Efficient Distributed Inner-Product Estimation via Pauli Sampling

M. Hinsche,[1,*,†] M. Ioannou,[1,*,‡] S. Jerbi,[1] L. Leone,[1] J. Eisert,[1,2] and J. Carrasco[1,*,§]

[1] *Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

[2] *Helmholtz-Zentrum Berlin für Materialien und Energie, 12489 Berlin, Germany*

(Received 4 August 2024; revised 15 April 2025; accepted 7 July 2025; published 18 September 2025)

Cross-platform verification is the task of comparing the output states produced by different physical platforms using solely local quantum operations and classical communication. While protocols have previously been suggested for this task, their exponential sample complexity renders them impractical even for intermediate-scale quantum systems. In this work, we propose a novel protocol for this task based on *Pauli sampling*, a subroutine that generates Paulis distributed according to their weight in the expansion of a quantum state in the Pauli basis. We show that our protocols for both Pauli sampling and cross-platform verification are efficient for pure states with low magic and entanglement [i.e., of the order $O(\log n)$]. Conversely, we show superpolynomial lower bounds on the complexity of both tasks for states with $\omega(\log(n))$ magic and entanglement. Interestingly, when considering states with real amplitudes, the requirements of our protocol for cross-platform verification can be significantly weakened.

DOI: 10.1103/g53f-z8cr

I. INTRODUCTION

Numerous institutions are currently engaged in the development of small and intermediate-sized quantum computers, each utilizing diverse physical platforms. The benchmarking of these devices is imperative for technological progress [1]. However, as we advance toward ever-larger-sized devices, their individual benchmarking based on certification methods such as tomography [2] and fidelity estimation [3,4] becomes increasingly impractical due to the exponential, or at least infeasible, scaling of the resources required. In some instances, a classical simulation of the quantum device is within reach but also here the scaling of resources is generically inhibitive. An alternative strategy involves *directly comparing* different quantum devices, bypassing classical benchmarks altogether. This approach gives rise to a central task known as *cross-platform verification*, which aims at estimating the fidelity of states prepared on two quantum computers in distinct laboratories, potentially operating on different physical platforms.

Establishing a high-quality quantum channel between distant laboratories operating on different physical platforms is currently beyond technological reach. This limitation emphasizes the importance of cross-platform verification protocols that do not rely on quantum communication between laboratories [5–9]. The first protocol in this vein has been proposed in Ref. [5] and has been experimentally implemented in Ref. [10]. It is based on randomized measurements coordinated between the two parties, i.e., the two parties agree on the measurement bases before data collection. Despite being less expensive than individual state tomography, this protocol is not scalable even to intermediate-sized systems, as the required number of state copies grows exponentially with the system size n. Recently, this scalability challenge has been further highlighted in Ref. [6] which has studied the task at the core of cross-platform verification: estimating the overlap $\mathrm{tr}(\rho\sigma)$ for quantum states ρ and σ prepared in separate experimental platforms without quantum communication. The authors refer to this task as distributed quantum inner-product estimation and show a $\Omega(\sqrt{d})$ lower bound on the required number of copies of ρ and σ, indicating unavoidable exponential scaling in general. Conversely, efficient protocols for the task require further assumptions on the input states ρ and σ.

This motivates the main question that this work addresses:

> *Which assumptions on ρ and σ allow for an efficient distributed estimation of $\mathrm{tr}(\rho\sigma)$?*

In this work, we approach this question from different angles. On the one hand, we propose new algorithms

*These authors contributed equally to this work.
†Contact author: m.hinsche@fu-berlin.de
‡Contact author: marios.ionannou@fu-berlin.de
§Contact author: jose.carrasco@fu-berlin.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

for the distributed estimation of $\mathrm{tr}(\rho\sigma)$ and demonstrate regimes in which these algorithms run efficiently. On the other hand, we show that the task remains hard even if we severely restrict the class of input states. Interestingly, both our hardness and easiness results connect to commonly studied properties of states, namely, their magic and their multipartite entanglement.

At the core of our proposed algorithms is a subroutine that we refer to as *Pauli sampling*. This subroutine operates on copies of an unknown state ρ and aims to approximately sample Pauli strings with probability proportional to their weight $\alpha_\rho(P)^2$ in the decomposition $\rho = 2^{-n}\sum_P \alpha_\rho(P)P$. We construct an algorithm for this subroutine by combining tools from two distinct fields. First, our algorithm incorporates Bell measurements across $\rho^{\otimes 2}$, a measurement scheme that has attracted recent attention from a quantum learning viewpoint due to its rich information content [11–17]. Second, the algorithm processes these data classically, drawing on concepts from the field of classical simulation of quantum circuits [18–20]. Our Pauli-sampling subroutine has applications beyond inner-product estimation; e.g., in the context of learning states prepared from Clifford and a few T gates [15,21,22].

A. Distributed inner-product estimation

We start by specifying the task studied in this work. We largely follow the formulation of Ref. [6] and consider the task of *distributed quantum inner-product estimation* (IP). This task involves two parties, Alice and Bob. A sketch of the setting is provided in Fig. 1.

Definition 1 (Inner-product estimation, IP). Alice is given k copies of an unknown state ρ and Bob is given k' copies of an unknown state σ. Their goal is to estimate the overlap $\mathrm{tr}(\rho\sigma)$ up to some desired additive error $\epsilon \in (0,1)$ with success probability at least $2/3$ using

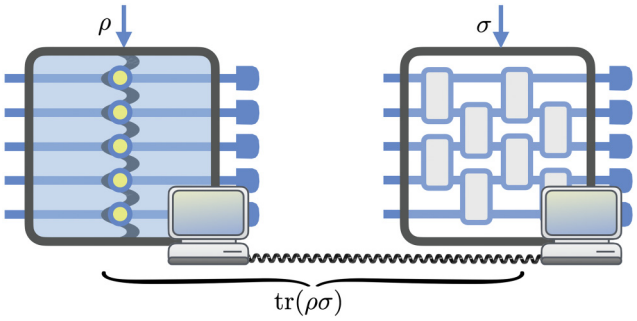


FIG. 1. A sketch of the setting of distributed inner-product estimation. Alice and Bob each have access to many copies of their respective states ρ and σ. Their goal is to obtain an estimate of $\mathrm{tr}(\rho\sigma)$ based on local quantum operations and classical communication, commonly abbreviated as LOCC. The two implementations may but do not have to be realized based on the same kind of physical architecture.

only *local quantum operations and classical communication* (LOCC). They are not allowed to have any quantum communication.

First, note that this task is interesting only in the distributed setting. If Alice and Bob could use quantum communication, then they could run a SWAP test to efficiently estimate the overlap $\mathrm{tr}(\rho\sigma)$. In the distributed setting with no quantum communication, a reasonable strategy is for Alice and Bob to coordinate the bases in which they measure their individual states. In this way, they can correlate the classical data that they collect. In Ref. [5], the authors have proposed a protocol based on such coordinated measurements. In their protocol, the measurement bases are selected by drawing local random unitaries. More recently, Anshu *et al.* [6] have proved that, in general, at least $\Omega(\sqrt{d})$ copies of the states ρ and σ acting on $\mathbb{C}^d$ are necessary to solve IP. The authors have further presented a variant of the approach laid out in Ref. [5], using coordinated measurement bases selected via drawing fully Haar-random unitaries. In both approaches, the random measurement bases are not tailored to the states ρ and σ. In our approach, we will instead rely on Pauli measurements that we tailor to the input states. In light of the lower bound proven by Ref. [6], this approach can only improve over fully random measurements under certain restrictions on the input states.

How can we select Pauli measurements that are tailored to the input states ρ and σ? This question motivates the study of the task that we call *Pauli sampling*.

B. Pauli sampling

Consider a state ρ expanded in the n-qubit Pauli basis as

$$\rho = \frac{1}{2^n}\sum_P \alpha_\rho(P)P. \tag{1}$$

When correctly normalized, the squared coefficients $\alpha_\rho(P)^2 = \mathrm{tr}(\rho P)^2$ form a probability distribution p_ρ over the set $\{I,X,Y,Z\}^n$, i.e., set of phaseless n-qubit Pauli strings. We refer to this distribution p_ρ as the *Pauli distribution*. We study in detail the following task that forms the core subroutine of our protocols for IP.

Definition 2 (Pauli sampling). Given access to copies of an unknown state ρ, sample from the *Pauli distribution*

$$p_\rho(P) = \frac{1}{2^n}\frac{\mathrm{tr}(\rho P)^2}{\mathrm{tr}\left(\rho^2\right)}. \tag{2}$$

We are also content with sampling from a distribution that approximates p_ρ up to error Δ in total-variation (TV) distance. Importantly, Pauli sampling generates Pauli strings with probability proportional to their weight in the

decomposition of the state [see Eq. (1)]. Thus, it is well suited to tailoring Pauli measurements to the input states ρ and σ in inner-product estimation.

In general, no efficient quantum algorithm exists to directly sample from p_ρ, given access to an unknown state ρ. Instead, one must measure copies of ρ to obtain classical data and then postprocess it to generate samples. However, if ρ is a stabilizer state or a real state (with real computational basis amplitudes), one can sample directly from p_ρ using Bell sampling or Bell difference sampling [11,12].

The Pauli distribution p_ρ of a state ρ is intimately connected to its nonstabilizerness, also known as the magic of the state ρ. In particular, recently, stabilizer Rényi entropies have been introduced as a measure of magic [23]. For pure states, they are defined as the Rényi entropies of the Pauli distribution p_ρ (up to a normalization). Interestingly, in Ref. [24], the complexity of *direct fidelity estimation* (DFE) has been found to scale exponentially with the stabilizer entropy of the target state. As explained in detail in Appendix C 2, we find a similar connection between the complexity of distributed inner-product estimation and the stabilizer entropy of the input states.

Lastly, we note that Pauli sampling has appeared in a different context in several works [25–27] trying to quantify nonstabilizerness in tensor-network states such as matrix product states (MPSs). However, these papers rely on a classical description of ρ in the form of a tensor network, while we work in the scenario in which the state ρ is not yet characterized.

II. MAIN RESULTS

In this section, we discuss our novel algorithms for distributed inner-product estimation and Pauli sampling as well as our complementary no-go results. These results connect to two key characteristics of input states ρ and σ: their magic (or nonstabilizerness) and their entanglement. We quantify magic through a magic measure $M(\rho)$ such as the stabilizer entropies [23] (for more details, see Appendix A 3) and we quantify entanglement in terms of the α-Rényi entropies of entanglement $\mathcal{S}_\alpha(\rho_{A:B})$.

A. Pauli sampling

The first main contribution of this work is the study of Pauli sampling and its resource requirements. More concretely, our focus is on characterizing the sample complexity (the required number of copies of the unknown input state ρ) and computational complexity across classes of input states. Note that our findings predominantly apply to classes of pure states.

1. No-go

We now state our first main result, a no-go theorem for efficient Pauli sampling.

Theorem 1 (Approximate Pauli sampling no-go—informal version of Theorem B1). There is no sample-efficient algorithm for approximate Pauli sampling with respect to the class of all pure quantum states. Moreover, let $\mathcal{C}$ be the class of n-qubit pure states ρ such that $M(\rho) < f(n)$ and $\mathcal{S}_\alpha(\rho_{A:B}) < g(n)$ for $0 \leq \alpha \leq \infty$ and for all bipartitions $(A:B)$, where $f(n), g(n) = \omega(\log n)$. Then, there is no sample-efficient algorithm for approximate Pauli sampling up to any constant TV distance $\Delta \leq 1/3$ with respect to $\mathcal{C}$.

Intuitively, Theorem 1 reveals that Pauli sampling is challenging not only for general states but also remains difficult when we restrict the class of input states to states that have magic and entanglement only slightly larger than $\log(n)$.

We obtain this result by connecting to so-called *pseudorandom states* (PRSs). This concept and the first constructions have been introduced in Ref. [28]. More recently, PRS constructions that feature tunable entanglement [29] and tunable magic [30] have been discovered. Our no-go results are based on these tunable PRS constructions. Concretely, we obtain Theorem 1 by demonstrating a reduction between Pauli sampling and imaginarity testing, i.e., testing if a state is either real or highly imaginary. We then make use of a proof strategy from Ref. [31], which has shown that imaginarity testing requires at least $\Omega\left(2^{n/2}\right)$ copies based on an ensemble of real pseudorandom states. We apply this strategy to the tunable PRS constructions mentioned above.

2. Algorithm

Turning to positive results, in Appendix B 2, we propose an algorithm for approximate Pauli sampling. To summarize, our algorithm applies the ancestral-sampling method—a standard tool in classical simulation of quantum computation [18–20]—to the Pauli distribution p_ρ. In ancestral sampling, a distribution p over $\{0,1\}^n$ is decomposed via the chain rule,

$$p(x) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_1,\ldots,x_{n-1}), \tag{3}$$

and a sample is generated bit by bit by sampling from the corresponding conditional or marginal distributions. Here, we represent n-qubit Pauli operators (from $\{I,X,Y,Z\}^n$) as bit strings in $\{0,1\}^{2n}$ and generate them qubit by qubit. Instead of computing the required marginals of p_ρ from a classical-state description of ρ, we estimate them from Bell-measurement data. Specifically, the marginals of the Pauli distribution p_ρ for a pure state ρ are given by

$$p_\rho(x_1,\ldots,x_k) = 2^{-k}\mathrm{tr}\left(P_{x_1}^{\otimes 2}\otimes\cdots\otimes P_{x_k}^{\otimes 2}\otimes\mathbb{S}^{\otimes n-k}\rho^{\otimes 2}\right), \tag{4}$$

where the observables—tensor products of two-copy single-qubit Paulis $P_i^{\otimes 2}$ and SWAP operators $\mathbb{S}$—are estimated from the Bell-measurement data.

We underscore the practicality of the algorithm by noting that it relies solely on Bell-measurement data and classical postprocessing. Our analysis (see Appendix B 4) builds a connection—flatness of the Pauli distribution reflects the magic of the state, while the size of its marginals corresponds to entanglement across cuts—and, through this connection, we are able to prove the following guarantee.

Theorem 2 (Efficient approximate Pauli sampling—informal version of Corollary B2). Let $\mathcal{C}$ be the class of n-qubit pure states ρ with $M(\rho) = O(\log(n))$ and $\mathcal{S}_0(\rho_{A:B}) = O(\log(n))$ for all bipartitions $(A:B)$. Then, our algorithm for approximate Pauli sampling with respect to $\mathcal{C}$ is both sample and computationally efficient.

Theorem 2 shows that approximate Pauli sampling via our algorithm is efficient when we limit both magic and entanglement sufficiently, namely, to at most $O(\log n)$. As explained in more detail in Appendix B 4, if we only require sample efficiency, our algorithm can cover an extended class of states for which bounded entanglement is only needed for a subset of bipartitions instead of all bipartitions.

We further emphasize that our algorithm can be applied to any state, including mixed states. Though Theorem 2 applies only to a specific set of pure states, the most general result that we obtain (see Theorem B3) characterizes the complexity of Pauli sampling with respect to properties of the underlying Pauli distribution. This characterization also holds for mixed states. However, in the mixed-state case, we are not yet able to connect the relevant properties of the Pauli distribution to mixed-state entanglement measures [32] and magic measures [33,34].

B. Distributed inner-product estimation

As the second main contribution of this work, we study which restricted classes of states allow for efficient distributed inner-product estimation, i.e., estimation of $\mathrm{tr}(\rho\sigma)$. Similarly to our results for Pauli sampling above, we again provide both positive and negative results.

1. No-go

We start by presenting a no-go theorem. Note that a first no-go has been demonstrated in Ref. [6], namely, that there is no sample-efficient algorithm for IP with respect to the class of all pure states. Here, we use the proof strategy from Ref. [6] and combine it with the tunable PRS constructions [29,30] also used to prove Theorem 1, in order to show that this no-go holds even for a much more restricted class of states. Note that the IP problem has two input states, ρ and σ. In the following informal-theorem statements, when referring to a single class of states $\mathcal{C}$, we will always assume that both ρ and σ belong to that class $\mathcal{C}$.

Theorem 3 (IP no-go—informal version of Theorem C3). Let $\mathcal{C}$ be the class of n-qubit pure states ρ such that $M(\rho) < f(n)$ and $\mathcal{S}_\alpha(\rho_{A:B}) < g(n)$ for $0 \leq \alpha \leq \infty$ and for all bipartitions $(A:B)$ where $f(n), g(n) = \omega(\log n)$. Then, there is no sample-efficient algorithm for IP up to error $\epsilon \in (0,1)$ with respect to $\mathcal{C}$.

2. Algorithm

Turning to positive results, in Appendix C 2, we propose two novel protocols for IP that employ coordinated Pauli-basis measurements selected via the Pauli-sampling subroutine. We refer to them as the *symmetric* and *asymmetric* protocols. In the symmetric protocol, both Alice and Bob perform Pauli sampling according to p_ρ and p_σ, respectively. In this way, they obtain a list of Pauli strings distributed according to $p_{\mathrm{mix}} = \frac{1}{2}\left(p_\rho + p_\sigma\right)$. Then, both parties measure the sampled Pauli operators on their respective states in order to estimate $\mathrm{tr}(\rho P), \mathrm{tr}(\sigma P)$. Lastly, by combining these estimates via classical communication, they can Monte Carlo estimate $\mathrm{tr}(\rho\sigma)$ via the rewriting

$$(1 + \mathrm{tr}(\rho\sigma)) = \mathop{\mathbb{E}}_{p_{\mathrm{mix}}} \left[G\left(\mathrm{tr}(\rho P), \mathrm{tr}(\sigma P)\right)\right], \tag{5}$$

where $G(x,y) := (x+y)^2/(x^2+y^2)$. The asymmetric protocol proceeds in a similar way; however, only one party performs Pauli sampling and the estimator is given by

$$\mathrm{tr}(\rho\sigma) = \mathop{\mathbb{E}}_{p_\rho} \left[\frac{\mathrm{tr}(\sigma P)}{\mathrm{tr}(\rho P)}\right]. \tag{6}$$

We give a detailed description of both protocols in Appendix C 2.

By analyzing the symmetric and asymmetric protocol, we find that their efficiency is linked to the magic of the input states. In particular, we are able to show the following guarantee.

Theorem 4 (Efficient IP via Pauli sampling—informal version of Corollary C2). Let $\mathcal{C}$ be the class of n-qubit pure states with $M(\rho) = O(\log(n))$. Then, assuming access to an oracle for approximate Pauli sampling up to total variation distance Δ, there exists an efficient algorithm for IP up to error $\epsilon + 2\Delta$ with respect to $\mathcal{C}$.

As remarked above, Theorem 4 assumes that both inputs ρ and σ to IP belong to the class $\mathcal{C}$. Furthermore, we point out that in Theorem 4, the total variation distance error Δ from the approximate Pauli-sampling subroutine

directly enters the total estimation error as $\epsilon + 2\Delta$. Here, ϵ is an additional contribution due to a standard finite sample error. Importantly, the IP algorithm has control over the value of Δ and can always suppress it by investing more resources.

Theorem 4 tells us that distributed inner-product estimation via Pauli sampling can be performed efficiently if we limit ourselves to states with magic of at most $O(\log n)$. Examples of low magic states are very common; they include eigenstates of so-called (perturbed) stabilizer Hamiltonians as shown in Ref. [35], such as, e.g., the toric code Hamiltonian. Concretely, it has been proven that these Hamiltonians admit eigenstates with bounded magic $M(\rho)$, making them interesting candidates for the application of our algorithm. By replacing the oracle access in Theorem 4 with concrete Pauli-sampling methods, we obtain guarantees on the complexity of IP for two classes of states. First, plugging in our proposed Pauli-sampling algorithm and its performance guarantee in Theorem 2, we obtain the following corollary.

Corollary 1 (Efficient IP for low magic and entanglement states—informal). Let $\mathcal{C}$ be the class of n-qubit pure states ρ with $M(\rho) = O(\log(n))$ and $\mathcal{S}_0(\rho_{A:B}) = O(\log(n))$ for all bipartitions $(A : B)$. Then, there exists an efficient algorithm for IP up to error ϵ with respect to $\mathcal{C}$.

Second, by using the fact that Bell sampling corresponds to Pauli sampling for real states [11] (states with real amplitudes with respect to the computational basis), we can get rid of the restriction of low entanglement and obtain the following corollary.

Corollary 2 (Efficient IP for real low magic states—informal). Let $\mathcal{C}$ be the class of real n-qubit pure states with $M(\rho) = O(\log(n))$. Then, there exists an efficient algorithm for IP up to error ϵ with respect to $\mathcal{C}$.

In particular, we note that the class of real states includes arbitrary eigenstates of any real Hamiltonian. Note that the overwhelming majority of many-body Hamiltonians considered in practice have real-valued coefficients with respect to the computational basis [36]. This includes, by construction, all stoquastic Hamiltonians [37] but, more importantly, also most well-known spin Hamiltonians. Further examples include Hamiltonians corresponding to Calderbank-Shor-Steane (CSS) codes, such as the toric code, and Rydberg-atom arrays [38]. Hence, this result applies to a broad class of states with a rich structure that has been studied in the literature.

Finally, we remark that our two protocols each have advantages that are unique to them. When using the *asymmetric* protocol, only one of the two input states ρ and σ has to be belong to the restricted class of states $\mathcal{C}$ with low magic and entanglement, whereas the other state is essentially unconstrained. On the other hand, for our symmetric protocol, the conditions on low magic and entanglement of the previous theorems can be weakened. Thus, to the best of our knowledge, the previous results provide a way to perform distributed inner-product estimation that goes beyond any existing methods.

III. NUMERICAL EXPERIMENTS

To support our theoretical findings, in this section, we present numerical results from the implementation of our symmetric and asymmetric protocols. For simplicity, we focus on the case in which $\rho = \sigma$ such that the true value of the inner product is known to be $\mathrm{tr}(\rho\sigma) = 1$. This choice not only makes it easier to assess the deviation of the protocol estimates from the true value but is also practically relevant for the case in which two devices prepare the same state.

Our choice of the input state ρ is closely linked to our main theoretical result, Theorem 4, which shows that the sample complexity of our protocols scales exponentially with the amount of magic in the input states (see also the formal statements in Corollaries C2 and C3). To validate this, we numerically study our protocols on an ensemble of n-qubit states with tunable magic. Specifically, we draw the state ρ randomly from the ensemble of t-doped states n-qubit state vectors given by

$$C(|T\rangle^{\otimes t} \otimes |0\rangle^{\otimes n-t}), \tag{7}$$

where C is a random global Clifford unitary and $0 \leq t \leq n$ and $|T\rangle = T|+\rangle$ is the standard single-qubit magic state. In particular, its Pauli decomposition is

$$|T\rangle\langle T| = \tfrac{1}{2}\left(I + (X+Y)/\sqrt{2}\right). \tag{8}$$

This ensemble provides an excellent benchmark because it is both generic and highly entangled due to the application of the random Clifford unitary. Moreover, since stabilizer entropy is additive under tensor products and invariant under Clifford operations, we can compute it analytically. In particular, from the Pauli decomposition of the $|T\rangle$ state, we can find that

$$M_0(|T\rangle) = \log_2(3) - 2 \approx 0.585, \tag{9}$$

$$M_1(|T\rangle) = \tfrac{1}{2} \tag{10}$$

and $M_0(|T\rangle^{\otimes t}) = t \times M_0(|T\rangle)$, and similarly for M_1. Hence the number t of T gates in the state preparation corresponds directly to the stabilizer entropy of the state ρ.

We now describe our numerical experiments. To evaluate the estimator in Eq. (5) for the symmetric protocol and the estimator Eq. (6) [for a more detailed version, see Eq. (C27)] for the asymmetric protocol, we have sampled $N_1 = 1000$ Pauli operators $P \in \{I, X, Y, Z\}^n$ from the Pauli

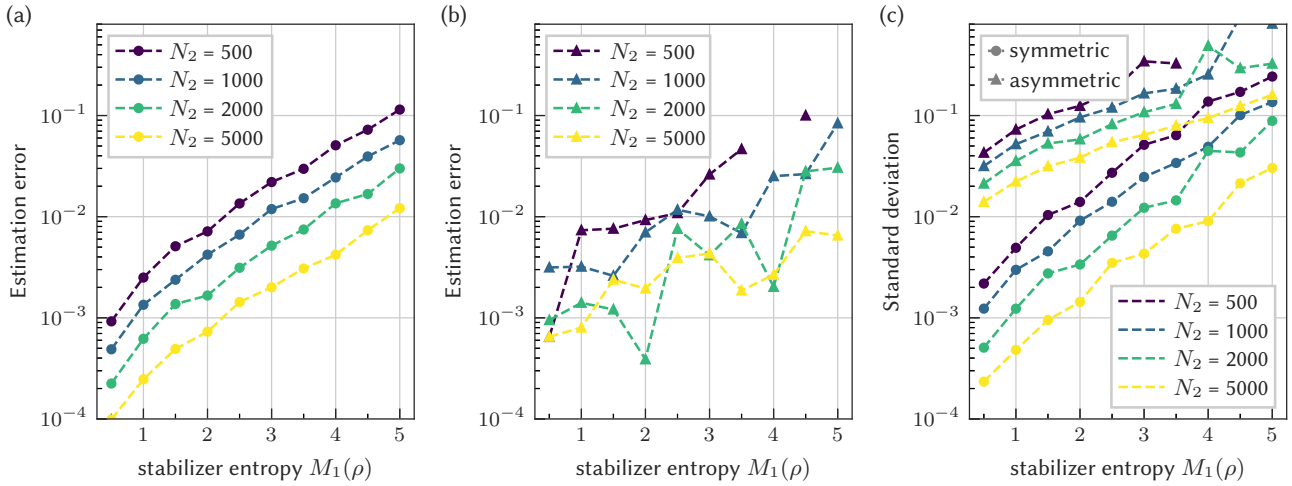


FIG. 2. The estimation error and standard deviation of the symmetric and asymmetric protocols as a function of the stabilizer entropy $M_1(\rho)$ and the number of shots per Pauli N_2. Here, ρ is drawn from the ensemble $C(|T\rangle^{\otimes t} \otimes |0\rangle^{\otimes n-t})$, with increasing t corresponding to increasing values of $M_1(\rho)$, the other parameters in the protocol are fixed to $N_1 = 1000$ Pauli samples, and the number of qubits is $n = 10$. (a) The estimation error of the symmetric protocol. (b) The estimation error of the asymmetric protocol. (c) The standard deviations of the estimators of the symmetric and asymmetric protocols, respectively.

distribution p_ρ and then estimated $\mathrm{tr}(\rho P)$ for each of the sampled Paulis P using N_2 state preparations of ρ and σ in the corresponding Pauli basis. The total number of shots is hence $2N_1N_2$.

In Fig. 2, we present the estimation error, i.e., the deviation of the estimate output by the protocol from the true value of $\mathrm{tr}(\rho\sigma)$, of the symmetric and asymmetric protocols as a function of the number of T gates (serving as a proxy for the amount of magic) and N_2. Alongside, we also compare the standard deviation of the estimators. Our numerical results align well with the theoretical analysis, confirming that the estimation error grows exponentially with the amount of magic but can be mitigated by increasing the number of shots in both protocols. Furthermore, when comparing the two approaches, the symmetric estimator demonstrates significantly lower standard deviation, indicating more stable behavior. This advantage arises from the fact that the symmetric estimator is inherently bounded, whereas the asymmetric estimator is not.

IV. COMPARISON TO LEARNING-BASED APPROACHES

The key challenge in cross-device verification is to determine whether two quantum devices prepare the same state without requiring direct quantum communication between them. One potential approach is to first learn a classical representation of the state produced by each device and then compute $\mathrm{tr}(\rho\sigma)$ based on these descriptions. However, in general, learning a full classical description scales exponentially with the number of qubits, making this approach infeasible. To mitigate this, methods such as MPS tomography [25,39] and learning-based techniques for stabilizer and t-doped stabilizer states [15, 21] provide efficient ways to infer compressed classical representations of quantum states for verification. However, these methods rely on stringent structural assumptions about the input states ρ and σ, which are often violated under realistic imperfect state preparations, limiting their practical applicability.

Our approach offers an alternative that bypasses the need for full state reconstruction. Instead of first approximating a classical description, we directly estimate $\mathrm{tr}(\rho\sigma)$ from measurement data, significantly reducing resource overhead. Since our method does not depend on any particular structural constraints, it is robustly applicable even when ρ and σ do not conform to specific assumptions, making it more resilient to experimental imperfections. In fact, our Pauli-based protocols enable efficient cross-device verification even for certain highly entangled, highly doped states, as demonstrated in Ref. [40]. In contrast, MPS-based and t-doped learning methods are prohibitively expensive in these regimes of large entanglement and doping via non-Clifford gates. Our findings establish our approach as a robust and scalable tool for quantum verification, significantly broadening its applicability beyond traditional learning-based techniques.

V. DISCUSSION AND FUTURE WORK

In this work, we propose algorithms for distributed inner-product estimation that employ Pauli basis measurements selected via the Pauli-sampling subroutine. In contrast to previous proposals [5,6], Pauli measurements are not only experimentally much more feasible but we also show how to tailor them to the input states in order to avoid previous no-go results. Our work thus opens up

the possibility of scalable distributed cross-platform verification on near-term devices. Furthermore, our results on Pauli sampling are relevant to other applications such as quantum learning tasks related to (t-doped)-stabilizer states [21] and experimental magic estimation [41]. Our results also reveal intimate connections to magic and entanglement, fundamental resources for quantum computation the interplay of which determines the performance of our Pauli-sampling algorithm.

Our work opens up several avenues for future research. For instance, one question is whether the relationship that we discovered between entanglement and magic—particularly in Lemmas A1–A3—can be extended to mixed states. Another open issue is finding an efficient method to identify "heavy" Pauli operators of a general state ρ the squared expectation value $\mathrm{tr}(\rho P)^2$ of which is nearly maximal. Although Bell-measurement data enable us to estimate all these values using only $O(n/\epsilon^2)$ samples, no efficient algorithm is known for selecting the high-value operators. Moreover, recent work [42,43] has examined the complexity of inner-product estimation in scenarios in which Alice and Bob share a limited amount of quantum information—a setting analogous to having a low-capacity quantum channel. Extending our protocol to this intermediate regime is another intriguing challenge. Finally, a promising direction is to investigate distributed inner-product estimation in other systems, such as qudits, or even bosonic and fermionic systems.

ACKNOWLEDGMENTS

We sincerely thank Matthias Caro and Tommaso Guaita for providing incredibly helpful feedback on the manuscript and for discussions. We thank David Gross, Gregory White, and Alexander Nietner for clarifying discussions on MPS-tomography schemes. We further thank Barbara Kraus and Janek Denzler for insightful discussions. This work has been supported by the Federal Ministry of Research, Technology and Space (BMFTR) (DAQC, MUNIQC-Atoms, QuSol, Hybrid++), the Munich Quantum Valley (K-4 and K-8), the Federal Ministry for Economic Affairs and Climate Action (BMWK) (EniQmA), the Quantum Flagship (PasQuans2, Millenion), QuantERA (HQCC), the Clusters of Excellence MATH+ and ML4Q, the German Research Foundation (DFG) (CRC183), the Einstein Foundation (Einstein Research Unit on Quantum Devices), Berlin Quantum and the European Research Council (ERC) (DebuQC).

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: PRELIMINARIES

We denote $[n] = \{1, \ldots, n\}$ and the TV distance between two probability distributions p and q by $\|p - q\|_{\mathrm{TV}}$. We denote drawing a sample x according to a distribution p by $x \sim p$. Further, we denote the probability weight of a distribution on a set S by $p(S) = \sum_{x \in S} p(x)$. Finally, we use log to denote the logarithm to base 2 and ln to denote the natural logarithm.

1. Bell basis and Bell sampling

The single-qubit Pauli matrices are denoted by $\{I, X, Y, Z\}$. Denoting with $|\Omega\rangle := \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$ a fiducial maximally entangled state vector, the two-qubit Bell basis is given by the four Bell state vectors $\left\{\left|\Phi^+\right\rangle, \left|\Phi^-\right\rangle, \left|\Psi^+\right\rangle, \left|\Psi^-\right\rangle\right\}$, which can expressed in terms of the single-qubit Pauli matrices as

$$2\left|\Phi^+\right\rangle = (I \otimes I)\,|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0,0\rangle + |1,1\rangle\right), \tag{A1}$$

$$\left|\Phi^-\right\rangle = (I \otimes Z)\,|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0,0\rangle - |1,1\rangle\right), \tag{A2}$$

$$\left|\Psi^+\right\rangle = (I \otimes X)\,|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0,1\rangle + |1,0\rangle\right), \tag{A3}$$

$$\left|\Psi^-\right\rangle = i\,(I \otimes Y)\,|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0,1\rangle - |1,0\rangle\right). \tag{A4}$$

The SWAP operator between two qubits can be expressed in terms of the single-qubit Pauli matrices as

$$\mathbb{S} = \frac{1}{2}\sum_{P \in \{I,X,Y,Z\}} P^{\otimes 2}. \tag{A5}$$

The four Bell states are eigenstates with eigenvalues in $\{-1, 1\}$ of $P \otimes P$ for any single-qubit Pauli matrix P as well as of the SWAP operator $\mathbb{S}$.

Let us now generalize the above definitions and observations to n qubits. We will label (phaseless) Pauli operators by bit strings of length $2n$ as follows. Let $x = (x_1, \ldots, x_n) = (v_1, w_1, \ldots, v_n, w_n) \in \{0,1\}^{2n}$, i.e., $x_i = (v_i, w_i)$. We then define

$$P_x = i^{v \cdot w}\left(X^{v_1} Z^{w_1}\right) \otimes \cdots \otimes \left(X^{v_n} Z^{w_n}\right) = i X^v Z^w, \tag{A6}$$

where the inner product $v \cdot w$ of the n-bit strings v and w on the phase in front of the tensor product is by convention meant to be an integer. Furthermore, we define the generalized $2n$-qubit canonical Bell state as follows. Let

$$\left|\Phi_n^+\right\rangle = 2^{-n/2}\sum_{v \in \{0,1\}^n} |v, v\rangle \tag{A7}$$

(by abuse of notation, we will often drop the subscript n). Then, the $2n$-qubit Bell basis is given by

$$\left\{|P_x\rangle := (P_x \otimes I)\left|\Phi^+\right\rangle \,|\, x \in \{0,1\}^{2n}\right\}. \tag{A8}$$

It is an orthonormal basis of $\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n}$. By virtue of the tensor-product structure, it is the eigenbasis of any $2n$-qubit Pauli string $P_x^{\otimes 2}$ as well as of the operator $\mathbb{S}^{\otimes n}$, as well as of strings containing both $P^{\otimes 2}$ and $\mathbb{S}$ factors, e.g., $X^{\otimes 2} \otimes \mathbb{S}^{\otimes n-1}$. In Appendix J, we comment on generalizations to other higher-dimensional systems.

By Bell sampling from a state ρ, we mean measuring $\rho^{\otimes 2}$ in the $2n$-qubit Bell basis, with the outcome of the measurement being a bit string x of length $2n$ that we can relate to the Pauli operator P_x. To be specific, to implement this measurement, one measures each qubit pair $(i, i+n)$ for $i \in [n]$ in the two-qubit Bell basis defined above.

2. Pauli distribution and its marginals

Next, we define the *Pauli distribution* of a quantum state ρ to be the distribution over $x \in \{0,1\}^{2n}$ with the probabilities given by

$$p_\rho(x) = \frac{1}{2^n} \frac{\operatorname{tr}(\rho P_x)^2}{\operatorname{tr}(\rho^2)}. \tag{A9}$$

In the following, we will often use the shorthand notation $\alpha_\rho(x) := \operatorname{tr}(\rho P_x)$. In the case of a pure state $\rho = |\psi\rangle\langle\psi|$, Eq. (A9) simplifies to

$$p_\psi(x) = \frac{1}{2^n} \operatorname{tr}(P_x |\psi\rangle\langle\psi|)^2 = \frac{\alpha_\psi(x)^2}{2^n}. \tag{A10}$$

Let us consider taking marginals of the Pauli distribution p_ρ. Let $A \subseteq [n]$ and let B be the complement of A in $[n]$. We can associate the subset A and its corresponding substring $x_A \in \{0,1\}^m$ of length $m = 2|A|$ with a marginal of the Pauli distribution as

$$\begin{aligned} p_\rho(x_A) &= \sum_{x_B} p_\rho(x) \\ &= \frac{1}{2^n \operatorname{tr}(\rho^2)} \sum_{x_B} \operatorname{tr}\left(P_{x_A}^{\otimes 2} \otimes P_{x_B}^{\otimes 2} \rho^{\otimes 2}\right). \end{aligned} \tag{A11}$$

Then, using $\operatorname{tr}(P_x\rho)^2 = \operatorname{tr}\left(P_x^{\otimes 2}\rho^{\otimes 2}\right)$, we can further express such marginals $p_\rho(x_A)$ in terms of the SWAP operator $\mathbb{S}$ as

$$p_\rho(x_A) = \frac{1}{2^{|A|}\operatorname{tr}(\rho^2)} \operatorname{tr}\left(P_{x_A}^{\otimes 2} \otimes \mathbb{S}_B \rho^{\otimes 2}\right). \tag{A12}$$

For instance, if $A = \{1, \ldots, k\}$, then the corresponding marginal is given by

$$\begin{aligned} p_\rho(x_A) = p_\rho(x_1, \ldots, x_k) &= \sum_{x_{k+1}} \cdots \sum_{x_n} p_\rho(x_1, \ldots, x_n) \\ &= \frac{1}{2^k \operatorname{tr}(\rho^2)} \operatorname{tr}\left(P_{x_1}^{\otimes 2} \otimes \cdots \otimes P_{x_k}^{\otimes 2} \otimes \mathbb{S}^{\otimes n-k} \rho^{\otimes 2}\right). \end{aligned} \tag{A13}$$

We note that in the case of a state vector $|\psi\rangle$, we can further rewrite Eq. (A12) in terms of the reduced state $\rho_A = \operatorname{tr}_B(|\psi\rangle\langle\psi|)$ of $|\psi\rangle\langle\psi|$ as

$$\begin{aligned} p_\psi(x_A) &= \frac{1}{2^{|A|}} \operatorname{tr}\left(|\psi\rangle\langle\psi| (P_A \otimes I_B) |\psi\rangle\langle\psi| (P_A \otimes I_B)\right) \\ &= \frac{1}{2^{|A|}} \operatorname{tr}(\rho_A P_A \rho_A P_A). \end{aligned} \tag{A14}$$

By Eqs. (A12) and (A14), it is clear that marginals $p(x_A)$ correspond to Pauli strings P_{x_A} of length $|A|$. Note that the x_A are bit strings of even length and thus we will in some instances refer to these as "even" marginals of p_ρ.

When viewing p_ρ as a distribution over $\{0,1\}^{2n}$, however, we also have "odd" marginals of p_ρ corresponding to substrings $x_m \in \{0,1\}^m$ for odd m. These do not directly correspond to Pauli strings. For instance, consider the marginal

$$\begin{aligned} p_\rho(x_1, \ldots, x_{k-1}, v_k) &= p_\rho((v_1, w_1), \ldots (v_{k-1}, w_{k-1}), v_k) \\ &= p_\rho(v_1, w_1, \ldots, v_k) \end{aligned} \tag{A15}$$

for some $k \in [n]$ [refer to Eq. (A6) for this notation]. This marginal is given by

$$\begin{aligned} p_\rho(v_1, w_1, \ldots, v_k) &= \sum_{w_k} \sum_{x_{k+1}} \cdots \sum_{x_n} \\ &\times p_\rho((v_1, w_1), \ldots, (v_{k-1}, w_{k-1}), (v_k, w_k), x_{k+1}, \ldots, x_n). \end{aligned} \tag{A16}$$

This odd marginal can be conveniently rewritten as the sum of two even marginals, as

$$\begin{aligned} p_\rho(v_1, w_1, \ldots, v_k) &= p_\rho(v_1, w_1, \ldots, v_k, w_k = 0) \\ &+ p_\rho(v_1, w_1, \ldots, v_k, w_k = 1). \end{aligned} \tag{A17}$$

3. Measures of magic and entanglement

In this section, we define the measures of magic, which quantity the nonstabilizerness of states, and measures of entanglement used in this work. The *Rényi entropy of a distribution* p over $\{0,1\}^n$ is defined as

$$H_\alpha(p) = \frac{1}{1-\alpha} \log\left(\sum_x p(x)^\alpha\right). \tag{A18}$$

The limiting value of $H_\alpha(p)$ for $\alpha \to 1$ is the Shannon entropy

$$H_1(p) = -\sum_x p(x) \log p(x). \tag{A19}$$

This definition of the classical Rényi entropy can be used to define both entanglement and magic measures. In particular, letting $(A:B)$ be a bipartition of the n qubits, the

Rényi entanglement entropies $\mathcal{S}_\alpha$ are defined as the Rényi entropy of the reduced density matrix

$$\mathcal{S}_\alpha(\rho_{A:B}) = \frac{1}{1-\alpha}\log\left(\mathrm{tr}\rho_A^\alpha\right) = H_\alpha(\vec{\lambda}), \tag{A20}$$

where $\rho_A = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$. The Rényi entanglement entropies measure bipartite pure-state entanglement over the cut $(A:B)$. Another important way to quantify entanglement in pure states is through the Schmidt rank. Any state vector $|\psi\rangle$ can be written in terms of the Schmidt decomposition as

$$|\psi\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}\,|i_A\rangle\,|i_B\rangle\,, \tag{A21}$$

where r is the known as the Schmidt rank of $|\psi\rangle$ with respect to the cut (A,B). For pure states, the Schmidt rank is related to the *Rényi-0 entanglement entropy* via

$$\mathcal{S}_0(\rho_{A:B}) = \log(r). \tag{A22}$$

The *stabilizer (Rényi) entropies* (SEs) [23] are (up to normalization) defined as the Rényi entropy of the corresponding Pauli distribution. They can be used to quantify the magic of a pure state [34]. In particular, for a pure n-qubit state ψ, its α-stabilizer entropy $M_\alpha(\psi)$ is given by

$$M_\alpha(\psi) = H_\alpha(p_\psi) - n. \tag{A23}$$

Here, the normalization is chosen such that for a stabilizer state, $M_\alpha(\psi) = 0$. Note that $M_\alpha(\psi)$ takes values in $[0,n]$.

Let us remark here that M_0 is not a robust measure of magic in the context of our work. Concretely, there exist pure states ψ that are very close to stabilizer states in trace distance but they exhibit almost maximal $M_0(\psi)$. So, while these states should be regarded as low-magic from the perspective of higher-order stabilizer entropies M_α with $\alpha > 1$ [26] or other magic monotones such as min-relative entropy of magic [33], according to $M_0(\psi)$, they are as magical as it gets. The different notions of magic captured by stabilizer entropies with $\alpha < 1$ and $\alpha > 1$ have also recently been discussed in Ref. [44].

Nevertheless, M_0 captures well the magic of some important classes of states, such as the class of t-doped stabilizer states [21]. These are states obtained from computational basis states by Clifford circuits doped with a finite number t of non-Clifford gates. This is because any such t-doped stabilizer state (where t is the number of single-qubit non-Clifford gates) satisfies $M_0(\psi) \leq 2t$.

4. Relating properties of the Pauli distribution to magic and entanglement

In this section, we collect a few lemmas connecting properties of the Pauli distribution p_ρ to magic and entanglement properties of the state ρ. These will be used notably in Appendix B 4 to obtain sufficient conditions on an unknown state for efficient Pauli sampling.

We start by relating the entanglement across a cut $(A:B)$ in a state ρ to the size of the corresponding marginals of p_ρ. Recall from Eq. (A14) that for pure states $\rho = |\psi\rangle\langle\psi|$, the marginals of the Pauli distribution are given by $p_\psi(x_A) = 1/2^{|A|}\mathrm{tr}(\rho_A P_A \rho_A P_A)$. In the following lemma, we show that we can provide lower bounds on these marginals in terms of the Schmidt rank of the corresponding bipartition. We have recently learned that an equivalent result has been independently derived in Ref. [45], in the context of MPSs and their stabilizer group.

Lemma A1 (Bounding marginals via Schmidt rank). Let $(A:B)$ be a bipartition of the set of n qubits. Let $\rho = |\psi\rangle\langle\psi|$ be a pure state and let ρ_A be its reduced state on the subsystem A. Let r be the Schmidt rank of ρ corresponding to the bipartition $(A:B)$. Then, for any n-qubit Pauli string $P = P_A \otimes P_B$, it holds that

$$p_\psi(x_A) = \frac{\mathrm{tr}(\rho_A P_A \rho_A P_A)}{2^{|A|}} \geq \frac{\mathrm{tr}(\rho P)^2}{2^{|A|} r}. \tag{A24}$$

Proof. Start by Schmidt decomposing the state vector $|\psi\rangle = \sum_{i=1}^r \sqrt{\lambda_i}\,|i_A\rangle\,|i_B\rangle$ in terms of the Schmidt coefficients $\sqrt{\lambda_i}$. The reduced state is then given by $\rho_A = \sum_{i=1}^r \lambda_i |i_A\rangle\langle i_A|$. Using this expression for the reduced state ρ_A, we find

$$\mathrm{tr}(\rho_A P_A \rho_A P_A) = \sum_{i=1}^{r}\sum_{j=1}^{r} \lambda_i\lambda_j\,|\langle i_A|\,P_A\,|j_A\rangle|^2\,. \tag{A25}$$

On the other hand, expanding $\mathrm{tr}(\rho P)$ in terms of the Schmidt decomposition, we find

$$\begin{aligned}\mathrm{tr}(\rho P) &= \mathrm{tr}(\rho P_A \otimes P_B)\\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} \sqrt{\lambda_i\lambda_j}\,\langle i_A|\,P_A\,|j_A\rangle\,\langle i_B|\,P_B\,|j_B\rangle\,.\end{aligned} \tag{A26}$$

Using the Cauchy-Schwarz inequality, we find

$$\begin{aligned}&\mathrm{tr}(\rho P)^2\\ &\leq \left(\sum_{i,j=1}^{r} \lambda_i\lambda_j\,|\langle i_A|\,P_A\,|j_A\rangle|^2\right)\left(\sum_{i,j=1}^{r} |\langle i_B|\,P_B\,|j_B\rangle|^2\right),\\ &= \mathrm{tr}(\rho_A P_A \rho_A P_A)\left(\sum_{i,j=1}^{r} |\langle i_B|\,P_B\,|j_B\rangle|^2\right).\end{aligned} \tag{A27}$$

Now, it remains to upper bound $\sum_{ij}^r |\langle i_B|\,P_B\,|j_B\rangle|^2$. To this end, note that since P_B has eigenvalues bounded by 1, we

have that

$$\sum_{ij}^{r} |\langle i_B| P_B |j_B\rangle|^2 \leq \sum_{ij}^{r} |\langle i_B| I_B |j_B\rangle|^2 = \sum_{ij}^{r} \delta_{ij} = r, \tag{A28}$$

where we have used orthonormality of the Schmidt vectors $|i_B\rangle$. This completes the proof. ■

Next, we consider the connection of the Pauli distribution p_ρ to the magic of the state ρ as measured by its stabilizer entropy. In particular, we are interested in the average size of $\alpha_\rho(x)^2 = \mathrm{tr}\,(\rho P_x)^2$ for $x \sim p_\rho$. To this end, we start with the following definition.

Definition A1. Let ρ be an n-qubit state and let $\tau \in [0,1]$. Denote $\alpha_\rho(x)^2 = \mathrm{tr}\,(\rho P_x)^2$. Then, we define the *cumulative distribution function* (CDF) associated with ρ as

$$\begin{aligned} F_\rho(\tau) &:= \underset{x\sim p_\rho}{\mathbf{Pr}}\left[\alpha_\rho(x)^2 < \tau\right] \\ &= \sum_{\{x:\alpha_\rho(x)^2<\tau\}} p_\rho(x) \\ &= \sum_{\{x:\alpha_\rho(x)^2<\tau\}} \frac{\alpha_\rho(x)^2}{2^n \mathrm{tr}\left(\rho^2\right)}. \end{aligned} \tag{A29}$$

Note that

$$\underset{x\sim p_\rho}{\mathbf{Pr}}\left(\alpha_\rho(x)^2 \geq \tau\right) = 1 - F_\rho(\tau). \tag{A30}$$

We can now relate the CDF to the stabilizer entropies M_α for different α. For $\alpha = 0$, the corresponding stabilizer entropy $M_0(\psi)$ is related to the size of the support of the Pauli distribution p_ψ via

$$\begin{aligned} M_0(\psi) &= \log|\{x : p_\psi(x) > 0\}| - n \\ &= \log|\{x : \alpha_\psi(x)^2 \neq 0\}| - n. \end{aligned} \tag{A31}$$

From this relation, we directly obtain the following lemma.

Lemma A2 (Bounding the CDF in terms of M_0). For any pure state ψ with stabilizer entropy $M_0(\psi)$ and $0 < \tau \leq 1$, it holds that

$$F_\psi(\tau) \leq 2^{M_0(\psi)}\tau. \tag{A32}$$

Proof. We have that

$$\begin{aligned} F_\psi(\tau) &= \sum_{\{x:\alpha_\rho(x)^2\leq\tau\}} \frac{\alpha_\psi(x)^2}{2^n} \\ &\leq \frac{\tau}{2^n} \sum_{\{x:\alpha_\rho(x)^2\leq\tau\}} 1 \leq \frac{\tau}{2^n} \sum_{\{x:\alpha_\rho(x)^2\neq 0\}} 1 \\ &= \frac{\tau}{2^n} 2^{n+M_0(\psi)} = 2^{M_0(\psi)}\tau. \end{aligned} \tag{A33}$$

■

Similarly, we can also obtain a bound on $F_\psi(\tau)$ in terms of $M_1(\psi)$ by means of Markov's inequality.

Lemma A3 (Bounding the CDF in terms of M_1). For any pure state ψ with stabilizer entropy $M_1(\psi)$ and $0 < \tau \leq 1$, it holds that

$$F_\psi(\tau) \leq \frac{M_1(\psi)}{\log(1/\tau)}. \tag{A34}$$

Proof. The proof uses Markov's inequality, $\mathbf{Pr}(X \geq a) \leq \mathbb{E}[X]/a$, which holds for any non-negative random variable X and $a > 0$. Now, in our case, let

$$X := \log\left(\frac{1}{\mathrm{tr}\,(P_x\psi)^2}\right), \tag{A35}$$

where x is drawn according to p_ψ. The bound $\mathrm{tr}\,(P_x\psi)^2 \leq 1$ implies that $X \geq 0$. Then,

$$\begin{aligned} \mathbb{E}[X] &= \sum_x p_\psi(x) \log\left(\frac{1}{\mathrm{tr}\,(P_x\psi)^2}\right) \\ &= H_1\left(p_\psi\right) - n = M_1(\psi). \end{aligned} \tag{A36}$$

So, using Markov's inequality, we find that for any $a > 0$,

$$\underset{x\sim p_\psi}{\mathbf{Pr}}\left(\mathrm{tr}\,(P_x\psi)^2 \leq 2^{-a}\right) \leq \frac{M_1(\psi)}{a} \tag{A37}$$

or, equivalently, letting $\tau = 2^{-a}$,

$$\underset{x\sim p_\psi}{\mathbf{Pr}}\left(\alpha_\psi(x)^2 \leq \tau\right) \leq -\frac{M_1(\psi)}{\log(\tau)}. \tag{A38}$$

■

5. Statistical versus computational indistinguishability

In this section, we provide a brief overview of the concepts of statistical indistinguishability and computational indistinguishability, which are frequently employed throughout the paper to establish no-go results.

Given a (discrete) ensemble of quantum states $\mathcal{E} = \{|\psi\rangle\}$, we define its corresponding mixture

$$\Psi_{\mathcal{E}}^{(M)} := \frac{1}{|\mathcal{E}|} \sum_{|\psi\rangle \in \mathcal{E}} |\psi\rangle \langle\psi|^{\otimes M}. \tag{A39}$$

This definition can be readily extended to continuous ensembles of quantum states. In particular, we will focus on the ensemble of all quantum pure states, which mixture is defined as

$$\Psi_{\text{Haar}}^{(M)} = \int_{\text{Haar}} d\psi \, |\psi\rangle \langle\psi|^{\otimes M}, \tag{A40}$$

where $\int_{\text{Haar}}$ is the integral with respect to the Haar measure on the set of pure quantum states (see also Ref. [28] for technical details).

Two ensembles $\mathcal{E}_1$ or $\mathcal{E}_2$ are said to be *statistically indistinguishable* if any distinguisher, provided with polynomially many copies of a quantum state vector $|\psi\rangle$ uniformly drawn from one of the two ensembles, cannot discern which ensemble the state originates from (with high probability). Equivalently, any algorithm capable of distinguishing the ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ (with high probability) must have a sample complexity $\omega(\text{poly}(n))$. It follows from the Holevo-Helstrom theorem [46, Theorem 3.4] that a sufficient condition for statistical indistinguishability of $\mathcal{E}_1$ and $\mathcal{E}_2$ is closeness in trace distance of their corresponding mixtures $\Psi_{\mathcal{E}_1}^{(M)}$, $\Psi_{\mathcal{E}_2}^{(M)}$.

Definition A2 (Statistical indistinguishability). The ensemble $\mathcal{E}_1$ is statistically indistinguishable from $\mathcal{E}_2$ if, for every $M = O(\text{poly}(n))$, the following holds:

$$\left\| \Psi_{\mathcal{E}_1}^{(M)} - \Phi_{\mathcal{E}_2}^{(M)} \right\|_1 = \text{negl}(n), \tag{A41}$$

where $\text{negl}(n) := o(\text{poly}^{-1}(n))$.

A strictly weaker notion is that of *computational indistignuishability*. Two ensembles, $\mathcal{E}_1$ and $\mathcal{E}_2$, are computationally indistinguishable if any distinguishing algorithm, provided with polynomially many copies of a quantum state vector $|\psi\rangle$ uniformly drawn from one of the two ensembles, and bounded to run in *polynomial time*, cannot discern which ensemble the state originates from (with high probability). Formally, we have the following.

Definition A3 (Computational indistinguishability). The ensemble $\mathcal{E}_1$ is computationally indistinguishable from $\mathcal{E}_1$ if, for any $M = O(\text{poly}(n))$ and any polynomial-time quantum algorithm $\mathcal{A}$, the following holds:

$$\left| \Pr_{|\psi\rangle \in \mathcal{E}_1} [\mathcal{A}(|\psi\rangle) = 1] - \Pr_{|\psi\rangle \in \mathcal{E}_2} [\mathcal{A}(|\psi\rangle) = 1] \right| = \text{negl}(n), \tag{A42}$$

where $\text{negl}(n) := o(\text{poly}^{-1}(n))$.

Computational indistinguishability is strictly weaker than statistical indistinguishability. That is, statistical indistinguishability implies computational indistinguishability but not the other way around.

Before concluding the section, let us provide some useful lemmas that will be used throughout the paper.

Lemma A4 (Statistical indistinguishability). Let $\mathcal{E}$ be an ensemble of quantum states that is statistically indistinguishable from the set of Haar random states. Then, for every unitary V, the set $\mathcal{E}_V : \{V|\psi\rangle \mid |\psi\rangle \in \mathcal{E}\}$ is statistically indistinguishable from Haar.

Proof. The proof uses the fact that the trace norm is unitarily invariant. Indeed,

$$\begin{aligned} \left\| \Psi_{\mathcal{E}_V} - \Psi_{\text{Haar}} \right\|_1 &= \left\| V^{\otimes M} \Psi_{\mathcal{E}}^{(M)} V^{\dagger \otimes M} - \Psi_{\text{Haar}}^{(M)} \right\|_1 \\ &= \left\| \Psi_{\mathcal{E}}^{(M)} - V^{\dagger \otimes M} \Psi_{\text{Haar}}^{(M)} V^{\otimes M} \right\|_1 \\ &= \left\| \Psi_{\mathcal{E}}^{(M)} - \Psi_{\text{Haar}}^{(M)} \right\|_1 = \text{negl}(n), \end{aligned} \tag{A43}$$

where we have used the fact that $V^{\dagger \otimes M} \Psi_{\text{Haar}}^{(M)} V^{\otimes M} = \Psi_{\text{Haar}}^{(M)}$ for the left or right invariance of the Haar measure. ■

Lemma A5. Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two ensembles that are statistically indistinguishable from an ensemble $\mathcal{E}_3$. Then, $\mathcal{E}_1$ is statistically indistinguishable from $\mathcal{E}_2$.

Proof. We can use the triangular inequality and write

$$\begin{aligned} \left\| \Psi_{\mathcal{E}_1}^{(M)} - \Psi_{\mathcal{E}_2}^{(M)} \right\|_1 &\le \left\| \Psi_{\mathcal{E}_1}^{(M)} - \Psi_{\mathcal{E}_3}^{(M)} \right\|_1 + \left\| \Psi_{\mathcal{E}_3}^{(M)} - \Psi_{\mathcal{E}_2}^{(M)} \right\|_1 \\ &= \text{negl}(n). \end{aligned} \tag{A44}$$

■

Lemma A6. Let $\mathcal{E}$ be an ensemble of quantum states that is computationally indistinguishable from the set of Haar-random states. Then for every poly-size circuit V, the set $\mathcal{E}_V : \{V|\psi\rangle \mid |\psi\rangle \in \mathcal{E}\}$ is computationally indistinguishable from Haar.

Proof. For the sake of contradiction, let us assume that there is a polynomial-time quantum algorithm $\mathcal{A}$ able to distinguish the ensemble $\mathcal{E}_V$ and Haar. Let us show that using $\mathcal{A}$, we can distinguish $\mathcal{E}$ from Haar in polynomial time. Let $|\psi\rangle$ be drawn either from $\mathcal{E}$ or from Haar. Let us apply a polynomial-size circuit V to the unknown state vector $|\psi\rangle$. Note that if $|\psi\rangle \in \mathcal{E}$, then $V|\psi\rangle \in \mathcal{E}_V$. Conversely, if $|\psi\rangle \in$ Haar, then $V|\psi\rangle \in$ Haar for the left invariance of the Haar measure. Therefore, we can use the algorithm $\mathcal{A}$ that is able to *efficiently* distinguish $\mathcal{E}_V$ from Haar to

effectively distinguish $\mathcal{E}$ from Haar. This is a contradiction. Therefore, we conclude that $\mathcal{E}_V$ is computationally indistinguishable from Haar. ■

Lemma A7. Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two ensembles of quantum states that are computationally indistinguishable from an ensemble $\mathcal{E}_3$. Then, $\mathcal{E}_1$ is computationally indistinguishable from $\mathcal{E}_2$.

Proof. Using the definition of computational indistinguishability in Eq. (A42), we can write

$$\begin{aligned}&\left|\Pr_{|\psi\rangle\in\mathcal{E}_1}[\mathcal{A}(|\psi\rangle)=1]-\Pr_{|\psi\rangle\in\mathcal{E}_2}[\mathcal{A}(|\psi\rangle)=1]\right|\\&\leq\left|\Pr_{|\psi\rangle\in\mathcal{E}_1}[\mathcal{A}(|\psi\rangle)=1]-\Pr_{|\psi\rangle\in\mathcal{E}_3}[\mathcal{A}(|\psi\rangle)=1]\right|\\&+\left|\Pr_{|\psi\rangle\in\mathcal{E}_3}[\mathcal{A}(|\psi\rangle)=1]-\Pr_{|\psi\rangle\in\mathcal{E}_2}[\mathcal{A}(|\psi\rangle)=1]\right|\\&=\mathrm{negl}(n)\end{aligned}\tag{A45}$$

for any polynomial-time quantum algorithm $\mathcal{A}$. ■

6. Subset phase states

In the following section, we introduce two discrete ensembles of states, *random subset phase states* and *pseudorandom subset phase states* that are, respectively, statistically and computationally indistinguishable from the ensemble of Haar-random states.

Definition A4 (Random subset phase states [29]). We denote the ensemble of *random subset phase states* $\mathcal{E}_{f,S}$ as the set of states

$$|\psi_{f,S}\rangle=\frac{1}{\sqrt{|S|}}\sum_{x\in S}(-1)^{f(x)}|x\rangle\tag{A46}$$

for some random subset of bit strings $S\subset\{0,1\}^n$ and $f:\{0,1\}^n\mapsto\{0,1\}$ a random Boolean function.

The ensemble of random subset phase states is statistically indistinguishable from Haar-random states for $|S|=\omega(\mathrm{poly}(n))$ [29]. Further, it has been shown that the ensemble of subset phase states features low magic and low entanglement, in contrast to the ensemble of Haar-random states (see Ref. [30]).

Lemma A8 (Restatement of Theorem 1 in Ref. [30] and Theorem 2.7 in Ref. [29]). For all $|S|$ such that $\omega(\log(n))\leq\log|S|\leq n$, there exists an ensemble of random subset phase states $\mathcal{E}_{f,S}$ such that:

(1) for all $\alpha=\Theta(1)$, $M_\alpha(\psi_{f,S})=O(\log|S|)$ and $M_\alpha(\psi_{f,S})=\omega(\log n)$

(2) for all α, $\mathcal{S}_\alpha\left(\psi_{f,S}\right)=O(\log|S|)$ and $\mathcal{S}_\alpha(\psi_{f,S})=\omega(\log n)$ over every cut $(A:B)$ [47]

with high probability over $|\psi_{f,S}\rangle\in\mathcal{E}_{f,S}$.

A limitation of subset phase states lies in their construction from random Boolean functions and random permutations, rendering their implementation inefficient. To address this limitation, we consider the ensemble of pseudorandom subset phase states. These differentiate themselves from random subset phase states only in the fact that S is taken as a pseudorandom subset and f a pseudorandom Boolean function. As outlined in Ref. [29], we can efficiently prepare pseudorandom subset phase states using pseudorandom permutations and pseudorandom Boolean functions, assuming the existence of *one-way functions* (OWFs)—Boolean functions. We denote the ensemble pseudorandom subset phase states as $\tilde{\mathcal{E}}_{f,S}$. In Ref. [29], it has been shown that the ensemble of pseudorandom subset phase states is computationally indistinguishable from random subset phase states for any $|S|=\omega(\mathrm{poly}(n))$ and thus is also computationally indistinguishable from Haar-random states.

APPENDIX B: PAULI SAMPLING

In this appendix, we discuss the core subroutine at the heart of our approach to distributed quantum inner-product estimation. We refer to this as (approximate) *Pauli sampling* and begin with a formal definition.

Definition B1 (Approximate Pauli sampling). Given $\Delta>0$ and access to the unknown quantum state ρ, we define approximate Pauli sampling as the task of sampling from a distribution q such that

$$\left\|q-p_\rho\right\|_{\mathrm{TV}}\leq\Delta.\tag{B1}$$

Here, p_ρ is the Pauli distribution corresponding to ρ, as defined in Eq. (A9).

We start our discussion with an observation regarding the feasibility of Pauli sampling for arbitrary states using single-copy algorithms. By single-copy algorithms, we mean algorithms without quantum memory, which use their access to copies of an unknown state ρ by processing each copy individually (see Ref. [48]). The following lemma is a direct consequence of the limitations of such single-copy algorithms, which have been demonstrated by Ref. [48].

Lemma B1 (Exponential sample complexity of Pauli sampling for general states for single-copy algorithms). Let $\Delta>0$ and let ρ be an unknown n-qubit (possibly mixed) quantum state. Then, any single-copy algorithm for

approximate Pauli sampling up to error $\Delta < 1/4$ requires $\Omega(2^{n/2})$ copies of ρ.

Proof. Consider the task of purity testing discussed in Ref. [48, Sec. 5.2]. Purity testing is the task of deciding if an unknown state ρ is either a pure state or the maximally mixed state. Reference [48] demonstrates a sample-complexity lower bound of $\Omega\left(2^{n/2}\right)$ for this distinguishing task that applies to any single-copy algorithm. Here, we note that this lower bound must also apply to the task of approximate Pauli sampling. To see this, consider the Pauli distribution of the maximally mixed state $\rho = I/2^n$,

$$p_{I/2^n}(x) = \begin{cases} 1, & x = 0^{2n}, \\ 0, & \text{else.} \end{cases} \tag{B2}$$

Now, assume that there is an algorithm for Δ-approximate Pauli sampling from an unknown state ρ that is either the maximally mixed state or a pure state. In case ρ is maximally mixed, this alleged algorithm would sample $x = 0^{2n}$ with probability at least $q(x = 0^{2n}) > 1 - \Delta$. On the other hand, for a pure state, we have $p_\rho(x = 0^{2n}) = 2^{-n}$, i.e., sampling the outcome $x = 0^{2n}$ is exponentially unlikely under p_ρ and so $q(x = 0^{2n}) < 1/2^n + \Delta$. Hence, the approximate Pauli-sampling algorithm facilitates deciding between these cases with high probability and thus inherits the lower bound for purity testing. ■

Lemma B1 shows that, in general, Pauli sampling requires multicopy algorithms such as Bell sampling, which is a two-copy algorithm. However, note that Lemma B1 crucially uses the fact that we require an algorithm that works for both pure and mixed states. If we assume from the start that the input state ρ is pure, then the limitation of Lemma B1 does not necessarily apply. This is an important caveat in the context of our work, as we mostly work only with pure states.

Next, let us remark on existing approaches to Pauli sampling. In particular, for certain classes of pure states, Pauli sampling has been shown to admit an efficient solution. In particular, if ρ is a real pure state, i.e., a state with real coefficients in the computational basis, then we can directly Pauli sample using Bell sampling by associating each outcome $x \in \{0,1\}^{2n}$ with its corresponding Pauli word [11]. Examples of such states are the ground states of stoquastic Hamiltonians, as can be seen by invoking the Perron-Frobenius theorem [37]. In fact, most natural many-body Hamiltonians considered in the literature have this feature [36].

The other class of states for which we can Pauli sample directly and efficiently are the pure stabilizer states. It has been shown in Refs. [11,12] that for these states, *Bell difference sampling* corresponds to Pauli sampling. Bell difference sampling is the sampling procedure consisting of independently performing Bell sampling on $\rho \otimes \rho$ and adding the two outcomes (mod 2), again associating the outcome with its corresponding Pauli operator.

Finally, it is sometimes possible to perform Pauli sampling based on an efficient classical representation of the state [25]. Therefore, one alternative approach to Pauli sampling relies on learning such a classical representation of the unknown state first. This strategy can be applied to stabilizer states [16] as well as states that admit an efficient MPS representation [39].

1. Hardness of approximate sampling even for bounded magic and entanglement

In this section, we provide a rigorous proof of Theorem 1, which we restate below for convenience. Our proof follows via a reduction between Pauli sampling and a property-testing problem. In particular, we show that Pauli sampling allows us to test the imaginarity of an unknown state. The *imaginarity* of a state vector $|\psi\rangle$ is defined as $I(|\psi\rangle) := 1 - |\langle\psi|\psi^*\rangle|^2$, where $|\psi^*\rangle$ denotes the conjugate state with respect to the computational basis. Imaginarity testing is the task of deciding whether an unknown state ψ has $I(|\psi\rangle) = 1$ or far from it. It has previously been shown that testing the imaginarity of an unknown state ρ requires exponentially many copies of ρ [31]. We defer the proofs of the auxiliary lemmas in this section to Appendix E.

Subset phase states as defined in Eq. (A46) are real states with $I(|\psi\rangle) = 0$. In Lemma B2, by conjugating these real states with randomly drawn local Clifford unitaries, we construct a new (pseudo)random ensemble of *imaginary* states and show that they are statistically indistinguishable from subset phase states. In Lemma B3, we show how access to samples from the Pauli distribution can be used to estimate the imaginarity of an unknown state. Putting these ingredients together, we obtain a lower bound on the sample complexity of Pauli sampling.

Lemma B2 (Pseudorandom states with high imaginarity). Let $U = \bigotimes_{i=1}^n C_i$ for C_i being random single-qubit Clifford unitaries. Then define the ensemble $\mathcal{E}_U = \{U|\psi_{f,S}\rangle \mid |\psi_{f,S}\rangle \in \mathcal{E}_{f,S}\}$, where $\mathcal{E}_{f,S}$ is an ensemble of random subset phase states. Denote $|\psi^*\rangle$ as the conjugate state in the computational basis. We have the following list of results:

(1) *Imaginarity.* $\mathbf{Pr}_{|\psi\rangle\sim\mathcal{E}_U}[I(|\psi\rangle) > 1/100] = 1 - \text{negl}(n)$.
(2) *Statistical indistinguishability.* For $|S| = \omega(\text{poly}(n))$, the two ensembles $\mathcal{E}_U$ and $\mathcal{E}_{f,S}$ are statistically indistinguishable.

Note that an analogous result (with computational rather than statistical indistinguishability) holds between ensembles $\tilde{\mathcal{E}}_U$ and $\tilde{\mathcal{E}}_{f,S}$, where $\tilde{\mathcal{E}}_U = \{U|\psi_{f,S}\rangle \mid |\psi_{f,S}\rangle \in \tilde{\mathcal{E}}_{f,S}\}$, and $\tilde{\mathcal{E}}_{f,S}$ are pseudorandom subset phase states.

Lemma B3 (Pauli sampling implies estimating imaginarity). Let $\rho = |\psi\rangle\langle\psi|$ be a pure quantum state and let p_ρ be its associated Pauli distribution. Then, given black-box access to an algorithm for sampling from a distribution q such that $\|p_\rho - q\|_{\mathrm{TV}} \leq \Delta$, for $0 \leq \Delta < 1$, there exists an efficient algorithm to estimate the imaginarity $I(|\psi\rangle) = 1 - |\langle\psi|\psi^*\rangle|^2$ within additive error $\epsilon > \Delta$ and failure probability δ using $2/(\epsilon - \Delta)^2 \ln 2/\delta$ samples from q.

Using the preceding lemmas, we are now ready to prove Theorem 1, which we restate below.

Theorem B1 (Formal version of Theorem 1). Let $|\psi\rangle$ be an unknown n-qubit pure state and let p_ψ be the Pauli distribution associated with $|\psi\rangle$. Then, there is no algorithm that can sample from a distribution q such that $\|p_\rho - q\|_{\mathrm{TV}} \leq 1/3$ using $O(\mathrm{poly}(n))$ copies of $|\psi\rangle$. In fact, this is the case even if one is promised that the state has bounded magic and entanglement, such that $M_\alpha(\psi) < g(n)$ (with $\alpha = O(1)$) and $S_\alpha(\psi) < f(n)$ (over all cuts $(A:B)$), where $f(n), g(n) = \omega(\log n)$.

Proof. By Lemma A8, for any two functions $f(n)$, $g(n) = \omega(\log n)$, there is an ensemble of subset phase states $\mathcal{E}_{f,S}$ with $|S| = \omega(\mathrm{poly}(n))$ that satisfies $M_\alpha(|\psi_{f,S}\rangle) = O(g(n))$ and $S_\alpha(|\psi_{f,S}\rangle) = O(f(n))$ over all cuts $(A:B)$. We can further construct an ensemble $\mathcal{E}_U = \{U|\psi_{f,S}\rangle \mid |\psi_{f,S}\rangle \in \mathcal{E}_{f,S}\}$, with U being a local random Clifford circuit.

Note that local Clifford operations on some state $|\psi\rangle$ preserve its entanglement and magic. Moreover, by virtue of Lemma B2, $\mathcal{E}_U$ is statistically indistinguishable from $\mathcal{E}_{f,S}$ and it contains states obeying $|\langle\psi|\psi^*\rangle|^2 < 1/100$ with high probability over the random U, while $|\langle\psi|\psi^*\rangle|^2 = 1$ for all states in $\mathcal{E}_{f,S}$.

By Lemma B3, we know that $O(1)$ samples from q suffice to obtain an estimate of $|\langle\psi|\psi^*\rangle|^2$ up to precision $\epsilon = 1/3 + 1/100$ with high probability. Therefore, assuming we could sample from q, we would also be able to distinguish between the two ensembles $\mathcal{E}_U$ and $\mathcal{E}_{f,S}$ with high probability, using only $O(\mathrm{poly}(n))$ copies of $|\psi\rangle$, depending on whether the estimate of $|\langle\psi|\psi^*\rangle|^2$ is in $[0, 1/3 + 1/200]$ or in $[2/3 - 1/100, 1]$. This contradicts the statistical indistinguishability of $\mathcal{E}_U$ and $\mathcal{E}_{f,S}$ as proven in Lemma B2 and thus concludes the proof. ■

In Theorem B1, we have demonstrated that there is no sample-efficient algorithm capable of sampling, even approximately, from the Pauli distribution of an unknown state $|\psi\rangle$ with bounded magic and entanglement. In what follows, we restrict the class of states to those states that are efficiently preparable by a quantum circuit of size scaling at most as $\mathrm{poly}(n)$. For this class of states, we can show that no computationally efficient algorithm exists that samples from the Pauli distribution modulo a cryptographic assumption. In particular, under the assumption on the existence of OWFs, we can construct *pseudorandom* subset phase states efficiently (see Appendix A 6).

Theorem B2. Let $|\psi\rangle$ be an unknown, yet efficiently preparable, n-qubit pure state and let p_ψ be the Pauli distribution associated with $|\psi\rangle$. Moreover, let $|\psi\rangle$ be such that $M_\alpha(\psi) < g(n)$ (with $\alpha = O(1)$) and $S_\alpha(\psi) < f(n)$ (over all cuts $(A:B)$) where $f(n), g(n) = \omega(\log n)$. Then, there is no polynomial-time quantum algorithm that can sample from a distribution $q(x)$ such that $\|p_\rho - q\|_{\mathrm{TV}} \leq 1/3$ that uses $O(\mathrm{poly}(n))$ copies of $|\psi\rangle$, assuming the existence of OWFs.

Proof. The proof follows identically to the one of Theorem B1. However, it uses the computational indistinguishability of the two ensembles of states $\tilde{\mathcal{E}}_U$ and $\tilde{\mathcal{E}}_{f,S}$, as opposed to the (stronger) statistical indistinguishability of $\mathcal{E}_U$ and $\mathcal{E}_{f,S}$. ■

2. Sampling algorithm

We have seen that for "complex states," i.e., states with high entanglement and high magic, Pauli sampling cannot be performed efficiently. Now, we turn our attention to the complementary regime and we will provide an efficient algorithm for Pauli sampling for states with bounded entanglement and magic.

a. Estimating marginals

Our starting point for the description of the algorithm is the insight that Bell-measurement data can be used to estimate marginals of the Pauli distribution p_ρ. In particular, in Appendix A 1, we have introduced the concept of Bell sampling as the measurement routine in which we perform Bell measurements across two copies of a state ρ. Now, recall that using the decomposition of the SWAP operator in Eq. (A5), we can express the marginals as [see Eq. (A12)]

$$p_\rho(x_1, \ldots, x_k) = p_\rho(x_{1:k}) = \frac{1}{2^k \mathrm{tr}(\rho^2)} \langle P_{x_{1:k}}^{\otimes 2} \otimes \mathbb{S}^{\otimes n-k} \rangle_{\rho\otimes\rho}. \tag{B3}$$

Here, we use the shorthand notation $x_{1:k}$ to denote $x_1, \ldots, x_k$. Since the Bell states are eigenstates of the SWAP operator $\mathbb{S}$ as well as of $P^{\otimes 2}$ for any single-qubit Pauli P with eigenvalues ± 1, we can estimate any expectation value of the form $\langle P_{x_{1:k}}^{\otimes 2} \otimes \mathbb{S}^{\otimes n-k} \rangle_{\rho\otimes\rho}$ to within ϵ precision using $O(1/\epsilon^2)$ Bell-sampling outcomes. Furthermore, since the purity is given by $\mathrm{tr}(\rho^2) = \langle \mathbb{S}^{\otimes n} \rangle_{\rho\otimes\rho}$, it can also be estimated from the Bell-measurement data. Hence, we can obtain an estimate for the right-hand side of Eq. (B3) by estimating $\langle P_{x_{1:k}}^{\otimes 2} \otimes \mathbb{S}^{\otimes n-k} \rangle_{\rho\otimes\rho}$ and $\mathrm{tr}(\rho^2)$ from the Bell data. For more details on the estimation, we refer the reader

to Appendix D. Importantly, the same measurement data can be used to estimate all marginals. That is, we can collect the data once and then classically estimate $p_\rho(x_{1:k})$ for any given string $x_{1:k}$ and $1 \leq k \leq n$ in postprocessing. In the following lemma, we record the sample complexity for simultaneously estimating all such marginals. We note that a similar observation about simultaneous estimation of all $\langle P_x^{\otimes 2}\rangle_{\rho\otimes\rho}$ for all $x \in \{0,1\}^{2n}$ has already been made in Ref. [49, Appendix E]. Here, we extend this observation to observables also containing $\mathbb{S}$ factors.

Lemma B4 (Estimating all even marginals). Let $\epsilon > 0, \delta > 0$ and let ρ be a pure state. Then, $N = O\left(n \log(1/\delta)/\epsilon^2\right)$ pairs of copies $\rho \otimes \rho$ suffice to produce, with probability $1-\delta$, estimates $\pi_\rho(x_{1:k})$ such that

$$\left|\pi_\rho(x_{1:k}) - p_\rho(x_{1:k})\right| \leq \frac{\epsilon}{2^k} \tag{B4}$$

for all $x \in \{0,1\}^{2n}$ and all $1 \leq k \leq n$.

The proof of this lemma is presented in Appendix D. There, we also present the slight generalization of this lemma to mixed states leading to an overall sample complexity of $N = O\left(n \log(1/\delta)/(\epsilon^2 \mathrm{tr}(\rho^2)^2)\right)$. Note that the purity $\mathrm{tr}(\rho^2)$ enters into the sample complexity because the size of the marginals scales with the purity [see Eq. (B3)].

Let us remark on a subtlety concerning Lemma B4. Recall that $x = (v_1, w_1, \ldots, v_n, w_n) \in \{0,1\}^{2n}$ such that $x_i = (v_i, w_i)$. Hence, the guarantees given in Lemma B4 only apply to *even* marginals of p_ρ, i.e., those marginals that correspond to a Pauli string $P_{x_{1:k}}$ of length k. It turns out, however, that the algorithm we will use for sampling from the Pauli distribution also requires odd marginals. These correspond to bit strings of length $2k-1$. Using the decomposition

$$\begin{aligned} p_\rho(v_1, w_1, \ldots, v_k) = {} & p_\rho(v_1, w_1, \ldots, v_k, w_k = 0) \\ & + p_\rho(v_1, w_1, \ldots, v_k, w_k = 1), \end{aligned} \tag{B5}$$

we can compute estimates for the odd marginals from estimates of the even ones. We thus obtain the following result, which applies to both even and odd marginals.

Corollary B1 (Estimating even and odd marginals). Let $\epsilon > 0, \delta > 0$ and let ρ be a pure state. Then, $N = O\left(n \log(1/\delta)/\epsilon^2\right)$ pairs of copies $\rho \otimes \rho$ suffice to produce, for all $1 \leq k \leq n$ with probability $1-\delta$,

(a) estimates $\pi_\rho(v_1, w_1, \ldots, v_k, w_k)$ such that

$$\begin{aligned} &\left|\pi_\rho(v_1, w_1, \ldots, v_k, w_k) - p_\rho(v_1, w_1, \ldots, v_k, w_k)\right| \\ &\quad \leq \frac{\epsilon}{2^k} \end{aligned} \tag{B6}$$

(b) estimates $\pi_\rho(v_1, w_1, \ldots, v_k)$ such that

$$\left|\pi_\rho(v_1, w_1, \ldots, v_k) - p_\rho(v_1, w_1, \ldots, v_k)\right| \leq \frac{2\epsilon}{2^k} \tag{B7}$$

for all $x = (v_1, w_1, \ldots, v_n, w_n) \in \{0,1\}^{2n}$ and all $1 \leq k \leq n$.

b. Ancestral sampling

The algorithm that we propose for Pauli sampling is based on the so-called ancestral-sampling algorithm and, in particular, a variant thereof put forward in Ref. [19]. We will begin by reviewing the standard algorithm and the adaptation thereof. Consider a distribution p over $\{0,1\}^n$. The ancestral-sampling algorithm makes use of the fact that the joint distribution $p(x)$ can be factorized in terms of conditionals as follows:

$$p(x) = p(x_1)\, p(x_2|x_1)\, p(x_3|x_1, x_2) \cdots p(x_n|x_1, \ldots, x_n). \tag{B8}$$

The sampling algorithm proceeds by sampling a bit string x in a bit-by-bit fashion. That is, it samples x_1 according to $p(x_1)$, then samples x_2 according to $p(x_2|x_1)$, where x_1 has been fixed by the previous step, and so on and so forth, until x is completely sampled. Furthermore, we can express the conditional probabilities in terms of marginals

$$p(x_k|x_1, \ldots, x_{k-1}) = \frac{p(x_1, \ldots, x_k)}{p(x_1, \ldots, x_{k-1})}. \tag{B9}$$

Hence, the ancestral algorithm can be run whenever we have access to marginals of the target distribution p from which we want to sample.

3. Adapted ancestral sampling

In previous works featuring the ancestral-sampling algorithm, access to these marginals $p(x_1, \ldots, x_k)$ is by some explicit computation. For instance, in the classical simulation literature, the marginals are computed from, say, the circuit description of the quantum circuit to be simulated (see, e.g., Refs. [18–20]). In contrast, in the context of our work, these marginals are estimated from measurement data as outlined above. Let $\pi(x_{1:k})$ denote such an estimate for $p(x_{1:k})$. Then, in particular, the approximation guarantees of Lemma B4 and Corollary B1 apply to $\pi(x_{1:k})$.

Note that, in general, additive error estimates $\pi(x_{1:k})$ of the marginal probabilities can take negative values. This would be an issue when trying to run the standard ancestral-sampling algorithm based on such possibly negative estimates. To deal with this issue, we turn to a variant of the ancestral-sampling algorithm put forward in Ref. [19], which we refer to as the *adapted ancestral-sampling algorithm*. The authors' adapted algorithm can

also be run based on possibly negative approximations to the marginals.

To allow such negative inputs, the algorithm differs in one aspect from the standard ancestral algorithm discussed above. Recall that in the kth step of the ancestral algorithm, we set the kth bit to 0/1 according to the probabilities $p(x_{1:k-1}, x_k = 0)$ and $p(x_{1:k-1}, x_k = 1)$, both of which are non-negative. With the adapted algorithm, taking the approximations $\pi(x_{1:k})$ as input, there are now two possible cases:

(1) either both estimates are non-negative, $\pi(x_{1:k-1}, x_k = 0), \pi(x_{1:k-1}, x_k = 1) \geq 0$
(2) or one of the two estimates is negative, so $\pi(x_{1:k-1}, x_k = 0) < 0$ or $\pi(x_{1:k-1}, x_k = 1) < 0$

In the first case, the algorithm functions in the same way as the standard ancestral-sampling algorithm described above; namely, it sets $x_k = 0$ with probability given by the ratio

$$\mathfrak{p} = \frac{\pi(x_{1:k-1}, x_k = 0)}{(\pi(x_{1:k-1}, x_k = 0) + \pi(x_{1:k-1}, x_k = 1))}, \tag{B10}$$

and it set $x_k = 1$ with probability $1 - \mathfrak{p}$. In the second case, however, if $\pi(x_{1:k-1}, x_k = 0) < 0$, the algorithm sets $x_k = 1$ (and, respectively, if $\pi(x_{1:k-1}, x_k = 1) < 0$, $x_k = 0$). In the pictorial representation of Fig. 3, this corresponds to deterministically taking the step along the path corresponding to the positive estimate of the marginal.

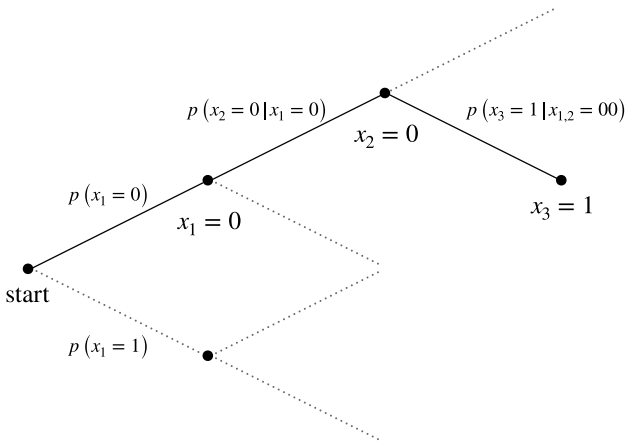


FIG. 3. A schematic example run of the ancestral-sampling algorithm. The algorithm can be visualized as a walk through a binary tree with its output corresponding to the entire path taken. The kth layer of the tree corresponds to the kth bit and each node corresponds to either of the (binary) values assigned to the corresponding bit. For example, in the second layer, the upper node corresponds to $x_2 = 0$ and the lower to $x_2 = 1$. The algorithm proceeds by traversing the tree from left to right, according to the conditional probabilities $p(x_k|x_1, \ldots, x_{k-1})$. In the figure, we see in a dark color the example path giving rise to the sample $x = 0001$, while the dotted lines show (parts) of other potential paths that could have been taken, leading to different outputs.

Below, we will provide a set of sufficient conditions for approximate Pauli sampling. We will first state these conditions with respect to an arbitrary distribution p over $\{0,1\}^n$, as we believe that our performance guarantees can be of independent interest for other settings in which one wants to sample a distribution p given access to additive error estimates of its marginals. We then proceed to reconnect to Pauli sampling by considering Pauli distributions p_ρ. We then give sufficient conditions on properties of the state ρ in terms of its magic and entanglement to give a characterization of efficiently sampleable states.

To state our performance guarantee, we need one additional definition.

Definition B2 (Corresponding marginals and $\mathfrak{f}(\gamma)$*).* Let p be a distribution over $\{0,1\}^n$ and let $\gamma > 0$. Then, for a fixed outcome $x \in \{0,1\}^n$, we say that the marginals $p(x_{1:k})$ for $k \in [n]$ are the marginals *corresponding* to x. Now, consider the following set:

$$S_\gamma = \left\{ x \in \{0,1\}^n : \forall k \in [n], p(x_{1:k}) \geq \frac{\gamma}{2^k} \right\}. \tag{B11}$$

In words, an outcome x is in S_γ if all its corresponding marginals are lower bounded as $p(x_{1:k}) \geq \frac{\gamma}{2^k}$. We define

$$\mathfrak{f}(\gamma) := 1 - p(S_\gamma). \tag{B12}$$

To exemplify this definition, take, e.g., the outcome $x = 001$. Then, $p(x_1 = 0)$, $p(x_1 = 0, x_2 = 0)$, and $p(x_1 = 0, x_2 = 0, x_3 = 1)$ are its corresponding marginals.

We are now in a position to state the main theorem of this section.

Theorem B3 (Performance guarantee of adapted ancestral-sampling algorithm). Let p be a distribution over $\{0,1\}^n$ and $\gamma > 0$. For all $\epsilon < \gamma/2$, given black-box access to estimates of the marginals such that

$$|\pi(x_{1:k}) - p(x_{1:k})| \leq \frac{\epsilon}{2^k}, \tag{B13}$$

the adapted ancestral algorithm samples from a distribution q that satisfies

$$\|q - p\|_{\mathrm{TV}} \leq \mathfrak{f}(\gamma) + \exp\left(\frac{4\epsilon n}{\gamma}\right) - 1. \tag{B14}$$

Note that the overall TV distance can be made arbitrarily small for any given distribution p by choosing γ sufficiently small and further choosing $\epsilon = O(\gamma/n)$. In particular, this choice allows one to suppress the second error term on the right-hand side of Eq. (B14) arbitrarily.

For this choice of ϵ, we will also apply the theorem in the subsequent section.

A detailed proof of Theorem B3 is presented in Appendix G. The key insight underlying the proof is the consideration of the set S_γ of outcomes x with k-body marginals $\pi(x_{1:k})$ larger than $\gamma/2^k$ (see Definition B2). On the one hand, the individual contribution to the final error coming from estimates of outcomes $x \in S_\gamma$ is multiplicative and can thus be bounded in terms of γ. On the other hand, the total contribution to the final error coming from the estimates of the rest of outcomes $x \notin S_\gamma$ can be bounded to be $f(\gamma)$ (see Definition B2).

4. Efficiently sampleable states

In Theorem B3, we have established sufficient conditions for the approximate sampling from a distribution when given estimates to its marginals. In this section, we want to establish sufficient conditions on properties of a pure state ρ that would result in efficient approximate Pauli sampling. In particular, we will show that as long as the state ρ has bounded entanglement and magic, we can approximately Pauli sample from p_ρ efficiently using the adapted ancestral-sampling algorithm discussed in the previous section.

First, let us explain our approximate Pauli-sampling algorithm. It proceeds via the following steps:

(1) *Data acquisition.* Use $2N$ copies of the unknown state ρ and perform Bell sampling on pairs of copies $\rho^{\otimes 2}$, obtaining a list of N outcomes, namely, $\{y_1, \ldots, y_N \in \{0,1\}^{2n}\}$.
(2) *Classical postprocessing.* Run the adapted ancestral-sampling algorithm on the estimates $\pi_\rho(x_{i:k})$ of the marginals $p_\rho(x_{i:k})$ to produce samples $x \in \{0,1\}^{2n}$. As discussed in Corollary B1, all required marginals can be estimated from the Bell-measurement data obtained in the data-acquisition phase.

The ancestral-sampling algorithm samples a bit string $x \in \{0,1\}^{2n}$ in a bit-by-bit fashion. Alternatively, since each pair of bits $x_i = (v_i, w_i)$ corresponds to a single-qubit Pauli P_{x_i}, the algorithm can be viewed as sampling a full n-qubit Pauli string P_x by drawing single-qubit Paulis qubit by qubit. In each pass of the ancestral algorithm, we have to estimate the marginal probabilities $p_\rho(x_1)$, $p_\rho(x_1, x_2)$, etc., which correspond to bipartitions $(1|2,\ldots,n)$, $(1,2|3,\ldots,n)$, etc. of the system of n qubits. It then becomes clear that the algorithm presupposes a certain qubit ordering, i.e., an assignment of which qubit is sampled first, which second, and so on. This choice of ordering can be crucial for the efficiency of the algorithm in our case. To see that, note that every ordering gives rise to a different sequence of bipartitions. By Lemma A1, we know that the size of the marginals crucially depends on the entanglement across these bipartitions. Hence, by Theorem B3, a path giving rise to bipartitions with little entanglement across is necessary in order for the sampling algorithm to be efficient.

To make this observation rigorous, we give the following definition.

Definition B3 (Qubit ordering). We identify a qubit ordering with a permutation $\pi \in S_n$, where S_n is the symmetric group. It acts on state vectors via

$$\pi\,|a_1,\ldots,a_n\rangle = \left|a_{\pi(1)},\ldots,a_{\pi(n)}\right\rangle. \tag{B15}$$

Hence, given a state ρ with respect to the canonical ordering corresponding to the identity permutation, the state with respect to a different ordering is given by $\pi\rho\pi^\dagger$.

Next, we define a measure of (multipartite) entanglement that is ordering specific and captures entanglement across the sequence of bipartitions $(1|2,\ldots,n)$, $(1,2|3,\ldots,n),\ldots,(1,\ldots,n-1|n)$.

Definition B4 (Path entanglement). Let ρ be an n-qubit pure state and let $\pi \in S_n$ specify a qubit ordering. Then, we define the *path entanglement* of ρ with respect to this ordering π as

$$E_0^\pi(\rho) = \max\left(\mathcal{S}_0(\rho_{\pi(1)}), \mathcal{S}_0(\rho_{\pi(1),\pi(2)}),\ldots,\mathcal{S}_0(\rho_{\pi(1),\ldots,\pi(n-1)})\right). \tag{B16}$$

Here, $\rho_{\pi(1)}, \rho_{\pi(1),\pi(2)},\ldots$ denote reduced states.

In words, this is the maximum Rényi-0 entanglement entropy across all bipartitions encountered by the ancestral-sampling algorithm when a certain qubit ordering is fixed.

To convey the importance of the choice of ordering, we present a few illustrative examples of states:

(1) *Product states.* In the case of product states (e.g., $|0\rangle^{\otimes n}$), the Pauli distribution p_ρ takes a product form such that $p_\rho(x) = p_\rho(x_1)\ldots p_\rho(x_n)$, which makes the choice of ordering irrelevant. Accordingly, the entanglement entropy is 0 along all bipartitions.
(2) *Bell pairs.* An interesting case in which the choice of ordering can render our algorithm efficient or inefficient is that of $n/2$ Bell pairs, i.e., $\rho = \left|\Phi_n^+\right\rangle\left\langle\Phi_n^+\right|$ [see Eq. (A7)]. Here, the Bell pairs are between qubits $(1, n/2+1), (2, n/2+2),\ldots,(n/2, n)$. In this case, the canonical ordering, corresponding to sampling qubit 1, then 2, then 3, and so on, leads to large path entanglement and hence requires an exponential number of Bell-sampling data. However, choosing the ordering π to be matching up with the pairing of the Bell pairs, such that one samples first qubit 1, then $n/2+1$,

then 2, then $n/2+2$, and so on, leads to a path entanglement $E_0^\pi(\rho)=2$ and hence to an efficient algorithm.

(3) *Two-dimensional (2D) cluster states.* Another interesting case is that of 2D cluster states [50]

$$|G\rangle = \prod_{(i,j)\in E} \mathrm{CZ}_{i,j} \bigotimes_{k\in V} |+\rangle_k\,, \tag{B17}$$

where V and E are the vertices and edges of a $\sqrt{n}\times\sqrt{n}$ square lattice G. For these states, all choices of orderings lead to a path entanglement of the order $E_0^\pi(\rho)=O(\sqrt{n})$, which renders our algorithm inefficient. One way to see this is by considering the bipartition at the halfway point of the ordering, i.e., $(\pi(1),\ldots,\pi(n/2):\pi(n/2+1),\ldots,\pi(n))=(A:B)$. Regardless of how the $n/2$ qubits in A are chosen, the entanglement entropy corresponding to this bipartition is exactly the number of edges in E at the boundary between A and B in the lattice G. Moreover, it is easy to see that any choice of ordering will lead to a boundary of size $O(\sqrt{n})$ at this halfway point.

Now, we finally state our performance guarantee for the approximate Pauli-sampling algorithm described above.

Theorem B4 (Performance guarantee for the approximate Pauli-sampling algorithm). Let $1>\Delta>0,\delta>0$. Let ρ be a pure n-qubit state with magic $M_0(\rho)$ (or $M_1(\rho)$) and let there be a qubit ordering $\pi\in S_n$ with path entanglement $E_0^\pi(\rho)$. Then, there exists an algorithm sampling from a distribution q such that

$$\|q-p_\rho\|_{\mathrm{TV}}\le\Delta, \tag{B18}$$

with probability at least $1-\delta$ provided that

$$N=O\left(\frac{n^3 2^{2E_0^\pi(\rho)}2^{2M_0(\rho)}\log(1/\delta)}{\Delta^4}\right) \tag{B19}$$

or, alternatively,

$$N=O\left(\frac{n^3 2^{2E_0^\pi(\rho)}2^{4M_1(\rho)/\Delta}\log(1/\delta)}{\Delta^4}\right). \tag{B20}$$

Bell samples of the state $\rho^{\otimes 2}$ are taken. Producing a single sample takes $O(Nn^2)$ time in classical postprocessing on the obtained Bell-measurement data. The success probability of $1-\delta$ is with respect to the randomness in the outcomes of the Bell measurements.

Proof of Theorem B4. The general idea going into proving the theorem is that states with bounded magic give rise to Pauli distributions supported on "few" elements the respective probabilities of which are relatively large. Concretely, we obtain sufficient conditions on the state ρ such that the conditions of Theorem B3 are satisfied. Namely, we show that:

(a) An upper bound on the magic of the state translates into a lower bound on the overall probability weight distributed among Paulis with high expectation values $\alpha_\rho(x)^2$. Put differently, it translates into an upper bound on the CDF associated with the state, $F_\rho(\tau)$ (see Lemmas A2 and A3).

(b) An upper bound on the path entanglement leads to a lower bound on the marginals of the Pauli distribution (see Lemma A1).

Let $R:=2^{E_0^\pi(\rho)}$. It is the maximal Schmidt rank across the set of bipartitions corresponding to the assumed qubit ordering π. Now, consider the set

$$S=\{x\in\{0,1\}^{2n}:\mathrm{tr}\,(\rho P_x)^2>\gamma R\}. \tag{B21}$$

By Lemma A1, for all $x\in\{0,1\}^n$ in S, all their corresponding marginals satisfy $p_\rho(x_{1:k})\ge\gamma/2^k$. Hence, $S\subseteq S_\gamma$, where S_γ has been defined in Definition B2 and also $\mathfrak{f}(\gamma)=1-p_\rho(S_\gamma)$. This implies that $p_\rho(S)\le p_\rho(S_\gamma)$ and therefore $\mathfrak{f}(\gamma)\le 1-p_\rho(S)$.

By Lemma A2, we have that

$$\begin{aligned} &1-p_\rho(S)=F_\rho(\gamma R)\le 2^{M_0(\rho)}\gamma R\\ \Rightarrow\quad &\mathfrak{f}(\gamma)\le 1-p_\rho(S)\le 2^{M_0(\rho)}\gamma R. \end{aligned} \tag{B22}$$

Similarly, by Lemma A3, we find that

$$\begin{aligned} &1-p_\rho(S)=F_\rho(\gamma R)\le \frac{M_1(\rho)}{\log\left(\frac{1}{\gamma R}\right)}\\ \Rightarrow\quad &\mathfrak{f}(\gamma)\le 1-p_\rho(S)\le \frac{M_1(\rho)}{\log\left(\frac{1}{\gamma R}\right)}. \end{aligned} \tag{B23}$$

From Theorem B3, we have that the adapted ancestral-sampling algorithm samples from a distribution q within TV distance

$$\|q-p_\rho\|_{\mathrm{TV}}\le\mathfrak{f}(\gamma)+\exp\left(\frac{4\epsilon n}{\gamma}\right)-1. \tag{B24}$$

So, to obtain the bound in terms of $M_0(\rho)$ in Eq. (B19), we choose

$$\gamma=\frac{\Delta}{2\times 2^{M_0(\rho)}R},\quad \epsilon\le\frac{\gamma\Delta}{4\times 4n} \tag{B25}$$

or, alternatively, to obtain the bound in terms of $M_1(\rho)$ in Eq. (B20), we choose

$$\gamma = \frac{1}{2^{2M_1(\rho)/\Delta}R}, \quad \epsilon \leq \frac{\gamma\Delta}{4\times 4n}. \tag{B26}$$

These choices guarantee that $\mathfrak{f}(\gamma) \leq \Delta/2$ and $\exp(4\epsilon n/\gamma) \leq \exp(\Delta/4) \leq 1 + (\Delta/2)$ (using $\Delta < 1$ and $\exp(x) \leq 1 + 2x$ for $x \in [0, 1]$), and hence $\|q - p_\rho\|_{\mathrm{TV}} \leq \Delta$.

Finally, by Corollary B1, the number of copies N required to obtain $\epsilon/2^k$ additive estimates for all marginals with probability $1 - \delta$ is given as $N = O\left(n\log(1/\delta)/\epsilon^2\right)$. Hence,

$$N = O\left(\frac{n^3\log(1/\delta)}{\Delta^2\gamma^2}\right), \tag{B27}$$

which, combined with Eqs. (B25) and (B26), gives the stated complexities for N.

For the time-complexity bound, note that to produce a single sample x via the algorithm, $O(n)$ marginals have to be estimated from the Bell-measurement data. Following the proof of Lemma B4 in Appendix D—in particular, Eq. (D5)—we see that each estimated marginal is written as a sum of N terms, each a product of n numbers, and can be computed in time $O(nN)$. Hence, the total time is of the order $O(n^2N)$. ■

Note that the sample complexity of the algorithm scales exponentially with both the magic and the entanglement. Nonetheless, we can efficiently sample from the Pauli distribution of ρ as long as it is a state with magic and entanglement bounded by $O(\log(n))$.

Corollary B2. Let $\mathcal{C}$ be the class of n-qubit pure states such that, for every $\rho \in \mathcal{C}$, we have $M_0(\rho) = O(\log n)$ and $E_0^\pi(\rho) = O(\log(n))$ for all qubit orderings $\pi \in S_n$. Then, there exists an approximate Pauli-sampling algorithm, consuming poly(n) Bell measurements, with success probability at least $1 - \delta$, for $\mathcal{C}$.

Note that a stronger statement is possible if we are willing to drop computational efficiency. Namely, it suffices that there exists *some* ordering such that $E_0^\pi(\rho) = O(\log(n))$. The reason why we cannot provide guarantees on the computational complexity in this case is because it might be computationally inefficient to find such an ordering even if it exists.

5. Beyond pure states

In the previous section, we have established sample-complexity upper bounds for Pauli sampling of pure states in terms of their magic and entanglement. In this section, we argue that our algorithm extends naturally to mixed states. However, the characterization in terms of magic and entanglement gets more difficult.

First, we emphasize that the key ingredient to our bound derived in the previous section, Theorem B3, applies to any distribution p. In particular, Theorem B3 considers as input the access to estimates $\pi(x_{1:k})$ of the marginals $p(x_{1:k})$ of the distribution p that are within an additive error of at most $\epsilon/2^k$. As we show in Lemma D1, in the case of Pauli distributions p_ρ, we can obtain such estimates for both pure and mixed states, using Bell measurements, with an overhead in the sample complexity that scales inversely with the purity of the state ρ. Hence, Theorem B3 can be readily applied to Pauli distributions of mixed states as well.

However, on a high level, Theorem B3 captures the finding that the complexity of our algorithm depends on the size of the marginals of the distribution in question. For pure states we were able to connect the size of the marginals of the Pauli distribution p_ρ to the entanglement and magic of ρ through Lemmas A1–A3, respectively. Here, we have made use of the fact that there are pure-state entanglement and magic measures given directly in terms of the Pauli distribution p_ρ. For mixed states, however, it is no longer clear how to quantify magic and entanglement in terms of the Pauli distribution p_ρ (see, e.g., Ref. [32] for a review on mixed-state entanglement measures and Refs. [33,34] for mixed-state measures of magic). Hence, we also lack generalizations of the above-mentioned lemmas to mixed states. We leave this direction to future work.

While in the mixed-state case, the connection of the Pauli distribution to magic and entanglement is not so clear, we emphasize that we can still analyze the complexity of our approximate Pauli-sampling algorithm in terms of the Pauli distribution p_ρ of a mixed state ρ directly. This will be sufficient for many practical purposes, e.g., when trying to understand the robustness of our algorithm to experimental imperfections. To exemplify this, we demonstrate how to analyze the complexity of our Pauli sampling algorithm when applied to copies of a mixed state ρ that we assume to be the result of a Pauli noise channel Λ acting on an ideal pure state σ. That is, $\rho = \Lambda(\sigma)$ and we are interested in sampling from p_ρ. In this case, we can easily relate p_ρ to p_σ and hence obtain a performance guarantee in terms of the pure input state σ and the noise parameter of the Pauli channel.

Theorem B5. Let σ be a pure state and let Λ be a Pauli channel such that

$$\rho := \Lambda(\sigma) = (1-\xi)\sigma + \sum_{y\in\{0,1\}^{2n}\setminus\{0^{2n}\}} \xi_y P_y \sigma P_y, \tag{B28}$$

where, for all y, we have that $\xi_y \geq 0$ and $\sum_y \xi_y = \xi$ and $0 \leq \xi \leq 1$. Then, there exists an algorithm for sampling, with probability at least $1 - \delta$, from a distribution q_ρ such

that

$$\|p_\rho - q_\rho\|_{TV} \leq \Delta, \tag{B29}$$

for any $\Delta > 4\xi$, using

$$N = N_\sigma/(1-2\xi)^4 \tag{B30}$$

Bell samples of the state $\rho^{\otimes 2}$. Here, N_σ denotes the number of Bell samples necessary to achieve the same TV distance Δ if the algorithm had been applied to the ideal pure state σ instead of ρ.

Note that N_σ can be read off from Theorem B4. We can obtain Theorem B5 since we can directly relate the Pauli distributions p_ρ and p_σ via $p_\rho(x) \geq (1-\xi)^2 p_\sigma(x)$, which holds for all x. The full proof of Theorem B5 is deferred to Appendix G 1.

APPENDIX C: DISTRIBUTED INNER-PRODUCT ESTIMATION

In this appendix, we focus on the task of distributed inner-product estimation (IP): estimating the overlap $\mathrm{tr}(\rho\sigma)$ between states ρ and σ, with the states ρ and σ being distributed between two parties that can only use local quantum operations and classical communication.

One strategy for this task relies on either of the two parties (say, Alice) learning a full classical representation of her unknown state ρ and sending it to the other (Bob). He could then perform direct fidelity estimation [4] in order to estimate $\mathrm{tr}(\rho\sigma)$. Alternatively, he could also learn a classical representation of his state σ and then compute the overlap classically based on the classical representations of ρ and σ. Both approaches only result in efficient protocols when applied to certain classes of states such as stabilizer states, t-doped stabilizer states, and states that admit efficient MPS representations. Other approaches based on randomized measurements have been proposed [5], and in Ref. [6], general lower bounds on the sample complexity of distributed inner-product estimation have been proven.

Following the analysis of Ref. [6], we begin by providing a lower bound on the sample complexity of IP even for restricted classes of states, namely, those with $\omega(\log(n))$ entanglement and magic. We further provide two protocols for IP based on coordinated Pauli measurements on the states ρ and σ and provide sufficient conditions for the protocols to be efficient.

1. Hardness of IP for states with large magic and entanglement

We are mainly interested in the IP estimation problem defined in Definition 1. However, for establishing lower bounds, it will be convenient to work with a decision version of this problem. Again, following the formulation of Ref. [6], we define the *decisional inner-product estimation problem*, abbreviated as DIPE.

Definition C1 (Decisional inner-product estimation, DIPE). Alice and Bob are each given k copies of a pure state in $\mathbb{C}^d$. They are promised that one of the following two cases hold:

(1) Alice and Bob both have $|\phi\rangle^{\otimes k}$, where $|\phi\rangle$ is a uniformly random state vector from an ensemble $\mathcal{E}$.
(2) Alice has $|\phi\rangle^{\otimes k}$ and Bob has $|\psi\rangle^{\otimes k}$, where $|\phi\rangle$ and $|\psi\rangle$ are independent uniformly random state vectors in $\mathcal{E}$.

Their goal is to decide which case they are in with success probability at least 2/3, using an interactive protocol that involves local quantum operations and classical communication.

Anshu *et al.* [6] have considered an instance of DIPE in which the ensemble $\mathcal{E}$ is taken to be all quantum states in $\mathbb{C}^d$. For this general instance, they have shown a $k = \Omega(\sqrt{d})$ sample-complexity lower bound.

Theorem C1 (Lower bound on DIPE, Theorem 4 in Ref. [6]). $k = \Omega(\sqrt{d})$ copies are necessary for Alice and Bob to solve DIPE, when they are allowed arbitrary interactive protocols (or arbitrary LOCC operations).

Then, via a chain of reductions, they arrive at a sample-complexity lower bound $k = \Omega(\sqrt{d}/\epsilon)$ to solve IP up to additive error ϵ, assuming that the input state vectors $|\psi\rangle, |\phi\rangle$ are unrestricted in $\mathbb{C}^d$.

We generalize these lower bounds, both for DIPE and IP, to the restricted case in which the states $|\psi\rangle$ and $|\phi\rangle$ have bounded magic and entanglement.

We first consider an instance of DIPE in which the ensemble $\mathcal{E}$ is taken to be the subset phase states $\mathcal{E}_{f,S} = \left\{|\psi_{f,S}\rangle = 1/\sqrt{|S|}\sum_{x\in S}(-1)^{f(x)}|x\rangle\right\}_{f,S}$, defined in Appendix A 6. We call this instance DIPE*. Showing the hardness of DIPE* is rather straightforward, as it follows from the hardness of DIPE for the Haar measure and the statistical (and computational) indistinguishability of the ensemble $\mathcal{E}_{f,S}$ from the Haar measure.

Theorem C2 (Lower bound on DIPE).* $k = \omega(\mathrm{poly}(n))$ copies are necessary for Alice and Bob to solve DIPE*, when they are allowed arbitrary interactive protocols (or arbitrary LOCC operations).

Proof. We prove this statement by contradiction. Assume that there exists an interactive protocol to solve DIPE* using $k = O(\mathrm{poly}(n))$ copies. Call this protocol $\mathcal{A}$. Then,

$$\begin{aligned}
&\left|\Pr_{\substack{|\psi\rangle\sim\mathcal{E}_{f,S}\\|\phi\rangle\sim\mathcal{E}_{f,S}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]-\Pr_{\substack{|\psi\rangle\sim\text{Haar}\\|\phi\rangle\sim\text{Haar}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]\right|\\
&\leq\left|\Pr_{\substack{|\psi\rangle\sim\mathcal{E}_{f,S}\\|\phi\rangle\sim\mathcal{E}_{f,S}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]-\Pr_{\substack{|\psi\rangle\sim\text{Haar}\\|\phi\rangle\sim\mathcal{E}_{f,S}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]\right|\\
&\quad+\left|\Pr_{\substack{|\psi\rangle\sim\text{Haar}\\|\phi\rangle\sim\mathcal{E}_{f,S}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]-\Pr_{\substack{|\psi\rangle\sim\text{Haar}\\|\phi\rangle\sim\text{Haar}}}[\mathcal{A}(|\psi\rangle^{\otimes k},|\phi\rangle^{\otimes k})=1]\right|\\
&\leq 2\,\text{negl}(n),
\end{aligned}\tag{C1}$$

using the triangular inequality and the computational indistinguishability of $\mathcal{E}_{f,S}$ from the Haar measure (a consequence of their statistical indistinguishability; see Appendix A 5). This would imply that $\mathcal{A}$ would also solve DIPE for Haar-random inputs using $k=O(\text{poly}(n))$ copies, therefore contradicting Theorem C1. ■

From here, we extend this hardness result to IP*, the analog instance of IP in which input states are restricted to have bounded magic and entanglement. This is rather straightforward in the regime of additive error $\epsilon\in\left(0,\frac{1}{2}-O\left(1/\text{poly}(n)\right)\right)$.

Lemma C1. Suppose that Alice has input $|\phi\rangle^{\otimes k}$ and Bob has input $|\psi\rangle^{\otimes k}$, for arbitrary unknown pure-state vectors $|\phi\rangle,|\psi\rangle$ with bounded magic and entanglement, i.e., $M_\alpha(\psi)<f(n)$ (with $\alpha=O(1)$) and $S_\alpha(\psi)<g(n)$ (over all cuts $(A:B)$) where $f(n),g(n)=\omega(\log n)$. Then, $k=\omega(\text{poly}(n))$ copies are necessary for them to estimate $|\langle\phi|\psi\rangle|^2$ up to additive error $\epsilon\in\left(0,\frac{1}{2}-O\left(1/\text{poly}(n)\right)\right)$ with success probability $2/3$, when they are allowed arbitrary interactive protocols (or arbitrary LOCC operations).

Proof. This comes from the observation that DIPE* is a special case of IP*. To see this, note that the overlap between $|\psi\rangle$ and $|\phi\rangle$ in Definition C1 is either $|\langle\phi|\psi\rangle|^2=1$ in case 1, while in case 2,

$$\mathop{\mathbf{E}}_{|\psi\rangle,|\phi\rangle\sim\mathcal{E}_{f,S}}[|\langle\phi|\psi\rangle|^2]=\text{negl}(n).\tag{C2}$$

This last equation comes from the fact that for Haar-random states, we have

$$\mathop{\mathbf{E}}_{|\psi\rangle,|\phi\rangle\sim\text{Haar}}\left[|\langle\phi|\psi\rangle|^2\right]=\frac{1}{d},\tag{C3}$$

and by considering, e.g., the SWAP test between state vector $|\psi\rangle$ and $|\phi\rangle$ as an algorithm $\mathcal{A}$, we have

$$\begin{aligned}
&\left|\Pr_{|\psi\rangle,|\phi\rangle\sim\mathcal{E}_{f,S}}[\mathcal{A}(|\psi\rangle,|\phi\rangle)=1]\right.\\
&\qquad\left.-\Pr_{|\psi\rangle,|\phi\rangle\sim\text{Haar}}[\mathcal{A}(|\psi\rangle,|\phi\rangle)=1]\right|\\
&=\frac{1}{2}\left|\mathop{\mathbf{E}}_{|\psi\rangle,|\phi\rangle\sim\mathcal{E}_{f,S}}[|\langle\phi|\psi\rangle|^2]-\mathop{\mathbf{E}}_{|\psi\rangle,|\phi\rangle\sim\text{Haar}}[|\langle\phi|\psi\rangle|^2]\right|\\
&\leq\text{negl}(n).
\end{aligned}\tag{C4}$$

Therefore, by applying IP* with accuracy, say, $\epsilon=0.1$, one can distinguish between cases 1 and 2 with high probability and therefore solve DIPE*. Similarly, applying IP* with an arbitrary accuracy in $\epsilon\in\left(0,\frac{1}{2}-O\left(1/\text{poly}(n)\right)\right)$ would still allow us to differentiate an inner product of 1 from $\text{negl}(n)$, as both estimates would be separated from $\frac{1}{2}$. The $k=\omega(\text{poly}(n))$ lower bound then comes from Theorem C2. ■

From here, similarly to Anshu *et al.*, we can extend this lower bound to the case

$$\epsilon\in\left(\frac{1}{2}-O\left(\frac{1}{\text{poly}(n)}\right),1\right),\tag{C5}$$

by using the same chain of reductions from DIPE to IP. This, however, involves considering input-state vectors of the form

$$\left|\psi_{\epsilon,\theta}\right\rangle=\sqrt{1-\epsilon}e^{i\theta}\,|0\rangle\left|0^{\otimes n}\right\rangle+\sqrt{\epsilon}\,|1\rangle\,|\psi\rangle\,,\tag{C6}$$

$$\left|\phi_{\epsilon,\theta'}\right\rangle=\sqrt{1-\epsilon}e^{i\theta'}\,|0\rangle\left|0^{\otimes n}\right\rangle+\sqrt{\epsilon}\,|1\rangle\,|\phi\rangle\,,\tag{C7}$$

for θ and θ' independent uniformly random phases in $[0,2\pi]$, and $|\psi\rangle,|\phi\rangle\in\mathcal{E}_{f,S}$ as opposed to being Haar random in Ref. [6]. So, in order to extend Lemma C1, we

need to show that these states also have bounded magic and entanglement. We establish this in the following lemma, proven in Appendix F.

Lemma C2. Let $|\in\rangle\,\mathcal{E}_{f,S}$ be a subset phase state that satisfies $S_\alpha(|\psi\rangle) = O(\log|S|)$ over all cuts $(A:B)$ and $M_\alpha(|\psi\rangle) = O(\log|S|)$, for $\alpha \geq 0$. Then, for $\epsilon \in (0,1), \theta \in [0, 2\pi]$, its corresponding tilted state $\left|\psi_{\epsilon,\theta}\right\rangle$, defined in Definition C6, also satisfies $S_\alpha(\left|\psi_{\epsilon,\theta}\right\rangle) = O(\log|S|)$ over all cuts $(A:B)$ and $M_\alpha(\left|\psi_{\epsilon,\theta}\right\rangle) = O(\log|S|)$, for $\alpha \geq 0$.

With this, we arrive at our general lower-bound statement for IP*.

Theorem C3 (Formal version of Theorem 3). Suppose that Alice has input $|\phi\rangle^{\otimes k}$ and Bob has input $|\psi\rangle^{\otimes k}$, for two arbitrary unknown state vectors $|\phi\rangle$ and $|\psi\rangle$ with bounded magic and entanglement, i.e., $M_\alpha(\psi) < f(n)$ (with $\alpha = O(1)$) and $S_\alpha(\psi) < g(n)$ (over all cuts $(A:B)$), where $f(n), g(n) = \omega(\log n)$. Then, $k = \omega(\mathrm{poly}(n))$ copies are necessary for them to estimate $|\langle\phi|\psi\rangle|^2$ up to additive error $\epsilon \in (0,1)$ with success probability $2/3$, when they are allowed arbitrary interactive protocols (or arbitrary LOCC operations).

Proof. The same chain of reductions of Ref. [6, Sec. 5.3] from DIPE to IP apply from DIPE* to IP*.

This is because the only property of the states $|\psi\rangle, |\phi\rangle \in \mathcal{E}$ that they use is that $|\langle\phi|\psi\rangle| \leq \frac{1}{200}$ holds with high probability for d larger than some constant. We also have this property for $\mathcal{E} = \mathcal{E}_{f,S}$, following Eq. (C2). The bounded magic and entanglement of the states used in the reduction follows from Lemma C2. ■

We can restrict the class of states that we consider such that, in addition to having bounded magic and entanglement, they are furthermore efficiently preparable. Similarly to our lower bounds on the sample complexity of approximate Pauli sampling (Appendix B 1), we obtain a lower bound on the computational complexity by utilizing pseudorandom states.

Theorem C4. Suppose that Alice has input $|\phi\rangle^{\otimes k}$ and Bob has input $|\psi\rangle^{\otimes k}$, for two efficiently preparable unknown state vectors $|\phi\rangle$ and $|\psi\rangle$ with bounded magic and entanglement, i.e., $M_\alpha(\psi) < f(n)$ (with $\alpha = O(1)$) and $S_\alpha(\psi) < g(n)$ (over all cuts $(A:B)$), where $f(n), g(n) = \omega(\log n)$. Then, $k = \omega(\mathrm{poly}(n))$ samples necessary for them to estimate $|\langle\phi|\psi\rangle|^2$ up to additive error $\epsilon \in (0,1)$ with success probability $2/3$, when they are allowed arbitrary polynomial-time interactive protocols (or arbitrary polynomial-time LOCC operations), assuming the existence of OWFs.

Proof. The proof follows the same steps as the proof of Theorem C3 with the difference that instead of considering the ensemble of random subset phase states $\mathcal{E}_{f,S}$, we consider the ensemble of pseudorandom phase states $\tilde{\mathcal{E}}_{f,S}$ (see Appendix A 6). More precisely, we construct similar titled states to those in Definition C6 but for $|\psi\rangle, |\phi\rangle \in \tilde{\mathcal{E}}_{f,S}$, for which Lemma C2 can be straightforwardly generalized, and the same proof of Theorem C3 can be applied. ■

2. Algorithm for IP based on Pauli sampling

In this section, we describe our approach to distributed inner-product estimation based on Pauli sampling. Specifically, we will present two variants of a protocol based on coordinated Pauli measurements. Throughout this section, we will use the shorthand notation $\alpha_\rho(x) = \mathrm{tr}(\rho P_x)$. Essentially, our approach is motivated by the expansion of $\mathrm{tr}(\rho\sigma)$ in the Pauli basis as

$$\begin{aligned}\mathrm{tr}(\rho\sigma) &= \frac{1}{2^n}\sum_{x\in\{0,1\}^{2n}} \mathrm{tr}(\rho P_x)\,\mathrm{tr}(\sigma P_x)\\ &= \frac{1}{2^n}\sum_{x\in\{0,1\}^{2n}} \alpha_\rho(x)\alpha_\sigma(x). \end{aligned} \tag{C8}$$

Assuming that ρ and σ are pure states such that $\mathrm{tr}\left(\rho^2\right) = \mathrm{tr}\left(\sigma^2\right) = 1$, this can be further rewritten as

$$\begin{aligned}\mathrm{tr}(\rho\sigma) &= \sum_{x\in\{0,1\}^{2n}} \frac{\mathrm{tr}(\rho P_x)^2}{2^n}\frac{\mathrm{tr}(\sigma P_x)}{\mathrm{tr}(\rho P_x)}\\ &= \sum_{x\in\{0,1\}^{2n}} p_\rho(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}. \end{aligned} \tag{C9}$$

This expression suggests a protocol for Monte Carlo estimation of $\mathrm{tr}(\rho\sigma)$ through sampling x from the Pauli distribution p_ρ and averaging $\alpha_\sigma(x)/\alpha_\rho(x)$. We note that this is the same idea as in the popular direct-fidelity-estimation protocol [3,4]. As we discuss below in Appendix C 4, this estimation protocol can also be adapted to the distributed setting of this work and we will henceforth refer to this protocol as the *asymmetric* protocol. A down side of this protocol is that the estimator $\alpha_\sigma(x)/\alpha_\rho(x)$ is essentially unbounded. Thus, the implementation requires dealing with the event where a sampled x is such that $|\alpha_\rho(x)|$ is very small. Otherwise, the sample complexity of the protocol would explode, leading to an inefficient protocol. In fact, this is also the case for the DFE protocol mentioned above. We propose to address such "bad events" with very small $|\alpha_\rho(x)|$ by applying a filtering function that rescales the estimator (for more details, see Appendix C 4). However, rescaling the estimator, in turn, can lead to a large bias in the final estimate.

Here, as our main contribution, we propose a different protocol for the distributed estimation of $\mathrm{tr}(\rho\sigma)$ which we

will call the *symmetric* protocol. Crucially, this newly proposed protocol does not require filtering or postselecting bad events because it uses a bounded estimator.

The symmetric protocol is based on expressing $\mathrm{tr}(\rho\sigma)$ in terms of the mixture distribution $p_{\mathrm{mix}} = \frac{1}{2}\left(p_\rho + p_\sigma\right)$. In particular, one can check that for pure ρ and σ

$$f := \frac{1}{2}\left(1 + \mathrm{tr}\left(\rho\sigma\right)\right) = \sum_{x\in\{0,1\}^{2n}} p_{\mathrm{mix}}(x) g(x), \tag{C10}$$

where

$$g(x) = G\left(\alpha_\rho(x), \alpha_\sigma(x)\right) = \frac{1}{2}\frac{\left(\alpha_\rho(x) + \alpha_\sigma(x)\right)^2}{\alpha_\rho(x)^2 + \alpha_\sigma(x)^2}. \tag{C11}$$

This expression suggests that we can estimate f and hence $\mathrm{tr}(\rho\sigma)$ by sampling x from $p_{\mathrm{mix}} = \frac{1}{2}\left(p_\rho + p_\sigma\right)$ and averaging $g(x)$. In the following, we give an informal overview of how to implement this idea in the distributed setting of IP. For concreteness, we assume that Alice has access to copies of ρ and that Bob has access to copies of σ:

(1) In a first step, Alice and Bob both perform (approximate) Pauli sampling according to p_ρ and p_σ, respectively. This allows them to approximately sample from the mixture p_{mix} and obtain a list $L = \{x_1, \ldots, x_{N_1}\}$ of bit strings corresponding to N_1 many Pauli operators, which they communicate to each other.
(2) In the next step, Alice and Bob will both measure, N_2 times, each of the Paulis in L on copies of their respective states. From these measurement data, Alice computes estimates $\hat{\alpha}_\rho(x_i)$ for $\alpha_\rho(x_i)$ and Bob computes estimates $\hat{\alpha}_\sigma(x_i)$ for $\alpha_\sigma(x_i)$. Bob then sends his list of estimates $\{\hat{\alpha}_\sigma(x_1), \ldots \hat{\alpha}_\sigma(x_{N_1})\}$ to Alice.
(3) In a third step, Alice uses their combined data to obtain an estimate of $\mathrm{tr}(\rho\sigma)$. She does so by computing $\hat{g}(x_i) = G\left(\hat{\alpha}_\rho(x_i), \hat{\alpha}_\sigma(x_i)\right)$. She repeats this for all Paulis in the list and averages the estimates to obtain the final estimate of f and hence $\mathrm{tr}(\rho\sigma)$ via Eq. (C10).

We note that step 2 of the above protocol, which involves the estimation of expectation values of many Pauli operators, can be improved by making use of more clever measurement strategies, such as grouping strategies [51,52] or locally biased classical shadows [53].

We call this protocol symmetric because, in the first step, both Alice and Bob perform Pauli sampling according to their respective states in order to sample from the mixture p_{mix}. Additionally, in the third step, the quantity to be averaged, $G\left(\alpha_\rho(x), \alpha_\sigma(x)\right)$, is symmetric in $\alpha_\rho(x), \alpha_\sigma(x)$. Importantly, the function $G\left(\alpha_\rho(x), \alpha_\sigma(x)\right)$ is bounded between 0 and 1 and hence makes for a better-behaved estimator in comparison to $\alpha_\sigma(x)/\alpha_\rho(x)$.

In the following, we will analyze both the symmetric and the asymmetric protocols for IP in detail and state their performance guarantees. As mentioned at the beginning of Appendix B, depending on the class of states, different approaches to Pauli sampling are more suitable than others. To account for this, in the following, we will explain and analyze our protocols assuming black-box access to an algorithm performing approximate Pauli sampling.

On a high level, our findings are that both the asymmetric and the symmetric protocol run efficiently for input states with low magic. However, we find that that symmetric protocol is efficient for a larger class of states, including those with $M_1(\rho) = O(\log(n))$ (see Corollary C2), whereas the asymmetric protocol has the stricter requirement that $M_0(\rho) = O(\log(n))$ (see Corollary C3). This is an important distinction, as $M_0(\rho)$ is not a robust measure of magic, as we have remarked at the end of Appendix A 3. Importantly, this distinction also implies that the symmetric protocol can deal with states beyond those generated from Clifford circuits doped with a few T gates. Furthermore, due to the bounded estimator, the symmetric protocol exhibits a better scaling of the error with the number of copies required.

On the flip side, the asymmetric protocol also has its advantages. It is arguably much less demanding to implement, since only one party needs to perform (approximate) Pauli sampling, which typically requires two-copy measurements such as Bell measurements. Moreover, due to the asymmetric nature of the protocol, only one of the input states ρ and σ needs to have bounded magic in order for the protocol to run efficiently, namely, the state of the party performing Pauli sampling. The other state is essentially unconstrained.

3. The symmetric protocol

In the symmetric protocol, both Alice and Bob perform approximate Pauli sampling in the first step. In particular, they can sample from the mixture $\tilde{p}_{\mathrm{mix}} = \frac{1}{2}(\tilde{p}_\rho + \tilde{p}_\sigma)$ by sampling from $\tilde{p}_\rho$ and $\tilde{p}_\sigma$, respectively. Here, $\tilde{p}$ denotes an approximate version of p. The full estimation protocol is given in Algorithm 1. Note that the final estimator reads

$$f\left(N_1, N_2\right) := \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{g}(x_i) = \frac{1}{N_1}\sum_{i=1}^{N_1} G\left(\hat{\alpha}_\rho(x_i), \hat{\alpha}_\sigma(x_i),\right) \tag{C12}$$

where the x_i are drawn independent identically distributed (IID) from $\tilde{p}_{\mathrm{mix}}$ and $\hat{\alpha}_\rho(x_i)$ is an estimate of $\alpha_\rho(x) = \mathrm{tr}(\rho P_x)$ obtained from N_2 Pauli measurements. In the protocol, Alice and Bob each use a total of $N_1 \times N_2$ many copies of their respective states.

ALGORITHM 1. Symmetric protocol.

Input: k copies of unknown pure states ρ, σ.
Output: an estimate of $f = \frac{1}{2}(1 + \mathrm{tr}(\rho\sigma))$.

1: Alice and Bob sample N_1 times from the mixture distribution $\tilde{p}_{\mathrm{mix}}$ and obtain $L = \{x_1, \ldots, x_{N_1}\}$.
2: **for all** $x \in L$ **do**
3: Alice measures N_2 times P_x on ρ and obtains estimate $\hat{\alpha}_\rho(x)$.
4: Bob measures N_2 times P_x on σ and obtains estimate $\hat{\alpha}_\sigma(x)$.
5: **end for**
6: **Return** $\frac{1}{N_1}\sum_{i=1}^{N_1} G\left(\hat{\alpha}_\rho(x_i), \hat{\alpha}_\sigma(x_i)\right)$.

The estimator in Eq. (C12), although bounded, is a not an unbiased estimator because the function $G : [-1,1]^2 \setminus \{0,0\} \to [0,1]$ given by

$$G(u,v) = \frac{1}{2}\frac{(u+v)^2}{u^2+v^2} \tag{C13}$$

is nonlinear. This is why we have to take into account properties of $\tilde{p}_{\mathrm{mix}}$ when bounding the error in the estimate $f(N_1, N_2)$, as we briefly describe below (for more details, see Appendix H 1).

We note that the function G varies rapidly for arguments (u,v) around the origin, i.e., when $\|(u,v)\|_2$ is small. Hence, $G\left(\hat{\alpha}_\rho(x_i), \hat{\alpha}_\sigma(x_i)\right)$ will be close to $G\left(\alpha_\rho(x_i), \alpha_\sigma(x_i)\right)$, only if x_i is such that at least one of $|\alpha_\rho(x_i)|, |\alpha_\sigma(x_i)|$ is reasonably large. In Appendix A 4, we have shown, on a high level, that the more magical a state ρ is, the more weight can accumulate in the tail of the distribution p_ρ. This tail probability is captured by the behavior of the CDF F_ρ introduced in Definition A1. Thus, the smaller the magic of the states, the smaller is the probability that we will sample a Pauli P_{x_i} such that both $|\alpha_\rho(x_i)|, |\alpha_\sigma(x_i)|$ are small and therefore the smaller is the error that we make (for a fixed amount of samples taken). This is the intuitive reason for why F_ρ and F_σ appear in the error contribution of our estimate.

We now state our performance guarantees for the symmetric protocol in terms of the CDFs F_ρ and F_σ. The proof of the following theorem is given in Appendix H.

Theorem C5 (Performance guarantee of the symmetric protocol). Let $\epsilon_1 > 0, \epsilon_2 > 0$ and $\delta > 0$. Let $\tilde{p}_{\mathrm{mix}}$ be a distribution such that $\|p_{\mathrm{mix}} - \tilde{p}_{\mathrm{mix}}\|_{\mathrm{TV}} < \Delta$. Let $f(N_1, N_2)$ be our estimate for $f = \frac{1}{2}(1 + \mathrm{tr}(\rho\sigma))$ as defined in Eq. (C12). Then,

$$|f(N_1,N_2) - f| \leq 2\epsilon_1 + 2\sqrt{\epsilon_2} + \frac{F_\rho(\epsilon_2)}{2} + \frac{F_\sigma(\epsilon_2)}{2} + 2\Delta \tag{C14}$$

with probability at least $1-\delta$ provided that

$$N_1 \geq (2\epsilon_1^2)^{-1}\ln(8/\delta), \tag{C15}$$

$$N_2 \geq (2/\epsilon_2^2)\ln(8N_1/\delta). \tag{C16}$$

Here, F_ρ and F_σ denote the CDFs previously introduced in Definition A1.

While the functions F_ρ and F_σ precisely capture the performance of the symmetric protocol, they are also, in general, as complex as their underlying distributions p_ρ and p_σ. To provide a more coarse-grained and informative statement, we now formulate a corollary to Theorem C5 in terms of the stabilizer entropies, which associate a single number to a state.

Corollary C1 (Performance guarantee of the symmetric protocol in terms of M_0). Let ρ and σ be n-qubit pure states and let $\epsilon > 0$. Let $p_{\mathrm{mix}} = \frac{1}{2}\left(p_\rho + p_\sigma\right)$ and $\tilde{p}_{\mathrm{mix}}$ be a distribution such that $\|p_{\mathrm{mix}} - \tilde{p}_{\mathrm{mix}}\|_{\mathrm{TV}} < \Delta$. Let $M_0 = \max(M_0(\rho), M_0(\sigma))$. Let $f(N_1, N_2)$ be our estimate for $f = \frac{1}{2}(1 + \mathrm{tr}(\rho\sigma))$ as defined in Eq. (C12). Then,

$$|f(N_1,N_2) - f| \leq 5\epsilon + 2\Delta \tag{C17}$$

with probability at least $1-\delta$ provided that

$$N_1 \geq \frac{1}{2\epsilon^2}\log(8/\delta), \tag{C18}$$

$$N_2 \geq \frac{2}{\min(\epsilon^4, \epsilon^2 2^{-2M_0})}\log(8N_1/\delta). \tag{C19}$$

Proof. Let us write $M_0 = \max(M_0(\rho), M_0(\rho))$. Consider Theorem C5 and recall from Lemma A2 that $F_\rho(\epsilon_2) \leq 2^{M_0(\rho)}\epsilon_2$ and $F_\sigma(\epsilon_2) \leq 2^{M_0(\sigma)}\epsilon_2$, so that $(F_\rho(\epsilon_2)/2) + (F_\sigma(\epsilon_2)/2) \leq 2^{M_0}\epsilon_2$ and Theorem C5 implies that

$$|f(N_1,N_2) - \mathrm{tr}(\rho\sigma)| \leq 2\epsilon_1 + 2\sqrt{\epsilon_2} + 2^{M_0}\epsilon_2 + 2\Delta \tag{C20}$$

with probability at least $1-\delta$ provided that $N_1 \geq (2\epsilon_1^2)^{-1}\log(8/\delta)$ and $N_2 \geq (2/\epsilon_2^2)\log(8N_1/\delta)$. Then, we take $\epsilon_1 = \epsilon$ and $\epsilon_2 = \min(\epsilon^2, \epsilon/2^{M_0})$ to obtain

$$|f(N_1,N_2) - \mathrm{tr}(\rho\sigma)| \leq 5\epsilon + 2\Delta \tag{C21}$$

with probability at least $1-\delta$ provided that $N_1 \geq (2\epsilon^2)^{-1}\log(8/\delta)$ and $N_2 \geq (2/\min(\epsilon^4, \epsilon^2 2^{-2M_0})\log(8N_1/\delta)$. ■

Corollary C2 (Performance guarantee of the symmetric protocol in terms of M_1—formal version of Theorem 4). Let ρ and σ be n-qubit pure states and let $\epsilon > 0$. Let $p_{\mathrm{mix}} = \frac{1}{2}\left(p_\rho + p_\sigma\right)$ and $\tilde{p}_{\mathrm{mix}}$ be a distribution such that $\|p_{\mathrm{mix}} - \tilde{p}_{\mathrm{mix}}\|_{\mathrm{TV}} < \Delta$. Let $M_1 = \max(M_1(\rho), M_1(\sigma))$. Let

$f(N_1, N_2)$ be our estimate for $f = \frac{1}{2}(1 + \mathrm{tr}(\rho\sigma))$ as defined in Eq. (C12). Then,

$$|f(N_1, N_2) - f| \leq 5\epsilon + 2\Delta \tag{C22}$$

with probability at least $1 - \delta$ provided that

$$N_1 \geq \frac{1}{2\epsilon^2}\log(8/\delta), \tag{C23}$$

$$N_2 \geq \frac{2}{\min(\epsilon^4, \epsilon^2 2^{-2M_1/\epsilon})}\log(8N_1/\delta). \tag{C24}$$

Proof. Let us write $M_1 = \max(M_1(\rho), M_1(\sigma))$. In this case, we have, from Lemma A3, that $F_\rho(\epsilon_2) \leq M_1(\rho)/\log_2(1/\epsilon_2)$ and $F_\sigma(\epsilon_2) \leq M_1(\rho)/\log_2(1/\epsilon_2)$, so that we have that $(F_\rho(\epsilon_2)/2) + (F_\sigma(\epsilon_2)/2) \leq M_1/\log_2(1/\epsilon_2)$ and Theorem C5 implies that

$$|f(N_1, N_2) - \mathrm{tr}(\rho\sigma)| \leq 2\epsilon_1 + 2\sqrt{\epsilon_2} + \frac{M_1}{\log_2(1/\epsilon_2)} + 2\Delta \tag{C25}$$

with probability at least $1 - \delta$ provided that $N_1 \geq (2\epsilon_1^2)^{-1}\log(8/\delta)$ and $N_2 \geq (2/\epsilon_2^2)\log(8N_1/\delta)$. Now choose $\epsilon_1 = \epsilon$ and $\epsilon_2 = \min(\epsilon^2, 2^{-M_1/\epsilon})$ to obtain

$$|f(N_1, N_2) - \mathrm{tr}(\rho\sigma)| \leq 5\epsilon + 2\Delta \tag{C26}$$

with probability at least $1 - \delta$ provided that $N_1 \geq (2\epsilon^2)^{-1}\log(8/\delta)$ and $N_2 \geq (2/\min(\epsilon^4, \epsilon^2 2^{-2M_1/\epsilon})\log(8N_1/\delta)$. ■

Remark C1. The conditions of low magic $M_0(\psi) = O(\log n)$ and/or $M_1(\psi) = O(\log n)$ for $\psi = \rho, \sigma$ are not necessary. For instance, given a desired error τ, by Theorem C5, the following condition is clearly also sufficient: $F_\psi(\epsilon) \leq \tau$ for a sufficiently small $\epsilon = 1/\mathrm{poly}(n)$. This however, does not imply that the magics $M_0(\psi)$ and/or $M_1(\psi)$ are $O(\log n)$. In Appendix I, we give the example of a pure state ψ with $M_0(\psi)$ larger than $n - 1$, $M_1(\psi)$ larger than $\sqrt{n}$ and $F(\tau^2) \leq 2\tau$ for $\tau = 1/\sqrt{n}$.

4. The asymmetric protocol

In this section, we will analyze the behavior of the asymmetric protocol as described earlier. A complete description of this protocol is given in Algorithm 2. The asymmetric protocol is closer to the *direct fidelity estimation* [4], protocol as only one party is required to sample Paulis. As we will see, the main drawback of the asymmetric protocol is that the naive protocol uses an estimator that is *a priori* unbounded. This has indeed already been noted and discussed in the setting of direct fidelity estimation. One strategy to overcome this issue has been suggested in Refs. [3,4] and consists of postselecting the sampled strings x on having $|\alpha_\rho(x)| > \lambda$ for some threshold λ of our choice. Here, we take a slightly different approach, in which we apply a "filtering" function to the estimates $\hat{\alpha}_\rho(x_i)$, ensuring that the overall estimate remains bounded. More precisely, in the estimation protocol given in Algorithm 2, we use the estimator $f(N_1, N_\sigma, N_\rho)$ that reads

ALGORITHM 2. Asymmetric protocol.

Input: k copies of unknown pure states ρ, σ.
Output: An estimate of $f = \mathrm{tr}(\rho\sigma)$.
1: Alice samples N_1 many bit strings x from $\tilde{p}_\rho$ and obtains $L = \{x_1, \ldots, x_{N_1}\}$.
2: **for all** $x \in L$ **do**
3: Alice measures N_ρ times P_x on ρ and obtains an estimate $\hat{\alpha}_\rho(x)$.
4: Bob measures N_σ times P_x on σ and obtains an estimate $\hat{\alpha}_\sigma(x)$.
5: Bob computes $\frac{\hat{\alpha}_\sigma(x)}{s_\lambda(\hat{\alpha}_\rho(x))}$.
6: **end for**
7: **Return** $\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\hat{\alpha}_\sigma(x_i)}{s_\lambda(\hat{\alpha}_\rho(x_i))}$.

$$f(N_1, N_\sigma, N_\rho) := \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\hat{\alpha}_\sigma(x_i)}{s_\lambda(\hat{\alpha}_\rho(x_i))}, \tag{C27}$$

with the function s_λ defined as

$$s_\lambda(z) = \begin{cases} z, & |z| > \lambda, \\ \mathrm{sign}(z)\lambda, & |z| \leq \lambda, \end{cases} \tag{C28}$$

and where the x_i are drawn IID from $\tilde{p}_\rho$, $\hat{\alpha}_\rho(x_i)$ is an estimate of $\alpha_\rho(x) = \mathrm{tr}(\rho P_x)$ obtained from N_ρ Pauli measurements, and $\hat{\alpha}_\sigma(x_i)$ is an estimate of $\alpha_\sigma(x) = \mathrm{tr}(\sigma P_x)$ obtained from N_σ Pauli measurements. In this protocol, Alice and Bob use a total of $N_1 \times N_\rho$ and $N_1 \times N_\sigma$ copies of their states, respectively. For comparison with the symmetric protocol, one should compare N_2 with $\max(N_\sigma, N_\rho)$, although it is interesting to note that due to the asymmetry of this protocol, we generally have $N_\rho > N_\sigma$.

We now state the performance guarantee for the asymmetric protocol. Importantly, in contrast to the symmetric protocol, the error depends directly on $M_0(\rho)$.

Theorem C6 (Performance guarantee of the asymmetric protocol). Let $\epsilon_\rho > 0$, $\epsilon_\sigma > 0$, $\delta > 0$, $\gamma > 0$, and $\lambda > 0$. Let $\tilde{p}_\rho$ be a distribution such that $||p_\rho - \tilde{p}_\rho||_{\mathrm{TV}} < \Delta$. We write $f(N_1, N_\sigma, N_\rho)$ for our estimate of $\mathrm{tr}(\rho\sigma)$, using N_1 samples from $\tilde{p}_\rho$, and N_ρ, N_σ measurements per sample for ρ and σ, respectively. One has

$$\begin{aligned}&|f(N_1, N_\sigma, N_\rho) - \mathrm{tr}(\rho\sigma)| \\ &\leq \frac{1}{\lambda}\left(\epsilon_\sigma + \frac{\epsilon_\rho}{\lambda}\right) + \gamma + \frac{\Delta}{\lambda} + \frac{F_\rho(\lambda^2)}{\lambda} + 2^{M_0(\rho)}\lambda\end{aligned} \tag{C29}$$

with probability at least $1-\delta$ provided that [54]

$$N_1 \geq \frac{3(1+\Delta/\lambda^2)}{\gamma^2\delta}, \tag{C30}$$

$$N_\sigma \geq \frac{2}{\epsilon_\sigma^2}\ln\left(\frac{6N_1}{\delta}\right), \tag{C31}$$

$$N_\rho \geq \frac{2}{\epsilon_\rho^2}\ln\left(\frac{6N_1}{\delta}\right), \tag{C32}$$

where $F_\rho(\cdot)$ denotes the previously introduced CDF of ρ and $M_0(\rho)$ is its stabilizer entropy for $\alpha = 0$.

Analogously to the symmetric protocol, we have the following corollary.

Corollary C3 (Performance guarantee of the asymmetric protocol in terms of M_0). Let $\epsilon > 0, \delta > 0$. Assuming that the state ρ has bounded magic scaling as $M_0(\rho) \leq c\log(n)$ and that we can sample Paulis from a distribution $\tilde{p}_\rho$ such that $||p_\rho - \tilde{p}_\rho||_{\mathrm{TV}} < O(\epsilon^2/n^{2c})$, then our asymmetric protocol returns an estimate $f(N_1, N_\sigma, N_\rho)$ of $\mathrm{tr}(\rho\sigma)$ using N_1 (approximate) Pauli samples and N_ρ, N_σ measurements per sample for ρ and σ, respectively, that satisfies

$$\left|f(N_1, N_\sigma, N_\rho) - \mathrm{tr}(\rho\sigma)\right| \leq \epsilon \tag{C33}$$

with probability at least $1-\delta$,

$$N_1 = O\left(\frac{1}{\epsilon^2\delta}\right), \tag{C34}$$

$$N_\sigma = O\left(\frac{n^{4c}\log\left(N_1\delta^{-1}\right)}{\epsilon^4}\right), \tag{C35}$$

$$N_\rho = O\left(\frac{n^{8c}\log\left(N_1\delta^{-1}\right)}{\epsilon^6}\right). \tag{C36}$$

We defer the proof of Theorem C6 to Appendix H 2. The proof follows a similar approach to the one for the symmetric protocol, with one main difference, namely, that there are several error terms of the form of approximately $1/\lambda$, which can be attributed to the fact that the range of the estimator is $[-1/\lambda, 1/\lambda]$.

APPENDIX D: ESTIMATION FROM BELL-MEASUREMENT DATA

In this appendix, for the sake of the work being self-contained, we explain how to estimate the marginals of the Pauli distribution from Bell-measurement data obtained via Bell sampling. A similar exposition can be found in Ref. [49, Appendix E].

Recall that a single round of Bell sampling on two copies $\rho^{\otimes 2}$ results in a bit string $y = (y_1, \ldots, y_n)$ of length $2n$ corresponding to a Bell state vector $|P_y\rangle$. For convenience, we define the 4×4 matrix $\boldsymbol{M}$ with entries $M_{a,b} = \mathrm{tr}\left(P_a^{\otimes 2}|P_b\rangle\langle P_b|\right)$, where a and b are two-bit strings labeling the single-qubit Paulis and two-qubit Bell basis:

$$\boldsymbol{M} = \left(\begin{array}{c|cccc} & \Phi^+ & \Phi^- & \Psi^+ & \Psi^- \\ \hline I & +1 & +1 & +1 & +1 \\ Z & +1 & +1 & -1 & -1 \\ X & +1 & -1 & +1 & -1 \\ Y & -1 & +1 & +1 & -1 \end{array}\right) = \left(\begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 00 & +1 & +1 & +1 & +1 \\ 01 & +1 & +1 & -1 & -1 \\ 10 & +1 & -1 & +1 & -1 \\ 11 & -1 & +1 & +1 & -1 \end{array}\right). \tag{D1}$$

Furthermore, we define the vector $\boldsymbol{s}$ with entries $s_a = \mathrm{tr}\left(\mathbb{S}|P_a\rangle\langle P_a|\right)$ so that

$$\boldsymbol{s} = \begin{pmatrix} +1 & +1 & +1 & -1 \end{pmatrix}. \tag{D2}$$

We now prove the following lemma asserting that we can estimate all the marginals of p_ρ from a few Bell-measurement data.

Lemma D1 (Lemma B4 restated). Let $\epsilon > 0$ and $\delta > 0$ and let ρ be a quantum state on n qubits. Then, $N = O\left(n\log(1/\delta)/(\epsilon\mathrm{tr}(\rho^2))^2\right)$ pairs of copies $\rho\otimes\rho$ suffice to produce, with probability $1-\delta$, estimates $\pi_\rho(x_{1:k})$ such that

$$\left|\pi_\rho(x_{1:k}) - p_\rho(x_{1:k})\right| \leq \frac{\epsilon}{2^k} \tag{D3}$$

for all $x \in \{0,1\}^{2n}$ and all $1 \leq k \leq n$.

Proof. Let $\{y^{(1)}, \ldots, y^{(N)}\}$ be the N bit strings of length $2n$ that were sampled independently via Bell sampling on $\rho^{\otimes 2}$. In this proof, it is useful to introduce the following quantities for the sake of clarity. We introduce

$$Q_k(x_{1:k}) = \langle P_{x_{1:k}}^{\otimes 2}\otimes S^{\otimes n-k}\rangle_{\rho\otimes\rho}, \tag{D4}$$

$$\hat{Q}_k(x_{1:k}) = \frac{1}{N}\sum_{i=1}^{N}\left(\prod_{j=1}^{k} M_{x_j, y_j^{(i)}}\prod_{l=k+1}^{n} s_{y_l^{(i)}}\right), \tag{D5}$$

i.e., $Q_k(x_{1:k})$ is the ideal expectation value $\langle P_{x_{1:k}}^{\otimes 2}\otimes S^{\otimes n-k}\rangle_{\rho\otimes\rho}$ and $\hat{Q}_k(x_{1:k})$ is our estimate of $Q_k(x_{1:k})$ computed from N Bell-measurement outcomes obtained from copies of $\rho\otimes\rho$. This is an empirical estimate, because all the operators $P_{x_{1:k}}^{\otimes 2}\otimes S^{\otimes n-k}$ acting on two copies commute and a common eigenbasis is precisely the Bell basis with

eigenvalues $\prod_{j=1}^{k} M_{x_j,y_j^{(i)}} \prod_{l=k+1}^{n} s_{y_l^{(i)}}$. Note that the previous expressions also work for $k=0$, so that $Q_0 := \mathrm{tr}\rho^2$ is the purity and $\hat{Q}_0$ is our estimate for the purity. Both are again obtained from Bell measurements on N copies of $\rho\otimes\rho$. Explicitly,

$$Q_0 = \langle S^{\otimes n}\rangle_{\rho\otimes\rho}, \tag{D6}$$

$$\hat{Q}_0 = \frac{1}{N}\sum_{i=1}^{N}\left(\prod_{l=1}^{n} s_{y_l^{(i)}}\right). \tag{D7}$$

Now, any single $Q_k(x_{1:k})$ can be estimated to precision ϵ_1 with probability at least $1-\delta$ using $N = O(\log(1/\delta)/\epsilon_1^2)$ copies of $\rho^{\otimes 2}$. Hence, by a union bound over all $x\in\{0,1\}^{2n}$, using $N = O(n\log(1/\delta)/\epsilon_1^2)$ copies suffices to estimate the quantities $Q_k(x_{1:k})$ for all $x\in\{0,1\}^{2n}$ and $1\le k\le n$ simultaneously up to an additive error smaller than ϵ_1 with probability larger than $1-\delta$. Now, it remains to study how the error ϵ_1 in $\hat{Q}_k(x_{1:k})$ propagates in the estimates $\pi(x_{1:k})$ of the marginals $p(x_{1:k})$. Here, recall that the marginals and our estimates are given by

$$p(x_{1:k}) = \frac{Q_k(x_{1:k})}{2^k Q_0}, \tag{D8}$$

$$\pi(x_{1:k}) = \frac{\hat{Q}_k(x_{1:k})}{2^k \hat{Q}_0}. \tag{D9}$$

Now, we shall use the simple inequalities,

$$\frac{Q_k(x_{1:k})-\epsilon_1}{Q_0+\epsilon_1} \le 2^k\pi(x_{1:k}) \le \frac{Q_k(x_{1:k})+\epsilon_1}{Q_0-\epsilon_1}, \tag{D10}$$

and let $\epsilon_1 = \epsilon Q_0/4$, where $0<\epsilon<1$ is the final desired precision. The proof will be completed by showing that

$$\begin{aligned}\frac{Q_k(x_{1:k})}{Q_0} - \epsilon &\le \frac{Q_k(x_{1:k})-\epsilon_1}{Q_0+\epsilon_1} \le 2^k\pi(x_{1:k}) \\ &\le \frac{Q_k(x_{1:k})+\epsilon_1}{Q_0-\epsilon_1} \le \frac{Q_k(x_{1:k})}{Q_0} + \epsilon\end{aligned} \tag{D11}$$

when $\epsilon_1 = \epsilon Q_0/4$. Putting everything together, the upper bound reads

$$\begin{aligned}\frac{Q_k+\epsilon_1}{Q_0-\epsilon_1} &= \frac{Q_k-\epsilon_1+2\epsilon_1}{Q_0-\epsilon_1} \\ &= \frac{Q_k-\epsilon_1}{Q_0-\epsilon_1} + \frac{2\epsilon_1}{Q_0-\epsilon_1} \le \frac{Q_k}{Q_0} + \frac{2\epsilon_1}{Q_0}\times\frac{1}{1-\epsilon_1/Q_0} \\ &\le \frac{Q_k}{Q_0} + \frac{2\epsilon_1}{Q_0}\times\left(1+\frac{2\epsilon_1}{Q_0}\right) \le \frac{Q_k}{Q_0}+\epsilon.\end{aligned} \tag{D12}$$

For the lower bound, we have, analogously,

$$\begin{aligned}\frac{Q_k-\epsilon_1}{Q_0+\epsilon_1} &= \frac{Q_k+\epsilon_1-2\epsilon_1}{Q_0+\epsilon_1} = \frac{Q_k+\epsilon_1}{Q_0+\epsilon_1} - \frac{2\epsilon_1}{Q_0+\epsilon_1} \\ &\ge \frac{Q_k}{Q_0} - \frac{2\epsilon_1}{Q_0}\times\frac{1}{1+\epsilon_1/Q_0} \ge \frac{Q_k}{Q_0}-\epsilon.\end{aligned} \tag{D13}$$

Recalling that $2^k p(x_{1:k}) = Q_k(x_{1:k})/Q_0$, we have shown that $2^k\pi(x_{1:k})$ approximates $2^k p(x_{1:k})$ within additive error ϵ with $N = O(n\log(1/\delta)/\epsilon_1^2)$ copies, where $\epsilon_1 = \epsilon Q_0/4$. In other words, a sufficient number of copies is $N = O(n\log(1/\delta)/(\epsilon Q_0)^2)$, as claimed. ■

APPENDIX E: PROOFS OF APPENDIX B 1: INDISTINGUISHABLE ENSEMBLES WITH IMAGINARITY GAP

In this appendix, we prove Lemmas B2 and B3. In the following lemma, we give a construction of two statistically indistinguishable ensembles such that states from the one ensemble are real while states from the other have high imaginarity $I(|\psi\rangle)$.

Lemma E1 (Lemma B2 restated). Let $U = \bigotimes_{i=1}^{n} C_i$ for C_i being random single-qubit Clifford unitaries. Then define the ensemble $\mathcal{E}_U = \{U|\psi_{f,S}\rangle \mid |\psi_{f,S}\rangle \in \mathcal{E}_{f,S}\}$, where $\mathcal{E}_{f,S}$ is an ensemble of random subset phase states. Denote $|\psi^*\rangle$ as the conjugate state in the computational basis. We have the following list of results:

(1) *Imaginarity:* $\mathbf{Pr}_{|\psi\rangle\sim\mathcal{E}_U}[I(|\psi\rangle)>1/100] = 1-\mathrm{negl}(n)$.
(2) *Statistical indistinguishability:* For $|S| = \omega(\mathrm{poly}(n))$, the two ensembles $\mathcal{E}_U$ and $\mathcal{E}_{f,S}$ are statistically indistinguishable.

Proof. To prove the first point, let $\psi = |\psi\rangle\langle\psi|$ and note that $\psi^* = \psi^T$, where T denotes the transposition. Then, using the replica trick, we can rewrite the overlap as

$$|\langle\psi|\psi^*\rangle|^2 = \mathrm{tr}(\psi\psi^*) = \mathrm{tr}(\mathbb{S}\psi\otimes\psi^T) = \mathrm{tr}(\mathbb{S}^{T_2}\psi^{\otimes 2}), \tag{E1}$$

where $\mathbb{S}$ is the SWAP operator in $\mathcal{H}^{\otimes 2}$, and we use the invariance of the trace under partial transpose T_2, i.e., the transposition on the second copy of ψ only. Note that

$$\mathbb{S}^{T_2} = 2^n|\Omega\rangle\langle\Omega|^{\otimes n}, \tag{E2}$$

where $|\Omega\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$ as defined above. Let us average over the choice of U using standard Haar-measure

techniques. We find that for any pure state $|\psi\rangle$, it holds that

$$\begin{aligned}&\mathbb{E}_U[|\langle\psi|\psi^*\rangle|^2]\\&=\left(\prod_{i=1}^{n}\mathbb{E}_{C_i}\right)\mathrm{tr}\left[\bigotimes_{i=1}^{n}(C_i^{\otimes 2}\,|\Omega\rangle\,\langle\Omega|\,C_i^{\dagger\otimes 2})\psi^{\otimes 2}\right]\\&=\frac{1}{3^n}\mathrm{tr}\left[\bigotimes_{i=1}^{n}(I_i+\mathbb{S}_i)\psi^{\otimes 2}\right]\leq\left(\frac{2}{3}\right)^n.\end{aligned}\tag{E3}$$

The last inequality can be derived by upper bounding each term by 1. Using Markov's inequality, we have that, for any $|\psi\rangle$, it holds that

$$\mathop{\mathbf{Pr}}_{U}\left[|\langle\psi|\psi^*\rangle|^2<\frac{1}{100}\right]\geq 1-100\left(\frac{2}{3}\right)^n.\tag{E4}$$

Now, note that we can write $\psi=U\psi_{f,S}U^\dagger$ for every $|\psi\rangle\in\mathcal{E}_U$, where $\psi_{f,S}=|\psi_{f,S}\rangle\langle\psi_{f,S}|$. Since $\psi_{f,S}\sim\mathcal{E}_{f,s}$ and U are drawn independently, we find

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{|\psi\rangle\sim\mathcal{E}_U}[I(|\psi\rangle)>1/100]\\&=\sum_{f,S}\mathop{\mathbf{Pr}}_{|\psi_{f,S}\rangle\sim\mathcal{E}_{f,s}}[\psi_{f,S}]\,\mathop{\mathbf{Pr}}_{U}[I(U\psi_{f,S}U^\dagger)>1/100]\\&\geq 1-100\left(\frac{2}{3}\right)^n.\end{aligned}\tag{E5}$$

This shows that while the ensemble $\mathcal{E}_{f,s}$ is strictly real, the Clifford-conjugated version $\mathcal{E}_U$ is highly imaginary.

To show the second point, i.e., that $\mathcal{E}_U$ is statistically indistinguishable from $\mathcal{E}_{f,S}$, we can use Lemmas A4 and A5 by setting $|S|=\omega(\mathrm{poly}(n))$. Using Lemmas A6 and A7, we conclude that $\tilde{\mathcal{E}}_U$ is computationally indistinguishable from $\mathcal{E}_{f,S}$ for $|S|=\omega(\mathrm{poly}(n))$. ■

In the following lemma, we show that given samples from a distribution close to the Pauli distribution, this allows us to estimate the imaginarity of the state using a number of samples that scales inverse polynomially with the estimation error.

Lemma E2 (Lemma B3 restated). Let $\rho=|\psi\rangle\langle\psi|$ be a pure quantum state and let p_ρ be its associated Pauli distribution. Then, given black-box access to an algorithm for sampling from a distribution q such that $\|p_\rho-q\|_{\mathrm{TV}}\leq\Delta$, for $0\leq\Delta<1$, there exists an efficient algorithm to estimate the imaginarity $I(|\psi\rangle)=1-|\langle\psi|\psi^*\rangle|^2$ within additive error $\epsilon>\Delta$ and failure probability δ using $2/(\epsilon-\Delta)^2\ln 2/\delta$ samples from q.

Proof. Let us first expand $I(\psi)$ in the Pauli basis:

$$I(\psi)=\frac{1}{d}\sum_x\mathrm{tr}(P_x\psi)\mathrm{tr}(P_x\psi^*)=\sum_x\frac{\mathrm{tr}(P_x\psi^*)}{\mathrm{tr}(P_x\psi)}p_\psi(x).\tag{E6}$$

Note that the components $\mathrm{tr}(P_x\psi^*)$ can be easily computed by $\mathrm{tr}(P_x\psi)$ via

$$\mathrm{tr}(P_x\psi^*)=\mathrm{tr}(P_x^T\psi)=(-1)^{\sum_{i=1}^n s_i^x t_i^x}\mathrm{tr}(P_x\psi),\tag{E7}$$

where we have used (i) the invariance of the trace under transposition and (ii) the fact that all Paulis are symmetric, except Y, for which $Y^T=-Y$. Indeed, in the above formula, s_i^P and t_i^P are binary variables corresponding to the Gottesman encoding of Pauli operators and $\sum_{i=1}^n s_i^P t_i^P$ is counting the number of Ys in the expression of P. Therefore, we have

$$\mathrm{tr}(\psi\psi^*)=\sum_x(-1)^{\sum_i s_i^x t_i^x}p_\psi(x)=\langle(-1)^{\sum_i s_i^x t_i^x}\rangle_{x\sim p_\psi(x)}.\tag{E8}$$

In other words, we can sample from $p_\psi(x)$ and then sum up signs $(-1)^{\sum_i s_i^x t_i^x}$ depending on the Pauli P_x sampled. Let us say that we sample k times from the distribution p_ψ, obtaining the random variables $v_i\in\{\pm 1\}$ for each sample $i=1,\ldots,k$. An unbiased estimator $\tilde{I}$ for the quantity $I(\psi)$ is

$$\tilde{I}=\frac{1}{k}\sum_{i=1}^k v_i\tag{E9}$$

and, using Hoeffding's inequality and the fact that $|s_i|=1$ for all i, we can write

$$\mathbf{Pr}[|I-\tilde{I}|\geq\epsilon]=\mathbf{Pr}[|\mathbf{E}[\tilde{I}]-\tilde{I}|\geq\epsilon]\leq 2e^{-k\epsilon^2/2},\tag{E10}$$

which in turn says that $k=2\epsilon^{-2}\ln(2/\delta)$ are sufficient to estimate I within an error ϵ and failure probability δ.

However, we are not sampling from $p_\psi(x)$ directly but, rather, from a distribution $q(x)$ such that $TV(p_\psi,q)=\sum_x|p_\psi(x)-q(x)|\leq\epsilon'$. Therefore, a similar estimator $\hat{I}$ to that in Eq. (E9) constructed out of samples from $q(x)$ satisfies

$$\begin{aligned}&\left|\mathbf{E}[\tilde{I}]-\mathbf{E}[\hat{I}]\right|\\&=\left|\sum_x p_\psi(x)(-1)^{\sum_i s_i^x t_i^x}-\sum_x q(x)(-1)^{\sum_i s_i^x t_i^x}\right|\\&\leq\sum_x|p_\psi(x)-q(x)|\leq\Delta,\end{aligned}\tag{E11}$$

where the first inequality follows from the triangular inequality. Similarly to Eq. (E10), it also satisfies

$$\mathbf{Pr}[|\mathbf{E}[\hat{I}] - \hat{I}| \geq \epsilon - \Delta] \leq 2e^{-k(\epsilon-\Delta)^2/2}, \tag{E12}$$

which means that $k = 2(\epsilon - \Delta)^{-2} \ln(2/\delta)$ samples are sufficient to guarantee that $|\mathbf{E}[\hat{I}] - \hat{I}| \leq \epsilon - \Delta$ with failure probability δ and therefore

$$\begin{aligned} |I - \hat{I}| &= |I - \mathbf{E}[\hat{I}] + \mathbf{E}[\hat{I}] - \hat{I}| \\ &\leq |I - \mathbf{E}[\hat{I}]| + |\mathbf{E}[\hat{I}] - \hat{I}| \leq \epsilon, \end{aligned} \tag{E13}$$

by applying the triangular inequality and Eq. (E11). ■

APPENDIX F: BOUNDING MAGIC AND ENTANGLEMENT OF TILTED SUBSET PHASE STATES

In this appendix, we prove Lemma C2, restated below for convenience.

Lemma F1 (Lemma C2 restated). Let $|\in\rangle \mathcal{E}_{f,S}$ be a subset phase state that satisfies $S_\alpha(|\psi\rangle) = O(\log|S|)$ over all cuts $(A:B)$ and $M_\alpha(|\psi\rangle) = O(\log|S|)$, for $\alpha \geq 0$. Then, for $\epsilon \in (0,1), \theta \in [0, 2\pi]$, its corresponding tilted state $|\psi_{\epsilon,\theta}\rangle$ defined in Definition C6 also satisfies $S_\alpha(|\psi_{\epsilon,\theta}\rangle) = O(\log|S|)$ over all cuts $(A:B)$ and $M_\alpha(|\psi_{\epsilon,\theta}\rangle) = O(\log|S|)$, for $\alpha \geq 0$.

Proof. Given a state $|\phi\rangle \in \mathbb{C}^{\otimes 2^n}$, call $Q(|\phi\rangle) = \{P \in \mathcal{P}_n | \langle\phi| P |\phi\rangle \neq 0\}$. Then, $M_0(|\phi\rangle) = \log(|Q(|\phi\rangle)|/2^n)$.

In the case of the subset phase state $|\psi\rangle$ in the statement of the lemma, we have

$$M_0(|\psi\rangle) = \log\left(\frac{|Q(|\psi\rangle)|}{2^n}\right) \leq O(\log|S|). \tag{F1}$$

From this, we aim to bound $M_0(|\psi_{\epsilon,\theta}\rangle)$ for the tilted state

$$|\psi_{\epsilon,\theta}\rangle = \sqrt{1-\epsilon}e^{i\theta}|0\rangle|0^{\otimes n}\rangle + \sqrt{\epsilon}|1\rangle|\psi\rangle. \tag{F2}$$

We do so by upper bounding the size of $Q(|\psi_{\epsilon,\theta}\rangle)$. Consider $P \in \mathcal{P}_{n+1}$ and divide it into $P = P_1 \otimes P_n$. We have

$$\begin{aligned} \langle\psi_{\epsilon,\theta}| P_1 \otimes P_n |\psi_{\epsilon,\theta}\rangle &= \epsilon \langle 1| P_1 |1\rangle \langle\psi| P_n |\psi\rangle \\ &+ (1-\epsilon)\langle 0| P_1 |0\rangle \langle 0^{\otimes n}| P_n |0^{\otimes n}\rangle \qquad \text{(F3)} \\ &+ \sqrt{\epsilon(1-\epsilon)}[\langle 0| P_1 |1\rangle \langle 0^{\otimes n}| P_n |\psi\rangle + \text{c.c.}]. \qquad \text{(F4)} \end{aligned}$$

Consider first the case in which $P_1 \in \{I, Z\}$. Then, $\langle 0| P_1 |1\rangle = 0$, which means that the terms in the second line [Eq. (F4)] vanish. In the first line, the terms do not vanish if $P_n \in Q(|\psi\rangle)$ or $\langle 0^{\otimes n}| P_n |0^{\otimes n}\rangle \neq 0$ (i.e., $P_n \in \{I, Z\}^{\otimes n}$). This implies that there are at most $2(Q(|\psi\rangle) + 2^n)$ terms in $Q(|\psi_{\epsilon,\theta}\rangle)$ such that $P_1 \in \{I, Z\}$. As for the case in which $P_1 \in \{X, Y\}$, we have $\langle 0| P_1 |0\rangle = \langle 1| P_1 |1\rangle = 0$, which means that the terms in the first line [Eq. (F3)] vanish. Now note that $\langle 0^{\otimes n}| P_n |\psi\rangle$ is always of the form $\langle x|\psi\rangle$ (up to a potential phase i) for a certain $x \in \{0,1\}^n$ that depends on P_n and that each x has 2^n Paulis P_n that lead to it. Given that the subset phase state vector $|\psi\rangle$ is a superposition of $|S|$ computational basis states, then this means that there are at most $2|S|2^n$ terms in $Q(|\psi_{\epsilon,\theta}\rangle)$ such that $P_1 \in \{X, Y\}$. Overall, this leads to

$$|Q(|\psi_{\epsilon,\theta}\rangle)| \leq 2(Q(|\psi\rangle) + 2^n + |S|2^n), \tag{F5}$$

which, from Eq. (F1), results in $M_0(|\psi_{\epsilon,\theta}\rangle) \leq O(\log|S|)$. Since $M_\alpha(|\phi\rangle) \leq M_0(|\phi\rangle)$, for all $|\phi\rangle$ and $\alpha \geq 0$, then $M_\alpha(|\psi_{\epsilon,\theta}\rangle) \leq O(\log|S|)$.

As discussed around Eq. (A21), the Rényi-0 entanglement entropy over a cut $(A:B)$ is related to the Schmidt rank r of a quantum state over that cut as

$$S_0(|\psi\rangle) = \log(r). \tag{F6}$$

Looking at the Schmidt decomposition of the subset phase state $|\psi\rangle$ in the statement of the lemma,

$$|\psi\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}|i_A\rangle|i_B\rangle, \tag{F7}$$

we can derive the Schmidt decomposition of the tilted state vector,

$$|\psi_{\epsilon,\theta}\rangle = \sum_{i=1}^{r} \sqrt{\epsilon\lambda_i}|1\rangle|i_A\rangle|i_B\rangle + \sqrt{1-\epsilon}e^{i\theta}|0\rangle|0^{\otimes n}\rangle, \tag{F8}$$

irrespective of the cut $(A:B)$, since the tilted state is a superposition of two orthogonal subspaces. This means that the Schmidt rank of $|\psi_{\epsilon,\theta}\rangle$ is simply $r+1$ and therefore $S_0(|\psi_{\epsilon,\theta}\rangle) \leq O(\log|S|)$. Since $S_\alpha(|\phi\rangle) \leq S_0(|\phi\rangle)$, for all $|\phi\rangle$ and $\alpha \geq 0$, then $S_\alpha(|\psi_{\epsilon,\theta}\rangle) \leq O(\log|S|)$. ■

APPENDIX G: PAULI SAMPLING VIA ANCESTRAL SAMPLING

In this appendix, we present the proof of Theorem B3, which we restate here for convenience.

Theorem G1 (Theorem B3 restated). Let p be a distribution over $\{0,1\}^n$ and $\gamma > 0$. For all $\epsilon < \gamma/2$, given

black-box access to estimates of the marginals such that

$$|\pi(x_{1:k}) - p(x_{1:k})| \leq \frac{\epsilon}{2^k}, \tag{G1}$$

the adapted ancestral algorithm samples from a distribution q that satisfies

$$\|q - p\|_{\mathrm{TV}} \leq \mathfrak{f}(\gamma) + \left|\exp\left(\frac{4\epsilon n}{\gamma}\right) - 1\right|. \tag{G2}$$

Here, $\mathfrak{f}(\gamma)$ is defined in Definition B2.

To prove this theorem, we will need two additional lemmas. The first lemma gives a bound on the total variation distance between two distributions the individual probabilities of which are multiplicatively close on a large subset $S \subseteq \{0,1\}^n$ of the sample space.

Lemma G1 (Bound on TV distance via multiplicative error approximation on a heavy subset). Let $\tau \geq 0$ and let p and q be distributions over $\{0,1\}^n$. Assume that there is a subset $S \subseteq \{0,1\}^n$ such that

$$p(S) = \sum_{x \in S} p(x) \geq 1 - \tau, \tag{G3}$$

and that for all $x \in S$, we have $|p(x) - q(x)| \leq \Delta p(x)$. Then,

$$\|q - p\|_{\mathrm{TV}} \leq \tau + \Delta \tag{G4}$$

Proof. We start from

$$\|q - p\|_1 = \sum_{x \in S} |p(x) - q(x)| + \sum_{x \notin S} |p(x) - q(x)| \tag{G5}$$

$$\leq \Delta p(S) + \sum_{x \notin S} |p(x) - q(x)|. \tag{G6}$$

As for the second term, we have

$$\sum_{x \notin S} |p(x) - q(x)|$$

$$= \sum_{\substack{x \notin S \\ p(x) \geq q(x)}} p(x) - q(x) + \sum_{\substack{x \notin S \\ p(x) < q(x)}} q(x) - p(x) \tag{G7}$$

$$\leq \sum_{\substack{x \notin S \\ p(x) \geq q(x)}} p(x) + \sum_{\substack{x \notin S \\ p(x) < q(x)}} q(x) \tag{G8}$$

$$\leq 1 - p(S) + \sum_{x \notin S} q(x) \tag{G9}$$

and

$$\sum_{x \notin S} q(x) = 1 - \sum_{x \in S} q(x) \leq 1 - (1 - \Delta) \sum_{x \in S} p(x)$$

$$= 1 - (1 - \Delta) p(S), \tag{G10}$$

so that

$$\sum_{x \notin S} |p(x) - q(x)| \leq 2 - 2p(S) + \Delta p(S) \tag{G11}$$

and, overall,

$$\|q - p\|_1 \leq 2[1 - p(S) + \Delta p(S)]$$

$$\leq 2[\tau + \Delta(1 - \tau)] \leq 2[\tau + \Delta]. \tag{G12}$$

■

The second lemma shows that the adapted ancestral sampling algorithm can sample from a distribution q that approximates p multiplicatively on individual outcomes $x \in \{0,1\}^n$ such that $|p(x) - q(x)| < \Delta p(x)$. The condition for this is that the additive error on the estimates of the marginals, to which the algorithm is given access, is sufficiently small compared to the size of the marginals.

Lemma G2. Let $\gamma > 0$ and $\epsilon < \gamma/2$ and let $x \in \{0,1\}^n$ be an outcome such that for all corresponding marginals, it holds that

$$p(x_1, \ldots, x_k) \geq \frac{\gamma}{2^k}. \tag{G13}$$

Assume that for all $k \in [n]$, we are given access to additive error approximations $\pi(x_{1:k})$, i.e., access to all the corresponding marginals, such that

$$|\pi(x_{1:k}) - p(x_{1:k})| \leq \frac{\epsilon}{2^k}. \tag{G14}$$

Then, the adapted ancestral algorithm samples from a distribution q such that

$$|p(x) - q(x)| \leq \left|\exp\left(\frac{4\epsilon n}{\gamma}\right) - 1\right| p(x), \tag{G15}$$

in time $t = O(n)$.

Proof. Consider the kth step of the sampling procedure so that $x_1, \ldots, x_{k-1}$ are already determined. There are two possible scenarios: either both marginals $a = p(x_1, \ldots, x_k = 0)$ and $b = p(x_1, \ldots, x_k = 1)$ are positive or at most one of them is negative. We will treat each scenario independently.

Case 1: $a > 0, b > 0$, such that $q(x_k = 0|x_{1:k-1}) = a/(a+b) = q(x_{1:k-1}, x_k = 0)/q(x_{1:k-1})$. Using the closeness of the estimates $\pi(x_{1:k})$ from Eq. (G14), we have that

$$\frac{p(x_{1:k-1}, x_k = 0) - \frac{\epsilon}{2^k}}{p(x_{1:k-1}) + \frac{\epsilon}{2^{k-1}}} \leq \frac{q(x_{1:k-1}, x_k = 0)}{q(x_{1:k-1})} \leq \frac{p(x_{1:k-1}, x_k = 0) + \frac{\epsilon}{2^k}}{p(x_{1:k-1}) - \frac{\epsilon}{2^{k-1}}} \quad \text{(G16)}$$

and similarly for $q(x_{1:k-1}, x_k = 1)/q(x_{1:k-1})$. We note that due to the assumption on the size of the marginals from Eq. (G13), the additive errors in the numerator and denominator are of the same order of magnitude as the marginal. Thus, the additive errors can be turned into a multiplicative errors as follows:

$$\frac{\left(1 - \frac{\epsilon}{\gamma}\right)}{\left(1 + \frac{\epsilon}{\gamma}\right)} \frac{p(x_{1:k-1}, q_k = 0)}{p(x_{1:k-1})} \leq \frac{q(x_{1:k-1}, q_k = 0)}{q(x_{1:k-1})} \leq \frac{\left(1 + \frac{\epsilon}{\gamma}\right)}{\left(1 - \frac{\epsilon}{\gamma}\right)} \frac{p(x_{1:k-1}, q_k = 0)}{p(x_{1:k-1})}. \quad \text{(G17)}$$

Hence, using our assumption that $(\epsilon/\gamma) < 1/2$, we obtain

$$|q(x_k = 0|x_{1:k-1}) - p(x_k = 0|x_{1:k-1})| \leq 4\frac{\epsilon}{\gamma} p(x_k = 0|x_{1:k-1}) = \tilde{\epsilon} p(x_k = 0|x_{1:k-1})' \quad \text{(G18)}$$

where we set $\tilde{\epsilon} = 4\epsilon/\gamma$.

In the second scenario, one of the two marginals a and b is negative. By the properties of the ancestral sampling algorithm, it is clear that if the outcome of interest has $x_k = 0$, it has to be the case that $a > 0$ and, equivalently, if $x_k = 1$, it has to be the case that $b > 0$. Without loss of generality, we will only treat the first case here.

Case 2: $a > 0, b < 0$, such that $q(x_k = 0|x_{1:k-1}) = 1$ and $q(x_k = 0|x_{1:k-1}) = 0$. Note that

$$\begin{aligned} &|p(x_k = 0|x_{1:k-1}) - q(q_k = 0|x_{1:k-1})| \\ &= \left| p(x_k = 1|x_{1:k-1}) - \underbrace{q(x_k = 1|x_{1:k-1})}_{=0} \right| \\ &= p(x_k = 1|x_{1:k-1}) \end{aligned} \quad \text{(G19)}$$

and

$$p(x_k = 1|x_{1:k-1}) = \frac{p(x_{1:k-1}, x_k = 1)}{p(x_{1:k-1})} \leq \frac{\frac{\epsilon}{2^k}}{\frac{\gamma}{2^{k-1}}} = \frac{1}{2}\frac{\epsilon}{\gamma} \leq \frac{\epsilon}{\gamma}, \quad \text{(G20)}$$

where for the first inequality, we have again used the bound on the marginals from Eq. (G13) and the fact that $\pi(x_{1:k-1}, x_k) \leq 0$ implies that $p(x_{1:k-1}, x_k) \leq \epsilon/2^k$. Hence,

$$p(x_k = 0|x_{1:k-1}) \geq 1 - \frac{\epsilon}{\gamma} \quad \text{(G21)}$$

and again, since we have assumed $\epsilon/\gamma < 1/2$, we obtain

$$|p(x_k = 0|x_{1:k-1}) - q(x_k = 0|x_{1:k-1})| \leq \frac{\frac{\epsilon}{\gamma}}{1 - \frac{\epsilon}{\gamma}} p(x_k = 0|x_{1:k-1}) \leq \tilde{\epsilon} p(x_k = 0|x_{1:k-1}). \quad \text{(G22)}$$

We conclude that in both cases, we can show multiplicative closeness of the conditionals. Now, if we have multiplicative closeness for all $k \in [n]$ conditionals

$$(1 - \tilde{\epsilon}) p(x_k|x_1, \ldots, x_{k-1}) \leq q(x_k|x_1, \ldots, x_{k-1}) \leq (1 + \tilde{\epsilon}) p(x_k|x_1, \ldots, x_{k-1}), \quad \text{(G23)}$$

then it follows that the probabilities are also close:

$$\begin{aligned} |q(x) - p(x)| &= |q(x_n|x_1, \ldots, x_{n-1}) \cdots q(x_1) \\ &\quad - p(x_n|x_1, \ldots, x_{n-1}) \cdots p(x_1)| \\ &\leq |(1 + \tilde{\epsilon})^n - 1| p(x) \\ &\leq |\exp(\tilde{\epsilon} n) - 1| p(x) \\ &= |\exp(4\epsilon n/\gamma) - 1| p(x). \end{aligned} \quad \text{(G24)}$$

For the time-complexity bound, we note that the algorithm needs to flip n biased coins to output an n-bit string x. ■

We end this section by providing the proof of Theorem G1.

Proof of Theorem G1. The proof now follows immediately by combining Lemma G1 with Lemma G2. In particular, in Lemma G1, take the set S to be S_γ as defined in Definition B2. Then, by Lemma G2, for all $x \in S_\gamma$, the adapted ancestral-sampling algorithm distribution q approximates p via

$$|p(x) - q(x)| \leq \left| \exp\left(\frac{4\epsilon n}{\gamma}\right) - 1 \right| p(x). \quad \text{(G25)}$$

Therefore, the TV distance bounded due to Lemma G1 is

$$\begin{aligned} \|q - p\|_{\mathrm{TV}} &\leq (1 - p(S_\gamma)) + \left| \exp\left(\frac{4\epsilon n}{\gamma}\right) - 1 \right| \\ &= \mathrm{f}(\gamma) + \left| \exp\left(\frac{4\epsilon n}{\gamma}\right) - 1 \right|. \end{aligned} \quad \text{(G26)}$$

■

1. Robustness to Pauli noise

The following is a restatement of Theorem B5 from the main text.

Theorem G2. Let σ be a pure state and let Λ be a Pauli channel such that

$$\rho := \Lambda(\sigma) = (1-\xi)\sigma + \sum_{y\in\{0,1\}^{2n}\setminus\{0^{2n}\}} \xi_y P_y \sigma P_y, \quad \text{(G27)}$$

where for all y, we have that $\xi_y \geq 0$ and $\sum_y \xi_y = \xi$ and $0 \leq \xi \leq 1$. Then, there exists an algorithm for sampling, with probability at least $1-\delta$, from a distribution q_ρ such that

$$\|p_\rho - q_\rho\|_{TV} \leq \Delta, \quad \text{(G28)}$$

for any $\Delta > 4\xi$, using

$$N = O\left(\frac{n^3 \log(1/\delta) 2^{M_0(\sigma)} 2^{4E_0^\pi(\rho)}}{\Delta^4 (1-2\xi)^4}\right) \quad \text{(G29)}$$

Bell samples of the state $\rho^{\otimes 2}$.

Proof. By the action of the channel Λ, we have that

$$\mathrm{tr}(\rho P_x) = (1-\xi)\mathrm{tr}(\sigma P_x) + \sum_y \xi_y \mathrm{tr}(P_y \sigma P_y P_x), \quad \text{(G30)}$$

$$= \mathrm{tr}(\sigma P_x)\left[(1-\xi) + \sum_y (-1)^{[x,y]}\xi_y\right], \quad \text{(G31)}$$

where we have used that $\mathrm{tr}(P_y\sigma P_y P_x) = \pm\mathrm{tr}(\sigma P_x)$, depending on whether P_x and P_y commute or anticommute. Their commutation relation is denoted via $[x,y]=0$ if they commute and via $[x,y]=1$ if they anticommute. It follows that for any P_x,

$$|\mathrm{tr}(\rho P_x)| \geq (1-2\xi)|\mathrm{tr}(\sigma P_x)|. \quad \text{(G32)}$$

We can therefore obtain the following lower bound on the individual probabilities of the Pauli distribution p_ρ. For all $x \in \{0,1\}^{2n}$,

$$p_\rho(x) = \frac{\mathrm{tr}(\rho P_x)^2}{2^n \mathrm{tr}(\rho^2)} \geq (1-2\xi)^2 \frac{\mathrm{tr}(P_x\sigma)^2}{2^n} \geq (1-2\xi)^2 p_\sigma(x), \quad \text{(G33)}$$

where we have used that $\mathrm{tr}(\rho^2) \leq 1$. Further, we obtain for any marginal that

$$p_\rho(x_{1:k}) = \sum_{x_{k+1:n}} p_\rho(x) \geq (1-2\xi)^2 \sum_{x_{k+1:n}} p_\sigma(x)$$
$$= (1-2\xi)^2 p_\sigma(x_{1:k}). \quad \text{(G34)}$$

It thus follows that

$$\forall\, x \in S_\gamma(\sigma) \Rightarrow x \in S_{\gamma'}(\rho), \quad \text{(G35)}$$

with $\gamma' \leq (1-2\xi)^2\gamma$. Hence, from Theorem G1, we see that

$$\|q_\rho - p_\rho\|_{\mathrm{TV}} \leq \left[\exp\left(\frac{4\epsilon n}{\gamma'}\right) - 1\right] + \left[1 - \sum_{x\in S_{\gamma'}(\rho)} p_\rho(x)\right]$$
$$\leq \exp\left(\frac{4\epsilon n}{\gamma'}\right) - (1-2\xi)^2 \sum_{x\in S_\gamma(\sigma)} p_\sigma(x). \quad \text{(G36)}$$

Now, using $1 \geq \sum_{x\in S_\gamma(\sigma)} p_\sigma(x) \geq 1 - 2^{M_0(\sigma)}\gamma R$ and choosing

$$\epsilon \leq \frac{\Delta(1-2\xi)^2\gamma}{4\times 4n}, \quad \gamma = \frac{\Delta - 4\xi}{2\times 2^{M_0(\sigma)}R}, \quad \text{(G37)}$$

we obtain

$$\|q_\rho - p_\rho\|_{\mathrm{TV}} \leq \Delta. \quad \text{(G38)}$$

To obtain the sample complexity, note that $\mathrm{tr}(\rho^2) \geq (1-\xi)^2$ and then make use of Lemma D1. ■

APPENDIX H: ERROR ANALYSIS OF THE PROTOCOLS

Here, we show Theorems C5 and C6. To this end, it will be useful to consider the following two basic lemmas that we just recall without proof.

Lemma H1. Given $0 < r < 1$, consider the compact set $C_r \subseteq [-1,1]^2$ defined as

$$C_r = \left\{(x,y) \in [-1,1]^2 : x^2 + y^2 \geq r^2\right\}. \quad \text{(H1)}$$

The function $G : C_r \to \mathbb{R}$ defined as $G(x,y) = (1/2)(x+y)^2/(x^2+y^2)$ is Lipschitz continuous with Lipschitz constant $L(r)$ at most $1/r$.

Lemma H2. Given $0 < \lambda < 1$, consider the compact set $C_\lambda \subseteq [-1,1]^2$ defined as

$$C_\lambda = \left\{(x,y) \in [-1,1]^2 : x^2 \geq \lambda^2\right\}. \quad \text{(H2)}$$

The function $H : C_\lambda \to \mathbb{R}$ defined as $H(x,y) = y/x$ is Lipschitz continuous with Lipschitz constant $L(\lambda)$ at most $\sqrt{2}/\lambda$.

1. Proof of Theorem C5

Here, we prove the following (equivalent) version of Theorem C5.

Theorem H1 (Theorem C5 restated). Let $\epsilon_1 > 0$ and $\epsilon_2 > 0$. Let $\tilde{p}_{\text{mix}}$ be a distribution such that $||p_{\text{mix}} - \tilde{p}_{\text{mix}}||_{\text{TV}} < \Delta$. We write $f\,(N_1, N_2)$ for our estimate of $\text{tr}(\rho\sigma)$ with N_1 samples $\{x_1, \ldots, x_{N_1}\}$ from $\tilde{p}_{\text{mix}}$ and N_2 measurements per sample. One has

$$\begin{aligned}\mathbf{Pr}\left(|f\,(N_1,N_2) - \text{tr}(\rho\sigma)| > 2\epsilon_1 + 2\sqrt{\epsilon_2} + \frac{F_\rho(\epsilon_2)}{2} + \frac{F_\sigma(\epsilon_2)}{2} + 2\Delta\right)\\ < 4\exp\left(-2\epsilon_1^2 N_1\right) + 4N_1\exp\left(-\epsilon_2^2 N_2/2\right),\end{aligned} \tag{H3}$$

where F_ρ and F_σ denote the CDFs previously introduced in Definition A1.

Proof. The main tools needed are (i) the triangular inequality, (ii) Lipschitz continuity of the two-variable function $G(u, v)$, and (iii) Hoeffding's inequality together with the union bound. Therefore, first, we use the triangular inequality to expand the error $|f\,(N_1, N_2) - f\,|$ as follows. Recall that $f\, = (1 + \text{tr}(\rho\sigma))/2$ and $f\,(N_1, N_2)$ is defined in Eq. (C10):

$$\begin{aligned}|f\, - f\,(N_1,N_2)| \le &\left|f\, - \sum_x \tilde{p}_{\text{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x))\right|\\ &+ \left|\sum_x \tilde{p}_{\text{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) - \frac{1}{N_1}\sum_{i=1}^{N_1} G(\alpha_\rho(x_i), \alpha_\sigma(x_i))\right|\\ &+ \frac{1}{N_1}\sum_{i=1}^{N_1}\left|G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right|.\end{aligned}$$

Now, observe that the function $G(u, v)$ is Lipschitz continuous when restricted to a domain that does not include the origin. This motivates the definition of the following set:

$$I = \left\{i : \{\alpha_\rho(x_i)^2 > \epsilon_2\} \cup \{\alpha_\sigma(x_i)^2 > \epsilon_2\}\right\}. \tag{H4}$$

Now, we split the sum in the third term of the right-hand side above into two parts: the contribution from I and the contribution from its complement. One has

$$\begin{aligned}&\frac{1}{N_1}\sum_{i=1}^{N_1}\left|G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right|\\ &= \frac{1}{N_1}\sum_{i\in I}\left|G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right|\\ &\quad + \frac{1}{N_1}\sum_{i\notin I}\left|G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right|.\end{aligned}$$

When we consider the contribution to the sum of the terms $i \in I$, one can use the fact that $G(u, v)$ is Lipschitz continuous with Lipschitz constant $\sqrt{2/\epsilon_2}$ (see Lemma H1). When we consider the contribution coming from the terms $i \notin I$, we simply use the bound $|G(u, v) - G(u', v')| \le 1$, which is always true. So one finally writes

$$\begin{aligned}|f\, - f\,(N_1,N_2)| \le &\left|\text{tr}(\rho\sigma) - \sum_x \tilde{p}_{\text{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x))\right|\\ &+ \left|\sum_x \tilde{p}_{\text{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) - \frac{1}{N_1}\sum_{i=1}^{N_1} G(\alpha_\rho(x_i), \alpha_\sigma(x_i))\right|\\ &+ \frac{1}{N_1}\sum_{i\in I}\left|G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right| + \frac{1}{N_1}\sum_{i\notin I} 1.\end{aligned} \tag{H5}$$

Now, we analyze the four terms in the right-hand side of Eq. (H5) above separately in the following. To begin with, the first term is easily bounded, since $\|p - \tilde{p}_{\mathrm{mix}}\|_{\mathrm{TV}} \leq \Delta$ and $|G(u,v)| \leq 1$. One has

$$\left| f - \sum_x \tilde{p}_{\mathrm{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) \right| = \left| \sum_x p_{\mathrm{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) - \sum_x \tilde{p}_{\mathrm{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) \right| \leq \Delta. \tag{H6}$$

Now, one can use Hoeffding's inequality to bound the second and fourth terms in the right-hand side of Eq. (H5). Indeed,

$$\begin{aligned} \mathbf{Pr}\left(\left| \sum_x \tilde{p}_{\mathrm{mix}}(x) G(\alpha_\rho(x), \alpha_\sigma(x)) - \frac{1}{N_1} \sum_{i=1}^{N_1} G(\alpha_\rho(x_i), \alpha_\sigma(x_i)) \right| > \epsilon_1 \right) &< 2\exp\left(-2N_1\epsilon_1^2\right) \\ \mathbf{Pr}\left(\frac{1}{N_1} \sum_{i \notin I} 1 > \epsilon_1 + \Delta + \frac{F_\rho(\epsilon_2)}{2} + \frac{F_\sigma(\epsilon_2)}{2} \right) &< 2\exp\left(-2N_1\epsilon_1^2\right), \end{aligned} \tag{H7}$$

where, in the second line, we have used that $\mathbb{E}[N_1^{-1} \sum_{i \notin I} 1] \leq \Delta + F_\rho(\epsilon_2)/2 + F_\sigma(\epsilon_2)/2$. To see this, note that

$$\left| \mathbb{E}[N_1^{-1} \sum_{i \notin I} 1] - \sum_{x \notin S} p_{\mathrm{mix}}(x) \right| \leq \Delta, \tag{H8}$$

where $S = \{x : \{\alpha_\rho(x)^2 > \epsilon_2\} \cup \{\alpha_\sigma(x)^2 > \epsilon_2\}\}$ and

$$\sum_{x \notin S} p_{\mathrm{mix}}(x) = \frac{1}{2} \sum_{x \notin S} p_\rho(x) + \frac{1}{2} \sum_{x \notin S} p_\sigma(x) \leq \sum_{\{x : \alpha_\rho(x)^2 \leq \epsilon_2\}} \frac{p_\rho(x)}{2} + \sum_{\{x : \alpha_\sigma(x)^2 \leq \epsilon_2\}} \frac{p_\sigma(x)}{2} = \frac{1}{2} F_\rho(\epsilon_2) + \frac{1}{2} F_\sigma(\epsilon_2). \tag{H9}$$

(recall Definition A1). Finally, we consider the third term in the right-hand side of Eq. (H5). Note first that, using Hoeffding's inequality again,

$$\begin{aligned} \mathbf{Pr}\left(\left| \alpha_\rho(x_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} m_\rho(i,j) \right| > \epsilon_2 \right) &< 2\exp\left(-N_2\epsilon_2^2/2\right), \\ \mathbf{Pr}\left(\left| \alpha_\sigma(x_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} m_\sigma(i,j) \right| > \epsilon_2 \right) &< 2\exp\left(-N_2\epsilon_2^2/2\right), \end{aligned} \tag{H10}$$

for each index $i \in [N_1]$, so that, by the union bound,

$$\begin{aligned} &\mathbf{Pr}\left(\bigcup_{i=1}^{N_1} \left\{ \left| \alpha_\rho(x_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} m_\rho(i,j) \right| > \epsilon_2 \right\} \cup \left\{ \left| \alpha_\sigma(x_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} m_\sigma(i,j) \right| > \epsilon_2 \right\} \right) \\ &< 4N_1 \exp\left(-N_2\epsilon_2^2/2\right). \end{aligned} \tag{H11}$$

Now, note that for each $i \in I$, we can assume that $|\alpha_\rho(x_i)| > \sqrt{\epsilon_2}$ without loss of generality [see the definition of the set I in Eq. (H4)], so that, with probability at least $1 - 4N_1 \exp\left(-N_2\epsilon_2^2/2\right)$, we have $|N_2^{-1} \sum_{j=1}^{N_2} m_\rho(i,j)| > \sqrt{\epsilon_2}(1 - \sqrt{\epsilon_2})$. In

general, one has [55]

$$\max\left\{\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right\} > \sqrt{\frac{\epsilon_2}{2}} \tag{H12}$$

for all $i \in I$ with probability at least $1 - 4N_1 \exp\left(-N_2\epsilon_2^2/2\right)$. Thus,

$$\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j)\right)^2 + \left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)^2 > \frac{\epsilon_2}{2} \tag{H13}$$

for all $i \in I$ with probability at least $1 - 4N_1 \exp\left(-N_2\epsilon_2^2/2\right)$. Now, according to Lemma H1, the Lipschitz constant of $G(u,v)$ can be taken to be $\sqrt{2/\epsilon_2}$ when $u^2 + v^2 > \epsilon_2/2$, so that

$$\begin{aligned}&\left|G(\alpha_\rho(x_i),\alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right| \\ &< \sqrt{\frac{2}{\epsilon_2}}\sqrt{\left(\alpha_\rho(x_i) - \frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j)\right)^2 + \left(\alpha_\sigma(x_i) - \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)^2} < 2\sqrt{\epsilon_2}\end{aligned} \tag{H14}$$

for all $i \in I$ with probability at least $1 - 4N_1 \exp\left(-N_2\epsilon_2^2/2\right)$. Equivalently,

$$\begin{aligned}&\mathbf{Pr}\left(\frac{1}{N_1}\sum_{i\in I}\left|G(\alpha_\rho(x_i),\alpha_\sigma(x_i)) - G\left(\frac{1}{N_2}\sum_{j=1}^{N_2} m_\rho(i,j), \frac{1}{N_2}\sum_{j=1}^{N_2} m_\sigma(i,j)\right)\right| > 2\sqrt{\epsilon_2}\right) \\ &< 4N_1 \exp\left(-N_2\epsilon_2^2/2\right)\end{aligned} \tag{H15}$$

and the proof is complete. ■

2. Proof of Theorem C6

Note that Corollary H1 below corresponds to Theorem C6 from the main text. In this section, we will prove this theorem. It follows immediately from the following theorem.

Theorem H2. Let $p_\rho(x) = \alpha_\rho(x)^2/2^n$, let $\{x_1, \ldots, x_{N_1}\}$ be sampled from a distribution q such that $||p_\rho - q||_{\mathrm{TV}} \leq \Delta$, and let $\lambda > 0$. Consider the estimator

$$f(N_1, N_\sigma, N_\rho) = \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{\hat{\alpha}_\sigma(x_i, N_\sigma)}{s_\lambda(\hat{\alpha}_\rho(x_i, N_\rho))}, \tag{H16}$$

where the function $s_\lambda(\cdot)$ is defined as

$$s_\lambda(z) = \begin{cases} z, & |z| > \lambda, \\ \mathrm{sign}(z)\lambda, & |z| \leq \lambda, \end{cases} \tag{H17}$$

and $\hat{\alpha}_\sigma(x, N_\sigma)$ and $\hat{\alpha}_\rho(x, N_\rho)$ are estimates of $\alpha_\sigma(x) = \langle P_x\rangle_\sigma$ and $\alpha_\rho(x) = \langle P_x\rangle_\rho$, respectively, estimated using N_σ and N_ρ measurements each. Then, for arbitrary $\epsilon_\sigma > 0$, $\epsilon_\rho > 0$, and $\gamma > 0$, one has

$$\begin{aligned}&\mathbf{Pr}\left(\left|f(N_1, N_\sigma, N_\rho) - \mathrm{tr}(\rho\sigma)\right| > \frac{1}{\lambda}\left(\epsilon_\sigma + \frac{\epsilon_\rho}{\lambda}\right) + \gamma + \frac{\Delta}{\lambda} + \frac{F_\rho(\lambda^2)}{\lambda} + 2^{M_0(\rho)}\lambda\right) \\ &< 2N_1 \exp\left(-\epsilon_\sigma^2 N_\sigma/2\right) + 2N \exp\left(-\epsilon_\rho^2 N_\rho/2\right) + (1 + \Delta/\lambda^2)/(N_1\gamma^2).\end{aligned} \tag{H18}$$

Corollary H1 (Theorem C6 restated). With the same notation, one has

$$\begin{aligned}&\left|f\left(N_1,N_\sigma,N_\rho\right)-\mathrm{tr}(\rho\sigma)\right|\\&\quad\le\frac{1}{\lambda}\left(\epsilon_\sigma+\frac{\epsilon_\rho}{\lambda}\right)+\gamma+\frac{\Delta}{\lambda}+\frac{F_\rho(\lambda^2)}{\lambda}+2^{M_0(\rho)}\lambda\end{aligned}\tag{H19}$$

with probability at least $1-\delta$ provided that

$$N_1\ge\frac{3(1+\Delta/\lambda^2)}{\gamma^2\delta},\tag{H20}$$

$$N_\sigma\ge\frac{2}{\epsilon_\sigma^2}\ln\left(\frac{6N_1}{\delta}\right),\tag{H21}$$

$$N_\rho\ge\frac{2}{\epsilon_\rho^2}\ln\left(\frac{6N_1}{\delta}\right).\tag{H22}$$

Corollary H2 (Corollary C3 restated). With the same notation, one has

$$\left|f\left(N_1,N_\sigma,N_\rho\right)-\mathrm{tr}(\rho\sigma)\right|\le\epsilon\tag{H23}$$

with probability at least $1-\delta$ provided that

$$N_1\ge\frac{216}{\epsilon^2\delta},\tag{H24}$$

$$N_\sigma\ge\frac{2592\times16^{M_0(\rho)}}{\epsilon^4}\ln\left(\frac{6N_1}{\delta}\right),\tag{H25}$$

$$N_\rho\ge\frac{93\,312\times256^{M_0(\rho)}}{\epsilon^6}\ln\left(\frac{6N_1}{\delta}\right).\tag{H26}$$

Proof. Note that $F_\rho(\lambda^2)\le2^{M_0(\rho)}\lambda^2$ (Lemma A2) and take

$$\gamma=\frac{\epsilon}{6},\tag{H27}$$

$$\lambda=\frac{\epsilon}{6\times2^{M_0(\rho)}},\tag{H28}$$

$$\Delta=\lambda^2=\frac{\epsilon^2}{36\times4^{M_0(\rho)}},\tag{H29}$$

$$\epsilon_\sigma=\frac{\lambda\epsilon}{6}=\frac{\epsilon^2}{36\times2^{M_0(\rho)}},\tag{H30}$$

$$\epsilon_\rho=\frac{\lambda^2\epsilon}{6}=\frac{\epsilon^3}{216\times16^{M_0(\rho)}}.\tag{H31}$$

■

Proof of Theorem H2. We expand the error in four terms as

$$\begin{aligned}&\left|f\left(N_1,N_\sigma,N_\rho\right)-\mathrm{tr}(\rho\sigma)\right|\\&\quad\le\underbrace{\left|\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\hat\alpha_\sigma(x_i,N_\sigma)}{s_\lambda(\hat\alpha_\rho(x_i,N_\rho))}-\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\alpha_\sigma(x_i)}{s_\lambda(\alpha_\rho(x_i))}\right|}_{\text{(i)}}\\&\qquad+\underbrace{\left|\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\alpha_\sigma(x_i)}{s_\lambda(\alpha_\rho(x_i))}-\sum_x q(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|}_{\text{(ii)}}\\&\qquad+\underbrace{\left|\sum_x q(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}-\sum_x p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|}_{\text{(iii)}}\\&\qquad+\underbrace{\left|\sum_x p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}-\sum_x p(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}\right|}_{\text{(iv)}}\end{aligned}\tag{H32}$$

and we analyze the right-hand side of the previous inequality term by term.

We start with term (i) and first bound the probability that all the $2N$ estimates $\hat\alpha_\sigma(x_i,N_\sigma),\hat\alpha_\rho(x_i,N_\rho)$ are within additive error ϵ_σ and ϵ_ρ, respectively, as a function of N_σ and N_ρ. By Hoeffding's inequality, we have

$$\begin{aligned}\mathbf{Pr}\left(|\hat\alpha_\sigma(x_i,N_\sigma)-\alpha_\sigma(x_i)|>\epsilon_\sigma\right)&<2\exp\left(-N_\sigma\epsilon_\sigma^2/2\right)\\\mathbf{Pr}\left(|\hat\alpha_\rho(x_i,N_\rho)-\alpha_\rho(x_i)|>\epsilon_\rho\right)&<2\exp\left(-N_\rho\epsilon_\rho^2/2\right)\end{aligned}\tag{H33}$$

for each $i\in[N]$. Combined with the union bound, we obtain

$$\begin{aligned}&\mathbf{Pr}\left(\bigcup_{i=1}^{N_1}\left\{|\hat\alpha_\sigma(x_i,N_\sigma)-\alpha_\sigma(x_i)|>\epsilon_\sigma\right\}\cup\left\{|\hat\alpha_\rho(x_i,N_\rho)-\alpha_\rho(x_i)|>\epsilon_\rho\right\}\right)\\&\quad\le\sum_{i=1}^{N_1}\mathbf{Pr}\left(|\hat\alpha_\sigma(x_i,N_\sigma)-\alpha_\sigma(x_i)|>\epsilon_\sigma\right)+\mathbf{Pr}\left(|\hat\alpha_\rho(x_i,N_\rho)-\alpha_\rho(x_i)|>\epsilon_\rho\right)\\&\quad<2N_1\exp\left(-N_\sigma\epsilon_\sigma^2/2\right)+2N_1\exp\left(-N_\rho\epsilon_\rho^2/2\right).\end{aligned}\tag{H34}$$

Now, note that whenever $\hat{\alpha}_\sigma(x_i,N_\sigma),\hat{\alpha}_\rho(x_i,N_\rho)$ are within additive error ϵ_σ and ϵ_ρ, we have

$$\begin{aligned}\left|\frac{\hat{\alpha}_\sigma(x_i,N_\sigma)}{s_\lambda(\hat{\alpha}_\rho(x_i,N_\rho))}-\frac{\alpha_\sigma(x_i)}{s_\lambda(\alpha_\rho(x_i))}\right| &\le \left|\frac{\epsilon_\sigma}{s_\lambda(\hat{\alpha}_\rho(x_i,N_\rho))}\right| + |\alpha_\sigma(x_i)|\left|\frac{s_\lambda(\alpha_\rho(x_i))-s_\lambda(\hat{\alpha}_\rho(x_i,N_\rho))}{s_\lambda(\hat{\alpha}_\rho(x_i,N_\rho))s_\lambda(\alpha_\rho(x_i))}\right| \\ &\le \frac{\epsilon_\sigma}{\lambda}+\frac{\epsilon_\rho}{\lambda^2},\end{aligned} \tag{H35}$$

where we have used the fact that $s_\lambda(\cdot)\ge\lambda$ and $|\alpha_\sigma(\cdot)|\le 1$. Therefore, by combining Eqs. (H34) and (H35), we obtain

$$\begin{aligned}&\mathbf{Pr}\left(\left|\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\hat{\alpha}_\sigma(x_i,N_\sigma)}{s_\lambda(\hat{\alpha}_\rho(x_i,N_\rho))}-\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\alpha_\sigma(x_i)}{s_\lambda(\alpha_\rho(x_i))}\right|>\frac{1}{\lambda}\left(\epsilon_\sigma+\frac{\epsilon_\rho}{\lambda}\right)\right) \\ &\le \mathbf{Pr}\left(\bigcup_{i=1}^{N_1}\left\{|\hat{\alpha}_\sigma(x_i,N_\sigma)-\alpha_\sigma(x_i)|>\epsilon_\sigma\right\}\cup\left\{|\hat{\alpha}_\rho(x_i,N_\rho)-\alpha_\rho(x_i)|>\epsilon_\rho\right\}\right) \\ &\le 2N_1\exp\left(-N_\sigma\epsilon_\sigma^2/2\right)+2N_1\exp\left(-N_\rho\epsilon_\rho^2/2\right).\end{aligned} \tag{H36}$$

Now, we move to term (ii). We use Chebyshev's inequality on the random variable $X_\lambda(x)=\alpha_\sigma(x)/s_\lambda(\alpha_\rho(x))$, distributed according to q. To use this inequality, we bound the variance of the random variable:

$$\begin{aligned}\mathrm{Var}_q[X_\lambda(x)] &\le \sum_x q(x)\left(\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right)^2 \\ &\le \sum_x p(x)\left(\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right)^2+\sum_x|q(x)-p(x)|\left(\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right)^2 \\ &\le \mathbb{E}_p[X_\lambda(x)^2]+\frac{||p-q||_{\mathrm{TV}}}{\lambda^2} \\ &\le 1+\frac{\Delta}{\lambda^2},\end{aligned} \tag{H37}$$

where we have used the positivity of $X_\lambda(x)^2$ in the second line, $s_\lambda(\cdot)\ge\lambda$ in the third line, and $\mathbb{E}_p[X_\lambda(x)^2]\le\mathbb{E}_p[X(x)^2]=\sum_x\alpha_\sigma(x)^2/2^n=1$ in the last line, for $X(x)=\alpha_\sigma(x)/\alpha_\rho(x)$, by noting that $X_\lambda(x)^2\le X(x)^2,\forall x$. Now, for any γ, we have

$$\mathbf{Pr}\left(\left|\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{\alpha_\sigma(x_i)}{s_\lambda(\alpha_\rho(x_i))}-\sum_x q(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|>\gamma\right)<\frac{1+\Delta/\lambda^2}{N_1\gamma^2}. \tag{H38}$$

For term (iii), we have

$$\left|\sum_x q(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}-\sum_x p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|\le\sum_x|p(x)-q(x)|\left|\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|\le\frac{\Delta}{\lambda}, \tag{H39}$$

where we have again used that $s_\lambda(\cdot)\ge\lambda$ and $|\alpha_\sigma(\cdot)|\le 1$, and the assumption on the TV distance $||p-q||_{\mathrm{TV}}\le\Delta$.

Finally, term (iv) is

$$\begin{aligned}\left|\sum_x p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}-\sum_x p(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}\right| &= \left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}}p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}-\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}}p(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}\right| \\ &\le \left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}}p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right|+\left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}}p(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}\right|.\end{aligned} \tag{H40}$$

For clarity, let us deal with each term in the expression above independently. First,

$$\left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}} p(x)\frac{\alpha_\sigma(x)}{s_\lambda(\alpha_\rho(x))}\right| \le \frac{1}{\lambda}\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}} p(x) = \frac{1}{\lambda}\sum_{\{x:\alpha_\rho(x)^2\le\lambda^2\}} p(x) = \frac{F_\rho(\lambda^2)}{\lambda}, \tag{H41}$$

where we have again used that $s_\lambda(\cdot)\ge\lambda$ and $|\alpha_\sigma(\cdot)|\le 1$, and the definition of the CDF. Second,

$$\left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}} p(x)\frac{\alpha_\sigma(x)}{\alpha_\rho(x)}\right| = \left|\sum_{\{x:|\alpha_\rho(x)|\le\lambda\}} \frac{\alpha_\rho(x)\,\alpha_\sigma(x)}{2^n}\right| \le \frac{\lambda}{2^n}\sum_{\{x:\alpha_\rho(x)\neq 0\}} 1 = \frac{2^{M_0(\rho)+n}\lambda}{2^n} = 2^{M_0(\rho)}\lambda, \tag{H42}$$

from $|\alpha_\sigma(\cdot)|\le 1$ once more and the definition of $M_0(\rho)$ involving the support of the characteristic distribution, i.e., $M_0(\rho)=\log|\{x:\alpha_\rho(x)\neq 0\}|-n$. We can now put together the bounds on terms (i)–(iv) as

$$\begin{aligned}
&\mathbf{Pr}\left(\left|f\left(N_1,N_\sigma,N_\rho\right)-\mathrm{tr}(\rho\sigma)\right| > \frac{1}{\lambda}\left(\epsilon_\sigma+\frac{\epsilon_\rho}{\lambda}\right)+\gamma+\frac{\Delta}{\lambda}+\frac{F_\rho(\lambda^2)}{\lambda}+2^{M_0(\rho)}\lambda\right)\\
&< \mathbf{Pr}\left(\left\{(\mathrm{i}) > \frac{1}{\lambda}\left(\epsilon_\sigma+\frac{\epsilon_\rho}{\lambda}\right)\right\}\cup\{(\mathrm{ii})>\gamma\}\right)\\
&< \mathbf{Pr}\left((\mathrm{i}) > \frac{1}{\lambda}\left(\epsilon_\sigma+\frac{\epsilon_\rho}{\lambda}\right)\right)+\mathbf{Pr}\left((\mathrm{ii})>\gamma\right)\\
&< 2N_1\exp\left(-\epsilon_\sigma^2 N_\sigma/2\right)+2N_1\exp\left(-\epsilon_\rho^2 N_\rho/2\right)+(1+\Delta/\lambda^2)/(N_1\gamma^2).
\end{aligned} \tag{H43}$$

where the second line follows from the complements of these probabilities and the third line follows from the union bound. This completes the proof. ■

APPENDIX I: PAULI DISTRIBUTIONS WITH LARGE M_1 WHERE THE ALGORITHM IS EFFICIENT

For any $0<\tau<1$, consider the vector $|\phi_0\rangle = \sqrt{\tau}\,|0\cdots 0\rangle+\sqrt{1-\tau}\,|+\cdots+\rangle$ and the corresponding normalized pure state $|\phi\rangle = K^{-1/2}\,|\phi_0\rangle$, where the normalization constant $K=\langle\phi_0|\phi_0\rangle$ can be expressed as

$$K=\langle\phi_0|\phi_0\rangle = 1+2\sqrt{\frac{\tau(1-\tau)}{2^n}} = 1+2m, \tag{I1}$$

with $m=\sqrt{\tau(1-\tau)/2^n}$. We have the following result.

Lemma I1. Consider the CDF F_{ϕ_τ} (see Definition A1) of the n-qubit normalized pure state $|\phi_\tau\rangle = K^{-1/2}\,|\phi_0(\tau)\rangle$, where $|\phi_0(\tau)\rangle = \sqrt{\tau}\,|0\cdots 0\rangle+\sqrt{1-\tau}\,|+\cdots+\rangle$ and $K=1+2m$ with $m=\sqrt{\tau(1-\tau)/2^n}$. One has

$$F_\phi(\eta^2)\le 2\tau(1-\tau), \qquad \forall\eta\le\tau. \tag{I2}$$

Proof. There will be four important subsets, S_1, S_2, S_3, and S_4, that identify Pauli strings of the form $X_1^{a_1}Z_1^{b_1}\cdots X_n^{a_n}Z_n^{b_n}$, which we label, as usual, by the vector $(\vec{a},\vec{b})\in\{0,1\}^{2n}$. More precisely, we have

$$S_1=\{(\vec{a},\vec{b})=(\vec{0},\vec{0})\}, \tag{I3}$$

$$S_2=\{(\vec{a},\vec{0}):\vec{a}\neq\vec{0}\}, \tag{I4}$$

$$S_3=\{(\vec{0},\vec{b}):\vec{b}\neq\vec{0}\}, \tag{I5}$$

$$S_4=\{(\vec{a},\vec{b}):\vec{a}\neq\vec{0},\vec{b}\neq\vec{0},\langle\vec{a},\vec{b}\rangle\equiv 0 \mod 2\}. \tag{I6}$$

Note that $|S_1|=1$, as it only contains the identity. On the other hand, $|S_2|=|S_3|=2^n-1$, since S_2 and S_3 contain, respectively, nontrivial Pauli strings of the form $X_1^{a_1}\cdots X_n^{a_n}$ and $Z_1^{b_1}\cdots Z_n^{b_n}$. Regarding the cardinality of S_4, consider the set $R=\{(\vec{a},\vec{b}):\langle\vec{a},\vec{b}\rangle\equiv 0 \mod 2\}$. It is clear that $S_4\subseteq R$ and, indeed, one can write

$$\begin{aligned}
|S_4| &= |R|-|S_1|-|S_2|-|S_3|\\
&= |R|-2\times(2^n-1)-1 = |R|-2\times 2^n+1.
\end{aligned} \tag{I7}$$

Now, note that R contains Pauli strings of the form $P_1\otimes\cdots\otimes P_n$, with $P_i\in\{1,X,Y,Z\}$, where the number of Y factors is even. Therefore,

$$|R| = \sum_{k=0}^{n}\binom{n}{k}\frac{1+(-1)^k}{2}3^{n-k} = \frac{4^n+2^n}{2} \tag{I8}$$

and $|S_4| = |R| - 2 \times 2^n + 1 = 2^{2n-1} - 3 \times 2^{n-1} + 1$. One can explicitly check the following expressions from the definition of $|\phi_0\rangle$:

$$\alpha_0(\vec{a},\vec{b}) := \langle\phi_0| X_1^{a_1} Z_1^{b_1} \cdots X_n^{a_n} Z_n^{b_n} |\phi_0\rangle = \begin{cases} 1+2m, & (\vec{a},\vec{b}) \in S_1, \\ 1-\tau+2m, & (\vec{a},\vec{b}) \in S_2, \\ \tau+2m, & (\vec{a},\vec{b}) \in S_3, \\ 2m, & (\vec{a},\vec{b}) \in S_4, \end{cases} \tag{I9}$$

from which we can compute the quantity in which we are interested, namely,

$$\alpha(\vec{a},\vec{b}) := \langle\phi| X_1^{a_1} Z_1^{b_1} \cdots X_n^{a_n} Z_n^{b_n} |\phi\rangle = \frac{\alpha_0(\vec{a},\vec{b})}{1+2m}. \tag{I10}$$

Recalling that $m = \sqrt{\tau(1-\tau)/2^n}$, we now assume that $\tau \leq 1/2$ and $2m < \tau$. Note that for the important case of $\tau = 1/\sqrt{n}$, one indeed has $2m < \tau$ for all n. Now, noting the inequalities

$$\frac{1-\tau+2m}{1+2m} \geq \frac{\tau+2m}{1+2m} > \tau > 2m > \frac{2m}{1+2m}, \tag{I11}$$

it follows that the set S_4 can be characterized as

$$\begin{aligned} S_4 &= \left\{(\vec{a},\vec{b}) : 0 < |\alpha(\vec{a},\vec{b})| \leq \frac{2m}{1+2m} \leq 2m\right\} \\ &= \left\{(\vec{a},\vec{b}) : 0 < |\alpha(\vec{a},\vec{b})| \leq \tau\right\}. \end{aligned} \tag{I12}$$

In other words, $0 < |\alpha(\vec{a},\vec{b})| \leq \tau$ if and only if $(\vec{a},\vec{b}) \in S_4$ if and only if $0 < |\alpha(\vec{a},\vec{b})| \leq 2m$. We have

$$\begin{aligned} F_{\phi_\tau}(\tau^2) &= \sum_{\{x:\alpha(x)^2 \leq \tau^2\}} p_\phi(x) = \sum_{(\vec{a},\vec{b}) \in S_4} \frac{\alpha(\vec{a},\vec{b})^2}{2^n} \\ &\leq \frac{4m^2}{2^n} \sum_{(\vec{a},\vec{b}) \in S_4} 1 = \frac{4m^2}{2^n} |S_4| \\ &= \frac{4\tau(1-\tau)}{4^n} |S_4| < 2\tau(1-\tau), \end{aligned} \tag{I13}$$

where we have used the fact that $|S_4| < 4^n/2$. Since the CDF F_{ϕ_τ} satisfies $F_{\phi_\tau}(x) \leq F_{\phi_\tau}(y)$ whenever $x \leq y$, this concludes the proof. ■

The previous result shows that there are states ϕ with $F_\phi(\tau^2) \leq 2\tau(1-\tau) \leq 2\tau$. Consider Theorem C5 with $\epsilon_1 = \tau$ and $\epsilon_2 = \tau^2$, assuming that $F_\rho(\tau^2) \leq 2\tau$ and $F_\sigma(\tau^2) \leq 2\tau$. One has

$$\left| f(N_1, N_2) - \mathrm{tr}(\rho\sigma) \right| \leq 6\tau + 2\Delta \tag{I14}$$

with probability at least $1-\delta$ provided that $N_1 \geq (2\tau^2)^{-1} \log(8/\delta)$ and $N_2 \geq (2/\tau^4) \log(8N_1/\delta)$.

1. Magic and stabilizer entropies

Clearly, we have $M_0(\phi) = \log|R| - n$, where $R = \{(\vec{a},\vec{b}) : \langle\vec{a},\vec{b}\rangle \equiv 0 \mod 2\}$ and we have shown that its cardinality is $|R| = 2^{n-1}(1+2^n)$. It follows that $M_0(\phi)$, which is between 0 and n, is very close to its maximum possible value for all values of the parameter τ. Indeed,

$$\begin{aligned} M_0(\phi) &= \log\left(2^{n-1}(1+2^n)\right) - n \\ &= \log\left(1+2^n\right) - 1 > \log\left(2^n\right) - 1 = n-1. \end{aligned} \tag{I15}$$

We now fix $\tau = 1/\sqrt{n}$ (since we are assuming $\tau \leq 1/2$, let us assume $n \geq 4$) and consider

$$p_\phi(\vec{a},\vec{b}) = \frac{|\alpha(\vec{a},\vec{b})|^2}{2^n} = \begin{cases} \frac{1}{2^n}, & (\vec{a},\vec{b}) \in S_1, \\ \frac{1}{2^n}\left(\frac{1-\tau+2m}{1+2m}\right)^2, & (\vec{a},\vec{b}) \in S_2, \\ \frac{1}{2^n}\left(\frac{\tau+2m}{1+2m}\right)^2, & (\vec{a},\vec{b}) \in S_3, \\ \frac{1}{2^n}\left(\frac{2m}{1+2m}\right)^2, & (\vec{a},\vec{b}) \in S_4. \end{cases} \tag{I16}$$

Now recall that $|S_1| = 1$, $|S_2| = |S_3| = 2^n - 1$, and $|S_4| = 4^n/2 - 3 \times 2^n/2 + 1$. Since the parameter m is just $\sqrt{\tau(1-\tau)/2^n}$, the previous expressions allows one to straightforwardly evaluate $M_1(\phi)$ for any value of τ and n. In particular, it is not hard to see that $M_1(\phi) > \sqrt{n}$ when $\tau = 1/\sqrt{n}$.

APPENDIX J: HIGHER-DIMENSIONAL SYSTEMS

In this appendix, we finally note that several of the above statements carry over to systems beyond quantum systems consisting of qubits but of higher-dimensional quantum systems, to exemplify the generality of the approach pursued. Specifically, for any local dimension d, one can pick a basis of unitaries $\{U_j | j = 1, \ldots, d\}$ with $U_j \in U(d)$ that are orthonormal with respect to the Hilbert-Schmidt scalar product as

$$\mathrm{tr}(U_j^\dagger U_k) = d\delta_{j,k} \tag{J1}$$

for $j, k = 1, \ldots, d$. Such a basis actually exists for any integer dimension d and generalizes Pauli operators to arbitrary dimensions [56], albeit not necessarily as Hermitian operators. Again denoting with

$$|\Omega\rangle := \frac{1}{\sqrt{d}} \sum_{v=0}^{d-1} |v, v\rangle \tag{J2}$$

a fiducial maximally entangled state vector in the $d \times d$-dimensional bipartite system, one finds a basis of maximally entangled state vectors as

$$\{|\psi_j\rangle := \{(U_j \otimes I)\,|\Omega\rangle\,|j = 0, \ldots, d-1\} \tag{J3}$$

as a basis in $\mathbb{C}^d \otimes \mathbb{C}^d$. It is easy to see that the orthonormality of the unitary operator basis is inherited by the orthonormality of this basis of state vectors.

In order to estimate the purity via maximally entangled state sampling for pure states $\rho = |\psi\rangle\langle\psi|$ with real state vectors $|\psi\rangle$, prepare $\rho \otimes \rho$. Then, measure in the maximally entangled basis $\{|\psi_j\rangle\}$, giving rise to outcomes $x \in \{0, \ldots, d-1\}$. It takes a moment of thought to see that the statistics of the maximally entangled measurements are given by

$$p_\rho(x) = \frac{1}{d}|\langle\psi|\,U_x\,|\psi^*\rangle|^2 \tag{J4}$$

in terms of a complex conjugation arising from transposition, which for real state vectors is

$$p_\rho(x) = \frac{1}{d}|\mathrm{tr}(\rho U_x)|^2. \tag{J5}$$

That is to say, by preparing two copies and performing suitable measurements in a maximally entangled basis, for any local dimension d, one can sample from the distribution p_ρ in this case. This means that many of the above statements carry over to this situation. If a multiqubit system is seen as a single higher-dimensional one, the measurements involve entangled measurements over several qubits.

It is also worth noting that all eigenvectors of $U_j \otimes U_j$ for all $j = 1, \ldots, d$ with eigenvalue $+1$ are all maximally entangled. Since, for all j,

$$(U_j \otimes U_j)|\psi\rangle\langle\psi|(U_j \otimes U_j)^\dagger = |\psi\rangle\langle\psi|, \tag{J6}$$

we have that

$$U_j\,\mathrm{tr}_2(|\psi\rangle\langle\psi|)U_j^\dagger = \mathrm{tr}_2(|\psi\rangle\langle\psi|), \tag{J7}$$

by partial trace. Invoking the unitary design property of the collection of unitaries [57], we find

$$\frac{1}{d^2}\sum_j U_j\,\mathrm{tr}_2(|\psi\rangle\langle\psi|)U_j^\dagger = \frac{I}{d} = \mathrm{tr}_2(|\psi\rangle\langle\psi|), \tag{J8}$$

which means that $|\psi\rangle$ is maximally entangled.

[1] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).

[2] R. O'Donnell and J. Wright, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16 (Association for Computing Machinery, New York, 2016), p. 899.

[3] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[4] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[5] A. Elben, B. Vermersch, R. van Bijnen, C. Kokail, T. Brydges, C. Maier, M. Joshi, R. Blatt, C. F. Roos, and P. Zoller, Cross-platform verification of intermediate scale quantum devices, Phys. Rev. Lett. **124**, 010504 (2020).

[6] A. Anshu, Z. Landau, and Y. Liu, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2022), p. 44.

[7] J. Carrasco, A. Elben, C. Kokail, B. Kraus, and P. Zoller, Theoretical and experimental perspectives of quantum verification, PRX Quantum **2**, 010102 (2021).

[8] Y. Qian, Y. Du, Z. He, M.-h. Hsieh, and D. Tao, Multimodal deep representation learning for quantum cross-platform verification, Phys. Rev. Lett. **133**, 130601 (2024).

[9] C. Zheng, X. Yu, and K. Wang, Cross-platform comparison of arbitrary quantum processes, npj Quantum Inf. **10**, 4 (2024).

[10] D. Zhu, Z.-P. Cian, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. H. Alderete, N. H. Nguyen, Q. Wang, A. Maksymov, Y. Nam, M. Cetina, N. M. Linke, M. Hafezi, and C. Monroe, Cross-platform comparison of arbitrary quantum computations, Nat. Commun. **13**, 6620 (2022).

[11] A. Montanaro, Learning stabilizer states by Bell sampling, ArXiv:1707.04012.

[12] D. Gross, S. Nezami, and M. Walter, Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations, Commun. Math. Phys. **385**, 1325 (2021).

[13] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, Science **376**, 1182 (2022).

[14] D. Hangleiter and M. J. Gullans, Bell sampling from quantum circuits, Phys. Rev. Lett. **133**, 020601 (2024).

[15] S. Grewal, V. Iyer, W. Kretschmer, and D. Liang, Efficient learning of quantum states prepared with few non-Clifford gates, ArXiv:2305.13409.

[16] S. Grewal, V. Iyer, W. Kretschmer, and D. Liang, in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024 (Association for Computing Machinery, New York, 2024), p. 1352.

[17] F. E. Gutiérrez, Simple algorithms to test and learn local Hamiltonians, ArXiv:2404.06282.

[18] S. Bravyi, D. Gosset, and Y. Liu, How to simulate quantum measurement without computing marginals, Phys. Rev. Lett. **128**, 220503 (2022).

[19] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Achieving quantum supremacy with sparse and noisy commuting quantum computations, Quantum **1**, 8 (2017).

[20] P. Clifford and R. Clifford, in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '18 (Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2018), p. 146.

[21] L. Leone, S. F. E. Oliviero, and A. Hamma, Learning t-doped stabilizer states, Quantum **8**, 1361 (2024).

[22] N.-H. Chia, C.-Y. Lai, and H.-H. Lin, Efficient learning of t-doped stabilizer states with single-copy measurements, Quantum **8**, 1250 (2024).

[23] L. Leone, S. F. E. Oliviero, and A. Hamma, Stabilizer Rényi entropy, Phys. Rev. Lett. **128**, 050402 (2022).

[24] L. Leone, S. F. E. Oliviero, and A. Hamma, Nonstabilizerness determining the hardness of direct fidelity estimation, Phys. Rev. A **107**, 022429 (2023).

[25] G. Lami and M. Collura, Nonstabilizerness via perfect Pauli sampling of matrix product states, Phys. Rev. Lett. **131**, 180401 (2023).

[26] T. Haug and L. Piroli, Stabilizer entropies and nonstabilizerness monotones, Quantum **7**, 1092 (2023).

[27] P. S. Tarabunga, E. Tirrito, T. Chanda, and M. Dalmonte, Many-body magic via Pauli-Markov chains—from criticality to gauge theories, PRX Quantum **4**, 040317 (2023).

[28] Z. Ji, Y.-K. Liu, and F. Song, in *Advances in Cryptology—CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, California, USA, August 19–23, 2018, Proceedings, Part III* (Springer-Verlag, Berlin, Heidelberg, 2018), p. 126.

[29] S. Aaronson, A. Bouland, B. Fefferman, S. Ghosh, U. Vazirani, C. Zhang, and Z. Zhou, in *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)* (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2024).

[30] A. Gu, L. Leone, S. Ghosh, J. Eisert, S. F. Yelin, and Y. Quek, Pseudomagic quantum states, Phys. Rev. Lett. **132**, 210602 (2024).

[31] T. Haug, K. Bharti, and D. E. Koh, Pseudorandom unitaries are neither real nor sparse nor noise-robust, Quantum **9**, 1759 (2025).

[32] M. B. Plenio and S. Virmani, An introduction to entanglement measures, Quantum Inf. Comput. **7**, 1 (2007).

[33] Z.-W. Liu and A. Winter, Many-body quantum magic, PRX Quantum **3**, 020333 (2022).

[34] L. Leone and L. Bittel, Stabilizer entropies are monotones for magic-state resource theory, Phys. Rev. A **110**, L040403 (2024).

[35] A. Gu, S. F. E. Oliviero, and L. Leone, Doped stabilizer states in many-body physics and where to find them, Phys. Rev. A **110**, 062427 (2024).

[36] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 2011), 2nd ed.

[37] S. Bravyi, D. P. Divincenzo, R. Oliveira, and B. M. Terhal, The complexity of stoquastic local Hamiltonian problems, Quantum Inf. Comput. **8**, 361 (2008).

[38] Y. Teng, R. Samajdar, K. Van Kirk, F. Wilde, S. Sachdev, J. Eisert, R. Sweke, and K. Najafi, Learning topological states from randomized measurements using variational tensor network tomography, ArXiv:2406.00193.

[39] M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).

[40] J. Denzler, S. Varona, T. Guaita, and J. Carrasco, Highly-entangled, highly-doped states that are efficiently cross-device verifiable, Phys. Rev. Res. **7**, 023320 (2025).

[41] T. Haug, S. Lee, and M. S. Kim, Efficient quantum algorithms for stabilizer entropies, Phys. Rev. Lett. **132**, 240602 (2024).

[42] W. Gong, J. Haferkamp, Q. Ye, and Z. Zhang, On the sample complexity of purity and inner product estimation, ArXiv:2410.12712.

[43] S. Arunachalam and L. Schatzki, in *42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025)*, Leibniz International Proceedings in Informatics (LIPIcs) (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2025), Vol. 327, p. 11:1.

[44] T. Haug, L. Aolita, and M. S. Kim, Probing quantum complexity via universal saturation of stabilizer entropies, Quantum **9**, 1801 (2025).

[45] G. Lami and M. Collura, Unveiling the stabilizer group of a matrix product state, Phys. Rev. Lett. **133**, 010602 (2024).

[46] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, United Kingdom, 2018).

[47] The bound $\mathcal{S}_\alpha(\psi_{f,S}) = \omega(\log n)$ applies only to cuts where the subsystem size is also at least $\omega(\log n)$.

[48] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society, Los Alamitos, CA, USA, 2022), p. 574.

[49] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, Phys. Rev. Lett. **126**, 190505 (2021).

[50] H. J. Briegel and R. Raussendorf, Persistent entanglement in arrays of interacting particles, Phys. Rev. Lett. **86**, 910 (2001).

[51] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[52] B. Wu, J. Sun, Q. Huang, and X. Yuan, Overlapped grouping measurement: A unified framework for measuring quantum states, Quantum **7**, 896 (2023).

[53] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).

[54] A median-of-means estimator could be used to reduce the complexity of N_1 to $2(1+\Delta/\lambda^2)/\gamma^2 \ln(3/\delta)$ (or $O\left(\log\left(\delta^{-1}\right)/\epsilon^2\right)$), but for ease of exposition, we opt for a simpler empirical mean estimator.

[55] We now assume that $\epsilon_2 < 1/16$, so that $\sqrt{\epsilon_2} < 1/4$ and $1 - \sqrt{\epsilon_2} > 3/4 > 1/\sqrt{2}$.

[56] R. F. Werner, All teleportation and dense coding schemes, J. Phys. A: Math. Gen. **34**, 7081 (2001).

[57] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, J. Math. Phys. **48**, 052104 (2007).