exist functions $\{g^\zeta_{x_{<t}}(\cdot)\}_{t \in [N], x_{<t} \in \mathcal{U}^{t-1}, \zeta \in \mathrm{supp}(\mathcal{D})}$, such that for any $\zeta, t, x_{\leq t}$, the Radon-Nikodym derivative satisfies

$$\frac{\mathrm{d}p_1^{\leq t}|\zeta}{\mathrm{d}p_0^{\leq t}}(x_{\leq t}) = \prod_{i=1}^{t}\left(1 + g^\zeta_{x_{<i}}(x_i)\right). \tag{1}$$

We refer to the $g^\zeta_{x_{<t}}(\cdot)$ functions as likelihood ratio factors*.*

We emphasize that neither $p_0^{\leq N}$ nor any of the alternatives $p_1^{\leq N}|\zeta$ is necessarily a product measure. Indeed, this is one of the crucial difficulties of proving lower bounds in the adaptive setting. In the non-adaptive setting, the picture of Definition III.3 simplifies substantially:

Definition III.6 (Non-adaptive Testing Lower Bound Setup)**.** *In this case, in the notation of Definition III.3, the measurement schedule A is nonadaptive, so $p_0^{\leq N}$ and all $p_1^{\leq N}|\zeta$ are product measures. Consequently, the functions $g^\zeta_{x_{<t}}$ will depend only on t and not on the particular transcript $x_{<t}$, so we will denote the functions by $\{g_t^\zeta(\cdot)\}_{t \in [N], \zeta \in \mathrm{supp}(\mathcal{D})}$.*

Paninski's lower bound for classical uniformity testing [9] is an instance of the non-adaptive setup of Definition III.6:

Example III.7. *Let us first recall Paninski's construction. Here the set $\mathcal{S}$ is the set of distributions over $[d]$. Uniformity testing is the property $S_0 = \{U\}, S_1 = \{U' : d_{\mathrm{TV}}(U, U') \geq \epsilon\}$, where U is the uniform distribution over $[d]$. In the classical "sampling oracle" model of distribution testing, the measurements M_i simply take a distribution $D \in \mathcal{S}$ and output an independent sample from D. In particular, $\mathcal{U} = [d]$.*

To form Paninski's lower bound instance, take $\mathcal{D}$ to be the uniform distribution over $\{\pm 1\}^{d/2}$. Let the null hypothesis be D_0, and let the set of alternate hypotheses be given by $\{D_z\}_{z \in \{\pm 1\}^{d/2}}$, where D_z the distribution over $[d]$ whose x-th marginal is $D_z(x) = \frac{1}{d} + (-1)^x \cdot \frac{\epsilon}{d} \cdot z_{\lceil x/2 \rceil}$ for any $x \in [d]$. Clearly $D_z \in \mathcal{S}_1$ for all z.

There is no obviously no adaptivity in what the tester does after seeing each new sample. So the family of likelihood ratio factors $\{g_t^z(\cdot)\}$ for which (1) *holds is given by*

$$g_t^z(x) = g^z(x) \triangleq \epsilon(-1)^x \cdot z_{\lceil x/2 \rceil}. \tag{2}$$

The definition of $p_0^{\leq N}, p_1^{\leq N}$ in our proofs will be straightforward (see Construction 1), and by Fact III.4, the key technical difficulty is to upper bound the total variation distance between $p_0^{\leq N}, p_1^{\leq N}$ in terms of N. After recording some notation in Section I-C, in Section III-A, we overview our approach for doing so in the non-adaptive setting of Definition III.6, and in Section III-B, we describe our techniques for extending these bounds to the generic, adaptive setting of Definition III.3.

A. Non-Adaptive Lower Bounds

It is a standard trick to upper bound total variation distance between two distributions in terms of the χ^2-divergence, which is often more amenable to calculations. These calculations are especially straightforward in the non-adaptive setting of Definition III.6.

Lemma III.8. *Let $p_0^{\leq N}, p_1^{\leq N}, \mathcal{D}, \{g_t^\zeta(\cdot)\}_{t \in \mathbb{N}, \zeta \in \mathrm{supp}(\mathcal{D})}$ be defined as in Definition III.6. As $p_0^{\leq N}$ is therefore a product*

measure, for every $t \in [N]$ denote its t-th marginal by p_0^t. Then

$$\frac{1}{2\ln 2} d_{\mathrm{TV}}\left(p_1^{\le N}, p_0^{\le N}\right)^2 \le \chi^2\left(p_1^{\le N} \| p_0^{\le N}\right) \le Z \max_t \mathop{\mathbb{E}}_{\zeta,\zeta'}\left[\left(1 + \mathop{\mathbb{E}}_{x_t \sim p_0^t}\left[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)\right]\right)^N\right] - 1.$$

Proof. The first inequality is just Pinsker's and the fact that chi-squared divergence upper bounds KL divergence. For the latter inequality, it will be convenient to define

$$g_S^\zeta(x_S) \triangleq \prod_{t \in S} g_t^\zeta(x_t).$$

Then for any ζ, ζ', S, the product structure implies

$$\mathop{\mathbb{E}}_{x_{\le N} \sim p_0^{\le N}}\left[g_S^\zeta(x_S) g_S^{\zeta'}(x_S)\right] = \prod_{t \in S} \mathop{\mathbb{E}}_{x_t \sim p_0^t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)] \quad (3)$$

We then get that

$$\begin{aligned}
&\chi^2\left(p_1^{\le N} \| p_0^{\le N}\right) \\
&= \mathop{\mathbb{E}}_{x_{\le N}, \zeta, \zeta'}\left[\sum_{\emptyset \ne S, S' \subseteq [N]} g_S^\zeta(x_S) g_{S'}^{\zeta'}(x_{S'})\right] \\
&= \mathop{\mathbb{E}}_{x_{\le N}, \zeta, \zeta'}\left[\sum_{S \ne \emptyset} g_S^\zeta(x_S) g_S^{\zeta'}(x_S)\right] \\
&= \mathop{\mathbb{E}}_{\zeta,\zeta'}\left[\prod_{t=1}^N \left(1 + \mathop{\mathbb{E}}_{x_t \sim p_0^t}\left[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)\right]\right)\right] - 1 \\
&\le \max_t \mathop{\mathbb{E}}_{\zeta,\zeta'}\left[\left(1 + \mathop{\mathbb{E}}_{x_t \sim p_0^t}\left[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)\right]\right)^N\right] - 1, \quad (4)
\end{aligned}$$

where the third step follows by (3), the last step follows by Holder's, and the second step follows by the fact that for $S \ne S'$ and any ζ, ζ', $\mathbb{E}_{x_{\le N}}[g_S^\zeta(x_S) g_{S'}^{\zeta'}(x_{S'})] = 0$. □

The upshot of (4) is that the fluctuations of the quantities $\mathbb{E}_{x_t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)]$ with respect to the randomness of ζ, ζ' dictate how large N must be for $p_0^{\le N}$ and $p_1^{\le N}$ to be distinguishable.

Example III.9. *Recalling (2), the quantities $\mathbb{E}_{x_t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)]$ take a particularly nice form in Paninski's setting:*

$$\begin{aligned}
\mathop{\mathbb{E}}_{x_t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)] &= \epsilon^2 \cdot \mathop{\mathbb{E}}_{x \sim [d]}\left[z_{\lceil x/2 \rceil} \cdot z'_{\lceil x/2 \rceil}\right] \\
&= \frac{\epsilon^2}{d} \sum_{x=1}^d \mathbb{1}\left[z_{\lceil x/2 \rceil} = z'_{\lceil x/2 \rceil}\right] = \frac{2\epsilon^2}{d} \langle z, z' \rangle \quad (5)
\end{aligned}$$

Because $\langle z, z' \rangle$ is distributed as a shifted, rescaled binomial distribution, $\mathbb{E}_{x_t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)]$ has sub-Gaussian tails and fluctuations of order $O(\epsilon^2/\sqrt{d})$, implying that for N as large as $o(\sqrt{d}/\epsilon^2)$, $\chi^2\left(p_1^{\le N} \| p_0^{\le N}\right) = o(1)$. While this is not exactly how Paninski's lower bound was originally proven, concentration of the binomial random variable $\langle z, z' \rangle$ lies at the heart of the lower bound and formalizes the usual intuition for the $\sqrt{d}$ scaling in the lower bound: to tell whether a distribution is far from uniform, it is necessary to draw $\Omega(\sqrt{d})$ samples just to see some element of $[d]$ appear twice.

In Section V, we will show how to use Lemma III.8 to prove Theorem I.1. As it turns out, understanding the fluctuations of the random variable $\mathbb{E}_{x_t}[g_t^\zeta(x_t) g_t^{\zeta'}(x_t)]$ that arises in that setting will be one of the primary technical challenges of this work, both for our adaptive and non-adaptive lower bounds (see Section VIII).

B. Adaptive Lower Bounds

As was discussed previously and is evident from the proof of Lemma III.8, the lack of product structure for $p_0^{\le N}$ and $p_1^{\le N} | \zeta$ in the adaptive setting of Definition III.3 makes it infeasible to directly estimate $\chi^2\left(p_1^{\le N} \| p_0^{\le N}\right)$. Inspired by the literature on bandit lower bounds [10], [11], we instead upper bound KL $\left(p_1^{\le N} \| p_0^{\le N}\right)$, for which we can appeal to the chain rule to tame the extra power afforded by adaptivity. To handle the mixture structure of $p_1^{\le N}$, we will upper bound each of the resulting *conditional* KL divergence terms by their corresponding conditional χ^2 divergence.

First, we introduce some notation essential to the calculations in this work.

Definition III.10 (Key Quantities)**.** *In the generic setup of Definition III.3, for any $x_{\le t} \in \mathcal{U}^t$, define*

$$\begin{aligned}
\Delta(x_{\le t}) &\triangleq \frac{\mathrm{d}p_1^{\le t}}{\mathrm{d}p_0^{\le t}}(x_{\le t}) \\
\phi_{x_{\le t}}^{\zeta,\zeta'} &\triangleq \mathop{\mathbb{E}}_{x \sim p_0^t(\cdot | x_{\le t})}\left[g_{x_{\le t}}^\zeta(x) g_{x_{\le t}}^{\zeta'}(x)\right] \\
\Psi_{x_{\le t}}^{\zeta,\zeta'} &\triangleq \prod_{i=1}^t (1 + g_{x_{<i}}^\zeta(x_i))(1 + g_{x_{<i}}^{\zeta'}(x_i)) \quad (6)
\end{aligned}$$

The following is a key technical ingredient of this work.

Lemma III.11. *Let $p_0^{\le N}, p_1^{\le N}, \mathcal{D}, \{g_{x_{<t}}^\zeta(\cdot)\}$ be defined as in Definition III.3. Then*

$$\begin{aligned}
\frac{1}{2\ln 2} d_{\mathrm{TV}}\left(p_0^{\le N}, p_1^{\le N}\right)^2 &\le KL\left(p_1^{\le N} \| p_0^{\le N}\right) \\
&\le \sum_{t=1}^N \mathop{\mathbb{E}}_{x_{<t} \sim p_0^{\le t-1}}\left[\frac{1}{\Delta(x_{<t})} \mathop{\mathbb{E}}_{\zeta,\zeta' \sim \mathcal{D}}\left[\phi_{x_{<t}}^{\zeta,\zeta'} \cdot \Psi_{x_{<t}}^{\zeta,\zeta'}\right]\right].
\end{aligned}$$

Proof. The first inequality is Pinsker's. For the second, by the chain rule for KL divergence and the fact that chi-squared divergence upper bounds KL, KL $\left(p_1^{\le (N)} \| p_0^{\le N}\right)$ can be written as

$$\begin{aligned}
&\sum_{t=1}^N \mathop{\mathbb{E}}_{x_{<t} \sim p_1^{\le t-1}}\left[\mathrm{KL}\left(p_1^t(\cdot | x_{<t}) \| p_0^t(\cdot | x_{<t})\right)\right] \\
&\qquad \le \sum_{t=1}^N \mathop{\mathbb{E}}_{x_{<t} \sim p_1^{\le t-1}}\left[\chi^2\left(p_1^t(\cdot | x_{<t}) \| p_0^t(\cdot | x_{<t})\right)\right].
\end{aligned}$$

By definition, the conditional densities $p_0^t(\cdot|x_{<t}), p_1^t(\cdot|x_{<t})$ satisfy

$$p_i^t(x_t|x_{<t}) = \frac{p_i^{\leq t}(x_{<t} \circ x_t)}{p_i^{\leq t-1}(x_{<t})} \quad \text{for } i = 0, 1. \tag{7}$$

Therefore, $\mathbb{E}_{x_{<t} \sim p_1^{\leq t-1}} \left[\chi^2 \left(p_1^t(\cdot|x_{<t}) \| p_0^t(\cdot|x_{<t})\right)\right]$ equals

$$\mathop{\mathbb{E}}_{x_{<t} \sim p_1^{\leq t-1}} \left[\mathop{\mathbb{E}}_{x_t \sim p_0^t(\cdot|x_{<t})} \left[\left(\frac{\Delta(x_{<t} \circ x_t)}{\Delta(x_{<t})} - 1 \right)^2 \right] \right] =$$

$$\mathop{\mathbb{E}}_{x_{<t} \sim p_0^{\leq t-1}} \left[\frac{1}{\Delta(x_{<t})} \mathop{\mathbb{E}}_{x_t \sim p_0^t(\cdot|x_{<t})} \left[\left(\Delta(x_{<t} \circ x_t) - \Delta(x_{<t})\right)^2 \right] \right]$$

by (7) and, in the last step, a change of measure in the outer expectation.

By the assumption (1) and the definition of $\Delta(\cdot)$, $\Delta(x_{<t}) = \mathbb{E}_\zeta \left[\prod_{i=1}^{t-1} (1 + g^\zeta(x_i)) \right]$. This yields

$$\mathop{\mathbb{E}}_{x_t \sim p_0^t(\cdot|x_{<t})} \left[\left(\Delta(x_{<t} \circ x_t) - \Delta(x_{<t})\right)^2 \right]$$

$$= \mathop{\mathbb{E}}_{x_t \sim p_0^t(\cdot|x_{<t})} \left[\mathop{\mathbb{E}}_{\mathcal{D}} \left[\prod_{i=1}^{t-1} (1 + g^\zeta_{x_{<i}}(x_i)) \cdot g^\zeta_{x_{<t}}(x_t) \right]^2 \right],$$

from which the lemma follows. □

IV. Lower Bound Instance

In this section we provide some preliminary notions and calculations that are essential to understanding the proofs of Theorem I.1 and I.2. We first formalize the notion of quantum property testing with unentangled, possibly adaptive measurements in Section IV-A. Then in Section IV-B, we give our lower bound construction and instantiate it in the generic setup of Definition III.3. Finally, in Section IV-C, we give some intuition for some of the key quantities that arise.

A. Testing with Unentangled Measurements

We first formally define the notion of a POVM with possibly infinite outcome set.

Definition IV.1. *Given space Ω with Borel σ-algebra $\mathcal{B}(\Omega)$, let μ be a regular positive real-valued measure μ on $\mathcal{B}(\Omega)$, and let $M : \Omega \to \mathbb{C}^{d \times d}$ be a measurable function taking values in the set of psd Hermitian matrices. We will denote the image of $x \in \Omega$ under M by M_x.*

We say that the pair (μ, M) specifies a POVM *$\mathcal{M}$ if $\int_\Omega M \, \mathrm{d}\mu = \mathbf{I}_{d \times d}$ and, for any $d \times d$ density matrix ρ, the map $B \mapsto \int_B \langle M_x, \rho \rangle \, \mathrm{d}\mu$ for $B \in \mathcal{B}(\Omega)$ specifies a probability measure over Ω. We call the distribution given by this measure the* distribution over outcomes from measuring ρ with $\mathcal{M}$.[2]

Given a POVM $\mathcal{M}$, we will refer to the space of measurement outcomes as $\Omega(\mathcal{M})$.

[2]This definition looks diferent from standard ones because we are implicitly invoking the Radon-Nikodym theorem for POVMs on finite-dimensional Hilbert spaces, see e.g. Theorem 3 from [39] or Lemma 11 from [40].

With no meaningful loss in understanding, the reader may simply imagine that all POVMs mentioned henceforth have finitely many outcomes so that a POVM is simply the data of some finite set of positive semidefinite Hermitian matrices $\{M_x\}_{x \in \Omega}$ for which $\sum_x M_x = \mathbf{I}_{d \times d}$, though our arguments extend to the full generality of Definition IV.1.

Definition IV.2. *Let $N \in \mathbb{N}$. An* unentangled, possibly adaptive POVM schedule *$\mathcal{S}$ is a type of measurement schedule specified by a (possibly infinite) collection of POVMs $\{\mathcal{M}^{x_{<t}}\}_{t \in [N], x_{<t} \in \mathcal{T}_t}$ where $\mathcal{T}_1 \triangleq \{\emptyset\}$, and for every $t > 1$, $\mathcal{T}_t$ denotes the set of all possible transcripts of measurement outcomes $x_{<t}$ for which $x_i \in \Omega(\mathcal{M}^{x_{<i}})$ for all $1 \leq i \leq t-1$ (recall that $x_{<i} \triangleq (x_1, ..., x_{i-1})$). The schedule works in the natural manner: at time t for $t = 1, \ldots, N$, given a transcript $x_{<t} \in \mathcal{T}_t$, it measures the t-th copy of ρ using the POVM $\mathcal{M}^{x_{<t}}$.*

If in addition the resulting schedule is also a nonadaptive measurement schedule, we say it is an unentangled, nonadaptive POVM schedule.

B. Lower Bound Instance

Let $\mathcal{D}$ be the Haar measure over the unitary group $U(d)$. In place of ζ from Definition III.3, we will denote elements from $\mathcal{D}$ by $\mathbf{U}$. $\mathbb{P}_{\mathbf{U}}[\cdot]$ and $\mathbb{E}_{\mathbf{U}}[\cdot]$ will be with respect to $\mathcal{D}$ unless otherwise specified.

Construction 1. *Let $\mathbf{X} \in \mathbb{R}^{d \times d}$ denote the diagonal matrix whose first $d/2$ diagonal entries are equal to ϵ, and whose last $d/2$ diagonal entries are equal to $-\epsilon$. Let $\mathbf{X}' \triangleq \frac{1}{\epsilon}\mathbf{X}$. Let $\mathbf{\Lambda} \triangleq \frac{1}{d}(\mathbf{I} + \mathbf{X})$.*

Our lower bound instance will be the distribution over densities $\mathbf{U}^\dagger \mathbf{\Lambda} \mathbf{U}$ for $\mathbf{U} \sim \mathcal{D}$. We remark that this instance, the quantum analogue of Paninski's lower bound instance [9] for classial uniformity testing, has appeared in various forms throughout the quantum learning and testing literature [1], [7], [3].

Given $N \in \mathbb{N}$, define $\boldsymbol{\rho}_0^{\leq N} \triangleq \rho_{\mathsf{mm}}^{\otimes N}$ and $\boldsymbol{\rho}_1^{\leq N} \triangleq \mathbb{E}_{\mathbf{U} \sim \mathcal{D}}[(\mathbf{U}^\dagger \mathbf{\Lambda} \mathbf{U})^{\otimes N}]$. Take any POVM schedule $\mathcal{S} = \{\mathcal{M}^{x_{<t}}\}_{t \in [N], x_{<t} \in \mathcal{T}_t}$. Given $t \leq N$, define $p_0^{\leq t}$ and $p_1^{\leq t}$ to be the distribution over the measurement outcomes when the first t steps of these POVM schedules are applied to the first t parts of $\boldsymbol{\rho}_0^{\leq N}$ and $\boldsymbol{\rho}_1^{\leq N}$ respectively. Equivalently, $p_1^{\leq t}$ can be regarded as the distribution over sequences of t measurement outcomes arising from first sampling $\mathbf{U}$ according to the Haar measure $\mathcal{D}$ and then applying the first t steps of POVM schedule $\mathcal{S}$ to t copies of $\rho \triangleq \mathbf{U}^\dagger \mathbf{\Lambda} \mathbf{U}$.

Lemma IV.3. *For any POVM $\mathcal{M}$, define*

$$g^{\mathbf{U}}_{\mathcal{M}}(x) \triangleq \langle \widehat{M}^{x_{<t}}_x, \mathbf{U}^\dagger \mathbf{X} \mathbf{U} \rangle. \tag{8}$$

$p_1^{\leq N}$ is absolutely continuous with respect to $p_0^{\leq N}$, and the family of likelihood ratio factors $\{g^{\mathbf{U}}_{x_{<t}}(\cdot)\}$ for which (1) *holds for $p_0^{\leq N}$ and $p_1^{\leq N}$ defined in Construction 1 is given by $g^{\mathbf{U}}_{x_{<t}}(\cdot) \triangleq g^{\mathbf{U}}_{\mathcal{M}^{x_{<t}}}$.*

The proof of this is straightforward and we defer it to the full version.

For any $\mathbf{U}, \mathbf{U}' \in U(d)$, the quantities $\Psi_{x<t}^{\mathbf{U},\mathbf{U}'}$ and $\phi_{x<t}^{\mathbf{U},\mathbf{U}'}$ are given by (6). Given a POVM $\mathcal{M}$, also define $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ in the obvious way. Lastly, we record the following basic facts:

Fact IV.4. *For any POVM $\mathcal{M}$,*

(I) $\mathbb{E}_{x\sim p}[g_{\mathcal{M}}^{\mathbf{U}}(x)] = 0$ *for any* $\mathbf{U} \in U(d)$.

(II) For any measurement outcome x and $\mathbf{U}, \mathbf{U}' \in U(d)$, $|g_{\mathcal{M}}^{\mathbf{U}}(x)| \le \epsilon$ and thus $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'} \le \epsilon^2$.

C. Intuition for $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$

Recall from Example III.9 that for classical uniformity testing, $\phi^{z,z'} = \frac{2\epsilon^2}{d}\langle z, z'\rangle$, and by Lemma III.8, the $O(\epsilon^2/\sqrt{d})$ fluctuations of $\phi^{z,z'}$ as a random variable in z, z' precisely dictate the sample complexity of uniformity testing.

One should therefore think of the distribution of the quantity $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ as a "quantum analogue" of the binomial distribution whose fluctuations are closely related to the scaling of the copy complexity of mixedness testing.

As we will show in Theorem V.1, $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ has $O(\epsilon^2/d^{3/2})$ fluctuations and concentrates well, from which it will follow by integration by parts that N can be taken as large as $o(d^{3/2}/\epsilon^2)$, yielding the lower bound of Theorem I.1.

To get some intuition for where these $O(\epsilon^2/d^{3/2})$ fluctuations come from, suppose $\mathcal{M}$ were the orthogonal POVM given by the standard basis. Then

$$\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'} = \frac{1}{d}\sum_{i=1}^{d} \epsilon^2 \cdot \delta(\mathbf{U}_i) \cdot \delta(\mathbf{U}_i'),$$

where

$$\delta(v) \triangleq \sum_{i=1}^{d/2} v_i^2 - \sum_{i=d/2+1}^{d} v_i^2. \tag{9}$$

For any fixed i, $\mathbf{U}_i, \mathbf{U}_i'$ are independent random unit vectors, and the variance of $\delta(\mathbf{U}_i)\cdot\delta(\mathbf{U}_i')$ is $O(1/d^2)$ (see Fact VIII.2). If $\mathbf{U}_1, \mathbf{U}_1' ..., \mathbf{U}_d, \mathbf{U}_d'$ were all independent, then $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ would thus have variance ϵ^4/d^3, suggesting $O(\epsilon^2/d^{3/2})$ fluctuations as claimed. Of course we do not actually have this independence assumption; in addition, the other key technical challenges we must face to get Theorem V.1 are 1) to go beyond just a second moment bound and show sufficiently strong concentration of $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$, and 2) to show this is the case for *all POVMs*. We do this in Section VIII.

V. Proof of Non-Adaptive Lower Bound

In this section we prove Theorem I.1 by applying Lemma III.8; the technical crux of the proof (and of our proof of Theorem I.2 in the next section) is the following tail bound, whose proof we defer to Section VIII:

Theorem V.1. *Fix any POVM $\mathcal{M}$. There exists an absolute constant $c'' > 0$ such that for any $t > \Omega(\epsilon^2/d^{1.99})$, we have*

$$\mathop{\mathbb{P}}_{\mathbf{U},\mathbf{U}'\sim\mathcal{D}}\left[\left|\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}\right| > t\right] \le \exp\left(-c''\left\{\frac{d^3t^2}{\epsilon^4} \wedge \frac{d^2t}{\epsilon^2}\right\}\right)$$

Proof of Theorem I.1. By Fact III.4, it suffices to show that no nonadaptive POVM schedule can solve the distinguishing task given by Construction 1, unless $N = \Omega(d^{3/2}/\epsilon^2)$. For a nonadaptive POVM schedule $\mathcal{S}$, let $\{\mathcal{M}^1, ..., \mathcal{M}^N\}$ denote the sequence of POVMs that are used. Recalling (8), the likelihood ratio factors $\{g_t^{\mathbf{U}}(\cdot)\}_{\mathbf{U}\in U(d), t\in[N]}$ for which (1) holds in the nonadaptive setting of Definition III.6 are given by $g_{\mathcal{M}^t}^{\mathbf{U}}(\cdot)$. Similarly, denote $\phi_{x<t}^{\mathbf{U},\mathbf{U}'}$ by $\phi_t^{\mathbf{U},\mathbf{U}'}$.

By Lemma III.8, we have

$$\frac{1}{2\ln 2} d_{\mathrm{TV}}\left(p_{\overline{1}}^{\le N}, p_{\overline{0}}^{\le N}\right)^2 \le \max_t \mathop{\mathbb{E}}_{\zeta,\zeta'}\left[\left(1+\phi_t^{\mathbf{U},\mathbf{U}'}\right)^N\right] - 1.$$

To finish the proof, we will show that

$$\sup_{\mathcal{M}} \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\left(1+\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}\right)^N\right] = 1 + o(1)$$

for $N = o(d^{3/2}/\epsilon^2)$, from which the proof is complete by (4).

We would like to apply integration by parts (Fact I.4) to the random variable $Z \triangleq 1+\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ and the function $f(Z) \triangleq Z^N$. By Part (II) of Fact IV.4, this random variable is supported in $[1-\epsilon^2, 1+\epsilon^2]$. We can take the parameters in Fact I.4 as follows: set $a \triangleq 1+\epsilon/(N^{1/2}d^{3/4})$, $b \triangleq 1+\epsilon^2$, and tail bound function $\tau(x) = \exp\left(-c''\left\{\frac{d^3(x-1)^2}{\epsilon^4} \wedge \frac{d^2(x-1)}{\epsilon^2}\right\}\right)$. Note that for $N = o(d^{3/2}/\epsilon^2)$, $(1+\tau(a))f(a) = 1+o(1)$. So by Fact I.4 and Theorem V.1,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\left(1+\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}\right)^N\right] - (1+o(1)) \\
&\quad\le \int_{\frac{\epsilon^2}{d^{3/2}}}^{\epsilon^2} N(1+x)^{N-1}\left(e^{-\frac{c''d^3x^2}{\epsilon^4}} + e^{-\frac{c''d^2x}{\epsilon^2}}\right)\,\mathrm{d}x \\
&\quad= (N/2)!(c''d^3/\epsilon^4)^{-N/2} + N!(c''d^2/\epsilon^2)^N,
\end{aligned}$$

and because $N = o(d^{3/2}/\epsilon^2)$, this is $1+o(1)$. □

VI. A Chain Rule Proof of Paninski's Theorem

As discussed previously, the proof of Theorem I.1 completely breaks down when the POVM schedule $\mathcal{S}$ is adaptive, so we will instead use the chain rule, via Lemma III.11, to prove Theorem I.2.

As a warmup, in this section we will show how to use Lemma III.11 to prove a lower bound for *classical* uniformity testing. As it turns out, it is possible to recover Paninski's optimal $\Omega(\sqrt{d}/\epsilon^2)$ lower bound with this approach, the details of which we give in the appendix of the full version, but in this section we opt to present a proof which achieves a slightly weaker bound. The reason is that in our proof of Theorem VI.1, we will make minimal use of the kind of precise cancellations that would yield a tight bound but which, unfortunately, are specific to the product structure of the distribution of random signs z. As such, these steps will be general-purpose enough to extend to the quantum setting where the Haar measure over $U(d)$ enjoys no such product structure.

Specifically, we use the chain rule to show the following:

Theorem VI.1 (Weaker Paninski Theorem)**.** *$\Omega(d^{1/3}/\epsilon^2)$ samples are necessary to test whether a distribution p is ϵ-far from the uniform distribution.*

In this section, let $p_0^{\le N}, p_1^{\le N}$ denote the distributions defined in Example III.7. Recalling the notation from Example III.7 and Definition III.10, as well as the identities (2) and (5), we immediately get the following from Lemma III.11:

Lemma VI.2. *KL* $\left(p_1^{\le N} \| p_0^{\le N}\right) \le \sum_{t=1}^N Z_t$ *for*

$$Z_t \triangleq \mathop{\mathbb{E}}_{x_{<t}\sim U^{\otimes t-1}}\left[\frac{1}{\Delta(x_{<t})}\mathop{\mathbb{E}}_{z,z'\sim\{\pm1\}^{d/2}}\left[\phi^{z,z'}\cdot\Psi^{z,z'}_{x_{<t}}\right]\right]. \quad (10)$$

We will also need the following estimates (see full version).

Lemma VI.3. $\Delta(x_{<t}) \ge \left(1-\epsilon^2\right)^{(t-1)/2}$ *for any* $x_{<t}$.

Lemma VI.4. $\mathbb{E}_{x_{<t}\sim U^{\otimes t-1}}[(\Psi^{z,z'}(x_{<t}))^2] \le (1+O(\epsilon^2))^{t-1}$ *for any* $z, z' \in \{\pm1\}^{d/2}$.

We now describe how to use these to bound the summands Z_t in (10). As discussed in Example III.9, $\phi^{z,z'} = \frac{2\epsilon^2}{d}\langle z, z'\rangle$ has $O(\epsilon^2/\sqrt{d})$ fluctuations. If we pretended $\phi^{z,z'}$ was of this magnitude with probability one, then Z_t would be of order

$$O(\epsilon^2/\sqrt{d})\cdot\mathop{\mathbb{E}}_{x_{<t}\sim U^{\otimes t-1}}\left[\frac{1}{\Delta(x_{<t})}\mathop{\mathbb{E}}_{z,z'\sim\{\pm1\}^{d/2}}\left[\Psi^{z,z'}_{x_{<t}}\right]\right],$$

which is just $O(\epsilon^2/\sqrt{d})$ because $\Delta(x_{<t})^2 = \mathbb{E}_{z,z'}[\Psi^{z,z'}_{x_{<t}}]$ and the likelihood ratio between two distributions always integrates to 1. Then by (10) we would in fact even recover Theorem I.3.

Unfortunately, in reality $\phi^{z,z'}$ can be as large as order ϵ^2, albeit with exponentially small probability, so instead we will partition the space of $z, z' \in \{\pm1\}^{d/2}$ into those for which $\phi^{z,z'}$ is either less than some threshold τ or greater. When $\phi^{z,z'} \le \tau$, we can bound the total contribution to Z_t of such z, z' by τ. When $\phi^{z,z'} > \tau$, we will use the pointwise estimates from Lemmas VI.3 and VI.4 and argue that because $\mathbb{P}[\phi^{z,z'} > \tau]$ is so small, these z, z' contribute negligibly to Z_t. The reason we only get an $\Omega(d^{1/3}/\epsilon^2)$ lower bound in the end is that we must take τ slightly larger than the fluctuations of $\phi^{z,z'}$ to balance the low probability of $\phi^{z,z'}$ exceeding τ with the pessimistic pointwise estimates of Lemmas VI.3 and VI.4.

Proof of Theorem VI.1. We fill in the details of the strategy outlined above. We will use Fact III.4 with the construction in Example III.7. Given a transcript $x_{<t}$ and $z, z' \in \{\pm1\}^{d/2}$, let $\mathbb{1}\left[\mathcal{E}^{z,z'}(\tau)\right]$ denote the indicator of whether $\phi^{z,z'} > \tau$. Then $\mathbb{E}_{z,z'}\left[\Psi^{z,z'}_{x_{<t}}\cdot\phi^{z,z'}\right]$ is at most

$$\begin{aligned}&\epsilon^2\cdot\mathop{\mathbb{E}}_{z,z'}\left[\Psi^{z,z'}_{x_{<t}}\cdot\mathbb{1}\left[\mathcal{E}^{z,z'}(\tau)\right]\right]+\tau\cdot\mathop{\mathbb{E}}_{z,z'}\left[\Psi^{z,z'}_{x_{<t}}\cdot\mathbb{1}\left[\mathcal{E}^{z,z'}(\tau)^c\right]\right]\\ &\le \epsilon^2\cdot\underbrace{\mathop{\mathbb{E}}_{z,z'}\left[\Psi^{z,z'}_{x_{<t}}\cdot\mathbb{1}\left[\mathcal{E}^{z,z'}(\tau)\right]\right]}_{\text{(B)}_{x_{<t}}}+\tau\cdot\underbrace{\mathop{\mathbb{E}}_{z,z'}\left[\Psi^{z,z'}_{x_{<t}}\right]}_{\text{(G)}_{x_{<t}}},\end{aligned}$$

where in the second step we used Part (II) of Fact IV.4. Note that for any transcript $x_{<t}$, $\Delta(x_{<t})^2 = \mathbb{E}_{z,z'}[\Psi^{z,z'}_{x_{<t}}] = \text{(G)}_{x_{<t}}$, so by this and the fact that the likelihood ratio between two distributions always integrates to 1,

$$\mathop{\mathbb{E}}_{x_{<t}\sim U^{\otimes t-1}}\left[\frac{1}{\Delta(x_{<t})}\cdot\text{(G)}_{x_{<t}}\right] = \mathop{\mathbb{E}}_{x_{<t}}\left[\Delta(x_{<t})\right] = 1. \quad (11)$$

We conclude that $Z_t \le \epsilon^2\cdot(1+\epsilon^2)^{(t-1)/2}\,\mathbb{E}_{x_{<t}}[\text{(B)}_{x_{<t}}]+\tau$, where the second step follows by Lemma VI.3 and (11). It remains to show that τ is the dominant quantity above, for appropriately chosen τ.

Pick $\tau = \Omega(\epsilon^2/d^{1/3})$. To upper bound $\mathbb{E}_{x_{<t}}[\text{(B)}_{x_{<t}}]$, first apply Cauchy-Schwarz to get

$$\begin{aligned}\mathop{\mathbb{E}}_{x_{<t}}[\text{(B)}_{x_{<t}}] &\le \mathop{\mathbb{E}}_{x_{<t},z,z'}\left[\left(\Psi^{z,z'}_{x_{<t}}\right)^2\right]^{1/2}\mathop{\mathbb{P}}_{x_{<t},z,z'}\left[\phi^{z,z'}>\tau\right]^{1/2}\\ &\le (1+O(\epsilon^2))^{(t-1)/2}\cdot\exp(-\Omega(d^{1/3})),\end{aligned}$$

where the second step follows by Lemma VI.4, (5), and standard binomial tail bounds. For $t = o(d^{1/3}/\epsilon^2)$, this quantity is indeed negligible, concluding the proof that $Z_t \le O(\epsilon^2/d^{1/3})$ and, by Lemma VI.2, that $\chi^2\left(p_1^{\le N}\|p_0^{\le N}\right) = o(1)$.. □

a) Parallels to Proof of Theorem I.2: Lastly, we comment on how these ingredients carry over to our proof of Theorem I.2. Lemma VI.2 translates verbatim to the quantum setting (see Lemma VII.1), as does the final part of the proof where we partition based on the value of $\phi^{z,z'}$.

Lemma VII.2 will be the quantum analogue of Lemma VI.3, and its proof uses a similar trick of AM-GM plus averaging with an involution.

Lemma VII.4 will be the quantum analogue of Lemma VI.4. Unfortunately, as we will see later in Section VII, an analogously naive bound will not suffice in our proof of Theorem I.2. The workaround is somewhat technical, and we defer the details to Lemma VII.4 and the discussion preceding it.

Finally, as in Section III-A, the central technical ingredient in the proof of Theorem VI.1 is the concentration of $\phi^{z,z'}$. Analogously, in the proof of Theorem I.2, we will need sufficiently strong tail bounds for $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$, which we show in Theorem V.1.

VII. An Adaptive Lower Bound for Mixedness Testing

In this section we prove our main result, Theorem I.2.

First, recalling the notation from Construction 1 and Definition III.10, as well as the identity (8), we immediately get the following from Lemma III.11:

Lemma VII.1. *KL* $\left(p_1^{\le N}\|p_0^{\le N}\right) \le \sum_{t=1}^N Z_t$ *for*

$$Z_t \triangleq \mathop{\mathbb{E}}_{x_{<t}\sim p_0^{\le t-1}}\left[\frac{1}{\Delta(x_{<t})}\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t}}\cdot\phi^{\mathbf{U},\mathbf{U}'}_{x_{<t}}\right]\right] \quad (12)$$

Take any $t \le N$. To bound Z_t in (12), we first estimate the likelihood ratio Δ for an arbitrary transcript, in analogy with Lemma VI.3 from Section VI respectively:

Lemma VII.2. $\Delta(x_{<t}) \ge \left(1-O(\epsilon^2/d)\right)^{t-1}$ *for any* $x_{<t}$.

Proof. By convexity of the exponential function and the fact that $1 + g^{\mathbf{U}}_{x_{<i}}(x_i) > 0$ for all $\mathbf{U}, i, x_i$,

$$\Delta(x_{<t}) \geq \exp\left(\mathop{\mathbb{E}}_{U\sim\mathcal{D}}\left[\sum_{i=1}^{t-1}\ln\left(1+g^{\mathbf{U}}_{x_{<i}}(x_i)\right)\right]\right)$$
$$= \prod_{i=1}^{t-1}\exp\left(\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}\left[\ln(1+g^{\mathbf{U}}_{x_{<i}}(x_i))\right]\right). \quad (13)$$

Define the unitary block matrix $\mathbf{T} = \begin{pmatrix}\mathbf{0} & \mathbf{I}_{d/2}\\ \mathbf{I}_{d/2} & \mathbf{0}.\end{pmatrix}$ As $\mathcal{D}$ is invariant with respect to left-multiplication by $\mathbf{T} \in U(d)$, for all $i < t$ we can write $\exp\left(\mathbb{E}_{\mathbf{U}\sim\mathcal{D}}\left[\ln(1+g^{\mathbf{U}}_{x_{<i}}(x_i))\right]\right)$ as

$$\exp\left(\frac{1}{2}\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}\left[\ln(1-g^{\mathbf{U}}_{x_{<i}}(x_i)^2)\right]\right)$$
$$\geq 1+\frac{1}{2}\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}\left[\ln(1-g^{\mathbf{U}}_{x_{<i}}(x_i)^2)\right]$$
$$\geq 1-\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{D}}\left[g^{\mathbf{U}}_{x_{<i}}(x_i)^2)\right] \quad (14)$$

Finally, note that for any trace-one psd matrix M, we may write $M = \sum\lambda_i v_i v_i^\dagger$, and for any unit vector $v \in \mathbb{C}^n$, $\mathbb{E}_{\mathbf{U}}[\langle vv^\dagger, \mathbf{U}^\dagger\mathbf{X}\mathbf{U}\rangle^2] = O(\epsilon^2/d)$. So $\mathbb{E}_U[\langle M, \mathbf{U}^\dagger\mathbf{X}\mathbf{U}\rangle^2]$ equals

$$\sum_{i,j}\lambda_i\lambda_j\,\mathbb{E}\Big[\langle v_i v_i^\dagger, \mathbf{U}^\dagger\mathbf{X}\mathbf{U}\rangle\langle v_j v_j^\dagger, \mathbf{U}^\dagger\mathbf{X}\mathbf{U}\rangle\Big] \leq O(\epsilon^2/d),$$

by Cauchy-Schwarz. From this we conclude that $\mathbb{E}_{\mathbf{U}\sim\mathcal{D}}[g^{\mathbf{U}}_{x_{<i}}(x_i)^2] \leq O(\epsilon^2/d)$ for all $i, x_{<i}, x_i$, and the lemma follows by (13) and (14). □

Next, in analogy with Lemma VI.4, we would like to control the expectation of $(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t}})^2$. We remark that like in the proof of Lemma VI.4, one can obtain a naive estimate of $(1+O(\epsilon^2))^{t-1}$ using just Fact IV.4, but unlike in the proof of Theorem VI.1, such a bound would not suffice here. Instead, we will need the following important moment bound, whose proof we defer to Section VIII:

Theorem VII.3. *For any POVM $\mathcal{M}$, let p denote the distribution over outcomes from measuring ρ_{mm} with $\mathcal{M}$, and let $\gamma > 0$ be an absolute constant. Define the random variable*

$$K^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}} \triangleq \mathop{\mathbb{E}}_{x\sim p}\left[\left(g^{\mathbf{U}}_{\mathcal{M}}(x) + g^{\mathbf{U}'}_{\mathcal{M}}(x)\right)^2\right]$$

Then for any $n = o(d^2/\epsilon^2)$, we have that

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\left(1+\gamma\cdot K^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}}\right)^n\right] \leq \exp(O(\gamma n\epsilon^2/d)) \quad (15)$$

We will use this and a series of Holder's to prove the following sufficiently strong analogue of Lemma VI.4:

Lemma VII.4. $\mathbb{E}_{x_{<t},\mathbf{U},\mathbf{U}'}[(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t}})^2] \leq \exp(O(t\cdot\epsilon^2/d))$ *for* $t = o(d^2/\epsilon^2)$.

Proof. Consider any $a, b \in \mathbb{Z}$ for which $a \geq b$ and $a \geq 2$. For any $x_{<t-1}$, let p denote the distribution over measurement outcomes when the POVM $\mathcal{M}^{x_{<t-1}}$ is applied to ρ_{mm}. We have by Part (II) of Fact IV.4 that

$$\mathop{\mathbb{E}}_{x\sim p}[g^{\mathbf{U}}_{x_{<t-1}}(x)^a\cdot g^{\mathbf{U}'}_{x_{<t-1}}(x_t)^b] \leq \epsilon\mathop{\mathbb{E}}_{x\sim p}[g^{\mathbf{U}}_{x_{<t-1}}(x)^2].$$

Recalling Part (I) of Fact IV.4, we conclude that for any $x_{<t-1}$ and constant degree $c \geq 2$,

$$\mathop{\mathbb{E}}_{x}[(1+g^{\mathbf{U}}_{x_{<t-1}}(x))^c(1+g^{\mathbf{U}'}_{x_{<t-1}}(x))^c] \leq 1+Z^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}(c). \quad (16)$$

for $Z^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}(c) \triangleq O_c(\mathbb{E}_x[g^{\mathbf{U}}_{x_{<t-1}}(x)^2]) + O_c(\mathbb{E}_x[g^{\mathbf{U}'}_{x_{<t-1}}(x)^2]) + O_c(\phi^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}})$. By abuse of notation, for POVM $\mathcal{M}$, define $Z^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}}(c)$ in the obvious way.

For $\alpha_i \triangleq 2\cdot\left(\frac{t-1}{t-2}\right)^i$, we can bound $\mathbb{E}_{x_{<t},\mathbf{U},\mathbf{U}'}\left[\left(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t}}\right)^{\alpha_i}\right]$ by

$$\mathop{\mathbb{E}}_{x_{<t-1},\mathbf{U},\mathbf{U}'}\left[\left(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}\right)^{\alpha_i}\cdot\left(1+Z^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}(\alpha_i)\right)\right],$$

by (16). By Holder's, we can upper bound this by the product of $\mathbb{E}_{x_{<t-1},\mathbf{U},\mathbf{U}'}\left[\left(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}\right)^{\alpha_i(t-1)/(t-2)}\right]^{(t-2)/(t-1)}$ and $\mathbb{E}_{x_{<t-1},\mathbf{U},\mathbf{U}'}\left[\left(1+Z^{\mathbf{U},\mathbf{U}'}_{x_{<t-1}}(\alpha_i)\right)^{t-1}\right]^{1/(t-1)}$. Unrolling this,

$$\mathop{\mathbb{E}}_{x_{<t},\mathbf{U},\mathbf{U}'}\left[\left(\Psi^{\mathbf{U},\mathbf{U}'}_{x_{<t}}\right)^2\right]$$
$$\leq\prod_{i=1}^{t-1}\mathop{\mathbb{E}}_{x_{<i},\mathbf{U},\mathbf{U}'}\left[\left(1+Z^{\mathbf{U},\mathbf{U}'}_{x_{<i}}(\alpha_{t-1-i})\right)^{t-1}\right]^{1/(t-1)}$$
$$\leq\prod_{i=1}^{t-1}\mathop{\mathbb{E}}_{x_{<i},\mathbf{U},\mathbf{U}'}\left[\left(1+Z^{\mathbf{U},\mathbf{U}'}_{x_{<i}}(2e)\right)^{t-1}\right]^{1/(t-1)}, \quad (17)$$
$$\leq\sup_{\mathcal{M}}\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\left(1+Z^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}}(2e)\right)^{t-1}\right]$$

where (17) follows by the fact that for $1 \leq i \leq t-1$, $\alpha_{t-1-i} \leq 2\left(1+\frac{1}{t-2}\right)^{t-2} \leq 2e$, and the supremum in the last step is over all POVMs $\mathcal{M}$. The proof is complete upon noting that $Z^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}}$ is at most a constant multiple of $K^{\mathbf{U},\mathbf{U}'}_{\mathcal{M}}$ defined in (15) and invoking Theorem VII.3. □

We can now complete the proof of Theorem I.2. The argument is very similar to the argument we used to complete the proof of Theorem VI.1, so we defer it to the full version.

VIII. Haar Tail Bounds

In this section we sketch the proof of the two key estimates, Theorems VII.3 and V.1, which were crucial to our proof of Theorem I.2, deferring the formal proofs to the full version. The following concentration inequality is key to our analysis:

Theorem VIII.1 ([20], Corollary 17, see also [41], Corollary 4.4.28)**.** *Equip $M \triangleq U(d)^k$ with the L_2-sum of Hilbert-Schmidt metrics. If $F : M \to \mathbb{R}$ is L-Lipschitz, then*

$$\mathbb{P}[|F(\mathbf{U}_1,...,\mathbf{U}_k) - \mathbb{E}[F(\mathbf{U}_1,...,\mathbf{U}_k)]| \geq t] \leq e^{-dt^2/12L^2},$$

for any $t > 0$, where $\mathbf{U}_1,...,\mathbf{U}_k$ are independent unitary matrices drawn from the Haar measure.

A. Proof of Theorem VII.3

For convenience, Theorem VII.3 is restated below:

Theorem VII.3. *For any POVM $\mathcal{M}$, let p denote the distribution over outcomes from measuring ρ_{mm} with $\mathcal{M}$, and let $\gamma > 0$ be an absolute constant. Define the random variable*

$$K_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'} \triangleq \mathop{\mathbb{E}}_{x\sim p}\left[\left(g_{\mathcal{M}}^{\mathbf{U}}(x) + g_{\mathcal{M}}^{\mathbf{U}'}(x)\right)^2\right]$$

Then for any $n = o(d^2/\epsilon^2)$, we have that

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}\left[\left(1 + \gamma \cdot K_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}\right)^n\right] \leq \exp(O(\gamma n \epsilon^2/d)) \tag{15}$$

To get intuition for this, here we will only prove this in the special case where $\mathcal{M}$ is an orthogonal POVM given by an orthonormal basis of $\mathbb{C}^d$. Then p is uniform over $[d]$ and

$$K_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'} = \frac{\epsilon^2}{d}\sum_{i=1}^{d}(\delta(\mathbf{U}_i)+\delta(\mathbf{U}'_i))^2 \leq \frac{2\epsilon^2}{d}\sum_{i=1}^{d}(\delta(\mathbf{U}_i)^2+\delta(\mathbf{U}'_i)^2) \tag{18}$$

where $\delta(\cdot)$ is defined in (9). The following is a standard fact:

Fact VIII.2. *For random $v \in \mathbb{S}^{d-1}$, $\mathbb{E}[\delta(v)^2] = \frac{1}{d+1}$.*

While this follows immediately from moments of random unit vectors, for pedagogical purposes we will give a proof using Weingarten calculus, as it will be a crucial ingredient in the full proof. Recall that for every $q \in \mathbb{N}$, there exists a corresponding Weingarten function $\mathrm{Wg}(\cdot, d) : \mathcal{S}_q \to \mathbb{R}$ [42], [43]. In the special case of $q = 2$, the symmetric group $\mathcal{S}_q$ consists of two elements e, τ^*, namely, the identity and non-identity permutation, respectively, and we have that $\mathrm{Wg}(e, d) = \frac{1}{d^2-1}$ and $\mathrm{Wg}(\tau^*, d) = -\frac{1}{d(d^2-1)}$. We then have:

Lemma VIII.3 (Degree-2 case of [43], Lemma 4.3)**.** *Let e, τ^* denote the identity and non-identity permutation of $\mathcal{S}_2$ respectively. For $d \geq 2$ and any $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{d\times d}$, we have that*[3]

$$\mathop{\mathbb{E}}_{\mathbf{U}}[Tr((\mathbf{A}\mathbf{U}^\dagger\mathbf{B}\mathbf{U})^2)] = \sum_{\sigma,\tau\in\mathcal{S}_2}\langle\mathbf{A}\rangle_\sigma\langle\mathbf{B}\rangle_\tau \,\mathrm{Wg}(\sigma\tau^{-1}, d)\ .$$

Proof of Fact VIII.2. Let $\mathbf{\Pi} \triangleq e_1 e_1^\dagger$ and note that $\delta(v)$ is identical in distribution to the quantity $\mathrm{Tr}(\mathbf{\Pi}\mathbf{U}^\dagger\mathbf{X}'\mathbf{U})$. By Lemma VIII.3,

$$\mathop{\mathbb{E}}_{v}[\delta(v)^2] = \sum_{\sigma,\tau\in\mathcal{S}_2}\langle\mathbf{\Pi}\rangle_\sigma\langle\mathbf{X}'\rangle_\tau \,\mathrm{Wg}(\sigma\tau^{-1}, d).$$

$\langle\mathbf{X}'\rangle_\tau = d\cdot\mathbb{1}[\tau = \tau^*]$ and $\langle\mathbf{\Pi}\rangle_\sigma = 1$ for all $\sigma \in \mathcal{S}_2$, so

$$\mathop{\mathbb{E}}_{v}[\delta(v)^2] = d\left(\frac{1}{d^2-1} - \frac{1}{d(d^2-1)}\right) = \frac{1}{d+1}$$

as claimed. □

Furthermore, it is known that $\delta(v)^2$ concentrates around its expectation. So if the columns of $\mathbf{U}$ were actually *independent* random unit vectors, we would conclude that $K_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'} = O(\epsilon^2/d)$ with high probability and obtain (15) for the special case where $\mathcal{M}$ is orthogonal.

To circumvent the issue of dependence among the columns of Haar-random $\mathbf{U}$, we will invoke Theorem VIII.1. The following is a toy version of the more general result that we show in the full proof of Theorem VII.3.

Lemma VIII.4. *For any $t > 0$,*

$$\mathop{\mathbb{P}}_{\mathbf{U}\sim\mathcal{D}}\left[\left(\sum_{i=1}^{d}\delta(\mathbf{U}_i)^2\right)^{1/2} \geq 1 + t\right] \leq \exp\left(-\Omega(dt^2)\right).$$

Proof. By Jensen's and Fact VIII.2,

$$\mathbb{E}\left[\left(\sum_{i=1}^{d}\delta(\mathbf{U}_i)^2\right)^{1/2}\right] \leq \mathbb{E}\left[\sum_{i=1}^{d}\delta(\mathbf{U}_i)^2\right]^{1/2} \leq 1.$$

We wish to invoke Theorem VIII.1, so it suffices to show that $G : \mathbf{U} \mapsto (\sum_{i=1}^{d}\delta(\mathbf{U}_i)^2)^{1/2}$ is $O(1)$-Lipschitz. Recalling the definition of $\mathbf{X}'$ from Construction 1, note that

$$\left(\sum_{i=1}^{d}\delta(\mathbf{U}_i)^2\right)^{1/2} = \|\mathrm{diag}(\mathbf{U}^\dagger\mathbf{X}'\mathbf{U})\|_{HS}.$$

Take any $\mathbf{U}, \mathbf{V} \in U(d)$ and note

$$\begin{aligned} G(\mathbf{U}) - G(\mathbf{V}) &\leq \sqrt{\sum_{i=1}^{d}|(\mathbf{U}^\dagger\mathbf{X}'\mathbf{U})_{ii} - (\mathbf{V}^\dagger\mathbf{X}'\mathbf{V})_{ii}|^2} \\ &\leq \|\mathbf{U}^\dagger\mathbf{X}'\mathbf{U} - \mathbf{V}^\dagger\mathbf{X}'\mathbf{V}\|_{HS} \\ &= \|\mathbf{U}^\dagger\mathbf{X}'(\mathbf{U}-\mathbf{V}) + (\mathbf{V}-\mathbf{U})^\dagger\mathbf{X}'\mathbf{V}\|_{HS} \\ &\leq 2\|\mathbf{X}'\|_2\|\mathbf{U}-\mathbf{V}\|_{HS} = 2\|\mathbf{U}-\mathbf{V}\|_{HS}, \end{aligned}$$

where the first step follows by Cauchy-Schwarz. So $G(\mathbf{U})$ is 2-Lipschitz as desired. □

Eq. (18), Fact VIII.2, and Lemma VIII.4, together with integration by parts, allow us to conclude Theorem VII.3 in the special case where $\mathcal{M}$ is orthogonal.

B. Proof of Theorem V.1

For convenience, Theorem V.1 is restated below. Recall from the discussion in Section IV-C that this can be thought of as the "quantum analogue" of binomial tail bounds:

Theorem V.1. *Fix any POVM $\mathcal{M}$. There exists an absolute constant $c'' > 0$ such that for any $t > \Omega(\epsilon^2/d^{1.99})$, we have*

$$\mathop{\mathbb{P}}_{\mathbf{U},\mathbf{U}'\sim\mathcal{D}}\left[\left|\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}\right| > t\right] \leq \exp\left(-c''\left\{\frac{d^3t^2}{\epsilon^4} \wedge \frac{d^2t}{\epsilon^2}\right\}\right)$$

Proof of Theorem V.1. Define $G : \mathbf{U} \mapsto \mathbb{E}_{x\sim p}\left[g_{\mathcal{M}}^{\mathbf{U}}(x)^2\right]^{1/2}$. Fix any $\mathbf{U}'$ and consider the function $F_{\mathbf{U}'} : \mathbf{U} \mapsto \phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$. Note

$$\mathop{\mathbb{E}}_{\mathbf{U}}[F_{\mathbf{U}'}(\mathbf{U})] = \mathop{\mathbb{E}}_{x\sim p}[g_{\mathcal{M}}^{\mathbf{U}'}(x)\cdot\mathop{\mathbb{E}}_{\mathbf{U}}[g_{\mathcal{M}}^{\mathbf{U}}(x)]] = 0$$

by Part (I) of Fact IV.4. Next, note that by Cauchy-Schwarz,

$$F_{\mathbf{U}'}(\mathbf{U}) - F_{\mathbf{U}'}(\mathbf{V}) \leq \mathop{\mathbb{E}}_{x\sim p}\left[\left(g_{\mathcal{M}}^{\mathbf{U}}(x) - g_{\mathcal{M}}^{\mathbf{V}}(x)\right)^2\right]^{1/2}\cdot G(\mathbf{U}'),$$

[3] Note that this looks different from the statement in [43] only because they work with normalized trace $\mathrm{tr}(\cdot) \triangleq \frac{1}{d}\mathrm{Tr}(\cdot)$.

which we show in the full proof of Theorem VII.3 is $O(\epsilon/\sqrt{d})\cdot G(\mathbf{U}')$-Lipschitz. So for any fixed $\mathbf{U}'$, Theorem VIII.1 implies

$$\mathop{\mathbb{P}}_{\mathbf{U}}[|F_{\mathbf{U}'}(\mathbf{U})| > t] \leq \exp\left(-C \cdot \frac{d^2 t^2}{\epsilon^2 G(\mathbf{U}')^2}\right)$$

for some absolute constant $C > 0$. We would like to integrate over $\mathbf{U}'$ to get a tail bound for $\phi_{\mathcal{M}}^{\mathbf{U},\mathbf{U}'}$ as a function of both $\mathbf{U}$ and $\mathbf{U}'$.

To this end, we can apply Fact I.4 to the random variable $Y \triangleq G(\mathbf{U}') \in [0, \epsilon]$ to conclude the proof (see the full version for details). □

ACKNOWLEDGMENTS

The authors would like to thank Ofer Zeitouni for pointing out the existence of Theorem VIII.1, and Robin Kothari and Jeongwan Haah for helpful preliminary discussions about quantum tomography.

REFERENCES

[1] R. O'Donnell and J. Wright, "Quantum spectrum testing," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 529–538.

[2] ——, "Efficient quantum tomography," in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016, pp. 899–912.

[3] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5628–5641, 2017.

[4] R. O'Donnell and J. Wright, "Efficient quantum tomography ii," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2017, pp. 962–974.

[5] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner, "Measuring quantum entropy," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 3012–3016.

[6] C. Bădescu, R. O'Donnell, and J. Wright, "Quantum state certification," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 503–514.

[7] J. Wright, "How to learn a quantum state," Ph.D. dissertation, Carnegie Mellon University Pittsburgh, PA, 2016.

[8] J. Cotler and F. Wilczek, "Quantum overlapping tomography," *Physical Review Letters*, vol. 124, no. 10, p. 100401, 2020.

[9] L. Paninski, "A coincidence-based test for uniformity given very sparsely sampled discrete data," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4750–4755, 2008.

[10] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[11] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[12] S.-O. Chan, I. Diakonikolas, P. Valiant, and G. Valiant, "Optimal algorithms for testing closeness of discrete distributions," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 1193–1203.

[13] I. Diakonikolas, D. M. Kane, and V. Nikishkin, "Testing identity of structured distributions," in *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 1841–1854.

[14] C. L. Canonne, I. Diakonikolas, T. Gouleakis, and R. Rubinfeld, "Testing shape restrictions of discrete distributions," *Theory of Computing Systems*, vol. 62, no. 1, pp. 4–62, 2018.

[15] P. Brouwer and C. Beenakker, "Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems," *Journal of Mathematical Physics*, vol. 37, no. 10, pp. 4904–4934, 1996.

[16] C. W. Beenakker, "Random-matrix theory of quantum transport," *Reviews of modern physics*, vol. 69, no. 3, p. 731, 1997.

[17] Y. M. Blanter and M. Büttiker, "Shot noise in mesoscopic conductors," *Physics reports*, vol. 336, no. 1-2, pp. 1–166, 2000.

[18] B. Khoruzhenko, D. Savin, and H.-J. Sommers, "Systematic approach to statistics of conductance and shot-noise in chaotic cavities," *Physical Review B*, vol. 80, no. 12, p. 125301, 2009.

[19] V. Al Osipov and E. Kanzieper, "Statistics of thermal to shot noise crossover in chaotic cavities," *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 47, p. 475101, 2009.

[20] E. Meckes and M. Meckes, "Spectral measures of powers of random matrices," *Electronic communications in probability*, vol. 18, 2013.

[21] A. Montanaro and R. deWolf, "A survey of quantum property testing," *Theory of Computing*, no. Graduate Surveys, 2016.

[22] A. Chefles, "Quantum state discrimination," *Contemporary Physics*, vol. 41, no. 6, pp. 401–424, 2000.

[23] K. M. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, no. 1, pp. 251–283, 2008.

[24] S. M. Barnett and S. Croke, "Quantum state discrimination," *Advances in Optics and Photonics*, vol. 1, no. 2, pp. 238–278, 2009.

[25] S. T. Flammia and Y.-K. Liu, "Direct fidelity estimation from few pauli measurements," *Physical review letters*, vol. 106, no. 23, p. 230501, 2011.

[26] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators," *New Journal of Physics*, vol. 14, no. 9, p. 095022, 2012.

[27] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, "Practical characterization of quantum devices without tomography," *Physical Review Letters*, vol. 107, no. 21, p. 210404, 2011.

[28] L. Aolita, C. Gogolin, M. Kliesch, and J. Eisert, "Reliable quantum certification of photonic state preparations," *Nature communications*, vol. 6, no. 1, pp. 1–8, 2015.

[29] N. Yu, "Quantum closeness testing: A streaming algorithm and applications," 2019.

[30] C. L. Canonne, "A survey on distribution testing," 2017.

[31] O. Goldreich, *Introduction to property testing*. Cambridge University Press, 2017.

[32] C. L. Canonne, D. Ron, and R. A. Servedio, "Testing probability distributions using conditional samples," *SIAM Journal on Computing*, vol. 44, no. 3, pp. 540–616, 2015.

[33] S. Chakraborty, E. Fischer, Y. Goldhirsh, and A. Matsliah, "On the power of conditional samples in distribution testing," *SIAM Journal on Computing*, vol. 45, no. 4, pp. 1261–1296, 2016.

[34] C. Canonne, D. Ron, and R. A. Servedio, "Testing equivalence between distributions using conditional samples," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 1174–1192.

[35] J. Acharya, C. L. Canonne, and G. Kamath, "A chasm between identity and equivalence testing with conditional queries," *arXiv preprint arXiv:1411.7346*, 2014.

[36] R. Bhattacharyya and S. Chakraborty, "Property testing of joint distributions using conditional samples," *ACM Transactions on Computation Theory (TOCT)*, vol. 10, no. 4, pp. 1–20, 2018.

[37] G. Kamath and C. Tzamos, "Anaconda: A non-adaptive conditional sampling algorithm for distribution testing," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2019, pp. 679–693.

[38] L. LeCam, "Convergence of estimates under dimensionality restrictions," *The Annals of Statistics*, vol. 1, no. 1, pp. 38–53, 1973.

[39] B. Moran, S. Howard, and D. Cochran, "Positive-operator-valued measures: a general setting for frames," in *Excursions in Harmonic Analysis, Volume 2*. Springer, 2013, pp. 49–64.

[40] G. Chiribella, G. M. D'Ariano, and D. Schlingemann, "Barycentric decomposition of quantum measurements in finite dimensions," *Journal of mathematical physics*, vol. 51, no. 2, p. 022111, 2010.

[41] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge university press, 2010, vol. 118.

[42] D. Weingarten, "Asymptotic behavior of group integrals in the limit of infinite rank," *Journal of Mathematical Physics*, vol. 19, no. 5, pp. 999–1001, 1978.

[43] B. Collins, "Moments and cumulants of polynomial random variables on unitarygroups, the itzykson-zuber integral, and free probability," *International Mathematics Research Notices*, vol. 2003, no. 17, pp. 953–982, 2003.