Estimating expectation values using approximate quantum states

Marco Paini[1], Amir Kalev[2], Dan Padilha[1], and Brendan Ruck[3]

[1]Rigetti Computing, 138 Holborn, London, EC1N 2SW, UK.

[2]Information Sciences Institute, University of Southern California, Arlington, VA 22203, USA.

[3]Rigetti Computing, 2919 Seventh St, Berkeley, CA 94710, USA.

We introduce an approximate description of an N–qubit state, which contains sufficient information to estimate the expectation value of any observable to a precision that is upper bounded by the ratio of a suitably–defined seminorm of the observable to the square root of the number of the system's identical preparations M, with no explicit dependence on N. We describe an operational procedure for constructing the approximate description of the state that requires, besides the quantum state preparation, only single–qubit rotations followed by single–qubit measurements. We show that following this procedure, the cardinality of the resulting description of the state grows as $3MN$. We test the proposed method on Rigetti's quantum processor unit with 12, 16 and 25 qubits for random states and random observables, and find an excellent agreement with the theory, despite experimental errors.

1 Introduction

Quantum density operators represent our knowledge of the state of quantum systems and give us a way to calculate expectation values and predict experimental results via the Born rule. Our knowledge of a system's density operator is equivalent to our ability to calculate the expectation value of any observable of the system. However, for an N–body system, reconstructing the system's density operator requires the knowledge of exponentially many numbers in N. This in turn translates into practical difficulties in estimating

Marco Paini: mpaini@rigetti.com

Amir Kalev: amirk@isi.edu

and storing density operators, in the form of density matrices, even for systems composed of just a few tens of qubits.

Over the last few years, there has been a surge of interest in devising protocols to reduce the resources required to estimate and classically store quantum states [1–9]. There have been numerous approaches to reduce the cost of performing quantum state tomography by assuming that the state of the system has a specific structure. These approaches include, for example: matrix product states tomography [3], where the state is assumed to have an efficient matrix product state representation as a matrix; neural network tomography [6, 7], where the state is assumed to have an efficient deep neural network representation; and compressed sensing tomography [4, 5], where the state is assumed to have a low-rank density matrix representation. The theory of Probably Approximately Correct (PAC) learning was utilised in [2] to show that if, instead of full tomography, we are interested in approximating with high probability the expectation values of observables drawn from a distribution, then only polynomially–many samples in N of the observables' expectation values are required. It was later also shown [8] that estimating with high probability the expectation values of a set of observables can be achieved using only polynomially–many copies in N of the system. As for storage, it is well established [1] that quantum states that can be represented as low bond–dimension matrix product states can be efficiently stored on classical computers. Recently it was shown [9] that a general quantum state of N qubits can be approximately described using its inner product with an order of $\sqrt{2^N}$ stabiliser states. This result thus provides a polynomial improvement for classically storing quantum states over a brute–force approach of storing

density matrices. In Sec. 4 we discuss these and other works in connection to the results presented in this paper.

The growth of the number of copies of the system in N for estimation and the considerable requirements in N for classically storing the information associated with a generic quantum state lead to the question of whether a protocol that significantly reduces or eliminates the dependence on N is admissible. In this work, we give an affirmative answer to the question. Instead of focusing on the estimation and storage of the density matrix, we start by considering the equivalent problem of calculating the expectation value of any observable of the system up to a certain precision. In this article we show that, for a system of N qubits, the expectation value of any observable can be calculated with arbitrary precision with a number of copies of the system independent of N, but dependent on a properly–defined seminorm of the observable. The information obtained through the measurements on the copies of the system will be used to form an *approximate description of the quantum state* – a description with a cardinality that grows linearly in N and thus can be constructed and stored for large systems. The method for estimating expectation values, and the exact definition of an observable seminorm and of the storage approximate quantum state will be given in the next section.

2 Expectation value estimation and the approximate quantum state

We will consider specifically a system of N qubits. However, the considerations that will be made are applicable in general to other systems, with finite and infinite–dimensional Hilbert spaces, since they are based on the general theory of irreducible representations of groups which provide explicit forms for a basis in the space of linear bounded operators defined on a Hilbert space. We refer the reader to Refs. [10, 11] where these considerations have been developed.

We start by presenting the method to estimate the expectation value of any observable and notion of the approximate state for a single–qubit system. We then generalise these ideas for the more general N–qubit case.

2.1 One qubit

A single–qubit density operator can be written as [10, 11]

$$\rho = \int_\Sigma \frac{d\vec{n}}{4\pi} \sum_{m=\pm 1} p(m, \vec{n}) K_1(m, \vec{n}), \qquad (1)$$

with the unitary spherical surface Σ as integration domain, $\vec{n} = (\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta)$, with $\vartheta \in [0, \pi], \varphi \in [0, 2\pi)$, is a unit vector in $\mathbb{R}^3$, $p(m, \vec{n})$ is the probability of obtaining the result $m \in \{-1, +1\}$ when measuring the qubit in the eigenbasis of $\vec{\sigma} \cdot \vec{n}$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a vector of the Pauli operators, and $K_1(m, \vec{n})$ is the 1–qubit kernel operator given by

$$K_1(m, \vec{n}) = \frac{1}{2}(\mathbb{1} + 3\, m\, \vec{\sigma} \cdot \vec{n}), \qquad (2)$$

see App. A for more details.

Let us define a single–qubit (unbiased) *estimator function*

$$R_1[\,\cdot\,](m, \vec{n}) = \text{Tr}\Big[\,\cdot\, K_1(m, \vec{n})\Big], \qquad (3)$$

where the $\cdot$ represents the argument of the function R_1, i.e. a single–qubit observable (for a lighter notation we sometime omit the explicit dependence of the estimator on $(m, \vec{n})$). In particular, the expectation value of the Pauli operator σ_i, with $i = 0, 1, 2, 3$ and $\sigma_0 := \mathbb{1}$, $\sigma_1 := \sigma_x$, $\sigma_2 := \sigma_y$, and $\sigma_3 := \sigma_z$, can be written in terms of $R_1[\sigma_i]$ as

$$\langle \sigma_i \rangle = \int_\Sigma \frac{d\vec{n}}{4\pi} \sum_{m=\pm 1} p(m, \vec{n}) R_1[\sigma_i](m, \vec{n}). \qquad (4)$$

Calculating the estimator of the Pauli operators we arrive at

$$R_1[\mathbb{1}](m, \vec{n}) = 1, \qquad (5)$$
$$R_1[\sigma_\alpha](m, \vec{n}) = 3\, m\, n_\alpha, \text{ for } \alpha = x, y, z. \qquad (6)$$

The estimator $R_1[\sigma_\alpha](m, \vec{n})$ is a random variable due to its dependence on the stochastic measurement outcome m and the random unit vector $\vec{n}$. Hereafter, unless noted otherwise, we use the subscript α to index x, y, z and subscript i to index $0, 1, 2, 3$. In App. A we show that the variance of the estimator $R_1[\sigma_\alpha]$ is upper bounded by 3,

$$\text{Var}(R_1[\sigma_\alpha](m, \vec{n})) \leqslant 3. \qquad (7)$$

Since $R_1[\mathbb{1}](m, \vec{n})$ is independent of m and $\vec{n}$, it is deterministic and thus has a null variance.

Equation (4) implies that we can construct an *estimate* of $\langle \sigma_i \rangle$ using the following Monte Carlo sampling procedure:

1. Sample a point $\vec{n}$ uniformly from the surface of a unit sphere,

2. prepare the qubit in the state ρ and measure it in the eigenbasis of $\vec{\sigma} \cdot \vec{n}$,

3. record the outcome $m \in \{-1, 1\}$,

4. repeat steps 1–3 M times.

The estimate of $\langle \sigma_i \rangle$, which we denote by $\langle \sigma_i \rangle_M$, is simply given by the average of corresponding estimator $R_1[\sigma_i](m, \vec{n})$ over the data set:

$$\langle \sigma_i \rangle_M = \frac{1}{M} \sum_{j=1}^{M} R_1[\sigma_i](m_j, \vec{n}_j), \qquad (8)$$

where $\vec{n}_j = (\cos \varphi_j \sin \vartheta_j, \sin \varphi_j \sin \vartheta_j, \cos \vartheta_j)$ is the unit vector sampled at the jth iteration and m_j is the corresponding measurement outcome. Note that the estimate $\langle \mathbb{1} \rangle_M = 1$ for any value of M and it has zero variance. Moreover, by construction, in the limit $M \to \infty$, $\langle \sigma_\alpha \rangle_M \to \langle \sigma_\alpha \rangle$, for $\alpha = x, y, z$. The value of M determines the precision of the estimation. For M sufficiently large, $\langle \sigma_\alpha \rangle_M$ has a normal statistical error, which due to Eq. (7) is smaller than or equal to $\sqrt{3/M}$.

The measurement in the eigenbasis of $\vec{\sigma} \cdot \vec{n}$, at step 2 above, can be implemented by first applying a unitary operator U, that rotates this basis to the eigenbasis of σ_z, and then measuring in the eigenbasis σ_z (the computational basis). This unitary operator is given by $U = e^{i \frac{\vartheta}{2} \vec{\sigma} \cdot \vec{n}_\perp}$, where $\vec{n}_\perp = (-\sin \varphi, \cos \varphi, 0)$. We call the vector $s_j = (m_j, \vartheta_j, \varphi_j)$ the j-th *snapshot* of the qubit system. We refer to the collection of the M snapshots, $\mathcal{S} = \{s_j\}_{j=1}^{M}$, that we obtain following the sampling procedure above, as the *approximate quantum state*, as it enables us to estimate the expectation value of any single–qubit observable to a certain precision as we now describe.

Let O be a single–qubit observable, expanded in the Pauli basis $\{\sigma_i\}_{i=0}^{3}$,

$$O = \sum_{i=0}^{3} a_i \, \sigma_i. \qquad (9)$$

Given the estimated expectation values of $\{\sigma_i\}_{i=0}^{3}$, by linearity of the estimator function (3), we can construct an estimate of $\langle O \rangle$ as

$$\langle O \rangle_M = \sum_{i=0}^{3} a_i \, \langle \sigma_i \rangle_M. \qquad (10)$$

We prove in App. A that for M sufficiently large, $\langle O \rangle_M$ has a normal statistical error smaller than or equal to $\|O\|/\sqrt{M}$, that is

$$\mathrm{Std}(\langle O \rangle_M) \leqslant \frac{\|O\|}{\sqrt{M}}, \qquad (11)$$

where Std stands for 'standard deviation' and

$$\|O\| = \sqrt{3 \sum_{i=1}^{3} a_i^2} \qquad (12)$$

is a seminorm of O that depends on its expansion in the Pauli operator basis.

2.2 N qubits

The method developed for one qubit in the previous subsection is easily generalised to N qubits, in the assumptions that the qubits are distinguishable and that a qubit–specific measurement capability is available. The space of the N–qubit system is the tensor product of the single–qubit spaces and thus we can write a general N–qubit density operator as [10, 11]

$$\rho = \left(\frac{1}{(4\pi)^N} \prod_{k=1}^{N} \int_{\Sigma_k} d\vec{n}_k \sum_{m_k} \right) p(\{m, \vec{n}\}) K(\{m, \vec{n}\}), \qquad (13)$$

where $\{m, \vec{n}\}$ is a shorthand for $(m_1, \vec{n}_1, \ldots, m_N, \vec{n}_N)$, the N–qubit kernel operator $K(\{m, \vec{n}\})$ equals to the tensor product of single–qubit kernels

$$K(\{m, \vec{n}\}) = \bigotimes_{k=1}^{N} K_1(m_k, \vec{n}_k), \qquad (14)$$

and $p(\{m, \vec{n}\})$ represents the joint probability of finding the outcome $(m_1, \ldots, m_N) \in \{-1, 1\}^N$ when measuring the qubits in the product eigenbasis of $\vec{\sigma}_1 \cdot \vec{n}_1 \otimes \cdots \otimes \vec{\sigma}_N \cdot \vec{n}_N$. Similarly to the single qubit case, this measurement can be implemented with single–qubit rotations followed by a measurement in the computational basis on each qubit.

We define the N–qubit (unbiased) estimator function

$$R[\,\cdot\,](\{m, \vec{n}\}) = \mathrm{Tr}\Big[\,\cdot\, K(\{m, \vec{n}\})\Big]. \qquad (15)$$

As a result of the tensor product structure of the kernel operator, $R[P_i]$, where $P_i = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_N}$,

$i = (i_1, \dots i_N)$, is a monomial of Pauli operators (a 'Pauli monomial' or a 'Pauli string' for short), and hereafter $i_k = 0, \dots, 3$, is simply the product of the corresponding single–qubit random variables,

$$R[P_i](\{m, \vec{n}\}) = \prod_{k=1}^{N} R_1[\sigma_{i_k}](m_k, \vec{n}_k), \qquad (16)$$

where $R_1[\sigma_{i_k}](m_k, \vec{n}_k)$ is given in Eqs. (5) and (6).

Similarly to the single qubit case of the previous section, we can write the expectation value of P_i with respect to the state of the system in terms of $R[P_i]$ as

$$\langle P_i \rangle = \left(\frac{1}{(4\pi)^N} \prod_{k=1}^{N} \int_{\Sigma_k} d\vec{n}_k \sum_{m_k} \right) \qquad (17)$$
$$\times p(\{m, \vec{n}\}) R[P_i](\{m, \vec{n}\}).$$

Eq. (17) implies that an estimate $\langle P_i \rangle_M$ of $\langle P_i \rangle$ can be calculated by performing the single–qubit tomographic procedure introduced in the above section, in parallel on all of the N qubits, as in Fig. 1. We call the vector $s_j = (m_{1j}, \vec{n}_{1j}, \dots, m_{Nj}, \vec{n}_{Nj})$ the jth *snapshot* of a system of N qubits, and we refer to a set $\mathcal{S}$ of such M independent snapshots as the *approximate state* of the N–qubit system.

The use of single–qubit Haar–random basis measurements was recently proposed, and experimentally implemented, to probe certain properties of an N–qubit system, such as purity [12, 13] and the Frobenius inner product of two states [14]. In light of these results, as we now show, the measurement protocol we introduced above is a universal protocol for estimating the expectation value of any N–qubit observable. The protocol is efficient in terms of time and space resources, as long as the observable admits an efficient representation in the Pauli basis and has a small seminorm (e.g. not growing exponentially with N).

To start with, we can use the approximate state $\mathcal{S}$ to calculate an estimate $\langle P_i \rangle_M$ of $\langle P_i \rangle$ as

$$\langle P_i \rangle_M = \frac{1}{M} \sum_{j=1}^{M} R[P_i](\{m, \vec{n}\})$$
$$= \frac{1}{M} \sum_{j=1}^{M} \prod_{k=1}^{N} R_1[\sigma_{i_k}](m_{kj}, \vec{n}_{kj}), \qquad (18)$$

where the first equality follows from Eq. (16). We show in App. B that for M sufficiently large,

$\langle P_i \rangle_M$ $(i \neq 0_v \doteq (0, \dots, 0))$ has a normal statistical error that is smaller than or equal to $\sqrt{3^{r_i}/M}$, where r_i represents the weight of P_i, i.e., the number of qubits on which it acts non–trivially. Note that the estimated expectation value for the N–qubit identity operator $\langle \mathbb{1}^{\otimes N} \rangle_M = 1$ and it has zero variance.

With the ability to efficiently calculate the estimated expectation value of any given Pauli monomial operator, we can estimate the expectation value of N–qubit observables. Let O be an N–qubit observable such that

$$O = \sum_i a_i P_i, \qquad (19)$$

From linearity of the estimator function (15) and from Eq. (18), the estimated expectation value of O is given by

$$\langle O \rangle_M = \sum_i a_i \langle P_i \rangle_M$$
$$= \frac{1}{M} \sum_{j=1}^{M} \sum_i a_i \prod_{k=1}^{N} R_1[\sigma_{i_k}](m_{kj}, \vec{n}_{kj}). \qquad (20)$$

Similar to $\langle P_i \rangle_M$, $\langle O \rangle_M$ is an unbiased estimate, i.e., $\langle O \rangle_M \to \langle O \rangle$ in the limit $M \to \infty$. Practically, for the efficient calculation of (20), the sum on i should have at most a polynomial number of terms in N.

While we assumed in Eq. (19) that the observable admits an efficient representation in terms of Pauli monomials, as a result of the tensor product structure of the kernel operator, Eq. (14), our procedure is efficient to calculate the expectation value of any operator of the form

$$O = \sum_i a_i O_{i_1} \otimes O_{i_2} \cdots \otimes O_{i_N}, \qquad (21)$$

where O_{i_k} is a single-qubit operator (on the kth qubit space). If this is the case then,

$$\langle O \rangle_M = \frac{1}{M} \sum_i a_i \sum_{j=1}^{M} \prod_{k=1}^{N} R_1[O_{i_k}]. \qquad (22)$$

Note that any Pauli monomial P_i is a special case of $O_{i_1} \otimes O_{i_2} \cdots \otimes O_{i_k}$.

We show in App. B that the standard deviation of $\langle O \rangle_M$ has the upper bound

$$\mathrm{Std}(\langle O \rangle_M) \leqslant \frac{\|O\|}{\sqrt{M}}, \qquad (23)$$

with the seminorm of O defined as

$$\|O\| = \sqrt{\sum_{i,j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} |a_i||a_j|}, \qquad (24)$$

where the sum is extended to all values of i and j except $0_v \doteq (0, 0, \ldots, 0)$, r_{ij} is the number of indices k for which $i_k \neq 0 \wedge j_k \neq 0$, $\Delta_{ij} = 0$ if there exists a k such that $i_k \neq 0 \wedge j_k \neq 0 \wedge i_k \neq j_k$ and $\Delta_{ij} = 1$ otherwise. Eq. (24) reduces to Eq. (12) when O is an observable on a single–qubit space.

Importantly, in this procedure, the expectation values estimation error has no global dependence on the system size N. The error only depends on the observable seminorm and the number of snapshots. One could argue that the presence in the seminorm of $3^{r_{ij}}$ reflects an exponential growth of the estimates' variance in N, as, for example, for an observable given by the tensor product of N Pauli operators, which has seminorm $3^{N/2}$. However, the seminorm is a property of the observable and not of the dimension of the space: an N–qubit observable O and its extension to a larger $(N+L)$–qubit space, $O \otimes \mathbb{1}^{\otimes L}$, have the same seminorm (24), independent of L. Irrespective of N, because of Eq. (23), the expectation values of all observables with unit seminorm can be obtained with statistical error bounded by $1/\sqrt{M}$. Any observable with seminorm different from zero can be obtained from an observable with unit seminorm through multiplication by a real number, representing in fact the observable's seminorm. For a given number of identical preparations and measurements, the statistical error in the expectation value of the generic observable has a bound only dependent on how "large" the observable is. In App. B, we also explain how the seminorm $\|O\|_2$, defined by

$$\|O\|_2 = \sqrt{\sum_{i \neq 0_v} 3^{r_i} a_i^2}, \qquad (25)$$

divided by $\sqrt{M}$, with r_i representing the weight of P_i, is generally a good approximation for the standard deviation of $\langle O \rangle_M$. This result may be important for applications of the method since $\|O\|_2$ can be significantly smaller than $\|O\|$ and since, for a chosen statistical error, the number of identical preparations required is inversely proportional to the variance of $\langle O \rangle_M$. See App. C for further details.

(a)

(b)
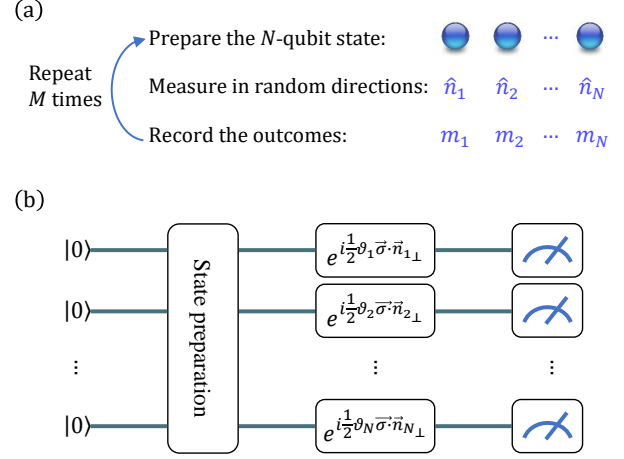
Figure 1: **Obtaining the approximate quantum state.** (a) To construct the approximate description of an N–qubit state we repeat the following steps M times: (1) prepare a copy of the N–qubit state, (2) measure each qubit in a random direction $\vec{n}_j = (\cos \varphi_j \sin \vartheta_j, \sin \varphi_j \sin \vartheta_j, \cos \vartheta_j)$, (3) record the measurement outcome of each qubit. (b) A circuit realization of the protocol. The single–qubit rotations are defined by ϑ_j and $\vec{n}_{j\perp} = (-\sin \varphi_j, \cos \varphi_j, 0)$, $j = 1, \ldots, N$.

We close this part with a remark on the notion of the approximate quantum state, $\mathcal{S}$. We have shown that the approximate description of the state, as we have defined it, comes with an explicit operational procedure to estimate the expectation value of any generic observable with a rigorous bound on the estimation statistical error. Therefore, the approximate state together with the rule (20) provides a description of a physical system in terms of statistical inferences based on observed experimental data and, in principle, could be assumed as primitive concepts instead, or, at least, as not requiring the existence of density operators and the Born rule.

3   Experimental results

As a proof–of–concept, we have tested the theory on Rigetti Computing quantum processor unit (QPU), Aspen7-28Q-A, with three systems, involving $N = 12, 16$ and 25 qubits, respectively. In each experiment the system is prepared in a random state. The state preparation circuit is composed of two layers of gates applied to the computational state $|0\rangle^{\otimes N}$. In the first layer, the gates are chosen at random from a flat distribution over the single–qubit gate set $\{X, Y, Z, H, S, T\}$

and applied to $N/2$ qubits, chosen at random. While in the second layer the two–qubit gate $XY(\alpha) = e^{-i\alpha(X\otimes X + Y\otimes Y)}$ is applied to 50% of pairs of qubits, chosen at random, with different values of α sampled uniformly from $[0, 2\pi)$. This gate composition results in shallow quantum circuits with $N/2$ random single–qubit gates and $N/4$ random two–qubit gates, to ensure that the experimentally prepared state has a high fidelity with the target state.

For each system, an approximate description of the prepared state was obtained using $M = 10^4$ snapshots (to check the results to 0.01 accuracy). The approximate state was then used to estimate the expectation value of two sets of observables: (1) 20 random observables, (2) 20 random projectors onto computational basis states (we have selected 20 for the number of states so that it would be visually 'easy' to check the results in Fig. 2 against the second standard deviation given it is at ∼95%). Each random observable in the first set is a linear combination of 20 Pauli monomials. Each Pauli monomial is constructed by randomly sampling from the set $\{\sigma_i\}_{i=0}^{3}$ for each qubit, whereas the linear combination coefficients (the a_i's) are sampled uniformly from the $(0, 1]$. The projectors onto computational basis states were sampled uniformly from the set $\{0, 1\}^N$. For simplicity in the analysis, the 20 random observables are normalised to have a unit seminorm (the projectors already have unit seminorm with probability exponentially close to 1 as a function of N, see App. C).

Fig. 2 shows the results for the 25–qubit experiment. In this figure, we compare the estimated expectation values of random observables and of random computational basis projectors to their expectation values with respect to the 25–qubit target state. The results are plotted as a function of the number of snapshots. Table 1 summarises the experimental results of the 12–, 16– and 25–qubit systems for 10^4 snapshots. The figure and the table show that the experimental results generally mirror the theory, with the fraction of estimates within one and two standard deviations from their true values not far from 68% and 95% respectively. Other than statistical considerations, the cases with a fraction of the estimates greater than the expected percentages from a normal distribution can be explained in terms of how tightly the observables' seminorms bound

the standard deviation of the estimators, in particular for the case of the projectors.

(a) Random observables

Number of qubits	12	16	25
Estimation within 1 Std	60%	65%	70%
Estimation within 2 Std's	93%	100%	95%

(b) Random projectors

Number of qubits	12	16	25
Estimation within 1 Std	70%	90%	85%
Estimation within 2 Std's	95%	95%	100%

Table 1: The fraction of the estimated expectation values, of (a) random observables and (b) random projectors onto computational basis states, that are within one and two standard deviations from their true expectation values for 10^4 snapshots.

To complete the picture, including an explanation for the cases of a lower fraction of the estimates than predicted, we need to start examining the effects of hardware errors. It is worth noting that we obtain agreement with the theory despite an order of $p_{\text{err}} = 5\%$ (averaged) readout error per qubit, which amounts to flipping the measurement outcome from $+1$ to -1 and vice versa. The resilience of our protocol to single–qubit readout errors can be attributed to the randomness of the measurement bases, as was discussed in [15]. Moreover, the following considerations provide additional insights in the robustness of the estimation protocol to readout errors.

Firstly, let us consider the estimation of the expectation value of single–qubit observable. Let $\mathcal{S}$ be the approximate state description composed of M snapshots. Suppose that in M_{err} of the snapshots we obtained the erroneous measurement outcome, $M_{\text{err}}/M \approx p_{\text{err}}$. The estimation of the expectation value of any single–qubit Pauli operator σ_α $(\alpha = x, y, z)$ is

$$
\begin{aligned}
\langle \sigma_\alpha \rangle_M &= \frac{1}{M} \sum_{j=1}^{M} 3 m_j n_{\alpha j} \\
&= \frac{M - M_{\text{err}}}{M} \Big(\frac{1}{M - M_{\text{err}}} \sum_{j=1}^{M - M_{\text{err}}} 3 m_j n_{\alpha j} \Big) \\
&\quad + \frac{M_{\text{err}}}{M} \Big(\frac{1}{M_{\text{err}}} \sum_{j=1}^{M_{\text{err}}} 3 m_j^{\text{err}} n_{\alpha j} \Big),
\end{aligned} \tag{26}
$$

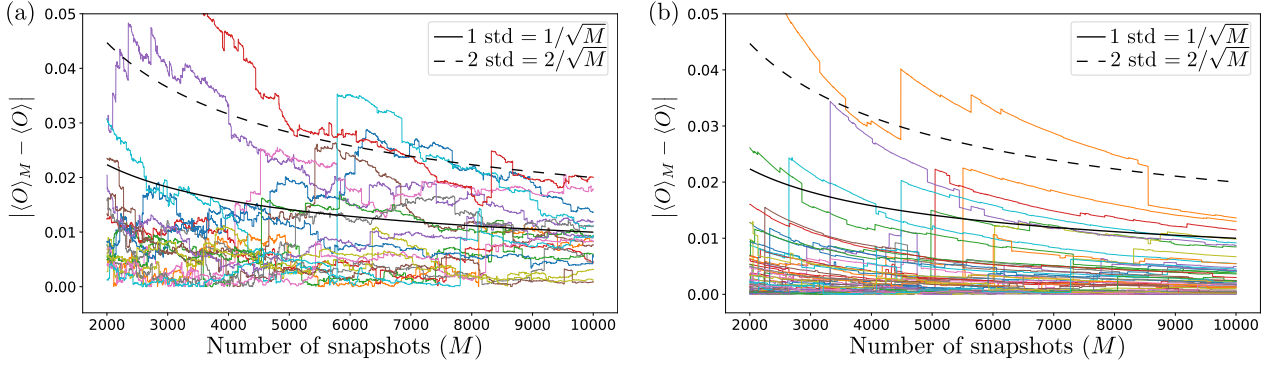where m_j^{err} is the erroneous outcome. Since

Figure 2: The error in the estimated expectation value of (a) 20 random observable and (b) 20 random projectors onto computational basis states in a 25–qubit system as a function of the number of snapshots.

$m_j^{\mathrm{err}} = -m_j$ we arrive at

$$\langle \sigma_\alpha \rangle_M = \frac{M - M_{\mathrm{err}}}{M}\Big(\frac{1}{M - M_{\mathrm{err}}}\sum_{j=1}^{M-M_{\mathrm{err}}} 3m_j n_{\alpha j}\Big)$$
$$- \frac{M_{\mathrm{err}}}{M}\Big(\frac{1}{M_{\mathrm{err}}}\sum_{j=1}^{M_{\mathrm{err}}} 3m_j n_{\alpha j}\Big), \qquad (27)$$

which, for sufficiently large M and M_{err}, can be approximated by

$$\langle \sigma_\alpha \rangle_M \approx (1 - p_{\mathrm{err}})\langle \sigma_\alpha \rangle - p_{\mathrm{err}}\langle \sigma_\alpha \rangle$$
$$= (1 - 2p_{\mathrm{err}})\langle \sigma_\alpha \rangle. \qquad (28)$$

Therefore, for a single–qubit observable $O = \sum_i a_i \sigma_i$ we arrive at

$$\langle O \rangle_M \approx a_0 \mathbb{1} + \sum_{i=1}^{3} a_i (1 - 2p_{\mathrm{err}})\langle \sigma_i \rangle. \qquad (29)$$

Thus, the readout error introduces an attenuator $(1 - 2p_{\mathrm{err}})$ to each coefficient different than a_0 in the expectation value.

Moving to the N–qubit case, we first note that only odd numbers of readout errors per snapshot contribute to the estimation error of an N–qubit expectation value. This is because the estimated expectation value is calculated via a product of the measurement outcomes which take values ± 1, c.f. Eq. (18). Therefore, if a given snapshot includes readout errors on an even number of qubits, there are even number of erroneous outcomes m_j^{err} such that $m_j^{\mathrm{err}} = -m_j$, and the estimated expectation value remains unaffected by these errors. Let $p_{\mathrm{odd}}(r)$ be the probability that in a given snapshot there is an odd number of readout errors on r qubits. Given a readout error

p_{err} per qubit,

$$p_{\mathrm{odd}}(r) = \sum_{k\ \mathrm{odd}} \binom{r}{k} p_{\mathrm{err}}^k (1 - p_{\mathrm{err}})^{r-k}$$
$$= \frac{1}{2}\Big(1 - (1 - 2p_{\mathrm{err}})^r\Big). \qquad (30)$$

Similarly to the single qubit analysis, one can then show that the expectation value of a Pauli monomial P_i with weight r_i is

$$\langle P_i \rangle_M \approx (1 - 2p_{\mathrm{odd}}(r_i))\langle P_i \rangle = (1 - 2p_{\mathrm{err}})^{r_i}\langle P_i \rangle. \qquad (31)$$

We therefore obtain that the estimated expectation value of an N–qubit observable $O = \sum_i a_i P_i$ is given by

$$\langle O \rangle_M \approx \sum_i a_i (1 - 2p_{\mathrm{err}})^{r_i}\langle P_i \rangle. \qquad (32)$$

This equation indicates that the readout error does not depend on N. It introduces an attenuator $(1 - 2p_{\mathrm{err}})^{r_i}$ to each term in the expectation value. Since the attenuator is an exponentially decreasing function in the weight r_i, only low–weight Pauli monomials contribute to the expectation value $\langle O \rangle_M$. Thus, the proposed method is robust against readout error when estimating expectation values that have contributions dominantly from low–weight Pauli monomials (this does not necessarily exclude observables, such as projectors, that have contributions from all Pauli monomials: the contribution of higher weight Pauli monomials could be small on a class of states of interest, as in the case of the low entanglement states of this section's random projectors experiment). Finally, we remark that other errors, e.g. gate operation errors, could be included in this framework in a similar fashion and that, clearly, (32) is easily generalisable to the case of different readout errors per qubit.

4 Related work

Quantum state tomography. In general, estimating an N–qubit state to error ϵ in trace distance requires $\mathcal{O}(2^N/\epsilon^2)$ copies of the state [16, 17]. The estimation of the density matrix with our tomographic procedure does not break this lower bound. If we choose the basis of N–qubit Pauli operators, Eq. (23) tells us that a few matrix elements will require M to be exponentially large in N. A different basis, such as the standard basis $|i\rangle\langle j|$, has exponentially many matrix elements bounded by 1. However, calculating the trace distance requires adding exponentially many of them, creating again the need for an exponentially large M. The key point of this paper, however, is to be able to estimate approximately and efficiently expectation values without resorting to the notion of density matrix.

Shadow tomography. In shadow tomography the task is to estimate with high probability the expectation values of a set of K observables, to a certain precision ϵ. It was shown that [8, 18] this can be done using $M = \mathcal{O}(\log(K)^2 N^2 \epsilon^{-8})$ copies of the state, by implementing a non–local measurement on all copies of the state. Recently, these results were improved [19]. The authors of [19] proposed an algorithm where $M = \mathcal{O}(B\log(K)/\epsilon^2)$ copies of the state, with $B \geqslant \max_i \|O_i\|_{\mathrm{shadow}}^2$, allow one to predict the expectation values of K observables, $\{O_i\}_{i=1}^K$, to a precision ϵ. The norm $\|\cdot\|_{\mathrm{shadow}}$ depends on the particular measurement procedure used in the algorithm. For example, for random Pauli measurements the authors show (Proposition 3 there) that for a k-local observable $\|O\|_{\mathrm{shadow}} \leqslant 2^k \|O\|_\infty$, where $\|\cdot\|_\infty$ denotes the spectral norm.

The algorithm of [19] has a similar structure to the algorithm proposed in this work. Similarly to [19], the choice of randomised single–qubit measurement directions of Eq. (13) is not the only possible one. The framework of [10, 11] generates valid formulas for any irreducible projective representation in the space of the system, including for example the Clifford group and the representation of $SU(2)$ in the space of a spin $s = (2^N - 1)/2$. Furthermore, one could also easily derive a general expression for the norm, bound of the standard deviation of the estimator, as $\sqrt{\mathrm{Tr}[O^2]}$ multiplied by a factor only dependent on the group representation chosen. It is important to note, however, that the use of the Cauchy-Schwarz inequality in the derivation would generate suboptimal norms, often exponentially larger than the true standard deviation of the estimator. A tight norm (or seminorm) is key, as it gives a guarantee of performance with a lower number of preparations, especially if the separation from a suboptimal one is exponential. The closeness to the standard deviation of the estimator is what justifies the importance of $\|O\|_2$ of (25): $\|O\|_2$ can be exponentially smaller than norms depending exponentially on the locality or on the dimension of the system (App. C for the case of projectors onto the computational basis).

Finally, while the task in shadow tomography is to estimate with high probability the expectation values of a set of K observables, to a certain precision ϵ, we focus here on the estimation with high probability the expectation values of a given, yet arbitrary, observable to a certain precision ϵ. We show that accomplishing this task requires $M = \mathcal{O}(\|O\|^2/\epsilon^2)$ copies of the state. If a certain precision is required for the combined evaluation of expectation values of several observables, one would simply calculate the joint probability of several normal variables of being smaller than a certain value which would lead to a log contribution to the error in the number of observables.

5 Conclusions and outlook

We have introduced a concept of approximate description of a quantum state for a system of N qubits. We have shown that the approximate state comes with an explicit operational procedure to estimate the expectation value of any observable that admits an efficient representation in the Pauli basis with statistical error only dependent on the cardinality of the approximate quantum state and on the observable's seminorm, independent of N and growing at most exponentially with the locality of the observable. The operational procedure to obtain the approximate state, other than the initial state preparations, only requires single–qubit operations and measurement. The natural robustness of random single–qubit operations to experimental noise, as we saw in the experimental results and also discussed in [15], suggests that the proposed protocol could be appropriate for NISQ computers.

We expect the notion of the approximate quantum state to enable improvement over existing

8

variational optimisation methods, as it would not require separate iterations for each set of values of the parameters. Machine learning optimisations, in particular for generative models known as Born Machines [20], may be an interesting case, since the explicit dependence of the expectation values of the projectors onto computational basis states on the variational parameters [21] may be obtained even for a large number of qubits, since, as shown in App. B, the estimators for the projectors onto computational basis state have, with high probability, variance bounded by 1 for all values of N.

The definition of the snapshots, constituting an approximate quantum state, can be generalised, too. The snapshots can in fact be derived from a group different from $SU(2)$. This generalisation has interest that goes beyond the purely mathematical exercise. As we have seen, $SU(2)$ tomography works well for observables expressed as linear combinations of products of Pauli operators, especially if the weights of the Pauli monomials are low. Electronic Hamiltonians can be represented as linear combinations of products of Pauli operators, but at the cost, in general, of introducing Pauli monomials with weights increasing with N. In second quantisation, electronic Hamiltonians contain linear combinations of products of only two or four ladder operators for any N. The bosonic tomographic formulas of [10] can be extended to the fermionic case, choosing the additive group of Grassmann numbers [22], instead of the additive group of complex numbers, as tomographic group and expressing the snapshots as vectors of sampled real numbers descending from Grassmann numbers and of measured presence or absence of electrons in orbitals. The estimators for ladder operators and for their linear combinations can then be calculated directly without the need for the mapping to Pauli operators and without the consequent appearance of weights of Pauli monomials growing with N.

6 Acknowledgments

References

[1] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91:147902, Oct 2003. URL https://doi.org/10.1103/PhysRevLett.91.147902.

[2] S. Aaronson. The learnability of quantum states. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007. URL https://doi.org/10.1098/rspa.2007.0113.

[3] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu. Efficient quantum state tomography. *Nat. Comm.*, 1:149, 2010. URL https://doi.org/10.1038/ncomms1147.

[4] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15):150401, 2010. URL https://doi.org/10.1103/PhysRevLett.105.150401.

[5] A. Kalev, R. Kosut, and I. Deutsch. Quantum tomography protocols with positivity are compressed sensing protocols. *NPJ Quantum Information*, 1:15018, 2015. URL https://doi.org/10.1038/npjqi.2015.18.

[6] X. Gao and L.-M. Duan. Efficient representation of quantum many-body states with deep neural networks. *Nat. Comm.*, 8:662, 2017. URL https://doi.org/10.1038/s41467-017-00705-2.

[7] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo. Neural-network quantum state tomography. *Nat. Phys.*, 18:447, 2017. URL https://doi.org/10.1038/s41567-018-0048-5.

[8] S. Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 325–338, New York, NY, USA, 2018. ACM. ISBN 978-

1-4503-5559-9. URL https://doi.org/10.1145/3188745.3188802.

[9] D. Gosset and J. Smolin. A compressed classical description of quantum states. In Wim van Dam and Laura Mancinska, editors, *14th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2019)*, volume 135 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:9, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-112-2. URL https://doi.org/10.4230/LIPIcs.TQC.2019.8.

[10] M. Paini. Quantum tomography via group theory. arXiv:quant-ph/0002078, 2000. URL https://arxiv.org/pdf/quant-ph/0002078.pdf.

[11] G. M. D'Ariano, L. Maccone, and M. Paini. Spin tomography. *J. Opt. B: Quantum and Semiclassical Optics*, 5(1):77–84, jan 2003. URL https://doi.org/10.1088/1464-4266/5/1/311.

[12] T. Brydges, A. Elben, P. Jurcevic, C. Vermersch, B.and Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos. Probing Rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, 2019. URL https://doi.org/10.1126/science.aau4963.

[13] A. Elben, J. Yu, G. Zhu, M. Hafezi, F. Pollmann, P. Zoller, and B. Vermersch. Many-body topological invariants from randomized measurements in synthetic quantum matter. *Science Advances*, 6(15), 2020. URL https://doi.org/10.1126/sciadv.aaz3666.

[14] A. Elben, B. Vermersch, R. van Bijnen, C. Kokail, T. Brydges, C. Maier, M. K. Joshi, R. Blatt, C. F. Roos, and P. Zoller. Cross-platform verification of intermediate scale quantum devices. *Phys. Rev. Lett.*, 124: 010504, Jan 2020. URL https://doi.org/10.1103/PhysRevLett.124.010504.

[15] B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller. Probing scrambling using statistical correlations between randomized measurements. *Phys. Rev. X*, 9: 021061, Jun 2019. URL https://doi.org/10.1103/PhysRevX.9.021061.

[16] R. O'Donnell and J. Wright. Efficient quantum tomography. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 899–912, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341325. URL https://doi.org/10.1145/2897518.2897544.

[17] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 913–925, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341325. URL https://doi.org/10.1145/2897518.2897585.

[18] S. Aaronson and G. N. Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 322–333, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367059. URL https://doi.org/10.1145/3313276.3316378.

[19] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, Jun 2020. ISSN 1745-2481. URL https://doi.org/10.1038/s41567-020-0932-7.

[20] S. Cheng, J. Chen, and L. Wang. Information perspective to probabilistic modeling: Boltzmann machines versus born machines. *Entropy*, 20(8), 2018. URL https://doi.org/10.3390/e20080583.

[21] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, nov 2019. URL https://doi.org/10.1088/2058-9565/ab4eb5.

[22] K. E. Cahill and R. J. Glauber. Density operators for fermions. *Phys. Rev. A*, 59:1538–1555, Feb 1999. URL https://doi.org/10.1103/PhysRevA.59.1538.

[23] Y. Li. Talk 5 : Averages over the unitary group. In *Selected topics in Mathematical Physics : Quantum Information Theory*, 2013.

[24] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum

world. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, pages 129–140, Wash-

ington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-2780-9. URL https://doi.org/10.1109/CCC.2007.26.

A One qubit

Based on group-theoretical considerations, it was shown in Refs. [10, 11] that the density operator on a Hilbert space of spin-s particle (a $2s + 1$ dimensional space) can be written as

$$\rho = \sum_{m_s=-s}^{s} \int_{\Sigma} \frac{d\vec{n}}{4\pi} p(m_s, \vec{n}) K(m_s, \vec{n}), \tag{33}$$

with the unitary spherical surface Σ as integration domain, $\vec{n} = (\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta)$, with $\vartheta \in [0, \pi], \varphi \in [0, 2\pi)$, $p(m, \vec{n})$ probability of obtaining the result $m_s \in [-s, s]$ when measuring the spin in the eigenbasis of $\vec{s}\cdot\vec{n}$, where $\vec{s} = (s_x, s_y, s_z)$ is spin of the system, and $K(m_s, \vec{n})$ is a kernel function given by

$$K(m_s, \vec{n}) = \frac{2s+1}{\pi} \int_0^{2\pi} d\psi \sin^2 \frac{\psi}{2} e^{i\psi(m_s - \vec{s}\cdot\vec{n})}. \tag{34}$$

Specialising Eq. (34) for a spin-1/2 particle where $m_s = \pm 1/2 \doteq m/2$ and $\vec{s} = (\sigma_x, \sigma_y, \sigma_z)/2 \doteq \vec{\sigma}/2$ we can write

$$K_1(m, \vec{n}) = \frac{2}{\pi} \int_0^{2\pi} d\psi \sin^2 \frac{\psi}{2} e^{i\frac{\psi}{2}(m - \vec{\sigma}\cdot\vec{n})}, \tag{35}$$

where the subscript of K_1 is used to emphasise that is the kernel operator on a single qubit. We calculate (35) utilising the identity $e^{-i\frac{\psi}{2}\vec{\sigma}\cdot\vec{n}} = \mathbb{1}\cos\frac{\psi}{2} - i\,\vec{\sigma}\cdot\vec{n}\,\sin\frac{\psi}{2}$ and directly obtain

$$K_1(m, \vec{n}) = \frac{1}{2}(\mathbb{1} + 3\, m\, \vec{\sigma}\cdot\vec{n}). \tag{36}$$

In the main text we have defined

$$R_1[\sigma_\alpha](m, \vec{n}) \doteq \text{Tr}[\sigma_\alpha K_1(m, \vec{n})] = 3\, m\, n_\alpha, \tag{37}$$

an unbiased estimator of σ_α, with $\alpha = x, y, z$ and n_α defined by $\vec{n} = (n_x, n_y, n_z)$. The variance of $R_1[\sigma_\alpha]$ is

$$\text{Var}(R_1[\sigma_\alpha](m, \vec{n})) = \langle R_1^2[\sigma_\alpha](m, \vec{n})\rangle - \langle R_1[\sigma_\alpha](m, \vec{n})\rangle^2 \leqslant \langle R_1^2[\sigma_\alpha](m, \vec{n})\rangle. \tag{38}$$

We calculate the last term of (38) utilising (37):

$$\langle R_1[\sigma_\alpha](m, \vec{n}) R_1[\sigma_\beta](m, \vec{n})\rangle = \int_{\Sigma} \frac{d\vec{n}}{4\pi} \sum_{m=\pm 1} p(m, \vec{n}) R_1[\sigma_\alpha](m, \vec{n}) R_1[\sigma_\beta](m, \vec{n}) \tag{39}$$

$$= 9 \int_{\Sigma} \frac{d\vec{n}}{4\pi} n_\alpha n_\beta \sum_{m=\pm 1} p(m, \vec{n}) = 9 \int_{\Sigma} \frac{d\vec{n}}{4\pi} n_\alpha n_\beta = 3\, \delta_{\alpha\beta},$$

with $\delta_{\alpha\beta}$ is Kronecker delta. To calculate the variance of $R_1[O](m, \vec{n})$, we expand O in the Pauli

operator basis $O = \sum_{i=0}^{3} a_i \, \sigma_i$, so that

$$\text{Var}(R_1[O](m,\vec{n})) = \text{Var}\left(a_o \, R_1[\mathbb{1}](m,\vec{n}) + \sum_{i=1}^{3} a_i \, R_1[\sigma_i](m,\vec{n})\right) \tag{40}$$

$$= a_0^2 \, \underbrace{\text{Var}(R_1[\mathbb{1}](m,\vec{n}))}_{=0} + \text{Var}\left(\sum_{i=1}^{3} a_i \, R_1[\sigma_i](m,\vec{n})\right)$$

$$+ 2 \, \underbrace{\text{Cov}\left(R_1[\mathbb{1}](m,\vec{n}), \sum_{i=1}^{3} a_i \, R_1[\sigma_i](m,\vec{n})\right)}_{=0}$$

$$= \text{Var}\left(\sum_{i=1}^{3} a_i \, R_1[\sigma_i](m,\vec{n})\right) \leqslant \left\langle \left(\sum_{i=1}^{3} a_i \, R_1[\sigma_i](m,\vec{n})\right)^2 \right\rangle$$

$$= \sum_{i=1}^{3} a_i^2 \, \langle R_1^2[\sigma_i](m,\vec{n})\rangle + \sum_{i \neq j(\neq 0)} a_i \, a_j \langle R_1[\sigma_i](m,\vec{n}) R_1[\sigma_j](m,\vec{n})\rangle$$

$$= 3 \sum_{i=1}^{3} a_i^2 + 3 \sum_{i \neq j(\neq 0)} a_i \, a_j \, \delta_{ij} = 3 \sum_{i=1}^{3} a_i^2.$$

We define a seminorm of O as

$$\|O\| = \sqrt{3 \sum_{i=1}^{3} a_i^2} \tag{41}$$

and obtain

$$\text{Var}(R_1[O](m,\vec{n})) \leqslant \|O\|^2. \tag{42}$$

Therefore, for M sufficiently large, $\langle O \rangle_M$ has a normal statistical error smaller than or equal to $\|O\|/\sqrt{M}$.

B N qubits

For a system of N qubits, we expand the generic observable O as in (19)

$$O = \sum_i a_i \, P_i \tag{43}$$

where $P_i = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_N}$, $i = (i_1, \ldots i_N)$, is a Pauli monomial operator and $i_k = 0, \ldots, 3$, and proceed in the calculation of the variance of $R[O]$ in a similar way to the one qubit case (the dependence of $R[\cdot]$ on $(m_1, \vec{n}_1, \ldots, m_N, \vec{n}_N)$ is not explicitly indicated for a lighter notation):

$$\text{Var}(R[O]) = a_0^2 \, \text{Var}(R[\mathbb{1}]) + \text{Var}\left(\sum_{i \neq 0_v} a_i \, R[P_i]\right) + 2 \, \text{Cov}\left(R[\mathbb{1}], \sum_{i \neq 0_v} a_i \, R[P_i]\right) \tag{44}$$

$$= \text{Var}\left(\sum_{i \neq 0_v} a_i \, R[P_i]\right) \leqslant \left\langle \left(\sum_{i \neq 0_v} a_i \, R[P_i]\right)^2 \right\rangle = \sum_{i,j(\neq 0_v)} a_i \, a_j \langle R[P_i] R[P_j]\rangle,$$

where the last sum on i and j is extended to all values of i and j except $0_v \doteq (0,0,\ldots,0)$. In general, for every k, the indices i_k and j_k appearing in each $\langle R[P_i] R[P_j]\rangle$ of (44) can assume any integer value between 0 and 3. For given i and j, let S be the set of indices k for which $i_k \neq 0 \wedge j_k \neq 0$, s the

Accepted in ⟨ ⟩uantum 2021-03-08, click title to verify. Published under CC-BY 4.0.

12

cardinality of S and $\overline{S}$ the set containing the remaining indices k of $1, \ldots, N$ with cardinality $\bar{s} = N - s$. We use (16) for a Pauli monomial to arrive at

$$
\begin{aligned}
\langle R[P_i]R[P_j]\rangle &= \left(\prod_{h\in\overline{S}} \int_{\Sigma_h} \frac{d\vec{n}_h}{4\pi} \sum_{m_h} R_1[\sigma_{i_h}](m_{i_h}, \vec{n}_{i_h}) R_1[\sigma_{j_h}](m_{j_h}, \vec{n}_{j_h}) \right) \\
&\times \left(\prod_{k\in S} \int_{\Sigma_k} \frac{d\vec{n}_k}{4\pi} \sum_{m_k} R_1[\sigma_{i_k}](m_{i_k}, \vec{n}_{i_k}) R_1[\sigma_{j_k}](m_{j_k}, \vec{n}_{j_k}) \right) p(\{m, \vec{n}\}) \\
&= \left(\prod_{h\in\overline{S}} \int_{\Sigma_h} \frac{d\vec{n}_h}{4\pi} \sum_{m_h} R_1[\sigma_{i_h}](m_{i_h}, \vec{n}_{i_h}) R_1[\sigma_{j_h}](m_{j_h}, \vec{n}_{j_h}) \right) \\
&\times \left(9^s \prod_{k\in S} \int_{\Sigma_k} \frac{d\vec{n}_k}{4\pi} n_{i_k} n_{j_k} \right) \left(\prod_{k\in S} \sum_{m_k} p(\{m, \vec{n}\}) \right) \\
&= \left(\prod_{h\in\overline{S}} \int_{\Sigma_h} \frac{d\vec{n}_h}{4\pi} \sum_{m_h} R_1[\sigma_{i_h}](m_{i_h}, \vec{n}_{i_h}) R_1[\sigma_{j_h}](m_{j_h}, \vec{n}_{j_h}) \right) \\
&\times \left(3^s \prod_{k\in S} \delta_{i_k j_k} \right) p(m_{h_1}, \vec{n}_{h_1}, \ldots, m_{h_{\bar{s}}}, \vec{n}_{h_{\bar{s}}}) \\
&= \left(\prod_{h\in\overline{S}} \int_{\Sigma_h} \frac{d\vec{n}_h}{4\pi} \sum_{m_h} R_1[\sigma_{i_h}](m_{i_h}, \vec{n}_{i_h}) R_1[\sigma_{j_h}](m_{j_h}, \vec{n}_{j_h}) \right) \\
&\times \left(3^s \prod_{k\in S} \delta_{i_k j_k} \int_{\Sigma_k} \frac{d\vec{n}_k}{4\pi} \sum_{m_k} R_1[\sigma_{i_k}^2] \right) p(\{m, \vec{n}\}) \\
&= 3^s \left(\prod_{k\in S} \delta_{i_k j_k} \right) \left(\prod_{h\in\overline{S}} \int_{\Sigma_h} \frac{d\vec{n}_h}{4\pi} \sum_{m_h} R_1[\sigma_{i_h}\sigma_{j_h}](m_{i_h}, \vec{n}_{i_h}, m_{j_h}, \vec{n}_{j_h}) \right) \\
&\times \left(\prod_{k\in S} \int_{\Sigma_k} \frac{d\vec{n}_k}{4\pi} \sum_{m_k} R_1[\sigma_{i_k}\sigma_{j_k}](m_{i_k}, \vec{n}_{i_k}, m_{j_k}, \vec{n}_{j_k}) \right) p(\{m, \vec{n}\}) \\
&= 3^s \left(\prod_{k\in S} \delta_{i_k j_k} \right) \langle R[P_i P_j]\rangle = 3^s \left(\prod_{k\in S} \delta_{i_k j_k} \right) \mathrm{Tr}[P_i P_j \rho] \leqslant 3^s \left(\prod_{k\in S} \delta_{i_k j_k} \right) \doteq 3^{r_{ij}} \Delta_{ij}.
\end{aligned}
\tag{45}
$$

If we introduce a seminorm of O as

$$
\|O\| = \sqrt{\sum_{i,j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} |a_i||a_j|},
\tag{46}
$$

Eq. (44) becomes

$$
\mathrm{Var}(R[O](\{m, \vec{n}\})) \leqslant \|O\|^2.
\tag{47}
$$

The seminorm (46) is a tighter upper bound for the variance of $R[O]$ than the formally simpler seminorm

$$
\|O\|_1 = \sum_{i\neq 0_v} \sqrt{3^{r_i}} |a_i|,
\tag{48}
$$

with r_i representing the weight of P_i, since

$$
\|O\|^2 = \sum_{i,j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} |a_i||a_j| \leqslant \sum_{i,j(\neq 0_v)} \sqrt{3^{r_i} 3^{r_j}} |a_i||a_j| = \left(\sum_{i\neq 0_v} \sqrt{3^{r_i}} |a_i| \right)^2 = \|O\|_1^2
\tag{49}
$$

This is not surprising, since the seminorm (48) can be shown to be a bound for the variance of $R[O]$ by expanding $\mathrm{Var}(\sum_{i\neq 0_v} a_i R[P_i])$ of (44) and using Schwarz inequality without resorting to explicit

properties of R as in (45). The seminorm (46) can also be expressed as

$$\|O\|^2 = \|O\|_2^2 + \sum_{i \neq j (\neq 0_v)} 3^{r_{ij}} \Delta_{ij} |a_i||a_j|, \tag{50}$$

with the seminorm $\|O\|_2$ defined by

$$\|O\|_2 = \sqrt{\sum_{i \neq 0_v} 3^{r_i} a_i^2}. \tag{51}$$

We use (45) and rewrite (44) as

$$\mathrm{Var}(R[O]) \leqslant \sum_{i,j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} \mathrm{Tr}[P_i P_j \rho] a_i a_j = \|O\|_2^2 + \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} \mathrm{Tr}[P_i P_j \rho] a_i a_j. \tag{52}$$

Eq. (52) shows how the mixed terms of (50) are a weak upper bound for the last term of (52). For a randomly chosen O or, in the same way, for a random ρ, the addends in the last term of (52) have mixed positive and negative signs. Furthermore, for most states ρ, the expectation values of the expressions $P_i P_j$ tend to become smaller for higher values of r_{ij}. As a result, $\|O\|_2$ is generally a good approximation of the standard deviation of $R[O]$.

To formalise this statement, we start by showing that the expectation value of the mixed terms of (52) over the space of the density operators of the system is equal to zero. The probability of choosing a density operator from the space of density operators is the probability of choosing a set of eigenvalues multiplied by the probability of choosing a density operator with specific eigenvalues conditional to the choice of the eigenvalues. If for any choice of the eigenvalues, the expectation value over the density operators with the chosen eigenvalues is zero, then the expectation value over all density operators is zero. All density operators with the same eigenvalues can be obtained from one with a unitary transformation and, vice versa, all unitary transformations applied to a density operator preserve the eigenvalues. Therefore, we can calculate the expectation value of the mixed terms of (52) over the density operators with given eigenvalues by averaging over the unitary group [23] with the unitary transformations applied to an arbitrary density operator ρ_0

$$\int_U dU \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} a_i a_j \mathrm{Tr}[P_i P_j U \rho_0 U^\dagger] = \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} a_i a_j \mathrm{Tr}\left[P_i P_j \int_U dU \, U \rho_0 U^\dagger \right] \tag{53}$$
$$= \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} a_i a_j \mathrm{Tr}\left[\frac{P_i P_j}{2^N} \right] = 0,$$

which implies

$$\left\langle \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} \Delta_{ij} \mathrm{Tr}[P_i P_j \rho] a_i a_j \right\rangle_\rho = 0. \tag{54}$$

To gain further information on how close $\|O\|_2$ is likely to be to $\|O\|$ for a random state, we now consider the variance of the mixed terms of (50) over the space of the density operators of the system. For simplicity, we specialise the analysis to pure states. From the unitary group notation of (53), we switch to the average over the states expressed as in [24] and recall the following identities:

$$\int_\psi d\psi |\psi\rangle\langle\psi|^{\otimes 2} = \frac{2}{2^N(2^N + 1)} \mathbb{1}_{\mathrm{sym}}, \tag{55}$$

with $\mathbb{1}_{\mathrm{sym}}$ projector onto the symmetric subspace of $\mathcal{H}^{\otimes 2}$, and

$$\mathrm{Tr}[A \otimes B \, \mathbb{1}_{\mathrm{sym}}] = \frac{1}{2}(\mathrm{Tr}[A]\mathrm{Tr}[B] + \mathrm{Tr}[AB]), \tag{56}$$

for any operator A and B. The variance of the mixed terms of (52) over the states of the system can be expressed as

$$\mathrm{Var}\left(\sum_{i\neq j(\neq 0_v)} 3^{r_{ij}}\,\Delta_{ij}\,\mathrm{Tr}[P_iP_j|\psi\rangle\langle\psi|]\,a_i\,a_j\right)_\psi = \int_\psi d\psi\left(\sum_{i\neq j(\neq 0_v)} a_i\,a_j\,3^{r_{ij}}\,\Delta_{ij}\,\mathrm{Tr}[P_iP_j|\psi\rangle\langle\psi|]\right)^2 \quad (57)$$

$$= \sum_{i\neq j(\neq 0_v)}\sum_{i'\neq j'(\neq 0_v)} a_i\,a_j\,a_{i'}a_{j'}\,3^{r_{ij}}\,3^{r_{i'j'}}\,\Delta_{ij}\,\Delta_{i'j'}\int_\psi d\psi\,\left(\mathrm{Tr}[P_iP_j|\psi\rangle\langle\psi|]\,\mathrm{Tr}[P_{i'}P_{j'}|\psi\rangle\langle\psi|]\right)$$

$$= \sum_{i\neq j(\neq 0_v)}\sum_{i'\neq j'(\neq 0_v)} a_i\,a_j\,a_{i'}a_{j'}\,3^{r_{ij}}\,3^{r_{i'j'}}\,\Delta_{ij}\,\Delta_{i'j'}\,\mathrm{Tr}\left[(P_iP_j)\otimes(P_{i'}P_{j'})\int_\psi d\psi|\psi\rangle\langle\psi|^{\otimes 2}\right]$$

$$= \frac{2}{2^N(2^N+1)}\sum_{i\neq j(\neq 0_v)}\sum_{i'\neq j'(\neq 0_v)} a_i\,a_j\,a_{i'}a_{j'}\,3^{r_{ij}}\,3^{r_{i'j'}}\,\Delta_{ij}\,\Delta_{i'j'}\,\mathrm{Tr}\left[(P_iP_j)\otimes(P_{i'}P_{j'})\,\mathbb{1}_{\mathrm{sym}}\right]$$

$$= \frac{1}{2^N(2^N+1)}\sum_{i\neq j(\neq 0_v)}\sum_{i'\neq j'(\neq 0_v)} a_i\,a_j\,a_{i'}a_{j'}\,3^{r_{ij}}\,3^{r_{i'j'}}\,\Delta_{ij}\,\Delta_{i'j'}\,\mathrm{Tr}[P_iP_jP_{i'}P_{j'}]$$

$$= \frac{1}{(2^N+1)}\sum_{i\neq j(\neq 0_v)}\sum_{i'\neq j'(\neq 0_v)} a_i\,a_j\,a_{i'}a_{j'}\,3^{r_{ij}}\,3^{r_{i'j'}}\,\Delta_{ij}\,\Delta_{i'j'}\,\overline{\Delta}_{iji'j'},$$

with $\overline{\Delta}_{iji'j'} = 1$ when i, j, i', j' are such that $P_iP_jP_{i'}P_{j'} = \mathbb{1}$, otherwise $\overline{\Delta}_{iji'j'} = 0$. A general upper bound of the last expression is likely to be weak, given its strong dependence on O. The specific case of O given by a projector onto a computational basis state is examined in App. C.

C   Projectors onto the computational basis states

In this section we examine properties of the seminorm for an important special case, namely, the case of projectors onto a generic element of the computational basis state,

$$|x\rangle\langle x| = \bigotimes_{k=1}^N \frac{1}{2}(\mathbb{1}\pm\sigma_{z_k}), \tag{58}$$

where the $\pm$ of each factor is defined by the choice of the element of the computational basis $|x\rangle$. This case is important not only for its practical significance in a machine learning context, as briefly indicated in Sec. 5, but also because it is representative of the cases with bounded $\|O\|_2$ and diverging $\|O\|$ for increasing N, which makes the determination of whether the statistical error of the estimator can be approximated by $\|O\|_2$ of particular interest. To calculate the seminorms of any projector onto a computational basis state, we expand the tensor product of (58) and use the seminorm definitions. For $\||x\rangle\langle x|\|_2$ this simply means

$$\||x\rangle\langle x|\|_2^2 = \frac{1}{4^N}\left(\binom{N}{N-1}3^1 + \binom{N}{N-2}3^2 + \ldots + \binom{N}{0}3^N\right) = \frac{1}{4^N}((1+3)^N) - \frac{1}{4^N} = 1 - \frac{1}{4^N}. \tag{59}$$

A more involved combinatorial calculation would show that $\||x\rangle\langle x|\|^2 \leqslant (3/2)^N$ and $\||x\rangle\langle x|\|^2 = O((3/2)^N)$. We come back to (57) for any projector onto computational basis state and observe that, for given i and j, there are 2^N ordered pairs i', j' such that $\overline{\Delta}_{iji'j'} = 1$. The operator P_iP_j is in fact the tensor product of single–qubit identities and operators σ_z. Each of them becomes the identity when multiplied by itself and this can be obtained with two choices for every qubit of $P_{i'}$ and $P_{j'}$. Unfortunately, to complicate things, $r_{i'j'}$ is not independent of the choice of i' and j' for given i and j;

however, we can simplify the evaluation of (57) by considering a limit case and determining an upper bound. In particular, for every i and j, with $i \neq j$ and i and j different from 0_v,

$$\sum_{i',j'} 3^{r_{i'j'}} \overline{\Delta}_{iji'j'} < \sum_{i',j'} 3^{r_{i'j'}} \overline{\Delta}_{iii'j'} = \sum_{l=0}^{N} \binom{N}{l} 3^l = 4^N. \tag{60}$$

With the substitutions $a_i = (1/2)^N$ and $\Delta_{ij} = 1$ valid for any projector $|x\rangle\langle x|$ and any value of the indexes, the last term of Eq. (57) becomes

$$\frac{1}{(2^N+1)} \sum_{i \neq j(\neq 0_v)} \sum_{i' \neq j'(\neq 0_v)} a_i \, a_j \, a_{i'} a_{j'} \, 3^{r_{ij}} \, 3^{r_{i'j'}} \, \Delta_{ij} \, \Delta_{i'j'} \, \overline{\Delta}_{iji'j'} \tag{61}$$

$$< \frac{1}{(2^N+1) \, 4^N} \sum_{i \neq j(\neq 0_v)} 3^{r_{ij}} = \frac{\||x\rangle\langle x|\|^2}{2^N+1} < \left(\frac{3}{4}\right)^N.$$

Eq. (61) shows that, for any projector $|x\rangle\langle x|$, $\mathrm{Var}(R[|x\rangle\langle x|])$ for a random state is bounded by $\||x\rangle\langle x|\|_2^2 < 1$ with probability exponentially close to 1 as a function of N.