

Sample-optimal tomography of quantum states

Jeongwan Haah,¹ Aram W. Harrow,¹ Zhengfeng Ji,^{2,3,4} Xiaodi Wu,¹ and Nengkun Yu^{2,5,6}

¹*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*

²*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada*

³*School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada*

⁴*State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China.*

⁵*Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Ontario, Canada*

⁶*Department of Mathematics & Statistics, University of Guelph, Guelph, Ontario, Canada*

(Dated: 7 August 2015)

It is a fundamental problem to decide how many copies of an unknown mixed quantum state are necessary and sufficient to determine the state. Previously, it was known only that estimating states to error ϵ in trace distance required $O(dr^2/\epsilon^2)$ copies for a d -dimensional density matrix of rank r . Here, we give a theoretical measurement scheme (POVM) that requires $O((dr/\delta)\ln(d/\delta))$ copies of ρ to error δ in infidelity, and a matching lower bound up to logarithmic factors. This implies $O((dr/\epsilon^2)\ln(d/\epsilon))$ copies suffice to achieve error ϵ in trace distance. For fixed d , our measurement can be implemented on a quantum computer in time polynomial in n .

INTRODUCTION

Given n copies of an unknown d -dimensional quantum state ρ , how accurately can ρ be estimated? This fundamental question arises both in quantum information theory and in the interpretation of experimental results. Since ρ has $d^2 - 1$ real parameters, it is reasonable to conjecture that $\Theta(d^2)$ measurements are necessary and sufficient to estimate ρ to constant accuracy. On the other hand, even distinguishing a fair coin from a coin biased to obtain heads with probability $1/2 + \epsilon$ requires $\Omega(1/\epsilon^2)$ measurements. In this paper we show that the number of copies required to estimate ρ with precision ϵ scales roughly with both d^2 and $1/\epsilon^2$. More precisely, if the fidelity goal is $1 - \delta$, we prove an $\Omega(d^2/\delta)$ lower bound and an $O((d^2/\delta)\ln(d/\delta))$ upper bound on the number of required copies. When the state ρ is guaranteed to have rank $\leq r$ we show an $O((dr/\delta)\ln(d/\delta))$ upper bound and an $\Omega((dr/\delta)/\ln(d/r\delta))$ lower bound.

Notation — We use the convention that $\Omega(x)$ means a function that is asymptotically $\geq c_1x$ for a constant $c_1 > 0$, $O(x)$ means $\leq c_2x$ for a constant $c_2 > 0$ and $\Theta(x)$ means both $O(x)$ and $\Omega(x)$. Notation $\tilde{O}()$ means that we neglect the \ln term. \ln and \exp are base- e . The fidelity of two quantum states ρ, σ is $F(\rho, \sigma) := \text{tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}}$, the “infidelity” is $1 - F^2$, represented by δ , and their trace distance is $T(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_1$, represented by ϵ . These are related by [1]

$$1 - F \leq T \leq \sqrt{1 - F^2}. \quad (1)$$

Accuracy measure — We derive an upper bound in terms of fidelity and a lower bound in terms of trace distance, in each case implying a near-optimal bound in terms of the other quantity. Here we discuss why fidelity is in many ways a natural quantity for tomography [2]. Tomography is essentially a state discrimination procedure where one distinguishes $\rho^{\otimes n}$ from $\sigma^{\otimes n}$. The sta-

tistical distinguishability of these states is measured by the trace distance $T_n = T(\rho^{\otimes n}, \sigma^{\otimes n})$, which is in general much larger than $T(\rho, \sigma)$; this amplification is what enables the tomography. The asymptotic behavior of T_n can be quantified as

$$\frac{1}{2}F(\rho, \sigma)^{2n} \leq 1 - T_n \leq F(\rho, \sigma)^n$$

by Eq. (1) and $F(\rho^{\otimes n}, \sigma^{\otimes n}) = F(\rho, \sigma)^n$. This means that $\ln(1/F)$ or infidelity gives nearly sharp bounds on the rate at which T_n converges to 1; the actual rate¹ is between $\ln(1/F)$ and $2\ln(1/F)$. In particular, for fixed d , the state discrimination is possible to infidelity δ using $n = \Theta(1/\delta)$ copies. Our upper bound on n in terms of fidelity proves that the POVM we will present in this paper indeed accomplishes the discrimination task using $n = \tilde{O}(1/\delta)$ copies. On the contrary, the corollary upper bound in terms of trace distance sometimes over-estimates the sufficient number of samples by an unbounded amount. As a simple example, consider qubit states

$$\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \sigma = \begin{pmatrix} 1 - \epsilon & 0 \\ 0 & \epsilon \end{pmatrix},$$

between which the trace distance is ϵ and the infidelity is also ϵ . The trace distance bound only says $n = \tilde{O}(1/\epsilon^2)$ copies are sufficient to distinguish them, whereas the fidelity bound says $n = \tilde{O}(1/\epsilon)$ copies are sufficient.

Previous Results — Quantum state estimation has been extensively studied, going back at least to the work of Helstrom [5], Holevo [6] and others from around 1970. Many of the rigorous results are for the special cases when

¹ The exact scaling of $1 - T_n$ for large n is known to be C^n where $C = C(\rho, \sigma) = \inf_{0 \leq s \leq 1} \text{tr}(\rho^s \sigma^{1-s})$, and $-\log C$ is called the quantum Chernoff distance [3, 4].

$d = 2$ or $r = 1$, or give an uncontrolled or suboptimal d dependence (e.g. with n scaling as $f(d)/\delta$ for unknown f) or discuss related problems such as spectrum estimation, parameter estimation or determining the identity of a state drawn from a discrete set. In this paper we will consider optimal measurements (also called “collective” measurements) and will not discuss the extensive literature on independent or adaptive measurements.

For $d = 2$ (i.e. qubits), the optimal infidelity was shown in [7–11] to scale as $1/n$. This scaling was generalized to qudits in [12] (see also Section 6.4 of [13]), but with an uncontrolled dependence on d (i.e. n scales as $f(d)/\delta$ for unknown $f(\cdot)$); see also [14]. In many settings (e.g. minimax estimation) one can show that covariant measurements are optimal. If one further assumes that ρ is pure then the optimal estimation strategy has a simple form and n should scale as $\Theta(d/\delta)$ [6, 15]; see also [16] where further connections were made to cloning and de Finetti theorems.

Another major theme in recent work has been the study of various forms of restricted measurements, e.g. independent measurements with a limited number of measurement settings. Here a sequence of works [17–20] showed that $n = O(dr^2/\epsilon^2)$ copies are sufficient to obtain trace distance $\leq \epsilon$ with high probability.²

In many cases it is not necessary to determine the full state ρ but only to estimate some parameters of the state. This is an extremely general problem which includes results such as a quantum version of the Cramér-Rao bound [5, 21, 22]. One special case that uses similar representation-theory techniques to our work is the problem of spectrum estimation. Here, the optimal covariant measurement was described by Keyl and Werner [23], its large-deviation properties were derived in [24] (see also [25]), and it was analyzed further in [26, 27]. Ref. [27] in particular showed (among other results) that the Keyl-Werner algorithm required

$$\Omega\left(\frac{d^2}{\epsilon^2}\right) \leq n \leq O\left(\frac{d^2}{\epsilon^2} \ln \frac{d}{\epsilon}\right).$$

Our results improve the upper bound by using the same number of copies to obtain a full estimate of ρ instead of merely its spectrum. We also improve the lower bound by showing that it applies to *all* estimation strategies, not only the Keyl-Werner algorithm; on the other hand, our lower bound is for the harder problem of state estimation, while the lower bound of Ref. [27] is for the problem of spectrum estimation. We improve both bounds in the case when $r \ll d$.

The problem of quantum state estimation can be thought of as a special case of minimax estimation

(i.e. choosing an estimator that minimizes the expected loss when we maximize over input states) when the loss function is given by the infidelity. Other loss functions have also been considered [28, 29]. For example, with the 0-1 loss function (assuming ρ is drawn from a finite set) the goal is to maximize the probability of guessing ρ correctly. Here a powerful heuristic is to use the so-called “pretty good measurement” or PGM [30], whose error is never worse than twice that of the optimal measurement for any ensemble [31]. While the PGM requires a prior distribution, prior-free versions can also be constructed [32]. We will describe two closely related measurements in this paper: first one closely related to the PGM and then one (with roughly equivalent performance) that corresponds precisely to a PGM over an appropriately chosen “uniform” ensemble of density matrices. In each case, we analyze the measurements directly, without making use of the results of [31, 32] or other prior work.

STATE TOMOGRAPHY

Representation theory — The symmetry of our problem implies that our estimators should be invariant under permuting the n systems and covariant under collective rotation by elements of $U(d)$. More precisely, any estimator can be replaced by one that is invariant/covariant as described above without sacrificing any performance. Thus it is natural to make use of the representation theory of the symmetric and unitary groups.

Schur-Weyl duality is a statement regarding joint representations of a matrix group and the symmetric group. This is standard material [33] in representation theory, but for the reader’s convenience we explain parts that are relevant to our results. Consider the Hilbert space $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$ of n qudits of d -dimensions. This space admits representations of the general linear group $GL(d)$ and the symmetric group \mathbb{S}_n . The matrix group acts by simultaneous “rotation” as $U^{\otimes n}$ for any $U \in GL(d)$, and the symmetric group acts by permuting tensor factors. Concretely, a permutation $\pi \in \mathbb{S}_n$ is represented by

$$P_\pi = \sum_{\{j_i\}} |j_{\pi^{-1}(1)} j_{\pi^{-1}(2)} \cdots j_{\pi^{-1}(n)}\rangle \langle j_1 j_2 \cdots j_n|.$$

Two actions $U^{\otimes n}$ and P_π obviously commute with each other, and hence \mathcal{H} admits a representation of $G = GL(d) \times \mathbb{S}_n$. Generally, an irreducible representation (irrep) of G is given by the tensor product of an irrep of $GL(d)$ and an irrep of \mathbb{S}_n . For both groups, the irreps are specified by Young diagrams, or equivalently, partitions $\lambda = (\lambda_1, \dots, \lambda_n)$ of $n = \sum_i \lambda_i$, where λ is sorted to be non-increasing. The Schur-Weyl duality asserts that the decomposition of the space \mathcal{H} into irreps of G has a

² The earlier papers [17, 18] achieved $n = \tilde{O}(d^2 r^2 / \epsilon^2)$. The improved $n = O(dr^2 / \epsilon^2)$ performance is achieved by analyzing Theorem 2 of [20]; see Appendix.

simple structure. Namely,

$$(\mathbb{C}^d)^{\otimes n} = \bigoplus_{\lambda \vdash n} \Pi_\lambda (\mathbb{C}^d)^{\otimes n} = \bigoplus_{\lambda \vdash n} \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$$

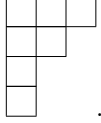
where \mathcal{Q}_λ is the irrep of $GL(d)$ and \mathcal{P}_λ is the irrep of \mathbb{S}_n corresponding to the Young diagram λ , and Π_λ is the projector onto the component $\mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$. Direct consequences of the decomposition are that

$$\Pi_\lambda X^{\otimes n} \Pi_\lambda \cong \mathbf{q}_\lambda(X) \otimes \text{id}_{\mathcal{P}_\lambda} \quad (2)$$

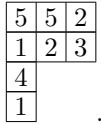
$$\Pi_\lambda X^{\otimes n} = X^{\otimes n} \Pi_\lambda \quad (3)$$

for any $d \times d$ matrix X , where we have defined $\mathbf{q}_\lambda(X)$ to mean the representing matrix of X . In fact, this is the main reason we are dealing with $GL(d)$, which is dense in the set of all matrices, rather than the more familiar $U(d)$. The space \mathcal{Q}_λ is also an irrep of the unitary group $U(d)$, and our discussion of Schur-Weyl duality could have been formulated entirely with $U(d)$; however, in this case X would be restricted to be unitary.

For our results it is important to understand the characters of the irrep \mathcal{Q}_λ of $GL(d)$. We identify a partition λ with a *Young diagram* in which there are λ_i boxes in the i^{th} row, e.g. the diagram for $\lambda = (3, 2, 1, 1)$ is as follows



Define a Young tableau T with shape λ to be a way of filling each box in λ with a number, e.g.



A *standard Young tableau* (SYT) is one in which each number from $1, \dots, n$ appears exactly once and numbers strictly increase from left to right and from top to bottom, while in a *semi-standard Young tableau* (SSYT) numbers weakly increase from left to right and strictly increase from top to bottom. Associated with a standard Young tableau T there are two subgroups A_T and B_T of \mathbb{S}_n . A_T is the set of all permutations that permute numbers within the rows of T , and B_T is the set of all permutations that permute numbers within the columns of T . The Young symmetrizer is then defined as

$$Y_T = \sum_{a \in A_T, b \in B_T} \text{sgn}(b) P_a P_b.$$

It can be shown that Y_T is proportional to an orthogonal projector, and it turns out that $Y_T \mathcal{H}$ is an irrep of $GL(d)$ and is isomorphic to \mathcal{Q}_λ . Since every T with the same λ gives rise to an isomorphic irrep of $GL(d)$, let us set

T to be the SYT where $1, 2, \dots, n$ are written in order from the upper left box towards right and down. To understand the basis of \mathcal{Q}_λ , let $|1\rangle, |2\rangle, \dots, |d\rangle$ form the standard orthonormal basis of \mathbb{C}^d . We may regard each basis vector $|E\rangle = |j_1, \dots, j_n\rangle$ of \mathcal{H} as a Young tableau E of shape λ . The Young symmetrizer Y_T projects this basis vector to a vector of \mathcal{Q}_λ . If there is any repetition along a column of E , then Y_T will annihilate it, thanks to the antisymmetric sum over P_b for $b \in B_T$. It follows that $\mathcal{Q}_\lambda = 0$ whenever λ has more than d rows. More precisely, let $\nu_i = \nu_i(E)$ denote the number of times the basis element $|i\rangle$ appears in the tableau E (also known as the *weight* of E), and let ν^\downarrow be the vector obtained by sorting ν into non-increasing order. Then Y_T annihilates E whenever $\sum_{i=1}^m \nu_i^\downarrow > \sum_{i=1}^m \lambda_i$ for some $m = 1, \dots, d-1$. The negation of the last condition is often denoted as

$$\nu \prec \lambda \Leftrightarrow \begin{cases} \sum_{i=1}^m \nu_i^\downarrow \leq \sum_{i=1}^m \lambda_i & (1 \leq m < d) \\ \sum_{i=1}^d \nu_i^\downarrow = \sum_{i=1}^d \lambda_i \end{cases}$$

and we say that ν is *majorized by* λ . The surviving tableaux E with $\nu(E) \prec \lambda$ form a spanning set for \mathcal{Q}_λ , or if we restrict to SSYT, they form a basis.

Now we can derive an expression for the characters of \mathcal{Q}_λ . Since $\text{tr } \mathbf{q}_\lambda(X)$ must be a function of eigenvalues of X , we may assume without loss of generality that X is a diagonal matrix with eigenvalues x_1, \dots, x_d associated with the standard basis elements $|1\rangle, \dots, |d\rangle$. The basis vectors of \mathcal{Q}_λ we just constructed are eigenvectors of diagonal $X^{\otimes n}$; $X^{\otimes n} Y_T |E\rangle = x_1^{\nu_1} \dots x_d^{\nu_d} Y_T |E\rangle =: x^\nu Y_T |E\rangle$, where $x^\nu := x_1^{\nu_1} \dots x_d^{\nu_d}$. Hence, the character value $\text{tr } \mathbf{q}_\lambda(X)$ is the sum of these eigenvalues:

$$\text{tr } \mathbf{q}_\lambda(X) = \sum_{\nu} K_{\lambda\nu} x^\nu =: s_\lambda(x). \quad (4)$$

Here $K_{\lambda\nu}$ is called the Kostka number and denotes the number of SSYT with weight ν and shape λ . One can show that $K_{\lambda\nu} > 0$ if and only if $\nu \prec \lambda$. We also define here the *Schur polynomial* $s_\lambda(x)$, which is a homogeneous polynomial in d variables of degree $\sum_i \nu_i = n$. Because the character $\text{tr } \mathbf{q}_\lambda(X)$ depends only on the eigenvalues, we will overload notation and denote this character also by $s_\lambda(X)$. For the same reason, it follows that $s_\lambda(XY) = s_\lambda(YX)$. The number of terms of the Schur polynomial is equal to

$$s_\lambda(\text{id}_d) = \text{tr } \mathbf{q}_\lambda(\text{id}_d) = \dim \mathcal{Q}_\lambda = \prod_{i < j} \frac{\lambda_i - \lambda_j + j - i}{j - i}.$$

Bound on Schur polynomials— Let ρ and σ be $d \times d$ density matrices. Suppose ρ has rank r . The central technical inequality in this paper is the following:

$$s_\lambda(\rho\sigma) \begin{cases} \leq (\dim \mathcal{Q}_\lambda) e^{-2nH(\bar{\lambda})} F^{2n} \\ = 0 & \text{if } \lambda_{r+1} > 0, \end{cases} \quad (5)$$

where

$$F = F(\rho, \sigma) = \text{tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \quad (6)$$

is the fidelity, and $H(\bar{\lambda}) = -\sum_i \bar{\lambda}_i \ln \bar{\lambda}_i$ is the Shannon entropy of $\bar{\lambda} = \lambda/n$.

Proof. Consider a positive semi-definite matrix X and a number $k \geq 0$. The largest term in the Schur polynomial $s_\lambda(X^k)$ at eigenvalues $x_1 \geq \dots \geq x_d \geq 0$ of X is

$$x_1^{k\lambda_1} \dots x_d^{k\lambda_d} = e^{-nkH(\bar{\lambda})} e^{-nkD(\bar{\lambda}||\bar{x})} (\text{tr } X)^{kn}$$

where $\bar{x} = (x_1, \dots, x_d)/\text{tr}(X)$, and $D(p||q) = \sum_i p_i \ln(p_i/q_i)$ is the relative entropy. This is because majorization implies that

$$\max_{\nu \prec \lambda} x^\nu = x^\lambda,$$

i.e. the maximum is attained by putting the largest number x_1 with the largest possible exponent $\nu_1 = \lambda_1$ and the second largest x_2 with $\nu_2 = \lambda_2$ and so on, subject to the majorization condition $\nu \prec \lambda$.

It follows that

$$s_\lambda(X^k) \leq \dim \mathcal{Q}_\lambda \cdot e^{-nkH(\bar{\lambda})} e^{-nkD(\bar{\lambda}||\bar{x})} (\text{tr } X)^{kn}. \quad (7)$$

Now, we set $X = \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}}$ and observe $s_\lambda(\rho\sigma) = s_\lambda(X^2)$. Using the fact that $D(\bar{\lambda}||\bar{x})$ is always non-negative and $= +\infty$ when the rank of $\bar{\lambda}$ is larger than that of \bar{x} , we arrive at Eq. (5) \square

Note that since $s_\lambda(\bar{\lambda})$ is a sum of non-negative terms, we have

$$s_\lambda(\bar{\lambda}) \geq e^{-nH(\bar{\lambda})}. \quad (8)$$

Tomography— Suppose we are given with $\rho^{\otimes n}$, n copies of an unknown density matrix ρ . What is the best strategy to learn about ρ ? The input state has a trivial symmetry \mathbb{S}_n under the permutations of the tensor factors. So, the POVM elements of the optimal strategy can be taken to commute with P_π without loss of generality. Additionally since we do not assume any distribution over ρ , our measurement should not perform differently when ρ is replaced by $U\rho U^\dagger$. This means that if M_σ is the outcome corresponding to σ then we should have

$$M_{U\sigma U^\dagger} = (U^\dagger)^{\otimes n} M_\sigma U^{\otimes n}.$$

These observations, along with the Schur-Weyl decomposition, motivate us to define positive semi-definite operators

$$M(\lambda, U) := \frac{\dim \mathcal{Q}_\lambda}{s_\lambda(\bar{\lambda})} \Pi_\lambda (U \bar{\lambda} U^\dagger)^{\otimes n} \Pi_\lambda, \quad (9)$$

for each unitary U and Young diagram λ that partitions n with at most d rows. As before, $\bar{\lambda}$ denotes the diagonal matrix with entries λ/n .

We first show that the $M(\lambda, U)dU$ constitute a POVM, where dU is the Haar probability measure on $\mathbb{U}(d)$. It suffices to check $\int dU M(\lambda, U) = \Pi_\lambda$, for $\sum_\lambda \Pi_\lambda = I$. Since $\int dU M(\lambda, U)$ is invariant under any unitary conjugation or permutation, we only need to check the traces of both sides.

$$\begin{aligned} \int dU \text{tr } M(\lambda, U) &= \frac{\dim \mathcal{Q}_\lambda \dim \mathcal{P}_\lambda}{s_\lambda(\bar{\lambda})} \int dU \text{tr } \mathbf{q}_\lambda(U \bar{\lambda} U^\dagger) \\ &= \frac{\dim \mathcal{Q}_\lambda \dim \mathcal{P}_\lambda}{s_\lambda(\bar{\lambda})} \int dU \text{tr } \mathbf{q}_\lambda(\bar{\lambda}) \\ &= \text{tr } \Pi_\lambda \end{aligned}$$

Note that $M(\lambda, U)$ is redundant; obviously $M(\lambda, U) = M(\lambda, e^{i\phi}U)$, and any degeneracy in λ renders some block of U ineffective. The redundancy is actually accounted for by the Haar measure, and thus will not concern us.

Next, we bound the probability of measuring $M(\lambda, U)$. Let $F = F(\rho, U \bar{\lambda} U^\dagger)$ be the fidelity. We claim

$$\text{tr}(M(\lambda, U) \rho^{\otimes n}) \leq (n+1)^{2dr} F^{2n}, \quad (10)$$

where r is the rank of ρ .

To show this, we need a bound on $\dim \mathcal{P}_\lambda$:

$$\dim \mathcal{P}_\lambda \leq e^{nH(\bar{\lambda})}. \quad (11)$$

This follows from

$$\dim \mathcal{P}_\lambda \prod_i \bar{\lambda}_i^{\lambda_i} \leq \frac{n!}{\prod_i \lambda_i!} \prod_i \bar{\lambda}_i^{\lambda_i} = \frac{n!}{n^n} \prod_i \frac{\lambda_i^{\lambda_i}}{\lambda_i!} \leq 1. \quad (12)$$

The first inequality is by the ‘‘hook length formula’’ [33]. For the last inequality we note that the function $f(z) = z \ln z - \ln \Gamma(z+1)$ satisfies $f(0) = 0$ and $f''(z) > 0$ for $z > 0$ [34]. Hence, $\sum_{i=1}^d f(\lambda_i)$ with $\sum_{i=1}^d \lambda_i = n$ is maximum if and only if $\lambda_1 = n$, in which case the inequality is saturated. Eqs. (8) and (11) now imply that

$$\begin{aligned} \text{tr}(M(\lambda, U) \rho^{\otimes n}) &= \frac{\dim \mathcal{Q}_\lambda \dim \mathcal{P}_\lambda}{s_\lambda(\bar{\lambda})} s_\lambda(\rho U \bar{\lambda} U^\dagger) \\ &\leq \dim \mathcal{Q}_\lambda \cdot e^{2nH(\bar{\lambda})} s_\lambda(\rho U \bar{\lambda} U^\dagger). \end{aligned}$$

By Eq. (5), this is nonzero only if $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_d = 0$. In this case, we have $\dim \mathcal{Q}_\lambda \leq (n+1)^{dr}$, and arrive at Eq. (10).

The output of our POVM is $\hat{\rho} = U \bar{\lambda} U^\dagger$. The probability of obtaining $\hat{\rho}$ where $\hat{\rho}$ has small fidelity, say infidelity δ , to the true state ρ can be estimated by integrating Eq. (10) over all pairs (λ, U) such that $F(\rho, U \bar{\lambda} U^\dagger)^2 \leq 1 - \delta$. Since $\sum_\lambda \int dU < (n+1)^d$, we see that

$$\text{Pr}[F(\hat{\rho}, \rho)^2 \leq 1 - \delta] \leq (n+1)^{3dr} e^{-n\delta}. \quad (13)$$

Therefore,

$$n = O\left(\frac{dr}{\delta} \ln \frac{d}{\delta}\right) \quad (14)$$

copies of ρ suffice to estimate ρ with error δ in infidelity with high probability. In terms of the trace distance, using Eq. (1) we have

$$\Pr \left[\frac{1}{2} \|\hat{\rho} - \rho\|_1 > \epsilon \right] \leq (n+1)^{3rd} e^{-n\epsilon^2}, \quad (15)$$

so the number of required copies scales as $\tilde{O}(dr/\epsilon^2)$. This is an asymptotic improvement in the number of needed copies of ρ over all previously considered POVM's for full state tomography [18–20], and as we will see below matches the lower bound up to a log factor.

Construction via PGM — Recall that given an ensemble $\{(p_1, \phi_1), \dots, (p_m, \phi_m)\}$, the PGM has measurement operators $M_i := \bar{\phi}^{-1/2} p_i \phi_i \bar{\phi}^{-1/2}$ with $\bar{\phi} := \sum_i p_i \phi_i$ [30]. A relevant ensemble for us is the one in which ϕ_i is equal to $\sigma_i^{\otimes n}$, and the index i should run over all state space; our ensemble is determined by n and a probability measure $d\sigma$ on the whole state space $\{\sigma\}$. Demanding the unitary invariance of $d\sigma$, we have

$$\begin{aligned} \bar{\phi} &= \int d\sigma \sigma^{\otimes n} = \sum_{\lambda} \frac{\int d\sigma s_{\lambda}(\sigma)}{\dim \mathcal{Q}_{\lambda}} \Pi_{\lambda}, \\ M_{\sigma} d\sigma &= \sum_{\lambda} \frac{\dim \mathcal{Q}_{\lambda}}{\mathbb{E} s_{\lambda}} \Pi_{\lambda} \sigma^{\otimes n} \Pi_{\lambda} d\sigma, \end{aligned} \quad (16)$$

where $\mathbb{E} s_{\lambda} = \int d\sigma s_{\lambda}(\sigma)$. It follows that the probability density of measuring M_{σ} given a state ρ of rank at most r is

$$\begin{aligned} \text{tr}(M_{\sigma} \rho^{\otimes n}) d\sigma &= \sum_{\lambda} \frac{(\dim \mathcal{Q}_{\lambda} \cdot \dim \mathcal{P}_{\lambda}) s_{\lambda}(\sigma \rho)}{\mathbb{E} s_{\lambda}} d\sigma \\ &\leq \sum_{\lambda: \lambda_{r+1}=0} \frac{(\dim \mathcal{Q}_{\lambda})^2}{e^{nH(\bar{\lambda})} \mathbb{E} s_{\lambda}} F^{2n} d\sigma \end{aligned}$$

where the inequality is by Eq. (5) and (11). This is the same scaling in n up to constants as Eq. (10), provided

$$e^{nH(\bar{\lambda})} \mathbb{E} s_{\lambda} \geq (nd)^{-O(dr)}.$$

Indeed, if we choose a uniform distribution over the simplex of spectra of σ , then

$$\begin{aligned} e^{nH(\bar{\lambda})} \int d\sigma s_{\lambda}(\sigma) &\geq e^{nH(\bar{\lambda})} \frac{\lambda_1! \cdots \lambda_d! (d-1)!}{(n+d-1)!} \\ &\geq (n+d)^{-d}, \end{aligned}$$

where in the first inequality we lower bound the Schur polynomial by its largest term, and in the second we use Eq. (12). We conclude that this PGM defined by the uniform spectrum distribution achieves the same bound on the sufficient number of copies for tomography.

LOWER BOUND

Our tomography scheme is the most precise up to logarithmic factors, among all possible measurement schemes

given n copies of unknown state ρ . We prove this using information theory.

Theorem 1. *Let $\epsilon \in (0, 1)$ and $\eta \in (0, 1)$. Suppose there exists a POVM $\{M_{\sigma} d\sigma\}$ on $(\mathbb{C}^d)^{\otimes n}$ such that for any d -dimensional density matrix ρ with rank $\leq r$,*

$$\int_{\frac{1}{2} \|\sigma - \rho\|_1 \leq \epsilon/2} d\sigma \text{tr}[M_{\sigma} \rho^{\otimes n}] \geq 1 - \eta. \quad (17)$$

Then,

$$n \geq C \frac{dr (1 - \epsilon)^2}{\epsilon^2 \ln(d/r\epsilon)}$$

for C a constant depending only on η . In addition, if $r = d$, then

$$n \geq C \frac{d^2}{\epsilon^2} (1 - \epsilon)^2.$$

The restriction that $\epsilon < 1$ is because outputting the constant estimate I/d will always achieve trace distance $\leq 1/2$. This theorem implies that achieving infidelity $\delta = 1 - F^2$ requires $n \geq \tilde{\Omega}(dr/\delta)$. For both trace distance and fidelity these lower bounds match our upper bounds up to the log factors.

Proof. We will show that any measurement satisfying (17) will imply the existence of a communication protocol that can reliably send a large message. Holevo's theorem [35] can then be used to obtain a lower bound on n .

Following convention, call the sender Alice and the receiver Bob. We will show in Lemma 2 below that there exists a states ρ_1, \dots, ρ_N each with rank $\leq r$ such that

$$\frac{1}{2} \|\rho_i - \rho_j\|_1 \geq \epsilon \quad \forall i \neq j. \quad (18)$$

The set $\{\rho_1, \dots, \rho_N\}$ is known as an ϵ -packing net. Fix such a net, along with a measurement $\{M_{\sigma} d\sigma\}$ satisfying (17).

We will now construct a communication protocol. Alice will choose a message $x \in [N] := \{1, \dots, N\}$ which she will encode by sending $\rho_x^{\otimes n}$. Bob will use the state estimation scheme $\{M_{\sigma}\}$ to attempt to guess x . If σ is within $\epsilon/2$ trace distance of some ρ_y then Bob will guess y . By (18), there is always at most one ρ_y satisfying this condition. If no such ρ_y exists, Bob will output failure. This results in the POVM with measurement outcomes

$$\tilde{M}_y = \int_{\frac{1}{2} \|\sigma - \rho_y\|_1 \leq \epsilon/2} d\sigma M_{\sigma} \quad (19)$$

$$\tilde{M}_{\text{fail}} = \text{id} - \sum_{y \in [N]} \tilde{M}_y. \quad (20)$$

Define $\Pr[y|x] = \text{tr}[\tilde{M}_y \rho_x^{\otimes n}]$. From (17) we have that $\Pr[x|x] \geq 1 - \eta$. In other words, Bob has a $\geq 1 - \eta$

chance of correctly decoding Alice's message. By Fano's inequality [36], this implies that

$$I(X : Y) \geq (1 - \eta) \ln(N) - \ln(2). \quad (21)$$

On the other hand, Holevo theorem [35] states that $I(X : Y) \leq \chi$ where χ is the Holevo information:

$$\chi = S\left(\frac{1}{N} \sum_{x \in [N]} \rho_x^{\otimes n}\right) - \frac{1}{N} \sum_{x \in [N]} S(\rho_x^{\otimes n}). \quad (22)$$

In Lemma 2 below we will argue that there exists a packing net with large N and small χ . Specifically, we will bound $\chi \leq n\chi_0$ where

$$\chi_0 = S(\mathbb{E}_U U \rho_x U^\dagger) - S(\rho_x),$$

for an appropriate Haar random unitary U , and prove $\chi_0 = \tilde{O}(\epsilon^2)$. This will imply that

$$n \geq \frac{(1 - \eta) \ln(N) - \ln(2)}{\chi_0}.$$

Our result then follows from Lemma 2 below. \square

Lemma 2. *There exist ϵ -packing nets I,II,III of d -dimensional states (i.e. satisfying (18)) characterized in the following table.*

	rank	$\chi_0/c \leq$	$c \ln N \geq$	restriction
I	r	$\epsilon^2 \ln(d/r\epsilon)$	rd	$\epsilon \leq 2^{-4}, r < d/3$
II	d	ϵ^2	d^2	$\epsilon \leq 2^{-3}, d \text{ even}$
III	r	$\ln(d/r)$	$rd(1 - \epsilon)$	$r < d(1 - \epsilon)/6$

where $c > 0$ is a sufficiently large constant; $c = 1000$ is good enough.

We remark that packing nets of size $\exp(\Omega(dr))$ for rank- r states have been achieved as early as 1981 [37, 38]; see also [39, 40] which used them for applications in communication complexity. These imply an $\Omega(dr)$ lower bound on the number of copies needed when ϵ is constant [39–41] and has been used in [18] to argue an $\tilde{\Omega}(r^2 d^2)$ lower bound on the number of copies needed for constant accuracy using adaptive Pauli measurements. Our main new contribution here is to analyze at the same time the Holevo capacity corresponding to these ensembles, in order to obtain bounds with simultaneously optimal scaling with r , d and ϵ .

Construction of nets

This subsection constitutes the proof of Lemma 2. To give some intuition for the construction, recall the arguments in the introduction for lower bounds of $\Omega(1/\epsilon^2)$ and $\Omega(d^2)$ (or $\Omega(dr)$ in the rank- r case). In terms of our

information theoretic strategy, these have two implications. The first one is that an ensemble of states that are contained in a radius t ball around a fixed full-rank state will have vanishing Holevo information in the limit $t \rightarrow 0$. In this regime, χ is analytic and has a local minimum at $t = 0$; thus, it should scale as $O(t^2)$ for small t .

The second one is that radius- t ball has volume that scales like t^D , where D is the dimension of the manifold of allowed states. For rank- r states this is $\Theta(dr)$. Even if our ensemble has small diameter (say t) if we demand precision that is smaller by a constant factor (say $t/3$) then there will be $\exp(\Omega(D))$ well-separated states. Indeed this is the approach used in [37, 38].

In order to find the states, we use a probabilistic existence argument. We will define a set of states $\rho_U = U \rho_I U^\dagger$ where U is any element of some subgroup $G \subseteq \mathbb{U}(d)$. Suppose

$$\Pr_U [\|\rho_U - \rho_I\|_1 \leq \epsilon] \leq \zeta$$

for Haar random $U \in G$. We wish to find a set $\{U_i\}$ of unitaries with cardinality at least $\lceil 1/\zeta \rceil$ such that $\|\rho_{U_i} - \rho_{U_j}\|_1 > \epsilon$ whenever $i \neq j$. This can be done inductively starting with the singleton $\{I\}$. Since Haar measure is left-invariant, $\Pr_U [\|\rho_U - \rho_V\|_1 \leq \epsilon] \leq \zeta$ for any unitary $V \in G$. If $m < \lceil 1/\zeta \rceil$ unitaries are chosen, the probability of choosing a unitary U such that ρ_U is ϵ -close to any previously chosen ρ_{U_i} is at most ζm , which is strictly smaller than 1. This proves the existence of one more desired unitary, and we obtain a set of $\lceil 1/\zeta \rceil$ elements. The probability ζ will be repeatedly estimated using the following fact.

Lemma 3 (Lemma III.5 of Ref. [42]). *Let P and Q be projectors on \mathbb{C}^d of rank p and q , respectively. Let $U \in \mathbb{U}(d)$ be Haar random. It holds that*

$$\begin{aligned} \forall z > 0 : \Pr_U \left[\frac{d}{pq} \text{tr} Q U P U^\dagger \geq 1 + z \right] &\leq \exp[-pqf(z)], \\ \forall z \in (0, 1) : \Pr_U \left[\frac{d}{pq} \text{tr} Q U P U^\dagger \leq 1 - z \right] &\leq \exp[-pqf(-z)], \end{aligned}$$

where

$$f(z) = z - \ln(1 + z) \geq \begin{cases} (1 + z)/2 & z \in [5, \infty) \\ (1 - \ln 2) z^2 & z \in (-1, 1] \\ z^2/2 & z \in (-1, 0] \end{cases}$$

Ref. [42] does not explicitly cover the $z > 1$ case for the first inequality, though it is implicitly covered in their proof. We will reprove the lemma in the appendix below.

Packing net I — Suppose $3r < d$. Let

$$U = \begin{pmatrix} I_r & 0 & 0 \\ 0 & A_{r \times r} & B_{r \times (d-2r)} \\ 0 & C_{(d-2r) \times r} & D_{(d-2r) \times (d-2r)} \end{pmatrix} \quad (23)$$

be a unitary matrix of $\mathbb{U}(d-r)$ with blocks as indicated, embedded into $\mathbb{U}(d)$. For $0 \leq t \leq 1$, define

$$\rho_{t,I} = \begin{pmatrix} (1-t^2)I_r/r & t\sqrt{1-t^2}I_r/r & 0 \\ t\sqrt{1-t^2}I_r/r & t^2I_r/r & 0 \\ 0 & 0 & 0_{d-2r} \end{pmatrix}, \quad (24)$$

$$\rho_{t,U} = U\rho_{t,I}U^\dagger.$$

It is a maximally mixed state on an r -dimensional subspace. The distance between $\rho_{t,U}$ satisfies

$$\|\rho_{t,U} - \rho_{t,I_{d-r}}\|_1 \geq \frac{t\sqrt{1-t^2}}{r} \text{tr } C^\dagger C \quad (25)$$

where C is as in Eq. (23). This is because $\|\rho_{t,U} - \rho_{t,I_{d-r}}\|_1 \geq \text{tr}[(\rho_{t,U} - \rho_{t,I_{d-r}})V]$ where

$$V = \begin{pmatrix} A & 0 & BF \\ 0 & E & 0 \\ C & 0 & DF \end{pmatrix}$$

and $E \in \mathbb{U}(r)$ and $F \in \mathbb{U}(d-2r)$ are arbitrary. Direct computation with optimized E and F proves the claim.

Lemma 4. *If $0 < t < 1/2$ and $r < d/3$, there exists a finite subset $\{U_i\} \subset \mathbb{U}(d-r)$ of cardinality $N \geq \exp(dr/54)$ such that $\|\rho_{t,U_i} - \rho_{t,U_j}\|_1 > t/4$ for any $i \neq j$. The Holevo χ_0 of $\{\rho_{t,U_i}\}_{i=1}^N$ fulfills $\chi_0 \leq t^2 \ln \frac{ed}{t^2 r}$.*

Proof. Lemma 3 states that if U is a Haar random unitary matrix of dimension k , then any $k_1 \times k_2$ subblock K of U satisfies

$$\Pr \left[\frac{k}{k_1 k_2} \text{tr}(K^\dagger K) < 1 - z \right] \leq \exp(-k_1 k_2 z^2/2)$$

for $z \in (0, 1)$. Eq. (25) says that $\|\rho_{t,I_{d-r}} - \rho_{t,U}\|_1 \leq t/4$ implies $\frac{d-r}{r(d-2r)} \text{tr } C^\dagger C \leq \frac{1}{\sqrt{3}} < 1 - \frac{1}{3}$. Therefore,

$$\Pr[\|\rho_{t,I_{d-r}} - \rho_{t,U}\|_1 \leq t/4] \leq e^{-r(d-2r)/18} < e^{-rd/54},$$

and we resort to the probabilistic existence argument.

Next, we estimate the Holevo information χ . Since U is unitary, we have $S(\rho_{t,U}) = S(\rho_{t,I_{d-r}}) = \ln r$. By the concavity of entropy, the ensemble average may be replaced with $\bar{\rho}_t = \int dU \rho_{t,U}$, only to increase the entropy. By Schur's lemma, the matrix $\bar{\rho}_t$ is diagonal, and has entropy

$$S(\bar{\rho}_t) = H(t^2) + (1-t^2) \ln r + t^2 \ln(d-r),$$

where $H(t^2) = -t^2 \ln(t^2) - (1-t^2) \ln(1-t^2)$ is the binary entropy. Combining, we have $\chi/n \leq H(t^2) + t^2 \ln \frac{d-r}{r}$. Using $H(z) \leq z \ln(e/z)$, we finish the proof. \square

Packing nets II & III — Assume that d is an even number, and fix a projector $Q = \text{diag}(1, \dots, 1, 0, \dots, 0)$ of rank $r \leq d/2$. For any $d \times d$ unitary U and $0 \leq t \leq 1$, define

$$\tau_{t,U} = \frac{1+t}{2r} UQU^\dagger + \frac{1-t}{2(d-r)} (I_d - UQU^\dagger). \quad (26)$$

Given an ensemble $\{\tau_{t,U}\}$, the entropy of the ensemble average is certainly at most $\ln d$. The entropy of $\tau_{t,U}$ is equal to $H((1+t)/2) + \frac{1+t}{2} \ln r + \frac{1-t}{2} \ln(d-r)$, where $H(\cdot)$ is the binary entropy. Therefore, the Holevo χ_0 is bounded as

$$\chi_0 \leq \frac{1}{2} \ln \frac{d^2}{r(d-r)} + \frac{t}{2} \ln \frac{d-r}{r} - H\left(\frac{1+t}{2}\right). \quad (27)$$

Next, if A denotes the upper-left $r \times r$ and C the lower-left $(d-r) \times r$ submatrix of U , we have

$$\text{tr } AA^\dagger + \text{tr } CC^\dagger = r \quad (28)$$

$$\|\tau_{t,U} - \tau_{t,I_d}\|_1 \geq \left(\frac{1+t}{r} - \frac{1-t}{d-r} \right) \text{tr } CC^\dagger. \quad (29)$$

This follows from direct calculation of $\tau_{t,U} - \tau_{t,I}$ with observation that it has trace zero.

Lemma 5. *Suppose $r = d/2$. Then, there exists a finite subset $\{U_i\} \subset \mathbb{U}(d)$ of cardinality $N \geq \exp(d^2/32)$ such that $\|\tau_{t,U_i} - \tau_{t,U_j}\|_1 > t/2$ for any $i \neq j$. The Holevo χ_0 fulfills $\chi \leq t^2$.*

Proof. Eq. (27) becomes $\chi/n \leq \ln 2 - H((1+t)/2) \leq t^2$. Eq. (29) says that if $\|\tau_{t,U} - \tau_{t,I}\|_1 \leq t/2$, then $(4/d) \text{tr } CC^\dagger \leq 1/2$. Lemma 3 states that this happens with probability at most $\exp(-d^2/32)$. The probabilistic existence argument applies. \square

Lemma 6. *Set $t = 1$. Suppose $\epsilon \in (0, 1)$, and $r < d(1-\epsilon)/6$. Then, there exists a finite subset $\{U_i\} \subset \mathbb{U}(d)$ of cardinality $N \geq \exp((1-\epsilon)rd/2)$ such that $\|\tau_{1,U_i} - \tau_{1,U_j}\|_1 > 2\epsilon$ for any $i \neq j$. The Holevo χ_0 fulfills $\chi_0 \leq \ln(d/r)$.*

Proof. Eq. (27) becomes $\chi_0 \leq \ln(d/r)$. Eq. (29) says that if $\|\tau_{t,U} - \tau_{t,I}\|_1 \leq 2\epsilon$, then $\frac{d}{r^2} \text{tr } AA^\dagger \geq (1-\epsilon)d/r$, which is greater than 6 when $r < d(1-\epsilon)/6$. By Lemma 3, this happens with probability at most $\exp(-r^2(1-\epsilon)d/2r) = \exp(-rd(1-\epsilon)/2)$. The probabilistic existence argument applies. \square

IMPLEMENTATION ON A QUANTUM COMPUTER

In this section we informally describe how our tomography strategy can be implemented in time $n^{O(dr)}$ on a quantum computer.

Our measurement involves a POVM with a continuously infinite number of outcomes. However, it can be

approximated with a finite POVM using ideas from [43]. The first step is to measure λ , as proposed by Keyl-Werner [23]. This can be done efficiently using the Schur transform [44] or the quantum Fourier transform over the symmetric group [45, 46].

Next, we would like to find a collection of unitaries U_1, \dots, U_m such that

$$\frac{1}{m} \sum_{i=1}^m M(\lambda, U_i) \approx \Pi_\lambda.$$

This can be done by choosing $m = \tilde{O}(\dim \mathcal{Q}_\lambda / \epsilon^2)$ random unitaries, as proven in [43], which in turn was based on [47]. The resulting measurement can be implemented by the isometry

$$V = m^{-1/2} \sum_{i=1}^m \sqrt{M(\lambda, U_i)} \otimes |i\rangle.$$

Using the Schur transform, this reduces to performing the isometry

$$\tilde{V} = C \sum_{i=1}^m \sqrt{\mathbf{q}_\lambda(U_i \bar{\lambda} U_i^\dagger)} \otimes |i\rangle,$$

where C is a normalizing constant. This isometry can be implemented using $O((\dim \mathcal{Q}_\lambda)^2 m^2)$ gates [48], which is $\tilde{O}(n^{2dr} / \epsilon^2)$.

We conjecture that run-time $\text{poly}(n, d, \log(1/\epsilon))$ is possible, but do not know how to achieve this, even in the relatively simple case of $r = 1$.

RECENT RELATED WORK

Independent of this paper, another work has achieved similar results. Ref. [41] analyzes the Keyl measurement strategy [14] as well as the measurement proposed in this paper, and shows that each requires $O(d/\gamma^2)$ copies in order to achieve expected 2-norm distance γ . This implies an $O(dr/\epsilon^2)$ upper bound for trace distance, which improves on our result for trace distance by removing the \ln term. However, the result does not obviously imply our fidelity bound, which appears to be incomparable to theirs. They also observe a lower bound of $\Omega(dr)$ for constant ϵ using packing nets; our use of Holevo's theorem is what lets us combine this with the $\Omega(1/\epsilon^2)$ bound.

DISCUSSION

Both POVMs in Eqs. (9) and (16) are inspired by the pretty good measurement, and indeed the measurement operator corresponding to the estimate σ is like a distorted version of $\sigma^{\otimes n}$. Variants of the PGM have been proposed in which the measurement operators are

distorted versions of higher powers of the state $p_i \sigma_i$, i.e. $M_i = X^{-1/2} (p_i \sigma_i)^k X^{-1/2}$ where $X \equiv \sum_i (p_i \sigma_i)^k$. When $k = 1$ this is the PGM, but the cases $k = 2$ and $k = 3$ have also been found useful in specific settings; see [49] for a review. If we take $k \rightarrow \infty$ here then this corresponds precisely to the Keyl “rotated-highest-weight” strategy. It is possible that this framework could be used to formally compare the performance of these different strategies.

Even though the sample complexity of the quantum tomography problem is nearly resolved here, many open questions remain. Can this measurement be made efficient? How well can product or adaptive measurements do? What is the rate of convergence to the Local Asymptotic Normality approximation of [10]?

ACKNOWLEDGMENTS

We thank Robin Blume-Kohout, Steve Flammia, Masahito Hayashi, Debbie Leung, and John Watrous for discussions. We also thank Ryan O'Donnell and John Wright for sharing their draft of [41] with us. JH is supported by Pappalardo Fellowship in Physics at MIT. AWH was funded by NSF grants CCF-1111382 and CCF-1452616 and ARO contract W911NF-12-1-0486. ZJ and NY's research was supported by NSERC, NSERC DAS, CRC, and CIFAR. XW's research was funded by ARO contract W911NF-12-1-0486 and by the NSF Waterman Award of Scott Aaronson. Part of the research was conducted when XW was visiting Institute for Quantum Computing (IQC), University of Waterloo and XW thanks IQC for its hospitality.

Appendix: Overlap of random projectors

Here, we provide a self-contained proof of Lemma 3 (Lemma III.5 of Ref. [42]). We follow the ideas of Ref. [42] and [50].

Lemma 7. *Let \mathcal{D} be the set of all $d \times d$ normalized density matrices of rank p , and Δ be the set of all probability vectors η of length p . Suppose \mathcal{D} has a $\mathbb{U}(d)$ -invariant probability measure $d\rho$. Then, there exists a permutation-symmetric probability measure $d\eta$ on Δ such that*

$$\iint d\eta dU f(U\eta U^\dagger) = \int d\rho f(\rho)$$

for any continuous function f on \mathcal{D} where dU is the normalized Haar measure on $\mathbb{U}(d)$, and η in between U and U^\dagger denotes the diagonal matrix with entries $(\eta_1, \dots, \eta_p, 0, \dots, 0)$.

This means that the eigenvalues and the eigenvectors can be treated as if they were “independent random variables.” Strictly speaking, $d\eta$ and dU are *not* derived from

ρ ; we just find that they induce the measure $d\rho$ on \mathcal{D} by the map $(\eta, U) \mapsto U\eta U^\dagger$.

Proof. Since *sorted* eigenvalues are continuous functions of the matrix, we have a map $\lambda : \mathcal{D} \rightarrow \Delta^\downarrow$, which induces a measure $d\lambda$ on Δ^\downarrow , the set of all sorted non-negative p real numbers summing to 1. The defining equation for the induced measure is $\int d\rho g(\lambda(\rho)) = \int d\lambda g(\lambda)$ for any continuous function g . Here, we have identified a vector with a diagonal matrix padded with $(d-p)$ zeros. Define

$$\bar{f}(\rho) = \int dU f(U\rho U^\dagger)$$

so that $\bar{f}(\rho) = \bar{f}(V\rho V^\dagger)$ for any $V \in \mathbb{U}(d)$. Since $d\rho$ is unitary invariant, $\int d\rho f(\rho) = \int d\rho f(U\rho U^\dagger)$. Integrating the both sides over U , $\int d\rho f(\rho) = \int dU \int d\rho f(U\rho U^\dagger) = \int d\rho \bar{f}(\rho)$. (All spaces are compact, so integration order never matters.) We can now prove an analogous version of the lemma for Δ^\downarrow :

$$\begin{aligned} \int d\rho f(\rho) &= \int d\rho \bar{f}(\rho) = \int d\rho \bar{f}(\lambda(\rho)) \\ &= \int d\lambda \bar{f}(\lambda) = \int d\lambda dU f(U\lambda U^\dagger) \end{aligned}$$

In order to finish the proof, all we need is to divide Δ into $p!$ pieces, each of which is mapped to Δ^\downarrow by permuting components up to measure zero sets, and assign measure to each piece by $d\lambda/p!$. Thus defined $d\eta$ on Δ is permutation-invariant. \square

Lemma 8. *Let x_1, x_2, \dots be independent gaussian random variables with mean 0 and variance $\frac{1}{2}$. Let U be a Haar random unitary of dimension d , and P and Q be d -dimensional projectors of rank p and q , respectively. For any real number ξ , it holds that*

$$\mathbb{E}_{x_i} \exp \left[\xi \sum_{i=1}^{2pq} x_i^2 \right] \geq \mathbb{E}_U \exp [\xi d \operatorname{tr}(QUPU^\dagger)].$$

Proof. Consider $\mathbb{C}^{dp} = \mathbb{C}^d \otimes \mathbb{C}^p$, and define $Q' = Q \otimes I_p$ to be the projector of rank qp . Without loss of generality, we assume that P, Q are diagonal. The random tuple (x_1, \dots, x_{2dp}) has the probability density $\frac{1}{\pi^{dp}} \exp(-r^2) d^{2dp}x$ where $r^2 = \sum_{i=1}^{2dp} x_i^2$. This means in particular that the magnitude variable r and the direction variable $\hat{x} = (x_1, \dots, x_{2dp})/r$ are independent. The direction variable \hat{x} defines a normalized pure state $|\hat{x}\rangle$ on $\mathbb{C}^d \otimes \mathbb{C}^p$, and the sum $\sum_{i=1}^{2pq} \hat{x}_i^2$ can be regarded as the squared norm of $Q'|\hat{x}\rangle$.

$$\sum_{i=1}^{2pq} x_i^2 = r^2 \langle \hat{x} | Q' | \hat{x} \rangle = r^2 \operatorname{tr} Q\rho$$

where ρ is the reduced density matrix of $|\hat{x}\rangle$ on \mathbb{C}^d .

As a random variable, ρ defines a $\mathbb{U}(d)$ -invariant measure on the set of all density operators of rank at most

p . By Lemma 7, ρ may be replaced with a random vector variable η and a Haar random U . Due to the permutation invariance and the normalization, we have $\mathbb{E}\eta_i = \mathbb{E}\eta_j = 1/p$, so $\mathbb{E}_\eta \sum_i \eta_i |i\rangle \langle i| = P/p$.

By the convexity of \exp ,

$$\begin{aligned} \mathbb{E}_{x_i} \exp \left[\xi \sum_{i=1}^{2pq} x_i^2 \right] &= \mathbb{E}_r \mathbb{E}_\eta \mathbb{E}_U \exp [\xi r^2 \operatorname{tr} QUPU^\dagger] \\ &\geq \mathbb{E}_U \exp [\xi (\mathbb{E}_r r^2) \mathbb{E}_\eta \operatorname{tr} QUPU^\dagger] \\ &= \mathbb{E}_U \exp [\xi (dp) \operatorname{tr} QU(P/p)U^\dagger]. \end{aligned}$$

we complete the proof. \square

Proof of Lemma 3. Recall Markov's inequality: For non-negative real random variable X and $a > 0$, $\Pr[X \geq a] \leq \mathbb{E}X/a$. This is easily seen once we define $Y = a$ if $X \geq a$ and $Y = 0$ if $X < a$, so $Y \leq X$. Then, $\Pr[X \geq a] = \Pr[Y = a] = \mathbb{E}Y/a \leq \mathbb{E}X/a$.

Let us abbreviate $\frac{d}{pq} \operatorname{tr} QUPU^\dagger$ as Z . For any $\xi > 0$ and $z > 0$,

$$\begin{aligned} \Pr[Z \geq 1+z] &= \Pr[e^{\xi Z} \geq e^{\xi(1+z)}] \\ &\leq \mathbb{E}_U e^{\xi Z} e^{-\xi(1+z)} \\ &\leq \mathbb{E}_{x_i} \exp \left[\frac{\xi}{pq} \sum_{i=1}^{2pq} x_i^2 \right] e^{-\xi(1+z)} \\ &= e^{-\xi(1+z)} \left(1 - \frac{\xi}{pq} \right)^{-pq} \end{aligned}$$

The last equality is directly evaluated with PDF $\frac{1}{\sqrt{\pi}} e^{-z^2}$. The best bound is when $\xi = pqz/(1+z) > 0$. Substituting this value for ξ , we prove the first inequality in the theorem.

The opposite direction goes similarly. Let $\xi > 0$ and $z \in (0, 1)$.

$$\begin{aligned} \Pr[Z \leq 1-z] &= \Pr[e^{-\xi Z} \geq e^{-\xi(1-z)}] \\ &\leq \mathbb{E} e^{-\xi Z} e^{\xi(1-z)} \\ &\leq \mathbb{E} \exp \left[-\frac{\xi}{pq} \sum_{i=1}^{2pq} x_i^2 \right] e^{\xi(1-z)} \\ &= e^{\xi(1-z)} \left(1 + \frac{\xi}{pq} \right)^{-pq} \end{aligned}$$

The best bound is when $\xi = pqz/(1-z) > 0$. Substituting this value for ξ , we prove the second inequality in the theorem.

The last inequality can be proved by examining extreme values of, for example, $g(z) = z - \ln(1+z) - (1-\ln 2)z^2$. The minimum values in the range $z \in (-1, 1]$ occur at $z = 0, 1$, where $g(z) = 0$. \square

Appendix: Sample complexity in [51]

The previously best achievable sample complexity for state tomography was described in [51]. Their setting does not naturally translate into our framework, so for convenience we sketch here how that is achievable. First we restate one of their main theorems:

Theorem 9. *There are universal constants $C_1, C_2, C_3 > 0$ such that the following holds for any r, d . Let $a_1, \dots, a_m \in \mathbb{C}^d$ be independent standard Gaussian vectors; i.e. normalized such that $\mathbb{E}[|a_i\rangle\langle a_j|] = I_d \delta_{ij}$. If $m \geq C_1 dr$, then with probability $\geq 1 - e^{-C_2 m}$ our choice of a_1, \dots, a_m is “good” in a sense we will define below.*

For X a matrix, define $\mathcal{A}(X) = \sum_j \langle a_j | X | a_j \rangle |j\rangle \in \mathbb{R}^m$. Given a d -dimensional density matrix ρ , a vector $b \in \mathbb{R}^m$ and a noise parameter η , define σ be any minimum of the following convex program:

$$\min \|\sigma\|_1 \text{ subject to } \|\mathcal{A}(\sigma) - b\|_2 \leq \eta.$$

Suppose further that $\|\mathcal{A}(\rho) - b\|_2 \leq \eta$. If the vectors a_1, \dots, a_m are good, then we have

$$\|\rho - \sigma\|_2 \leq C_3 \frac{\eta}{\sqrt{m}}. \quad (30)$$

To translate this into a quantum measurement, observe that by the operator Chernoff bound [47], we have $\frac{1}{m} \sum_{i=1}^m |a_i\rangle\langle a_i| \approx I_d$ with high probability. (For the purpose of this analysis, we neglect the error here.) We can then define a POVM with elements $E_i = |a_i\rangle\langle a_i|/m$. Measuring this POVM yields outcome i with probability $p_i := \text{tr}[E_i \rho]$; in the notation of [51] we have $p = \mathcal{A}(\rho)/m$. We will define the vector b of observed probabilities by measuring n independent copies of ρ using this POVM. If the resulting vector of frequencies is f , i.e., outcome i occurs f_i times, then we define $b = \frac{m}{n} f$. Thus b is an unbiased estimator of $\mathcal{A}(\rho)$; i.e. $\mathbb{E}[b] = \frac{m}{n} \mathbb{E}[f] = \frac{m}{n} np = \mathcal{A}(\rho)$. We can also estimate the error by

$$\mathbb{E} \|b - \mathbb{E}[b]\|_2^2 = \frac{m^2}{n^2} \sum_{i=1}^m \text{Var}[f_i] \leq \frac{m^2}{n^2} \sum_{i=1}^m np_i = \frac{m^2}{n}.$$

We thus have $\eta \leq O(m/\sqrt{n})$ with high probability. According to (30) we then have $\|\rho - \sigma\|_2 \leq O(\sqrt{m/n}) = O(\sqrt{dr/n})$. Assuming, without loss of generality, that σ has rank $\leq r$ we can then bound $\|\rho - \sigma\|_1 \leq \sqrt{dr^2/n}$. In other words, trace-distance error ϵ can be achieved with $n = O(dr^2/\epsilon^2)$. While this bound is significantly worse than our bound of $\tilde{O}(dr/\epsilon^2)$, their approach does have the significant advantage of not requiring entangled measurements. The improved performance of our bound (as well as that of [41]) can be seen as the advantage that entangled measurements yield for tomography.

- [1] C. A. Fuchs and J. van de Graaf, *IEEE Trans. Inf. Theory* **45**, 1216 (1999), [quant-ph/9712042](#).
- [2] W. K. Wootters, *Phys. Rev. D* **23**, 357 (1981).
- [3] M. Nussbaum and A. Szkoła, *The Annals of Statistics* **37**, 1040 (2009), [quant-ph/0607216](#).
- [4] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, *Phys. Rev. Lett.* **98**, 160501 (2007), [quant-ph/0610027](#).
- [5] C. W. Helstrom, *Journal of Statistical Physics* **1**, 231 (1969).
- [6] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, Publications of the Scuola Normale Superiore (Scuola Normale Superiore, 2011).
- [7] E. Bagan, M. Baig, R. Muñoz Tapia, and A. Rodriguez, *Phys. Rev. A* **69**, 010304 (2004), [quant-ph/0307199](#).
- [8] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz Tapia, *Phys. Rev. A* **73**, 032301 (2006), [quant-ph/0510158](#).
- [9] M. Guță and J. Kahn, *Phys. Rev. A* **73**, 052108 (2006), [quant-ph/0512075](#).
- [10] M. Guță and J. Kahn, *Communications in Mathematical Physics* **277**, 127 (2008), [quant-ph/0608074](#).
- [11] M. Hayashi and K. Matsumoto, *Journal of Mathematical Physics* **49**, 102101 (2008), [quant-ph/0411073](#).
- [12] J. Kahn and M. Guță, *Communications in Mathematical Physics* **289**, 597 (2009), [0804.3876](#).
- [13] M. Hayashi, *Quantum information: an introduction* (Springer-Verlag, 2006).
- [14] M. Keyl, *Reveiw in Mathematical Physics* **18**, 19 (2006), [quant-ph/0412053](#).
- [15] M. Hayashi, *Journal of Physics A: Mathematical and General* **31**, 4633 (1998), [quant-ph/9704041](#).
- [16] G. Chiribella, in *Proceedings of the 5th conference on Theory of quantum computation, communication, and cryptography*, TQC'10 (Springer-Verlag, Berlin, Heidelberg, 2011) pp. 9–25, [1010.1875](#).
- [17] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Phys. Rev. Lett.* **105** (2010), [0909.3304](#).
- [18] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, *New J. Phys.* **14**, 095022 (2012), [1205.2300](#).
- [19] V. Voroninski, “Quantum tomography from few full-rank observables,” (2013), [1309.7669](#).
- [20] R. Kueng, H. Rauhut, and U. Terstiege, “Low rank matrix recovery from rank one measurements,” (2014), [1410.6913](#).
- [21] R. D. Gill and S. Massar, *Phys. Rev. A* **61**, 042312 (2000), [quant-ph/9902063](#).
- [22] M. Hayashi, *American Mathematical Society Translations 2*, **277**, 99 (2009), [quant-ph/0608198](#).
- [23] M. Keyl and R. F. Werner, *Phys. Rev. A* **64**, 052311 (2001), [quant-ph/0102027](#).
- [24] M. Hayashi and K. Matsumoto, *Phys. Rev. A* **66**, 022311 (2002), [quant-ph/0202001](#).
- [25] M. Christandl and G. Mitchison, *Commun. Math. Phys.* **261**, 789 (2006).
- [26] A. Childs, A. W. Harrow, and P. Wocjan, in *Proc. of STACS, LNCS*, Vol. 4393 (2007) pp. 598–609, [quant-ph/0609110](#).
- [27] R. O’Donnell and J. Wright, in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC ’15 (2015) pp. 529–538, [1501.05028](#).

- [28] R. D. Gill, “Conciliation of bayes and pointwise quantum state estimation: Asymptotic information bounds in quantum statistics,” (2005), [math/0512443](#).
- [29] F. Tanaka, “Quantum minimax theorem,” (2014), [1410.3639](#).
- [30] P. Hausladen and W. K. Wootters, *Journal of Modern Optics* **41**, 2385 (1994).
- [31] H. Barnum and E. Knill, *J. Math. Phys.* **43**, 2097 (2002), [quant-ph/0004088](#).
- [32] A. W. Harrow and A. J. Winter, *IEEE Trans. Inf. Theory* **58**, 1 (2012), [quant-ph/0606131](#).
- [33] W. Fulton and J. Harris, *Representation Theory: A first course*, Graduate Texts in Mathematics, Vol. 129 (Springer, 2004).
- [34] N. Batir, *Archiv der Mathematik* **91**, 554 (2008).
- [35] A. S. Holevo, *Problems of Information Transmission* **9**, 177 (1973).
- [36] R. M. Fano, *The transmission of information* (M.I.T. Press and John Wiley and Sons, New York and London, 1961).
- [37] S. J. Szarek, in *Proceedings of Research Workshop on Banach Space Theory*, edited by B.-L. Lin (The University of Iowa, 1981) pp. 169–185.
- [38] S. Szarek, *Acta Mathematica* **151**, 153 (1983).
- [39] A. Winter, *Quantum Inf. Comput.* **4**, 563 (2004), [quant-ph/0401060](#).
- [40] T. Lee, I. Villanueva, Z. Wei, and R. de Wolf, In Preparation (2015).
- [41] R. O’Donnell and J. Wright, “Efficient quantum tomography,” (2015), due to appear on the arXiv at the same time as our paper.
- [42] P. Hayden, D. W. Leung, and A. Winter, *Commun. Math. Phys.* **265**, 95 (2006), [quant-ph/0407049](#).
- [43] A. Winter, “Compression of sources of probability distributions and density operators,” (2002), [quant-ph/0208131](#).
- [44] D. Bacon, I. L. Chuang, and A. W. Harrow, in *Proc. of SODA* (2007) pp. 1235–1244, [quant-ph/0601001](#).
- [45] R. Beals, in *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation (STOC)* (ACM Press, El Paso, Texas, 1997) pp. 48–53.
- [46] A. W. Harrow, *Applications of coherent classical communication and Schur duality to quantum information theory*, Ph.D. thesis, M.I.T., Cambridge, MA (2005), [quant-ph/0512255](#).
- [47] R. Ahlswede and A. Winter, *IEEE Trans. Inf. Theory* **48**, 569 (2002), [quant-ph/0012127](#).
- [48] R. Iten, R. Colbeck, I. Kukuljan, J. Home, and M. Christandl, “Quantum circuits for isometries,” (2015), [1501.06911](#).
- [49] J. Tyson, *Phys. Rev. A* **79**, 032343 (2009), [0907.1884](#).
- [50] P. Hayden, D. Leung, P. W. Shor, and A. Winter, *Commun. Math. Phys.* **250**, 371 (2004), [quant-ph/0307104](#).
- [51] R. Kueng, H. Rauhut, and U. Terstiege, (2014), [1410.6913](#).