An infinite hierarchy of multi-copy quantum learning tasks

Jan Nöller[1], Viet T. Tran[2], Mariami Gachechiladze[1], and Richard Kueng[2]

[1]Quantum Computing Group, Department of Computer Science, Technical University of Darmstadt, Germany
[2]Department for Quantum Information and Computation at Kepler (QUICK), Johannes Kepler University Linz, Austria

Learning properties of quantum states from measurement data is a fundamental challenge in quantum information. The sample complexity of such tasks depends crucially on the measurement primitive. While shadow tomography achieves sample-efficient learning by allowing entangling measurements across many copies, it requires prohibitively deep circuits. At the other extreme, two-copy measurements already yield exponential advantages over single-copy strategies in tasks such as Pauli tomography. In this work we show that such sharp separations extend far beyond the two-copy regime: for every prime c we construct explicit learning tasks of degree c, which are exponentially hard with $(c-1)$-copy measurements but efficiently solvable with c-copy measurements. Our protocols are not only sample-efficient but also realizable with shallow circuits. Extending further, we show that such finite-degree tasks exist for all square-free integers c, pointing toward a general principle underlying their existence. Together, our results reveal an infinite hierarchy of multi-copy learning problems, uncovering new phase transitions in sample complexity and underscoring the role of reliable quantum memory as a key resource for exponential quantum advantage.

Introduction

Quantum state learning is the fundamental task of inferring useful information about an unknown quantum system from experimental data. The data are obtained through quantum measurements, which inevitably disturb the system. To collect sufficient statistics, the same quantum state must therefore be re-prepared many times, with each copy measured individually. The total number of preparations and measurements needed to estimate the desired properties with high confidence is known as the *sample complexity*. This parameter is central in determining the feasibility of quantum learning protocols.

The most general approach to state learning is *quantum state tomography* [1], which reconstructs the full density matrix of the system. For an n-partite quantum system, the density matrix is a Hermitian $d^n \times d^n$ matrix in a local dimension d, and tomography corresponds to accessing all of its informationally complete degrees of freedom. While this procedure yields complete knowledge of the state, it comes with a prohibitive cost: even the best-known tomographic techniques [2–5] require a number of samples that grows exponentially with n. This exponential overhead is unavoidable, as shown by fundamental lower bounds [3, 6].

In practice, however, one is rarely interested in reconstructing the entire density matrix. More often, the goal is to predict a family of properties of the state. This motivates the framework of *shadow tomography* [7], where the task is to approximate the expectation values of many observables $\mathrm{tr}(O_l\rho)$ up to accuracy ϵ, without fully reconstructing ρ. For L observables on an n-qubit system, refined shadow tomography protocols achieve this with a sample complexity of $\widetilde{O}(n\log^2(L)/\epsilon^4)$ copies [8], remarkably efficient and independent of both the state and the observables. The trade-off, however, is that the required quantum circuits are extremely demanding: they involve exponentially deep operations acting on all of these state copies simultaneously.

Jan Nöller: jan.noeller@tu-darmstadt.de

A natural way to reduce these overheads is to restrict either the class of states or the class of observables. *Classical shadows* [9–12] are a prominent example of the latter. A particularly important subclass of observables is given by the n-qubit Pauli operators, i.e., tensor products of single-qubit Pauli matrices. These play a central role in quantum information, quantum algorithms and qubit encodings of quantum chemistry problems, see e.g. [13]. Learning Pauli observables from data is commonly referred to as *Pauli shadow tomography*, and because it relies on sequential single-copy measurements, it requires exponentially many samples in the worst case [14, 15].

With the continued development of quantum hardware, these no-go results can be circumvented by changing the measurement paradigm: instead of measuring single copies in a sequential fashion, one may store multiple copies in a quantum memory and process them jointly using entangling measurements. Refs. [14, 15] have shown that this drastically changes the picture: With access to only two copies of a quantum state at a time and the ability to perform entangling measurements on both copies, it becomes possible to jointly estimate the absolute values of all 4^n squared Pauli expectation values, $|\mathrm{Tr}(P\rho)|^2$, in a sample-efficient manner. What is more, the quantum circuits involved only have constant depth. When combined with lower bounds showing that the same task is exponentially costly in the single-copy regime, this establishes a rigorous exponential sample advantage [14] that has been demonstrated on a 53-qubit quantum device [15].

Beyond Pauli tomography, similar two-copy advantages have been demonstrated for principal component analysis [16] and for quantum channel learning [17]. Follow-up works have analyzed this phenomenon in greater depth, including estimation of Pauli channels [18–20], Pauli observables [21], and purity estimation [22] (the latter being the original SWAP-test task introduced in Ref. [23]). More recently, refined techniques using martingale concentration inequalities have enabled a unified treatment of lower bounds for such tasks [21], which we also use as a central tool in this work.

All known exponential advantages fall into two broad categories. Some are fully realized with two-copy measurements [15, 18–20, 24], while others follow the paradigm of shadow tomography [7, 25], requiring joint measurements on a number of copies that grows with system size. In particular, Ref. [26] introduced a hierarchy of learning tasks, each of which is provably *c-copy-hard*, meaning that no algorithm limited to c or fewer copies can solve it efficiently. The best schemes however are again based on shadow tomography, requiring poly(nc)-copy measurements and correspondingly complex circuits. This leaves a large gap between the regimes where tasks remain provably hard and those where they become efficiently solvable.

The situation in Refs. [14, 15] is different, as it exhibits a sharp transition from exponential to linear sample complexity already at two-copy measurements. Motivated by this, we set out to investigate whether further examples of such sharp transitions exist. For clarity, we refer to the number of copies at which such a transition occurs as the *degree* of a learning task. A task is said to be of degree c if it can be solved efficiently with c-copy measurements, but remains provably hard for strategies with $(c-1)$ (or fewer) copies, even if one allows for adaptive protocols utilizing deep quantum circuits and/or arbitrarily expensive classical co-processing. This unifying point of view puts shadow tomography one one extreme of a spectrum ($c \gg 1$) and degree-2 tasks, like purity testing and Pauli shadow tomography, on the other extreme. Naturally, this posits the question whether there exist tasks of degree c for some $c > 2$, and whether one can identify infinite families of such tasks for increasing c.

Our results provide a positive answer to both questions. Specifically, for every prime c we construct explicit learning tasks of degree c, thereby establishing an infinite hierarchy that interpolates between the previously known degree-2 examples and shadow tomography, which can be regarded as a task of degree ∞ (in the limit of very large system sizes). On the technical side, we adapt the results developed in Ref. [21] to prove that no $(c-1)$-copy protocol (whether exponentially deep or adaptive) can solve the same ϵ-approximate learning problem without incurring an exponential sample complexity in the number n of identical quantum systems involved. Moreover, at the sharp transition, we show not only that the sample complexity becomes efficient, but also that the measurement protocols can be implemented with shallow circuits of (at most) logarithmic depth in c (and independent of n). This highlights the power of multi-copy quantum information processing. We explicitly construct the protocol and give examples of how it may be transpiled on a qubit-based quantum architecture. We further extend these constructions to all *square-free* integers c, i.e., those in which each prime factor appears with multiplicity at most one. This extension illustrates that finite-degree learning tasks are not confined to the special

role of prime dimensions, but instead arise in a much broader setting, pointing toward a general principle underlying their existence. In fact, since the natural density of square-free numbers is $\lim_{N\to\infty}|\{m \leq N : m \text{ square-free}\}|/N = 6/\pi^2 \approx 0.6$, as shown in [27], we expect that also *linear* degree-c learning tasks exist for every integer c.

Taken together, these results uncover a rich and previously unexplored hierarchy of quantum learning problems. Complementing the results of [28], they demonstrate that sharp phase transitions in sample complexity are not confined to the one-copy vs. two-copy setting, but occur broadly across higher-copy regimes. This perspective not only advances our theoretical understanding of the complexity of quantum learning, but also suggests new ways of exploiting multi-copy quantum resources to achieve exponential advantages and points at the importance of building reliable quantum memory.

Note added During the final preparation stages of this work, we became aware of a concurrent work, Ref. [28], which addresses the same overarching question from a different vantage point. On a high level, two avenues exist to establish complexity separations via degree-two learning tasks. The first, is the linearity of quantum mechanics, leading to advantages of 2-copy schemes, if one were to investigate quantities, that involve, or can be determined using the second moments of the distributed state ρ, such as purity estimation or principal component analysis. The second avenue builds on Heisenberg's uncertainty principle and related anticommutation relations that can change drastically if one moves to tensor products [14]. While [28] sets out to expand upon the former approach to establish degree-c learning tasks for any integer c, we follow the latter to also answer the posed question affirmatively. Our main results however are confined to *prime* degrees, although we subsequently demonstrate a derived construction for *square-free* integers c. Therefore, complementing the results of [28], we manage to provide an infinite hierarchy of linear learning tasks, that interpolates between the previously known degree-2 examples and shadow tomography, which can be regarded as a task of degree ∞.

Outline The remainder of this manuscript is organized as follows: We begin with a few definitions to properly summarize our results afterwards. We continue by setting up some preliminaries and defining the learning tasks we consider in Section 2. Following this, we first prove our hardness results in Section 3. Section 4 then continues to discuss the d-copy efficient scheme solving these tasks. In Section 5 we summarize additional results related to the learning task discussed in the previous sections. More technical details and proofs supporting the claims in the main text, as well as details to the numerical simulations related to Fig. 1 can be found in the Appendix.

1 Summary of results

In this section, we briefly introduce the notion of efficiently solving a learning task before we present the main results.

Definition 1. *Let $\mathcal{A} = \{\mathcal{A}_n\}_{n\in\mathbb{N}}$ be a learning task for an n-partite quantum state ρ. We say that there exists an efficient protocol ε-approximately solving this task for $\varepsilon > 0$ if it satisfies:*

- Sample efficiency: *the protocol uses at most $O(\text{poly}(n, 1/\varepsilon))$ copies of ρ;*
- Quantum efficiency: *all quantum subroutines can be implemented by circuits of size $O(\text{poly}(n))$;*
- Classical efficiency: *all classical postprocessing runs in $\text{poly}(n)$ time using $\text{poly}(n)$ memory.*

A protocol that has access to measurement data on at most c copies of ρ per update step and meets all the criteria above is called c-copy efficient.

On the other hand, to capture the minimal complexity of learning tasks, we use the following terminology (cf. [26]).

Definition 2. *A learning task $\mathcal{A}$ is called* c-copy hard *if there exists no sample-efficient protocol that solves $\mathcal{A}$ using adaptive measurements on at most c copies of the state ρ at a time.*

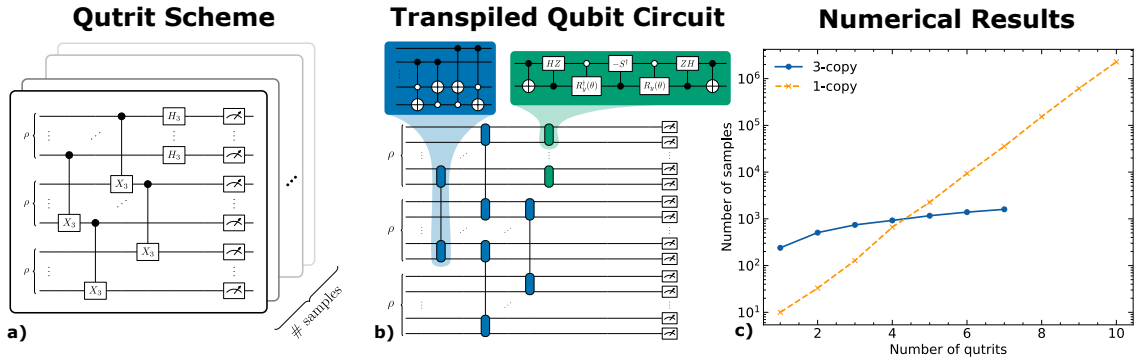


Figure 1: *Learning challenge with degree $c = 3$, i.e. an efficient 3-copy strategy exists, but it's hard for any 1- and 2-copy strategy.* The underlying learning task is formulated for n-qutrit systems ($d = 3$) and asks for estimating *all* generalized Pauli observables, also known as Weyl-Heisenberg matrices, in modulus. **a)** a qutrit-based 3-copy learning protocol that allows for jointly predicting all third powers of generalized Pauli observables (think: a qutrit extension of the destructive SWAP test for qubit systems). **b)** a transpilation of this qutrit-based circuit into qubits: $|0\rangle_3 \to |00\rangle$, $|1\rangle_3 \to |01\rangle$, $|2\rangle_3 \to |10\rangle$ (the same can be also done for the underlying learning challenge). **c)** Numerical simulations that compare the sample complexity of one particular single-copy strategy (orange) – local classical shadows for qutrit systems [30, 31] – and the advertised 3-copy strategy (blue) required to complete the learning task with (at least) 70% success probability. The overall plot shape resembles Figure 2 in Ref. [15] which addressed a degree-2 variant of this learning challenge on the Google sycamore chip.

Here, *sample-efficient* refers only to the number of state copies used; we place no restrictions on the quantum circuit depth or the amount of classical postprocessing. Finally, combining the two notions above gives rise to the concept of the *degree* of a learning task.

Definition 3. *A learning task $\mathcal{A}$ is said to have* degree *c if it is $(c-1)$-copy hard, but admits a c-copy efficient protocol.*

We now summarize our main result, which establishes the existence of a hierarchy of learning tasks of increasing degree. The specific tasks will be introduced in the following sections, together with the proofs of the corresponding complexity bounds.

Result 4. *For each prime number p, there exists a learning task $\mathcal{A}^{(p)} = \{\mathcal{A}_n^{(p)}\}_{n\in\mathbb{N}}$ of degree p. More precisely:*

(i) There is a p-copy efficient protocol for $\mathcal{A}^{(p)}$ that uses $\mathcal{O}(np\log(p)\ \mathrm{poly}(\varepsilon^{-2p}))$ samples and quantum circuits of size $\mathcal{O}(n)$ and constant depth.

(ii) Any protocol restricted to measurements on at most $c < p$ copies requires at least $\mathcal{O}((p-1)^n/(c\varepsilon^2))$ samples of ρ.

As a short technical detour, and similarly to Refs. [15, 21], the learning tasks we study correspond to predicting the *absolute values* of observables $A_1, \ldots, A_N$ that form a basis of the underlying Hilbert space. As shown in part (i) of Theorem 4, this can be achieved c-copy efficiently for suitable c. Moreover, if one wishes to recover the full state, including phase information, our protocol can be combined with the scheme of Ref. [29], as we explain in Section 5.2, though at the cost of substantially higher classical and quantum computational resources. In particular, while the tasks we focus on are not full tomography, our constructions can be extended to Weyl-Heisenberg tomography under p-copy measurements, at the expense of forfeiting computational efficiency while retaining sample efficiency.

A key feature of Theorem 4 is the sharp phase transition at $c = p$ copies: the exponential lower bound remains unchanged whether entangling measurements are allowed on two, three, or even $(p-1)$ copies, but collapses to polynomial complexity once p-copy measurements are permitted. This phenomenon can be traced back to counting solutions of quadratic equations of the form $ck^2 + ak + b = 0$ over the finite field $\mathbb{Z}_p$, where the quadratic term vanishes when $c = p$. Finally, we extend our constructions from primes to all *square-free* integers, showing that finite-degree learning tasks are not confined to the special role of prime dimensions but arise in a much broader setting.

Taken together, these results establish the first infinite hierarchy of multi-copy learning tasks, reveal sharp phase transitions in their sample complexity, and highlight the power of multi-copy quantum information processing as a fundamental resource for quantum learning.

2 Preliminaries and Notation

We now turn to the technical development. To set the stage, we first introduce the notation that will be used throughout the paper. With this in place, we revisit the central mathematical object underlying our constructions: the *Weyl–Heisenberg group*, which serves as the natural generalization of the Pauli group to higher dimensions and provides the structural backbone for the learning tasks we study.

Notation

For an operator A and a multiindex $I = (i_1, \ldots, i_m) \in \mathbb{N}^m$ we write $A^{\otimes I} := A^{i_1} \otimes A^{i_2} \otimes \cdots \otimes A^{i_m}$. If A is invertible this notation extends to negative-valued entries of I. Given a basis $|\psi_1\rangle, \ldots, |\psi\rangle_n$ of some Hilbert space, for $J \in \{1, \ldots, n\}^m$ we write $|\psi_J\rangle := |\psi_{j_1}\rangle \otimes |\psi_{j_2}\rangle \otimes \cdots \otimes |\psi_{j_m}\rangle$. For a multiindex $I = (i_1, \ldots, i_m)$ we abbreviate $|I| = \sum_{k=1}^m i_k$ and for the multiindex, which consists of copies of 1 only, we write $\mathbf{1} = (1, \ldots, 1)$. Throughout, d is always refereed to an odd prime, unless indicated differently. All arithmetic involving indices which indicate computational basis elements, are then to be understood modulo d, we will not always make this explicit. For a natural number n we write $[n] = \{0, 1, \ldots, n-1\}$. In order to avoid confusion with the well known Lie bracket $[\cdot, \cdot]$, in this work we will denote the group commutator via $[\![\cdot, \cdot]\!]$. Recall, that it is defined for invertible matrices A, B via $[\![A, B]\!] := ABA^{-1}B^{-1}$.

The Weyl-Heisenberg operators

This section introduces the Weyl-Heisenberg group as the natural generalization of the Pauli group to higher dimensions, and briefly reviews its relation to mutually unbiased bases (MUB). Even if the reader is familiar with these, we encourage not to skip this section, as some of our conventions will slightly differ from the standard ones. The Weyl-Heisenberg operators in dimension d are generated analogously to the qubit Pauli group by a shift operator $X := \sum_{k\in[d]} |k \oplus_d 1\rangle\langle k|$ and a boost operator $Z = \sum_{k\in[d]} = \omega^k |k\rangle\langle k|$, where $\omega = e^{2\pi \mathrm{i}/d}$. Here, and in the following, $\oplus_d$ denotes the addition modulo d. These elementary Weyl-Heisenberg operators obey the commutation relation $XZ = \omega^{-1} ZX$, as well as $X^d = Z^d = \mathrm{id}$. One can then derive that up to a suitable factor, all unitary operators generated by X and Z, have a projective representative of the form $W(p, q) = X^p Z^q$. It is not hard to verify that all these operators satisfy the following generalized anticommutation relations

$$W(p, q)W(\tilde{p}, \tilde{q}) = \omega^{p\tilde{q} - q\tilde{p}} W(\tilde{p}, \tilde{q})W(p, q). \tag{1}$$

Changing phases in front of operators neither adds complexity in determining expectation values, nor does it affect commutation relations and hence incompatibility of observables. For our purposes, we therefore take the freedom to choose the Weyl-Heisenberg operators in the following way: First, we define

$$W_a := XZ^a \text{ for } a \in [d] \quad \text{and} \quad W_\infty := Z, \tag{2}$$

and let $\mathcal{W}_a := \{W_a^k : k \in [d]\}$ be the cyclic group generated by W_a. We then obtain a full set of projective representatives by the union $\mathcal{W}^{(d)} := \bigcup_{a\in[d]} \mathcal{W}_a \cup \mathcal{W}_\infty$. We shall omit the superscript, when there is no danger of confusion about the underlying dimension. This particular way of grouping the Weyl-Heisenberg operators into distinct cyclic subgroups, which intersect only at the identity, will be convenient for subsequent calculations. In particular, to each $\mathcal{W}_a$ we can associate the orthonormal basis $\mathcal{B}_a = \{|\psi_0^a\rangle, \ldots, |\psi_{d-1}^a\rangle\}$ of eigenstates, where

$$|\psi_j^a\rangle = \sum_{k\in[d]} \omega^{-jk + ak(k-1)/2} |k\rangle, \text{ for a } \neq \infty, \tag{3}$$

and $\mathcal{B}_\infty = \{|k\rangle\}_{k\in[d]}$ is the computational basis. This way we can express each $W_a^k \in \mathcal{W}_a$ as

$$W_a^k = \sum_{j\in[d]} \omega^{kj} |\psi_j^a\rangle\langle\psi_j^a|. \tag{4}$$

In case the dimension d is prime, we observe essential additional features to obtain the hardness results. First of all, the $\mathcal{W}_a$ are the maximal (abelian) subgroups of $\mathcal{W}$ and two $W_1, W_2 \in \mathcal{W}$ commute if and only if they are contained in the same $\mathcal{W}_a$. Further, the eigenbases $\mathcal{B}_a$ corresponding to the cyclic subgroups $\mathcal{W}_a$ are exactly the $(d+1)$ mutually unbiased bases (MUBs) in dimension d [32]. This is also the maximal number of MUB's which can exist in a specified dimension. Since mutual unbiasedness of bases indicates that one basis measurement reveals as little information about the other measurement as possible, this hints at the fact, that we expect the task of learning different Weyl-Heisenberg observables to be particularly hard in prime dimensions beyond the special case of Pauli Tomography when $d = 2$.

Remark 5. Usually, in quantum information, observables are defined as Hermitian operators. Experimentally, the expectations of observables are measured by performing a measurement in the eigenbasis of the observable, and then weighting the measured probabilities with the corresponding eigenvalues. For Hermitian observables, the eigenvalues of course are always real-valued. However, there is no harm in also considering eigenvalue decompositions with complex-valued eigenvalues, leading to non-Hermitian observables. Expectation values can be formed as long as the operator is diagonalizable.

Following this remark, we see that in particular it is justified to treat the Weyl-Heisenberg operators as quantum mechanical observables. Complexity-theoretical results necessarily involve statements about systems that vary in size. One possibility here would be to consider Weyl-Heisenberg tomography for the different underlying qudit systems. However, the structure of the problem changes drastically in this setting, and is not suitable for our purpose. Instead, similarly to [14, 15], we consider the task of predicting the values of all such Weyl-Heisenberg *strings* on n-qudit systems for fixed d. Moreover, to establish a clean separation in the sense of Theorem 3, we consider a slightly simplified task, where we are only interested in the absolute value or *amplitude* of the expected value. Specifically, we consider the following task.

Task 1 (Weyl-Heisenberg-amplitude-estimation)**.** *For a fixed $\varepsilon > 0$, and unknown n-partite quantum state ρ with local dimension d, approximate $|\mathrm{Tr}(W\rho)|$ for all $W \in (\mathcal{W}^{(d)})^{\otimes n}$ up to additive error ε.*

In the following we first derive rigorous lower bounds on the sample complexity for protocols which use fewer than d copies per update step when d is a prime. Our analysis builds upon the work of [21]. Following up on this, we continue to outline the d-copy efficient protocol to solve Weyl-Heisenberg-amplitude estimation in section 4, (d being prime is not important here), and derive the upper bounds, as claimed in Theorem 4. To summarize, these results show that amplitude estimation for $\mathcal{W}^{(d)}$ is a degree-d learning task for d being prime.

3 Amplitude estimation for $\mathcal{W}^{(d)}$ is $(d-1)$-copy hard

In order to establish lower bounds on sample complexity, one must show that a task remains hard for any conceivable protocol, including those with adaptive strategies where each measurement can be chosen in response to the entire prior measurement history. A powerful abstraction for reasoning about such protocols is the framework of *learning trees*.

The idea is as follows. We model the learning algorithm as a machine equipped with an internal memory and the ability to perform measurements on quantum states. At time step t, given a memory state u, the machine performs a measurement M_u on (copies of) the unkonwn quantum state ρ. Depending on the outcome $a \in \{a_1, \ldots, a_{m(u)}\}$, the machine updates its memory to a new node v_a. This update is represented by a directed edge from u to v_a.

Because we are interested in lower bounds on sample complexity, we may without loss of generality assume the machine has unbounded memory. Consequently, no two directed edges need

to merge into the same node, since such merging would mean the algorithm discards information. The underlying structure is therefore a tree, not just a directed graph.

The problem of bounding the depth of such learning trees required to solve a given task with high success probability has been extensively studied [15, 21]. However, the learning tree framework is most naturally suited to *hypothesis testing*. To exploit this approach, we must therefore identify a hypothesis testing problem that reduces to Weyl-Heisenberg amplitude estimation while still yielding stringent lower bounds. Following the approach of [15, 21], a particularly effective choice is that of *many-versus-one distinguishing tasks*. These tasks are hypothesis tests in which one must decide between a single null hypothesis and a collection of alternative hypotheses, hence the name *many-versus-one*. In the quantum setting, such problems have been analyzed in great detail in Ref. [21], leading to general lower bounds on their sample complexity. For our purposes, we introduce directly the instance most relevant to our construction.

Task 2 (Many-versus-one distinguishing task)**.** *Fix* $0 < \varepsilon \leq 1/6$*. Distinguish between the following two hypotheses, each occurring with prior probability* $1/2$*:*

- *Null hypothesis: the state is maximally mixed,* $\rho = \rho_m := \mathrm{id}/d^n$*.*
- *Alternative hypothesis: the state is a fixed but random element of the ensemble*

$$\left\{ \rho_W^\varepsilon := \tfrac{1}{d^n}\left(\mathrm{id} + 3\varepsilon(W + W^\dagger)\right) \right\}_{\mathrm{id} \neq W \in \mathcal{W}^n}. \tag{5}$$

The constraint $\varepsilon \leq 1/6$ ensures that ρ_W^ε is always positive semidefinite, but can be relaxed for certain concrete dimensions, e.g., to $\varepsilon \leq 1/3$ when $d = 3$. The following theorem is a direct corollary of Theorem 1 in [21] applied to our setting. It relates the sample complexity of solving the many-versus-one-distinguishing Task 2 to the following quantity

$$\delta_{\varepsilon,c} := \min_{\mu} \max_{\mathcal{M}} \mathbb{E}_{W\sim\mu}[\chi^2_{\mathcal{M}}(\rho_W^{\otimes c} || \rho_m^{\otimes c})]. \tag{6}$$

Here, c refers to the number of state copies that can be processed simultaneously, the maximization runs over all (finite) POVM's $\mathcal{M}$, and the minimization over all probability distributions μ on $(\mathcal{W}^{(d)})^{\otimes n}$. $\chi^2_{\mathcal{M}}$ refers to the classical χ^2-divergence between the probability distributions obtained from the respective states via the measurement $\mathcal{M}$. Recall that for discrete probability distributions, the χ^2-divergence is defined as

$$\chi^2(p||q) := \sum_j \left(\frac{p_j - q_j}{q_j}\right)^2 q_j.$$

With these clarifications, we are now ready to state the Theorem.

Theorem 6 (Theorem 1 in [21])**.** *Consider the many-versus-one distinguishing task between the maximally mixed state* ρ_m*, and the ensemble of states* $\{\rho_W^\varepsilon\}$ *as in Eq. (5). Any learning protocol which can process up to* c *copies of the distributed state* ρ *simultaneously and solve said task correctly with probability* $p > 1/2$*, needs at least* $\Omega(c/\delta_{\varepsilon,c})$ *copies of* ρ*, with* $\delta_{\varepsilon,c}$ *as specified in Eq. (6)*

In order to apply these results to our problem of lower bounding the sample complexity for Weyl-Heisenberg amplitude estimation, we need to reduce the former to the latter. While the intuition should be relatively straightforward, we nevertheless give a short proof to convince ourselves of the correct choice of parameters in the following lemma.

Lemma 7 (Reduction to separation instance)**.** *Assume, that we can solve the Weyl-Heisenberg (WH) amplitude estimation with high probability for the precision parameter* $\varepsilon > 0$*. Then, with the same probability, we can also solve the separation instance of the many-versus one learning task specified in Task 2. In particular, the sample complexity for solving WH-amplitude estimation is lower bounded by* $\mathcal{O}(c/\delta_{c,\varepsilon})$ *with* $\delta_{\varepsilon,c}$ *as in Eq. (6).*

Proof. If one is given estimates u_W such that $\max_W |u_W - |\mathrm{Tr}(W\rho)|| \leq \varepsilon$, it is easy to see that we can distinguish the two cases displayed in Task 2. Case 1 ($\rho = \mathrm{id}/d^n$) implies $\mathrm{tr}(W\rho) = 0$ for

all W and therefore $|u_W - \mathrm{Tr}[\rho_m W]| = |u_W| \leq \varepsilon$. Case 2 instead yields substantially larger values for exactly two Weyl-Heisenberg operators (W and $W^\dagger$): $\mathrm{Tr}[W\rho_W] = \mathrm{Tr}[W^\dagger \rho_W] = 3\varepsilon$ and we find that $|u_W| \geq |\mathrm{Tr}[W\rho]| - \mathrm{Tr}[W\rho] \geq 2\varepsilon$. Therefore, if $\max_W u_W \leq \varepsilon$ we opt for the null hypothesis, and for the alternative hypothesis otherwise. Since we have $\mathrm{Tr}[W\rho^\varepsilon_{W'}] = 3\varepsilon(\delta_{W=W'} + \delta_{W^\dagger=W'})$ in case the alternative hypothesis is true, the hypothesis test is solved correctly in both cases. In particular, the probability of solving the hypothesis test correctly is equal to the probability of obtaining ε-correct estimates u_W for all $|\mathrm{Tr}[W\rho]|$ in the considered instance. □

This reduction reduces the task of deriving sample complexity lower bounds to the task of upper bounding the quantity $\delta_{c,\varepsilon}$ from Eq. (6). Inserting the c-fold copies of the two state families yields

$$\delta_{c,\varepsilon} = \min_{\mu} \max_{M} \mathbb{E}_{W\sim\mu} \left[\chi^2_M \left(\left(\frac{I + 3\varepsilon(W + W^\dagger)}{d^n} \right)^{\otimes c} \middle\| \left(\frac{I}{d^n} \right)^{\otimes c} \right) \right]. \tag{7}$$

To produce an upper bound, we proceed in essentially two steps. The key idea behind each step is summarized below, while rigorous proofs are provided in the Appendix. In order to bound the minimization of Eq. (6) it suffices to choose any probability distribution over $\mathcal{W}^n$ that serves our purpose. For us it will be convenient to consider the uniform distribution over all Weyl-Heisenberg strings that contain no $W \in \mathcal{W}_\infty$, i.e. the set $\hat{\mathcal{W}}^n$ with $\mathcal{W} \setminus \mathcal{W}_\infty$. This allows us to exploit some structure of the associated MUB's more conveniently while not significantly simplifying the underlying learning task.
Intuitively, choosing the uniform distribution renders the many-versus-one distinguishing task hard, because in this way one does not reveal any information about in which 'direction' to expect a deviation from the uniformly random statistics. With this choice, following very similar lines of reasoning as in [21], it is then relatively straightforward to adapt the arguments therein to obtain the following estimate:

$$\delta_{c,\varepsilon} \leq \left(\sum_{0\neq S\in[2]^c} \frac{(6\varepsilon)^{|S|}}{\sqrt{d(d-1)}^n} \sqrt{ \max_{\tau\in\{\pm1\}^{2|S|}} \left\| \sum_{W\in\hat{\mathcal{W}}} W^{\otimes\tau} \right\|_\infty^n } \right)^2. \tag{8}$$

We explicitly derive this bound step-by-step in Lemma B.1 in the Appendix. A subtle difference compared to the scenario of [21] is, that since the state ρ^ε_W also involves the inverse $W^\dagger$ of W to account for the non-hermiticity of W, we need to find bounds for all the averages $M_\tau = \sum_{W\in\hat{\mathcal{W}}} W^{\otimes\tau}$ for the various $\tau \in \{\pm1\}^{2m}$, not only for $W \mapsto W^{\otimes 2m}$. Fortunately, the norm of M_τ does not actually depend on the specific choice of τ.

Proposition 8. *Let d be a prime, $1 \leq m \leq d$ a natural number, and $\tau \in \{\pm1\}^{2m}$. Then we have*

$$\left\| \sum_{W\in\hat{\mathcal{W}}} W^{\otimes\tau} \right\|_\infty = \begin{cases} d & \text{for } m < d, \\ d(d-1) & \text{for } m = d. \end{cases} \tag{9}$$

We prove this result in Section B by deriving an explicit form of the operator M_τ. With this result we continue estimating $\delta_{c,\varepsilon}$ where we now assume $c < d$. In particular, $|S| < d$ for all $S \in [2]^c$, and by inserting the bounds derived in Theorem 8 into Eq. (8) accordingly, we find that

$$\delta_{c,\varepsilon} \leq \left(\sum_{0\neq S\in[2]^c} \frac{(6\varepsilon)^{|S|}}{\sqrt{d-1}^n} \right)^2 = \frac{((1+6\varepsilon)^c - 1)^2}{(d-1)^n} \leq \frac{(6c\varepsilon e^{6c\varepsilon})^2}{(d-1)^n}. \tag{10}$$

Altogether, we have proven the following theorem, which constitutes the second part of Theorem 4.

Theorem 9. *Let $d \geq 3$ be a prime number and $c < d$. Any c-copy protocol which solves the Weyl-Heisenberg amplitude estimation task for an arbitrary n-qudit state state ρ and hence, by extension, the many-versus-one distinguishing task Task 2, needs access to at least $N \geq \Omega((d-1)^n/(c\varepsilon^2)$ copies of ρ.*

4 Amplitude estimation for $\mathcal{W}^{(d)}$ is d-copy efficient

Similar to the Pauli-tomography task, one can bypass the exponential scaling of the sample complexity by a sufficiently large quantum memory that permits multi-copy measurements. In the same way, as two-copies are sufficient for Pauli-tomography, performing entangling measurements across d copies of the state allows to measure all Weyl-Heisenberg strings simultaneously. However, like for the Pauli tomography, this partially blurs the information, and only allows to predict the absolute value i.e. the amplitude reliably. If one additionally wishes to extract information about the phase, one has to go through a somewhat involved post-processing scheme, similar to the one proposed in [29] which is described in Section 5.2.

One might be tempted to simply estimate $|\mathrm{Tr}[W\rho]|$ via the conjugate representation $W \otimes \bar{W}$. However, this would require access to $\bar{\rho}$ as an additional resource, which cannot be obtained from ρ via any physical operation. The power of having access to $\bar{\rho}$ as an additional resource is discussed extensively in [33]. If however, the unknown state is assumed or believed to be real-valued, then $\bar{\rho} = \rho$ and this observation provides a significant shortcut. On the other hand, complex numbers are an intrinsic feature of quantum mechanics and generic states are certainly not exclusively real-valued [34], and this observation does not contradict our overall conclusion.

Let us first consider the case where $n = 1$, i.e. a single d-dimensional qudit. First we observe that according to Eq. (1), any two elements of $\mathcal{W}$ commute up to a power of ω as a prefactor. Hence, under the d-th power tensor representation $\Phi^d : W \mapsto W^{\otimes d}$, all elements in $\mathcal{W}$ commute. In particular, there is an orthonormal basis (ONB), in which all $W^{\otimes d}$, $W \in \mathcal{W}$ are simultaneously diagonal. A relatively straightforward calculation shows that this diagonalization is achieved by the following natural generalization of the Bell-basis to d-qudit systems

$$|\phi_{I,q}\rangle = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \omega^{kq} |(0, I) + k\mathbf{1}\rangle, \text{ where } q \in [d], I \in [d]^{d-1}.$$

It is also easy to check that the states $|\psi_{I,q}\rangle$ for $I \in [d]^{d-1}$, $q \in [d]\}$ form an ONB of $\left(\mathbb{C}^d\right)^{\otimes d}$. What is more, these orthonormal states are also eigenstates of all $W^{\otimes d}$. For this it sufficient to consider the generators $X, Z \in \mathcal{W}$:

$$X^{\otimes d}|\phi_{I,q}\rangle = \frac{1}{\sqrt{d}} \sum_{k} \omega^{kq} |(0, I) + k\mathbf{1}\rangle = \omega^{-q} |\psi_{I,q}\rangle, \tag{11}$$

$$Z^{\otimes d}|\phi_{I,q}\rangle = \frac{1}{\sqrt{d}} \sum_{k=0} \omega^{kq} \omega^{|I|+dk} |(0, I) + k\mathbf{1}\rangle = \omega^{|I|} |\psi_{I,q}\rangle. \tag{12}$$

Now observe the following: The dependence of the eigenvalue of any $(X^a Z^b)^{\otimes d}$ on the basis state $|\psi_{I,q}\rangle$ can be significantly coarse-grained. Specifically, the information on the multiindex I gets blurred up to the value $|I|$. Therefore, in order to determine the outcomes of all Weyl-Heisenberg observables, it suffices to consider the following projective d^2-outcome measurement

$$\{\Pi_{s,q} | s, q \in [d]\} \text{ where } \Pi_{s,q} = \sum_{I \in [d]^{d-1}, |I|=s} |\psi_{I,q}\rangle\langle\psi_{I,q}|, \tag{13}$$

rather than the the full generalized Bell-basis measurement. This measurement primitive readily extends from the single-qudit case to the n-qudits, by performing the (coarse-grained) Bell-basis measurements in parallel on the respective d-qudit clusters. Specifically, for $\mathbf{q} \in [d]^n, \mathbf{s} \in [d]^n$ we consider the PVM $\{\Pi_{\mathbf{q},\mathbf{s}}\}$ with $\Pi_{\mathbf{q},\mathbf{s}} := \bigotimes_{k=1}^n \Pi_{q_k,s_k}$. In particular, the expectations of the WH-string are determined as

$$\mathrm{Tr}[(X^{\otimes \mathbf{a}} Z^{\otimes \mathbf{b}})^{\otimes d} \rho^{\otimes d}] = \sum_{\mathbf{q},\mathbf{s}} \omega^{\langle \mathbf{b},\mathbf{s}\rangle - \langle \mathbf{a},\mathbf{q}\rangle} \mathrm{Tr}(\Pi_{\mathbf{q},\mathbf{s}} \rho^{\otimes d}).$$

From (independent repetitions of) this single basis measurement on dn qudits in total, the values of all observables $\mathrm{tr}\left(W^{\otimes d}\rho^{\otimes d}\right) = \mathrm{tr}\left(W\rho\right)^d$ can be jointly estimated. This, in turn, allows us to predict the absolute values $|\mathrm{Tr}[W\rho]|$ up to additive error ε, by estimating $\mathrm{Tr}[W^{\otimes d}\rho^{\otimes d}]$ up to precision ε^d

for all $W \in \mathcal{W}^n$. Note, that aside from the ability to process d, n-qudit states the protocol does not require quantum resources that hinder practical applicability. Specifically, the circuits performing the generalized Bell-state measurements can be executed in parallel as schematically depicted in Fig. 1a), and can be implemented in a way using at most two-qudit gates as non-local gates. In particular, the depth remains unaffected as a function of n. Explicit quantum circuits for these execution of these measurements, as well as a transpilation procedure for embedding them into qubit systems for the case of qu*trits* are depicted in Fig. 1b).

Sample complexity

For $W \in \mathcal{W}^n$ let $X_W^{(i)}$ denote the associated random variable taking values in $\{1, \omega, \dots, \omega^{d-1}\}$ according to the measured outcome in round i. In order to obtain estimates on the sample complexity, we employ Hoeffding's inequality (see Lemma D.1). For this, we first observe that if a complex number z satisfies $|z| \geq c$ for some $c > 0$, we have $\max\{|\Re|(z), |\Im(z)|\} \geq \frac{c}{\sqrt{2}}$. By employing a union bound for maximizing over real and imaginary parts as well as all Weyl-Heisenberg matrices, we find (since $|X_i - \mathbb{E}[X_i]| = \left|X_i - \mathrm{Tr}[\rho W]^d\right| \leq 2$)

$$\begin{aligned}
\Pr\left[\max_{W\in\mathcal{W}^d}\left|\sum_{i=1}^N \frac{X_W^{(i)} - \mathrm{Tr}(W\rho)^d}{N}\right| \geq \varepsilon\right] &\leq \sum_{W\in\mathcal{W}^d} \Pr\left[\max_{F\in\Re,\Im}\left|\sum_{i=1}^N \frac{F[X_i - \mathrm{Tr}(W\rho)^d]}{N}\right| \geq \frac{\varepsilon}{\sqrt{2}}\right] \\
&\leq \sum_{W\in\mathcal{W}^d}\sum_{F=\Re,\Im} \Pr\left[\left|\sum_{i=1}^N \frac{F[X_i - \mathrm{Tr}(W\rho)^d]}{N}\right| \geq \frac{\varepsilon}{\sqrt{2}}\right] \\
&\leq 2d^{2n} 2e^{-\varepsilon^2 N/4}.
\end{aligned}$$

Since $||\mathrm{Tr}(W\rho)| - r| \leq \left|\mathrm{Tr}(W\rho)^d - r^d\right|^{1/d}$ for any $0 \leq r \leq 1$, this shows that altogether $S \geq d\log(\sqrt{2})\varepsilon^{-2d} n \log d \log(\delta^{-1})$ state copies are sufficient to determine all $|\mathrm{Tr}[W\rho]|$ up to ε-precision simultaneously with confidence $1-\delta$.

Quantum circuit complexity

As we showed in the previous section, it is possible to estimate all Weyl-Heisenberg-amplitudes *sample-efficiently*. Overall however, precious little would be gained if this advantage was only achieved by significantly increasing the resource requirements on the quantum hardware. As we clarified already, we perform measurements across a *constant* number of states, which is a significant advantage over shadow tomography. Additionally, we observe that while the measurement must act as a entangling measurement on the global system of dimension d^{dn}, the circuit size scales only linearly in n, since the diagonal basis can be prepared as the n-fold tensor product of generalized Bell states on every group of d qudits. Moreover, preparing the generalized Bell states (as in Fig. 1a)) in terms of single- and two-qudit Cliffords is straighforward and requires circuits of size $\mathcal{O}(d)$ and *constant* depth 2. While the depth and size of the circuits preparing individual generalized Bell-basis measurements will in general change depending on the underlying d if one were interested in transpiling the qudit architecture to qubit systems, this does not affect the linear scaling in n. To summarize our findings of this section until now, we have also proven the first part of Theorem 4.

Theorem 10. *By performing parallel generalized Bell measurements, we find a d-copy efficient protocol, which uses $\mathcal{O}(nd\log(d)\varepsilon^{-2d})$ samples of ρ to solve the Weyl-Heisenberg amplitude estimation up to additive error $\varepsilon > 0$. The quantum circuits to realize these measurements in terms of single- and two-qudit Clifford gates have size $\mathcal{O}(n)$ and constant depth, see Section E.*

Remark on composite d

In our previous derivation of the lower bound, the proof of Theorem 8 critically depends on d being a prime number. The hardness was essentially derived by only considering a subset of all Weyl-Heisenberg strings, namely $\hat{\mathcal{W}}^n$, where $\hat{\mathcal{W}} = \mathcal{W} \setminus \mathcal{W}_\infty$. Complementing this, we observe the

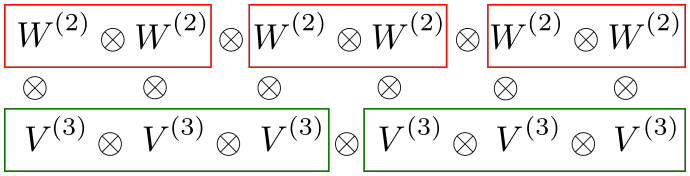


Figure 2: Visualization of the efficient 6-copy protocol for the combined learning task to obtain estimates of $\mathcal{W}^{(2)}\otimes\mathcal{W}^{(3)}$ as elaborated in Section 5.1. The boxes indicate the parallel execution of the generalized-Bell-basis measurements on the respective sites, to predict all the corresponding values of the Weyl-operators.

following: For a composite number $d = p_{\min}q$, where $p_{\min}$ denotes the smallest prime factor of d, estimating the amplitudes of all strings with Weyl-Heisenberg strings with values in $\hat{\mathcal{W}}^{(d)}$ is $p_{\min}$-copy-efficient. The key idea is to find a decomposition

$$\hat{\mathcal{W}}^n = \bigcup_{j=1}^{m} S_j \quad \text{s.t. } W,\tilde{W}\in S_j \Rightarrow [\![W,\tilde{W}]\!]\in\{\omega^{kq}\mathbb{1}|k\in[d]\}. \tag{14}$$

Then, for a fixed S_j all $W\in S_j$ commute under the $p_{\min}$-fold tensor representation, and hence can then be measured simultaneously. If m can be chosen independently of n, the scaling to predict all values $|\mathrm{Tr}(W\rho)|$ then remains linear in the system size n. Indeed, we can find such a decomposition with $m = q$ as follows. For $\mathbf{a}\in[d]^n$, let $W_{\mathbf{a}} = \bigotimes_{k=1}^n W_{a_k}$ the generator of the cyclic group $\mathcal{W}_{\mathbf{a}}$. Now observe that

$$[\![W_{\mathbf{a}},W_{\mathbf{b}}]\!] = [\![X^{\otimes\mathbf{1}}Z^{\otimes\mathbf{a}},X^{\otimes\mathbf{1}}Z^{\otimes\mathbf{b}}]\!] = \omega^{|\mathbf{b}|-|\mathbf{a}|}\mathbb{1}.$$

Since the $W_{\mathbf{a}}$ are the generators of the corresponding cyclic subgroups $\mathcal{W}_{\mathbf{a}}$, we can conclude, that for any two $\mathbf{a},\mathbf{b}\in[d]^n$ where $|\mathbf{a}|-|\mathbf{b}|$ is divisible by q, we have $[\![W,\tilde{W}]\!]\in\{\omega^{kq}\mathbb{1}\}$. Therefore, we indeed find a decomposition as above Eq. (14) via

$$\hat{\mathcal{W}}^n = \bigcup_{k\in[q]} S_k, \quad S_k := \bigcup_{|\mathbf{a}|=_q k} \mathcal{W}_{\mathbf{a}}. \tag{15}$$

Of course, when bringing back $\mathcal{W}_\infty$ into the picture, the argument breakes down, and despite considerable efforts we did not manage to provide a decomposition as in Eq. (14) without allowing m to scale with n. It would be interesting to understand if such a protocol can actually be found or if this can be attributed to the inclusion of $\mathcal{W}_\infty$ rendering learning provably hard for $(d-1)$-copy measurement schemes also for composite d, which we leave open to further research.

5 Extensions of the Main Results

This section discusses two additional results related to solving the Weyl-Heisenberg amplitude estimation task. First, we show how the results from the previous two sections, can be used to construct learning tasks also for any square-free degree. Unrelated to this, we explore how one can extend the efficient protocol for WH-amplitude estimation to full WH-tomography in Section 5.2.

5.1 Construction of learning tasks with square-free degree

Our method is not applicable to derive hardness results for $\mathcal{W}^{(d)}$-amplitude estimation if d is composite, as elaborated above. However, we can construct such degree-d learning tasks for every d which is square-free, i.e. when d has a prime factor decomposition $d = p_1\cdot\ldots\cdot p_k$ where each prime appears at most once. With d factorized as above, we consider the learning task of determining the amplitudes of all $W\in(\mathcal{W}^{(\mathbf{p})})^{\otimes n}$, where $\mathcal{W}^{(\mathbf{p})} := \mathcal{W}^{(p_1)}\otimes\cdots\otimes\mathcal{W}^{(p_k)}$. For this, we have the following result:

Theorem 11. *Let d be a square-free integer. The task to estimate $|\mathrm{Tr}(W\rho)|$ for all $W\in(\mathcal{W}^{(\mathbf{p})})^{\otimes n}$ up to fixed additive error $\varepsilon>0$ is of degree d.*

Proof. Since d is a common multiple of all its factors, it is straightforward to convince oneself, that all $W, \tilde{W}$ commute up to a power of the d-th root of unity ω_d. The d-copy measurement scheme outlined above can then be readily applied up to minor modifications, and therefore the task is d-copy efficient. On the other hand, the associated quantity $\delta_{\varepsilon,c}$ as defined in Eq. (6) can be estimated via Eq. (8) completely analogous, where here we choose $\hat{\mathcal{W}}^{(\mathbf{p})} := \hat{\mathcal{W}}^{(p_1)} \otimes \cdots \otimes \hat{\mathcal{W}}^{(p_k)}$ with $\hat{\mathcal{W}}^{(p)} = \mathcal{W}^{(p)} \setminus \mathcal{W}_\infty^{(p)}$ as before for $p \geq 3$, and $\hat{\mathcal{W}}^{(2)} = \mathcal{W}^{(2)}$. Now, observe that for any $\tau \in \{\pm 1\}^{2m}$ we have

$$\left\| \sum_{W \in \hat{\mathcal{W}}^{(\mathbf{p})}} W^{\otimes \tau} \right\|_\infty = \prod_{l=1}^{k} \left\| \sum_{W \in \mathcal{W}^{(p_l)}} W^{\otimes \tau} \right\|_\infty = \prod_{l=1}^{k} p_l \varphi(p_l)^{\delta(m=0 \mod p_l)}, \tag{16}$$

where $\varphi(p) = p - 1$ when $p \neq 2$, and $\varphi(2) = 2$ (see Lemma B.2 in Section B). In particular, we find the corresponding average:

$$\frac{1}{\left|\hat{\mathcal{W}}^{(\mathbf{p})}\right|} \left\| \sum_{W \in \hat{\mathcal{W}}^{(\mathbf{p})}} W^{\otimes \tau} \right\|_\infty = \prod_{l=1}^{k} \frac{p_l \varphi(p_l)^{\delta(m=0 \mod p_l)}}{p_l \varphi(p_l)} = \prod_{l=1}^{k} \varphi(p_l)^{-\delta(m \neq 0 \mod p_l)}. \tag{17}$$

Now, it is important to observe, that because d is square-free, for every $m < d$, there is at least one of the $p_1, \ldots, p_k$ which does not divide m. In particular, with $p_{\min}$ being the smallest prime factor of d, we find for any $m < d$ and $\tau \in \{\pm 1\}^{2m}$ that

$$\frac{1}{\left|\hat{\mathcal{W}}^{(\mathbf{p})}\right|} \left\| \sum_{W \in \hat{\mathcal{W}}^{(\mathbf{p})}} W^{\otimes \tau} \right\|_\infty \leq \varphi(p_{\min})^{-1}. \tag{18}$$

This ensures that, as before, the bound on $\delta_{\varepsilon,c}$ decreases exponentially in n, as long as $c < d$, and thus renders the learning task $d-1$-copy hard, which concludes the proof. □

5.2 Extension of efficient WH-amplitude estimation to full WH-tomography

The protocol outlined in Section 4 only provides estimates of the amplitude of $\mathrm{Tr}(W\rho)$ for all Weyl-Heisenberg strings W, rather than the actual value. If for whatever application one is interested in, this is not enough, and additionally one needs the phase information there is a way to expand the protocol to estimate the actual value of all $\mathrm{Tr}(W\rho)$ in a sample-efficient manner. However this comes at the expense of requiring access to significantly increased classical and quantum computational resources. The idea is to prepare copies of a known state σ, which has the property that $|\mathrm{Tr}(W\sigma)|$ is significant whenever $|\mathrm{Tr}(W\rho)|$ is, as determined beforehand via WH-amplitude estimation. Specifically, given $d-1$ copies of such a *mimicking state* σ together with one copy of ρ, we can jointly estimate $\mathrm{Tr}(W^{\otimes d}\sigma^{\otimes d-1} \otimes \rho)$ for all W. Then, since the values of $\mathrm{Tr}(W\sigma)$ are known, we can efficiently determine the actual values of all $\mathrm{Tr}(W\rho)$. The computational bottleneck for this procedure is clearly to find such a mimicking state σ. This can be achieved based on Hamiltonian update rules via the matrix multiplicative weight algorithm, following ideas of [29, 35]. Note, that our scenario again requires minor adaptions compared to these mentioned results. The modified algorithm, the correctness of which we prove formally in Lemma B.3, is stated below as Protocol 1. Finally, let us briefly analyze how the various error thresholds propagate during the procedure. As we prove in the Appendix, given $\varepsilon > 0$ and $\varepsilon/6$-accurate estimates of all $|\mathrm{Tr}(W\rho)|$, with high probability, Protocol 1 outputs a state σ, such that $|\mathrm{Tr}(W\rho)| \geq \varepsilon \Rightarrow |\mathrm{Tr}(W\sigma)| \geq \varepsilon/3$. Let z_W be ε^d-accurate estimators of $\mathrm{Tr}(W^{\otimes d}\sigma^{d-1} \otimes \rho)$. Then

$$\left| \mathrm{Tr}[W\rho] - \frac{z}{\mathrm{Tr}[W\sigma]^{d-1}} \right| = \frac{\left| \mathrm{Tr}[W\sigma]^{d-1} \mathrm{Tr}[W\rho] - z \right|}{|\mathrm{Tr}[W\sigma]|^{d-1}} \leq \frac{\varepsilon^d}{(1/3)^{d-1}\varepsilon^{d-1}} = 3^{d-1}\varepsilon.$$

In particular, we see that the procedure outlined in this section requires only $\mathcal{O}(nd\log(d)\varepsilon^{-2d})$ samples of ρ over all involved steps to estimate all the values $\mathrm{Tr}(W\rho)$ up to additive error $\varepsilon > 0$.

Protocol 1 Mimicking state construction

Given: estimates u_W s.t. $|u_W - |\mathrm{Tr}(\rho W)|| \leq \varepsilon/6$.
Output: σ, such that $|\mathrm{Tr}(W\sigma)| \geq \varepsilon/3$ whenever $u_W \geq \varepsilon$.
Initialise: $T = \lfloor 75n/\varepsilon^2 \rfloor + 2$; $\beta = \sqrt{n/T}$; $\sigma_0 = \mathbb{1}/d^n$.
For $t = 0, \ldots, T-1$ do the following:

1. Search for $W_t \in \mathcal{W}^n$ s.t. $u_t := u_{W_t} \geq \varepsilon$ and $||\mathrm{Tr}(W_t\sigma_t)| - u_t| \geq 2\varepsilon/3$.
2. If no such W_t is found, output $\sigma = \sigma_t$.
3. Otherwise:
   (a) Use $\mathcal{O}(\log(T)/\varepsilon^2)$ copies of ρ to produce an estimate z_t such that $|\mathrm{Tr}(W_t\rho) - z_t| \leq \varepsilon/25$ with probability $p \geq 1 - 0.01/|T|$.
   (b) Determine $\mathrm{argmax}_{F\in\{\Re,\Im\}} |F(\mathrm{Tr}(W_t\sigma_t) - z_t)|$.
   If argmax $= \Re$ then: $\hat{W}_t := \frac{W_t + W_t^\dagger}{2}$, and $\hat{z}_t := \Re[z_t]$
   If argmax $= \Im$ then: $\hat{W}_t := \frac{W_t - W_t^\dagger}{2i}$, and $\hat{z}_t := \Im[z_t]$
   $M_t := \mathrm{sgn}(\mathrm{Tr}(\hat{W}_t\sigma_t) - \hat{z}_t) \cdot \hat{W}_t$
   (c) $\sigma_{t+1} := \frac{1}{\mathcal{N}} \exp\left(-\beta \sum_{\tau=1}^{t} M_\tau\right)$ with suitable normalization factor $\mathcal{N}$.

Output $\sigma = \sigma_T$.

6 Conclusion and Outlook

We have successfully demonstrated the existence of degree-d learning tasks for every prime number d, that is learning tasks which are hard to solve if the measurement device can process only up to $d-1$ copies of an unknown state at one, but become efficiently solvable, reducing to a linear scaling in the system size, if d-copies can be measured together.

We achieved this by the natural generalization of the degree-2 task of determining the amplitudes of all Pauli strings, as discussed in [15, 21], to the Weyl-Heisenberg group. The generalization of the efficient d-copy scheme is pretty much straightforward, by considering the natural extension of Bell-basis measurements to qudit-systems, as we demonstrated in Section 4. On the other hand, we reduced the learning task to a hypothesis test in the framework of many-versus-one-distinguishing tasks, and exploited the machinery developed in [21] to derive lower on the number of iterations/measurements necessary to solve this type of hypothesis test reliably. This is where our assumption on d being prime is essential. Based on these results we also provided a construction to generate degree-d-learning tasks for all d which are *square-free*. Finally, we also briefly outlined how one can extend this scheme to the full tomography of all Weyl-Heisenberg strings, at the expense of exponentially increasing computational cost, while remaining sample efficient.

Our learning challenges can either be formulated in terms of Weyl-Heisenberg shadow tomography on n-partite qudit systems ($d \geq 2$), or they can readily be transpiled into equivalent learning challenges on $O(\log(d)n)$-qubit systems. The latter is conceptually closer to the thematically related, but independent, submission [28], as well as previous work [14, 15] (which we recover in the special case of $d = 2$). The former interpretation, however, opens up interesting future research directions as well. Qudit-based classical simulation techniques could be used to analyze the susceptibility of these learning strategies vis a vis to noise. This can, in turn, inform the design of quidt-based noise mitigation strategies that may then allow an actual implementation of Weyl-Heisenberg tomography on a native qudit-based quantum processor [36]. It would also be interesting to compare the performance of such genuine qudit-based implementations with an equivalent learning protocol transpiled for qubit-based architectures.

On a more conceptual level, we believe that our generalizations of the two-copy learning results from Refs. [14, 15] can also form the basis of a generalization of several more sophisticated learning protocols that build upon two-copy Bell sampling (also known as the destructive SWAP test).

Concrete examples are triply-efficient shadow tomography [29] (which we have already partially addressed in Algorithm 1 above), as well as alternative protocols for auxiliary-free replica shadow estimation [37].

7 Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project numbers 441423094, 236615297 - SFB 1119, the Austrian Science Fund (FWF) via the SFB BeyondC (10.55776/FG7) and an FWF START award, as well as the European Research Council (ERC) via the Starting grant q-shadows.

Appendix

The Appendix is organized as follows: Section A briefly reviews the qudit-Clifford group, supplemental to the introduction of the Weyl-Heisenberg group in the main text. In Section B we provide the proofs of some results which we claimed but did not prove in the main part, where other technical results are outsourced further to Section D. Finally, Section E features the quantum circuits necessary to run the d-copy-efficient protocol from Section 4.

A Qudit quantum computing

Quantum computing in qubits makes heavily use of the Pauli matrices, which are "classical" quantum operations, in the sense, that if one splits a qubit into a amplitude and a phase information, they merely act as the flip operator, they themselves can neither create nor destroy coherence in the computational basis. In the main text, we already provided the analogue of the Pauli group in higher dimensions, the Weyl-Heisenberg group. Similarly to the qubit case, we can then proceed to define the qudit-Clifford group as the symmetry group of the Weyl-Heisenberg matrices.

$$\mathcal{C}l = \{U | UWU^\dagger \in \mathcal{W} \text{ for all } W \in \mathcal{W}\}. \tag{19}$$

Similar to the qubit case, the Clifford group is the a maximal finite subgroup of the unitary group $\mathrm{U}(d)$. In particular, in the single-qudit case it has only two generators, which resemble the Hadamard and the S-gate. For the n-qudit Clifford group, next to the local generators S and H on each subsystem, on additionally needs $n-1$ generators which generalise the role of the controlled-NOT operation in the multi-qudit case, see e.g. [38]. We now explicitly state these generators, and some of their properties inside the Clifford group. First, consider the analogue of the Hadamard-gate, defined via

$$H = \frac{1}{\sqrt{d}} \sum_{k,j\in[d]} \omega^{kj} |k\rangle\langle j|. \tag{20}$$

It has order four $H^4 = \mathbb{1}$, and interchanges between the X and Z bases such that $HZH^\dagger = X$. The analogue to the S-gate in the qubit Clifford group, is the following:

$$S = \sum_{j\in[d]} \omega^{j(j-1)/2} |j\rangle\langle j|. \tag{21}$$

It satisfies $S^d = \mathbb{1}$, and together with H generates Z via $Z = H^2 S^{d-1} H^2 S$. In particular, the operator S is constructed such that $|\psi_j^a\rangle = S^a |\psi_j^0\rangle$ for $a \neq \infty$. Additionally we have the following two-qudit gate, which serves as a generalization of the CNOT gate, where the addition on the second qudit is now simply performed modulo d:

$$CX = \sum_{k,l\in[d]} |k\rangle\langle k| \otimes |k+l\rangle\langle l|. \tag{22}$$

Together, the local generators S and H on every qudit and the two-qudit gates $CX^{(j,j+1)}$ for $j = 1, \ldots, n-1$ generate the entire n-qudit Clifford group.

B Outsourced Proofs from Main part

We begin with providing the proofs of the two main steps for deriving the lower bound on the sample complexity in Section 3.

Lemma B.1. *[Compare with [21], Thm. 15, slightly extended and modified vesrion thereof]*

$$\delta_{c,\varepsilon} \leq \left(\sum_{0\neq S\in[2]^c} \frac{(6\varepsilon)^{|S|}}{\sqrt{d(d-1)}^n} \sqrt{\max_{\tau\in\{\pm1\}^{2|S|}} \left\| \sum_{W\in\hat{\mathcal{W}}} W^{\otimes\tau} \right\|^n} \right)^2. \tag{23}$$

Proof. The proof is essentially analogous to the one of Theorem 15 in [21]. Due to subtle differences in several places, however, we conduct the proof here nonetheless. First of all, we spell out the χ^2-divergence for fixed W and measurement $M = \{M_a\}_{a\in A}$. Specifically, we have

$$\begin{aligned}\chi^2_M\left((\rho_W^\varepsilon)^{\otimes c} \,||\, \rho_m^{\otimes c}\right) &= \sum_{a\in A} \operatorname{Tr}(M_a\rho_m^{\otimes c}) \left(\frac{\operatorname{Tr}(M_a(\rho_W^\varepsilon)^{\otimes c})}{\operatorname{Tr}(M_a\rho_m^{\otimes c}} - 1\right)^2 \\ &= \sum_{a\in A} \frac{\operatorname{Tr}(M_a)}{d^{cn}} \left(\frac{\operatorname{Tr}(M_a(\mathbb{1} + 3\varepsilon(W+W^\dagger))^{\otimes c})}{\operatorname{Tr}(M_a)}\right)^2 \\ &= \sum_{a\in A} \frac{1}{\operatorname{Tr}(M_a)d^{cn}} \left(\sum_{0\neq S\in[2]^c} (3\varepsilon)^{|S|} \operatorname{Tr}(M_a(W+W^\dagger)^{\otimes S}\right)^2 .\end{aligned}$$

Next, we introduce arbitrary non-zero numbers $w_S > 0$ for each $0 \neq S \in [2]^c$ and apply Cauchy Schwarz in a particular way, to find that for any probability distribution μ:

$$\begin{aligned}&\max_M \mathbb{E}_{W\sim\mu} \left[\sum_{a\in A} \frac{1}{\operatorname{Tr}(M_a)d^{cn}} \left(\sum_{0\neq S\in[2]^c} \frac{\sqrt{w_S}}{\sqrt{w_S}} (3\varepsilon)^{|S|} \operatorname{Tr}(M_a(W+W^\dagger)^{\otimes S}\right)^2\right] \\ \leq &\max_M \mathbb{E}_{W\sim\mu} \left[\sum_{a\in A} \frac{1}{\operatorname{Tr}(M_a)d^{cn}} \sum_{0\neq S\in[2]^c} \frac{1}{w_S} (3\varepsilon)^{2|S|} \operatorname{Tr}[M_a(W+W^\dagger)^{\otimes S}] \sum_{0\neq S\in[2]^c} w_S\right] \\ \leq &\sum_{0\neq S\in[2]^c} w_S \sum_{0\neq S\in[2]^c} \frac{(3\varepsilon)^{2|S|}}{w_S} \max_M \mathbb{E}_{W\sim\mu} \left[\sum_a \frac{\operatorname{Tr}(M_a(W+W^\dagger)^{\otimes S})}{d^{cn}\operatorname{Tr}(M_a)}\right] \\ = &\left(\sum_{0\neq S\in[2]^c} \sqrt{\max_M \mathbb{E}_{W\sim\mu} \left[\sum_a \frac{\operatorname{Tr}(M_a(W+W^\dagger)^{\otimes S})}{d^{cn}\operatorname{Tr}[M_a]}\right]}\right)^2 .\end{aligned}$$

where we picked $w_S = \sqrt{\max_M \mathbb{E}_\mu \sum_a \frac{\operatorname{Tr}(M_a(W+W^\dagger)^{\otimes S})}{d^{cn}\operatorname{Tr}[M_a]}} > 0$ to arrive at the last term. For any distribution μ and (finite) positive operator-valued measure (POVM) $M = \{M_a\}_{a\in A}$ we can upper-bound

$$\delta_{c,\varepsilon} \leq \left(\sum_{0\neq S\in[2]^c} (3\varepsilon)^{|S|} \sqrt{\mathbb{E}_{W\sim\mu} \left[\sum_{a\in A} \frac{\operatorname{Tr}[M_a(W+W^\dagger)^{\otimes S}]^2}{d^{cn}\operatorname{Tr}[M_a]}\right]}\right)^2 . \tag{24}$$

Let us now pick the distribution $\hat{\mu}$ specified before and let $\{M_a\}_{a\in A}$ be an arbirtrary POVM acting on $\mathbb{C}^{dn}$. In order to expand the tensor product $(W+W^\dagger)^{\otimes S}$ we introduce the following index set: For a fixed $S \in [2]^c$, let $\mathcal{I}(S) = \{I \in \{-1,0,1\}^c | \operatorname{supp}(S) = \operatorname{supp}(I)\}$. Note that $|\mathcal{I}(S)| = 2^{|S|}$, where $|S|$ here refers to the number of non-zero entries of S. Due to unitarity, it then holds that $(W+W^\dagger)^{\otimes S} = \sum_{\tau\in\mathcal{I}(S)} W^{\otimes\tau}$. Via simple algebraic manipulations, and the fact that for any operators A, B where A is positive semidefinite, it holds that $|\operatorname{Tr}[AB]| \leq \|B\|_\infty \operatorname{Tr}[A]$,

we can derive an upper bound for the term inside the square root in Eq. (24):

$$\sum_{W\in\hat{\mathcal{W}}^n}\sum_{a\in A}\frac{\operatorname{Tr}[M_a(W+W^\dagger)^{\otimes S}]^2}{d^{cn}\operatorname{Tr}[M_a]} = \sum_{a\in A}\frac{1}{d^{cn}\operatorname{Tr}[M_a]}\sum_{W\in\hat{\mathcal{W}}^n}\operatorname{Tr}\left[M_a\sum_{\tau\in\mathcal{I}(S)}W^{\otimes\tau}\right]^2$$
$$=\sum_{a\in A}\frac{1}{d^{cn}\operatorname{Tr}[M_a]}\sum_{W\in\hat{\mathcal{W}}^n}\operatorname{Tr}\left[M_a\otimes M_a\sum_{\tau\in\mathcal{I}(S)}W^{\otimes\tau}\otimes\sum_{\tau\in\mathcal{I}(S)}W^{\otimes\tau}\right]$$
$$=\sum_{a\in A}\frac{1}{d^{cn}\operatorname{Tr}[M_a]}\sum_{\tau\in\mathcal{I}(S)^2}\operatorname{Tr}\left[M_a\otimes M_a\sum_{W\in\hat{\mathcal{W}}^n}W^{\otimes\tau}\right]$$
$$\leq\sum_{a\in A}\frac{\operatorname{Tr}[M_a\otimes M_a]}{d^{cn}\operatorname{Tr}[M_a]}\sum_{\tau\in\mathcal{I}(S)^2}\left\|\sum_{W\in\hat{\mathcal{W}}^n}W^{\otimes\tau}\right\|_\infty=\sum_{\tau\in\mathcal{I}(S)^2}\left\|\sum_{W\in\hat{\mathcal{W}}^n}W^{\otimes\tau}\right\|_\infty$$
$$\leq\sum_{\tau\in\mathcal{I}(S)^2}\left\|\left(\sum_{W\in\hat{\mathcal{W}}}W^{\otimes\tau}\right)^{\otimes n}\right\|_\infty=\sum_{\tau\in\mathcal{I}(S)^2}\left\|\sum_{W\in\hat{\mathcal{W}}^n}W^{\otimes\tau}\right\|_\infty^n=4^{|S|}\max_{\tau\in\{\pm1\}^{2|S|}}\left\|\sum_{W\in\hat{\mathcal{W}}^n}W^{\otimes\tau}\right\|_\infty^n.$$

Clearly, for sites k where $\tau_k=0$ we have $W^{\tau(k)}=\mathbb{1}$ on this sites for all W. Therefore the sites can be factored into a tensor product of the identity with the sum over the remaining sites. Since tensoring with the identity does not change the operator norm of any operator, we can therefore assume that the strings τ only take values in $\{\pm1\}$ and are of length $|S|$. However, since we are in a tensor can be can be pulled out of the summation. To summarize, since $\left|\hat{\mathcal{W}}^n\right|=(d(d-1))^n$ we end up with the estimate

$$\delta_{\varepsilon,c}\leq\left(\sum_{0\neq S\in[2]^c}\frac{(6\varepsilon)^{2|S|}}{\sqrt{d(d-1)}^n}\sqrt{\max\tau\in\{\pm1\}^{2|S|}\left\|\sum_{W\in\hat{\mathcal{W}}}W^{\otimes\tau}\right\|_\infty}\right)^2 \tag{25}$$

□

We state the following theorem slightly more general than in the main text, in order to also apply it to the later extensions. For convenience of the reader, we repeat the statement of Theorem 8 before giving its proof. In particularly, we state it slightly more general than in the main text, to highlight its applicability to the situation in Section 5.1.

Proposition 8 (restated)**.** *Let $d\geq3$ be a prime, m a natural number, and $\tau\in\{\pm1\}^{2m}$. Then we have*

$$\left\|\sum_{W\in\hat{\mathcal{W}}}W^{\otimes\tau}\right\|_\infty=\begin{cases}d(d-1) & \text{for } m=0 \mod d,\\ d & \text{otherwise.}\end{cases} \tag{26}$$

Proof. In order to avoid confusion from the start, all occurences of δ in this proof refer to the Kronecker delta, not to Eq. (6). Also, for simplicity we assume $m\leq d$, since only the parity of m modulo d is important here. For the purpose of our proof let $M_\tau:=\sum_{W\in\hat{\mathcal{W}}}W^{\otimes\tau}$. In order to keep the presentation of the proof as neat as possible, let $\kappa:=\langle\tau,\mathbf{1}\rangle/2\in\mathbb{Z}$ and $T(J):=\langle\tau,J\rangle$ for $J\in[d]^{2m}$. As we will show in the following, the respective exact expressions of M_τ are given by

$$M_\tau=d\sum_{T(J)\neq\kappa}|J+m^{-1}(\kappa-T(J))\tau\rangle\langle J| \qquad \text{for } m<d, \tag{27}$$

$$M_\tau=d\sum_{T(J)=\kappa}\sum_{a=1}^{d-1}|J+a\tau\rangle\langle J| \qquad \text{for } m=d. \tag{28}$$

From these expression it is not hard to derive the claim of the theorem. To prove that $\|M_\tau\|=d$ in the case $m<d$, it suffices to show that the mapping $J\mapsto J+m^{-1}(\kappa-T(J))\tau$ is injective: Let

J, K such that

$$J + m^{-1}(\kappa - T(J))\tau = K + m^{-1}(\kappa - T(K))\tau. \tag{29}$$

By applying $T(\cdot)$ to this condition we find that $T(J) = T(K)$ must hold, since $T(\tau) = \langle \tau, \tau \rangle = 2m$. Clearly, inserting this relation back into Eq. (29) it then also follows that $J = K$.

In the latter case $m = d$, first note, that M_τ is the sum (over a) of $d-1$ operators, each with operator norm d. This directly leads to the upper bound $\|M_\tau\| \leq d(d-1)$. On the other hand, consider the (unnormalized) state $|\psi\rangle = \sum_{a=0}^{d-1} |2^{-1}\mathbf{1} + a\tau\rangle$. Since $T(\tau) =_d 0$, $|\psi\rangle$ is contained in the subspace spanned by the $|J\rangle$ with $T(J) = \kappa$. It is easy checked that this is an eigenstate of M_τ:

$$M_\tau |\psi\rangle = d \sum_{a=1}^{d-1} \sum_{b=0}^{d-1} |2^{-1}\mathbf{1} + (a+b)\tau\rangle = d(d-1)|\psi\rangle.$$

In particular, this shows $\|M_\tau\| \geq d(d-1)$, so equality must hold. Let us now derive the expressions Eq. (27), Eq. (28). First of all, we may assume without loss of generality that $\tau_1 = 1$, since exchanging $W \leftrightarrow W^{-1} = W^\dagger$ in the summation gives the same result. We split the sum into summations the subgroups $\mathcal{W}_a = \langle W_a \rangle$ for $a = 0, \ldots, d-1$, (which only intersect at the identity) and express the operators in terms of their eigenprojectors.

$$\begin{aligned}
\sum_{W \in \hat{\mathcal{W}}} W^{\otimes \tau} &= \sum_{\alpha \in [d]} \sum_{k \in [d]} (W_\alpha^k)^{\otimes \tau} - d\mathbb{1} = \sum_\alpha \sum_k \sum_{J \in [d]^{2m}} \omega^{k\langle \tau, J \rangle} \psi_J^\alpha - d\mathbb{1} \\
&= d \sum_\alpha \sum_{\langle \tau, J \rangle = 0} \psi_J^\alpha - d\mathbb{1}.
\end{aligned}$$

For the next step, we exploit the fact that the ω^j-eigenstates of the different W_a are related to each other via the action of the S-gate, and spell out everything in the computational basis. Also, we abbreviate $f(K) = \sum_i k_i(k_i - 1) = \langle K, K - \mathbf{1} \rangle$, which is always an even number.

$$\begin{aligned}
\sum_\alpha \sum_{T(J)=0} \psi_J^\alpha &= \sum_\alpha \sum_{T(J)} (S^\alpha)^{\otimes 2m} \psi_J^0 (S^{-\alpha})^{\otimes 2m} \\
&= d^{-2m} \sum_\alpha \sum_{T(J)=0} \sum_{K,L} (S^\alpha)^{\otimes 2m} \omega^{-\langle J, K \rangle} \omega^{+\langle J, L \rangle} |K\rangle\langle L| (S^{-\alpha})^{\otimes 2m} \\
&= d^{-1} \sum_\alpha (S^\alpha)^{\otimes m} \sum_{K,L} \delta_{(l_1 - k_1)\tau = L - K} |K\rangle\langle L| (S^{-\alpha})^{\otimes 2m} \\
&= d^{-1} \sum_{\alpha, K, L} \delta_{\Delta\tau = K - L} \omega^{\alpha(f(K) - f(L))/2} |K\rangle\langle L| \\
&= \sum_{K,L} \delta_{\Delta\tau = K-L} \delta_{f(K) = f(L)} |K\rangle\langle L|.
\end{aligned}$$

In the third line we applied Lemma D.2, and in the fourth line we substituted $\Delta := k_1 - l_1$. Now we need to jointly solve these two equations. For this purpose, let L be fixed, and we determine all solutions Δ. We substitute the linear constraint $K = L + \Delta\tau$ into the quadratic equation $f(K) = f(L)$ and obtain the following condition:

$$\begin{aligned}
f(L + \Delta\tau) &= \langle L + \Delta\tau, L + \Delta\tau \rangle - \langle L + \Delta\tau, \mathbf{1} \rangle = \langle L, L \rangle - \langle L, \mathbf{1} \rangle = f(L), \\
&\Leftrightarrow\ 2\Delta\langle \tau, L \rangle + \Delta^2 \langle \tau, \tau \rangle - \Delta\langle \tau, \mathbf{1} \rangle = 0, \\
&\Leftrightarrow\ 2\Delta T(L) + 2m\Delta^2 - \Delta 2\kappa = 0.
\end{aligned}$$

Since d is prime ≥ 3, this is equivalent to the following quadratic equation which is central to the proof:

$$(m\Delta + T(L) - \kappa)\,\Delta = 0. \tag{30}$$

So far, all calculations worked independently of m. However, to analyze the solutions of Eq. (30) we need to distinguish two fundamentally different cases. When m is divisible by d, the quadratic

part of the equation vanishes, and for $T(L) - \kappa \neq 0$ we are left with the only solution $\Delta = 0$. However, for $T(L) = \kappa$, all $\Delta = 0, \ldots, d-1$ are solutions. Hence, we find

$$\begin{aligned} M_\tau &= d \left(\sum_{T(J)\neq\kappa} |J\rangle\langle J| + \sum_{T(J)=\kappa} \sum_{a\in[d]} |J + a\tau\rangle\langle J| - \mathbb{1} \right) \\ &= d \left(- \sum_{T(J)=\kappa} |J\rangle\langle J| + \sum_{T(J)=\kappa} \sum_{a\in[d]} |J + a\tau\rangle\langle J| \right) = d \sum_{T(J)=\kappa} \sum_{a=1}^{d-1} |J + a\tau\rangle\langle J|. \end{aligned}$$

On the other hand, when $d < m$, Eq. (30) has only the two solutions $\Delta = 0$ and $\Delta = m^{-1}(\kappa - T(L))$ which of course coincide when $T(L) = \kappa$. Therefore we find

$$M_\tau = d \left(\sum_J |J\rangle\langle J| + \sum_{T(J)\neq\kappa} |J + m^{-1}(\kappa - T(J))\tau\rangle\langle J| - \mathbb{1} \right) \tag{31}$$

$$= d \sum_{T(J)\neq\kappa} |J + m^{-1}(\kappa - T(J))\tau\rangle\langle J|. \tag{32}$$

This concludes the proof. □

Note, that the result above covers only the situation where d is an odd prime, and we are averaging over the reduced Weyl-Heisenberg group $\hat{\mathcal{W}}_d$. To obtain our result in Section 5.1 we also need a result for the even prime $d = 2$, which we state here.

Lemma B.2. *Let k be an integer and $\mathcal{P} = \{I, X, Y, Z\}$ the set of the four Pauli matrices. Then:*

$$\left\| \sum_{P\in\mathcal{P}} P^{\otimes 2k} \right\|_\infty = 3 + (-1)^k. \tag{33}$$

Proof. This is a straightforward calculation: We substitute $X = |0\rangle\langle 1| + |1\rangle\langle 0|$, $Y = i|0\rangle\langle 1| - i|1\rangle\langle 0|$, and I, Z analogously. Then

$$\sum_{P\in\{I,X,Y,Z\}} P^{\otimes 2k} = \sum_{I\in[2]^{2k}} |I\rangle\langle I| + (-1)^{|I|}|I\rangle\langle I| + |I\rangle\langle I \oplus \mathbf{1}| + i^{2k}(-1)^{|I|}|I\rangle\langle I \oplus \mathbf{1}| \tag{34}$$

$$= \sum_{I\in[2]^{2k}} (1 + (-1)^{|I|})|I\rangle\langle I| + (1 + (-1)^{k+|I|})|I\rangle\langle I \oplus \mathbf{1}|. \tag{35}$$

In particular, when k is odd, for every $I \in [2]^k$ either of the two factors will vanish, leading to $\left\|\sum_{P\in\mathcal{P}} P^{\otimes 2k}\right\|_\infty = 2$. On the other hand, when k is even, with $|GHZ_n\rangle$ being the GHZ-state on n qubits, we clearly find $\sum_{P\in\mathcal{P}} P^{\otimes 2k}|GHZ_{2k}\rangle = 4|GHZ_{2k}\rangle$, which concludes the proof. □

Finally, to conclude this section, we provide the proof that the mimicking state algorithm Protocol 1 works correctly.

Lemma B.3. *Let $\varepsilon > 0$ and ρ a fixed quantum states, and u_W estimates of $|\mathrm{Tr}(W\rho)|$ up to additive error $\varepsilon/6$. Then with probability $p \geq 0.99$ the output of Protocol 1 provides a classical description of a state σ such that $\mathcal{W}_{\varepsilon,\rho} \subset \mathcal{W}_{\varepsilon/3,\sigma}$.*

Proof. First of all, the algorithm is guaranteed to be correct (with high probability) whenever no violating $W \in \mathcal{W}^n$ is found in step 1.
Let us therefore assume that the algorithm iterates over all $t = 0, \ldots, T$ and reaches the last step $t = T$. Let $y_t = \mathrm{Tr}(\hat{W}_t\sigma_t)$ for $0 \leq t \leq T$. By Lemma 3.1 from [29] we are guaranteed that after T steps of the algorithm:

$$\begin{aligned} \sum_{t=1}^T \mathrm{sgn}(y_t - \hat{z}_t) \cdot y_t &\leq \lambda_{\min} \left(\sum_{t=1}^T \mathrm{sgn}(y_t - \hat{z}_t) \cdot \hat{W}_t \right) + 2\sqrt{nT} \\ &\leq \sum_{t=1}^T \mathrm{sgn}(y_t - \hat{z}_t)\, \mathrm{Tr}(\hat{W}_t\rho) + 2\sqrt{nT}. \end{aligned}$$

This can be equivalently expressed as

$$\sum_{t=1}^{T} \mathrm{sgn}(y_t - \hat{z}_t)\left(y_t - \mathrm{Tr}(\hat{W}_t\rho)\right) \leq 2\sqrt{nT}. \tag{36}$$

Next, observe, that we can lower bound the difference

$$|\mathrm{Tr}(W_t\sigma_t) - z_t| \geq ||\mathrm{Tr}(W_t\sigma_t)| - u_t| - ||z_t| - u_t| \geq \varepsilon/2 - (\varepsilon/8 + \varepsilon/30) > \varepsilon/3.$$

This in turn leads to the following estimates:

$$|y_t - \hat{z}_t| = \left|\mathrm{Tr}(\hat{W}_t\sigma_t) - \hat{z}_t\right| \geq \frac{1}{\sqrt{2}}\left|\mathrm{Tr}(W_t\sigma_t) - z_t\right| \geq \frac{\varepsilon}{3\sqrt{2}}, \tag{37}$$

$$\left|y_t - \mathrm{Tr}(\hat{W}_t\rho)\right| \geq |y_t - \hat{z}_t| - \left|\hat{z}_t - \mathrm{Tr}[\hat{W}\rho]\right| \geq \frac{\varepsilon}{3\sqrt{2}} - \frac{\varepsilon}{30} > \frac{\varepsilon}{5}. \tag{38}$$

The first estimate, together with $|z_t - \mathrm{Tr}(W_t\rho)| \leq \varepsilon/30$ can be used to show that $\mathrm{sgn}(y_t - \hat{z}_t) = \mathrm{sgn}(y_t - \mathrm{Tr}(\hat{W}_t\rho))$. This can be easily shown by distinguishing the following two cases:

$$\begin{aligned}
&\text{If } \; y_t - \hat{z}_t \geq \frac{\varepsilon}{3\sqrt{2}} \;\Rightarrow\; y_t - \hat{z}_t - (\mathrm{Tr}(\hat{W}_t\rho) - \hat{z}_t) \geq \frac{\varepsilon}{3\sqrt{2}} - \frac{\varepsilon}{30} > 0,\\
&\text{If } \; \hat{z}_t - y_t \geq \frac{\varepsilon}{3\sqrt{2}} \;\Rightarrow\; \hat{z}_t - y_t - (\hat{z}_t - \mathrm{Tr}(\hat{W}_t\rho)) \geq \frac{\varepsilon}{3\sqrt{2}} - \frac{\varepsilon}{30} > 0.
\end{aligned}$$

We use this fact, together with Eq. (38) to lower bound the left hand side of Eq. (36):

$$\begin{aligned}
&2\sqrt{nT} \geq \sum_{t=1}^{T} \mathrm{sgn}(y_t - \hat{z}_t)\left(y_t - \mathrm{Tr}(\hat{W}_t\rho)\right) = \sum_{t=1}^{T}\left|y_t - \mathrm{Tr}(\hat{W}_t\rho)\right| \geq \frac{\varepsilon}{5}T,\\
\Leftrightarrow\; &T \leq 100n/\varepsilon^2.
\end{aligned}$$

This however is a contradiction to the way T was chosen beforehand, i.e. the algorithm never reaches the last step. Let us finally estimate the probability that this algorithm fails. The only place where it can lead to erroneous results is when estimating the z_t in stap 3 a). The proability that all of these estimates are correct is $p \geq (1 - 0.01/|T|)^{|T|} \geq 1 - 0.01$. Hence it follows, that the algorithm outputs a correct mimicking state σ in 99 out of 100 runs. □

C Remarks on Numerical Simulations

To quantify the sampling requirements of the proposed certification protocol, we performed numerical simulations for systems of up to 10 qutrits. For each system size, we determine the minimal number of samples $N_{\min}$ that guarantees a success probability $p_{\mathrm{succ}} \geq 0.7$. The success probability measures how often the protocol correctly identifies all tested expectation values $\mathrm{tr}(W\rho)$ within some tolerance.

C.1 Wilson confidence strategy

To decide whether a given number of samples is sufficient, we employ the Wilson confidence-interval strategy for binomial success probabilities. Given t repetitions of the reconstruction experiment at fixed sample size N with s observed successes, the estimated probability is $\hat{p} = s/t$. The Wilson interval [39] provides an adjusted confidence bound that remains reliable even for small t,

$$\Pr\left[p \in (w_+, w_-)\right] = 1 - \alpha \quad \text{with } w_\pm = \frac{n\hat{p} + \frac{z^2}{2}}{t + z^2} \pm \frac{z\sqrt{t}}{t + z^2}\sqrt{\hat{p}(1-\hat{p}) + \frac{z^2}{4t}}, \tag{39}$$

where z is the standard normal quantile corresponding to the desired confidence level $1 - \alpha = 90\%$. The lower bound of this interval serves as a conservative estimate of the true success probability.

For a target success probability $p_\star = 0.7$ we certify that the true success rate at sample size N satisfies $p_N \geq p_\star$ whenever $w_- \geq p_\star$. At each fixed N we apply the following stopping rule:

$$\text{Decision} = \begin{cases} \textbf{Accept}, & \text{if } w_- \geq p_\star, \\ \textbf{Early-reject}, & \text{if } w_+ < p_\star, \\ \text{Continue sampling (increase } t\text{) up to } t_{\max}, & \text{otherwise.} \end{cases} \tag{40}$$

If t reaches $t_{\max}$ while $w_- < p_\star \leq w_+$ (undecided), we conservatively treat this N as a reject.

Across sample sizes, we perform an exponential search: starting at N_0, increase $N \leftarrow \max\{N+1, \lfloor gN \rfloor\}$ (with growth factor $g > 1$) whenever the decision at the current N is reject or undecided, until we encounter the first accepted N (call it N_{hit}). If a last failing point N_{fail} is known, we optionally refine via bisection on the integer interval $(N_{\text{fail}}, N_{\text{hit}}]$ using the same Wilson decision at each midpoint, returning the minimal accepted N. This value is reported as $N_{\min}$.

C.2 Three-copy strategies

In the three-copy setting we estimate, for an n-qutrit state ρ, the expectations $\langle W \rangle = \operatorname{tr}(W\rho)$ of Weyl strings $W \in \mathcal{W}^{\otimes n}$, where $\mathcal{W}$ denotes the single-qutrit Weyl–Heisenberg set (of size $|\mathcal{W}| = d^2 = 9$). All $|\mathcal{W}|^n = 9^n$ strings are verified simultaneously from the same 3-copy data. A reconstruction experiment counts as success if for every $W \in \mathcal{W}^{\otimes n}$ the reconstructed value $\hat{y}$ sufficiently matches the theoretical target. That is $\max_W |\operatorname{Tr}(W^{\otimes 3}\rho^{\otimes 3}) - \hat{y}| < \delta$ where we used $\delta = 0.1$ in our experiments. At fixed sample size N, the accept/reject decision uses the Wilson rule of Sec. C.1 to certify $p_{\text{succ}} \geq 0.7$; the minimal accepted N is reported as $N_{\min}$.

C.3 Single-copy strategies

In the single-copy baseline we fix a hidden Weyl string $W^\star \in \mathcal{W}_n$ (which determines ρ) and draw measurement strings W uniformly at random from $\mathcal{W}_n = \{ X^{a_1}Z^{b_1} \otimes \cdots \otimes X^{a_n}Z^{b_n} : a_j, b_j \in \{1,2\} \}$, so that $|\mathcal{W}_n| = 4^n$ (each site uses one of the four non-identity pairs $(a,b) \in \{1,2\}^2$). A draw is a success iff $W = W^\star$ or $W = W^{\star\dagger}$. Hence, a single draw has hit probability $p_{\text{hit}} = 2/4^{-n}$ and, with N independent draws, the probability of at least one hit is:

$$p(N) = 1 - (1 - 2/4^{-n})^N. \tag{41}$$

For each fixed N, we repeat $t = 2000$ times and record the number of successful retrievals. The empirical success probability is $\hat{p}(N) = s/t$. We iterate over a fine grid of N and report

$$N_{\min}^{\text{emp}} = \min\{N : \hat{p}(N) \geq 0.7\}. \tag{42}$$

C.4 Simulation summary

Figure 1 summarizes the scaling of the minimal sample number $N_{\min}$ with the number of qutrits. The five curves correspond to the strategies described above: three-copy full check, three-copy random check (Wilson), three-copy random check (empirical), one-copy empirical, and one-copy theoretical. The data show that the three-copy protocol dramatically reduces the sampling complexity compared to the single-copy baseline, to achieve the prescribed success probability.

D Technical Lemmata

Lemma D.1 (Hoeffding). *Let $X_1, \ldots, X_N$ be independent real-valued random variables, not necessarily identically distributed, such that $a_i \leq X_i \leq b_i$ for suitable a_i, b_i almost surely. Then, for any $c \geq 0$ it holds that*

$$\Pr\left[\frac{1}{N}\left|\sum_{i=1}^{N} X_i - \mathbb{E}[X_i]\right| \geq c\right] \leq 2\exp\left(\frac{2c^2N^2}{\sum_i (b_i - a_i)^2}\right). \tag{43}$$

Lemma D.2. *Let d be a prime, ω a primitive $d-th$ root of unity, and $\mathbf{a}, \mathbf{b} \neq 0$ be vectors on $\mathbb{Z}_d^m$ and assume that $a_1 = 1$. Then*

$$\sum_{\langle \mathbf{a}, I\rangle = 0} \omega^{\langle \mathbf{b}, I\rangle} = d^{m-1}\delta_{b_1\mathbf{a}=\mathbf{b}}. \tag{44}$$

Proof. First of all, the condition $\langle a, I\rangle = 0$ translates to $i_1 = -\sum_{k\geq 2} i_k a_k$. Therefore, we express the summation as

$$\sum_{i_2,\ldots,i_m} \omega^{-b_1\sum_{k\geq 2} a_k i_k} \prod_{k\geq 2} \omega^{b_k i_k} = \sum_{i_2,\ldots,i_m} \prod_{k\geq 2} \omega^{i_k(b_k - b_1 a_k)} = d^{m-1} \prod_{k\geq 2} \delta_{b_k = b_1 a} = d^{m-1}\delta_{b_1\mathbf{a}=\mathbf{b}},$$

where we used that the condition $b_1 a_1 = b_1$ is trivially satisfied by construction in the last step. □

E Quantum Circuits

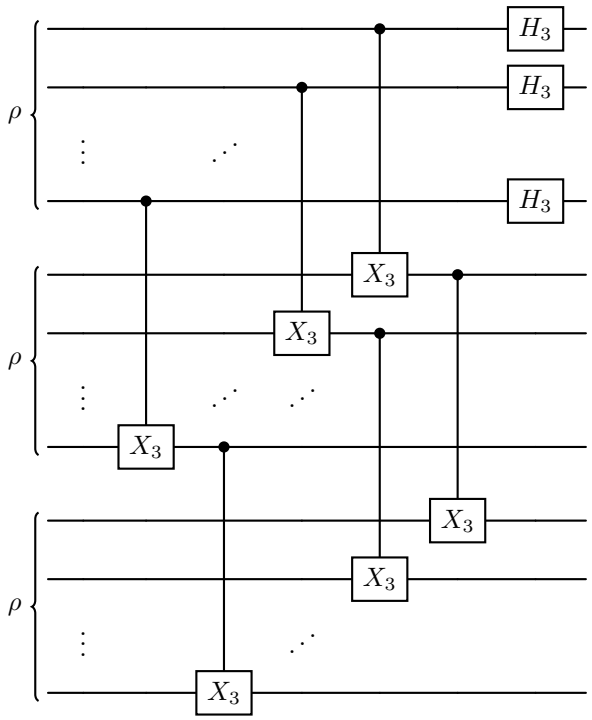
In this section we provide examples of quantum circuits to implement the measurement processes in the main part.

E.1 Quantum circuits for generalized Bell-basis measurements

In order to show, that the quantum circuits required to execute the generalized Bell-basis measurements are efficient, we sketch their structure, which strongly resembles the circuits for the preparation of ordinary Bell states in the circuits below. Specifically, for $n = 1$ we have that the diagonalizing unitary can be decomposed into single and two-qubit qudit Cliffords as

$$C = H^{(1)} \circ CX^{d-1,d} \circ CX^{d-2,d-1} \circ \cdots \circ CX^{1,2}. \tag{45}$$

For general n, the diagonalizing unitary is then simply $C^{\otimes n}$, which is schematically depicted in the circuit below for the example of qu*trits*.

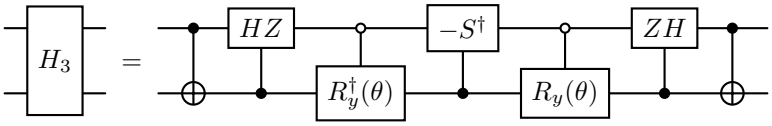


(46)

Here, the notation of the ordinary controlled X gate refers to the two-qudit unitary defined in Eq. (22) and H_3 the Hadamard/QFT as in Eq. (20).

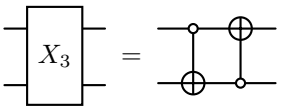
E.2 Transpilation to qubit architectures for the special case $d = 3$

Clearly, in order to faithfully represent qutrit gates, it suffices to embed one qutrit into two qubits. Here, we have to ensure, that the representation is reducible (into a trivial one-dimensional component and a three-dimensional irreducible one). Specifically, we choose the embedding via the binary representation, i.e. $|0\rangle \mapsto |00\rangle, |1\rangle \mapsto |01\rangle, |2\rangle \mapsto |10\rangle$. First, we consider the qutrit Hadamard-gate H_3. The eigenvalues of the Blockmatrix $\begin{pmatrix} H_3 & 0 \\ 0 & 1 \end{pmatrix}$ are given by $\{1, i, -1, 1\}$ with the corresponding (unnormalized) eigenvectors $\{(1-\sqrt{3}, 1, 1, 0), (0, 1, -1, 0), (1+\sqrt{3}, 1, 1, 0), (0, 0, 0, 1)\}$. Alltogether, we can therefore can express it as



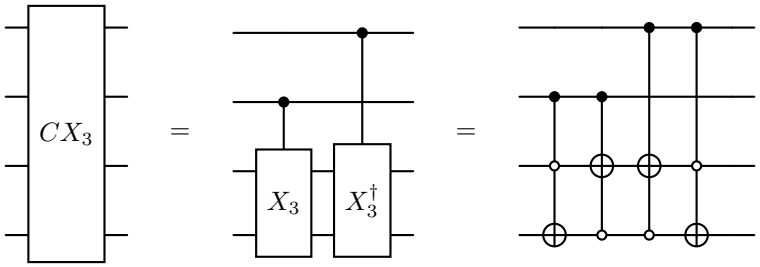
(47)

where $\theta = \arccos(1/\sqrt{3})$. Let us continue with a suitable implementation of the X_3-gate before moving on to its controlled version. It is straightforward to verify the following circuit identity

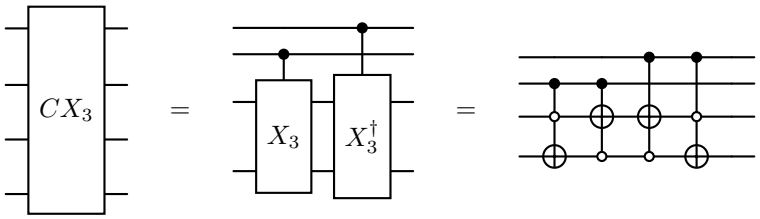


(48)

Since we are working with ternary logic, we know that $-k = 2k$, and we can realize the controlled-X operation conveniently as follows:



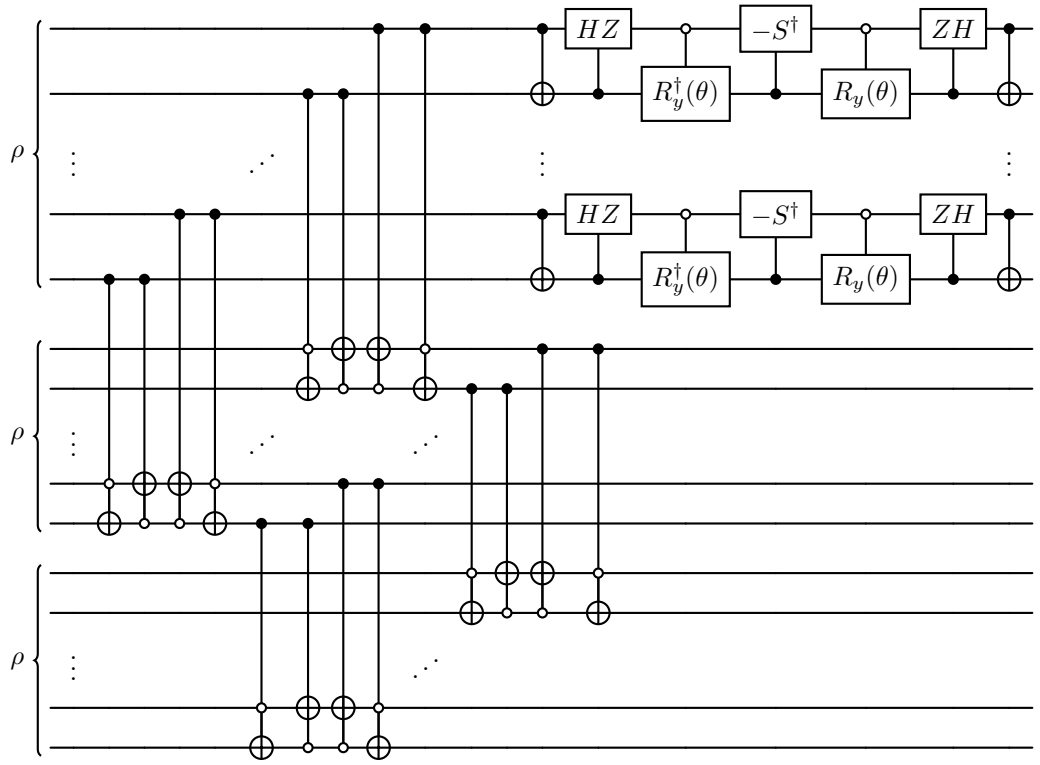
(49)



(50)

In particular, applying these transpilation rules to the circuit Eq. (46) we find the following measurement circuit for the d-copy efficient protocol on an n-qutrit state, where we omit the

computational basis measurement sign at the end of each line.



(51)

References

[1] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1(1):149, 2010.

[2] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *CoRR*, abs/1410.6913, 2014.

[3] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 913–925, 2016.

[4] Ryan O'Donnell and John Wright. Efficient quantum tomography. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 899–912. ACM, 2016.

[5] M Guţă, J Kahn, R Kueng, and J A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, apr 2020.

[6] Angus Lowe and Ashwin Nayak. Lower Bounds for Learning Quantum States with Single-Copy Measurements. *ACM Transactions on Computation Theory*, 17(1):1–42, March 2025.

[7] Scott Aaronson. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368–STOC18–394, 2020.

[8] Costin Badescu and Ryan O'Donnell. Improved quantum data analysis. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1398–1411. ACM, 2021.

[9] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020.

[10] Joshua Morris and Borivoje Dakić. Selective Quantum State Tomography, June 2020. arXiv:1909.05880 [quant-ph].

[11] Marco Paini and Amir Kalev. An approximate description of quantum states, November 2019. arXiv:1910.10543 [quant-ph].
[12] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, 2023.
[13] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Phys. Rev. Lett.*, 127:030503, Jul 2021.
[14] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19), May 2021.
[15] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, June 2022.
[16] Bálint Koczor. The dominant eigenvector of a noisy quantum state. *New Journal of Physics*, 23(12):123047, dec 2021.
[17] Zheng-Hao Liu, Romain Brunel, Emil E. B. Østergaard, Oscar Cordero, Senrui Chen, Yat Wong, Jens A. H. Nielsen, Axel B. Bregnsbo, Sisi Zhou, Hsin-Yuan Huang, Changhun Oh, Liang Jiang, John Preskill, Jonas S. Neergaard-Nielsen, and Ulrik L. Andersen. Quantum learning advantage on a scalable photonic platform. *Science*, 389(6767):1332–1335, 2025.
[18] Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. Tight bounds on pauli channel learning without entanglement. *Physical Review Letters*, 132(18), May 2024.
[19] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for pauli channel estimation. *Phys. Rev. A*, 105:032435, Mar 2022.
[20] Sitan Chen and Weiyuan Gong. Efficient pauli channel estimation with logarithmic quantum memory. *PRX Quantum*, 6:020323, May 2025.
[21] Sitan Chen, Weiyuan Gong, and Qi Ye. *Optimal Tradeoffs for Estimating Pauli Observables* . IEEE Computer Society, Los Alamitos, CA, USA, October 2024.
[22] Weiyuan Gong, Jonas Haferkamp, Qi Ye, and Zhihan Zhang. On the sample complexity of purity and inner product estimation, October 2024. arXiv:2410.12712 [quant-ph].
[23] H. Nakazato, T. Tanaka, K. Yuasa, G. Florio, and S. Pascazio. Measurement scheme for purity based on two two-body gates. *Physical Review A*, 85(4), April 2012.
[24] Sitan Chen, Jerry Li, Brice Huang, and Allen Liu. Tight Bounds for Quantum State Certification with Incoherent Measurements, October 2022.
[25] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential Separations Between Learning With and Without Quantum Memory . In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585, Los Alamitos, CA, USA, February 2022. IEEE Computer Society.
[26] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. A Hierarchy for Replica Quantum Advantage, December 2021. arXiv:2111.05874 [quant-ph].
[27] Huaning Liu and Wenpeng Zhang. On the squarefree and squarefull numbers. *Journal of Mathematics of Kyoto University*, 45(2):247 – 255, 2005.
[28] Qi Ye, Zhenhuan Liu, and Dong-Ling Deng. Exponential Advantage from One More Replica in Estimating Nonlinear Properties of Quantum States, September 2025. arXiv:2509.24000 [quant-ph].
[29] Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 914–946. SIAM, 2025.
[30] Chengsi Mao, Changhao Yi, and Huangjun Zhu. Qudit shadow estimation based on the clifford group and the power of a single magic gate. *Phys. Rev. Lett.*, 134:160801, Apr 2025.
[31] Jadwiga Wilkens, Kristina Kirova, Peter Tirler, Michael Meth, Martin Ringbauer, and Richard Kueng. Classical shadows for multi-qudit systems: theory and experiment. *in preparation*, 2025.
[32] William K Wootters and Brian D Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989.
[33] Robbie King, Kianna Wan, and Jarrod R. McClean. Exponential learning advantages with conjugate states and minimal quantum memory. *PRX Quantum*, 5:040301, Oct 2024.

[34] Marc-Olivier Renou, David Trillo, Mirjam Weilenmann, Thinh P. Le, Armin Tavakoli, Nicolas Gisin, Antonio Acín, and Miguel Navascués. Quantum theory based on real numbers can be experimentally falsified. *Nature*, 600(7890):625–629, December 2021.

[35] Viet T. Tran and Richard Kueng. One, Two, Three: One empirical evaluation of a two-copy shadow tomography scheme with triple efficiency, August 2025. arXiv:2508.11744 [quant-ph].

[36] Martin Ringbauer, Michael Meth, Lukas Postler, Roman Stricker, Rainer Blatt, Philipp Schindler, and Thomas Monz. A universal qudit quantum processor with trapped ions. *Nature Physics*, 18(9):1053–1057, Sep 2022.

[37] Qing Liu, Zihao Li, Xiao Yuan, Huangjun Zhu, and You Zhou. Auxiliary-free replica shadow estimation, July 2024. arXiv:2407.20865 [quant-ph].

[38] Mahnaz Jafarzadeh, Ya-Dong Wu, Yuval R Sanders, and Barry C Sanders. Randomized benchmarking for qudit clifford gates. *New Journal of Physics*, 22(6):063014, June 2020.

[39] Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–133, May 2001. Publisher: Institute of Mathematical Statistics.