Testing identity of collections of quantum states: sample complexity analysis

Marco Fanizza[1], Raffaele Salvia[2], and Vittorio Giovannetti[3]

[1]Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.

[2]Scuola Normale Superiore, I-56127 Pisa, Italy.

[3]NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56127 Pisa, Italy.

We study the problem of testing identity of a collection of unknown quantum states given sample access to this collection, each state appearing with some known probability. We show that for a collection of d-dimensional quantum states of cardinality N, the sample complexity is $O(\sqrt{N}d/\epsilon^2)$, with a matching lower bound, up to a multiplicative constant. The test is obtained by estimating the mean squared Hilbert-Schmidt distance between the states, thanks to a suitable generalization of the estimator of the Hilbert-Schmidt distance between two unknown states by Bădescu, O'Donnell, and Wright (https://dl.acm.org/doi/10.1145/3313276.3316344).

Contents

Marco Fanizza: marco.fanizza@uab.cat
Raffaele Salvia: raffaele.salvia@sns.it

1 Introduction

The closeness between quantum states can be quantified according to a variety of unitarily invariant distance measures, with different operational interpretations [Hay17c]. Given access to copies of some unknown states, a fundamental inference problem is to tell if the states are equal or distant more than ϵ, according to some unitarily invariant distance. Since this problem does not require to completely reconstruct the unknown states with tomography protocols, optimal algorithms require less copies than full tomography to answer successfully [BOW19]. Due to unitarily invariance, efficient algorithms can also be guessed by symmetry arguments [Hay17a].

In this work we study the problem of testing identity of a collection of unknown quantum states given sample access to the collection. We show that for a collection of d-dimensional quantum states of cardinality N, the *sample complexity* is $O(\sqrt{N}d/\epsilon^2)$, which is optimal up to a constant. We assume a *sampling model access*, where each state appears with some known probability, adapting [LRR13; DK16] to the quantum case. We also consider a Poissonized version of the sampling model, where the number of each copies of a state is a Poissonian random variable, and we show that the sample complexity of the two models is the same. This problem is an example of *property testing*, a concept developed in computer science [Gol17], and applied to hypothesis testing of distributions [Can20] and quantum states and channels [MW16]. At variance with optimal asymptotic error rates studied in statistical classical and quantum hypothesis testing [LR06; Hay17c], the sample complexity captures finite size effect in inference problems, as it expresses the number of samples required to successfully execute an inference task in terms of the extensive parameters of the problem, in our case the dimension d and the cardinality N of the collection. The interest in these kind of questions in the classical case has been motivated by the importance of the study of big data sources; a similar motivation holds for the quantum case, since outputs of fully functional quantum computers will also live in high-dimensional spaces.

1.1 Results

Given a collection of d-dimensional quantum states $\{\rho_i\}_{i=1,\ldots,N}$, and a probability distribution p_i ($0 < p_i < 1$), we consider a *sampling model* [LRR13; DK16] where we have access to M copies of the density matrix

$$\rho = \sum_{i=1}^{N} p_i \, |i\rangle\langle i| \otimes \rho_i, \tag{1}$$

where $\{|i\rangle\}_{i=1,\ldots,N}$ is an orthonormal basis of a N dimensional (classical) register. We are promised that one of the two following properties holds:

- **Case A:** $\rho_1 = \rho_2 = \ldots = \rho_N$, which can be equivalently stated by saying that there exists a d-dimensional state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) = 0$, with D_{Tr} the trace distance [Hay17c];

- **Case B:** For any d-dimensional state σ it holds $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon$.

Our goal is to find the values of M for which there is a two-outcome test that can discriminate the two cases with high probability of success. Explicitly, indicating with `"accept"` and `"reject"` the outcomes of the test, we require the probability of getting `"accept"` to be larger than 2/3 in case A, and smaller than 1/3 in case B, i.e.

$$
\begin{cases}
P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case A}) > 2/3 \,, \\[2mm]
P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case B}) < 1/3 \,.
\end{cases}
\tag{2}
$$

Note that the values 2/3 and 1/3 are by convention and, as long as we are interested in the sample complexity only up to a scaling factor, can be replaced by any pair of constants c, s, respectively, such that $1 > c > s > 0$. The main result of the paper is to provide an estimate of necessary and sufficient values of M to fulfill the above conditions. We use the notations $O(f(d, N, \epsilon))$ and $\Omega(g(d, N, \epsilon))$ to indicate respectively upper and lower bounds to sample complexities, up to multiplicative constants. If lower and upper bounds which differ by a multiplicative constant can be obtained, the sample complexity is considered to be determined and indicated as $\Theta(f(d, N, \epsilon)) = \Theta(g(d, N, \epsilon))$.

Specifically we prove the following results:

Theorem 1.1. *For any $\epsilon > 0$, given access to $O\left(\frac{\sqrt{N}d}{\epsilon^2}\right)$ samples of the density matrix ρ of Eq. (1), there is an algorithm which can distinguish with high probability whether*

- *$\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon$ for every state σ (Case B), or*

- *there exists a state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) = 0$ (that is, all the states ρ_i are equal, Case A).*

Theorem 1.2. *For any $\epsilon > 0$, any algorithm which can distinguish with high probability whether*

- *$\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon$ for every state σ (Case B), or*

- *there exists a state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) = 0$ (that is, all the states ρ_i are equal, Case A),*

given access to M copies of the density matrix ρ of Eq. (1), requires at least $M = \Omega\left(\frac{\sqrt{N}d}{\epsilon^2}\right)$ copies.

The proof of Theorem 1.2 is presented in Sec. 4 and it relies on the fact that a test working with M copies could be used to discriminate between two states which are close in trace distance unless $M = \Omega\left(\frac{\sqrt{N}d}{\epsilon^2}\right)$. These states are obtained as average inputs ρ_A and ρ_B of the form of Eq. (1) for two different set of collections of states: in the first case the set is made of only one collection consisting of maximally mixed states (thus satisfying case A), and in the second the set of collections is such that its elements satisfy case B with high probability. The technical contributions of this proof are (a) a lower bound on the probability that a collection of random states with spectrum $s_\epsilon = (\frac{1+\epsilon}{d}, \frac{1-\epsilon}{d}, ..., \frac{1+\epsilon}{d}, \frac{1-\epsilon}{d})$ has large average trace distance to their average state; (b) an upper bound on the distance between ρ_A and ρ_B being the average input state over collections of random states with spectrum s_ϵ. Both results could be useful elsewhere.

The derivation of the upper bound for M given in Theorem 1.1 is instead presented in Sec. 3 and it is obtained by constructing an observable $\mathcal{D}_M$ whose expected value is the mean squared Hilbert-Schmidt distance between the states ρ_i, and we bound the variance of the estimator. By relating the mean squared Hilbert-Schmidt distance to $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sum_i p_i \rho_i)$ we obtain the test of the theorem. This strategy follow closely the methods of [BOW19] (for $N = 2$), although

with some relevant changes due to the fact that we are not requiring a fixed number of copies of each state ρ_i, like in [BOW19]. This difference is relevant from a conceptual point of view, since having an arbitrary number of copies of each state is a stronger type of access with respect to the sampling model, and closer to the query model (we discussed the different applicability scenario in the following section). It is also relevant from a technical point of view, since it is not immediate to devise an estimator for which the analysis can be completed. In fact, the analysis exploits a *Poissonization* trick [LRR13] where the number of copies M is not fixed but a random variable, extracted from a Poisson distribution with average μ, $\mathrm{Poi}_\mu(M) := \frac{e^{-\mu}\mu^M}{M!}$ (summarized later on by the notation $M \sim \mathrm{Poi}_\mu$). We then look for a test which can be performed by a two-outcome POVM $\{E_0^{(M)}, E_1^{(M)}\}$ for each M. Poissonization is a standard technique that allows the for some useful simplification of the analysis by getting rid of unwanted correlations (more on this in Sec. 3.1). The equivalence of the Poissonized model with the original one is formalised in Appendix A.

Analogously to [BOW19] we can refine the upper bound when the states in the collection have low rank. Given the state ρ of Eq. (1), we define its reduced average density matrix

$$\bar{\rho} := \sum_{i=1}^N p_i \rho_i \,, \tag{3}$$

In particular, when $\bar{\rho}$ is η-close to rank k, that is, the sum of its k largest eigenvalues is larger than $1 - \eta$, we can refine Theorem 1.1:

Theorem 1.3. *If the density matrix $\bar{\rho}$ of Eq. (3) is η-close to rank k, given access to $O\left(\frac{\sqrt{N}k}{\epsilon^2}\right)$ samples of ρ there exists an algorithm which can distinguish with high probability whether $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon + \eta$ for every state σ, or there exists a state σ such that $\sum_i p_i D_{HS}(\rho_i, \sigma) < 8(2 - \sqrt{2})\epsilon$.*

1.2 Motivation of the setting

In this section we present a couple of physical settings which give rise to the sampling models discussed in Sec.1.1, as both the original model and the Poissonized model refer to natural scenarios for a certification task.

Independent sources setting (panel (a) of Figure 1). It is fair to assume that each copy of the states is produced by a device $\mathcal{S}_i$ that require some physical time to run, and produces the expected state with some probability. Moreover, assume that the number of produced copies of ρ_i by $\mathcal{S}_i$ at any time T is given by a Poisson distribution with rate r_i and average $r_i T$, i.e. $P_T(m_i) = \mathrm{Poi}_{r_i T}(m_i) = \frac{(r_i T)^{m_i} e^{r_i T}}{m_i!}$. With this assumption, it also holds that the probability that a total of $M = m_1 + ... + m_N$ copies is produced in the time T is $P_T(M) = \frac{(T\sum_{i=1}^N r_i)^M e^T \sum_{i=1}^N r_i}{M!}$. The Poissonized sampling model (where the probabilities of getting m_i copies of ρ_i are given by a Poisson distribution with average $p_i \mu$, see Eq. (38)) is an adequate representation of the setting where we want to do our certification test with all the copies that are produced in a certain timeframe T, see panel (a) of Figure 1.

On the other hand, if we decide to run the test as soon the total number of copies corresponds to the desired number M, we end up in the original sample model. Indeed, if T_M is the random variable equal to the time at which the total number of copies is M, we have that the probability of finding a vector $\vec{m} = (m_1, ..., m_N)$ of number of copies $\rho_1, ..., \rho_N$, respectively, conditioned on $m_1 + ... + m_N = M$ at the time T_M, is

$$P(\vec{m}|M) = \int_0^\infty p(T_M = T) P_T(\vec{m}|M) dT, \tag{4}$$

where $p(T_M = T)$ is the probability density for the stopping time T_M, and

$$P_T(\vec{m}|M) = P_T(\vec{m}, M)/P_T(M) = P_T(\vec{m})/P_T(M) = \prod_{j=1}^{N} P_T(m_j)/P_T(M) \qquad (5)$$

$$= \frac{M!}{m_1!...m_N!} \prod_{j=1}^{N} r_i^{m_i}. \qquad (6)$$

where the first equality comes from the definition of conditional probability, the second comes from the fact that M is completely determined by $\vec{m}$, the third comes from the fact that the components of $\vec{m}$ are independent when conditioning only on T, and the last equality comes from writing the probabilities explicitly. Finally, by integrating a constant function, we have

$$P(\vec{m}|M) = \frac{M!}{m_1!...m_N!} \prod_{j=1}^{N} r_i^{m_i}, \qquad (7)$$

which is the probability distribution of the copies of each ρ_i in the original sampling model with M total copies of ρ, provided that $r_i = p_i$.

These two situations can be compared with the setting of the *query model*, already considered in [Yu23]; in that case, we are allowed to ask for any number of copies of each state ρ_i in the collection, and the sample complexity is measured with respect to the total number of copies requested. This type of access is clearly stronger with respect to the sampling models, and indeed the sample complexity is lower, being $\Theta(d/\epsilon^2)$. However, assuming there is a finite rate of copies/time, the sampling model captures better the actual physical time required to generate the copies for the test.

On the other hand, the validity of the assumption that the number of copies of each state is generated by a Poisson distribution can be questioned. By the law of rare events, this is a realistic approximation if each source actually corresponds to many independent sources, each of which produces a copy of the state with very small probability, such that the total rate of production of state is finite. In particular, the following bound on the variational distance between the Poisson distribution and sum of independent Bernoulli random variables $X_i \sim (p_i, 1 - p_i)$ holds [LC60]: $\sum_{k=0}^{\infty} |P(\sum_{i=1}^{n} X_i = k) - \frac{(\sum_{i=1}^{n} p_i)^k e^{-(\sum_{i=1}^{n} p_i)}}{k!}| < 2(\sum_{i=1}^{n} p_i^2)$. The approximation with i.i.d. Bernoulli variables was considered, for example, for entanglement certification of single-photon pairs produced with spontaneus parametric down conversion [HTM08] in the asymptotic setting, with proposed tests implemented experimentally [Hay+06]. In these cases the single-photon pairs are produced with a very small probability from a single beam, but with a finite rate if the number of beams is large, and the distribution of the total number of pairs is approximated by a Poisson distribution. In any case, since any probabilistic model can be simulated or can simulate our sampling model, simply simulating the desired probability distribution on a classical computer and waiting for enough copies, our protocol gives respectively upper or lower bounds on the sample complexity. These bounds are tight if the simulation is efficient, that is it requires the same number of copies, up to a constant multiplicative factor. It would be interesting to characterize which sampling models can efficiently simulate or be simulated by the Poissonized model, but we will not discuss this issue here.

Noisy measurement setting (panel (b) of Figure 1). We point out another setting where the sampling model can represent a realistic situation in the lab: suppose that some preparation procedure $\mathcal{S}$ ends with some measurement, but different outcomes of the measurement are expected to correspond to the same desired state. An example could be the case if our preparation apparatus has interacted with an environment, and we measure the environment. Since
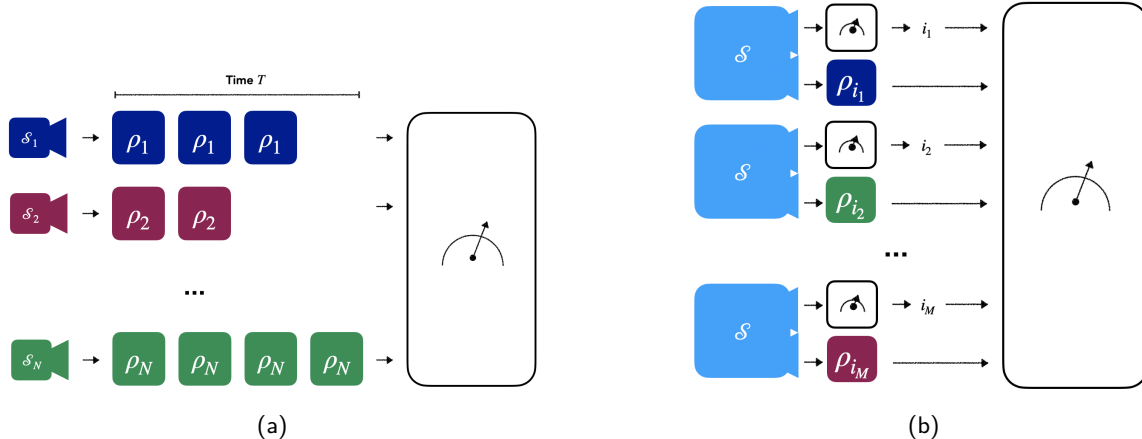
Figure 1: Two settings to which our model applies. In a) independent sources produce a random number, Poisson distributed, of copies of a state in a time T. In b), a measurement procedure prepares labeled states, each label appearing with some probability.

the outcome of the measurement at the preparation stage is random, the procedure prepares in principle different states for each measurement outcome. The classical-quantum state we obtain, possibly after post-selection of acceptable measurement outcomes, will have the form in Eq. (1). The goal of the test is to certify if a source of states of this kind is stable or not.

Finally, we point out similarities with the problem of quantum change point detection [AH11; Sen+16; SCMT17; SMVMT18; FHC23], in which a sequence of unknown states is presented, and it is asked if they are all equal or not. An additional question is to identify the points where the states change. With our algorithm, we are able to answer correctly to the change point problem when we know that there could be a change point among $N-1$ possible change points, which have distances between them given by Poissonian random variables. It would be interesting if the analysis and the techniques of the present paper could be extended to address the change point problem more directly.

1.3 Related work

1.3.1 Classical distribution testing

For an overview of learning properties of a classical distribution in the spirit of property testing, we refer to [Gol17; Can20]. We report a partial list of results which are of direct interest for this paper, about testing symmetric properties of distribution in total variation distance. We use the notation $[d]$ for the set $\{1, ..., d\}$. Learning a classical distribution over $[d]$ in total variation distance can be done in $O(d/\epsilon^2)$ samples [Gol17], therefore the interest in testing properties is to get a sample complexity $o(d)$. The problem of testing uniformity was addressed in [GR11] and established to be $O(\sqrt{d}/\epsilon^2)$ in successive works [Pan08; VV14]. More generally, the sample complexity of identity testing to a known distribution has been established to be $\Theta(\sqrt{d}/\epsilon^2)$ [VV14; DKN15]. Identity testing for two unknown distribution is $\Theta(\max(d^{1/2}/\epsilon^2, d^{2/3}/\epsilon^{4/3}))$ [Cha+14]. The problem of testing identity of collection of N distributions was introduced in the classical case in [LRR13] and solved in [DK16], obtaining $\Theta(\max(\sqrt{dN}/\epsilon^2, d^{2/3}N^{1/3}/\epsilon^{4/3}))$ for the sampling model, where at each sample the tester receives one of N distributions with probabilty p_i, and $\Theta(\max(\sqrt{d}/\epsilon^2, d^{2/3}/\epsilon^{4/3}))$ for the query model, where the tester can choose the distribution to call at each sample. A problem related to testing identity of collections is testing independence of a distribution on $\times_{i=1}^{l}[n_i]$, which was addressed by [Bat+01; LRR13; AD15] and solved in [DK16], which showed a tight sample

complexity $\Theta(\max_j(\prod_{i=1}^l n_i^{1/2}/\epsilon^{1/2}, n_j^{1/3} \prod_{i=1}^l n_i^{1/3}/\epsilon^{4/3}))$.

1.3.2 Quantum state testing

It has been shown that the reconstruction of the classical description of an unknown state, *quantum tomography*, requires $\Theta(d^2/\epsilon^2)$ copies of the state [Haa+17; OW16; OW17]. These algorithms often include, as a subroutine, spectrum learning [ARS88; KW01; HM02; Chr06; Key06], which has sample complexity $O(d^2/\epsilon^2)$ [OW16], although a matching lower bound is available only for the empirical Young diagram estimator [OW15]. These results have been refined in the case the state is known to be close to a state of rank less than k. Quantum entropy estimation has been studied in [Ach+20]. The property testing approach to quantum properties has been reviewed in [MW16], where it is also shown that testing identity to a pure state requires $O(1/\epsilon^2)$. Testing identity to the maximally mixed state takes $\Theta(d/\epsilon^2)$ [OW15], and the same is true for a generic state and for testing identity between unknown states (with refinements if the state can be approximated by a rank k state) [BOW19]. In [BOW19], identity testing between unknown states is done by first estimating their Hilbert-Schmidt distance with a minimum variance unbiased estimator, developing a general framework for efficient estimators of sums of traces of polynomials of states. This improves on a simple way to estimate the overlap $\text{Tr}[\rho\sigma]$ between two unknown states, the swap test [Buh+01], while optimal estimation of the overlap between pure states with average error figures of merit has been addressed by a series of works [BRS04; BIMT06; LSB06; GI06; Fan+20]. In all of these cases, the algorithms considered are classical post-processing of the measurement used to learn the spectrum of a state, possibly repeated on nested sets of inputs. This measurement can be efficiently implemented, with gate complexity $O(n, \log d, \log 1/\delta)$ [BCH06; Har05; Kro19], where n is the number of copies of the state, and δ is the precision of the implementation. This measurement is relevant for several quantum information tasks, for example in communication (see e.g. [Hay17a; Ben+14]). Testing identity of collections of quantum states in the query model has been established to be $\Theta(d/\epsilon^2)$ [Yu21], while the sampling model complexity was left open and is addressed in this paper. Independence testing is also addressed in [Yu21], obtaining a sample complexity $O(d_1 d_2/\epsilon^2)$, which is tight up to logarithmic factors, using the identity test of [BOW19] for testing independence of a state on $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$; similar results hold for the multipartite case (see also [HT16] for the asymptotic setting and [Bai+22] for an application of independence testing to identification of causal structure). Besides these optimality results, which are valid if one allows any measurement permitted by quantum mechanics, several results have been obtained in the case in which there are restrictions on the measurements: [BCL20] shows that the sample complexity for testing identity to the maximally mixed state with independent but possibly adaptive measurements is $\Omega(d^{4/3}/\epsilon^2)$ and $\Theta(d^{3/2}/\epsilon^2)$ for non-adaptive measurements, while the instance optimal case for the same problem is studied in [CLO22]; [Haa+17] shows that the sample complexity for tomography for non-adaptive measurements is $\Omega(d^3/\epsilon^2)$. Algorithms with Pauli measurements only have been considered [Yu21; Yu23], while a general review of the various approaches with attention to feasibility of the measurement can be found in [KR21].

2 Preliminaries

2.1 Distance measures for collection of distributions

Quantum states are positive operators in a Hilbert space, with trace one. In this work we consider states living in a Hilbert space of finite dimension d and we make use of the Schatten operator norms [Hay17c]: $||A||_p = \text{Tr}\left[\sqrt{A^\dagger A}^p\right]^{1/p}$. In particular, given ρ and σ two quantum

states of the system, we express their trace distance as $D_{\text{Tr}}(\rho, \sigma)$ and their Hilbert-Schmidt distance $D_{\text{HS}}(\rho, \sigma)$ as

$$D_{\text{Tr}}(\rho, \sigma) = \frac{||\rho - \sigma||_1}{2} , \qquad D_{\text{HS}}(\rho, \sigma) = ||\rho - \sigma||_2 . \tag{8}$$

These quantities are connected via the following inequalities

$$\frac{1}{2} D_{\text{HS}}(\rho, \sigma) \leq D_{\text{Tr}}(\rho, \sigma) \leq \frac{\sqrt{d}}{2} D_{\text{HS}}(\rho, \sigma). \tag{9}$$

We also recall that the trace distance admits a clear operational interpretation due to the Holevo-Helstrom theorem (see e.g. [Hay17c]): if a state is initialized as ρ with probability $1/2$ and σ with probability $1/2$, the maximum probability of success in identifying the state correctly is given by:

$$p_{succ}(\rho, \sigma) = \frac{1}{2} \left(1 + D_{\text{Tr}}(\rho, \sigma)\right). \tag{10}$$

For ρ and $\bar{\rho}$ as defined in Eq. (1) and (3), we introduce the quantity

$$\mathcal{M}_{\text{Tr}}(\rho) := \sum_{i=1}^{N} p_i D_{\text{Tr}}(\rho_i, \bar{\rho}) \leq \frac{1}{2} \sum_{i=1}^{N} p_i \sqrt{d D_{HS}^2(\rho_i, \bar{\rho})} . \tag{11}$$

We also define the mean squared Hilbert-Schmidt distance of the model as

$$\mathcal{M}_{HS}(\rho) := \left[\sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j D_{HS}^2(\rho_i, \rho_j) \right]^{1/2}, \tag{12}$$

observing that it can be equivalently expressed in terms of $\bar{\rho}$ as

$$\begin{aligned}
\mathcal{M}_{HS}^2(\rho) &:= \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j D_{HS}^2(\rho_i, \rho_j) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j \, \text{Tr}\left[(\rho_i - \rho_j)^2\right] \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j \, \text{Tr}\left[(\rho_i - \bar{\rho} + \bar{\rho} - \rho_j)^2\right] \\
&= 2 \sum_{i=1}^{N} p_i \, \text{Tr}\left[(\rho_i - \bar{\rho})^2\right] - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j \, \text{Tr}[(\rho_i - \bar{\rho})(\rho_j - \bar{\rho})] \\
&= 2 \sum_{i=1}^{N} p_i D_{HS}^2(\rho_i, \bar{\rho}). \tag{13}
\end{aligned}$$

Therefore we can derive the following important inequality

$$\begin{aligned}
\mathcal{M}_{\text{Tr}}(\rho) = \sum_{i=1}^{N} p_i D_{\text{Tr}}(\rho_i, \bar{\rho}) &\leq \frac{1}{2} \sum_{i=1}^{N} p_i \sqrt{d D_{HS}^2(\rho_i, \bar{\rho})} \\
&\leq \frac{1}{2} \sqrt{\sum_{i=1}^{N} p_i} \sqrt{\sum_{i=1}^{N} p_i d D_{HS}^2(\rho_i, \bar{\rho})} = \frac{\sqrt{d}}{2\sqrt{2}} \mathcal{M}_{HS}(\rho) , \tag{14}
\end{aligned}$$

which will be used in the next section to obtain a test for $\mathcal{M}_{\text{Tr}}(\rho)$ starting from a test for $\mathcal{M}_{HS}(\rho)$.

If the state σ is close to having rank k, in the sense that the sum of its largest k eigenvalues is larger than $1 - \eta$, then the following inequality (proven in section 5.4 of [BOW19]) holds

$$D_{\mathrm{Tr}}(\rho, \sigma) \le \frac{\sqrt{k}}{c} D_{HS}(\rho, \sigma) + \eta \,, \tag{15}$$

with $c = 2 - \sqrt{2}$. Therefore, in the special case in which the average state $\bar{\rho}$ is η-close to having rank k, the inequality (14) can be improved by

$$
\begin{aligned}
\mathcal{M}_{\mathrm{Tr}}(\rho) = \sum_{i=1}^{N} p_i D_{\mathrm{Tr}}(\rho_i, \bar{\rho}) &\le \sum_{i=1}^{N} p_i \left(\frac{1}{c} \sqrt{k D_{HS}^2(\rho_i, \bar{\rho})} + \eta \right) \\
&= \frac{1}{c} \sum_{i=1}^{N} p_i \sqrt{k D_{HS}^2(\rho_i, \bar{\rho})} + \eta = \frac{1}{c} \sum_{i=1}^{N} \sqrt{p_i} \sqrt{p_i k D_{HS}^2(\rho_i, \bar{\rho})} + \eta \\
&\le \frac{1}{c} \sqrt{\sum_{i=1}^{N} p_i} \sqrt{\sum_{i=1}^{N} p_i k D_{HS}^2(\rho_i, \bar{\rho})} + \eta = \frac{\sqrt{k}}{c\sqrt{2}} \mathcal{M}_{HS}(\rho) + \eta \,.
\end{aligned}
\tag{16}
$$

In our analysis we will also need the following divergences for classical distributions p,q: the *chi-squared* divergence, defined as $\mathrm{d}_{\chi^2}(p\|q) := \sum_i \frac{(p_i - q_i)^2}{q_i}$; the *Kullback-Leibler* divergence, defined as $\mathrm{d}_{KL}(p\|q) := \sum_i p_i \log_2 \frac{p_i}{q_i}$; and the total variation distance, defined as $\mathrm{d}_{TV}(p\|q) := \frac{1}{2} \sum_i |p_i - q_i|$, which corresponds to the trace distance between states which are diagonal in the same basis [CT05; SV16]. From the definition of Kullback-Leibler divergence, it follows that it is additive, i.e.

$$\mathrm{d}_{KL}\left(\prod_{j=1}^{N} p^{(j)} \| \prod_{j=1}^{N} q^{(j)} \right) = \sum_{j=1}^{N} \mathrm{d}_{KL}(p^{(j)} \| q^{(j)}) \,. \tag{17}$$

We remind also that the total variation distance is related to the Kullback-Leibler divergence by Pinsker's inequality:

$$\mathrm{d}_{TV}(p, q) \le \sqrt{\frac{1}{2} \mathrm{d}_{KL}(p\|q)} \,, \tag{18}$$

and that the Kullback-Leibler can be bounded in terms of the chi-squared divergence, as:

$$\mathrm{d}_{KL}(p, q) \le \ln\left[1 + \mathrm{d}_{\chi^2}(p, q) \right] \,. \tag{19}$$

2.2 Schur-Weyl duality

In this section we review some key facts in group representation theory that are useful to discuss properties of i.i.d. quantum states. Consider the state space of l, d-dimensional systems, $\mathcal{H}_d^{\otimes l}$. This space carries the action of two different groups; the special unitary group of $d \times d$ complex matrices, $\mathrm{SU}(d)$, and the permutation group of l objects, S_l. Specifically, the groups $\mathrm{SU}(d)$ and S_l act on a basis $\{|i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_l\rangle\}_{i_1,i_2...,i_l}$ of $\mathcal{H}_d^{\otimes l}$ via unitary representations $u_l : \mathrm{SU}(d) \to \mathrm{U}(\mathcal{H}_d^{\otimes l})$, and $s_l : S_l \to \mathrm{U}(\mathcal{H}_d^{\otimes l})$ as follows

$$
\begin{aligned}
u_l(U) |i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_l\rangle &= U^{\otimes l} |i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_l\rangle \\
&= U|i_1\rangle \otimes U|i_2\rangle \otimes ... \otimes U|i_l\rangle \,, \quad \forall U \in \mathrm{SU}(d) \tag{20} \\
s_l(\tau) |i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_l\rangle &= |\tau^{-1}(i_1)\rangle \otimes |\tau^{-1}(i_2)\rangle \otimes ... \otimes |\tau^{-1}(i_l)\rangle \,, \forall \tau \in S_l.
\end{aligned}
$$

9

Observe that $[U^{\otimes l}, s_l(\tau)] = 0$, $\forall U \in \mathrm{SU}(d)$, and $\forall \tau \in S_l$. Let $Y_{l,d}$ denote be the set of integer partitions of l in at most d parts written in decreasing order, pictorially represented by Young diagrams, where l boxes are arranged into at most d rows. $\lambda \in Y_{l,d}$ can then also be written as a vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_d)$ with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$. Schur-Weyl duality [Hay17b; Hay17a] states that the total state space $\mathcal{H}_d^{\otimes l}$ can be decomposed as

$$\mathcal{H}_d^{\otimes l} \cong \bigoplus_{\lambda \in Y_{l,d}} \mathcal{U}^{(\lambda)}(\mathrm{SU}(d)) \otimes \mathcal{V}^{(\lambda)}(S_l), \tag{21}$$

where the unitary irreducible representation (irrep) $u^{(\lambda)}$ of $\mathrm{SU}(d)$ acts non trivially on the factor $\mathcal{U}^{(\lambda)}(\mathrm{SU}(d))$ of dimension χ_λ and the irrep $s^{(\lambda)}$ of S_l acts non trivially on the factor $\mathcal{V}^{(\lambda)}(S_l)$ of dimension ω_λ. The use of the congruence sign in Eq. (21) indicates that this block decomposition is accomplished by a unitary transformation; in the case considered here this unitary is the Schur transform [BCH06; Har05; Kro19].

A state $\rho^{\otimes l} \in \mathcal{D}(\mathcal{H}_d^{\otimes l})$ commutes with $s_l(\sigma)$ for any σ. By Schur's lemma, $\rho^{\otimes l}$ can be decomposed in block diagonal form according to the isomorphism in Eq. (21).

$$\rho^{\otimes l} = \sum_{\lambda \in Y_{l,d}} \mathrm{SW}_\rho^l(\lambda) \rho_\lambda \otimes \frac{\mathbf{1}_\lambda}{\omega_\lambda}, \tag{22}$$

where $\mathrm{SW}_\rho^l(\lambda)$ is a probability distribution over the Young diagrams, which depends only on the number of copies l and on the spectrum of ρ, and ρ_λ are χ_λ-dimensional states. Applying $u_l(U)$ with U extracted from the Haar measure of $\mathrm{SU}(d)$ gives

$$\mathcal{G}_{\mathrm{SU}(d)}^l[\rho] := \int_{U \in \mathrm{SU}(d)} \mathrm{d}U \, U^{\otimes l} \, \rho^{\otimes l} \, U^{\dagger \otimes l} = \sum_{\lambda \in Y_{l,d}} \mathrm{SW}_\rho^l(\lambda) \frac{\mathbf{1}_\lambda}{\chi_\lambda} \otimes \frac{\mathbf{1}_\lambda}{\omega_\lambda} = \sum_{\lambda \in Y_{l,d}} \mathrm{SW}_\rho^l(\lambda) \frac{\Pi_\lambda}{\chi_\lambda \omega_\lambda}, \tag{23}$$

again by Schur's lemma, where we defined the orthogonal set of projectors $\{\Pi_\lambda\}_{\lambda \in Y_{l,d}}$. The projective measurement with these projectors is called weak Schur sampling [Har05; Kro19], and it can be executed with gate complexity $O(l, \log d, \log 1/\delta)$, where δ is the precision of the implementation (that is, the maximum trace distance between pairs of states obtained applying the actual circuit implementation of the measurement and the ideal operation to the same pure state). Finally, for any decomposition $\mathcal{H}_d^{\otimes l} = \otimes_{i=1}^N \mathcal{H}_d^{\otimes m_i}$ (where $\sum_{i=1}^N m_i = l$), one can define a family of weak Schur sampling projectors for each factor, $\{\Pi_\lambda^{(i)}\}_{\lambda \in Y_{m_i,d}}$. Since the elements of $\{\Pi_\lambda\}_{\lambda \in Y_{l,d}}$ commute with local permutations, they commute with the projectors $\{\otimes_{i=1}^N \Pi_{\lambda_i}^{(i)}\}_{\lambda_i \in Y_{m_i,d}}$. Indeed, we can decompose $\mathcal{H}_d^{\otimes l}$ according to irreducible representations of $S_{m_1} \times S_{m_2} \times ... \times S_{m_N}$; irreducible representations are labeled by $(\lambda_1, ..., \lambda_N)$, $\lambda_i \in Y_{m_i,d}$, appear in general with multiplicity, and the projector on all the irreducible components with label $(\lambda_1, ..., \lambda_N)$ is $\otimes_{i=1}^N \Pi_{\lambda_i}^{(i)}$. By Schur's lemma, $\{\Pi_\lambda\}_{\lambda \in Y_{l,d}}$ should be block diagonal according to the decomposition given by $\{\otimes_{i=1}^N \Pi_{\lambda_i}^{(i)}\}_{\lambda_i \in Y_{m_i,d}}$. Therefore local and global weak Schur sampling can be done with a unique projective measurement, and the probabilities of the outcomes are the same if the two projective measurements are executed in any order. Therefore, this nested weak Schur sampling is also efficient, and it will give an implementation of the measurement required by the test we study in this paper.

3 Upper bound on the sample complexity

In order to prove Theorem 1.1 here we show a stronger version of such statement, i.e.

Theorem 3.1. *Given access to $O(\frac{\sqrt{N}}{\delta})$ samples of the state ρ of Eq. (1), for $\delta > 0$ there is an algorithm which can distinguish with high probability whether $\mathcal{M}^2_{HS}(\rho) \leq 0.99\delta$ or $\mathcal{M}^2_{HS}(\rho) > \delta$.*

The connection with Theorem 1.1 follows by the relations between the functionals $\mathcal{M}_{HS}(\rho)$ and $\mathcal{M}_{\mathrm{Tr}}(\rho)$ discussed in Sec. 2.1. Specifically we note that $\mathcal{M}_{\mathrm{Tr}}(\rho) = 0$ (case A) implies $\mathcal{M}_{HS}(\rho) = 0$, while having $\mathcal{M}_{\mathrm{Tr}}(\rho) > \epsilon$ (a constraint that holds in Case B) implies $\mathcal{M}^2_{HS}(\rho) > \frac{8\epsilon^2}{d}$ by Eq. (14). Therefore a test satisfying the requests of Theorem 1.1 can be obtained by taking the algorithm identified by Theorem 3.1 with $\delta = \frac{8\epsilon^2}{d}$. [Incidentally we stress that the test can be performed by a two outcome POVMs $\{E_0^{(M)}, E_1^{(M)}\}$ when the number of copies of ρ is M (for any $M \geq 0$), obtained as projectors on the eigenvectors of the observable $\mathcal{D}_M$, defined in the following, with eigenvalues respectively larger or lower than a threshold; therefore, it is of the class of test on which we can apply Proposition A.1].

In a complete analogous way, Theorem 1.3 follows by calling the algorithm of Theorem 3.1 with $\delta = \frac{16(2-\sqrt{2})^2\epsilon^2}{k}$, and using the inequality (16).

The reminder of the section is hence devoted to the prove Theorem 3.1.

3.1 Building the estimator for $\mathcal{M}^2_{HS}$

To prove Theorem 3.1 we construct an unbiased estimator for $\mathcal{M}^2_{HS}$, generalizing the estimator of $D^2_{HS}(\rho, \sigma)$ discussed in [BOW19]. We start noticing that via permutations that operate on the quantum registers conditioned on measurements performed on the classical registers, the density matrix $\rho^{\otimes M}$ describing M sampling of the state ρ, can be cast in the following equivalent form

$$\rho^{(M)} := \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} |\vec{m}\rangle\langle\vec{m}| \otimes \rho^{\vec{m}}. \tag{24}$$

In this expression the summation runs over all vectors $\vec{m} = (m_1, m_2, \cdots, m_N)$ formed by integers that satisfy $m_1 + m_2 + \cdots + m_N = M$; while $\mathsf{M}(\vec{m})_{\vec{p},M}$ is the multinomial distribution with M extractions and probabilities $\vec{p} = (p_1, p_2, \cdots, p_N)$, i.e.

$$\mathsf{M}(\vec{m})_{\vec{p},M} := \frac{M!}{m_1!...m_N!} p_1^{m_1} p_2^{m_2} \cdots p_N^{m_N} ; \tag{25}$$

the vectors $|\vec{m}\rangle = |m_1, m_2, \cdots, m_N\rangle$ form an orthonormal set for the classical registers of the model; while finally

$$\rho^{\vec{m}} := \rho_1^{\otimes m_1} \otimes \rho_2^{\otimes m_2} \otimes ... \otimes \rho_N^{\otimes m_N} , \tag{26}$$

is a state of the quantum registers with m_i elements initialized into ρ_i, which formally operates on a Hilbert space with tensor product structure $\otimes_{i=1}^N \mathcal{H}_i$, with $\mathcal{H}_i = (\mathbb{C}^d)^{\otimes m_i}$, with $m_i = 0, ..., M$. Exploiting the representation of Eq. (24) we then introduce the observable

$$\mathcal{D}_M := \sum_{\vec{m} \in \mathcal{P}_M} |\vec{m}\rangle\langle\vec{m}| \otimes \mathcal{D}^{\vec{m},M}, \tag{27}$$

with

$$\mathcal{D}^{\vec{m},M} := \sum_{i \neq j} \mathcal{D}_{ij}^{m_i,m_j,M}, \tag{28}$$

and

$$\mathcal{D}_{ij}^{m_i,m_j,M} := \frac{m_i(m_i - 1)}{\mu^2 p_i} p_j \mathcal{O}_{ii}^{m_i,m_i} + \frac{m_j(m_j - 1)}{\mu^2 p_j} p_i \mathcal{O}_{jj}^{m_j,m_j} - 2 \frac{m_i m_j}{\mu^2} \mathcal{O}_{ij}^{m_i,m_j}. \tag{29}$$

11

In the above expression $\mu > 0$ is a free parameter that will be fixed later on. The operators $\mathcal{O}_{ij}^{m_i,m_j}$ are defined to be the average of all possible different transpositions $S \in S_{ij}^{m_i,m_j}$ between two local copies of $\mathbb{C}^d$ in the spaces $\mathcal{H}_i$ and $\mathcal{H}_j$, with i and j possibly equal, i.e.

$$\mathcal{O}_{ij}^{m_i,m_j} := \frac{1}{|S_{ij}^{m_i,m_j}|} \sum_{S \in S_{ij}^{m_i,m_j}} S \; . \tag{30}$$

Since each transposition is Hermitian, $\mathcal{O}_{ij}^{m_i,m_j}$ is Hermitian too. Note that $|S_{ij}^{m_i,m_j}| = m_i m_j$ when $i \neq j$, while $|S_{ii}^{m_i,m_i}| = m_i(m_i - 1)/2$.

The expectation values of $\mathcal{D}_M$ on $\rho^{(M)}$ can be formally computed by exploiting the relations

$$\text{Tr}\Big[\mathcal{O}_{ii}^{m_i,m_i} \rho^{\vec{m}}\Big] = \text{Tr}\Big[\mathcal{O}_{ii}^{m_i,m_i} \rho_i^{\otimes m_i}\Big] = \text{Tr}\Big[\rho_i^2\Big] \; , \tag{31}$$

where the first identity follows from the fact that $\mathcal{O}_{ii}^{m_i,m_i}$ acts nontrivially only on registers containing copies of ρ_i, and

$$\text{Tr}\Big[\mathcal{O}_{ij}^{m_i,m_j} \rho^{\vec{m}}\Big] = \text{Tr}\Big[\mathcal{O}_{ij}^{m_i,m_j} \rho_i^{\otimes m_i} \otimes \rho_j^{\otimes m_j}\Big] = \text{Tr}[\rho_i \rho_j] \; , \tag{32}$$

where the first identity follows from the fact that $\mathcal{O}_{ij}^{m_i,m_j}$ acts not trivially only on registers containing copies of ρ_i and ρ_j. Accordingly for $i \neq j$ we have

$$\text{Tr}\Big[\mathcal{D}_{ij}^{m_i,m_j,M} \rho^{\vec{m}}\Big] = \frac{m_i(m_i - 1)}{\mu^2 p_i} p_j \, \text{Tr}\Big[\rho_i^2\Big] + \frac{m_j(m_j - 1)}{\mu^2 p_j} p_i \, \text{Tr}\Big[\rho_j^2\Big] - 2\frac{m_i m_j}{\mu^2} \text{Tr}[\rho_i \rho_j] \; , \tag{33}$$

which leads to

$$\text{Tr}\Big[\mathcal{D}_M \rho^{(M)}\Big] = \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \sum_{i \neq j} \left(\frac{m_i(m_i-1)}{\mu^2 p_i} p_j \, \text{Tr}\Big[\rho_i^2\Big] + \frac{m_j(m_j-1)}{\mu^2 p_j} p_i \, \text{Tr}\Big[\rho_j^2\Big] - 2\frac{m_i m_j}{\mu^2} \text{Tr}[\rho_i \rho_j] \right) \; . \tag{34}$$

To simplify the analysis of the performance of a test based on $\mathcal{D}_M$ we can invoke the equivalence of Proposition A.1 between the original model and its Poissonized version where the value of M (and hence the density matrix $\rho^{(M)}$ that are presented to us) is randomly generated with probability $\text{Poi}_\mu(M)$ (notice that the mean value of the distribution is taken equal to parameter μ which enters the definition (29) of $D_{ij}^{m_i,m_j,M}$). Defining Γ_M the set of eigenvalues of the observables $\mathcal{D}_M$ (27), we then introduce a new estimator $\mathcal{D}$ that produces outputs $X \in \Gamma := \bigcup_M \Gamma_M$ with probabilities

$$P_X := \sum_{M=0}^{\infty} \text{Poi}_\mu(M) \sum_{x \in \Gamma_M} \delta_{x,X} P_x^{(M)} \; , \tag{35}$$

where $P_x^{(M)}$ is the probability of getting the outcome x from $\mathcal{D}_M$ when acting on $\rho^{(M)}$.

The following facts can then be proved:

Proposition 3.1 (Unbiasedness). *Given* $\mathbb{E}[\mathcal{D}] := \sum_{X \in \Gamma} X P_X$ *the mean value of the estimator* $\mathcal{D}$ *we have*

$$\mathbb{E}[\mathcal{D}] = \mathcal{M}_{HS}^2(\rho) \; . \tag{36}$$

Proof. From Eq. (35) and (34) we can write

$$
\begin{aligned}
\mathbb{E}[\mathcal{D}] &= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{x \in \Gamma_M} x P_x^{(M)} = \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \mathrm{Tr}\left[\mathcal{D}_M \rho^{(M)}\right] \\
&= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \\
&\quad \times \sum_{i \neq j} \left(\frac{m_i(m_i-1)}{\mu^2 p_i} p_j \mathrm{Tr}\left[\rho_i^2\right] + \frac{m_j(m_j-1)}{\mu^2 p_j} p_i \mathrm{Tr}\left[\rho_j^2\right] - 2\frac{m_i m_j}{\mu^2} \mathrm{Tr}[\rho_i \rho_j] \right) \\
&= \sum_{m_1=0}^{\infty} \cdots \sum_{m_N=0}^{\infty} \mathrm{Poi}_{p_1 \mu}(m_1) \cdots \mathrm{Poi}_{p_N \mu}(m_N) \\
&\quad \times \sum_{i \neq j} \left(\frac{m_i(m_i-1)}{\mu^2 p_i} p_j \mathrm{Tr}\left[\rho_i^2\right] + \frac{m_j(m_j-1)}{\mu^2 p_j} p_i \mathrm{Tr}\left[\rho_j^2\right] - 2\frac{m_i m_j}{\mu^2} \mathrm{Tr}[\rho_i \rho_j] \right) , \quad (37)
\end{aligned}
$$

where in the second identity we used $\sum_{x \in \Gamma_M} x P_x^{(M)} = \mathrm{Tr}\left[\mathcal{D}_M \rho^{(M)}\right]$, while in the last identity we exploit the fact that under Poissonization the random variables m_i become independent due to the property

$$
\sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \mathsf{M}(\vec{m})_{\vec{p},M} = \prod_{i=1}^{N} \mathrm{Poi}_{p_i \mu}(m_i) , \quad (38)
$$

with $\mathrm{Poi}_{p_i \mu}(m_i)$ being a Poisson distribution of mean $p_i \mu$. Equation (36) then finally follows from the identities

$$
\sum_{m_i=0}^{\infty} m_i \, \mathrm{Poi}_{p_i \mu}(m_i) = \mu p_i , \qquad \sum_{m_i=0}^{\infty} \frac{m_i(m_i - 1)}{p_i} \, \mathrm{Poi}_{p_i \mu}(m_i) = \mu^2 p_i . \quad (39)
$$

$\square$

Proposition 3.2 (Bound on the variance). *The variance of the estimator $\mathcal{D}$, $\mathrm{Var}[\mathcal{D}] := \sum_{X \in \Gamma} P_X (X - \mathbb{E}[\mathcal{D}])^2$, satisfies the inequality*

$$
\mathrm{Var}[\mathcal{D}] \leq O\left(\frac{N}{\mu^2}\right) + \frac{16 \mathcal{M}_{HS}^2(\rho)}{\mu} . \quad (40)
$$

Proof. See Appendix B. $\square$

With these ingredients we can prove Theorem 3.1, following the proof of Lemma 2.1 of [BOW19], which is an application of Chebyshev inequality. We reproduce here the reasoning. Let us put $c = \mathcal{M}_{HS}^2(\rho)$. By Chebyshev's inequality, $P(|\mathcal{D} - c| \geq \epsilon) \leq \frac{\mathrm{Var}[\mathcal{D}]}{\epsilon^2}$. If $c < 0.99\delta$, then we have, for $C > 0$ large enough and $\mu = C\frac{\sqrt{N}}{\delta}$,

$$
P(|\mathcal{D} - c| \geq 0.005\delta) \leq \frac{\mathrm{Var}[\mathcal{D}]}{(0.005)^2 \delta^2} \leq \left(O(1)\frac{1}{C^2} + \frac{16}{C\sqrt{N}}\right) \delta^2 \frac{1}{(0.005)^2 \delta^2} \leq \frac{1}{3}, \quad (41)
$$

therefore $\mathcal{D} \leq 0.99\delta + 0.005\delta = 0.995\delta$ with high probability. If $c \geq \delta$, then we have, for $C > 0$ large enough and $\mu = C\frac{\sqrt{N}}{\delta}$,

$$
P(|\mathcal{D} - c| \geq 0.005c) \leq \frac{\mathrm{Var}[\mathcal{D}]}{(0.005)^2 c^2} \leq \left(O(1)\frac{1}{C^2} + \frac{16}{C\sqrt{N}}\right) c^2 \frac{1}{(0.005)^2 c^2} \leq \frac{1}{3}, \quad (42)
$$

therefore $\mathcal{D} \geq c - 0.005c \geq 0.995\delta$ with high probability.

13

4 Lower bound on the sample complexity

We now explain the idea for proving the lower bound on M that follows from Theorem 1.2. First of all we limit ourselves to even d, since for odd d one can simply use the lower bound for $d-1$. We also choose the probability distribution p to be uniform, $p_i = 1/N$. The case $N=2$ is a straightforward consequence of the lower bound in [OW15], which gives a lower bound of $\Omega(d/\epsilon^2)$, noting that with access to M copies of ρ_ϵ one can simulate access to M copies of $\frac{1}{2}\left(\frac{I_d}{d}\otimes|1\rangle\langle1| + \rho_\epsilon\otimes|2\rangle\langle2|\right)$:

Lemma 4.1 (Corollary 4.3 of [OW15]). *Let ρ_ϵ be a quantum state with $d/2$ eigenvalues equal to $\frac{1+2\epsilon}{d}$ and the other $d/2$ eigenvalues equal to $\frac{1-2\epsilon}{d}$. Then any algorithm that can discern between the states $(I_d/d)^{\otimes M}$ and $\rho_\epsilon^{\otimes M}$ with a probability greater than 2/3 must require $M \geq 0.15d/\epsilon^2$.*

This is a lower bound for any N smaller than a constant, say $N < 10$. Therefore we consider $N \geq 10$ in the following. We define two sets of collections of N quantum states. The first set A contains only one collection, namely a collection where all the states are the maximally mixed states. Clearly, the only element of A is a collection satisfying the property of case A. For even d, the second set B contains all the collections of states having $d/2$ eigenvalues equal to $\frac{1+8\epsilon}{d}$ and $d/2$ eigenvalues equal to $\frac{1-8\epsilon}{d}$. This means that all the states in a collection of B can be written as $U_i\rho_0 U_i^\dagger$ for ρ_0 with the prescribed spectrum and U_i arbitrary. If each U_i is drawn independently according to the Haar measure of $\mathrm{SU}(d)$, we show that the elements of B satisfy property B with probability larger than a constant. We also show an upper bound on the trace distance between ρ_A and ρ_B, being respectively M samples for a collection of all maximally mixed states and the average input of M samples for collections in B. Explicitly, we have

$$\rho_A = \left(\frac{1}{N}\sum_{i=1}^N |i\rangle\langle i|\otimes\frac{I}{d}\right)^{\otimes M}, \tag{43}$$

$$\rho_B = \int_{U_1,\dots,U_N\in\mathrm{SU}(d)} \mathrm{d}U_1\dots\mathrm{d}U_N \left(\frac{1}{N}\sum_{i=1}^N |i\rangle\langle i|\otimes U_i\rho_0 U_i^\dagger\right)^{\otimes M}. \tag{44}$$

If a test capable of distinguishing with high probability between case A and case B exists, then it can be used to distinguish between ρ_A and ρ_B. Since the probability of success in the latter task has to be lower than what we obtain from the bound on the trace distance, we obtain a lower bound on the sample complexity.

Lemma 4.2. *Let $\{\rho_i\}_{i,\dots,N}$ be a collection of states such that $\frac{1}{N}\sum_{i=1}^N||\rho_i - \bar\rho||_1 > 4\epsilon$.*
Then $\frac{1}{N}\sum_{i=1}^N||\rho_i - \sigma||_1 > 2\epsilon$ for any σ.

Proof. Suppose that we have $\frac{1}{N}\sum_{i=1}^N||\rho_i - \sigma||_1 \leq 2\epsilon$ for some σ. By monotonicity of the trace distance, $||\bar\rho - \sigma||_1 \leq 2\epsilon$. Then

$$\frac{1}{N}\sum_{i=1}^N||\rho_i - \bar\rho||_1 = \frac{1}{N}\sum_{i=1}^N||\rho_i - \sigma + \sigma - \bar\rho||_1 \leq \frac{1}{N}\sum_{i=1}^N||\rho_i - \sigma||_1 + ||\sigma - \bar\rho||_1 \leq 4\epsilon \tag{45}$$

which is a contradiction. □

Lemma 4.3. *For $N > 10$, let $\{U_i\rho_0 U_i^\dagger\}_{i,\dots,N}$ be a collection of states in B and ρ as in Eq. (1), with $p_i = 1/N$. If each U_i is drawn independently according to the Haar measure of $\mathrm{SU}(d)$, the probability of having $\mathcal{M}_{\mathrm{Tr}}(\rho) \geq 2\epsilon$ is at least*

$$\mathop{\mathrm{P}}_{U_1,\dots,U_N\sim\mathbf{U}(d)}\left(\mathcal{M}_{\mathrm{Tr}}(\rho) > 2\epsilon\right) \geq \frac{11}{15}. \tag{46}$$

Proof. We denote $|k\rangle_{k=1,\dots,d}$ a basis of eigenvectors of ρ_0, such that

$$\langle k| \rho_0 |k\rangle = \frac{1 + (-1)^k 8\epsilon}{d} \ , \tag{47}$$

and define

$$\Theta := \sum_{k=1}^{d} (-1)^k |k\rangle\langle k| \ , \tag{48}$$

We can write

$$2\mathcal{M}_{\mathrm{Tr}}(\rho) = \frac{1}{N} \sum_{i=1}^{N} \|\rho_i - \bar{\rho}\|_1 = \frac{1}{N} \sum_{i=1}^{N} \left\| \rho_i - \frac{1}{N} \sum_{j=1}^{N} U_j \rho_0 U_j^\dagger \right\|_1 = \frac{1}{N} \sum_{i=1}^{N} \left\| U_i \rho_0 U_i^\dagger - \frac{1}{N} \sum_{j=1}^{N} U_j \rho_0 U_j^\dagger \right\|_1$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left\| \rho_0 - \frac{1}{N} \sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right\|_1 \geq \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{d} \left| \langle k| \rho_0 - \frac{1}{N} \sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right| \ . \tag{49}$$

We can observe now that, from (47) it follows that $\frac{1 \pm 8\epsilon}{d}$ are maximum/minimum eigenvalues of ρ_0, so that

$$\langle k| \rho_0 |k\rangle \, (k \text{ odd}) = \frac{1 - 8\epsilon}{d} \leq \langle k|U_i^\dagger U_j \rho_0 U_j^\dagger U_i|k\rangle \leq \frac{1 + 8\epsilon}{d} = \langle k| \rho_0 |k\rangle \, (k \text{ even}) \ ; \tag{50}$$

and therefore

$$\sum_{k=1}^{d} \left| \langle k|\rho_0 - \frac{1}{N} \sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right| = (-1)^k \left(\langle k|\rho_0|k\rangle - \langle k| \frac{1}{N} \sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right) \tag{51}$$

Replacing (51) into (49) we have

$$2\mathcal{M}_{\mathrm{Tr}}(\rho) \geq \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{d} (-1)^k \left(\langle k|\rho_0|k\rangle - \langle k| \frac{1}{N} \sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right)$$

$$= 8\epsilon - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{d} (-1)^k \sum_{j=1}^{N} \langle k|U_i^\dagger U_j \rho_0 U_j^\dagger U_i|k\rangle$$

$$= 8\epsilon - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{d} (-1)^k \langle k|U_i^\dagger U_j \rho_0 U_j^\dagger U_i|k\rangle$$

$$= 8\epsilon - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] \tag{52}$$

Now observe that $\rho_0 = \frac{1}{d}(I + 8\epsilon\Theta)$. Therefore

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] = \frac{1}{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{Tr}\left[\hat{\Theta} U_i^\dagger U_j U_j^\dagger U_i \right] + \frac{8\epsilon}{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{Tr}\left[\hat{\Theta} U_i^\dagger U_j \hat{\Theta} U_j^\dagger U_i \right]$$

$$= \frac{8\epsilon}{d} \sum_{i=1}^{N} \mathrm{Tr}\left[\left(\sum_{i=1}^{N} U_i \hat{\Theta} U_i^\dagger \right) \left(\sum_{j=1}^{N} U_j \hat{\Theta} U_j^\dagger \right) \right] \geq 0. \tag{53}$$

Since the latter term of (52) is always positive, we may use the Markov's inequality on it. Its expected value is:

$$\mathbb{E}_{U_1,...,U_N \sim \mathbf{U}(d)} \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \text{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] \right] = \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N \mathbb{E}_{U_1,...,U_N \sim \mathbf{U}(d)} \left[\text{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] \right]$$

$$= \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N 8\epsilon \delta_{ij} = 8\frac{\epsilon}{N} \ . \tag{54}$$

Therefore, using Markov inequality, we can write

$$\mathop{\mathbf{P}}_{U_1,...,U_N \sim \mathbf{U}(d)} \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \text{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] \geq 3\epsilon \right) \leq \frac{8}{3N} \tag{55}$$

Combining (55) with (52), we have

$$\mathop{\mathbf{P}}_{U_1,...,U_N \sim \mathbf{U}(d)} \left(\mathcal{M}_{\text{Tr}}(\rho) > 2\epsilon \right) \geq 1 - \frac{8}{3N} \geq \frac{11}{15}, \qquad N \geq 10 \tag{56}$$

$$\square$$

Lemma 4.4.

$$D_{\text{Tr}}(\rho_A, \rho_B) \leq 16 \frac{\epsilon^2 M}{d\sqrt{N}} \tag{57}$$

Proof. We have that

$$D_{\text{Tr}}(\rho_A, \rho_B) = E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \left[D\left(\left(\frac{I}{d}\right)^{\otimes M}, \int_{U_i \in \text{SU}(d)} \mathrm{d}U_1....\mathrm{d}U_N \bigotimes_{i=1}^N \left(U_i \rho_0 U_i^\dagger \right)^{\otimes m_i} \right) \right] \tag{58}$$

Using Schur-Weyl duality, we can write ρ_A and ρ_B as

$$\left(\frac{I}{d}\right)^{\otimes M} = \bigotimes_{i=1}^N \left(\sum_{\lambda \in Y_{m_i,d}} \text{SW}_{I/d}^{m_i}(\lambda) \frac{I_{d(\lambda,m_i) \times d(\lambda,m_i)}}{d(\lambda, m_i)} \right) \tag{59}$$

$$\int_{U_i \in \text{SU}(d)} \mathrm{d}U_1....\mathrm{d}U_N \bigotimes_{i=1}^N \left(U_i \rho_0 U_i^\dagger \right)^{\otimes m_i} = \bigotimes_{i=1}^N \left(\sum_{\lambda \in Y_{m_i,d}} \text{SW}_{\rho_0}^{m_i}(\lambda) \frac{I_{d(\lambda,m_i) \times d(\lambda,m_i)}}{d(\lambda, m_i)} \right), \tag{60}$$

where $Y_{m_i,d}$ is a set of Young diagrams and $\text{SW}_\rho^M(\lambda)$ is a probability distribution over Young diagrams which depends only on the spectrum of ρ. Defining

$$\mathfrak{D}_0^{\vec{m}} = \text{SW}_d^{m_1} \times \cdots \text{SW}_d^{m_i} \ , \qquad \mathfrak{D}_\epsilon^{\vec{m}} = \text{SW}_{\rho_0}^{m_1} \times \cdots \text{SW}_{\rho_0}^{m_i}, \tag{61}$$

we have

$$D_{\text{Tr}}(\rho_A, \rho_B) = E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} d_{TV}(\mathfrak{D}_0^{\vec{m}}, \mathfrak{D}_\epsilon^{\vec{m}}) \tag{62}$$

First of all we invoke the from [OW15]:

$$\mathrm{d}_{\chi^2}(\text{SW}_\rho^n || \text{SW}_{I/d}^n) \leq \exp\left(256 n^2 \epsilon^4 / d^2 \right) - 1 \tag{63}$$

16

Our first observation is that, when $m_i = 1$, (63) can be improved noticing that $d_{KL}(\mathrm{SW}^1_{\rho_i}, \mathrm{SW}^1_{I/d}) = 0$ for every possible state ρ_i (since there is only one possible partition of $n = 1$ - in other words, we gain no information on whether the state is mixed by measuring a single copy). This observation, together with (63) and (19), imply that

$$\mathrm{d}_{KL}(\mathrm{SW}^{m_i}_{\rho}||\mathrm{SW}^{m_i}_{I/d}) \leq 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2} \; . \tag{64}$$

Using (17) and (64) we can write

$$
\begin{aligned}
D_{\mathrm{Tr}}(\rho_A, \rho_B) &= E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} d_{TV}(\mathfrak{D}^{\vec{m}}_0, \mathfrak{D}^{\vec{m}}_\epsilon) \leq E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} d_{KL}(\mathfrak{D}^{\vec{m}}_0, \mathfrak{D}^{\vec{m}}_\epsilon)} \\
&= E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} \sum_{i=1}^N \mathrm{d}_{KL}(\mathrm{SW}^{m_i}_{\rho}||\mathrm{SW}^{m_i}_{I/d})} \leq E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} \sum_{i=1}^N 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2}} \\
&\leq \sqrt{E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \frac{1}{2} \sum_{i=1}^N 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2}} \leq \sqrt{E_{\vec{m} \sim \mathsf{M}_{\vec{p},N,M}} \sum_{i=1}^N 256 m_i(m_i - 1) \frac{\epsilon^4}{d^2}} \\
&\leq 16 \frac{\epsilon^2 M}{d\sqrt{N}},
\end{aligned}
\tag{65}
$$

where the first inequality is from Pinsker's inequality, the second equality is the additivity of the Kullback-Leibler divergence, the third inequality is from concavity of the square root. $\quad\square$

It is now immediate to prove Theorem 1.2

Proof of Theorem 1.2. If an algorithm as in Theorem 1.2 exists, one can use it to try to discriminate between ρ_A and ρ_B. By also invoking the Holevo-Helstrom bound Eq. (10), the probability of success has to satisfy

$$\frac{1}{2}\left(1 + 16 \frac{\epsilon^2 M}{d\sqrt{N}}\right) \geq p_{succ} \geq \frac{1}{2}\left(\frac{11}{15} + 1\right) \frac{2}{3} \; . \tag{66}$$

Therefore

$$M \geq 4 \cdot 10^{-3} \frac{\sqrt{N} d}{\epsilon^2}. \tag{67}$$

$\square$

5 Implementation of the optimal measurement

The measurement of the test defined in Section 3 to prove Theorem 1.1 can be implemented on a quantum computer with gate complexity $O(M, \log d, \log 1/\delta)$, where δ is the precision of the implementation, because it can be realized with a sequence of weak Schur sampling measurements. This was already shown for the observable of [BOW19] for $N = 2$ and it can be easily be shown to be true in the general case too. Indeed, in [BOW19] it is shown that $\mathcal{O}^{m_i,m_i}_{ii}$ can be written as

$$\mathcal{O}^{m_i,m_i}_{ii} = \sum_{\lambda \in Y_{m_i,d}} \mathsf{TN}(\lambda) \Pi^{(i)}_\lambda, \tag{68}$$

where $Y_{m_i,d}$ are Young diagrams, Π_λ a complete set of orthogonal projectors and $\mathsf{TN}(\lambda) = \frac{1}{n(n-1)} \sum_{i=1}^{d}((\lambda_i - i + 1/2)^2 - (-i + 1/2)^2)$. We now define $\mathcal{O}$ to be the average of all transposition on $\mathcal{H}_d^{\otimes M}$, for which we have:

$$\mathcal{O} = \sum_{\lambda \in Y_{M,d}} \mathsf{TN}(\lambda)\Pi_\lambda. \tag{69}$$

Using that

$$\frac{M(M-1)}{2}\mathcal{O} = \frac{1}{2}\sum_{i \neq j} m_i m_j \mathcal{O}_{ij}^{m_i, m_j} + \sum_{i=1}^{N} \frac{m_i(m_i - 1)}{2}\mathcal{O}_{ii}^{m_i, m_i}, \tag{70}$$

we have

$$\mathcal{D}^{\vec{m},M} := \sum_{i \neq j} \mathcal{D}_{ij}^{m_i, m_j, M} = \sum_{i=1}^{N} \frac{2m_i(m_i - 1)(1 - p_i)}{\mu^2 p_i}\mathcal{O}_{ii}^{m_i, m_i} - \sum_{i \neq j} 2\frac{m_i m_j}{\mu^2}\mathcal{O}_{ij}^{m_i, m_j}.$$

$$= \sum_{i=1}^{N} \frac{2m_i(m_i - 1)}{\mu^2 p_i}\mathcal{O}_{ii}^{m_i, m_i} - \frac{2M(M-1)}{\mu^2}\mathcal{O}. \tag{71}$$

Since $[\Pi_\lambda, \otimes_{i=1}^{N}\Pi_{\lambda_i}^{(i)}] = 0$, the measurement can be implemented efficiently by nested weak Schur sampling.

6 Conclusions and remarks

We have established the sample complexity of testing identity of collections of quantum states in the sampling model, with a test that can be also implemented efficiently in terms of gate complexity. Note that for this problem one could have used the independence tester of [Yu21], based on the identity test of [BOW19], since if the state in the collection are equal the input of our problem in Eq. (1) is a product state, and far from it otherwise. However, the guaranteed sample complexity in this case would have been $O(Nd/\epsilon^2)$, and to get $\sqrt{N}d/\epsilon^2$ we need to make use of the fact that the state in Eq. (1) is a classical-quantum state and that we know the classical marginal. This is a state of zero discord [HV01; OZ01; ABC16], and one could ask how the sample complexity differ if the discord is not zero, for example if the states $|i\rangle$ are not orthogonal. This could be seen as an example of quantum inference problem with quantum flags, proved useful in other contexts, e.g. the evaluation of quantum capacities [SSW08; LDS18; FKG20; KFG22; Wan21; FKG21]. More generally, an interesting problem would be to study the sample complexity of independence testing with constraints on the structure of the state, with a rich variety of scenarios possible.

7 Acknowledgment

References

[ABC16]   Gerardo Adesso, Thomas R. Bromley, and Marco Cianciaruso. "Measures and applications of quantum correlations". In: *Journal of Physics A: Mathematical and Theoretical* 49.47 (2016), p. 473001. DOI: 10.1088/1751-8113/49/47/473001. arXiv: 1605.00806.

[Ach+20] Jayadev Acharya, Ibrahim Issa, Nirmal V. Shende, and Aaron B. Wagner. "Estimating Quantum Entropy". In: *IEEE Journal on Selected Areas in Information Theory* 1.2 (2020), pp. 454–468. DOI: 10.1109/JSAIT.2020.3015235.

[AD15]   Jayadev Acharya and Constantinos Daskalakis. "Testing Poisson Binomial Distributions". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015, pp. 1829–1840. DOI: 10.1137/1.9781611973730.122. arXiv: 1507.05952.

[AH11]   Daiki Akimoto and Masahito Hayashi. "Discrimination of the change point in a quantum setting". In: *Physical Review A* 83.5 (2011), p. 052328. DOI: 10.1103/PhysRevA.83.052328. arXiv: 1102.2555.

[ARS88]   Robert Alicki, Slawomir Rudnicki, and Slawomir Sadowski. "Symmetry properties of product states for the system of N n-level atoms". In: *Journal of Mathematical Physics* 29.5 (1988), pp. 1158–1162. DOI: 10.1063/1.527958.

[Bai+22] Ge Bai, Ya-Dong Wu, Yan Zhu, Masahito Hayashi, and Giulio Chiribella. "Quantum causal unravelling". In: *npj Quantum Information* 8.1 (2022), p. 69. DOI: 10.1038/s41534-022-00578-4. arXiv: 2109.13166.

[Bat+01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. "Testing random variables for independence and identity". In: *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001, pp. 442–451. DOI: 10.1109/SFCS.2001.959920.

[BCH06]   Dave Bacon, Isaac L. Chuang, and Aram W. Harrow. "Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms". In: *Physical Review Letters* 97.17 (2006), p. 170502. DOI: 10.1103/PhysRevLett.97.170502. arXiv: 0407082 [quant-ph].

[BCL20]   Sebastien Bubeck, Sitan Chen, and Jerry Li. "Entanglement is Necessary for Optimal Quantum Property Testing". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 692–703. DOI: 10.1109/FOCS46700.2020.00070. arXiv: 2004.07869.

[Ben+14] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. "The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels". In: *IEEE Transactions on Information Theory* 60.5 (2014), pp. 2926–2959. DOI: 10.1109/TIT.2014.2309968. arXiv: arXiv:0912.5537.

[BIMT06]   E. Bagan, S. Iblisdir, and R. Muñoz-Tapia. "Relative states, quantum axes, and quantum references". In: *Physical Review A* 73.2 (2006), p. 022341. DOI: 10.1103/PhysRevA.73.022341. arXiv: 0508187 [quant-ph].

[BLM13]   Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities.* Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199535255.001.0001.

[BOW19]   Costin Bădescu, Ryan O'Donnell, and John Wright. "Quantum state certification". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing.* New York, NY, USA: ACM, 2019, pp. 503–514. DOI: 10.1145/3313276.3316344. arXiv: 1708.06002.

[BRS04]   Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Optimal measurements for relative quantum information". In: *Physical Review A* 70.3 (2004), p. 032321. DOI: 10.1103/PhysRevA.70.032321. arXiv: 0310009 [quant-ph].

[Buh+01]   Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. "Quantum Fingerprinting". In: *Physical Review Letters* 87.16 (2001), p. 167902. DOI: 10.1103/PhysRevLett.87.167902. arXiv: 0102001 [quant-ph].

[Can20]   Clement L. Canonne. "A Survey on Distribution Testing: Your Data is Big. But is it Blue?" In: *Theory of Computing* 1.1 (2020), pp. 1–100. DOI: 10.4086/toc.gs.2020.009.

[Cha+14]   Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. "Optimal Algorithms for Testing Closeness of Discrete Distributions". In: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 2014, pp. 1193–1203. DOI: 10.1137/1.9781611973402.88. arXiv: 1308.3946.

[Chr06]   Matthias Christandl. *The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography.* 2006. arXiv: 0604183 [quant-ph].

[CLO22]   Sitan Chen, Jerry Li, and Ryan O'Donnell. "Toward Instance-Optimal State Certification With Incoherent Measurements". In: *Proceedings of Thirty Fifth Conference on Learning Theory.* Ed. by Po-Ling Loh and Maxim Raginsky. Vol. 178. Proceedings of Machine Learning Research. https://proceedings.mlr.press/v178/chen22b.html. PMLR, 2022, pp. 2541–2596. arXiv: 2102.13098.

[CT05]   Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* 2005. DOI: 10.1002/047174882X.

[DK16]   Ilias Diakonikolas and Daniel M. Kane. "A New Approach for Testing Properties of Discrete Distributions". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS).* IEEE, 2016, pp. 685–694. DOI: 10.1109/FOCS.2016.78. arXiv: 1601.05557.

[DKN15]   Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin. "Testing Identity of Structured Distributions". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms.* Vol. 2015-Janua. January. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015, pp. 1841–1854. DOI: 10.1137/1.9781611973730.123.

[Fan+20]   M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. "Beyond the Swap Test: Optimal Estimation of Quantum State Overlap". In: *Physical Review Letters* 124.6 (2020), p. 060503. DOI: 10.1103/PhysRevLett.124.060503. arXiv: 1906.10639.

[FHC23]   Marco Fanizza, Christoph Hirche, and John Calsamiglia. "Ultimate Limits for Quickest Quantum Change-Point Detection". In: *Phys. Rev. Lett.* 131 (2 2023), p. 020602. DOI: 10.1103/PhysRevLett.131.020602. arXiv: 2208.03265.

[FKG20]   Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. "Quantum Flags and New Bounds on the Quantum Capacity of the Depolarizing Channel". In: *Physical Review Letters* 125.2 (2020), p. 020503. DOI: 10.1103/PhysRevLett.125.020503. arXiv: 1911.01977.

[FKG21]   Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. "Estimating Quantum and Private Capacities of Gaussian Channels via Degradable Extensions". In: *Phys. Rev. Lett.* 127 (21 2021), p. 210501. DOI: 10.1103/PhysRevLett.127.210501. arXiv: 2103.09569.

[GI06]   N. Gisin and S. Iblisdir. "Quantum relative states". In: *The European Physical Journal D* 39.2 (2006), pp. 321–327. DOI: 10.1140/epjd/e2006-00097-y. arXiv: 0507118 [quant-ph].

[Gol17]   Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. DOI: 10.1017/9781108135252.

[GR11]   Oded Goldreich and Dana Ron. "On Testing Expansion in Bounded-Degree Graphs". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2011, pp. 68–75. DOI: 10.1007/978-3-642-22670-0˙9.

[Haa+17]   Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. "Sample-optimal tomography of quantum states". In: *IEEE Transactions on Information Theory* 63.9 (2017), pp. 1–1. DOI: 10.1109/TIT.2017.2719044. arXiv: 1508.01797.

[Har05]   Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. 2005. arXiv: 0512255 [quant-ph].

[Hay+06]   Masahito Hayashi, Bao-Sen Shi, Akihisa Tomita, Keiji Matsumoto, Yoshiyuki Tsuda, and Yun-Kun Jiang. "Hypothesis testing for an entangled state produced by spontaneous parametric down-conversion". In: *Phys. Rev. A* 74 (6 2006), p. 062321. DOI: 10.1103/PhysRevA.74.062321.

[Hay17a]   Masahito Hayashi. *A Group Theoretic Approach to Quantum Information*. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-45241-8.

[Hay17b]   Masahito Hayashi. *Group Representation for Quantum Theory*. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-44906-7.

[Hay17c]   Masahito Hayashi. *Quantum Information Theory*. Ed. by Springer. Graduate Texts in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017. DOI: 10.1007/978-3-662-49725-8.

[HM02]   Masahito Hayashi and Keiji Matsumoto. "Quantum universal variable-length source coding". In: *Physical Review A* 66.2 (2002), p. 022311. DOI: 10.1103/PhysRevA.66.022311. arXiv: 0202001 [quant-ph].

[HT16]   Masahito Hayashi and Marco Tomamichel. "Correlation detection and an operational interpretation of the Rényi mutual information". In: *Journal of Mathematical Physics* 57.10 (2016), p. 102201. DOI: 10.1063/1.4964755. arXiv: 1408.6894.

[HTM08]   Masahito Hayashi, Akihisa Tomita, and Keiji Matsumoto. "Statistical analysis of testing of an entangled state based on the Poisson distribution framework". In: *New Journal of Physics* 10.4 (2008), p. 043029. DOI: 10.1088/1367-2630/10/4/043029.

[HV01]   L. Henderson and V. Vedral. "Classical, quantum and total correlations". In: *Journal of Physics A: Mathematical and General* 34.35 (2001), pp. 6899–6905. DOI: 10.1088/0305-4470/34/35/315. arXiv: 0105028 [quant-ph].

[Key06]   M. Keyl. "Quantum state estimation and large deviations". In: *Reviews in Mathematical Physics* 18.1 (2006), pp. 19–60. DOI: 10.1142/S0129055X06002565.

[KFG22]   Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. "Bounding the quantum capacity with flagged extensions". In: *Quantum* 6 (Feb. 2022), p. 647. DOI: 10.22331/q-2022-02-09-647. arXiv: 2008.02461.

[KR21]   Martin Kliesch and Ingo Roth. "Theory of Quantum System Certification". In: *PRX Quantum* 2.1 (2021), p. 010201. DOI: 10.1103/PRXQuantum.2.010201. arXiv: 2010.05925.

[Kro19]   Hari Krovi. "An efficient high dimensional quantum Schur transform". In: *Quantum* 3 (2019), p. 122. DOI: 10.22331/q-2019-02-14-122. arXiv: 1804.00055.

[KW01]   M. Keyl and R. F. Werner. "Estimating the spectrum of a density operator". In: *Physical Review A* 64.5 (2001), p. 052311. DOI: 10.1103/PhysRevA.64.052311. arXiv: 0102027 [quant-ph].

[LC60]   Lucien Le Cam. "An approximation theorem for the Poisson binomial distribution." In: *Pacific Journal of Mathematics* 10.4 (1960), pp. 1181–1197.

[LDS18]   Felix Leditzky, Nilanjana Datta, and Graeme Smith. "Useful States and Entanglement Distillation". In: *IEEE Transactions on Information Theory*. Vol. 64. 7. 2018, pp. 4689–4708. DOI: 10.1109/TIT.2017.2776907. arXiv: 1701.03081.

[LR06]   Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses.* Springer Science & Business Media, 2006.

[LRR13]   Reut Levi, Dana Ron, and Ronitt Rubinfeld. "Testing Properties of Collections of Distributions". In: *Theory of Computing* 9.8 (2013), pp. 295–347. DOI: 10.4086/toc.2013.v009a008.

[LSB06]   Netanel H. Lindner, Petra F. Scudo, and Dagmar Bruß. "Quantum estimation of relative information". In: *International Journal of Quantum Information* 4.1 (2006), pp. 131–149. DOI: 10.1142/S0219749906001657. arXiv: 0506223 [quant-ph].

[MW16]   Ashley Montanaro and Ronald de Wolf. "A survey of quantum property testing". In: *Theory of Computing* 1.1 (2016), pp. 1–81. DOI: 10.4086/toc.gs.2016.007. arXiv: 1310.2035.

[OW15]   Ryan O'Donnell and John Wright. "Quantum Spectrum Testing". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*. Vol. 14-17-June. New York, NY, USA: ACM, 2015, pp. 529–538. DOI: 10.1145/2746539.2746582. arXiv: 1501.05028.

[OW16]   Ryan O'Donnell and John Wright. "Efficient quantum tomography". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. Vol. 19-21-June. New York, NY, USA: ACM, 2016, pp. 899–912. DOI: 10.1145/2897518.2897544. arXiv: 1508.01907.

[OW17]   Ryan O'Donnell and John Wright. "Efficient quantum tomography II". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: ACM, 2017, pp. 962–974. DOI: 10.1145/3055399.3055454. arXiv: 1612.00034.

[OZ01]   Harold Ollivier and Wojciech H Zurek. "Quantum Discord: A Measure of the Quantumness of Correlations". In: *Physical Review Letters* 88.1 (2001), p. 017901. DOI: 10.1103/PhysRevLett.88.017901. arXiv: 0105072 [quant-ph].

[Pan08]   Liam Paninski. "A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data". In: *IEEE Transactions on Information Theory* 54.10 (2008), pp. 4750–4755. DOI: 10.1109/TIT.2008.928987.

[SCMT17]   Gael Sentís, John Calsamiglia, and Ramon Munoz-Tapia. "Exact Identification of a Quantum Change Point". In: *Physical Review Letters* 119.14 (2017). DOI: 10.1103/PhysRevLett.119.140506. arXiv: 1707.07769.

[Sen+16]   Gael Sentís, Emilio Bagan, John Calsamiglia, Giulio Chiribella, and Ramon Munoz-Tapia. "Quantum change point". In: *Physical Review Letters* 117.15 (2016). DOI: 10.1103/PhysRevLett.117.150502. arXiv: 1605.01916.

[SMVMT18] Gael Sentís, Esteban Martínez-Vargas, and Ramon Muñoz-Tapia. "Online strategies for exactly identifying a quantum change point". In: *Physical Review A* 98.5 (2018), p. 052305. DOI: 10.1103/PhysRevA.98.052305. arXiv: 1802.00280.

[SSW08]   Graeme Smith, John A. Smolin, and Andreas Winter. "The quantum capacity with symmetric side channels". In: *IEEE Transactions on Information Theory* 54.9 (2008), pp. 4208–4217. DOI: 10.1109/TIT.2008.928269. arXiv: 0607039 [quant-ph].

[SV16]   Igal Sason and Sergio Verdu. "f-Divergence Inequalities". In: *IEEE Transactions on Information Theory* 62.11 (2016), pp. 5973–6006. DOI: 10.1109/TIT.2016.2603151. arXiv: 1508.00335.

[VV14]   Gregory Valiant and Paul Valiant. "An Automatic Inequality Prover and Instance Optimal Identity Testing". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 2014, pp. 51–60. DOI: 10.1109/FOCS.2014.14.

[Wan21]   Xin Wang. "Pursuing the fundamental limits for quantum communication". In: *IEEE Transactions on Information Theory* 67.7 (2021), pp. 4524–4532. DOI: 10.1109/TIT.2021.3068818. arXiv: 1912.00931.

[Yu21]   Nengkun Yu. "Sample Efficient Identity Testing and Independence Testing of
           Quantum States". In: *12th Innovations in Theoretical Computer Science
           Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International
           Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss
           Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 11:1–11:20. DOI:
           10.4230/LIPIcs.ITCS.2021.11. arXiv: 1904.03218.

[Yu23]   Nengkun Yu. "Almost Tight Sample Complexity Analysis of Quantum Identity
           Testing by Pauli Measurements". In: *IEEE Transactions on Information
           Theory* 69.8 (2023), pp. 5060–5068. DOI: 10.1109/TIT.2023.3271206. arXiv:
           2009.11518.

A Equivalence of sampling model and Poissonized model

The equivalence of the Poisson model with the original one can be formalised in the following
propositions.

Proposition A.1. *Suppose that given access to M copies of the state ρ of Eq. (1), where M
is extracted from a Poisson distribution with mean μ, there is a test `Ptest` such that*

$$
\begin{cases}
P(\texttt{Ptest} \mapsto \texttt{"accept"} \mid Case\ \mathsf{A}) > 3/4 \,, \\[2mm]
P(\texttt{Ptest} \mapsto \texttt{"accept"} \mid Case\ \mathsf{B}) < 1/4 \,,
\end{cases}
\tag{72}
$$

*and it can be performed by a two-outcome POVM $\{E_0^{(M)}, E_1^{(M)}\}$ for each M. Then, provided
that μ is larger than a fixed constant, there is a test in the sampling model using 2μ copies of ρ
satisfying*

$$
\begin{cases}
P(\texttt{test} \mapsto \texttt{"accept"} \mid Case\ \mathsf{A}) > 2/3 \,, \\[2mm]
P(\texttt{test} \mapsto \texttt{"accept"} \mid Case\ \mathsf{B}) < 1/3 \,.
\end{cases}
\tag{73}
$$

Proof. Given 2μ copies of ρ, we construct the following test. We extract M from a Poisson
distribution with mean μ. If $M \leq 2\mu$, we perform the measurement $\{E_0^{(M)}, E_1^{(M)}\}$, otherwise
we declare failure. The difference of the acceptance probabilities of `test` and `Ptest` is

$$
\begin{aligned}
&P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \\
&= \sum_{M=0}^{2\mu} \mathrm{Poi}_\mu(M) \left(\mathrm{Tr}\!\left[E_0^{(M)} \rho^{\otimes M} \right] - \mathrm{Tr}\!\left[E_0^{(M)} \rho^{\otimes M} \right] \right) + \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) \left(\mathrm{Tr}\!\left[E_0^{(M)} \rho^{\otimes M} \right] - 0 \right) \\
&= \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) \, \mathrm{Tr}\!\left[E_0^{(M)} \rho^{\otimes M} \right],
\end{aligned}
\tag{74}
$$

which implies

$$
0 \leq P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \leq \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) = P_{M \sim \mathrm{Poi}_\mu}(M > 2\mu).
\tag{75}
$$

Invoking hence the Cramér-Chernoff tail bound on the Poisson distribution [BLM13], i.e.

$$
P_{M \sim \mathrm{Poi}_\mu}(M \geq t) \leq e^{-\mu h(t/\mu)} \qquad h(x) = (1+x)\log(1+x) - x \,,
\tag{76}
$$

and setting $\mu > 2$, from Eq. (75) we then get

$$0 \le P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \le e^{-\mu h(2)} < 1/10 \;, \tag{77}$$

from which the statement of the proposition follows. $\qquad\square$

Proposition A.2. *Suppose that given access to M copies of the state ρ of Eq. (1), there is a test* Ptest *such that*

$$\begin{cases} P(\texttt{Ptest} \;\mapsto\; \texttt{"accept"} \,|\, Case\ \mathsf{A}) > 3/4 \;, \\[2mm] P(\texttt{Ptest} \;\mapsto\; \texttt{"accept"} \,|\, Case\ \mathsf{B}) < 1/4 \;, \end{cases} \tag{78}$$

and it can be performed by a two-outcome POVM $\{E_0^{(M)}, E_1^{(M)}\}$. Then, provided that M is larger than a fixed constant, there is a test in the Poissonized sampling model using M' copies of ρ where M' is extracted from a Poisson distribution with mean $2M$, satisfying

$$\begin{cases} P(\texttt{test} \;\mapsto\; \texttt{"accept"} \,|\, Case\ \mathsf{A}) > 2/3 \;, \\[2mm] P(\texttt{test} \;\mapsto\; \texttt{"accept"} \,|\, Case\ \mathsf{B}) < 1/3 \;. \end{cases} \tag{79}$$

Proof. We have that [BLM13],

$$P_{M' \sim \mathrm{Poi}_{2M}}(M' \le t) \le e^{-2Mh(-t/(2M))} \qquad h(x) = (1+x)\log(1+x) - x \;. \tag{80}$$

Therefore, if $M > 16$

$$P_{M' \sim \mathrm{Poi}_{2M}}(M' \le M) \le e^{-2Mh(-1/2)} < 1/10. \tag{81}$$

with high probability $M' > M$ and we can use Ptest on M copies. $\qquad\square$

B Proof of Proposition 3.2

As in the proof of Proposition 3.1 we can invoke Eqs. (35), (34) and the identity $\sum_{x \in \Gamma_M} x^2 P_x^{(M)} = \mathrm{Tr}\!\left[\mathcal{D}_M^2 \rho^{(M)}\right]$ to write

$$\begin{aligned} \mathrm{Var}[\mathcal{D}] \;&=\; \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \, \mathrm{Tr}\!\left[\mathcal{D}_M^2 \rho^{(M)}\right] - \mathbb{E}[\mathcal{D}]^2 \\ &=\; \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \, \mathrm{Tr}\!\left[(D^{\vec{m},M})^2 \rho^{\vec{m}}\right] - \mathbb{E}[\mathcal{D}]^2 \;, \end{aligned} \tag{82}$$

where the last passage involves (27) and (24). Replacing Eqs. (24), (28), and (29) into $\mathrm{Tr}\!\left[(D^{\vec{m},M})^2 \rho^{\vec{m}}\right]$ reveals that such term can be written as a linear combination of the expectation values of the operators $\mathcal{O}_{ij}^{m_i,m_j}\mathcal{O}_{kl}^{m_k,m_l}$ on $\rho^{\vec{m}}$ which are complicated functions of of the random variable m_i and traces of powers of the ρ_i reported in the next subsection. Invoking hence (38) to decouple the averages over the m_i we can finally write

$$\mathrm{Var}[\mathcal{D}] = V_1 + V_2 \;, \tag{83}$$

where setting $\mathrm{Var}_\rho[O] := \mathrm{Tr}\big[(O - \mathrm{Tr}[O\rho])^2\rho\big]$, we defined

$$V_1 = \underset{\substack{m_l \sim \mathrm{Poi}(p_l\mu) \\ l=1,\dots,N}}{\mathbb{E}} \mathrm{Var}_{\rho^{\vec{m}}}\left[D^{\vec{m},M}\right], \tag{84}$$

$$V_2 = \underset{\substack{m_l \sim \mathrm{Poi}(p_l\mu) \\ l=1,\dots,N}}{\mathbb{E}} \left(\mathrm{Tr}\left[\mathcal{D}^{\vec{m},M}\rho^{\vec{m}}\right] - \sum_{i,j} p_i p_j D_{HS}^2(\rho_i,\rho_j)\right)^2, \tag{85}$$

(we remind that the expression $m_l \sim \mathrm{Poi}(p_l\mu)$ indicates that the random variables m_l are extracted from a Poisson distribution of mean $p_l\mu$).

B.1 Bound on V_1

The covariance of two observables O, O' on a state ρ is defined as

$$\mathrm{Cov}_\rho[O,O'] := \mathrm{Tr}\big[(O - \mathrm{Tr}[O\rho])(O' - \mathrm{Tr}[O'\rho])\rho\big]. \tag{86}$$

The covariances of the observables $\mathcal{O}_{ij}^{m_i,m_j}$ on $\rho^{\vec{m}}$, read:

$$\mathrm{Var}_{\rho^{\vec{m}}}[\mathcal{O}_{ii}^{m_i,m_i}] = \frac{2}{m_i(m_i-1)}(1 - (\mathrm{Tr}\left[\rho_i^2\right])^2) + \frac{4(m_i-2)}{m_i(m_i-1)}(\mathrm{Tr}\left[\rho_i^3\right] - (\mathrm{Tr}\left[\rho_i^2\right])^2) \tag{87}$$

$$\mathrm{Var}_{\rho^{\vec{m}}}[\mathcal{O}_{ij}^{m_i,m_j}] = \frac{1}{m_i m_j} + \frac{1 - m_i - m_j}{m_i m_j}\mathrm{Tr}[\rho_i\rho_j]^2$$
$$+ \frac{1}{m_j}\left(1 - \frac{1}{m_i}\right)\mathrm{Tr}\left[\rho_i^2\rho_j\right] + \frac{1}{m_i}\left(1 - \frac{1}{m_j}\right)\mathrm{Tr}\left[\rho_i\rho_j^2\right] \qquad i \neq j \tag{88}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{ii}^{m_i,m_i},\mathcal{O}_{ij}^{m_i,m_j}] = \frac{2}{m_i}\left(\mathrm{Tr}\left[\rho_i^2\rho_j\right] - \mathrm{Tr}\left[\rho_i^2\right]\mathrm{Tr}[\rho_i\rho_j]\right) \qquad i \neq j \tag{89}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{ij}^{m_i,m_j},\mathcal{O}_{ik}^{m_i,m_k}] = \frac{\mathrm{Tr}[\rho_i\rho_j\rho_k] - \mathrm{Tr}[\rho_i\rho_j]\mathrm{Tr}[\rho_i\rho_k]}{m_i} \qquad i \neq j \wedge i \neq k \wedge j \neq k \tag{90}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{ij}^{m_i,m_j},\mathcal{O}_{kl}^{m_k,m_l}] = 0 \qquad i,j,k,l \text{ all different,} \tag{91}$$

Replacing the above expressions into (84), we can rewrite it as

$$V_1 = \underset{\substack{m_l \sim \mathrm{Poi}(p_l\mu) \\ l=1,\dots,N}}{\mathbb{E}} \mathrm{Var}_{\rho^{\vec{m}}}\left[\sum_{i \neq j}\left(\frac{m_i(m_i-1)}{\mu^2 p_i}p_j\mathcal{O}_{ii}^{m_i,m_i} + \frac{m_j(m_j-1)}{\mu^2 p_j}p_i\mathcal{O}_{jj}^{m_j,m_j} - 2\frac{m_i m_j}{\mu^2}\mathcal{O}_{ij}^{m_i,m_j}\right)\right]$$

$$= \sum_i 4\underset{m_i \sim \mathrm{Poi}(p_i\mu)}{\mathbb{E}}\frac{m_i^2(m_i-1)^2}{\mu^4 p_i^2}(1-p_i)^2\mathrm{Var}[\mathcal{O}_{ii}^{m_i,m_i}] + 8\sum_{i \neq j}\underset{\substack{m_i \sim \mathrm{Poi}(p_i\mu) \\ m_j \sim \mathrm{Poi}(p_j\mu)}}{\mathbb{E}}\frac{m_i^2 m_j^2}{\mu^4}\mathrm{Var}[\mathcal{O}_{ij}^{m_i,m_j}]$$

$$- 16\sum_{i \neq j}\underset{\substack{m_i \sim \mathrm{Poi}(p_i\mu) \\ m_j \sim \mathrm{Poi}(p_j\mu)}}{\mathbb{E}}\frac{m_i^2(m_i-1)m_j}{\mu^4 p_i}(1-p_i)\mathrm{Cov}[\mathcal{O}_{ii}^{m_i,m_i},\mathcal{O}_{ij}^{m_i,m_j}]$$

$$+ 16\sum_{i \neq j \neq k \neq i}\underset{\substack{m_i \sim \mathrm{Poi}(p_i\mu) \\ m_j \sim \mathrm{Poi}(p_j\mu) \\ m_k \sim \mathrm{Poi}(p_k\mu)}}{\mathbb{E}}\frac{m_i^2 m_j m_k}{\mu^4}\mathrm{Cov}[\mathcal{O}_{ij}^{m_i,m_j},\mathcal{O}_{ik}^{m_i,m_k}] \tag{92}$$

Now we proceed to evaluate separately each term of (92).

From (87) we get

$$
\underset{m_i \sim \text{Poi}(p_i\mu)}{\mathbb{E}} \left[\frac{m_i^2(m_i-1)^2}{\mu^4 p_i^2}(1-p_i)^2 \text{Var}[\mathcal{O}_{ii}^{m_i,m_i}] \right]
$$

$$
= \underset{m_i \sim \text{Poi}(p_i\mu)}{\mathbb{E}} \left[\frac{m_i(m_i-1)}{\mu^4 p_i^2}(1-p_i)^2 [2(1-(\text{Tr}\left[\rho_i^2\right])^2) + 4(m_i-2)(\text{Tr}\left[\rho_i^3\right] - (\text{Tr}\left[\rho_i^2\right])^2] \right]
$$

$$
= \frac{\mu^2 p_i^2}{\mu^4 p_i^2}(1-p_i)^2 [2(1-(\text{Tr}\left[\rho_i^2\right])^2) + 4\mu p_i(\text{Tr}\left[\rho_i^3\right] - (\text{Tr}\left[\rho_i^2\right])^2]
$$

$$
\leq \frac{4p_i(1-p_i)^2}{\mu}(\text{Tr}\left[\rho_i^3\right] - (\text{Tr}\left[\rho_i^2\right])^2) + O(1/\mu^2) \tag{93}
$$

where in the third line we used the fact that $\mathbb{E}[m_i(m_i-1)] = \mu_i^2 p_i^2$ and $\mathbb{E}[m_i(m_i-1)(m_i-2)] = \mu_i^3 p_i^3$ for a Poisson distribution with mean $\mu_i p_i$.

Analougously, from (88) we have

$$
\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}} \left[\frac{m_i^2 m_j^2}{\mu^4} \text{Var}[\mathcal{O}_{ij}^{m_i,m_j}] \right]
$$

$$
= \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}} \left[\frac{m_i m_j}{\mu^4}(1 + (1-m_i-m_j)\text{Tr}[\rho_i\rho_j]^2 + (m_i-1)\text{Tr}\left[\rho_i^2\rho_j\right] + (m_j-1)\text{Tr}\left[\rho_i\rho_j^2\right]) \right]
$$

$$
\leq \frac{p_i p_j^2 \text{Tr}\left[\rho_i\rho_j^2\right] + p_j p_i^2 \text{Tr}\left[\rho_j\rho_i^2\right] - p_i p_j(p_i+p_j)\text{Tr}[\rho_i\rho_j]^2}{\mu} + \frac{2p_i p_j}{\mu^2} \tag{94}
$$

The corresponding contribution from (89) is

$$
\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}} \left[\frac{m_i^2(m_i-1)m_j}{\mu^4 p_i}(1-p_i)\text{Cov}[\mathcal{O}_{ii}^{m_i,m_i}, \mathcal{O}_{ij}^{m_i,m_j}] \right]
$$

$$
= \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}} \left[\frac{m_i(m_i-1)m_j}{\mu^4 p_i}(1-p_i)2\left(\text{Tr}\left[\rho_i^2\rho_j\right] - \text{Tr}\left[\rho_i^2\right]\text{Tr}[\rho_i\rho_j]\right) \right]
$$

$$
= \frac{(1-p_i)p_i p_j}{\mu}2\left(\text{Tr}\left[\rho_i^2\rho_j\right] - \text{Tr}\left[\rho_i^2\right]\text{Tr}[\rho_i\rho_j]\right) \tag{95}
$$

Finally, from (90) we have

$$
\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu) \\ m_k \sim \text{Poi}(p_k M)}}{\mathbb{E}} \left[\frac{m_i^2 m_j m_k}{\mu^4}\text{Cov}[\mathcal{O}_{ij}^{m_i,m_j}, \mathcal{O}_{ik}^{m_i,m_k}] \right]
$$

$$
= \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu) \\ m_k \sim \text{Poi}(p_k M)}}{\mathbb{E}} \left[\frac{m_i m_j m_k}{\mu^4}\left(\text{Tr}[\rho_i\rho_j\rho_k] - \text{Tr}[\rho_i\rho_j]\text{Tr}[\rho_i\rho_k]\right) \right]
$$

$$
= \frac{p_i p_j p_k}{\mu}\left(\text{Tr}[\rho_i\rho_j\rho_k] - \text{Tr}[\rho_i\rho_j]\text{Tr}[\rho_i\rho_k]\right) \tag{96}
$$

Inserting (93), (94) and (96) into (92) we can finally write

$$V_1 = 16 \sum_i \frac{p_i(1-p_i)^2}{\mu} (\text{Tr}\left[\rho_i^3\right] - (\text{Tr}\left[\rho_i^2\right])^2)$$

$$+ 8 \sum_{i \neq j} \frac{p_i p_j^2 \, \text{Tr}\left[\rho_i \rho_j^2\right] + p_j p_i^2 \, \text{Tr}\left[\rho_j \rho_i^2\right] - p_i p_j (p_i + p_j) \, \text{Tr}[\rho_i \rho_j]^2}{\mu}$$

$$- 32 \sum_{i \neq j} \frac{(1-p_i)p_i p_j}{\mu} \left(\text{Tr}\left[\rho_i^2 \rho_j\right] - \text{Tr}\left[\rho_i^2\right] \text{Tr}[\rho_i \rho_j] \right)$$

$$+ 16 \sum_{i \neq j \neq k \neq i} \frac{p_i p_j p_k}{\mu} (\text{Tr}[\rho_i \rho_j \rho_k] - \text{Tr}[\rho_i \rho_j] \, \text{Tr}[\rho_i \rho_k]) + O(N/\mu^2) \quad (97)$$

B.2 Bound on V_2

We start defining the quantities

$$o_{ii} = \left(\frac{m_i(m_i - 1)}{\mu^2 p_i} - p_i \right) \text{Tr}\left[\rho_i^2\right], \qquad o_{ij} = \left(\frac{m_i m_j}{\mu^2} - p_i p_j \right) \text{Tr}[\rho_i \rho_j], \qquad i \neq j. \quad (98)$$

Noting that

$$\text{Tr}\left[\mathcal{D}^{\vec{m},\mu} \rho^{\vec{m}}\right] - \sum_{i,j} p_i p_j D_{HS}^2(\rho_i, \rho_j) = \sum_{i \neq j} (p_j o_{ii} + p_i o_{jj} - 2o_{ij}) \ , \quad (99)$$

we can rewrite (85) as

$$V_2 = \sum_i 4(1-p_i)^2 \underset{m_i \sim \text{Poi}(p_i\mu)}{\mathbb{E}}[o_{ii}^2] + 8 \sum_{i \neq j} \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}}[o_{ij}^2]$$

$$+ 16 \sum_{i \neq j \neq k \neq i} \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu) \\ m_k \sim \text{Poi}(p_k\mu)}}{\mathbb{E}}[o_{ij} o_{ik}] - 16 \sum_{i \neq j} (1 - p_i) \underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}}[o_{ii} o_{ij}] \quad (100)$$

The expected values which appear in (100) can be easily computed:

$$\underset{m_i \sim \text{Poi}(p_i\mu)}{\mathbb{E}}[o_{ii}^2] = \frac{2(1 + 2\mu p_i)}{\mu^2} \text{Tr}\left[\rho_i^2\right]^2 \quad (101)$$

$$\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}}[o_{ij}^2] = \frac{(\mu p_i p_j (p_i + p_j) + p_i p_j)}{\mu^2} \text{Tr}[\rho_i \rho_j]^2, \qquad i \neq j \quad (102)$$

$$\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu) \\ m_k \sim \text{Poi}(p_k\mu)}}{\mathbb{E}}[o_{ij} o_{ik}] = \frac{p_i p_j p_k}{\mu} \text{Tr}[\rho_i \rho_j] \, \text{Tr}[\rho_i \rho_k], \qquad i \neq j \neq k \neq i \quad (103)$$

$$\underset{\substack{m_i \sim \text{Poi}(p_i\mu) \\ m_j \sim \text{Poi}(p_j\mu)}}{\mathbb{E}}[o_{ii} o_{ij}] = \frac{2p_i p_j}{\mu} \text{Tr}[\rho_i \rho_j] \, \text{Tr}\left[\rho_i^2\right] \qquad i \neq j. \quad (104)$$

Replacing (101), (102), (103) and (104) into (100), and then isolating the leading order, we can conclude that

28

$$V_2 = \sum_i \frac{8(1 + 2\mu p_i)}{\mu^2}(1 - p_i)^2 \operatorname{Tr}\left[\rho_i^2\right]^2 + \sum_{i \neq j} \frac{8(\mu p_i p_j(p_i + p_j) + p_i p_j)}{\mu^2} \operatorname{Tr}[\rho_i \rho_j]^2$$

$$+ \sum_{i \neq j \neq k \neq i} \frac{16(p_i p_j p_k)}{\mu} \operatorname{Tr}[\rho_i \rho_j] \operatorname{Tr}[\rho_i \rho_k] - \sum_{i \neq j} \frac{32 p_i p_j}{\mu}(1 - p_i) \operatorname{Tr}[\rho_i \rho_j] \operatorname{Tr}\left[\rho_i^2\right]$$

$$\leq \sum_i \frac{16}{\mu}(1 - p_i)^2 p_i \operatorname{Tr}\left[\rho_i^2\right]^2 + \sum_{i \neq j} \frac{8 p_i p_j(p_i + p_j)}{\mu} \operatorname{Tr}[\rho_i \rho_j]^2$$

$$+ \sum_{i \neq j \neq k \neq i} \frac{16(p_i p_j p_k)}{\mu} \operatorname{Tr}[\rho_i \rho_j] \operatorname{Tr}[\rho_i \rho_k] - \sum_{i \neq j} \frac{32 p_i p_j}{\mu}(1 - p_i) \operatorname{Tr}[\rho_i \rho_j] \operatorname{Tr}\left[\rho_i^2\right] + O(N/\mu^2) \,. \tag{105}$$

B.3   Bound on $V_1 + V_2$

We start by observing that

$$0 \leq \operatorname{Tr}\left[(\rho_i \sqrt{\rho_j} - \rho_k \sqrt{\rho_j})^\dagger(\rho_i \sqrt{\rho_j} - \rho_k \sqrt{\rho_j})\right]$$
$$\implies \operatorname{Tr}[\rho_i \rho_j \rho_k] + \operatorname{Tr}[\rho_i \rho_k \rho_j] \leq \operatorname{Tr}[\rho_i^2 \rho_j] + \operatorname{Tr}[\rho_k^2 \rho_j] \,. \tag{106}$$

Applying (106) to the sum and summing

$$\sum_{i \neq j \neq k \neq i} \frac{p_i p_j p_k}{\mu} \operatorname{Tr}[\rho_i \rho_j \rho_k] \leq \sum_{i \neq j} \frac{p_i p_j(1 - p_i - p_j) \operatorname{Tr}\left[\rho_i \rho_j^2\right]}{\mu} \tag{107}$$

Combining (97), (105) and using (107) we have

$$V_1 + V_2 = O\left(\frac{N}{\mu^2}\right) + 16 \sum_i \frac{p_i(1 - p_i)^2}{\mu} \operatorname{Tr}\left[\rho_i^3\right] + 8 \sum_{i \neq j} \frac{p_i p_j^2 \operatorname{Tr}\left[\rho_i \rho_j^2\right] + p_j p_i^2 \operatorname{Tr}\left[\rho_j \rho_i^2\right]}{\mu}$$

$$- 32 \sum_{i \neq j} \frac{(1 - p_i)p_i p_j}{\mu}\left(\operatorname{Tr}\left[\rho_i^2 \rho_j\right]\right) + 16 \sum_{i \neq j \neq k \neq i} \frac{p_i p_j p_k}{\mu}\left(\operatorname{Tr}[\rho_i \rho_j \rho_k]\right) + O(N/\mu^2)$$

$$\leq O\left(\frac{N}{\mu^2}\right) + 16\left(\sum_i \frac{p_i(1 - p_i)^2}{\mu} \operatorname{Tr}\left[\rho_i^3\right] + \sum_{i \neq j} \frac{p_i p_j[(p_j + 1 - p_i - p_j) \operatorname{Tr}\left[\rho_i \rho_j^2\right] - 2(1 - p_i) \operatorname{Tr}\left[\rho_i^2 \rho_j\right]]}{\mu}\right)$$

$$= O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu} \sum_{i \neq j} p_i p_j \operatorname{Tr}\left[((1 - p_i)\rho_i)(\rho_i - \rho_j)^2\right] \leq \sum_{i \neq j} p_i p_j \operatorname{Tr}\left[\|(1 - p_i)\rho_i\|_\infty(\rho_i - \rho_j)^2\right]$$

$$\leq O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu} \sum_{i \neq j} p_i p_j \operatorname{Tr}\left[(\rho_i - \rho_j)^2\right]$$

$$= O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu} \sum_{i \neq j} p_i p_j D_{HS}^2(\rho_i, \rho_j) = O\left(\frac{N}{\mu^2}\right) + \frac{16 \mathcal{M}_{HS}^2}{\mu}. \tag{108}$$