Sample-Optimal Tomography of Quantum States

Jeongwan Haah* Aram W. Harrow* Zhengfeng Ji† Xiaodi Wu‡ Nengkun Yu§

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

†Centre for Quantum Computation & Intelligent Systems, School of Software
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia
Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada
and
State Key Laboratory of Computer Science
Institute of Software, Chinese Academy of Sciences, Beijing, China

‡Department of Computer and Information Science, University of Oregon, Eugene, Oregon, USA

§Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada
Centre for Quantum Computation & Intelligent Systems
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia
and
Department of Mathematics & Statistics, University of Guelph, Guelph, Ontario, Canada

ABSTRACT

It is a fundamental problem to decide how many copies of an unknown mixed quantum state are necessary and sufficient to determine the state. This is the quantum analogue of the problem of estimating a probability distribution given some number of samples.

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 measurement scheme (POVM) that uses $O((dr/\delta)\ln(d/\delta))$ copies to estimate ρ to error δ in infidelity. This implies $O((dr/\epsilon^2)\cdot\ln(d/\epsilon))$ 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.

We also use the Holevo bound from quantum information theory to prove a lower bound of $\Omega(dr/\epsilon^2)/\log(d/r\epsilon)$ copies needed to achieve error ϵ in trace distance. This implies a lower bound $\Omega(dr/\delta)/\log(d/r\delta)$ for the estimation error δ in infidelity. These match our upper bounds up to log factors.

Our techniques can also show an $\Omega(r^2 d/\delta)$ lower bound for measurement strategies in which each copy is measured individually and then the outcomes are classically post-processed to produce an estimate. This matches the known achievability results and proves for the first time that such "product" measurements have asymptotically suboptimal scaling with d and r.

Categories and Subject Descriptors

F.2 [**ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY**]: Miscellaneous
; G.3 [**PROBABILITY AND STATISTICS**]: Probabilistic algorithms (including Monte Carlo)
; E.4 [**CODING AND INFORMATION THEORY**]: Data compaction and compression

General Terms

Theory

Keywords

Sample complexity, quantum state tomography, Schur-Weyl duality, Pretty Good Measurement

1. INTRODUCTION

Problem and Motivation.

The identification of the state of a quantum system is so fundamental that any physics experiment starts with it, and is sometimes entirely about it. Generally speaking, the preparation of an experimental setup must be accurate and consistent, and hence must involve some type of quantum state identification. For example, one may ask whether a state is close to a fixed target state (state verification), whether a state is pure, or whether a bipartite state is sufficiently entangled (property testing); see [MdW13] for a review.

We address the simplest and strongest form of the state identification problem, *(quantum) state tomography* or *(quantum) state estimation*, in which one is asked to output a classical description of the full state vector $|\psi\rangle \in \mathbb{C}^d$, or more generally the full density matrix $\rho \in \mathbb{C}^{d\times d}$ of the system, when given n identical copies of the quantum states. Assuming that ρ has rank $\le r$ lets us interpolate between the cases of pure states ($r = 1$, $\rho = |\psi\rangle\langle\psi|$) and general states ($r = d$).

Quantum state tomography can be viewed as the natural quantum extension of the classical problem of reconstructing probability distributions from independent samples. The latter problem is fundamental in many disciplines such as statistics and property testing, and remains an active research topic. Quantum state tomography, however, deals with a much harder scenario, in which any density operator contains both the spectrum information and the eigenbasis information. Moreover, it could also be viewed as a special and fundamental problem in quantum property testing, the study of which has recently attracted much attention [MdW13, OW15b]. Not only are the goals of quantum state tomography harder, but the choice of quantum measurements means that the set of available techniques is also more complicated. One of the fundamental questions is then the following:

Given d, r, δ, how many copies n are necessary and sufficient to output an estimate $\hat{\rho}$ with expected infidelity δ to the true state $\rho \in \mathbb{C}^{d\times d}$, given the promise that ρ has rank $\leq r$?

Since ρ has $O(rd)$ real parameters, it is reasonable to conjecture that $\Theta(rd)$ measurements are necessary and sufficient to estimate ρ to constant accuracy. To get a sense of the likely error scaling, even distinguishing a fair coin from a coin with heads probability $1/2 + \epsilon$ (i.e. for $d = 2$) requires $\Omega(1/\epsilon^2)$ measurements.

The accuracy is usually measured in the following two forms, discussed further in Section 1.2. The fidelity of two quantum states ρ, σ is $F(\rho,\sigma) := \operatorname{tr}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}$, the “infidelity” is $1 - F$, represented by δ, and their trace distance is $T(\rho,\sigma) := \frac{1}{2}\|\rho - \sigma\|_1$, represented by ϵ. These are related by [FvdG99].

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

1.1 Main Results

In this paper we show that the number of copies required to estimate ρ with precision ϵ (in trace distance) scales roughly with both dr and either $1/\delta$ (if error is measured in infidelity) or $1/\epsilon^2$ (if trace distance is used). More precisely, we show an $O((dr/\delta)\ln(d/\delta))$ upper bound and an $\Omega((dr/\delta)/\ln(d/r\delta))$ lower bound (which can be slightly improved to $\Omega(d^2/\delta)$ when $r = d$). When the accuracy is measured in trace distance ϵ, by Eq. (1), we can replace δ by ϵ^2. Both our upper and lower bounds improve upon previous results. We refer readers to Section 2 for the upper bounds and to Section 3 for the lower bounds. See also Table 1 for a summary.

General tomography schemes (including ours) require joint (or entangled) measurements over multiple copies of ρ, which is an arguably difficult task with current experimental technology. It would be desirable to have a tomography scheme that performs only independent measurements. In such a scheme each copy of the state is independently measured and the outcomes of the measurement are jointly processed by a classical computer. The best previously known scheme with independent measurements [KRT14] makes use of $n = O(dr^2/\epsilon^2)$ copies when ρ is guaranteed to have rank $\leq r$ and the accuracy ϵ is measured in trace distance. We demonstrate that any independent measurement scheme requires $\Omega(dr^2/\delta)$ copies when the accuracy δ is measured in infidelity (see Theorem 4). Our work thus shows for the first time that joint measurements are asymptotically more efficient for large d than independent measurements. Previously only the error scaling of joint measurements was known to outperform independent measurements, as we discuss below along with other prior work in Section 1.3.

We also discuss how to implement our measurement scheme on quantum computers in Section 4.

1.2 Discussion on 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 [Woo81]. Tomography is essentially a state discrimination procedure where one distinguishes $\rho^{\otimes n}$ from $\sigma^{\otimes n}$. The statistical 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 rate[1] 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[2]. 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, for any $0 < \xi < 1$,

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

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

1.3 Previous Results

Quantum state estimation has been extensively studied, going back at least to the work of Helstrom [Hel69], Holevo [Hol82] and others from around 1970. Many of the rigorous results are for the special cases when $d = 2$ 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 [BBMnTR04, BBG+06, GK06, GK08, HM08] to scale as $1/n$. This scaling was generalized to qudits in [KG09] (see also Section 6.4 of [Hay06]), but with an uncontrolled dependence on d (i.e.n scales as $f(d)/\delta$ for unknown $f(\cdot)$); see also [Key06]. 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 $\Theta(d/\delta)$ [Hay98, Hol82]; see

[1] The exact scaling of $1 - T_n$ for large n is known to be C^n where $C = C(\rho,\sigma) = \inf_{0\leq s\leq 1}\operatorname{tr}(\rho^s\sigma^{1-s})$, and $-\log C$ is called the quantum Chernoff distance [NS09, ACMnT+07].

[2] We adopt the convention that $\tilde{O}(\cdot)$ hides logarithmic terms.

Table 1: Conditions for the quantum state tomography with high success probability. δ denotes the accuracy goal measured in the infidelity $1 - F(\rho, \hat{\rho}) = 1 - \operatorname{tr}\sqrt{\sqrt{\rho}\,\hat{\rho}\sqrt{\rho}}$, and ϵ denotes that in the trace distance $T(\rho, \hat{\rho}) = \frac{1}{2}\|\rho - \hat{\rho}\|_1$. The upper bound in terms of the infidelity implies that in terms of trace distance; $n \leq O(d^2/\epsilon^2)\log(d/\epsilon)$. The lower bound in terms of the trace distance implies that in terms of infidelity; $n \geq \Omega(d^2/\delta)$. The previously known upper bound on n already used only independent measurements; thus our lower bounds show that this result was essentially optimal.

	Our result		Previous result
	for general $\rho \in \mathbb{C}^{d\times d}$	for ρ of rank at most r	
Sufficient	$n \leq O(d^2/\delta)\log(d/\delta)$	$n \leq O(rd/\delta)\log(d/\delta)$	$n \leq O(r^2 d/\epsilon^2)$ [KRT14] See App. E.
Necessary	$n \geq \Omega\left(d^2/\epsilon^2\right)$	$n \geq \Omega\left(rd/\epsilon^2\right)/\log(d/r\epsilon)$	$n \geq \Omega(1/\epsilon^2) + \tilde{\Omega}(rd)$ [FL11]
Necessary using independent measurements	$n \geq \Omega(d^3/\delta)$	$n \geq \Omega(dr^2/\delta)$	$n \geq \Omega(1/\delta^2 \log(1/\delta))$ See Sec. 1.3

also [Chi11] 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. Intermediate between independent measurements and unrestricted (also called "collective" or "entangled") measurements are *adaptive* measurements in which the copies of ρ are measured individually, but the choice of measurement basis can change in response to earlier measurements.

On the achievability side for independent measurements, a sequence of works [GLF$^+$10, FGLE12, Vor13, KRT14] showed that $n = O(dr^2/\epsilon^2)$ copies are sufficient to obtain trace distance $\leq \epsilon$ with high probability.[3] On the other hand, even for $d = 2$, adaptive and collective measurements are known to have asymptotically better error scaling, at least when measured in terms of infidelity. The usual intuition is that n should scale as $1/\delta^2$ for independent measurements and $1/\delta$ for adaptive or collective measurements; e.g. see [MRD$^+$13] for numerical evidence. Refs. [BBG$^+$06, HM08] showed that adaptive measurements could achieve $n = O(1/\delta)$ scaling. When a POVM contains a finitely many elements, the lower bound $1/\delta^2$ can be demonstrated by considering qubit tomography when the density matrix does not commute with POVM elements. We were unable to find a reference that proves this particular fact. [FBK15] gave an $\Omega(\frac{1}{\delta'^2\log(1/\delta')})$ lower bound for independent measurements with *relative entropy* δ' as accuracy measure without restriction that POVM should consist of finitely many elements. One major question left open by our work is to clarify the power of adaptive measurements, and in particular to determine whether there is an asymptotic separation between the power of adaptive and collective measurements.

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 [Hel69, GM00, Hay09] again going back to the early prehistory of quantum information. 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 [KW01], its large-deviation properties were derived in [HM02] (see also [CM06]), and it was analyzed further in [CHW07, OW15b]. Ref. [OW15b] 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. [OW15b] 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 [Gil05, Tan14]. 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 [Bel75b, Bel75a, HW94], whose error is never worse than twice that of the optimal measurement for any ensemble [BK02]. While the PGM requires a prior distribution, prior-free versions can also be constructed [HW12]. 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 [BK02, HW12] or other prior work.

Independent and simultaneous work.

Independent of this paper, another work has achieved similar results. Ref. [OW15a] analyzes the Keyl measurement strategy [Key06] as well as the first 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, and $O(dr/\delta^2)$ upper bound for infidelity, which improves on our result for trace distance by removing the log term. However, the result does

[3] The earlier papers [GLF$^+$10, FGLE12] achieved $n = \tilde{O}(d^2r^2/\epsilon^2)$. The improved $n = O(dr^2/\epsilon^2)$ performance is achieved by analyzing Theorem 2 of [KRT14]. This is not obvious from their theorem statement, but we explain the connection in Appendix E.

not imply our fidelity bound, which is incomparable to theirs. They also observe a lower bound of $\tilde{\Omega}(dr)$ for constant ϵ using packing nets; our use of Holevo's theorem allows us to derive a stronger $\tilde{\Omega}(dr/\epsilon^2)$ lower bound. Moreover, we also consider the lower bound of independent measurement schemes and the implementation of proposed measurements on quantum computers, which were not considered in [OW15a].

1.4 Techniques

The input to the problem is $\rho^{\otimes n}$ with no prior information about ρ except for its dimension d. The input has an obvious intrinsic symmetry $\mathbb{S}_n$ of permuting tensor factors. In addition, the measurement strategy should not perform differently for ρ and $U\rho U^\dagger$ for a unitary $U \in \mathbb{U}(d)$. This suggests that our POVM should be symmetric under $\mathbb{U}(d) \times \mathbb{S}_n$. It is natural to work in the basis where the symmetry action is block diagonal. Such a basis is called a Schur basis, and the decomposition of the total Hilbert space into irreducible representation (irrep) spaces $\Pi_\lambda(\mathbb{C}^d)^{\otimes n} = \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$ of $\mathbb{U}(d) \times \mathbb{S}_n$ is known as Schur-Weyl duality:

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \vdash n} \mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda, \tag{2}$$

where $\lambda = (\lambda_1, \ldots, \lambda_d)$ is a partition of $n = \sum_{i=1}^d \lambda_i$. We consider two POVM's on $(\mathbb{C}^d)^{\otimes n}$ which obey this symmetry:

$$\begin{aligned} M_1(\lambda, U)\mathrm{d}U &\propto \Pi_\lambda (U \mathrm{diag}(\lambda/n) U^\dagger)^{\otimes n} \Pi_\lambda \mathrm{d}U, \\ M_2(\eta)\mathrm{d}\eta &= \bar{\eta}^{-\frac{1}{2}} \eta^{\otimes n} \bar{\eta}^{-\frac{1}{2}} \mathrm{d}\eta \end{aligned}$$

where $\mathrm{d}U$ is the normalized Haar measure on $\mathbb{U}(d)$, $\mathrm{d}\eta$ is a $\mathbb{U}(d)$-invariant measure on the set of all states such that the distribution of the spectrum of η over the probability simplex is uniform, and $\bar{\eta} = \int \eta^{\otimes n} \mathrm{d}\eta$. The normalization for M_1 is such that $\int M_1(\lambda, U)\mathrm{d}U = \Pi_\lambda$. Each POVM is labeled by a state, $\sigma = U\bar{\lambda}U^\dagger$ or $\sigma = \eta$, which is our output $\hat{\rho} = \sigma$. For both of the POVM's we prove the concentration of the measurement outcome distribution, peaked around the true state ρ, yielding the upper bound on the sample-complexity of tomography. M_1 may be viewed as a two-stage measurement where the spectrum of ρ is measured first by Keyl-Werner scheme, and then the basis is estimated by a so-called *pretty good measurement* with the Haar random prior distribution. M_2 directly implements the pretty good measurement for a continuous ensemble of states of form $\eta^{\otimes n}$. The key ingredient in the concentration proof is the identification of the character $s_\lambda(\rho\sigma)$ of $\mathcal{Q}_\lambda$ in terms of fidelity $F(\rho, \sigma)$. See Section 2 for more details.

Lower bounds from information theory.

For lower bounds, we observe that any tomography scheme can be used to decode a classical message x that has been encoded as $\rho_x^{\otimes n}$, where $\rho_1, \ldots, \rho_N$ are states with

$$\min_{x \neq y} T(\rho_x, \rho_y) \geq \epsilon.$$

A classic result in quantum information theory is Holevo's 1973 bound on the capacity of a quantum channel to transmit classical information [Hol73]. If all message states are $O(\epsilon)$-close to one another, the Holevo capacity χ should vanish as $\epsilon \to 0$, and if the average state is an interior point then $\chi \sim \epsilon^2$. Meanwhile, the manifold of rank-r states is $\Theta(rd)$-dimensional, and in the interior, the ϵ-ball still contains $N := \exp(\Omega(rd))$ balls of radius $\epsilon/10$, which are distinguishable by the tomography procedure. These two observations lead us to the lower bound $\Omega(rd/\epsilon^2)$ that combines d and ϵ.

A similar argument (in Theorem 4) restricting to independent measurements leads to a lower bound of $\Omega(r^2 d/\delta)$. The idea there is that the random choice of eigenbasis in ρ is an additional source of noise, so that each measurement reveals only $O(1/r)$ bits of information about ρ. These results are discussed in more detail in Section 3.

1.5 Discussion and open questions

Our POVM constructions 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 [Tys09] for a review. If we take $k \to \infty$ 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 adaptive measurements do? What is the rate of convergence to the Local Asymptotic Normality approximation of [GK08]?

2. STATE TOMOGRAPHY

Schur-Weyl duality.

Given n i.i.d. samples of a classical probability distribution p, it is natural to analyze the outcomes in terms of their *type*, or empirical distribution. The Schur decomposition of $(\mathbb{C}^d)^{\otimes n}$ described in (2) can be thought of as a quantum method of types, and is similarly useful in analyzing $\rho^{\otimes n}$. The analogues of types are partitions $\lambda = (\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ and $\sum_i \lambda_i = n$.

We review the Schur decomposition in Appendix A. One useful piece of notation that will be introduced there is the Schur polynomial $s_\lambda(x)$, which (for λ defined as above) is a degree-n polynomial in d variables. It is also natural to define $s_\lambda(X) := s_\lambda(\mathrm{eig}(X))$ for any (not necessarily diagonalizable) matrix X, where $\mathrm{eig}(X)$ are the roots of the characteristic polynomial of X. Then we will see in the technical appendices that $s_\lambda(\rho)$ is proportional to the probability of observing partition λ given $\rho^{\otimes n}$ and $s_\lambda(\rho\sigma)$ is related to the probability that our measurement procedure outputs σ when the true state is ρ.

Bound on Schur polynomials.

Let ρ and σ be $d \times d$ density matrices. Suppose ρ has rank r. The central inequality in this paper (see Appendix B for a very short proof) is:

$$s_\lambda(\rho\sigma) \begin{cases} \leq (\dim \mathcal{Q}_\lambda) e^{-2nH(\bar{\lambda})} F^{2n} \\ = 0 \quad \text{if} \quad \lambda_{r+1} > 0, \end{cases} \tag{3}$$

where $F = F(\rho, \sigma) = \mathrm{tr}\,\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}$ is the fidelity and $H(\bar{\lambda}) = -\sum_i \bar{\lambda}_i \ln \bar{\lambda}_i$ is the Shannon entropy of $\bar{\lambda} = \lambda/n$. Note that since $s_\lambda(\bar\lambda)$ is a sum of non-negative terms, it is lower bounded by its largest term:

$$s_\lambda(\bar\lambda) \ge e^{-nH(\bar\lambda)}. \tag{4}$$

Tomography.

Since our input $\rho^{\otimes n}$ is symmetric under permutations of the tensor factors ($\mathbb{S}_n$), the POVM elements of the optimal strategy can be taken to commute with permutations 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

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

These observations, along with the Schur-Weyl decomposition (from Appendix A), motivate us to define positive semi-definite operators

$$M(\lambda, U) := \frac{\dim \mathcal{Q}_\lambda}{s_\lambda(\bar\lambda)} \Pi_\lambda (U\bar\lambda U^\dagger)^{\otimes n} \Pi_\lambda, \tag{5}$$

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

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

Note that $M(\lambda, U)$ is redundant; obviously we have $M(\lambda, U) = M(\lambda, e^{i\phi}U)$, 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. The following theorem is our achievability result on tomography.

Theorem 1. *Suppose a quantum state $\rho \in \mathbb{C}^{d\times d}$ has rank at most r. The measurement using the POVM $\{M(\lambda, U)\mathrm{d}U\}$ on $\rho^{\otimes n}$ outputs an estimate $\hat\rho = U\bar\lambda U^\dagger$ such that $F(\hat\rho, \rho) \ge 1 - \delta$ with probability at least $1 - (n+1)^{3dr} e^{-2n\delta}$, which is at least $2/3$ whenever $n \ge (10dr/\delta)\ln(d/\delta)$.*

In terms of the trace distance, using Eq. (1) we have

$$\Pr\left[\, \frac{1}{2}\|\hat\rho - \rho\|_1 > \epsilon \,\right] \le (n+1)^{3rd} e^{-n\epsilon^2}, \tag{6}$$

so the number of required copies scales as $\tilde O(dr/\epsilon^2)$. This is an asymptotic improvement in the number of copies of ρ over all previously considered POVM's for full state tomography [FGLE12, Vor13, KRT14], and as we will see below matches the lower bound up to log factors.

Proof. Let $F = F(\rho, U\bar\lambda U^\dagger)$ be the fidelity. We claim

$$\operatorname{tr}(M(\lambda, U)\rho^{\otimes n}) \le (n+1)^{2dr} F^{2n}, \tag{7}$$

where r is the rank of ρ. To show this, we need a bound on $\dim \mathcal{P}_\lambda$:

$$\dim \mathcal{P}_\lambda \le e^{nH(\bar\lambda)}, \tag{8}$$

which has implicitly appeared in [CM06]. This follows from

$$\dim \mathcal{P}_\lambda \prod_i \bar\lambda_i^{\lambda_i} \le \frac{n!}{\prod_i \lambda_i!} \prod_i \bar\lambda_i^{\lambda_i} = \frac{n!}{n^n} \frac{\prod_i \lambda_i^{\lambda_i}}{\prod_i \lambda_i!} \le 1. \tag{9}$$

The first inequality is by the "hook length formula" [FH04]. 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 $z > 0$ [Bat08]. 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. (4) and (8) now imply that

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

By Eq. (3), this is nonzero only if $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_d = 0$. In this case, we have $\dim \mathcal{Q}_\lambda \le (n+1)^{dr}$ by Eq. (19) in Appendix A. We arrive at Eq. (7).

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. (7) over all pairs (λ, U) such that $F(\rho, U\bar\lambda U^\dagger) \le 1 - \delta$. Since $\sum_\lambda \int \mathrm{d}U < (n+1)^d$, we see that

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

□

Construction via PGM.

Recall that given an ensemble $\{(p_1, \phi_1), \ldots, (p_m, \phi_m)\}$, the PGM has measurement operators $M_i := \bar\phi^{-1/2} p_i \phi_i \bar\phi^{-1/2}$ with $\bar\phi := \sum_i p_i \phi_i$ [Bel75b, Bel75a, HW94]. 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 $\mathrm{d}\sigma$ on the whole state space $\{\sigma\}$. Demanding the unitary invariance of $\mathrm{d}\sigma$, we have

$$\bar\phi = \int \mathrm{d}\sigma\ \sigma^{\otimes n} = \sum_\lambda \frac{\int \mathrm{d}\sigma s_\lambda(\sigma)}{\dim \mathcal{Q}_\lambda} \Pi_\lambda,$$

$$M_\sigma \mathrm{d}\sigma = \sum_\lambda \frac{\dim \mathcal{Q}_\lambda}{\mathbb{E}\, s_\lambda} \Pi_\lambda \sigma^{\otimes n} \Pi_\lambda \mathrm{d}\sigma, \tag{10}$$

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

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

where the inequality is by Eq. (3) and (8). This is the same scaling in n up to constants as Eq. (7), provided $e^{nH(\bar\lambda)}\, \mathbb{E}\, s_\lambda \ge (nd)^{-O(dr)}$. Indeed we show that this is the case if we choose a uniform distribution over the simplex of spectra of σ. First, we bound the Schur polynomial by its largest term:

$$\int \mathrm{d}\sigma s_\lambda(\sigma) \ge \frac{1}{f_d(\vec\lambda = 0)} \underbrace{\int_{s_i \ge 0,\ \sum_i s_i = 1} s_1^{\lambda_1} \cdots s_d^{\lambda_d} \mathrm{d}s}_{f_d(\vec\lambda)} .$$

By writing the integral explicitly, we see that

$$f_d(\lambda_1, \ldots, \lambda_d) = f_2\left(\lambda_1, d-2+\sum_{i=2}^{d}\lambda_i\right) f_{d-1}(\lambda_2, \ldots, \lambda_d),$$

$$f_2(a,b) = \frac{a!b!}{(a+b+1)!}.$$

This implies that $f_d(\vec{\lambda}) = \lambda_1! \cdots \lambda_d!/(n+d-1)!$. (We just calculated the normalization factor for the Dirichlet distribution.) Hence,

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

where in the second inequality we use Eq. (9). We conclude that this PGM defined by the uniform spectrum distribution achieves the same bound (up to constants) on the sufficient number of copies for tomography.

3. LOWER BOUNDS

General Measurements.

Our tomography scheme is the most precise up to logarithmic factors, among all possible measurement schemes given n copies of the unknown state ρ.

THEOREM 2. *Let $\epsilon \in (0,1)$ and $\eta \in (0,1)$. Suppose there exists a POVM $\{M_\sigma \mathrm{d}\sigma\}$ on $(\mathbb{C}^d)^{\otimes n}$ such that for any d-dimensional density matrix ρ with rank $\leq r$,*

$$\int_{\frac{1}{2}\|\sigma-\rho\|_1 \leq \epsilon/2} \mathrm{d}\sigma \,\mathrm{tr}[M_\sigma \rho^{\otimes n}] \geq 1-\eta. \qquad (11)$$

Then,

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

for C a constant depending only on η. In addition, if $r = d$, then $n \geq C\frac{d^2}{\epsilon^2}(1-\epsilon)^2$.

This theorem implies that achieving infidelity $\delta = 1 - F$ requires $n \geq \tilde{\Omega}(dr/\delta)$.

PROOF. We will show that any measurement satisfying (11) will imply the existence of a communication protocol that can reliably send a large message. Holevo's theorem [Hol73] 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 3 below that there exists states $\rho_1, \ldots, \rho_N$ each with rank $\leq r$ such that

$$\frac{1}{2}\|\rho_i - \rho_j\|_1 > \epsilon \quad \forall i \neq j. \qquad (12)$$

The set $\{\rho_1, \ldots, \rho_N\}$ is known as an ϵ-packing net. Fix such a net, along with a measurement $\{M_\sigma \mathrm{d}\sigma\}$ satisfying (11).

We will now construct a communication protocol. Alice will choose a message $x \in [N] := \{1, \ldots, N\}$ 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 (12), 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

$$\tilde{M}_y = \int_{\frac{1}{2}\|\sigma-\rho_y\|_1 \leq \epsilon/2} \mathrm{d}\sigma M_\sigma \ , \ \tilde{M}_{\mathrm{fail}} = \mathrm{id} - \sum_{y\in[N]} \tilde{M}_y.$$

Define $\Pr[y|x] = \mathrm{tr}[\tilde{M}_y \rho_x^{\otimes n}]$. From (11) 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 [Fan61], this implies that

$$I(X:Y) \geq (1-\eta)\ln(N) - \ln(2). \qquad (13)$$

On the other hand, Holevo theorem [Hol73] states that $I(X:Y) \leq \chi$ where χ is the Holevo information:

$$\chi = S\left(\frac{1}{N}\sum_{x\in[N]}\rho_x^{\otimes n}\right) - \frac{1}{N}\sum_{x\in[N]} S(\rho_x^{\otimes n}). \qquad (14)$$

In Lemma 3 below we will argue that there exists a packing net with large N and small χ. We will bound $\chi \leq n\chi_0$ by the subadditivity of entropy where

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

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 3 below. □

LEMMA 3. *There exist ϵ-packing nets I,II,III of d-dimensional states (i.e. satisfying (12)) 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 *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. (Proof in Appendix C.)

The packing net I and II cover small-ϵ regime, and III covers regime where ϵ is close to 1.

We remark that packing nets of size $\exp(\Omega(dr))$ for rank-r states have been achieved as early as 1981 [Sza81, Sza83]; see also [Win04, LVWdW15] 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 [Win04, LVWdW15, OW15a] and has been used in [FGLE12] to argue an $\tilde{\Omega}(r^2d^2)$ 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.

Independent Measurements.

Let us say that a POVM M_σ on $(\mathbb{C}^d)^{\otimes n}$ is a product measurement if it is equal to the tensor product of n POVM's $M^{(a)}$ on $\mathbb{C}^d$. Then we have,

THEOREM 4. *Let $\delta \in (0,1)$ and $\eta \in (0,1)$. Suppose there exists a product POVM $M_\sigma \mathrm{d}\sigma$ on $(\mathbb{C}^d)^{\otimes n}$ such that for any d-dimensional density matrix ρ with rank $\leq r$,*

$$\int_{1-F(\sigma,\rho)\leq\delta/4} \mathrm{d}\sigma \,\mathrm{tr}[M_\sigma \rho^{\otimes n}] \geq 1-\eta. \qquad (15)$$

Then,

$$n \geq C\frac{dr^2}{\delta}(1-\delta)^4$$

for C a constant depending only on η. (Proof in Appendix D.)

4. 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 [Win02]. The first step is to measure λ, as proposed by Keyl-Werner [KW01]. This can be done efficiently using the Schur transform [BCH07] or the quantum Fourier transform over the symmetric group [Bea97, Har05].

Next, we would like to find a collection of unitaries $U_1, \ldots, U_m$ 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 [Win02], which in turn was based on [AW02]). 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 [ICK+15], which is $\tilde{O}(n^{2dr}/\epsilon^2)$.

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

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 [OW15a] with us. JH was 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 were supported by NSERC, NSERC DAS, CRC, and CIFAR. XW was funded by ARO contract W911NF-12-1-0486 and by the NSF Waterman Award of Scott Aaronson when he was affiliated with Center for Theoretical Physics, Massachusetts Institute of Technology. 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.

5. REFERENCES

[ACMnT+07] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum chernoff bound. *Phys. Rev. Lett.*, 98:160501, Apr 2007.

[AW02] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002.

[Bat08] Necdet Batir. Inequalities for the gamma function. *Archiv der Mathematik*, 91(6):554–563, 2008.

[BBG+06] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz Tapia. Optimal full estimation of qubit mixed states. *Phys. Rev. A*, 73:032301, Mar 2006.

[BBMnTR04] E. Bagan, M. Baig, R. Muñoz Tapia, and A. Rodriguez. Collective versus local measurements in a qubit mixed-state estimation. *Phys. Rev. A*, 69:010304, Jan 2004.

[BCH07] D. Bacon, I. L. Chuang, and A. W. Harrow. The quantum Schur and Clebsch-Gordan transforms: I. Efficient qudit circuits. In *Proc. of SODA*, pages 1235–1244, 2007.

[Bea97] R. Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation (STOC)*, pages 48–53, El Paso, Texas, 1997. ACM Press.

[Bel75a] VP Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics: An International Journal of Probability and Stochastic Processes*, 1(1-4):315–345, 1975.

[Bel75b] VP Belavkin. Optimum distinction of non-orthogonal quantum signals. *Radio Engineering and Electronic Physics*, 20:39–47, 1975.

[BK02] H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *J. Math. Phys.*, 43(5):2097–2106, 2002.

[Chi11] Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de Finetti theorems. In *Proceedings of the 5th conference on Theory of quantum computation, communication, and cryptography*, TQC'10, pages 9–25, Berlin, Heidelberg, 2011. Springer-Verlag.

[CHW07] A.M. Childs, A. W. Harrow, and P. Wocjan. Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem. In *Proc. of STACS*, volume 4393 of *LNCS*, pages 598–609, 2007.

[CM06] Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group. *Commun. Math. Phys.*, 261:789–797, 2006.

[Fan61] Robert M Fano. *The transmission of information.* M.I.T. Press and John Wiley and Sons, New York and London, 1961.

[FBK15] Christopher Ferrie and Robin Blume-Kohout.

Minimax quantum tomography: the ultimate bounds on accuracy, 2015.

[FGLE12] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: Error bounds, sample complexity, and efficient estimators. *New J. Phys.*, 14:095022, May 2012.

[FH04] William Fulton and Joe Harris. *Representation Theory: A first course*, volume 129 of *Graduate Texts in Mathematics*. Springer, 2004.

[FL11] Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.*, 106,:230501, April 2011.

[FvdG99] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum mechanical states. *IEEE Trans. Inf. Theory*, 45:1216, 1999.

[Gil05] Richard D. Gill. Conciliation of Bayes and pointwise quantum state estimation: Asymptotic information bounds in quantum statistics, 2005.

[GK06] Mădălin Guţă and Jonas Kahn. Local asymptotic normality for qubit states. *Phys. Rev. A*, 73:052108, May 2006.

[GK08] Mădălin Guţă and Jonas Kahn. Optimal estimation of qubit states with continuous time measurements. *Communications in Mathematical Physics*, 277(1):127–160, 2008.

[GLF$^+$10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(150401), 2010.

[GM00] Richard D. Gill and Serge Massar. State estimation for large ensembles. *Phys. Rev. A*, 61:042312, Mar 2000.

[Har05] Aram W. Harrow. *Applications of coherent classical communication and Schur duality to quantum information theory*. PhD thesis, M.I.T., Cambridge, MA, 2005.

[Hay98] Masahito Hayashi. Asymptotic estimation theory for a finite-dimensional pure state model. *Journal of Physics A: Mathematical and General*, 31(20):4633, 1998.

[Hay06] Masahito Hayashi. *Quantum information: an introduction*. Springer-Verlag, 2006.

[Hay09] Masahito Hayashi. Quantum estimation and the quantum central limit theorem. *American Mathematical Society Translations*, 277:99–123, 2009.

[Hel69] Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, 1969.

[HLW06] Patrick Hayden, Debbie W. Leung, and Andreas Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 265(1):95–117, 2006.

[HM02] Masahito Hayashi and Keiji Matsumoto. Quantum universal variable-length source coding. *Phys. Rev. A*, 66:022311, 2002.

[HM08] Masahito Hayashi and Keiji Matsumoto. Asymptotic performance of optimal state estimation in qubit system. *Journal of Mathematical Physics*, 49(10):102101, 2008.

[Hol73] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

[Hol82] A.S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Publications of the Scuola Normale Superiore. North Holland, 1982.

[HW94] Paul Hausladen and William K. Wootters. A 'pretty good' measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

[HW12] A. W. Harrow and A. J. Winter. How many copies are needed for state discrimination? *IEEE Trans. Inf. Theory*, 58(1):1–2, 2012.

[ICK$^+$15] Raban Iten, Roger Colbeck, Ivan Kukuljan, Jonathan Home, and Matthias Christandl. Quantum circuits for isometries, 2015.

[Key06] M. Keyl. Quantum state estimation and large deviations. *Reveiws in Mathematical Physics*, 18(1):19–60, 2006.

[KG09] Jonas Kahn and Mădălin Guţă. Local asymptotic normality for finite dimensional quantum systems. *Communications in Mathematical Physics*, 289(2):597–652, 2009.

[KRT14] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements, 2014.

[KW01] M. Keyl and R. F. Werner. Estimating the spectrum of a density operator. *Phys. Rev. A*, 64:052311, 2001.

[LVWdW15] Troy Lee, Ignacio Villanueva, Zhaohui Wei, and Ronald de Wolf, 2015.

[MdW13] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing, October 2013.

[MRD$^+$13] D. H. Mahler, Lee A. Rozema, Ardavan Darabi, Christopher Ferrie, Robin Blume-Kohout, and A. M. Steinberg. Adaptive quantum state tomography improves accuracy quadratically. *Phys. Rev. Lett.*, 111:183601, Oct 2013.

[NS09] Michael Nussbaum and Arleta Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37(2):1040–1057, 2009.

[OW15a] Ryan O'Donnell and John Wright. Efficient quantum tomography, 2015.

[OW15b] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 529–538, 2015.

[Sza81] Stanislaw J. Szarek. Nets of Grassmann manifold and orthogonal group. In Bor-Luh Lin, editor, *Proceedings of Research Workshop on Banach Space Theory*, pages

169–185. The University of Iowa, June 29 - July 31 1981.

[Sza83] StanislawJ. Szarek. The finite dimensional basis problem with an appendix on nets of Grassmann manifolds. *Acta Mathematica*, 151(1):153–179, 1983.

[Tan14] Fuyuhiko Tanaka. Quantum minimax theorem, 2014.

[Tys09] Jon Tyson. Error rates of Belavkin weighted quantum measurements and a converse to Holevo's asymptotic optimality theorem. *Phys. Rev. A*, 79:032343, Mar 2009.

[Vor13] Vladislav Voroninski. Quantum tomography from few full-rank observables, 2013.

[Win02] A. Winter. Compression of sources of probability distributions and density operators, 2002.

[Win04] A Winter. Quantum and classical message identification via quantum channels. *Quantum Inf. Comput.*, 4(6&7):563–578, 2004.

[Woo81] W. K. Wootters. Statistical distance and Hilbert space. *Phys. Rev. D*, 23:357–362, Jan 1981.

APPENDIX

A. BACKGROUND ON SCHUR-WEYL DUALITY

The symmetry of our problem implies that our estimators should be invariant under permuting the n systems ($\mathbb{S}_n$) 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 $\mathbb{S}_n$ and $U(d)$.

Schur-Weyl duality is a statement regarding joint representations of a matrix group and the symmetric group. This is standard material [FH04] 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

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

The 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 $\lambda = (\lambda_1, \ldots, \lambda_n)$ 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 simple structure. Namely,

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

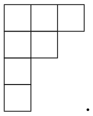
where $\mathcal{Q}_\lambda$ is the irrep of $\mathrm{GL}(d)$ and $\mathcal{P}_\lambda$ is the irrep of $\mathbb{S}_n$ corresponding to the Young diagram λ, and Π_λ is the projector onto the component $\mathcal{Q}_\lambda \otimes \mathcal{P}_\lambda$. Direct consequences of the decomposition are that

$$\Pi_\lambda X^{\otimes n} \Pi_\lambda \cong \mathbf{q}_\lambda(X) \otimes \mathrm{id}_{\mathcal{P}_\lambda} \tag{16}$$

$$\Pi_\lambda X^{\otimes n} = X^{\otimes n} \Pi_\lambda \tag{17}$$

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 $\mathbb{U}(d)$. The space $\mathcal{Q}_\lambda$ is also an irrep of the unitary group $\mathbb{U}(d)$, and our discussion of Schur-Weyl duality could have been formulated entirely with $\mathbb{U}(d)$; however, in this case X would be restricted to be unitary.

For our results it is important to understand the characters of the irrep $\mathcal{Q}_\lambda$ 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.

$$\begin{array}{|c|c|c|} \hline 5 & 5 & 2 \\ \hline 1 & 2 & 3 \\ \hline 4 \\ \cline{1-1} 1 \\ \cline{1-1} \end{array} \; .$$

A *standard Young tableau* (SYT) is one in which each number from $1, \ldots, n$ 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

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

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 $\mathcal{Q}_\lambda$. 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 $1, 2, \ldots, n$ are written in order from the upper left box towards right and down. To understand the basis of $\mathcal{Q}_\lambda$, let $|1\rangle, |2\rangle, \ldots, |d\rangle$ form the standard orthonormal basis of $\mathbb{C}^d$. We may regard each basis vector $|E\rangle = |j_1, \ldots, j_n\rangle$ of $\mathcal{H}$ as a Young tableau E of shape λ. The Young symmetrizer Y_T projects this basis vector to a vector of $\mathcal{Q}_\lambda$. 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 $\nu^\downarrow$ 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 $m = 1, \ldots, d-1$. The negation of the last condition is often denoted as

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

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

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

$$\operatorname{tr} \mathbf{q}_\lambda(X) = \sum_\nu K_{\lambda\nu} x^\nu =: s_\lambda(x). \tag{18}$$

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 $\operatorname{tr} \mathbf{q}_\lambda(X)$ 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

$$s_\lambda(\mathrm{id}_d) = \operatorname{tr} \mathbf{q}_\lambda(\mathrm{id}_d) = \dim \mathcal{Q}_\lambda = \prod_{i<j} \frac{\lambda_i - \lambda_j + j - i}{j - i}. \tag{19}$$

B. PROOF OF BOUND ON SCHUR POLYNOMIALS

Proof of Eq. (3). Consider a positive semi-definite matrix X and a number $k \ge 0$. The largest term in the Schur polynomial $s_\lambda(X^k)$ at eigenvalues $x_1 \ge \cdots \ge x_d \ge 0$ of X is

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

where $\bar x = (x_1, \ldots, x_d)/\operatorname{tr}(X)$, 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

$$s_\lambda(X^k) \le \dim \mathcal{Q}_\lambda \cdot e^{-nkH(\bar\lambda)} e^{-nkD(\bar\lambda\|\bar x)} (\operatorname{tr} X)^{kn}. \tag{20}$$

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. (3). □

C. CONSTRUCTION OF NETS

This section constitutes the proof of Lemma 3. 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 $t \to 0$. 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 [Sza81, Sza83].

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

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

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 $i \ne j$. This can be done inductively starting with the singleton $\{I\}$. Since Haar measure is left-invariant, $\Pr_U[\, \|\rho_U - \rho_V\|_1 \le \epsilon \,] \le \zeta$ 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 5 (Lemma III.5 of Ref. [HLW06]). *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*

$$\forall z > 0 : \Pr_U\left[\frac{d}{pq} \operatorname{tr} QUPU^\dagger \ge 1 + z\right] \le \exp[-pqf(z)],$$

$$\forall z \in (0,1) : \Pr_U\left[\frac{d}{pq} \operatorname{tr} QUPU^\dagger \le 1 - z\right] \le \exp[-pqf(-z)],$$

where

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

Ref. [HLW06] does not explicitly cover the $z > 1$ case for the first inequality, though it implicitly covered in their proof.

Packing net I.

Suppose $3r < d$. Let

$$U = \begin{pmatrix} I_r & 0 & 0 \\ 0 & A_{r\times r} & B_{r\times(d-2r)} \\ 0 & C_{(d-2r)\times r} & D_{(d-2r)\times(d-2r)} \end{pmatrix} \tag{21}$$

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

$$\rho_{t,I} = \begin{pmatrix} (1-t^2)I_r/r & t\sqrt{1-t^2}I_r/r & 0 \\ t\sqrt{1-t^2}I_r/r & t^2 I_r/r & 0 \\ 0 & 0 & 0_{d-2r} \end{pmatrix}, \tag{22}$$

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

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

$$\|\rho_{t,U} - \rho_{t,I_{d-r}}\|_1 \ge \frac{t\sqrt{1-t^2}}{r} \operatorname{tr} C^\dagger C \tag{23}$$

where C is as in Eq. (21). To prove this, observe that $\|\rho_{t,U} - \rho_{t,I_{d-r}}\|_1 \ge |\operatorname{tr}[(\rho_{t,U} - \rho_{t,I_{d-r}})V]|$ 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. Abbreviate as $\alpha = (1-t^2)/r$, $\beta = t\sqrt{1-t^2}/r$, and $\gamma = t^2/r$. Expanding the formula,

$$\begin{aligned} &\operatorname{tr}[(\rho_{t,U} - \rho_{t,I_{d-r}})V] \\ &= \operatorname{tr}\left[\begin{pmatrix} 0 & \beta(A^\dagger - I) & \beta C^\dagger \\ \beta(A-I) & \gamma(AA^\dagger - I) & \gamma A C^\dagger \\ \beta C & \gamma C A^\dagger & \gamma C C^\dagger \end{pmatrix} \begin{pmatrix} A & 0 & BF \\ 0 & E & 0 \\ C & 0 & DF \end{pmatrix}\right] \\ &= \operatorname{tr}\begin{pmatrix} \beta C^\dagger C & \star & \star \\ \star & (\gamma AA^\dagger - I)E & \star \\ \star & \star & (\beta CB + \gamma CC^\dagger D)F \end{pmatrix}. \end{aligned}$$

For some unitary E and F, the trace of the last two entries become the trace norm of the matrices in the parentheses, which are non-negative. This proves Eq. (23).

Lemma 6. *If $0 < t < 1/2$ and $r < d/3$, there exists a finite subset $\{U_i\} \subset \mathbb{U}(d-r)$ of cardinality $N \ge \exp(dr/54)$ 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 $\chi_0 \le t^2 \ln \frac{ed}{t^2 r}$.*

Proof. Lemma 5 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

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

for $z \in (0,1)$. Eq. (23) says that $\|\rho_{t,I_{d-r}} - \rho_{t,U}\|_1 \le t/4$ implies $\frac{d-r}{r(d-2r)} \operatorname{tr} C^\dagger C \le \frac{1}{\sqrt{3}} < 1 - \frac{1}{3}$. Therefore,

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

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 $\bar\rho_t = \int \mathrm{d}U \rho_{t,U}$, 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 $\chi/n \le H(t^2) + t^2 \ln \frac{d-r}{r}$. Using $H(z) \le z \ln(e/z)$, we finish the proof. □

Packing nets II & III.

Assume that d is an even number, and fix a projector $Q = \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)$ of rank $r \le d/2$. For any $d \times d$ unitary U and $0 \le t \le 1$, define

$$\tau_{t,U} = \frac{1+t}{2r} UQU^\dagger + \frac{1-t}{2(d-r)}(I_d - UQU^\dagger). \tag{24}$$

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

$$\chi_0 \le \frac{1}{2}\ln\frac{d^2}{r(d-r)} + \frac{t}{2}\ln\frac{d-r}{r} - H\left(\frac{1+t}{2}\right). \tag{25}$$

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

$$\begin{aligned} \operatorname{tr} AA^\dagger + \operatorname{tr} CC^\dagger &= r \\ \operatorname{tr} BB^\dagger + \operatorname{tr} DD^\dagger &= d - r \\ \operatorname{tr} CC^\dagger + \operatorname{tr} DD^\dagger &= d - r \end{aligned} \tag{26}$$

and

$$\begin{aligned} &\tau_{t,U} - \tau_{t,I_d} = \\ &\begin{pmatrix} \alpha AA^\dagger + \beta BB^\dagger - \alpha I_r & \star \\ \star & \alpha CC^\dagger + \beta DD^\dagger - \beta I_{d-r} \end{pmatrix} \end{aligned}$$

where $\alpha = (1+t)/2r$ and $\beta = (1-t)/2(d-r)$. Multiplying a unitary $\operatorname{diag}(-I_r, I_{d-r})$ on the right of $\tau_{t,U} - \tau_{t,I_d}$, we see that

$$\begin{aligned} &\|\tau_{t,U} - \tau_{t,I_d}\|_1 \\ &\ge \alpha \operatorname{tr}(CC^\dagger - AA^\dagger) + \beta(DD^\dagger - BB^\dagger) + (d-r)\beta - r\alpha \\ &= 2(\alpha - \beta)\operatorname{tr}(CC^\dagger) \qquad \text{by Eq. (26)} \\ &= \left(\frac{1+t}{r} - \frac{1-t}{d-r}\right)\operatorname{tr} CC^\dagger. \end{aligned} \tag{27}$$

Lemma 7. *Suppose $r = d/2$. Then, there exists a finite subset $\{U_i\} \subset \mathbb{U}(d)$ of cardinality $N \ge \exp(d^2/32)$ such that $\|\tau_{t,U_i} - \tau_{t,U_j}\|_1 > t/2$ for any $i \neq j$. The Holevo χ_0 fulfills $\chi \le t^2$.*

Proof. Eq. (25) becomes $\chi/n \le \ln 2 - H((1+t)/2) \le t^2$. Eq. (27) says that if $\|\tau_{t,U} - \tau_{t,I}\|_1 \le t/2$, then $(4/d)\operatorname{tr} CC^\dagger \le 1/2$. Lemma 5 states that this happens with probability at most $\exp(-d^2/32)$. The probabilistic existence argument applies. □

Lemma 8. *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 $N \ge \exp((1-\epsilon)rd/2)$ such that $\|\tau_{1,U_i} - \tau_{1,U_j}\|_1 > 2\epsilon$ for any $i \neq j$. The Holevo χ_0 fulfills $\chi_0 \le \ln(d/r)$.*

Proof. Eq. (25) becomes $\chi_0 \le \ln(d/r)$. Eq. (27) says that if $\|\tau_{t,U} - \tau_{t,I}\|_1 \le 2\epsilon$, then $\frac{d}{r^2}\operatorname{tr} AA^\dagger \ge (1-\epsilon)d/r$, which is greater than 6 when $r < d(1-\epsilon)/6$. By Lemma 5, this happens with probability at most $\exp(-r^2(1-\epsilon)d/2r) = \exp(-rd(1-\epsilon)/2)$. The probabilistic existence argument applies. □

D. PROOF OF INDEPENDENT MEASUREMENT LOWER BOUND

Proof of Theorem 4. Since $\sqrt{1-F}$ is a metric (Bures metric) on the space of states, if there is a set of states ρ_i such that $1 - F(\rho_i, \rho_j) > \delta$ for all $i \neq j$, then for any ρ there is at most one ρ_i such that $1 - F(\rho_i, \rho) \le \delta/4$. Thus, we will take an almost identical strategy as the proof of Theorem 2 to construct δ-packing net of states and compute the Holevo information. In the regime where δ is close to 1, we can use Packing Net III of Lemma 3. Since $1 - T \geq 1 - \sqrt{1 - F^2} \geq F^2/2$, we obtain a packing net of cardinality $N = \exp(\Omega(rd(1-\delta)^4))$.

In order to compute Holevo information and to account for the small δ regime, we consider the following set of states. Define for $t \in [0,1]$ and $U \in \mathbb{U}(d-r)$

$$\omega_{t,I} = \begin{pmatrix} (1-t)I_r/r & & \\ & tI_r/r & \\ & & 0_{d-2r} \end{pmatrix} \tag{28}$$

$$\omega_{t,U} = U \omega_{t,I} U^\dagger \tag{29}$$

where U is embedded into $\mathbb{U}(d)$ as in the paper. $\omega_{t,U}$ has rank $2r < d$. Applying the defining formula

$$F = \operatorname{tr} \sqrt{\sqrt{\omega_{t,I}} \omega_{t,U} \sqrt{\omega_{t,I}}}$$

with the observation that $\omega_{t,U}$ is a mixture of two orthogonal states, we obtain

$$1 - F(\omega_{t,U}, \omega_{t,I}) = t(1 - F(\tau_I, \tau_U)) \tag{30}$$

where $\tau_U = U \tau_I U^\dagger$ is the $(d-r) \times (d-r)$-maximally mixed state of rank r. Since

$$T^2 \leq 2(1-F)$$

by Eq. (1), we can apply the probabilistic existence argument to find a large set of states of cardinality $\exp(\Omega(rd))$ that are $\delta = \Omega(t)$-separated in infidelity.

Next, we bound the Holevo information. Let $\vec{M}^{(a)}$ denote a POVM on $\mathbb{C}^d$, measuring the a-th tensor component (a-th copy) of the input state $\rho^{\otimes n}$. The measurement outcome follows the product distribution

$$\operatorname{tr}\left(\rho^{\otimes n} \bigotimes_{a=1}^{n} \vec{M}^{(a)} \right).$$

The first term in the Holevo information is the Shannon entropy of the distribution

$$\mathfrak{p} = \mathop{\mathbb{E}}_{\omega} \operatorname{tr}\left(\omega^{\otimes n} \bigotimes_{a=1}^{n} \vec{M}^{(a)} \right) \tag{31}$$

whose marginal is equal to $\mathbb{E}_\omega \operatorname{tr}(\omega \vec{M}^{(a)})$. By the subadditivity of entropy, we have

$$H(\mathfrak{p}) \leq \sum_{a=1}^{n} H(\mathop{\mathbb{E}}_{\omega} \operatorname{tr}(\omega \vec{M}^{(a)})). \tag{32}$$

The other term is $-\sum_{a=1}^{n} \mathbb{E}_\omega H(\operatorname{tr}(\omega \vec{M}^{(a)}))$. We are going to bound

$$\chi_a = H(\mathop{\mathbb{E}}_{\omega} \operatorname{tr}(\omega \vec{M}^{(a)})) - \mathop{\mathbb{E}}_{\omega} H(\operatorname{tr}(\omega \vec{M}^{(a)})) \tag{33}$$

for each a, so we shall drop the superscript (a) from now on.

We may assume that $\mathbb{E}_U$ is over the Haar random unitary $U \in \mathbb{U}(d-r)$. Given the POVM $\vec{M}^{(a)}$, Alice and Bob choose a specific set of message carrier states such that Holevo information is minimal: Fix a δ-net ω_{U_j}, a finite set, and consider ω_{VU_j} for $V \in \mathbb{U}(d-r)$. There exists $W \in \mathbb{U}(d-r)$ such that

$$\chi(\omega_{WU_j}) \leq \min_V \sum_a \chi_a(VU_j) \leq \mathop{\mathbb{E}}_V \sum_a \chi_a(VU_j).$$

Using concavity of H, we are effectively replacing U_j in χ_a with Haar random U.

The state $\omega_{t,U}$ has $\mathbb{U}(r)$ symmetry acting on the upper-left corner. If M is any element of POVM, we have for any $V \in \mathbb{U}(r)$

$$\operatorname{tr}(M\omega_{t,U}) = \operatorname{tr}(MV\omega_{t,U}V^\dagger) = \int_{\mathbb{U}(r)} \mathrm{d}V \operatorname{tr}(V^\dagger M V \omega_{t,U}).$$

Therefore, without loss of generality, we can assume that any POVM commutes with $V \in \mathbb{U}(r)$. In addition, a POVM element is a positive semi-definite operator, it suffices to consider rank-1 POVM element since one can alwasy decompose it into rank-1 projectors of some positive weight. If $M_i = dw_i \left|a_i\right\rangle \left\langle a_i\right|$, then $\left|a_i\right\rangle \left\langle a_i\right|$ acts either (i) on the upper-left $r \times r$ corner or (ii) on the complement. In the first case $i = 1, \ldots, m'$,

$$p_i = \operatorname{tr}(M_i \omega_{t,U}) = dw_i \frac{1-t}{r} \tag{34}$$

and in the second case $i = m'+1, \ldots, m$,

$$p_i = \operatorname{tr}(M_i \omega_{t,U}) = dw_i \frac{t}{r} \operatorname{tr}(P_1 U P_r U^\dagger) =: \frac{tw_i d}{d-r} Z_i. \tag{35}$$

The dependence of Z_i on U is implicit. Let $\bar{Z}_i$ denote $\mathbb{E}_U Z_i$. Now,

$$\begin{aligned} \chi_a &= \sum_{i=1}^{m} -\bar{p}_i \ln \bar{p}_i + \mathop{\mathbb{E}}_{U} p_i \ln p_i &(36) \\ &= \sum_{i=m'+1}^{m} \frac{tw_i d}{d-r} (\mathop{\mathbb{E}}_{U} Z_i \ln Z_i - \bar{Z}_i \ln \bar{Z}_i) &(37) \\ &\leq \sum_{i=m'+1}^{m} \frac{tw_i d}{d-r} (\mathop{\mathbb{E}}_{U} Z_i^2 - \bar{Z}_i - \bar{Z}_i \ln \bar{Z}_i) &(38) \end{aligned}$$

By the invariance of the Haar measure, Z_i's now become identical and independent distributions. We only need to know $\sum_{i=m'+1}^{m} w_i$. This is equal to $(d-r)/d$ since $\sum_{i=m'+1}^{m} M_i$ must be equal to the projector onto the $(d-r)$-dimensional subspace. For the expectation value, we may use Lemma 5, but in the present situation we may also directly compute $\mathbb{E}_U Z_i$ and $\mathbb{E}_U Z_i^2$ using the fact that a column of Haar random unitary matrix is a normalized gaussian distribution. Let $d' = d - r$.

$$Z_i r/d' = \operatorname{tr} P_1 U P_r U^\dagger = |u_{11}|^2 + \cdots + |u_{1r}|^2 = \frac{x_1^2 + \cdots + x_{2r}^2}{x^2}$$

where $x_i \sim \mathcal{N}(0, 1/2)$. Since the direction and magnitude are independent, $\mathbb{E}\, Z_i r/d' = [\mathbb{E}(x_1^2 + \cdots + x_{2r}^2)]/\mathbb{E}\, x^2 = 2r/2d'$, so $\mathbb{E}_U Z_i = 1$. Similarly, $\mathbb{E}\, Z_i^2 r^2/d'^2 = [\mathbb{E}(x_1^2 + \cdots + x_{2r}^2)^2]/\mathbb{E}\, x^4 = (r^2 + r)/(d'^2 + d')$. Therefore,

$$\chi \leq \sum_{a=1}^{n} \chi_a \leq nt\left(\frac{1 + 1/r}{1 + 1/d'} - 1 \right) \leq \frac{nt}{r} \tag{39}$$

□

E. SAMPLE COMPLEXITY IN [KRT14]

The previously best achievable sample complexity for state tomography was described in [KRT14]. 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 $a_1, \ldots, a_m \in \mathbb{C}^d$ 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 $a_1, \ldots, a_m$ is "good."*

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 $a_1, \ldots, a_m$ are good, then we have

$$\|\rho - \sigma\|_2 \leq C_3 \frac{\eta}{\sqrt{m}}. \tag{40}$$

To translate this into a quantum measurement, observe that by the operator Chernoff bound [AW02], 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 $p_i := \operatorname{tr}[E_i \rho]$; in the notation of [KRT14] 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

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

We thus have $\eta \leq O(m/\sqrt{n})$ with high probability. According to (40) we then have $\|\rho - \sigma\|_2 \leq O(\sqrt{m/n}) = O(\sqrt{dr/n})$. It follows that

$$\begin{aligned} &\|\rho - \sigma\|_1 \\ &\leq 2\sqrt{\min(\operatorname{rank}(\rho), \operatorname{rank}(\sigma))}\|\rho - \sigma\|_2 \\ &\leq O(\sqrt{dr^2/n}). \end{aligned}$$

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 [OW15a]) can be seen as the advantage that entangled measurements yield for tomography.