Practical characterization of quantum devices without tomography

Marcus P. da Silva
Raytheon BBN Technologies, Cambridge, Massachusetts, USA and
Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, Canada

Olivier Landon-Cardinal and David Poulin
Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, Canada
(Dated: April 19, 2011)

The complexity of quantum tomography experiments presents a major obstacle for the characterisation of even moderately large quantum information devices. Not only does the number of experimental settings grows exponentially with the number of qubits, but so does the running time for processing the data generated by these experiments. Part of the problem is that tomography generates much more information than is usually seeked. Taking a more targeted approach, we demonstrate that it is posible to independently *(i)* verify the fidelity of an experiment to a theoretically ideal state or process, *(ii)* learn which quantum state or process within a reduced subset best matches the experimental data. Both these approaches yield a significant reduction in resources for both state and process tomography. In general this speed-up is quadratic, but for a number of special cases leading to universal quantum computation, we obtain exponential reduction in resources, as well as fidelity estimation experiments with resources which are independent of the system size.

I. INTRODUCTION

The building blocks for quantum computers have been demonstrated in a number of different physical systems [1–6]. In order to quantify how closely these demonstrations come to the ideal operations, the experiments are fully characterized via either *quantum state tomography* [7] or *quantum process tomography* [8], and then some distance measure between the estimated description of the experiment and the ideal target is computed. The main drawbacks of these approaches are that tomography fundamentally requires both experimental and data post-processing resources that increase exponentially with the number of particles n [9]. Here we demonstrate that by using targeted approaches, it is possible to efficiently characterize a large class of states and operations—including some that are universal resources for quantum computation—completely circumventing these exponential costs while still using only local measurements. More generally, we show that it is possible to directly estimate the distance between the experimental implementation and the ideal description with significantly less resources than needed by tomography for any state or operation.

It is important to realize that the exponential cost of tomography is not a problem restricted to a large number of qubits. For example, recent ion trap experiments characterizing an 8 qubit state required 10 hours of measurements, despite collecting only 100 samples per observable [3]. Surprisingly, the post-processing of the data obtained from these experiments took approximatelly a week [10]. Under similar time scales, the characterization of a 16 qubit state would take years of measurements, and over a century of data post-processing. This is clearly a major obstacle in the demonstration of working quantum computers, even at sizes moderately larger than what has been demonstrated to date.

Moreover, one of the key assumptions for the fault-tolerance theorems of quantum computation is that the noise on elementary components does not scale with the system size [11]. Therefore, despite the fact that universal quantum computation can be realized with one- and two-qubit elementary operations, it is not sufficient to characterize small gates—larger systems may have significant noise contributions from correlated sources as seen in recent experiments [6]. The characterization of multi-qubit states and operations acting on the entire quantum computer provides crucial information for the verification of these assumptions, and therefore the development of large quantum information processors.

Part of the problem with the usual approach is that tomography often provides more information than what is truly seeked. Given an experiment that prepares a quantum state represented by a density operator $\hat{\sigma}$, one usually extracts a complete description for $\hat{\sigma}$ via quantum tomography, and then compares this description to a theoretical state $\hat{\rho}$ by computing the fidelity $F(\hat{\rho}, \hat{\sigma})$—a single number. As this example illustrates, we often have an idea of what has been realized in the laboratory, so we are interested in asking for much less information—*e.g.*, we only want to know the distance to some particular theoretical target, or to measure some entanglement witness, or to learn the identity of the state or operation within a restricted set of possibilities.

Our main results, summarized in Table I, show that some of these data can be extracted efficiently for a wide class of states and operations—including some that are universal resources for quantum computation—without resorting to tomography. Moreover, these schemes require only elementary experimental procedures, such as single qubit measurements and the preparation of product states. In particular, it shows that estimating the fidelity to some theoretical target always requires drastically less resources than full tomography—in some important cases, it is an exponential reduction in resources. Even in the worst case, our scheme offers three significant advantages for the characterization of quantum states: *(1)* Its computational cost is bounded by $n4^n$, compared to 4^{3n} required for the simplest tomography procedure based on pseudo-inverses. *(2)* The number of distinct experimental settings it requires is constant—independent of the sys-

		Fidelity		**Learning**
		Sampling (**C1**)	Fluctuations (**C2**)	
States	Stabilizer	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$
	W	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$
	$\vert t_n\rangle$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$
	General MPS	$\mathcal{O}(n)$	?	$\mathcal{O}(n)$ [12]
	General pure state	$\mathcal{O}(n^2 2^{2n})$	$\mathcal{O}(2^n)$	$\mathcal{O}(2^{6n})$
Processes	Clifford	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(n^2)$
	MPS Choi matrix	$\mathcal{O}(n)$	?	$\mathcal{O}(n)$
	General unitary	$\mathcal{O}(n^2 2^{4n})$	$\mathcal{O}(2^{2n})$	$\mathcal{O}(2^{12n})$
Evolution	Local Hamiltonian	—	—	$\mathcal{O}(n)$
	Local Linbladian	—	—	$\mathcal{O}(n)$

TABLE I: Complexity of the characterization of various states and processes. Entries in red are efficient, *i.e.* require resources that grow at most polynomially with the number of qubits n. **Fidelity** refers to the task of estimating the fidelity to a theoretical target and **Learning** consists of identifying the state or operation within a restricted set of possibilities. When both fidelity estimate and learning are possible, the assumption that the state or operation belongs to a restricted set can be lifted since fidelity can testify of that assumption. The **Sampling** column gives the complexity of the classical processing required to sample from the relevance distribution, *c.f.* **C1**. The **Fluctuations** column gives the factor by which the measurement fluctuations are increased when evaluating the fidelity, *c.f.* **C2**. The **Learning** column gives the total number of measurements (including repetitions of the same measurement setting) required to learn the state; the classical processing is always a polynomial of that number. Stabilizer states, Clifford gates, local Hamiltonians and local Lindbladians are discussed in the main text. The W state has often been used as an experimental benchmark, *e.g.* [3]. The $|t_n\rangle$ state plays a key role in linear optics quantum computation [13]. Matrix product states (MPS) and operations describe 1D quantum systems [14]. Question marks indicate open problems, but they can be no worst than the general states and operations.

tem size and depending only on the desired accuracy of the estimate—compared the 4^n distinct measurements needed by tomography. *(3)* The total number of measurements (counting repeated measurements used to statistically estimate expectation values) of our scheme is bounded by 2^n, which is at least a quadratic improvement over what is required by tomography (equivalent statements hold for quantum operations).

To estimate the fidelity to some theoretical pure state $\hat\rho$ (quantum processes will be discussed similarly below), we use

$$F(\hat\rho,\hat\sigma) = \mathrm{tr}\,\hat\rho\hat\sigma = \sum_i \frac{\rho_i\sigma_i}{d}. \tag{1}$$

where $\rho_i = \mathrm{tr}\,\hat\rho\hat P_i$, $\sigma_i = \mathrm{tr}\,\hat\sigma\hat P_i$, d is the dimension of the Hilbert space, and $\hat P_i$ is some orthonormal Hermitian operator basis satisfying $\mathrm{tr}\,\hat P_i\hat P_j = d\delta_{ij}$. For a system composed of n qubits, the $\hat P_i$ could be the 4^n Pauli operators obtained by taking tensor products of the Pauli matrices and the identity. Defining the *relevance distribution* $\Pr(i) = \frac{\rho_i^2}{d}$, we can rewrite the fidelity as

$$F(\hat\sigma,\hat\rho) = \sum_i \Pr(i)\frac{\sigma_i}{\rho_i}, \tag{2}$$

where the sum is taken over only the i with $\rho_i \neq 0$. This expression leads to an experimental procedure to estimate the fidelity as follows: one generates N random indices $i_1, i_2, \ldots, i_N$ following the relevance distribution $\Pr(i)$ and estimates $\sigma_{i_k} = \langle\hat P_{i_k}\rangle_{\hat\sigma}$, the experimental expectation value of the observable $\hat P_{i_k}$. With high probability, the fidelity is close to $\frac{1}{N}\sum_{k=1}^N \frac{\sigma_{i_k}}{\rho_{i_k}}$ with an uncertainty that decreases as $\frac{1}{\sqrt{N}}$. The total number of distinct experimental settings is at most N, independent of the system size.

There are two important caveats to this technique:

C1 Generating an index i according to the relevance distribution $\Pr(i)$ can in general require exponential amount of computational resources.

C2 Each σ_{i_k} is estimated within some finite accuracy. To estimate the fidelity with accuracy ϵ therefore requires repeating the measurement of P_{i_k} roughly $(\epsilon\rho_{i_k})^{-2}$ times, which in the worst case grows exponentially with the number of qubits.

These are important limitations, and as a consequence our method will not scale polynomially for all quantum states and operations, but nevertheless always does significantly better than tomography. In addition, there are important classes of states and operations which avoid these two problems (see Table I). The remainder of this article is organized as follows: Sec. II-V focus of the Monte Carlo approach described here, from a detailed proof of the claims, to a dicussion of the algorithm performance with respect to the task of sampling random experiments, and to the task of measuring relevant observables with sufficient accuracy. Sec. VI details how a description of stabilizer states and Clifford operations can be obtained from efficient experiments and post-processing, while Sec. VII details how local Hamiltonians and Lindbladians can be characterized efficiently.

II. STATISTICAL BOUND FOR MONTE CARLO ESTIMATION OF THE FIDELITY

A rigorous statistical bound for the scaling of the error of the Monte Carlo fidelity estimate with respect to the number of measurement settings and the number of repetition of each measurement is given by the following theorem.

Theorem 1. *Let $\hat\rho = \sum_i \frac{\rho_i}{d}\hat P_i$ be the decomposition of the pure state $\hat\rho$ over the orthogonal Hermitian operator basis $\{\hat P_i\}$ where $\mathrm{tr}\,\hat P_i\hat P_j = d\delta_{ij}$ and the operator norm $\|\hat P_i\| \leq 1$. One can obtain an estimate $\bar F$ of the fidelity $F(\hat\rho,\hat\sigma)$ between $\hat\rho$ and $\hat\sigma$ with error $\epsilon = \epsilon_1 + \epsilon_2$ such that*

$$\Pr(|F-\bar F| \geq \epsilon) \leq \frac{1}{N_1\epsilon_1^2} + 2\exp\left[-\frac{\epsilon_2^2 N_1^2}{2}\left(\sum_{k=1}^{N_1}\frac{1}{\rho_{i_k}^2 N_2^{[k]}}\right)^{-1}\right] \tag{3}$$

where

- *$I = \{i_1 \dots i_{N_1}\}$ are N_1 indices sampled from $\Pr(i)$, corresponding to observables $\hat{P}_{i_k}$ to be measured experimentally on $\hat{\sigma}$*
- *$N_2^{[k]}$ is the number of experimental samples taken to estimate $\sigma_{i_k} = \operatorname{tr} \hat{P}_{i_k}\hat{\sigma}$*
- *ϵ_1 is the error associated to the Monte Carlo estimate*
- *ϵ_2 is the error associated to the experimental estimation of the $\{\sigma_i\}_{i\in I}$*

Proof. The fidelity $F(\hat{\rho}, \hat{\sigma})$ can be rewritten as

$$F(\hat{\rho}, \hat{\sigma}) = \sum_i{}' \frac{\rho_i^2}{d}\frac{\sigma_i}{\rho_i} \tag{4}$$

where prime indicates that the summation runs only over nonzero values of ρ_i. Since $\operatorname{tr}\hat{\rho}^2 = 1$ by assumption, $\Pr(i) = \rho_i^2/d$ is a normalized probability distribution. We can thus interpret the fidelity as the expectation value of a random variable X which takes value σ_i/ρ_i with probability $\Pr(i)$. Its variance is bounded by a constant, as

$$\mathrm{Var}(X) = \sum_i{}' \frac{\sigma_i^2}{d} - F^2 \leq \operatorname{tr}\hat{\sigma}^2 - F^2 \leq 1, \tag{5}$$

and thus, using Chebyshev's inequality, we obtain

$$\Pr(|F - \bar{F}_1| \geq \epsilon_1) \leq \frac{1}{N_1\epsilon_1^2}, \tag{6}$$

where $\bar{F}_1 = \sum_{i\in I}\sigma_i/\rho_i$ is the estimate of the fidelity by sampling N_1 realizations of X, *i.e.*, by drawing $I = \{i_1 \dots i_{N_1}\}$ indexes from the probability distribution $\Pr(i)$ and estimating $\mathbb{E}(X)$ by the realization of $\bar{X} = N_1^{-1}\sum_{i\in I} X_i$ where all X_i are independent and distributed as X. Thus, the number of measurements settings does not depend on the dimension of the system and scales as $\mathcal{O}(1/\epsilon_1^2)$.

However, the expectation value σ_i of each observables with respect to the experimental state $\hat{\sigma}$ can only be estimated up to finite precision. For each $i_k \in I$, the observable $\hat{P}_{i_k}$ is measured on the experimental state and yields a number $y_{i_k}^{[m]}$ whose absolute value is bounded by the operator norm of the observables. This measurement is repeated $N_2^{[k]}$ times and the approximate realization of X_k is $\tilde{\sigma}_{i_k}/\rho_{i_k} = \left(\rho_{i_k}N_2^{[k]}\right)^{-1}\sum_{m=1}^{N_2^{[k]}} y_{i_k}^{[m]}$. This estimation proceadure is then repeated for each of the N_1 observables. Hoeffding's bound [15] states that, if the *independent* real random variables Y_i are such that $a_i \leq Y_i \leq b_i$, then for $S = Y_1 + Y_2 + \cdots + Y_n$,

$$\Pr(|S - \langle S\rangle| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{7}$$

In our case, for all k, we have $-1/|\rho_{i_k}| \leq y_{i_k}^{[k]}/\rho_{i_k} \leq 1/|\rho_{i_k}|$ for all $N_2^{[k]}$ experimental measurement performed to estimate σ_{i_k} and we can apply the Hoeffding's inequality to all $N = \sum_{k=1}^{N_1} N_2^{[k]}$ experimental samples to bound the distance between the sum $\bar{F}$ of $\tilde{\sigma}_{i_k}/\rho_{i_k}$ by

$$\Pr(|\bar{F} - \bar{F}_1| \geq \epsilon_2) \leq 2\exp\left[-\frac{\epsilon_2^2 N_1^2}{2}\left(\sum_{k=1}^{N_1}\frac{1}{\rho_{i_k}^2 N_2^{[k]}}\right)^{-1}\right]. \tag{8}$$

Finally to reach eq. (3), one applies the union bound to $|F - \bar{F}| \leq |F - \bar{F}_1| + |\bar{F} - \bar{F}_1|$. □

A few remarks are in order. First, it is likely these bounds can be improved upon, as the union bound is rather loose, and is only used to provide an upper bound on the resources which scales well with the desired acuracy. Inequality (8) shows that the smaller $\rho_{i_k}^2$, the greater the number of samples $N_2^{[k]}$ has to be in order to efficiently estimate σ_{i_k}/ρ_{i_k}. However, this does not take into account the fact if $\rho_{i_k}^2$ is small the probability of drawing the index i_k from the probability distribution $\Pr(i) = \rho_i^2/d$ is also small. Using this fact, we will show in Sec. IV C how to improve the scaling of the statistical error. Note also that we have assumed that all experimental uncertainties are of statistical nature. Any additional uncertainty—e.g., a fundamental limitation of the measurement apparatus—could increase the deviation from the true fidelity. Moreover, when measurements are realized directly on an ensemble of particles, such as in NMR-based quantum devices, the number of samples N_2 will correspond to the number of systems in the ensemble, which vastly improves the convergence in practice. Finally, the assumption about the purity of $\hat{\rho}$ can be dropped and the theorem then applies to estimating the quantity $\operatorname{tr}\hat{\rho}\hat{\sigma}/\operatorname{tr}\hat{\rho}^2$ which does not correspond to fidelity but is nevertheless a reasonable similarity measure.

While we have focused on a particular relevance distribution, the relevance distribution can and should be adapted to the target state $\hat{\rho}$, as this can minimise the required experimental and postprocessing resources. For instance, any even positive function f of the expectation values $\{\rho_i\}$ defines a meaningful relevance distribution of the form $\Pr(i) = f(\rho_i)/\Gamma$ where $\Gamma = \sum_i f(\rho_i)$ is a normalization factor. The fidelity

$$F(\hat{\rho}, \hat{\sigma}) = \sum_i{}' \Pr(i)\frac{\Gamma}{d}\frac{\rho_i}{f(\rho_i)}\sigma_i \tag{9}$$

can again be interpreted as the mean value of a random variable X which takes value $\frac{\Gamma}{d}\frac{\rho_i}{f(\rho_i)}\sigma_i$ with probability $\Pr(i) = f(\rho_i)/\Gamma$. The relevant quantities in the theorem, *i.e.,* the variance and the scaling of the repeated measurements, become

$$\mathrm{Var}(X) = \sum_i{}' \frac{\sigma_i^2}{d}\frac{\Gamma}{d}\frac{\rho_i^2}{f(\rho_i)} - F^2 \tag{10}$$

$$N_2^{[k]} \propto \left(\frac{\Gamma}{d}\frac{\rho_{i_k}}{f(\rho_{i_k})}\right)^2. \tag{11}$$

For example, the alternate relevance distribution $\Pr(i) = |\rho_i|/\Gamma$ leads to a variance bounded by Γ/d and a number of required repeated measurements that scales as Γ^2/d^2.

III. SAMPLING FROM THE RELEVANCE DISTRIBUTION

Sampling from the relevance distribution $\Pr(i)$ is not trivial because the dimension of the operator space on n particles is exponentially large in n. Therefore, computing all $\rho_i = \operatorname{tr} \hat\rho \hat P_i$ for all observables $\hat P_i$ is unefficient. Furthermore, computing a given ρ_i can be a challenging task in itself. However, by choosing operators $\hat P_i = \hat p_i^{[1]} \otimes \ldots \otimes \hat p_i^{[n]}$ that are tensor products of single-particle operators—such as the Pauli operators for qubits—sampling can be simplified by recursively picking the observables for each particle as we now demonstrate.

A. Sampling using conditional probabilities

This divide and conquer approach, inspired by [16], relies on the computation of conditional probability distribution, *i.e.*, by computing the probability of picking a given observable knowing which observables have been picked on the previous particles.

Consider for concreteness a system composed of n qubits, and an operator basis $\hat P_i$ all consisting of tensor product of single qubit operators. The Hilbert space dimension is $d = 2^n$ (the argument straightforwardly carry over to particles with more degrees of freedom). The probability $q_{i_1,\ldots,i_n}$ of picking the observable $\hat P_i = \bigotimes_{m=1}^n \hat p_{i_m}^{[m]}$ can be rewritten as the product of conditional probabilities

$$q_{i_1,\ldots,i_n} = \prod_{k=1}^n q_{i_k|i_1,\ldots,i_{k-1}} \tag{12}$$

where the conditional probability $q_{i_k|i_1,\ldots,i_{k-1}}$ of drawing the observable $\hat p_{i_k}^{[k]}$ on particle k knowing which observables have been picked on the previous particles is

$$q_{i_k|i_1,\ldots,i_{k-1}} = q^{-1}_{i_1,\ldots,i_{k-1}} \sum_{I=i_{k+1},\ldots,i_n} q_{i_1,\ldots,i_k I}, \tag{13}$$

i.e., the sum of probabilities over all possible choice of observables $\hat P$ acting on the $n-k$ remaining particles, divided by the probability of picking the observable $\bigotimes_{m=1}^{k-1} \hat p_{i_m}^{[m]}$ on the $k-1$ previous particles. Using equation (12), sampling from the probability distribution reduces to sequentially picking an observable on each particle according to the conditional probability distribution (13) which can be written, up to a normalization, as

$$q_{i_k|i_1,\ldots,i_{k-1}} \propto \sum_{\hat P \in \mathcal{P}_{n-k}} \operatorname{tr}\left[\left(\hat\rho \times \left(\bigotimes_{m=1}^{k} \hat p_{i_m}^{[m]} \otimes \hat P\right)\right)^{\otimes 2}\right], \tag{14}$$

where the trace of two copies accounts for the square in the definition of $\Pr(i)$ since $\operatorname{tr} \hat A^{\otimes 2} = \left(\operatorname{tr} \hat A\right)^2$. Thus, we now have two copies of the state $\hat\rho = |\psi\rangle\langle\psi|$. It is convenient to number the $2n$ particles from 1 to n for the first copy and $n+1$ to $2n$ for the second copy. Grouping particles m and $n+m$ into pairs, eq. (14) reduces to

$$\operatorname{tr}\left[\hat\rho^{\otimes 2}\left(\bigotimes_{m=1}^{k} (\hat p_{i_m})^{\otimes 2} \otimes \sum_{\hat P \in \mathcal{P}_{n-k}} \hat P^{\otimes 2}\right)\right] \tag{15}$$

where the same observable acts on both particles in a pair. The sum over all duplicated observables $\hat P^{\otimes 2}$ can be written as the tensor product of operators acting on each pair $[m,\, n+m]$ of particles

$$2^{-(n-k)} \sum_{P \in \mathcal{P}_{n-k}} P \otimes P = \bigotimes_{m=k+1}^{n} \Omega^{[m,\, n+m]} \tag{16}$$

where $\Omega^{[i,j]} = \frac{1}{2}\sum_m \hat p_m^{[i]} \otimes \hat p_m^{[j]}$ is an observable acting on the pair of particles (i, j). For instance, for the Pauli operator basis, the Ω is the SWAP operator. Thus, the conditional probability is proportionnal to

$$\operatorname{tr}\left[\hat\rho^{\otimes 2}\left(\bigotimes_{m=1}^{k} (\hat p_{i_m})^{\otimes 2} \bigotimes_{m=k+1}^{n} \Omega^{[m,\, n+m]}\right)\right] \tag{17}$$

which is the *expectation value of a tensor product of 2-local observables* on the state $\hat\rho \otimes \hat\rho$ on $2n$ particles.

B. Bound on the complexity of sampling

The problem of sampling reduces to, for each of the n particles, *i)* computing conditional probabilities for each of the possible observables acting on that particle *ii)* pick one of those observables by generating a random number. Conditional probabilities can be expressed as expectation value through eq. (17). Thus, if computing expectation values on tensor product of local observables on states of n particles has complexity $q(n)$, generating an index $i = i_1 \ldots i_n$ from the relevance distribution Pr(i) has complexity at most $n \times q(2n)$.

For many states of interest, computing expectation values of observables can be performed in polynomial time, *i.e.*, $q(n) \in \operatorname{poly}(n)$. That is the case for many families of tensor-network states such as matrix product states (MPS) [17] which are known to represent faithfully ground states of interesting many-body Hamiltonians in 1D [14]. Their natural extension to 2D, projected entangled pair states (PEPS) [18] also allows the efficient heuristic computation of such expectation values.

As an example, we will show that for MPS, sampling from the relevance distribution $\Pr(i)$ can in fact be done in a time that scales linearly with n and polynomially with the bond dimension by contracting the tensor network represented on Fig. 1. In particular, this implies that the general bound $n \times q(2n)$, which would be n^2 for MPS, is not tight. In the case of MPS, the improvement is due to the possibility of precomputing the state $\bigotimes_{m=k+1}^{n} \Omega^{[m,\, n+m]}\, \hat\rho$ for all k. Precomputation of this family of states is useful since they are needed to compute conditional values, c.f. eq. (17). Crucially, for MPS, these states are described by a linear number of parameters and can thus be stored efficiently.

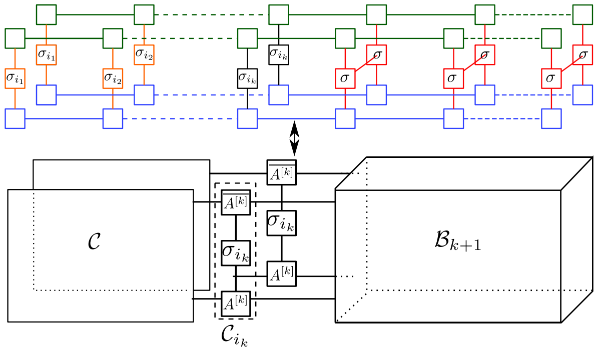

FIG. 1: Tensor network corresponding to eq. (17) if $\hat{\rho} = |\psi\rangle\langle\psi|$ is a MPS, *i.e.*, there exist a famility of matrices $\left\{A^{[k]}_{i_k}\right\}$ such that $|\psi\rangle = A^{[1]}_{i_1}\dots A^{[n]}_{i_n}|i_1\dots i_n\rangle$. The upper figure represent the individual tensors in the tensor network. Each square represent a tensor and outgoing legs represent the tensor indices. Two squares connected by a line are the contraction of the corresponding indices of two tensors. Red squares correspond to the Ω operators. Orange squares correspond to the Pauli operators already chosen on the $k-1$ previous qubits. Blue squares are the MPS tensors of the two copies of $|\psi\rangle$ while the green squares are the MPS tensors of the two copies of $\langle\psi|$. The lower figure correspond to the partial contraction of the tensor network.

Consider the partially contrated tensor network at the bottom of Fig. 1. The block $\mathcal{B}_{k+1}$ on the right corresponds to the application of the operator Ω on each of the pair of particles $[m,\, n+m]$ for $m > k$. In practice, all blocks $\{\mathcal{B}_k\}_{k\in\{2\dots n\}}$ are precomputed since they only depend on the MPS, not on the Pauli operator being picked. The two rectangles $\mathcal{C}$ on the left correspond to the partial computation of expectation values for the Pauli operators already drawn on the $k-1$ first qubits. The two columns $\mathcal{C}_{i_k}$ results from the contraction of σ_{i_k} with the k^{th} tensors of the MPS representation of $|\psi\rangle$ and $\langle\psi|$. The index i_k can take 4 possible values corresponding to the four Pauli operators.

Picking a Pauli operator for each site k is performed by computing four tensors $\mathcal{C}'_{i_k}$, each of which result from the contraction of the already computed tensor $\mathcal{C}$ with one of the four columns $\mathcal{C}_{i_k}$ corresponding to the 4 possibles choices of Pauli operator on the k^{th} qubit. Then, contracting two copies of $\mathcal{C}'_{i_k}$ with $\mathcal{B}_{k+1}$ yields the conditional probability of picking operator σ_{i_k} on the k^{th} qubit.

Repeating this procedure for each of the 4 possible values of i_k, we end up with four probabilities from which one Pauli operator $\sigma^{[k]}_{i_{k_0}}$ is picked. The column $\mathcal{C}$ is then updated according to $\mathcal{C} = \mathcal{C}'_{i_{k_0}}$ and the process is iterated for $k+1$. Note that all those tensors have a few outgoing bonds. Thus, the cost of contraction remains polynomial in the bond dimension.

A larger class of multi-qubit states for which sampling can be done efficiently by computing conditional probabilities are computationally tractable (CT) states [19]. CT states are states in which *(a)* the overlap with any element of the computational basis can be computed efficiently, and *(b)* it is possible to sample from the distribution of outcomes from measurements in the computational basis efficiently. For such states, it is possible to efficiently compute the expectation value of tensor products of Pauli observables which only permute elements of the computational basis and thus are basis preserving.

In the generic case of a state defined as a vector of the Hilbert space, computing the the expectation value of a single local observable will take time $\mathcal{O}(2^{2n})$ since we have to account for the Hilbert space of $2n$ qubits. A tensor product of local observables can be thought as the product of $\mathcal{O}(n)$ observables that act non-trivially on a few qubits. Thus, computing the expectation value given by equation (17) will take time $\mathcal{O}\left(n\, 2^{2n}\right)$. In order to sample, such a computation has to be repeated for each particles, leading to an overall complexity of sampling from the relevance distribution of $\mathcal{O}\left(n^2\, 2^{2n}\right)$ in the worst case.

IV. FLUCTUATION OF THE EXPECTATION VALUES

As pointed out in caveat **C2**, our method is sensitive to the distribution of expectations values $\{\rho_{i_k}\}$ of the target state. Namely, for a given measurement setting used to estimate σ_{i_k} on the experimental state, the number of repeated measurements scales roughly as $(\epsilon\rho_{i_k})^{-2}$ in order to estimate fidelity up to precision ϵ.

Naively, this implies that the number of measurement goes as $\max_{i|\rho_i\neq 0}\rho_i^{-2}$ in the worst case. Therefore, if the smallest non-zero ρ_i is polynomially small, our method is efficient. This very stringent condition holds true for many interesting states, as we discuss in the next section. We will show how to improve on this naive estimate by truncating the sum to leave out the entries with small $|\rho_i|$. This does not affect the fidelity estimate since such entries are not likely to show up when sampling the relevance distribution $\Pr(i) = \rho_i^2/d$.

A. Stabilizer states

Some of the most promising approaches to universal and scalable quantum computation are teleportation-based quantum computation [20] and measurement-based quantum computations [21]. Both these approaches rely heavily on the preparation of stabilizer states [22] and the application of quantum operations known as the Clifford group [20], which map stabilizer states to stabilizer states. Stabilizer states are also important for quantum computation in general because of their close relationship to a large class of quantum error correction codes known as *stabilizer codes*. Many of the experimental demonstrations of state preparation to date have been of stabilizer states such as states encoded into stabilizer codes [2], cluster states [4], and the GHZ state $|00\cdots0\rangle + |11\cdots1\rangle$ [5, 6].

Stabilizer states are defined to be $+1$ eigenstates of some set of commuting Pauli operators $\hat{S}_j$ that generate the stabilizer group, *i.e.* $\hat{S}_j|\psi\rangle = |\psi\rangle$ for all $j = 1,\dots n$. It follows that $\Pr(i) = 1/d$ if either of $\pm\hat{P}_i$ is in the stabilizer group and 0 otherwise. Sampling from $\Pr(i)$ thus amounts to generating an index i uniformly between 1 and d, avoiding the problem associated with caveat **C1**. For the same reasons, $\rho_i^2 = 1$ for all i with $\Pr(i) \neq 0$, so that the uncertainity in the estimation of σ_i is not amplified, avoiding the problem associated with caveat **C2**. It also follows that the fidelity $F(\hat{\sigma}, \hat{\rho})$ to a stabilizer state $\hat{\rho}$ can be estimated with error ϵ using $N = \mathcal{O}(\frac{1}{\epsilon^4})$ experiments involving only local projective measurements, *independently of the system size and without any prior knowledge of the experimental state $\hat{\sigma}$*. Since this result relies only on local measurements, it can immediatelly be generalized to states which are locally equivalent to stabilizer states.

B. States with polynomially small expectation values

Stabilizer states are the simplest example of states on n qubits for which non-zero ρ_i scale nicely since they all have value ± 1, leading to a uniform distribution. However, states for which non-zero ρ_i decrease as $1/\mathrm{poly}(n)$ also require only polynomially many measurements.

While sampling can be performed efficiently for MPS, the expectation values of a random MPS on a local operator basis can be exponentially small. However, for many states of interest, this is not the case. Two cases of interest, for which the smallest non-zero expectation value of Pauli operators scale as $1/n$ for n qubits, are the W state [3] and the $|t_n\rangle$ state used in linear optics for heralded teleportation with high success probability [13]. Both are MPS with bond dimension 2 and are uniform superpositions of a linear number of computational-basis states. The MPS representation of the W state is well-known [23] while the representation for $|t_n\rangle$ is

$$A_0^{[1]} = \begin{pmatrix} 1 & 1 \end{pmatrix} \quad A_0^{[k]} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad A_0^{[n]} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$A_1^{[1]} = \begin{pmatrix} 0 & 2 \end{pmatrix} \quad A_1^{[k]} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \quad A_1^{[n]} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

For both states, the expectation value of a Pauli operator $\hat{P}$ is given by

$$\langle\psi|\hat{P}|\psi\rangle = \alpha(n) \sum_{i,j} \langle i|\hat{P}|j\rangle \tag{18}$$

where $|\psi\rangle$ is either state, $\alpha(n)$ is $1/n$ for the W state and $1/(n+1)$ for $|t_n\rangle$, and the sum runs over computational states that appear in the decomposition of the state. For all i, j, there exists a Pauli operator σ_{ij} such that $|j\rangle = \hat{\sigma}_{ij}|i\rangle$. Since the Pauli operators form a group, $\hat{P}\hat{\sigma}_{ij}$ is another Pauli operator and all terms appearing in the sums are ± 1. Thus, the smallest non-zero Pauli expectation scales as $1/n$, and the number of samples required to estimate σ_i/ρ_i to constant accuracy scales as n^2 in the worst case.

C. Truncating negligible expectation values

As mentioned above, very small non-zero ρ_i are unlikely to show up when sampling of the relevance distribution $\Pr(i) = \rho_i^2/d$. Furthermore, small expectation values have a limited impact on the fidelity. To see this, define the set of negligible expectation values as $S \equiv \{\rho_i \text{ such that } |\rho_i| < d^{-\alpha}\}$ where α is a positive number to be determined. We split the fidelity estimate into negligible and relevant expectation values

$$\sum_i \frac{\rho_i \sigma_i}{d} = \sum_{\rho_i \notin S} \frac{\rho_i \sigma_i}{d} + \sum_{\rho_i \in S} \frac{\rho_i \sigma_i}{d} \tag{19}$$

and we bound the absolute value of the sum over negligible terms by

$$\left| \sum_{\rho_i \in S} \frac{\rho_i \sigma_i}{d} \right| \leq \sum_{\rho_i \in S} \frac{|\sigma_i|}{d} \max_{i \in S} |\rho_i| \leq d^{-(\alpha+1)} \sum_{\rho_i \in S} |\sigma_i|. \tag{20}$$

The sum of a subset of $|\sigma_i|$ is bounded by the sum over *all* $|\sigma_i|$. To bound $\sum_i |\sigma_i|$, we can use the constraint on the purity of the state $\sum_i \sigma_i^2 = d \operatorname{tr} \hat{\sigma}^2 \leq d$. The sum of absolute values is maximal when all absolute values are equal, which follows from standard Lagrange multiplier techniques. The purity constraint finally leads to

$$\sum_i |\sigma_i| \leq d\sqrt{d \operatorname{tr} \hat{\sigma}^2} \leq d^{3/2}. \tag{21}$$

Inserting this inequality that into eq. (20) yields

$$\left| \sum_{\rho_i \in N} \frac{\rho_i \sigma_i}{d} \right| \leq d^{1/2-\alpha}. \tag{22}$$

Hence, the sum over negligible ρ_i vanishes exponentially for $\alpha = (1+\epsilon)/2$, *i.e.*, when we drop all expectation values smaller than $d^{-\frac{1+\epsilon}{2}}$ in absolute value, for any constant $\epsilon > 0$.

We thus modify the sampling method in the following way. For each observable $\hat{P}_i$ picked from sampling the relevance distribution, compute the corresponding expectation value $\rho_i = \operatorname{tr} \hat{\rho}\hat{P}_i$. When ${\rho_i}^2 < d^{-1-\epsilon}$, reject this entry, otherwise you proceed as before. It is important to verify that this modification does not slow down the procedure, i.e. that we are not constantly rejecting samples. To see this, notice that the probability of choosing an element from the negligible set is bounded by

$$\sum_{\rho_i \in S} \frac{\rho_i^2}{d} \leq \sum_{\rho_i \in S} d^{-2-\epsilon} \leq d^{-\epsilon}. \tag{23}$$

Since we reject all negligible ρ_i, the maximum number of repeated measurements needed for a given experimental setting scales in the worst case as $d^{1+\epsilon}$. In particular, for qubits, the maximum number of measurements is $2^{n(1+\epsilon)}$. Moreover, since the number of measurement settings does not scale with the size of the system, the total number of measurements scales as $2^{n(1+\epsilon)}$ which is a *quadratic improvement over the number of measurements needed to perform brute-force tomography* on a generic state of n qubits.

D. States with more structure

The previous bound is fully general as it allows to truncate negligible expectation values for any quantum state, but as a consequence the scaling remains exponential. For some classes of states, taking advantage of their structure can result in a drastic improvement, as we have seen for stabilizer states or the W and $|t_n\rangle$ states. Since those last two examples are MPS, it is natural to investigate the behavior of the expectation values for a generic MPS.

A general MPS of bond dimension χ can be written

$$|\psi\rangle = \sum_{\alpha_1\ldots\alpha_{n-1}=1}^{\chi} |\phi^{[1]}_{\alpha_1}\rangle \lambda^{[1]}_{\alpha_1} |\phi^{[2]}_{\alpha_1\alpha_2}\rangle \ldots \lambda^{[n-1]}_{\alpha_{n-1}} |\phi^{[n]}_{\alpha_{n-1}}\rangle \quad (24)$$

where the real parameters $\lambda_{\alpha_k\alpha_{k+1}}$ are the eigenvalues of the reduced density matrix of the k^{th} particle. If we group particles into block of size m such that $\chi^2 < 2^m$, we will take advantage of the presence of zero eigenvalues in the reduced density matrix due to the MPS structure. For each block we can take the at most χ^2 eigenvectors corresponding to non-zero eigenvalues and choose a basis of operators for that block containing $2\chi^2$ operators. By taking tensor product of those local observables, we end up with at most $\left(2\chi^2\right)^{n/m} < 4^n$ observables on which $|\psi\rangle\langle\psi|$ can be decomposed. At the cost of measuring observables blocks of $m = 2\lfloor\log_2\chi\rfloor + 1$ qubits, we can reduce the number of non-zero ρ_i even if this number will remain exponential in the number of particles. However, numerically, the expectation values ρ_i were found to be exponentially small for generic states, resulting in a blow-up of repeated experimental measurements to estimate σ_i/ρ_i. Thus, it remains an open problem to assess whether there exists a way to tailor an observable basis for a given generic MPS.

V. MONTE CARLO PROCESS CERTIFICATION

Thanks to the Choi-Jamiołkowski isomorphism [24, 25], an algorithm similar to the one used to estimate state fidelity can be implemented for the certification of unitaries. The Choi-Jamiołkowski isomorphism states that any quantum process $\mathcal{U}$ can be uniquely described by its action on half of a maximally entangled state such as $|\phi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i\rangle\otimes|i\rangle$, leading to the state

$$\hat{\rho}_{\mathcal{U}} = (\mathrm{id}\otimes\mathcal{U})\left(|\phi\rangle\langle\phi|\right) \quad (25)$$

where id is the identity superoperator. The density operator $\hat{\rho}_{\mathcal{U}}$ is often called the *Choi matrix* of $\mathcal{U}$. As with state certification, the idea is to compare a target unitary $\mathcal{U}$ to an arbitrary quantum process $\tilde{\mathcal{U}}$ acting on the same space through a distance measure, namely the average output fidelity $\overline{F}$ given by [26]

$$\overline{F}(\mathcal{U},\tilde{\mathcal{U}}) = \frac{d\,F_e(\mathcal{U},\tilde{\mathcal{U}})+1}{d+1}, \quad (26)$$

where $F_e(\mathcal{U},\tilde{\mathcal{U}}) = \mathrm{tr}\,\hat{\rho}_{\mathcal{U}}\hat{\rho}_{\tilde{\mathcal{U}}}$ is called the *entanglement fidelity* between the two processes [27]. This reduces the problem of comparing two processes to the problem of comparing two states. Thus, techniques described for state fidelity estimation can be used for the estimation of the entanglement fidelity, which in turn leads to the average gate fidelity through eq. (26).

One should note that the operator basis coefficients of $\hat{\rho}_{\tilde{\mathcal{U}}}$ can be estimated *without the preparation of the maximally entangled state used in the isomorphism.* Naively, one would prepare a maximally entangled state, measure the first half of the state in some basis, apply $\tilde{\mathcal{U}}$ to the other half of the state, and then measure the second half in some basis. This naive approach corresponds to Fig. 2 a). A more practical method is to prepare an eigenstate of the first measurement uniformly at random, apply local unitaries that depend only on the maximally entangled state chosen, apply $\tilde{\mathcal{U}}$, and then measure the output with the second observable, as represented in Fig. 2 b). Averaging over an ensemble of such random choices leads to an estimate of any operator basis coefficient of $\hat{\rho}_{\tilde{\mathcal{U}}}$ without the preparation of the larger state.

The processes which are particularly tailored to our protocol are transformations in the Clifford group, that take stabilizer states to stabilizer states. Using this property, the procedure described for stabilizer states can be turned into a procedure to estimate the *average gate fidelity* between any Clifford ideal transformation and an experimental transformation, and with minor modifications this protocol can be made to only require the preparation of product states and single qubits measurements. In the case of Clifford transformations similar results can be obtained using "twirling" experiments [28–31], although the Monte Carlo approach described here generalizes to other cases.

Operations in the Clifford group are not sufficient to perform universal computation, either quantum or classical [20, 32]. In order to achieve quantum universality, one must consider at least one operation outside the Clifford group, such as a $45°$ qubit rotation. While arbitrary combinations of Clifford operations and these rotations may not be characterized efficiently using the scheme described here, Clifford operations followed by these local rotations can always be characterize efficiently thanks to local equivalence. Similarly, if these operations are implemented indirectly via teleportation with the so called "magic states" [20, 33], the resource states are locally equivalent to stabilizer states, and thus can also be certified efficiently using the protocol presented above.

The average output fidelity is not as powerful as other measures of distance between quantum processes such as the diamond norm [34] or the worst case or minimal fidelity [35]. These extremal measures allow for the strongest statements about the distinguishability of two quantum processes. However, we do not expect that it is possible to estimate them efficiently, since in a number of special cases this would reduce to efficient algorithms to solve QMA complete problems [36–38]. Furthermore, concentration of measure arguments [29, 39] guarantee that, for large dimensions, the vast majority of input states will have output fidelity close to the average. Similar results hold for estimation of the fidelity based on the χ matrix representation, where the χ matrix elements can be estimated individually but with more stringent

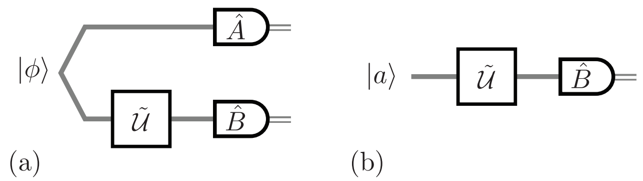

FIG. 2: Equivalent quantum circuits for the measurement of the expectation values of $\hat{A} \otimes \hat{B}$ for the Choi matrix of some process $\tilde{\mathcal{U}}$. Here $|a\rangle$ is locally equivalent to an eigenstate of the observable $\hat{A}$, and it is chosen uniformly at random.

experimental requirements [40, 41].

Other processes

Another quantum process for which the sampling of observables can be done efficiently is the approximate quantum Fourier transform (AQFT) [42] which is a key component in Shor's algorithm for efficient integer factoring [43]. The AQFT Choi matrix is a MPS since its circuit presents a staircase structure [44] which is known to lead to a MPS description [23]. This is equivalent to using the tensor network description of the AQFT [45] which has small treewidth. Its bond dimension χ is related to the degree of approximation of the AQFT. If the approximation is logarithmic in the number of qubits, which is known to guarantee polynomial accuracy in the output state, the bond dimension is polynomial in the number of qubits. This allows for efficient sampling of the expectation values. However, we have observed numerically that those expectation values in the Pauli basis are exponentially small. As in the case of general MPS, the existence of a tailored basis with better scaling remains an open problem. Nevertheless, since the classical processing is efficient, this approach would greatly simplify previous demonstrations of quantum Fourier transform circuits [46], and allow for the certification of implementations with more qubits.

Another larger class of operations for which sampling is efficient are operations which map CT states to CT states, known as *sparse-basis preserving (SBP) operations* [19]. Beside Clifford operations and the quantum Fourier transform (which have simpler sampling algorithms), circuits composed of classical logic circuits acting on qubits, among many others, are SBP operations [19]. Still, the best upper bound we can derive on the number of measurements is exponentially large for a general SBP operation..

VI. LEARNING STABILIZER STATES

Clifford operations can be specified by how they transform a generating set of the Pauli group and thus can be learned efficiently provided one can learn stabilizer states efficiently. We will first present an algorithm to learn a state guaranteed to be a stabilizer state before adapting it to any input state.

A. Ideal case

Due to their importance in fault-tolerant quantum computation and other tasks, stabilizer states have been studied extensively, and an efficient learning algorithm has been proposed by Gottesman and Aaronson [47]. This algorithm exploits the fact that $\langle \hat{P}_i \rangle = \pm 1$ if $\pm \hat{P}_i$ is in the stabilizer group of the state in question, and $\langle \hat{P}_i \rangle = 0$ otherwise. Choosing $\hat{P}_i$ at random is not efficient, as the Pauli group has size $d^2 = 4^n$, and the stabilizer group $\mathcal{G}$ has size $d = 2^n$. However, the Pauli group also has the property that any two Pauli operators either commute or they anti-commute. In fact, for a given Pauli operator (distinct from the identity), it commutes with half the Pauli operators and anticommutes with the other half—thus splitting the Pauli group in half through the commutation relations. This property can be harnessed to learn an unkown stabilizer state $\hat{\sigma}$ by identifying n generators of its stabilizer group $\mathcal{G}_{\hat{\sigma}}$ in the following way:

1. Choose a Pauli operator $\hat{Q}$ at random, under the constraints that: (a) $\hat{Q}$ must commute with all known generators of the stabilizer group while not being generated by them, and (b) $\hat{Q}$ must anti-commute with at least one of the operators known to be outside the stabilizer group and not be generated by them.

2. Experimentally measure the expectation value $\langle \hat{Q} \rangle_{\hat{\sigma}}$.

   (a) If $\langle \hat{Q} \rangle_{\hat{\sigma}} = 0$, then $\hat{Q}$ is not in the stabilizer group, and it should be added to the set of operators outside from the stabilizer group $\mathcal{G}_{\hat{\sigma}}$.

   (b) Otherwise, $\langle \hat{Q} \rangle_{\hat{\sigma}} = \pm 1$ and then $\pm \hat{Q}$ is in the stabilizer group $\mathcal{G}_{\hat{\sigma}}$ (depending on the sign of the expectation value), and $\hat{Q}$ should be added to the generators of stabilizer group $\mathcal{G}_{\hat{\sigma}}$.

3. Continue until n stabilizer generators are identified.

In this ideal setting, every measured expectation value halves the number of candidates to the stabilizer group. Thus, even though the Pauli group is initally exponentially large, the number of rounds scales linearly with n. This implies that *the total number of measurements is linear in n.* All classical processing tasks required in a round can be performed in poly(n) time, by using the symplectic representation of the Pauli group.

B. General scheme

Suppose that the experimental state $\hat{\sigma}$ is not a stabilizer state. We will now adapt the previous algorithm into one that *i)* aborts if the experimental state is far from any stabilizer state and *ii)* eliminates all but one candidate stabilizer state $\hat{E}(\hat{\sigma})$ that could be close to the experimental state. The fidelity of that candidate to the experimental state can be estimated in a second phase using the Monte Carlo scheme described earlier.

We adapt the previous algorithm by putting threshold around the values of $\langle\hat{Q}\rangle_{\hat{\sigma}}$, *i.e.,* we replace step 2 by

2. Experimentally measure the expectation value $\langle\hat{Q}\rangle_{\hat{\sigma}}$.
    (a) If $\left|\langle\hat{Q}\rangle_{\hat{\sigma}}\right| \leq \epsilon$, $\hat{Q}$ is not in the stabilizer group.
    (b) If $\left|\left|\langle\hat{Q}\rangle_{\hat{\sigma}}\right| - 1\right| \leq \epsilon$, $\pm\hat{Q}$ is in the stabilizer group.
    (c) Otherwise, abort the procedure.

To quantify the distance between the experimental state and stabilizer states, we use the following lemma.

Lemma. *For any states $\hat{\sigma}$,$\hat{\rho}$ and any Pauli operator $\hat{Q}$, the difference between the expectation values is bounded by twice the trace distance*

$$\left|\langle\hat{Q}\rangle_{\hat{\rho}} - \langle\hat{Q}\rangle_{\hat{\sigma}}\right| \leq 2\mathcal{D}(\hat{\rho},\, \hat{\sigma}) \tag{27}$$

Proof. Consider $\hat{\Pi} = \left(\mathrm{id} + \hat{Q}\right)/2$, the projector on the $+1$ eigenspace of $\hat{Q}$. The trace distance $\mathcal{D}(\hat{\rho},\, \hat{\sigma})$ is the maximal probability of distinguishing the two states [48]. $\hat{\Pi}$ is a projective measurement that could be uses to distinguish the two states. Thus $\left|\mathrm{tr}\,\hat{\Pi}\hat{\rho} - \mathrm{tr}\,\hat{\Pi}\hat{\sigma}\right| \leq \mathcal{D}(\hat{\rho},\, \hat{\sigma})$. □

Let us now consider the two possible outcomes of the procedure. If the procedure aborts, a least one expectation value $\langle\hat{Q}\rangle_{\hat{\sigma}}$ is ϵ-away from the expectation value of any stabilizer state. Thus, lemma VI B implies that the trace distance of the experimental state $\hat{\sigma}$ to any stabilizer state $\hat{\rho}$ is greater than $\epsilon/2$. If the procedure succeeds, it yields a candidate stabilizer state $\hat{E}(\hat{\sigma})$ whose distance to the experimental state $\hat{\sigma}$ will be checked by estimating their fidelity by our Monte-Carlo method *which requires resources independent of the system size since the target state is a stabilizer state*. It remains to determine what fidelity level guarantees that $\hat{E}(\hat{\sigma})$ is the closest stabilizer state.

Consider two distinct stabilizer states. There exists a Pauli operator belonging to the stabilizer group of one of the state and not the other. Expectation values with respect to this Pauli operator are 0 for one state and ± 1 for the other. Lemma VI B thus shows that any two distinct stabilizer states have trace distance greater than $1/2$. Therefore, if the experimental state is within $1/4$ to the candidate stabilizer state $\hat{E}(\hat{\sigma})$ in trace distance, then $\hat{E}(\hat{\sigma})$ is the closest stabilizer state. A sufficient condition on fidelity for this is

$$15/16 < F\left(\hat{E}(\hat{\sigma}),\, \hat{\sigma}\right). \tag{28}$$

In fact, we can improve criterion (28) in the case that the candidate stabilizer state comes from our identification procedure which has ruled out all other stabilizer states. For any of those rejected stabilizer state $\hat{\rho}$, there exists a Pauli measurement $\hat{Q}$ such that $\langle\hat{Q}\rangle_{\hat{\sigma}}$ is ϵ-close to 1 and for which $\langle\hat{Q}\rangle_{\hat{\rho}} = 0$. Thus, lemma VI B implies that $\mathcal{D}(\hat{\rho},\, \hat{\sigma}) \geq (1-\epsilon)/2$ and it is thus sufficient to certify that state $\hat{E}(\hat{\sigma})$ is within $(1-\epsilon)/2$ in trace distance to the experimental state, which turns into a less stringent condition on fidelity

$$\frac{(1+\epsilon)(3-\epsilon)}{4} < F\left(\hat{E}(\hat{\sigma}),\, \hat{\sigma}\right). \tag{29}$$

To sum up, if criteria (28) or (29) is fulfilled, $\hat{E}(\hat{\sigma})$ is the closest stabilizer state to the experimental state.

VII. LEARNING LOCAL HAMILTONIANS AND LINDBLADIANS

Other models of universal quantum computation exist where the idea of discrete gates is not a natural fit. These include adiabatic quantum computing (AQC) [49] and dissipation-driven quantum computing (DDQC) [50]. In order to perform state transformations in these models, an experimentalist must smoothly vary control fields to go from a simple local Hamiltonian (or Liouvillian, in the case of DDQC) to a Hamiltonian which encodes the final state of the computations as its ground state.

The initial state is often taken to be the $|00\cdots 0\rangle$, which can easily be certified. An experimental implementation of such a computational model must demonstrate that such smooth evolutions can be implemented reliably. The most direct way to do this is to estimate the Hamiltonian of the system, and explicitly check how it compares against the ideal target Hamiltonian. The scalability of this problem for arbitrary Hamiltonians fails in two ways: *(i)* The number of parameters in an arbitrary Hamiltonian operator grows exponentially with system size, *(ii)* the linearization of the exponential of the Hamiltonian is only valid for a short time compared to the inverse of the operator norm of the Hamiltonian, which is extensive, and thus larger system would require faster and faster measurements. As aluded to, this need not be the case. The Hamiltonian can be restricted to be 2-local and embedded in a square lattice and the model remains universal [51, 52]. This immediately reduces the number of parameters in the instantaneous Hamiltonian from exponential to linear. In order to see how the second problem can be dealt with, we must look more carefuly at the idea of locality in quantum mechanics.

We consider *local* Hamiltonians acting on n particles of the form $\hat{H} = \sum_X \hat{H}_X$ where X label subsets of n particles, each term has bounded norm $\|\hat{H}_X\| \leq E$, and acts on at most k neighbouring particles, such that $H_X = 0$ when $|X| > k$. The evolution of an operator is governed by the equation $\frac{\partial}{\partial t}\hat{A}(t) = i[\hat{H}, \hat{A}]$.

Local Lindbladians are defined similarly in terms of a local Hamiltonian H and dissipation operators $\hat{L}_X$ that are local ($\hat{L}_X = 0$ when $|X| > k$) and bounded $\|\hat{L}_X\| \leq E$. The Lindblad equation, governing the evolution of dissipative quantum systems, is

$$\begin{aligned}\frac{\partial}{\partial t}\hat{A}(t) &= i[\hat{H}, \hat{A}(t)] \\ &\quad + \sum_X \hat{L}_X^\dagger \hat{A}(t)\hat{L}_X - \frac{1}{2}(\hat{L}_X^\dagger \hat{L}_X \hat{A}(t) + \hat{A}(t)\hat{L}_X^\dagger \hat{L}_X).\end{aligned} \tag{30}$$

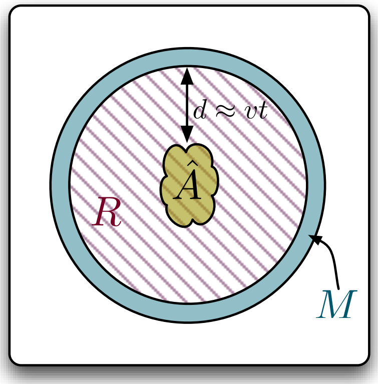

FIG. 3: The Lieb-Robinson bound [53, 54] states that, given an observable $\hat{A}$, evolution under a local Hamiltonian for time t will lead to correlations localized to a region R of radius $d = vt$ around $\hat{A}$, where the velocity v is an intensive quantity, and which is separated from the rest of the system by a membrane M.

We demonstrate here how to learn the value of a local Hamiltonian using only *(i)* the preparation of initial product states, *(ii)* the simultaneous measurement of a constant number of single-qubit operator, *(iii)* a number of experimental settings that grows linearly with the system size, *(iv)* and classical post-processing of complexity n^3 (inverting an $cn \times cn$ matrix for some constant c). The case of local Lindbladians is a straightforward generalization.

The crux of the argument uses the Lieb-Robinson bound [53, 54] that shows that a local Hamiltonian generates a causal evolution, with effects propagating at a finite velocity v (note that this bound has been generalized to the setting of dissipative systems [55], so our derivation holds for local Lindbladians as well). Break the Hamiltonian into $\hat{H} = \hat{H}_0 + \hat{H}_M$, where $\hat{H}_M$ contains all the terms $\hat{H}_X$ that intersect a membrane M surrounding the operator $\hat{A}$ in Fig. 3. The idea of this membrane is to disconnect its interior, denoted region R, from the rest of the particles. Indeed, $e^{i\hat{H}_0 t}\hat{A}e^{-i\hat{H}_0 t} = e^{i\hat{H}_R t}\hat{A}e^{-i\hat{H}_R t}$ where $\hat{H}_R$ in the Hamiltonian acting only inside the membrane, see Fig. 1 in the main body. The differential equation for $\hat{A}(t)$ is

$$\frac{\partial}{\partial t}\hat{A}(t) = i[\hat{H}_0, \hat{A}(t)] + i[\hat{H}_M, \hat{A}], \tag{31}$$

which has solution

$$\begin{aligned}\hat{A}(t) =& e^{i\hat{H}_0 t}\hat{A}(0)e^{-i\hat{H}_0 t}\\ &+ i\int_0^t e^{i\hat{H}_M(t-s)}[\hat{H}_M, \hat{A}(s)]e^{-i\hat{H}_M(t-s)}ds \qquad (32)\\ =& e^{-i\hat{H}_R t}\hat{A}(0)e^{i\hat{H}_R t}\\ &+ i\int_0^t e^{i\hat{H}_M(t-s)}[\hat{H}_M, \hat{A}(s)]e^{-i\hat{H}_M(t-s)}ds \qquad (33)\end{aligned}$$

as can be verified directly by differentiation. The commutator appearing in the second term can be bounded by the Lieb-Robinson bound:

$$\|[\hat{H}_M, \hat{A}(s)]\| \leq cV\|\hat{A}\|\|\hat{H}_M\| \exp\left(-\frac{d-vt}{\xi}\right) \tag{34}$$

where V is the number of sites in the support of the observable $\hat{A}$, and c, v, and ξ are constant that depend only on the microscopic details of the system, independent of the system size. Integrating, we obtain

$$\begin{aligned}&\|\hat{A}(t) - e^{i\hat{H}_R t}\hat{A}(0)e^{-i\hat{H}_R t}\| \qquad (35)\\ &\leq ctV\|\hat{A}\|\|\hat{H}_M\|\| \exp\left(-\frac{d-vt}{\xi}\right).\end{aligned}$$

Expanding the exponential to first order yields

$$\begin{aligned}&\|\hat{A}(t) - \hat{A}(0) - it[\hat{H}_R, A(0)]\| \qquad (36)\\ &\leq ctV\|\hat{A}\|\|\hat{H}_M\|\| \exp\left(-\frac{d-vt}{\xi}\right)\\ &+ c'\|\hat{A}\|\|\hat{H}_R\|^2 t^2.\end{aligned}$$

Because $\hat{H}_R$ and $\hat{H}_M$ represent respectively the Hamiltonian of a ball of radius d and the Hamiltonian for a constant thickness membrane around that ball, they grow proportionally to d^D and d^{D-1} respectively, where D is the spatial dimension, *i.e.*, $\|\hat{H}_R\| \leq \alpha d^D$ and $\|\hat{H}_M\| \leq \alpha d^{D-1}$ for some constant α. Choosing $d \approx vt + \log(cV/c't)$ such that

$$d^{D+1}\exp\left(\frac{d}{\xi}\right) \geq \frac{cV}{c't}\exp\left(\frac{vt}{\xi}\right), \tag{37}$$

we obtain

$$\|\hat{A}(t) - \hat{A}(0) - it[\hat{H}_R, \hat{A}(0)]\| \leq \kappa\|\hat{A}\|\left[vt + \log\left(\frac{cV}{c't}\right)\right]^2 t^2 \tag{38}$$

for some constant $\kappa = 2c'\alpha^2$.

From now on, we assume that t is chosen small enough to neglect the high-order correction and write $\hat{A}(t) = \hat{A}(0) + it[\hat{H}_R, \hat{A}(0)]$. We choose a set of k-local operators $\hat{P}_l$ to parametrize the Hamiltonian $\hat{H}$. The size of this set grows linearly with the number of particles. We can write $\hat{H} = \sum_l h_l\hat{P}_l$, and our goal is to learn the value of the h_l. This is achieved by a series of experiments that measure the expectation value of some local observables $\hat{A}_i(t)$ for different initial product states $\hat{\rho}_j$,

For every pair of observable $\hat{A}_i$ and state $\hat{\rho}_j$, define $W_{ij} = \langle\hat{A}_i(t)\rangle_{\hat{\rho}_j} - \mathrm{tr}\,\hat{A}_i\hat{\rho}_j = it\langle[\hat{H}, \hat{A}_i]\rangle_{\hat{\rho}_j}$. The term $\mathrm{tr}\,\hat{A}_i\hat{\rho}_j$ can be easily computed since $\hat{A}_i$ is local and $\hat{\rho}_j$ is a product state. The term $\langle\hat{A}_i(t)\rangle_{\hat{\rho}_j}$ represents the experimental expectation value of the observable $\hat{A}_i$ at time t for initial state ρ_j. It can be estimated within accuracy ϵ with a number of measurements equal to ϵ^{-2}. Thus, it is possible to efficiently learn the W_{ij} within some constant accuracy and a number of experiments proportional to the number of particle n.

The linearized evolution equation gives $W_{ij} = \sum_l T_{ij,l} h_l$ where $T_{ij,l} = it \operatorname{tr} \hat{\rho}_j [\hat{P}_l, \hat{A}_i]$. This trace formula can be efficiently evaluated because the commutator of two k-local operators is at most $2k$-local, and $\hat{\rho}_j$ is a product state. Note that $T_{ij,l} = T_{ij',l}$ when $\hat{\rho}_j$ and $\hat{\rho}_{j'}$ differ only outside a region of radius k away from the local observable $\hat{A}_i$. In addition, the T become linearly dependent—and thus redundant—when the input states are linearly dependent. Thus, for each observable $\hat{A}_i$, we only need to vary the initial state locally, so the total number of observable-state pairs (ij) grows linearly with the number of particles. Thus, learning the Hamiltonian—or equivalently the h_l—amounts to inverting the linear-size linear equation $W_{ij} = \sum_l T_{ij,l} h_l$. This can be done using different methods. When experimental errors ϵ_{ij} are known for the estimation of each M_{ij}, a least-χ^2 methods can be appropriate. It amounts to

$$\min_{h_l} \sum_{ij} \left(\frac{W_{ij} - \sum_l T_{ij,l} h_l}{\epsilon_{ij}} \right)^2 . \qquad (39)$$

In the case where the experimental error estimates are all the same for all observations, this problem reduces to computing $T^+_{ij,l}$, the Moore-Penrose pseudo inverse of $T_{ij,l}$ [56, 57].

Numerical experiments were performed for local Hamiltonians, and the results are plotted in Fig. 4 and Fig. 5. The systems we considered were small chains of qubits with random nearest neighbour interactions. The system evolution was calculated exactly for a short amount of time, and the linearized problem was inverted using the Moore-Penrose pseudoinverse. Since these Hamiltonians are drawn at random (but with maximum strength for each term independent of the system size), we calculate the average l_2 distance between the estimated Hamiltonian and the actual Hamiltonian (Fig. 4), as well as the quantiles for error propagation scaling factor of each of the elements of h_l, given by $\sqrt{\sum_{ij} |T^+_{ij,l}|^2}$ (Fig. 5). The results clearly indicate well behaved error scaling for these systems, even under finite statistical error in the estimation of observable expectations.

VIII. SUMMARY

We have demonstrated that by treating learning and verification of quantum states and processes as separate problems, and by using a targeted approach to the extraction of information from experiments, the exponential overhead of tomography can be completely avoided for a number of physically interesting cases—including a number of physical models that lead to universal quantum computation. In the cases where the exponential overhead cannot be avoided, our approach leads to significant practical advantages for verification experiments when compared to state and process tomography protocols.

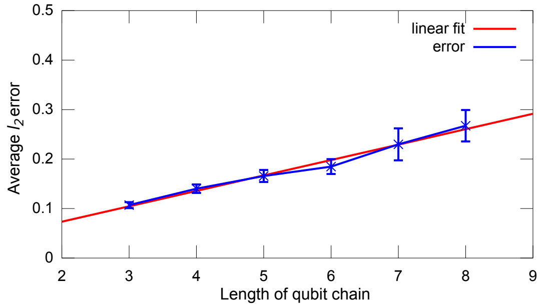

FIG. 4: Average l_2 distance of h_l coefficients of the estimated Hamiltoninan in numerical experiments with Moore-Penrose pseudoinversion of the linearized evolution. The systems consist of linear chains of qubits with randomly chosen 2-local Hamiltonians—each coefficient has strength uniformly distributed between 0.8 and 1.2, with random signs. It is evolved exactly for $t = 10^{-3}$, and the expectation of randomly chosen observables is measured with precision 10^{-4} (the initial states are chosen to be tensor products of Pauli operator eigenstates). The error scaling is well described by a linear function of the size of the system as it would be expected, since the number of parameters measured is linear in the system size.

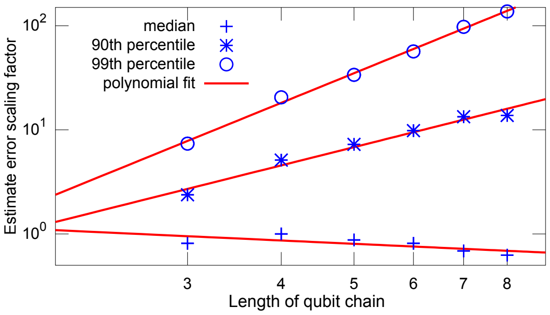

FIG. 5: Error scaling factors for each of the Hamiltonian parameters, using the Moore-Penrose pseudoinversion of the linearized evolution. The statistics of 40 numerical experiments for each chain length indicate that the error scaling for each of the Hamiltonian parameters is exponentially concentrated at small values.

IX. ACKNOWLEDGEMENTS

We would like to thank P. Sémon for many instructive discussions about Monte Carlo methods.

[1] T. Yamamoto, Y. A. Pashkin, O. Astafiev, Y. Nakamura, and J. S. Tsai, Nature **425**, 941 (2003).

[2] J. Chiaverini, D. Leibfried, T. Schaetz, M. D. Barrett, R. B. Blakestad, J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, et al., Nature **432**, 602 (2004).

[3] H. Haffner, W. Hansel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Korber, U. D. Rapol, M. Riebe, P. O. Schmidt, et al., Nature **438**, 643 (2005).

[4] P. Walther, K. J. Resch, T. Rudolph, E. Schenck, H. Weinfurter, V. Vedral, M. Aspelmeyer, and A. Zeilinger, Nature **434**, 169

(2005).

[5] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, et al., Nature **438**, 639 (2005).

[6] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **106**, 130506 (2011).

[7] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).

[8] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).

[9] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).

[10] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).

[11] P. Aliferis, D. Gottesman, and J. Preskill, Quant. Inf. Comput. **6**, 97 (2006).

[12] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nature Comm. **1**, 149 (2010).

[13] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46 (2001).

[14] F. Verstraete and J. I. Cirac, Phys. Rev. B **73**, 094423 (2006).

[15] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).

[16] L. G. Valiant, SIAM J. Comput. **31**, 1229 (2002).

[17] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987).

[18] F. Verstraete and J. Cirac, Arxiv preprint cond-mat/0407066 (2004).

[19] M. V. den Nest (2009), arxiv:0911.1624.

[20] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).

[21] R. Raussendorf and H. J. Briegel, Phys. Rev. Lett. **86**, 5188 (2001).

[22] D. Gottesman, Ph.D. thesis, California Institute of Technology (1997).

[23] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Quantum Inf. Comput. **7**, 401 (2007).

[24] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[25] M.-D. Choi, Lin. Alg. Appl. **10**, 285 (1975).

[26] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[27] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[28] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002), ISSN 0375-9601.

[29] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B: Quantum Semiclass. Opt. **7**, S347 (2005).

[30] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Science **317**, 1893 (2007).

[31] O. Moussa, M. P. da Silva, and R. Laflamme (2011), unpublished.

[32] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).

[33] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).

[34] A. Y. Kitaev, Russ. Math. Surv. **52**, 1191 (1997).

[35] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).

[36] D. Janzig, P. Wocjan, and T. Beth, Int. J. Quant. Info. **3**, 463 (2003).

[37] B. Rosgen and J. Watrous, in *Proc. of the 20th Conference on Computational Complexity* (2004).

[38] B. Rosgen (2009), arXiv:0910.3740.

[39] E. M. R. Blume-Kohout and J. Emerson (2009), arXiv:0910.1315v3.

[40] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).

[41] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. A **80**, 032116 (2009).

[42] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä, Phys. Rev. A **54**, 139 (1996).

[43] P. W. Shor, SIAM Review **41**, 303 (1999).

[44] N. Yoran and A. J. Short, Phys. Rev. A **76**, 042321 (2007).

[45] I. L. Markov and Y. Shi, SIAM J. Comput. **38**, 963 (2008).

[46] Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, S. Lloyd, and D. G. Cory, J. Chem. Phys. **121**, 6117 (2004).

[47] D. Gottesman and S. Aaronson (2008), unpublished.

[48] C. W. Helstrom, Journal of Statistical Physics **1**, 231 (1969), ISSN 0022-4715, 10.1007/BF01007479, URL `http://dx.doi.org/10.1007/BF01007479`.

[49] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser (2000), arXiv:quant-ph/0001106v1.

[50] F. Verstrate, M. M. Wolf, and J. I. Cirac, Nature Physics **5**, 633 (2009).

[51] J. Kempe, A. Kitaev, and O. Regev, SIAM J. Comput. **35**, 1070 (2004).

[52] R. Oliveira and B. M. Terhal, Quant. Inf. Comput. **8**, 900 (2008).

[53] E. H. Lieb and D. W. Robinson, Commun. Math. Phys. **28**, 251 (1972).

[54] M. B. Hastings, Phys. Rev. Lett. **93**, 140402 (2004).

[55] D. Poulin, Phys. Rev. Lett. **104**, 190401 (2010).

[56] E. H. Moore, B. Am. Math. Soc. **26** (1920).

[57] R. Penrose, P. Camb. Philol. Soc. **51**, 406 (1955).