

Learning to measure: adaptive informationally complete POVMs for near-term quantum algorithms

Guillermo García-Pérez,^{1,2,3} Matteo A. C. Rossi,^{1,3} Boris Sokolov,^{4,3} Francesco Tacchino,⁵ Panagiotis Kl. Barkoutsos,⁵ Guglielmo Mazzola,⁵ Ivano Tavernelli,⁵ and Sabrina Maniscalco^{4,6,3}

¹*QTF Centre of Excellence, Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014 Turun Yliopisto, Finland*

²*Complex Systems Research Group, Department of Mathematics and Statistics, University of Turku, FI-20014 Turun Yliopisto, Finland*

³*Algorithmiq Ltd, Linnankatu 55 K 329, 20100 Turku, Finland*

⁴*QTF Centre of Excellence, Department of Physics, Faculty of Science, University of Helsinki, Finland*

⁵*IBM Quantum, IBM Research – Zurich, 8803 Rüschlikon, Switzerland*

⁶*QTF Centre of Excellence, Center for Quantum Engineering, Department of Applied Physics, Aalto University School of Science, FI-00076 Aalto, Finland*

Variational quantum algorithms stand as the most promising approaches towards practical applications of near-term quantum computers. However, these methodologies usually require a large number of measurements, which represents an important roadblock for future real-world use cases. We introduce a novel approach to tackle this problem: a variational measurement scheme. We present an algorithm that optimises informationally complete POVMs on the fly in order to minimise the statistical fluctuations in the estimation of relevant cost functions. We use it in combination with the Variational Quantum Eigensolver to calculate ground-state energies of molecular Hamiltonians in numerical simulations and show that it is competitive with state-of-the-art measurement reduction approaches. We also highlight the potential of the informational completeness of the measurement outcomes by reusing the ground-state energy estimation data to perform high-fidelity reduced state tomography.

I. INTRODUCTION

Quantum computing is a rapidly growing multidisciplinary field with a very clear objective: understanding if, and to what extent, it is possible to build computing machines able to perform tasks that are impossible for conventional (classical) computers. Theoretically, milestone discoveries such as Shor’s and Grover’s quantum algorithms hint towards a positive answer to this question. These algorithms, exploiting quantum properties of the processor, can in principle outperform all currently existing classical methods. In practice, however, the implementation of such protocols in the regimes of interest will most probably require the use of ideal fault-tolerant universal quantum computers. At the same time, because of the extreme fragility of quantum information storage and processing in presence of environmental noise, error-correction techniques required to achieve fault-tolerance are still experimentally in their infancy.

Universal fault-tolerant quantum computers, however, are not the only type of quantum machines able to tackle computationally hard problems. In fact, we can reformulate the main quantum computing research question and ask ourselves: what are the useful problems that quantum computers can solve more efficiently than their classical counterparts and, specifically, which subclass of such problems are less demanding in terms of experimental requirements, given the current state-of-the-art quantum hardware? Notice that this question has a different starting point, namely it focuses on our current—or near-future—technologies and devices, and aims at

identifying, based on the current understanding, useful applications that may benefit from them.

There are at least two classes of problems that satisfy the requirements above. The first class has a longstanding history, dating back to Feynman (1982) [1] and Manin (1980) [2], who pointed out that simulating quantum systems is hard on classical computers, while, under certain conditions, they can be efficiently investigated by means of other quantum systems [3]. In fact, this can be done using either digital quantum simulators, namely specific-purpose quantum computers [4–6], or by employing analogue quantum simulators [7–10], namely other equivalent but easier-to-control quantum systems. The second class of problems emerges when we lift the requirement of finding “exact” solutions to a given problem. Approximate near-term quantum devices might be able, e.g., to find better solutions to certain worst-case instances of NP-hard problems, or find such approximate solutions faster.

A final ingredient to move towards the existing approximate noisy quantum devices [11, 12] is the combination of quantum and classical techniques to maximise performance. In this paper, we focus on variational quantum algorithms, which have emerged recently as the most suited paradigm to tackle the classes of problems identified above [13, 14] with approximate quantum computing. Specifically, these protocols are implemented by preparing a parametrised N -qubit trial state on a quantum device, extracting some observable quantities with suitable measurements and processing such measurement outcomes using a classical optimiser. The latter then re-

turns the small changes that need to be implemented to prepare, in the next step, an updated trial wavefunction. This cycle is repeated many times until it converges to a quantum state from which the desired approximate solution can be extracted.

This procedure can be used to solve problems in chemistry [15–18], for the design of new materials [19], and generally in every field of physics where one needs to extract the properties of many-body quantum correlated systems, e.g., interacting fermionic systems, typically hard to simulate on classical devices [20, 21]. In this case, these algorithms go by the name of Variational Quantum Eigensolvers (VQE) [15, 22, 23]. In essence, the quantum processor is used to explore the exponentially-large Hilbert space of the fermionic particles in order to find iteratively the ground state of the Hamiltonian, without solving the full diagonalisation problem. As an example, the knowledge of the ground state of a chemical compound as a function, e.g., of the nuclear coordinates allows one to extract crucial information such as the equilibrium bond length, bond angle, and dissociation energy. Notice that, at least in principle, a quantum computer with a few hundreds of qubits could already have the potential to solve useful quantum chemistry problems which are intractable on classical computers.

The application of VQE has already been demonstrated in many proof-of-principle experiments [15, 22, 24–26]. However, a few major challenges still need to be overcome along the path to useful quantum advantage. On one hand, the classical optimisation step which is associated with variational quantum algorithms can in general incur into high computational costs because of the existence of many local minima or due to the problem of vanishing gradients [27]. Some possible solutions have been proposed, combining techniques borrowed from classical optimisation theory with a careful design of the variational ansatz, such as the recently proposed ADAPT-VQE [28] and oo-VQE [29], and of the associated cost function [30]. On the other hand, the so-called measurement problem arises from the very high cost in terms of the number of observations that are typically needed to reconstruct the properties of interest, and specifically the expectation value of the Hamiltonian, on the quantum states constructed by variational means. In fact, as the size of the problem approaches the regime in which the VQE could compete with classical methods, the current approaches would lead to prohibitive requirements to reach the desired degree of accuracy [21, 31–33].

In this work, we tackle the second problem, by presenting a novel adaptive method that sensibly alleviates the demands on the number of measurements, thus paving the way for an increase of the affordable problem sizes in experimental realisations. On a fundamental level, our approach introduces a new perspective on how to improve the overall observables reconstruction strategy in VQE, and possibly in variational algorithms in general, by leveraging informationally complete quantum measurements. Before introducing our protocol, however, we

describe in more detail the measurement problem and briefly mention the main approaches that have been proposed in the literature to tackle it in the next section.

II. THE MEASUREMENT PROBLEM IN VQE

One of the most prominent differences between classical and quantum methods concerns the way in which information is extracted at the end of the algorithm’s execution. In a typical situation, the quantum circuit prepares a N -qubit quantum state $|\psi\rangle$ which is used to compute the expectation value of an operator $\langle\mathcal{O}\rangle = \langle\psi|\mathcal{O}|\psi\rangle$. Generally, it is not possible to measure \mathcal{O} directly in its eigenbasis. For instance if we are interested in finding the ground state of the Hamiltonian H , measuring in its eigenbasis requires solving the problem itself in advance.

The standard measurement protocol, henceforth named *Pauli* method, consists in writing the operator \mathcal{O} as a sum of K Pauli strings, $\mathcal{O} = \sum_{\mathbf{k}} c_{\mathbf{k}} P_{\mathbf{k}}$, where $P_{\mathbf{k}} = \bigotimes_{i=1}^N \sigma_{k_i}^{(i)}$ and $\sigma_0^{(i)} = \mathbb{I}^{(i)}$, $\sigma_1^{(i)} = \sigma_x^{(i)}$, \dots are Pauli operators. The expectation value of the operator is therefore obtained in terms of the weighted sum of K expectation values.

Unfortunately, this method leads to a sub-optimal measurement scheme, as the variance of \mathcal{O} is the sum of the weighted variances of the individual operators $P_{\mathbf{k}}$. More precisely, the error in the estimation is given by

$$\epsilon = \sqrt{\sum_{\mathbf{k}} |c_{\mathbf{k}}|^2 \text{Var}(P_{\mathbf{k}}) / S_{\mathbf{k}}}, \quad (1)$$

where $\text{Var}(P_{\mathbf{k}}) = \langle P_{\mathbf{k}}^2 \rangle - \langle P_{\mathbf{k}} \rangle^2$ is the variance of $P_{\mathbf{k}}$ and $S_{\mathbf{k}}$ is the number of measurements, i.e. wavefunction collapses, used to estimate term \mathbf{k} [32]. Interestingly, under such measurement scheme, even exactly prepared ground states do not enjoy the *zero-variance* property, such that statistical energy fluctuations always remain finite and large.

This constitutes a major source of problems for variational-based state preparation, where circuit parameters are optimised to minimise the expectation value of the energy. Given its significance, several efforts have been put forward to mitigate this problem. One simple strategy, henceforth named *Grouped Pauli* method, aims at identifying all the Pauli strings that can be measured simultaneously from the same dataset [15]. While this is not solving the issue, it reduces the computational overhead of the procedure. Promising approaches also involve the usage of a classical machine learning engine to perform an *approximate* reconstruction of the quantum state [34] using only the basis state defined by $P_{\mathbf{k}}$ [35], or *classical shadows* of a quantum state [36–38]. Other approaches based on grouping of commuting terms, effective measurement scheduling and optimised qubit tomography have been described in Refs. [39–47].

In this work, we present an algorithm for efficient observable estimation that exploits generalised quantum

measurements integrating three important components: a hybrid quantum-classical Monte Carlo, a method to navigate generalised measurement space towards efficient measurements, and a recipe to combine different estimations of the observable of interest. The result is a procedure in which the optimal measurement of an operator average is learnt in an adaptive fashion with no measurement overhead.

III. ADAPTIVE MEASUREMENT SCHEME

In this section, we explain our adaptive measurement scheme. In a nutshell, the idea is to use parametric informationally complete positive operator-valued measures (IC-POVMs), which can in principle be used to estimate any expectation value of our choice. We first introduce a hybrid Monte Carlo approach, which bypasses the need to use tomographic reconstructions of quantum states from the IC data. We then describe how, by using parametric families of POVMs, the measurement settings can be optimised to yield low statistical errors in the estimation of the target expectation values.

With respect to the second point, special attention must be devoted to achieving the desired POVM optimisation without incurring in additional overheads in terms of, e.g., the number of repetitions (also named *shots* in the following) of the state preparation and measurement routine. As we will explain in the following, an adaptive method, that is, an on-the-fly optimisation will serve this scope. In brief, the key is to use the IC data obtained with one given POVM twice: first, we use them to produce an estimation of the mean of the observable. Second, the same set of results can also be employed to find a better POVM for the next experiment. The collection of intermediate estimators of the target observable, each constructed along the process with a different POVM, is finally integrated together as to minimise the overall statistical uncertainty. As a result of this strategy, the measurement learning procedure improves over the initial POVM (which turns out to be already quite efficient, as shown in Sec. IV) with no additional measurement costs. The scheme is illustrated and summarised in Fig. 1.

It is important to stress that the method does not require any approximations whatsoever. In fact, it is completely agnostic to the nature of the operator \mathcal{O} to be measured, as long as it is given in terms of a linear combination of products of single-qubit observables (e.g., Pauli strings). While the algorithm is rather general, its performance is strongly dependent on the weight of such products (the number of non-identity single-qubit operators in every term), as we explain later, which makes quantum chemistry with low-weight fermion-to-qubit mappings, such as Bravyi-Kitaev [48] and the one recently introduced in Ref. [49], ideal use cases.

To ease the explanation of the algorithm, we present its three main components separately. We first introduce the hybrid quantum-classical Monte Carlo sampling

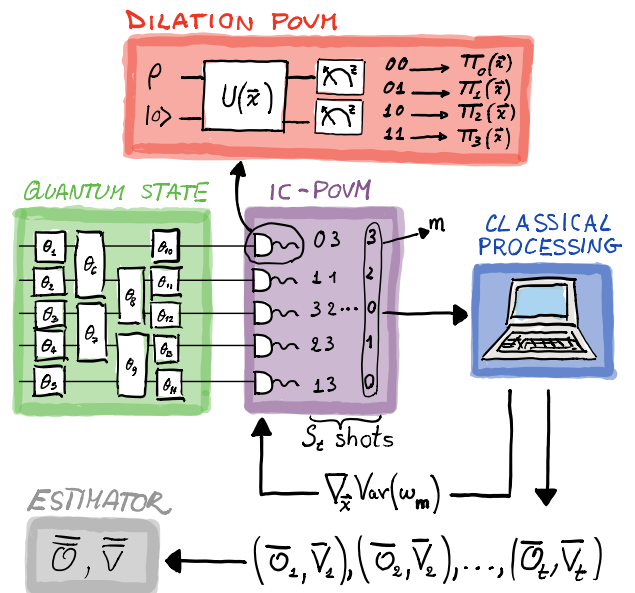


FIG. 1. **Adaptive measurement scheme.** The sketch represents a typical measurement step of a variational algorithm. The ansatz prepares a state $|\psi(\vec{\theta})\rangle$ (green box) for which the mean of some observable \mathcal{O} must be evaluated. Our algorithm is an efficient measurement subroutine in this process. It relies on parametric informationally complete POVMs (purple box) implemented with ancillary qubits (red box). These are explained in detail in App. A. Initially, we start by performing S_1 measurements using the POVM corresponding to parameters \vec{x}_1 , and obtain S_1 outcomes $\mathbf{m}_1, \dots, \mathbf{m}_{S_1}$. The measurement data are post-processed efficiently on a classical device (blue box) twice, with two different goals. First, we estimate the mean of the observable, $\bar{\mathcal{O}}_1$, and the corresponding error of the estimation, \bar{V}_1 , as explained in Sec. III A. Second, we calculate the gradient of the estimation variance, $\nabla_{\vec{x}} \text{Var}(\omega_{\mathbf{m}})$, in POVM parameter space, and thus find a better POVM for iteration 2 (see Sec. III B and App. B). At every step t , the variables $\bar{\mathcal{O}}$ and \bar{V} integrate all the estimations for $t' \leq t$ while minimising the overall statistical error (see Sec. III C and App. D). The process is repeated iteratively until \bar{V} is below some desired threshold.

for the estimation of expectation values of operators in Sec. III A. We then show in Sec. III B how to estimate the gradient in the space of POVMs without additional measurements, using only efficient classical post-processing and, lastly, in Sec. III C, we illustrate how to integrate all the data obtained from different POVMs to estimate mean values while minimising statistical fluctuations.

A. Hybrid quantum-classical Monte Carlo sampling

Our proposed algorithm relies on single-qubit (minimal) informationally complete (IC) POVMs, which can be realised by applying a two-qubit gate between a system qubit and an ancillary one, the latter in a known

state, and then measuring both qubits in the computational basis. The implementation of these POVMs on current quantum computers has been recently demonstrated experimentally on IBM Quantum devices [50, 51].

By definition, one such POVM is represented by four linearly independent positive operators $\{\Pi_i > 0, i = 0, \dots, 3\}$ adding up to identity, $\sum_i \Pi_i = \mathbb{I}$, and spanning the space of linear operators in the Hilbert space of the system qubit. Each of these operators, usually called *effects*, is associated with one of the four possible outcomes of the two-qubit measurement, with $\text{Tr}[\rho \Pi_i]$ being the probability of outcome i on the quantum state ρ of the target qubit. It is important to notice that different qubit-ancilla unitaries generally lead to different POVMs. Hence, by parametrising these unitaries, we can parametrise the corresponding family of POVMs (see App. A).

Let us consider the N -qubit case, with local and not necessarily identical POVMs associated to each qubit. The four effects associated with qubit i are denoted by $\Pi_m^{(i)}$, with m running from 0 to 3. The outcome of an experiment in which all qubits are measured via these local POVMs is a string $\mathbf{m} = (m_1, \dots, m_N)$, where $m_i \in \{0, \dots, 3\}$. The probability of such outcome given an N -qubit state ρ is $\text{Tr}[\rho \Pi_{\mathbf{m}}]$, with $\Pi_{\mathbf{m}} = \bigotimes_{i=1}^N \Pi_{m_i}^{(i)}$. The set of these 4^N effects $\Pi_{\mathbf{m}}$ is IC in $\mathcal{H}^{\otimes N}$.

As explained in previous sections, in VQE realisations one typically needs to measure an operator \mathcal{O} that can be decomposed in terms of K Pauli strings, $\mathcal{O} = \sum_{\mathbf{k}} c_{\mathbf{k}} P_{\mathbf{k}}$ (we assume $c_{\mathbf{k}} \in \mathbb{R}$, as is customary, although our results can be easily generalised to complex-valued coefficients). Given that each of the local POVMs is IC, we can express the Pauli operators acting on each qubit i in terms of the effects $\Pi_m^{(i)}$ as $\sigma_k^{(i)} = \sum_m b_{km}^{(i)} \Pi_m^{(i)}$, with which we can write

$$\begin{aligned} \mathcal{O} &= \sum_{\mathbf{k}} c_{\mathbf{k}} \bigotimes_{i=1}^N \sigma_{k_i}^{(i)} = \sum_{\mathbf{k}} c_{\mathbf{k}} \bigotimes_{i=1}^N \left(\sum_{m_i=0}^3 b_{k_i m_i}^{(i)} \Pi_{m_i}^{(i)} \right) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{k}} c_{\mathbf{k}} \prod_{i=1}^N b_{k_i m_i}^{(i)} \Pi_{\mathbf{m}} \equiv \sum_{\mathbf{m}} \omega_{\mathbf{m}} \Pi_{\mathbf{m}}. \end{aligned} \quad (2)$$

The above expression seems useless at first sight: we transformed a representation of \mathcal{O} in terms of K terms $c_{\mathbf{k}} P_{\mathbf{k}}$ into one with possibly 4^N terms $\omega_{\mathbf{m}} \Pi_{\mathbf{m}}$. However, the expectation value of the operator now reads

$$\langle \mathcal{O} \rangle = \text{Tr}[\rho \mathcal{O}] = \sum_{\mathbf{m}} \omega_{\mathbf{m}} \text{Tr}[\rho \Pi_{\mathbf{m}}] = \sum_{\mathbf{m}} \omega_{\mathbf{m}} p_{\mathbf{m}}, \quad (3)$$

where $p_{\mathbf{m}}$ is the probability to obtain outcome \mathbf{m} . In other words, the mean value of the operator is the average of $\omega_{\mathbf{m}}$ over the probability distribution $\{p_{\mathbf{m}}\}$, $\langle \mathcal{O} \rangle = \langle \omega_{\mathbf{m}} \rangle_{\{p_{\mathbf{m}}\}}$, so we can transform the original problem of measuring a linear combination of a large number of operators $P_{\mathbf{k}}$ into a Monte Carlo integral.

The strategy is now to repeat the measurement S times using the local POVMs to sample from the probability

distribution $\{p_{\mathbf{m}}\}$, resulting in a sequence of outcomes $\mathbf{m}_1, \dots, \mathbf{m}_S$, and compute

$$\bar{\mathcal{O}} = \frac{1}{S} \sum_{s=1}^S \omega_{\mathbf{m}_s}. \quad (4)$$

Each term $\omega_{\mathbf{m}_s} = \sum_{\mathbf{k}} c_{\mathbf{k}} \prod_{i=1}^N b_{k_i m_i}^{(i)}$ can be calculated in a polynomial time on a classical computer. This estimator converges to $\langle \mathcal{O} \rangle$ as $\sqrt{\text{Var}(\omega_{\mathbf{m}})/S}$, where $\text{Var}(\omega_{\mathbf{m}})$ is the variance of $\omega_{\mathbf{m}}$ over the probability distribution $\{p_{\mathbf{m}}\}$. Crucially, this method estimates the weighted average of all the Pauli strings $P_{\mathbf{k}}$ simultaneously, regardless of whether they commute or not, by exploiting IC data, yet circumventing any costly tomographic reconstruction of quantum states. In addition, in this Monte Carlo approach, the variance naturally takes into account the covariance between all these parallel measurements. In other words, the quantity $\sqrt{(\langle \omega_{\mathbf{m}}^2 \rangle_{\{p_{\mathbf{m}}\}} - \langle \omega_{\mathbf{m}} \rangle_{\{p_{\mathbf{m}}\}}^2)/S}$, which can be estimated efficiently from the data, accounts for the total statistical error. As we explain next, our strategy is to iteratively search for POVMs that minimise this error.

Importantly, the previous result holds for any operator \mathcal{O} , that is, the same sequence of outcomes $\mathbf{m}_1, \dots, \mathbf{m}_S$ can be used to estimate, using only classical post-processing, any expectation value. However, not all expectation values can be estimated with the same precision. In particular, notice that the products $\prod_{i=1}^N b_{k_i m_i}^{(i)}$ can potentially result in variances scaling exponentially in N . This limitation can be overcome for fermionic problems by using fermion-to-qubit mappings such as the Bravyi-Kitaev (BK) [48] and especially the one recently proposed in Ref. [49] by Jiang et al. (to which we will refer to as JKMN mapping), which lead to Pauli strings with logarithmic weight (that is, such that fermionic creation/annihilation operators are mapped onto Pauli strings with at most a logarithmic number of non-identity Pauli operators). Since the terms $b_{0m_i}^{(i)}$, corresponding to the decomposition of identity, are always equal to one (recall that $\sum_m \Pi_m^{(i)} = \mathbb{I}^{(i)}$), these mappings ensure that the products $\prod_{i=1}^N b_{k_i m_i}^{(i)}$ scale polynomially in N .

Regarding the method proposed in Ref. [49], it should be mentioned that our Monte Carlo approach, Eq. (4), offers some advantages over the latter. On the one hand, it bypasses the classical overhead needed for tomographic reconstructions. On the other hand, and more importantly, our approach does not disregard the covariance-induced statistical errors in the estimation of the average resulting from parallel measurements. These points are discussed in more details in App. C.

B. Classical gradient estimation for POVM optimisation

Modifying the POVM results in a different probability distribution $\{p_{\mathbf{m}}\}$, as well as different weights $\omega_{\mathbf{m}}$, hence

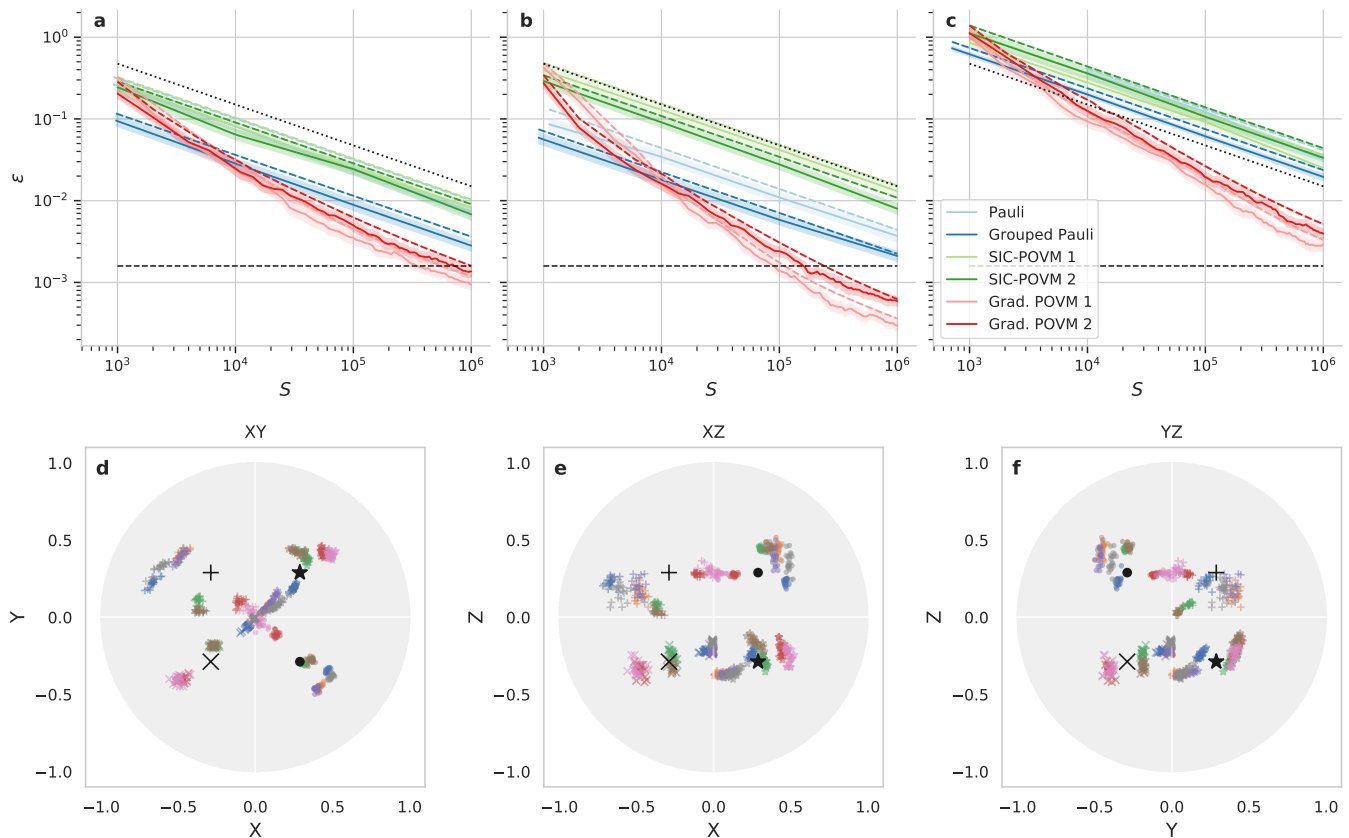


FIG. 2. **Adaptive measurement for energy estimation.** (a-c) solid lines indicate the error in the estimation of the ground energy for selected Hamiltonians as a function of the total number of shots, with a VQE state, for various methods: Pauli, Grouped Pauli, as well as the hybrid Monte Carlo without optimisation (SIC-POVM 1 and 2) and the gradient-optimised method (Grad. POVM 1 and 2) with two different initial conditions (see App. C). The dashed lines correspond to the estimated statistical error ($\bar{V}^{1/2}$ in the case of the POVM methods). The curves show the average error over 100 realisations of the numerical experiments and the shaded areas the 95% confidence interval obtained with bootstrapping. The horizontal dashed line indicates chemical accuracy, while the tilted one illustrates the $S^{-1/2}$ scaling. The three examples are (a) 8-qubit H_2 with JKMN mapping, (b) 8-qubit LiH with BK, and (c) 14-qubit H_2O with JKMN. (d-f) final POVM effects in the gradient optimisation process, when starting from the SIC-POVM 2, for a sample of 20 realisations from the data set of (a). Every POVM effect is mapped onto the three-dimensional unit-radius ball in a similar way as how single-qubit states are mapped onto the Bloch ball. In particular, the point $\vec{r} = (r_x, r_y, r_z)$, $|\vec{r}| \leq 1$, is associated with the effect $\Pi(\vec{r}) = (|\vec{r}| \mathbb{I} + \vec{r} \cdot \vec{\sigma})/2$ (notice the difference with the Bloch ball representation of quantum states; see App. A). In the figure, the colour indicates the qubit to which an effect corresponds, while the symbol identifies the effect itself among the possible four. The black symbols locate the initial effects, common to all realisations and qubits. Each panel presents the projection of the ball onto a different plane. The clustering of the points with equal colour and symbol reveal that all realisations reach approximately the same optimal measurement. However, the result of the optimisation is different for every qubit. Moreover, starting with SIC-POVM 1 instead leads to a very different measurement (see [52]).

potentially different $\text{Var}(\omega_{\mathbf{m}})$. This can be exploited to devise an adaptive algorithm in which the measurement of $\langle \mathcal{O} \rangle$ is optimised over the space of POVMs, that is, by finding one that minimises the variance $\text{Var}(\omega_{\mathbf{m}})$. We now propose a classical post-processing routine to navigate the space of POVMs towards low-variance ones. Essentially, besides using the outcomes obtained with the current POVM to construct an estimation of the target observable, the same set of data is also employed in a classical routine to assess the variance of *other* POVMs that were not previously implemented on the quantum

processor. Such procedure is explained in detail in the following.

Suppose that we want to evaluate the Monte Carlo variance $\text{Var}(\omega'_{\mathbf{r}})$ for a new POVM defined in terms of local POVMs with effects $\{\Gamma_{\mathbf{r}}^{(i)}\}$, that is,

$$\text{Var}(\omega'_{\mathbf{r}}) = \sum_{\mathbf{r}} \omega'_{\mathbf{r}}{}^2 \text{Tr}[\rho \Gamma_{\mathbf{r}}] - \left(\sum_{\mathbf{r}} \omega'_{\mathbf{r}}{}^2 \text{Tr}[\rho \Gamma_{\mathbf{r}}] \right)^2, \quad (5)$$

where the $\omega'_{\mathbf{r}}$ are given by the $b_{kr}^{(i)}$ matrices corresponding to these local POVMs, and $\Gamma_{\mathbf{r}} = \bigotimes_{i=1}^N \Gamma_{r_i}^{(i)}$. The

second term in Eq. (5) is the squared mean $\langle \mathcal{O} \rangle^2$, which does not depend on the POVM. The first term, i.e. the second moment of $\omega_{\mathbf{r}}$ over the probability distribution $\{q_{\mathbf{r}} \equiv \text{Tr}[\rho \Gamma_{\mathbf{r}}]\}$, $\langle \omega_{\mathbf{r}}'^2 \rangle_{\{q_{\mathbf{r}}\}}$, is the one that we must minimise. Suppose further that we have already run some experiments on the quantum computer with another IC-POVM given by the effects $\{\Pi_m^{(i)}\}$. Since this POVM is IC, we can write $\Gamma_{\mathbf{r}}^{(i)} = \sum_m d_{r_i m}^{(i)} \Pi_m^{(i)}$, where the $d_{r_i m}^{(i)}$ are real numbers. Inserting these decompositions into the expression for the second moment, we obtain

$$\begin{aligned} \langle \omega_{\mathbf{r}}'^2 \rangle_{\{q_{\mathbf{r}}\}} &= \sum_{\mathbf{r}} \omega_{\mathbf{r}}'^2 \text{Tr} \left[\rho \bigotimes_{i=1}^N \left(\sum_{m_i=0}^3 d_{r_i m_i}^{(i)} \Pi_{m_i}^{(i)} \right) \right] \\ &= \sum_{\mathbf{m}} p_{\mathbf{m}} \sum_{\mathbf{r}} \prod_{i=1}^N d_{r_i m_i}^{(i)} \omega_{\mathbf{r}}'^2. \end{aligned} \quad (6)$$

This last expression is also calculated in a hybrid Monte Carlo manner. More precisely, we can reuse the strings $\mathbf{m}_1, \dots, \mathbf{m}_S$ obtained from the measurements on the quantum computer (sampled from the probability distribution $\{p_{\mathbf{m}}\}$) to estimate the variances of other POVMs by calculating, for each \mathbf{m}_s , the corresponding $\sum_{\mathbf{r}} \prod_{i=1}^N d_{r_i m_i}^{(i)} \omega_{\mathbf{r}}'^2$ classically. Notice, however, that this last sum cannot always be computed efficiently, since it generally contains 4^N terms (both positive and negative), and involves products $\prod_{i=1}^N d_{r_i m_i}^{(i)}$ that can scale exponentially in N . To ensure the feasibility of the procedure, we use a gradient descent approach for the optimisation of the POVMs; in such case, only one of the terms in the product is different from one.

For concreteness, suppose that we used the effects $\{\Pi_m^{(i)}\}$ corresponding to the point \vec{x} in the POVM parameter space (see App. A) on the quantum computer and obtained S samples with which we can estimate the second moment $\langle \omega_{\mathbf{m}}^2 \rangle$. We can approximate the partial derivative of the second moment with respect to one of the parameters (for instance, the k -th), as $\partial_{x_k} \langle \omega_{\mathbf{m}}^2 \rangle \approx (\langle \omega_{\mathbf{r}}'^2 \rangle - \langle \omega_{\mathbf{m}}^2 \rangle)/h$, where $\langle \omega_{\mathbf{r}}'^2 \rangle$ is the estimated second moment corresponding to the POVM whose coordinates in parameter space \vec{x}' fulfil $x'_j = x_j + h \delta_{kj}$ (let us denote the corresponding effects by $\{\Gamma_{\mathbf{r}}^{(i)}\}$), and $h \ll 1$.

Since all single-qubit POVM are parametrised independently, $N-1$ of them are identical to the ones already used on the device, that is, $d_{r_i m}^{(i)} = \delta_{r_i m}$, $\forall i \neq l$, where l is the qubit whose POVM depends on the k -th parameter. Introducing this expression into Eq. (6), we obtain

$$\langle \omega_{\mathbf{r}}'^2 \rangle = \sum_{\mathbf{m}} p_{\mathbf{m}} \sum_{r_l=0}^3 d_{r_l m_l}^{(l)} \omega_{(m_1, \dots, m_{l-1}, r_l, m_{l+1}, \dots, m_N)}^2. \quad (7)$$

Using this method, all the partial derivatives can be calculated using classical post-processing, in polynomial time, of the same samples $\mathbf{m}_1, \dots, \mathbf{m}_S$ obtained from the quantum computer. Once the gradient has been esti-

ated, we can identify a new POVM with smaller expected variance than the previous one. We detail the gradient-based optimisation used in this work in App. B.

C. On-the-fly optimisation

An important aspect of the algorithm is that we do not need to first optimise the POVM (until it reaches a small-enough variance) before starting estimating the expected value of the observable. The intermediate POVMs used in the process are also IC, so they can be used for the estimation of $\langle \mathcal{O} \rangle$ as well. The strategy is to use the intermediate mean values obtained with every fixed choice of the POVM to calculate a weighted average. As we will show below, the latter is designed in a way that minimises the resulting variance in the overall estimation. The whole procedure can be carried out iteratively as the algorithm progresses, thus effectively making use of all measurement results obtained during the intermediate POVM optimisation steps for the reconstruction of $\langle \mathcal{O} \rangle$.

For the sake of clarity, let us first consider the situation in which we have used T different POVMs, each with $S_t, t \in [1, T]$ shots (i.e., statistical samples or repetitions of the measurement protocol), and we have produced T different estimations $\bar{\mathcal{O}}_t$ using Eq. (4) for each of the individual POVM choices. These estimations are, in fact, random variables with the same mean $\langle \mathcal{O} \rangle$ but different variance, $\text{Var}(\bar{\mathcal{O}}_t) = \text{Var}(\omega_{\mathbf{m}}^{(t)})/S_t$. Let us now define a new random variable, $\bar{\mathcal{O}}_2(\alpha) = \alpha \bar{\mathcal{O}}_2 + (1-\alpha) \bar{\mathcal{O}}_1$, where $\alpha \in (0, 1)$ is a parameter of our choice, whose mean is $\langle \bar{\mathcal{O}}_2(\alpha) \rangle = \alpha \langle \bar{\mathcal{O}}_2 \rangle + (1-\alpha) \langle \bar{\mathcal{O}}_1 \rangle = \langle \mathcal{O} \rangle$ for any value of α . Its variance is, however, a function of α . Since the $\bar{\mathcal{O}}_t$ are all independent random variables, we have $\text{Var}(\bar{\mathcal{O}}_2(\alpha)) = \alpha^2 \text{Var}(\bar{\mathcal{O}}_2) + (1-\alpha)^2 \text{Var}(\bar{\mathcal{O}}_1)$. While we do not know the actual values of the variances $\{\text{Var}(\bar{\mathcal{O}}_t)\}$, we can estimate them from the data; let us refer to the corresponding estimation of $\text{Var}(\bar{\mathcal{O}}_t)$ with the symbol \bar{V}_t . This allows us to estimate the variance of $\bar{\mathcal{O}}_2(\alpha)$ as $\text{Var}(\bar{\mathcal{O}}_2(\alpha)) \approx \alpha^2 \bar{V}_2 + (1-\alpha)^2 \bar{V}_1$. This quantity is minimised for $\alpha_{\text{opt}} = \bar{V}_1 / (\bar{V}_1 + \bar{V}_2)$, yielding an estimated variance $\text{Var}(\bar{\mathcal{O}}_2(\alpha_{\text{opt}})) \approx \bar{V}_2 \equiv \bar{V}_1 \bar{V}_2 / (\bar{V}_1 + \bar{V}_2)$, which is smaller than \bar{V}_1 and \bar{V}_2 . Thus, we have combined two different estimators of $\langle \mathcal{O} \rangle$ to produce a new one with smaller statistical error. Next, define the random variable $\bar{\mathcal{O}}_3(\alpha') = \alpha' \bar{\mathcal{O}}_3 + (1-\alpha') \bar{\mathcal{O}}_2(\alpha_{\text{opt}})$. Using the same arguments as before, the value of α' minimising the variance of $\bar{\mathcal{O}}_3$ is $\alpha'_{\text{opt}} = \bar{V}_2 / (\bar{V}_2 + \bar{V}_3)$, giving $\text{Var}(\bar{\mathcal{O}}_3(\alpha_{\text{opt}})) \approx \bar{V}_3 \equiv \bar{V}_2 \bar{V}_3 / (\bar{V}_2 + \bar{V}_3)$. The process can be iterated for the T estimators.

The above procedure can be recast in terms of an iterative algorithm as follows:

1. Initialize two variables, $\bar{\mathcal{O}}$ and \bar{V} , such that $\bar{\mathcal{O}}_1 \rightarrow \bar{\mathcal{O}}$ and $\bar{V}_1 \rightarrow \bar{V}$.

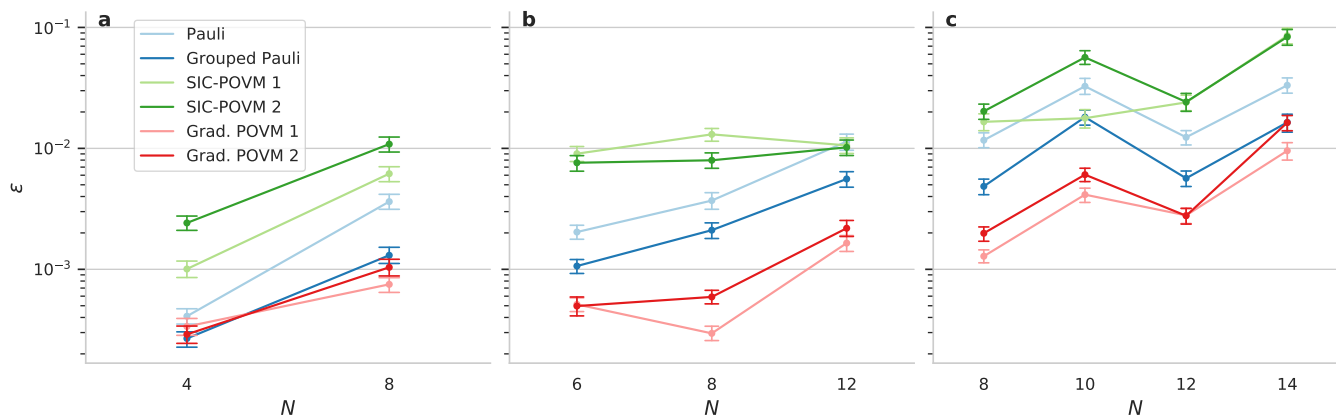


FIG. 3. **Performance of the adaptive measurement scheme.** Error in the estimation of the ground energy of the three molecules with the BK fermion-to-qubit mapping. Every plot compares the results for one molecule (H_2 (a), LiH (b), and H_2O (c)) with different measurement methods, with a total of $S = 10^6$ shots (for the Pauli and Grouped Pauli methods, we use the same number of shots $\lfloor 10^6/K \rfloor$ on every Pauli string, so the total number of shots is in fact $S = K \lfloor 10^6/K \rfloor$; this represents a deficit of at most 0.1% in the total number of shots in the examples considered). The ground state is approximated by optimising a VQE ansatz. The estimation error is the absolute difference between the simulation results and the exact value for the optimised ansatz. Points represent the average error over 100 realisations and the error bars show a 95% confidence interval obtained using bootstrapping. For H_2 , our algorithm offers little improvement, but the difference in performance becomes clearer with the other two molecules. Notice that the two initial POVMs yield slightly different results, with SIC-POVM 1 generally outperforming SIC-POVM 2. The results with other fermion-to-qubit mappings are available in App. E.

- At the end of each iteration $t \in (2, \dots, T)$ of the POVM optimisation, update them as $(\bar{\mathcal{O}}_t \bar{V} + \bar{\mathcal{O}} \bar{V}_t) / (\bar{V} + \bar{V}_t) \rightarrow \bar{\mathcal{O}}$ and $\bar{V} \bar{V}_t / (\bar{V} + \bar{V}_t) \rightarrow \bar{V}$.

At any point along the process, we have an estimated mean $\bar{\mathcal{O}}$ with estimated standard error $\bar{V}^{1/2}$ that minimises the overall error of the input data and can be easily updated with new ones. It is important to stress that this iterative mixing of the outcomes is unbiased, as we prove in App. D.

IV. NUMERICAL SIMULATIONS

In this section, we present the results of the numerical experiments that were run to test the feasibility and performance of our algorithm. Section IV A is aimed at illustrating the effect of the adaptive measurement. Section IV B presents a more in-depth analysis of the performance. Finally, in Sec. IV C we demonstrate an important feature of our approach: the IC data used for the estimation of the energy can be reused for other purposes.

A. Energy measurement learning

We start by measuring the ground-state energy of the H_2 , LiH and H_2O molecules. For the characterisation of each system, we use different number of molecular orbitals. The basis set used for H_2 is 6-31G [54–57] leading to 8 spin orbitals, while for the case of LiH and H_2O we used STO3G [58, 59] basis set leading to 12 and 14 spin

Method	Mapping		
	Parity	BK	JKMN
Pauli	6.0 ± 0.4	6.8 ± 0.5	6.2 ± 0.2
Grouped Pauli	5.5 ± 0.4	6.4 ± 0.5	5.7 ± 0.4
SIC-POVM 1	5.8 ± 0.7	5.7 ± 0.6	4.9 ± 0.5
SIC-POVM 2	5.4 ± 0.5	4.4 ± 0.4	4.7 ± 0.2
Grad. POVM 1	4.3 ± 0.6	4.6 ± 0.5	3.2 ± 0.3
Grad. POVM 2	4.0 ± 0.6	4.4 ± 0.5	3.4 ± 0.3

TABLE I. **Scaling exponents of the adaptive measurement scheme.** The points in Fig. 4, as well as their counterparts using other fermion-to-qubit mappings, are fitted to a function of the form $S_{\text{tar}} = aN^b$. The table contains the corresponding exponents. The exponent $b = 6$ of the Pauli method, as well as the mild reduction $b \approx 5.6$ offered by Grouped Pauli, are consistent with the values reported in Ref. [53] for other molecules. The POVM-based method without optimisation already outperforms these results, with $b \approx 4.8$ using the JKMN mapping [49]. The adaptive strategy results in a considerably smaller exponent $b \approx 3.3$.

orbitals, respectively. We use the Bravyi-Kitaev [48], the Jiang et al. [49], and the parity [60] mapping transformations. The latter has an intrinsic property, deriving from spin up and spin down electron conservation, that reduces the number of qubits required by two [60]. We also leverage different symmetries present in each system to reduce further the qubit count [60]. For the case of LiH and H_2O we also freeze the core orbitals allowing us to exclude another two spin orbitals from our calcula-

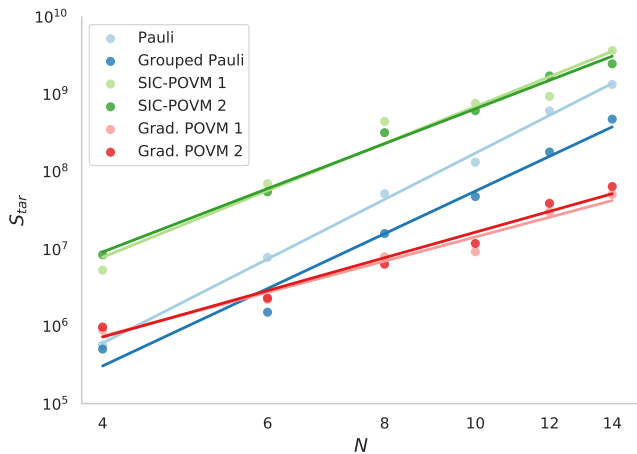


FIG. 4. **Scaling of the adaptive measurement scheme.** Number of shots S_{tar} required to achieve a target error of $\epsilon_{\text{tar}} = 0.5$ mHa for H chains as a function of the number of qubits N . The qubit Hamiltonian is obtained using the JKMN mapping. For each method and molecule, we use up to $S_{\text{lim}} \approx 10^6$ runs, as in Fig. 3. If the average error with S_{lim} shots, ϵ_{lim} , is still larger than ϵ_{tar} , we estimate the required number of shots needed to reach it by assuming a scaling $\epsilon \sim S^{-1/2}$, that is, we use $S_{\text{tar}} = S_{\text{lim}} \epsilon_{\text{lim}}^2 / \epsilon_{\text{tar}}^2$. While this procedure saves us considerable computing time, it also overestimates the number of measurements needed by our algorithm: indeed, our method’s convergence to ϵ_{tar} is faster than $\epsilon \sim S^{-1/2}$ unless it has already converged to the optimal POVM (see Fig. 2). Thus, these results are to be regarded as an upper bound to the total measurement cost of the learning POVM method. The curves depict least squares fits to the data with functions of the form $S_{\text{tar}} = aN^b$. The corresponding values of a and b for each method are reported in Table I. Notice that the values found for the Pauli and the Grouped Pauli methods are consistent with the ones reported in Ref. [41]. Moreover, the performance of our algorithm is similar to that of the state-of-the-art method proposed in Ref. [41], especially for the lower values of N , for which the overestimation of S_{tar} is less significant.

tion (refer to the table in Fig. 6 for more details on the Hamiltonians and qubit reductions considered). Each of these molecular Hamiltonians is mapped into qubits using one or more of the aforementioned techniques, hence producing several qubit Hamiltonians with varied number of qubits, which are then used to simulate the energy measurement process in a VQE experiment near convergence.

We proceed as follows. First, for each qubit Hamiltonian H , we numerically approximate the ground state with a hardware-efficient ansatz $|\psi(\vec{\theta})\rangle$ introduced in Ref. [15]. This generates a trial wavefunction by combining repetitive layers of single qubit R_y gates and entangling blocks composed of two-qubit operations (CNOT gates). The single qubit rotations are parametrised with a set of angles (aka variational parameters) that are iteratively updated, with the help of a classical optimisation routine, in order to minimise the energy expectation

value. Once we have the optimal parameters for which the variational form $|\psi(\vec{\theta}_{\text{opt}})\rangle$ approximates the ground-state wavefunction, we calculate the corresponding exact expected energy $\langle E \rangle = \langle \psi(\vec{\theta}_{\text{opt}}) | H | \psi(\vec{\theta}_{\text{opt}}) \rangle$. We then simulate different energy evaluation methods as a function of the number of state preparations (shots) $\bar{E}(S)$, and compute the corresponding errors $|\bar{E}(S) - \langle E \rangle|$. We also calculate the estimated statistical error for each approach, that is, the estimated error when the exact value $\langle E \rangle$ is not available (for the gradient-descent algorithm, this error is given by $\bar{V}^{1/2}$ as defined in Sec. III C). These quantities are depicted in Fig. 2 (a-c) for three selected examples.

The effect of the measurement learning results in the error decreasing faster than $S^{-1/2}$, especially for small S . This is a consequence of the fact that, after each batch of runs, the next POVM used in the sequence is in principle more efficient (i.e. leads to a smaller variance) than the previous one. Importantly, even if the starting efficiency is lower than that of other methods, our algorithm eventually takes over and reaches better accuracy at lower costs. Moreover, as we discuss in detail in the next subsection, even the use of Eq. (4) with the initial POVM without optimisation tends to give better performances than with the Pauli and the Grouped Pauli methods, as the size of the problem increases. The results also reveal that $\bar{V}^{1/2}$, as introduced in Sec. III C, gives the correct estimation of the statistical error in the evaluation of the energy [61].

The learning process is also illustrated in Fig. 2 (d-f), where we depict graphically the result of the optimisation in terms of a geometric representation of the effects akin to the Bloch sphere for single-qubit states (see App. A for details). We only include the results for one example in the paper, but the results for all the Hamiltonians analysed in this work, as well as their animated version, can be found here [52]. Interestingly, while the optimisation eventually converges and different realisations with the same initial condition lead to the same minimum (modulo small fluctuations), the two initial conditions considered here (see App. C) result in different optimal POVMs with slightly different performance. This suggests the potential existence of better initial conditions than the ones explored here. This subject will be considered in future work.

B. Performance and scaling

While the previous results illustrate the working principles of the algorithm with three molecular Hamiltonians, we now turn our attention towards the analysis of its performance. In Fig. 3 and in App. E, we collect the errors of similar estimations for several other Hamiltonians corresponding to the same molecules (under different qubit-reduction schemes) for a total number of measurements $S \approx 10^6$, from which it can be seen that

our algorithm is advantageous in almost all cases, and particularly for LiH and H₂O.

In order to study the performance of the algorithm for larger Hamiltonians, we analyse the number of measurements required to reach an accuracy of 0.5 mHa as a function of the number of qubits for hydrogen chains with increasing number of atoms. Due to limitations in computational power, we run our simulations for a limited number of measurements and extrapolate the total number required for such precision (see Fig. 4 for results using the JKMN mapping and caption for a detailed explanation). Even though this method overestimates the actual number of shots needed by our algorithm, we see a considerable improvement with respect to the Pauli and Grouped Pauli methods. Interestingly, the bare hybrid quantum-classical Monte Carlo method without optimisation, despite yielding higher errors for the small sizes considered here, also shows a more favourable scaling than the former methods.

To provide a more quantitative evaluation, we further fit each set of results into a function of the form aN^b . We report the corresponding values of the exponent b , also including those for other fermion-to-qubit mappings, in Table I. Notice that, while other mappings are added for completeness, the optimal performance of our algorithm is expected with the mapping from Ref. [49] (see Sec. III A), as confirmed by the results. Importantly, we can see that our method thus benefits from two improvements: the Monte Carlo approach results in a considerable reduction in the exponent, followed by a second scaling improvement stemming from the learning strategy. The result is an overall efficiency comparable to state-of-the-art methods [41, 49].

C. Exploiting informationally complete data

Further numerical experiments demonstrate that the IC data collected for the estimation of the energy can indeed be reused for other purposes. As explained in Sec. III A, the same IC outputs can be post-processed to calculate any expectation value of our choice, the only limitation being that, as it is reasonable to expect, the optimisation procedure targeting a particular observable may worsen the estimation of other specific quantities. In what follows, rather than focusing on particular additional observables, we consider an arguably more costly task: state tomography. More precisely, we address the reconstruction of all the k -qubit density operators in the system for all $k \leq K$. Reduced tomography has recently attracted some interest in the quantum information literature for diverse purposes [39–47, 51, 62].

We thus proceed in a similar manner as in the previous subsections. We approximate the ground states by training VQE ansätze and then estimate the energy using the adaptive algorithm. The resulting data is then used to reconstruct all the k -qubit reduced density matrices using likelihood maximisation. In particular, for every subset

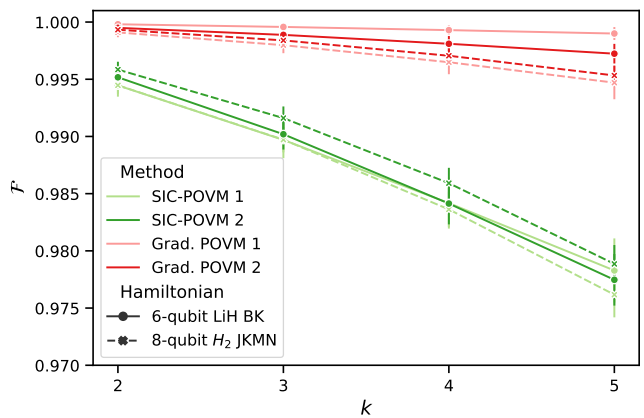


FIG. 5. **Reduced tomography from energy estimation data.** Average k -wise state fidelity \mathcal{F} , as a function of k , of the tomographic reconstructions obtained from the same data used for the estimation of the ground energy with four different methods. The green lines represent the results for the SIC-POVM 1 (lighter) and 2 (darker) (see App. C) without optimisation, while the red lines show the results for the gradient descent-optimised POVM 1 (lighter) and 2 (darker) data starting with the same initial condition. The style identifies the molecular Hamiltonian used. The solid lines are obtained for 6-qubit LiH with BK mapping; the dashed lines, 8-qubit H₂ with JKMN mapping. Every realisation involves a total of $S = 10^5$ measurements. The points indicate the average over 20 realisations, while the error bars the standard deviations.

of k qubits in the system, we marginalise the outcomes over the subset and then use the algorithm introduced in Ref. [63] to reconstruct the density operator. Since we must integrate IC data from T different POVMs in the likelihood maximisation procedure, we define a collective POVM with $T \times 4^N$ effects $\{\Xi_{(t,\mathbf{m})} = \prod_{\mathbf{m}}^{(t)} S_t / S, t \in [1, T]\}$, where the index t indicates the POVM optimisation step, S_t represents the number of measurements carried out in iteration t , and $S = \sum_t S_t$ [64]. Once a k -qubit density matrix ρ_{tomo} is reconstructed, we compute its fidelity $\mathcal{F}(\rho_{\text{tomo}}, \rho_{\text{exact}}) = \text{Tr}[\sqrt{\sqrt{\rho_{\text{tomo}}}\rho_{\text{exact}}\sqrt{\rho_{\text{tomo}}}}]^2$, with respect to the exact one ρ_{exact} (obtained by tracing out all other qubits in the trained VQE ansatz).

In Fig. 5, we show the resulting average k -wise fidelity for the ground states of two molecules, H₂ and LiH, as a function of k , with and without gradient-based POVM optimisation. The comparison between these two methods reveals an interesting effect of the optimisation. On one hand, the data resulting from the measurement learning strategy yields better reconstructions. On the other hand, the relative results of each method on the two problems are reversed: the non-optimised data give better or nearly equal k -wise fidelity for H₂ than for LiH, while the opposite is true for the gradient descent-based data. In all cases, however, fidelity decreases with k , as expected.

V. CONCLUSIONS

We have introduced an algorithm for efficient observable estimation that exploits informationally complete generalised quantum measurements integrating three important components: a hybrid quantum-classical Monte Carlo, an efficient method to navigate POVM space towards low-variance measurements, and a recipe to combine different estimations of the observable of interest. The result is a procedure in which an optimised measurement of an operator average is learnt in an adaptive fashion with no measurement overhead. Consequently, the overall measurement cost is drastically reduced with respect to the initial POVM considered. This is particularly interesting for real applications, considering that the initial SIC-POVMs used already offer a significant improvement over other widely used methods, such as Grouped Pauli. Importantly, the method does not require any exponentially scaling classical computations, although it involves a modest polynomial classical overhead.

We have illustrated the potential of the approach with several proof-of-principle numerical experiments by reconstructing the ground-state energies of several molecular Hamiltonians. Importantly, our simulations suggest that this adaptive method exhibits scaling performances comparable to the most efficient measurement reduction techniques in the current literature. While confirming this calls for a more thorough analysis and simulations, possibly including more general operators than molecular Hamiltonians, it is also important to point out that there is still substantial room for improvement in our algorithm, specially in the parametrisation of the POVMs and in the gradient descent-based update schedule.

Our algorithm also offers some other intrinsic advantages. Being completely agnostic to the nature of the qubit Hamiltonian, and not inspired by quantum chemistry but by quantum information alone, the proposed procedure may find interesting applications beyond VQE calculations. The method is also formally exact, as no approximations are made at any point, except for using the estimated variances as proxies of the actual ones. Moreover, the informationally complete data produced during the measurement process for a particular observable can in principle be reused to calculate many other properties of the underlying quantum state, including its tomographic reconstruction. We have provided evidence of the feasibility of this prospect by performing high-fidelity reduced state tomography with no additional measurements.

In this paper, we have only considered the task of estimating a given observable on a fixed quantum state. This typically represents a single step of, e.g., a VQE calculation. In perspective, one could however easily integrate our proposed method as a subroutine of the whole ansatz optimisation method. In such case, it might actually be helpful to use the optimal POVM from the previous VQE step, or a slight modification of it, as the starting

point of the measurement optimisation on the updated ansatz, given that the trial wavefunction should undergo relatively small changes between consecutive iterations. While this stands as a hypothesis for now, it might reduce the average number of steps required to adjust the POVM settings, hence leading to an even larger reduction in the measurement costs associated to the overall ansatz optimisation process.

Admittedly, our contribution presents a drawback: it requires twice as many qubits. However, it is important to discuss what this entails in practice. The ancillary qubits used for the implementation of the POVM remain in the ground state until the very end of the circuit execution. Hence, the algorithm introduced here does not require preparing $2N$ -qubit states behaving coherently during the whole computation, which would amplify the detrimental effect of decoherence. Instead, the additional N qubits should be regarded as nothing more than part of the measurement apparatus.

ACKNOWLEDGEMENTS

G.G.-P., M.A.C.R., B.S, and S.M. acknowledge financial support from the Academy of Finland via the Centre of Excellence program (Project no. 336814). The computer resources of the Finnish IT Center for Science (CSC) and the Finnish Grid and Cloud Infrastructure (FGCI) project (Finland) are acknowledged. G.G.-P. acknowledges support from the emmy.network foundation under the aegis of the Fondation de Luxembourg. IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. The current list of IBM trademarks is available at <https://www.ibm.com/legal/copytrade>.

Appendix A: POVM parametrisation and circuit implementation

As stated in the main text, the algorithm relies on parametrised, informationally complete POVMs implemented through the application of two-qubit unitaries with ancillary qubits, followed by projective measurements on the computational basis. To explain the parametrisation used in this work, it is easier to start by identifying the POVM characterising one such measurement when applying an arbitrary unitary gate U between some system qubit q in state ρ and an ancilla a in state $|0\rangle\langle 0|$. Since the two qubits are eventually measured projectively in the computational basis, there are four possible outcomes (b_q, b_a) with $b_q \in \{0, 1\}$ (and similarly for b_a). Each outcome occurs with probability $p_{(b_q, b_a)} = \langle b_q b_a | U \rho \otimes |0\rangle\langle 0| U^\dagger |b_q b_a\rangle$. Writing $U = \sum_{ijkl} u_{kl}^{ij} |ij\rangle\langle kl|$, this expression becomes $p_{(b_q, b_a)} =$

$\sum_{kk'} u_{k0}^{b_q b_a} (u_{k'0}^{b_q b_a})^* \langle k | \rho | k' \rangle = \text{Tr} [|\pi_{(b_q, b_a)}\rangle \langle \pi_{(b_q, b_a)} | \rho]$, where we have defined $|\pi_{(b_q, b_a)}\rangle = \sum_k (u_{k0}^{b_q b_a})^* |k\rangle$. Hence, the corresponding POVM is given by the set of effects $\{\Pi_i = |\pi_i\rangle \langle \pi_i|, i \in [0, 3]\}$, where we have relabelled the outcomes using $i = 2b_q + b_a$.

The previous calculation suggests our strategy for the POVM parametrisation: parametrise the unitary U , and compute the resulting POVM. The following observations are important. Firstly, not all the components $u_{kl}^{b_q b_a}$ are relevant for the measurement, as the initial state of the ancilla deems those with $l = 1$ irrelevant (provided that U is unitary). Secondly, global phases on $|\pi_i\rangle$ have no effect on the resulting operator Π_i , so we are free to set $u_{00}^{b_q b_a} \in \mathbb{R}$. Thirdly, $U^\dagger U = \mathbb{I}$ implies $\sum_{b_q b_a} (u_{k0}^{b_q b_a})^* u_{k'0}^{b_q b_a} = \delta_{kk'}$, that is, $u_{00}^{b_q b_a}$ and $u_{10}^{b_q b_a}$ are the components of two orthonormal vectors, which we may call \mathbf{u}_0 and \mathbf{u}_1 in what follows, in \mathbb{C}^4 . Before we proceed any further, let us count the total number of available degrees of freedom. On the one hand, we have four real numbers whose squares add up to one for \mathbf{u}_0 , which amounts to 3 degrees of freedom. For \mathbf{u}_1 , we have four complex numbers with three constraints (one for normalisation and two for the orthogonality with \mathbf{u}_0), which results in 5 degrees of freedom. In total, we need 8 parameters per system qubit.

Our parametrisation for single-qubit POVMs thus consists of 8 real numbers $\vec{x} = (x_0, \dots, x_7)$, with $x_i \in (0, 1), \forall i$ (in practice, we constrain the values further, see App. B). We start by using the first three of these to produce the set of angles $(\pi x_0, \pi x_1, 2\pi x_2)$, which identify (uniquely) a point on a 3-sphere \mathbb{S}^3 with unit radius embedded in \mathbb{R}^4 . The corresponding Euclidean coordinates in the embedding space are four real numbers whose squares add up to one, hence generating $u_{00}^{b_q b_a}$. Defining $u_{10}^{b_q b_a}$ from the other five parameters is slightly more involved. To guarantee that the vector \mathbf{u}_1 is orthogonal to \mathbf{u}_0 , we construct it as a linear combination of orthonormal vectors orthogonal to \mathbf{u}_0 , that is, $\mathbf{u}_1 = \sum_i z_i \mathbf{u}_i^\perp$; the orthonormal basis $\{\mathbf{u}_i^\perp\}$ can be found by means of the Gram-Schmidt orthonormalisation. The components z_i , which must also be normalised, are determined by the remaining parameters: once again, we define a list of angles $(\pi x_3, \dots, \pi x_6, 2\pi x_7)$ and calculate the Euclidean coordinates of the corresponding point in \mathbb{S}^5 . These six real numbers $\{r_i, i \in [0, 5]\}$ are then used to define three components $\{z_k = r_{2k} + ir_{2k+1}\}$. The result of this procedure is a vector $\mathbf{u}_1 \in \mathbb{C}^4$ whose components can be identified with $u_{10}^{b_q b_a}$.

Finally, we must find two more vectors $\mathbf{u}_2, \mathbf{u}_3 \in \mathbb{C}^4$ to complete the missing terms $u_{k1}^{b_q b_a}$ in the definition of the unitary. This can be done by using the Gram-Schmidt orthonormalisation once more. Once the unitary U is defined, we can not only calculate the corresponding set of effects $\{\Pi_i\}$, but also implement it in a given circuit. Indeed, the algorithms to find the circuit decomposition of unitary U are known and readily implemented

in Qiskit [65] (also, notice that any two-qubit gate can be decomposed in up to three CNOT gates).

Admittedly, this methodology is more complicated than simply parametrising arbitrary two-qubit gates U and then calculating the corresponding POVM. However, as discussed above, our procedure avoids the use of unnecessary or redundant parameters, which could make the POVM optimisation harder. Nevertheless, it is likely that other parametrisations, more suitable for the adaptive optimisation algorithm, exist. These refinements, as well as improving the gradient descent protocol (see App. B), will be the subject of future work.

To conclude this section, let us also explain the geometric representation of the POVMs in Fig. 2. Notice that the effects introduced above are rank-1, so they can be written as $\Pi_i = \gamma \tilde{\Pi}_i$, where $\gamma = \text{Tr}[\Pi_i]$ and $\tilde{\Pi}_i$ are single-qubit pure states (also, $\gamma \leq 1$). As a result, we can exploit the Bloch sphere representation and write $\Pi_i(\vec{r}) = \gamma(\mathbb{I} + \vec{q} \cdot \vec{\sigma})/2$ with $\vec{q} = (\text{Tr}[\tilde{\Pi}_i \sigma_x], \text{Tr}[\tilde{\Pi}_i \sigma_y], \text{Tr}[\tilde{\Pi}_i \sigma_z])$ and $|\vec{q}| = 1$. Thus, defining $\vec{r} = (\text{Tr}[\Pi_i \sigma_x], \text{Tr}[\Pi_i \sigma_y], \text{Tr}[\Pi_i \sigma_z]) = \gamma \vec{q}$, we have $\Pi_i(\vec{r}) = (|\vec{r}|\mathbb{I} + \vec{r} \cdot \vec{\sigma})/2$.

Appendix B: Gradient descent protocol

Along the measurement process, we iteratively update the POVM parameters as well as the number of shots per experiment. In particular, we gradually increase the number of shots in order to have more precise estimations of the second moment as the POVM parameters approach a minimum and, consequently, the gradient decreases in magnitude. In this section, we briefly outline the protocol used in our numerical experiments.

As explained in the main text, the POVM-based measurements allow us to estimate the gradient $\nabla_{\vec{x}} \langle \omega_{\mathbf{m}}^2 \rangle$ classically from the outcomes of an experiment run with the POVM corresponding to parameters \vec{x}_t , where t labels the iteration (for the finite-difference partial derivatives $\partial_{x_k} \langle \omega_{\mathbf{m}}^2 \rangle \approx (\langle \omega_{\mathbf{r}}^2 \rangle - \langle \omega_{\mathbf{m}}^2 \rangle)/h$, we use $h = 10^{-3}$). With these elements, we determine the POVM to be used in the $(t+1)$ -th iteration through $\vec{x}_{t+1} = \vec{x}_t - \nu \nabla_{\vec{x}} \langle \omega_{\mathbf{m}}^2 \rangle / \max(|\nabla_{\vec{x}} \langle \omega_{\mathbf{m}}^2 \rangle|)$, where $|\nabla_{\vec{x}} \langle \omega_{\mathbf{m}}^2 \rangle|$ is to be understood as the set of absolute values of the components of $\nabla_{\vec{x}} \langle \omega_{\mathbf{m}}^2 \rangle$. Hence, ν is the magnitude of the largest change, in absolute value, of the POVM parameters. It should also be mentioned that, to avoid numerical instabilities, we further constrain every parameter to be between $[\delta, 1 - \delta]$, with $\delta = 0.05$. We start our simulations with $S_1 = 1000$ shots, and we use $\nu = 0.05$. Every three iterations, we update $S_t + 1000 \rightarrow S_{t+1}$ and $\nu/1.2 \rightarrow \nu$. Hence, as the algorithm approaches the minimum, we obtain more precise estimations of the gradient (larger S_t) and we make smaller changes to the parameters (smaller ν).

This parameter updating schedule is rather heuristic and still leaves room for improvement. Designing a more

theory-driven approach, or using more sophisticated optimisation techniques, will be the subject of future work.

Appendix C: Symmetric IC-POVMs as initial measurements and correlated estimators

In the absence of prior knowledge about the state of the qubit register, it is desirable to use a so-called symmetric informationally complete POVM (SIC-POVM) on every system qubit. Symmetric here means that its single-qubit effects, when rescaled as $\tilde{\Pi}_i = 2\Pi_i$ yield a set of projectors $\{\tilde{\Pi}_i : \tilde{\Pi}_i^2 = \tilde{\Pi}_i\}$ fulfilling $\text{Tr}[\tilde{\Pi}_i\tilde{\Pi}_j] = (2\delta_{ij} + 1)/3, \forall i, j$. Hence, the projectors $\{\tilde{\Pi}_i\}$ form a regular tetrahedron in the Bloch sphere.

In this work, we have considered two different SIC-POVMs as initial conditions for the adaptive algorithm. The first one is the classic example of single-qubit SIC-POVM, defined in terms of the projectors $\{\tilde{\Pi}_i = |\tilde{\pi}_i\rangle\langle\tilde{\pi}_i| : |\tilde{\pi}_0\rangle = |0\rangle, |\tilde{\pi}_k\rangle = (|0\rangle + \sqrt{2}e^{i2\pi(k-1)/3}|1\rangle)/\sqrt{3}, k \in [1, 3]\}$. The second SIC-POVM used in this paper is the one introduced by Jiang et al. [49]. In order to use them in our algorithm, we must first find the parameters \vec{x} of each of them in the POVM space (see App. A). This can be done numerically; the resulting parameters are reported in the computer code accompanying this paper [66].

It is worth discussing some properties of this second SIC-POVM when used in our hybrid quantum-classical Monte Carlo algorithm, Eq. (4). In this case, all the $b_{km}^{(i)}$ exhibit the nice feature $b_{km}^{(i)} \in \{-\sqrt{3}, \sqrt{3}\}, \forall k > 0$ (for $k = 0$, these are equal to one, since the effects add up to identity). This, in turn, has interesting implications. Let us consider the statistical error in the estimation of the expectation value of an observable given by a single Pauli string $P_{\mathbf{k}}$ with weight l , that is, only l Pauli operators in $P_{\mathbf{k}}$ are different from identity. In this case, the variance of the Monte Carlo is given by $\text{Var}(\omega_{\mathbf{m}}) = 3^l - \langle P_{\mathbf{k}} \rangle^2 \leq 3^l$. Hence, if S measurements are performed, the variance of the estimator $\bar{P}_{\mathbf{k}}$ is $\text{Var}(\bar{P}_{\mathbf{k}}) \leq 3^l/S$. This is indeed consistent with Ref. [49].

While we can reuse the IC data from the quantum computer to calculate the expectation value of other Pauli strings $P_{\mathbf{k}'}$ with similar statistical error (assuming they have the same weight l), we must take into account that the resulting estimators $\bar{P}_{\mathbf{k}}$ and $\bar{P}_{\mathbf{k}'}$ can be correlated. In practice, this means that, if we are to use them to calculate the expectation value of an operator defined in terms of a linear combination of Pauli strings, $\mathcal{O} = \sum_{\mathbf{k}} c_{\mathbf{k}} P_{\mathbf{k}}$, the variance of the estimator $\bar{\mathcal{O}} = \sum_{\mathbf{k}} c_{\mathbf{k}} \bar{P}_{\mathbf{k}}$ depends on the potentially non-zero covariance between distinct terms, so we cannot assume that $\text{Var}(\bar{\mathcal{O}}) = \sum_{\mathbf{k}} |c_{\mathbf{k}}|^2 \text{Var}(\bar{P}_{\mathbf{m}})$.

The estimation based on the Monte Carlo method, Eq. (4), naturally takes into account these correlations when accounting for the statistical error of the approach, hence yielding the correct estimation. This is important

for two reasons. On the one hand, it provides a meaningful assessment of how far the algorithm is from reaching the required accuracy at any given point of its execution. On the other hand, since the Monte Carlo variance is the quantity that our adaptive strategy seeks to minimise, the algorithm presented here can potentially find POVMs for which the negative impact of these correlations on the estimated mean is reduced.

Appendix D: Sequential and one-step mixing equivalence

In this section we prove that the sequential estimation mixing presented in the main text is unbiased. To show this, let us first compute the unbiased one-step mixing estimation. Suppose that, after the different experiments have been run, we are left with a set of T estimated means $\{\bar{\mathcal{O}}_t\}$ and variances $\{\bar{V}_t\}$. We would like to find a set of weights $\{\alpha_t > 0\}$, with $\sum_t \alpha_t = 1$, that minimises the variance $\bar{V}_T = \sum_t \alpha_t^2 \bar{V}_t$ of $\bar{\mathcal{O}}_T = \sum_t \alpha_t \bar{\mathcal{O}}_t$. To do so, we can introduce a Lagrange multiplier λ and define

$$\mathcal{L} = \sum_{t=1}^T \alpha_t^2 \bar{V}_t - \lambda \left(\sum_t \alpha_t - 1 \right), \quad (\text{D1})$$

so that $\partial_{\lambda} \mathcal{L} = 0$ imposes the constraint $\sum_t \alpha_t = 1$. From $\partial_{\alpha_t} \mathcal{L} = 0$ we obtain $\alpha_t = \lambda \bar{\rho}_t / 2$, where we have defined $\bar{\rho}_t \equiv 1/\bar{V}_t$ to ease the presentation, as inverse variances will appear throughout. Using now $\sum_t \alpha_t = 1$ yields $\lambda = 2/\sum_i \bar{\rho}_i$ and $\alpha_t = \bar{\rho}_t / \sum_i \bar{\rho}_i$. Hence, we arrive at

$$\bar{\mathcal{O}}_T = \frac{\sum_{t=1}^T \bar{\mathcal{O}}_t \bar{\rho}_t}{\sum_{t=1}^T \bar{\rho}_t} \quad \text{and} \quad \bar{V}_T = \frac{1}{\sum_{t=1}^T \bar{\rho}_t}. \quad (\text{D2})$$

To assess the result of the sequential algorithm, notice that the recurrence $\bar{V}_t = \bar{V}_{t-1} \bar{V}_t / (\bar{V}_{t-1} + \bar{V}_t)$ in the second step is equivalent to $\bar{\rho}_t = \bar{\rho}_{t-1} + \bar{\rho}_t$, with $\bar{\rho}_t \equiv 1/\bar{V}_t$. Iterating, we obtain $\bar{\rho}_T = \sum_{t=1}^T \bar{\rho}_t$, which is the rightmost term in Eq. (D2). Similarly, the recurrence for the mean, $\bar{\mathcal{O}}_t = (\bar{\mathcal{O}}_t \bar{V}_{t-1} + \bar{\mathcal{O}}_{t-1} \bar{V}_t) / (\bar{V}_{t-1} + \bar{V}_t)$, reads $\bar{\mathcal{O}}_t = [\bar{\mathcal{O}}_t \bar{\rho}_t + \bar{\mathcal{O}}_{t-1} \bar{\rho}_{t-1}] / \bar{\rho}_t$. Iterating once again, we obtain the expression for $\bar{\mathcal{O}}_T$ in Eq. (D2). Hence, both estimations are equivalent.

Appendix E: Other experiments

In the main text we presented a selection of the numerical results obtained in this work allowing us to showcase the main features of our method. For completeness, in this Appendix we report all the results of the simulations obtained with the H_2 , LiH and H_2O molecules, using different combinations of fermion-to-qubit mappings

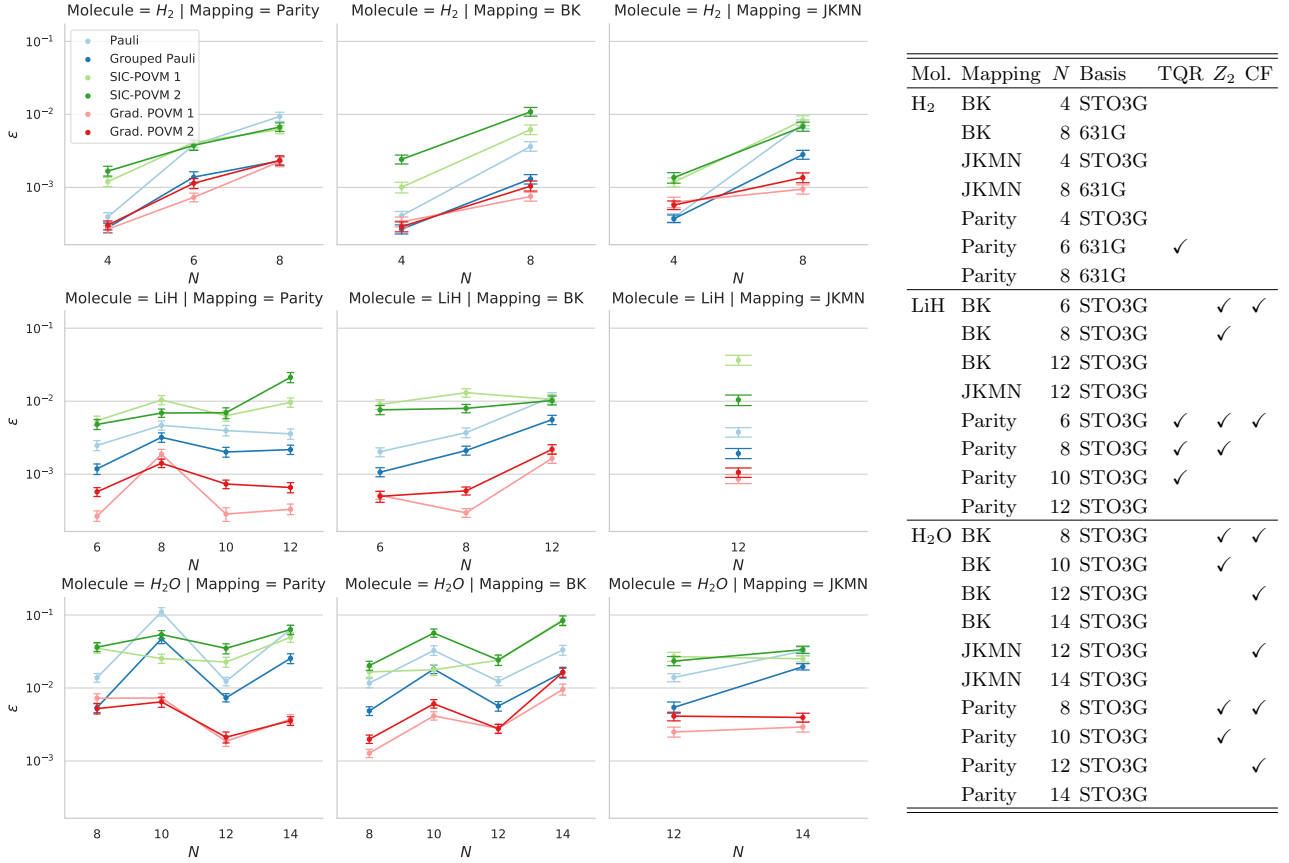


FIG. 6. (Left) Error in the estimation of the energy of an optimal VQE circuit for all the Hamiltonians reported in the table on the right. Every row compares the results for one molecule (H₂, LiH, and H₂O) with different measurement methods, with a total of $S = 10^6$ shots. Each column corresponds to a different mapping. Points represent the average error over 100 realisations and the error bars show a 95% confidence interval obtained using bootstrapping. (Right) A table of the various combinations of molecule, mapping, basis and qubit-reduction techniques considered, with the corresponding number of qubits N . TQR is the two-qubit reduction for the parity mapping, Z_2 refers to qubit reductions due to discrete symmetries [60] and CF stands for core freeze.

(namely parity, BK and JKMN) and qubit reduction techniques (namely two-qubit parity reduction, Z_2 symmetry and core freeze), where applicable. Figure 6 shows the average absolute error ε on the estimation of the en-

ergy with 10^6 shots, for the Hamiltonians reported in the table. The plots are in line with the figures in the main text, as our algorithm generally reaches smaller errors with the same total number of measurements.

[1] R. P. Feynman, “Simulating physics with computers,” *Int. J. Theor. Phys.* **21**, 467 (1982).
[2] Yu. I. Manin, “Computable and Noncomputable,” *Sov. Radio*, 13 (1980).
[3] I. M. Georgescu, S. Ashhab, and Franco Nori, “Quantum simulation,” *Rev. Mod. Phys.* **86**, 153–185 (2014).
[4] Seth Lloyd, “Universal quantum simulators,” *Science* **273**, 1073–1078 (1996).
[5] F. Tacchino, A. Chiesa, S. Carretta, and D. Gerace, “Quantum computers as universal quantum simulators: State-of-the-art and perspectives,” *Advanced Quantum Technologies* **3**, 1900052 (2020).
[6] A. Chiesa, F. Tacchino, M. Grossi, P. Santini, I. Tavernelli, D. Gerace, and S. Carretta, “Quantum hard-

ware simulating four-dimensional inelastic neutron scattering,” *Nat. Phys.* **15**, 455–459 (2019).
[7] Andrew A. Houck, Hakan E. Türeci, and Jens Koch, “On-chip quantum simulation with superconducting circuits,” *Nat. Phys.* **8**, 292–299 (2012).
[8] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z. X. Gong, and C. Monroe, “Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator,” *Nature* **551**, 601 (2017).
[9] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin, “Prob-

- ing many-body dynamics on a 51-atom quantum simulator,” *Nature* **551**, 579 (2017).
- [10] Ehud Altman, Kenneth R. Brown, Giuseppe Carleo, Lincoln D. Carr, Eugene Demler, Cheng Chin, Brian DeMarco, Sophia E. Economou, Mark A. Eriksson, Kai-Mei C. Fu, Markus Greiner, Kaden R.A. Hazzard, Randall G. Hulet, Alicia J. Kollár, Benjamin L. Lev, Mikhail D. Lukin, Ruichao Ma, Xiao Mi, Shashank Misra, Christopher Monroe, Kater Murch, Zaira Nazario, Kang-Kuen Ni, Andrew C. Potter, Pedram Roushan, Mark Saffman, Monika Schleier-Smith, Irfan Siddiqi, Raymond Simmonds, Meenakshi Singh, I.B. Spielman, Kristan Temme, David S. Weiss, Jelena Vučković, Vladan Vuletić, Jun Ye, and Martin Zwierlein, “Quantum simulators: Architectures and opportunities,” *PRX Quantum* **2**, 017003 (2021).
 - [11] John Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum* **2**, 79 (2018).
 - [12] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik, “Noisy intermediate-scale quantum (NISQ) algorithms,” (2021), [arXiv:2101.08448](https://arxiv.org/abs/2101.08448) [quant-ph].
 - [13] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, Abhinav Kandala, Antonio Mezzacapo, Peter Müller, Walter Riess, Gian Salis, John Smolin, Ivano Tavernelli, and Kristan Temme, “Quantum optimization using variational algorithms on near-term quantum devices,” *Quantum Sci. Technol.* **3**, 030503 (2018).
 - [14] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles, “Variational quantum algorithms,” (2020), [arXiv:2012.09265](https://arxiv.org/abs/2012.09265) [quant-ph].
 - [15] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature* **549**, 242–246 (2017).
 - [16] Panagiotis Kl. Barkoutsos, Jerome F. Gonthier, Igor Sokolov, Nikolaj Moll, Gian Salis, Andreas Fuhrer, Marc Ganzhorn, Daniel J. Egger, Matthias Troyer, Antonio Mezzacapo, Stefan Filipp, and Ivano Tavernelli, “Quantum algorithms for electronic structure calculations: Particle-hole hamiltonian and optimized wavefunction expansions,” *Phys. Rev. A* **98**, 022322 (2018).
 - [17] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan, “Quantum computational chemistry,” *Rev. Mod. Phys.* **92**, 015003 (2020).
 - [18] Pauline J. Ollitrault, Abhinav Kandala, Chun-Fu Chen, Panagiotis Kl. Barkoutsos, Antonio Mezzacapo, Marco Pistoia, Sarah Sheldon, Stefan Woerner, Jay M. Gambetta, and Ivano Tavernelli, “Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor,” *Phys. Rev. Res.* **2**, 043140 (2020).
 - [19] Panagiotis Kl. Barkoutsos, Fotios Gkritis, Pauline J. Ollitrault, Igor O. Sokolov, Stefan Woerner, and Ivano Tavernelli, “Quantum algorithm for alchemical optimization in material design,” *Chem. Sci.*, – (2021).
 - [20] Guglielmo Mazzola, Pauline J. Ollitrault, Panagiotis Kl. Barkoutsos, and Ivano Tavernelli, “Nonunitary operations for ground-state calculations in near-term quantum computers,” *Phys. Rev. Lett.* **123**, 130501 (2019).
 - [21] Zhenyu Cai, “Resource estimation for quantum variational simulations of the hubbard model,” *Phys. Rev. Appl.* **14**, 014059 (2020).
 - [22] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nat. Commun.* **5**, 4213 (2014).
 - [23] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik, “The theory of variational hybrid quantum-classical algorithms,” *New J. Phys.* **18**, 023023 (2016).
 - [24] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos, “Quantum chemistry calculations on a trapped-ion quantum simulator,” *Phys. Rev. X* **8**, 031022 (2018).
 - [25] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta, “Error mitigation extends the computational reach of a noisy quantum processor,” *Nature* **567**, 491–495 (2019).
 - [26] M. Ganzhorn, D.J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, “Gate-efficient simulation of molecular eigenstates on a quantum computer,” *Phys. Rev. Appl.* **11**, 044092 (2019).
 - [27] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications* **9**, 4812 (2018).
 - [28] Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall, “An adaptive variational algorithm for exact molecular simulations on a quantum computer,” *Nat. Commun.* **10**, 3007 (2019).
 - [29] I.O. Sokolov, P.K. Barkoutsos, P.J. Ollitrault, D Greenberg, J. Rice, M Pistoia, and I. Tavernelli, “Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?” *J. Chem. Phys.* **152**, 124107 (2020).
 - [30] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature Communications* **12**, 1791 (2021).
 - [31] Jarrod R. McClean, Ryan Babbush, Peter J. Love, and Alán Aspuru-Guzik, “Exploiting locality in quantum computation for quantum chemistry,” *J. Phys. Chem. Lett.* **5**, 4368–4380 (2014).
 - [32] Dave Wecker, Matthew B. Hastings, and Matthias Troyer, “Progress towards practical quantum variational algorithms,” *Phys. Rev. A* **92**, 042303 (2015).
 - [33] Ryan Babbush, Nathan Wiebe, Jarrod McClean, James McClain, Hartmut Neven, and Garnet Kin-Lic Chan, “Low-depth quantum simulation of materials,” *Phys. Rev. X* **8**, 011044 (2018).
 - [34] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo, “Neural-network quantum state tomography,” *Nat.*

- Phys.* **14**, 447–450 (2018).
- [35] Giacomo Torlai, Guglielmo Mazzola, Giuseppe Carleo, and Antonio Mezzacapo, “Precise measurement of quantum observables with neural-network estimators,” *Phys. Rev. Res.* **2**, 022060 (2020).
- [36] Hsin-Yuan Huang, Richard Kueng, and John Preskill, “Predicting many properties of a quantum system from very few measurements,” *Nat. Phys.* **16**, 1050–1057 (2020).
- [37] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo, “Measurements of quantum Hamiltonians with locally-biased classical shadows,” (2020), [arXiv:2006.15788 \[quant-ph\]](https://arxiv.org/abs/2006.15788).
- [38] Hsin-Yuan Huang, Richard Kueng, and John Preskill, “Efficient estimation of pauli observables by derandomization,” (2021), [arXiv:2103.07510 \[quant-ph\]](https://arxiv.org/abs/2103.07510).
- [39] Andrew Jena, Scott Genin, and Michele Mosca, “Pauli Partitioning with Respect to Gate Sets,” (2019), [arXiv:1907.07859 \[quant-ph\]](https://arxiv.org/abs/1907.07859).
- [40] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F. Izmaylov, “Measuring all compatible operators in one series of a single-qubit measurements using unitary transformations,” (2019), [arXiv:1907.09386 \[quant-ph\]](https://arxiv.org/abs/1907.09386).
- [41] William J. Huggins, Jarrod McClean, Nicholas Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush, “Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers,” (2019), [arXiv:1907.13117 \[quant-ph\]](https://arxiv.org/abs/1907.13117).
- [42] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T. Chong, “Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families,” (2019), [arXiv:1907.13623 \[quant-ph\]](https://arxiv.org/abs/1907.13623).
- [43] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley, “Efficient quantum measurement of Pauli operators,” (2019), [arXiv:1908.06942 \[quant-ph\]](https://arxiv.org/abs/1908.06942).
- [44] Andrew Zhao, Andrew Tranter, William M. Kirby, Shu Fay Ung, Akimasa Miyake, and Peter Love, “Measurement reduction in variational quantum algorithms,” (2019), [arXiv:1908.08067 \[quant-ph\]](https://arxiv.org/abs/1908.08067).
- [45] Marco Pains and Amir Kalev, “An approximate description of quantum states,” (2019), [arXiv:1910.10543 \[quant-ph\]](https://arxiv.org/abs/1910.10543).
- [46] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O’Brien, “Nearly optimal measurement scheduling for partial tomography of quantum states,” *Phys. Rev. X* **10**, 031064 (2020).
- [47] Jordan Cotler and Frank Wilczek, “Quantum overlapping tomography,” *Phys. Rev. Lett.* **124**, 100401 (2020).
- [48] Sergey B. Bravyi and Alexei Yu. Kitaev, “Fermionic quantum computation,” *Ann. Phys.* **298**, 210–226 (2002).
- [49] Zhang Jiang, Amir Kalev, Wojciech Mruzekiewicz, and Hartmut Neven, “Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning,” *Quantum* **4**, 276 (2020).
- [50] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała, “Simulating all quantum measurements using only projective measurements and postselection,” *Phys. Rev. A* **100**, 012351 (2019).
- [51] Guillermo García-Pérez, Oskari Kerppo, Matteo A. C. Rossi, and Sabrina Maniscalco, “Experimentally accessible non-separability criteria for multipartite entanglement detection,” in preparation (2021).
- [52] Animations of the evolution of the POVM effects available online at https://bosoko.github.io/povm_for_vqe_videos/.
- [53] Jérôme F. Gonthier, Maxwell D. Radin, Corneliu Buda, Eric J. Daskocil, Clena M. Abuan, and Jhonathan Romero, “Identifying challenges towards practical quantum advantage through resource estimation: the measurement roadblock in the variational quantum eigensolver,” (2020), [arXiv:2012.04001 \[quant-ph\]](https://arxiv.org/abs/2012.04001).
- [54] R Ditchfield, W Hehre, and John A. Pople, “Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules,” *J. Chem. Phys.* **54**, 724 (1971).
- [55] Warren J Hehre, Robert Ditchfield, and John A Pople, “Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules,” *J. Chem. Phys.* **56**, 2257–2261 (1972).
- [56] Praveen C Hariharan and John A Pople, “The influence of polarization functions on molecular orbital hydrogenation energies,” *Theor. Chim. Acta* **28**, 213–222 (1973).
- [57] PC Hariharan and John A. Pople, “Accuracy of AH_n equilibrium geometries by single determinant molecular orbital theory,” *Mol. Phys.* **27**, 209–214 (1974).
- [58] W. J. Hehre, R. F. Stewart, and J. A. Pople, “Self Consistent Molecular Orbital Methods. I. Use of Gaussian Expansions of Slater Type Atomic Orbitals,” *J. Chem. Phys.* **51**, 2657–2664 (1969).
- [59] John B. Collins, Paul von R. Schleyer, J. Stephen Binkley, and John A. Pople, “Self consistent molecular orbital methods. XVII. geometries and binding energies of second row molecules. a comparison of three basis sets,” *J. Chem. Phys.* **64**, 5142–5151 (1976).
- [60] Sergey Bravyi, Jay M. Gambetta, Antonio Mezzacapo, and Kristan Temme, “Tapering off qubits to simulate fermionic Hamiltonians,” (2017), [arXiv:1701.08213 \[quant-ph\]](https://arxiv.org/abs/1701.08213).
- [61] Notice that the small difference between the measured and the estimated errors is entirely due to the definition of error used, i.e. $\langle |\bar{E}(S) - \langle E \rangle| \rangle$. Indeed, using $\sqrt{\langle (\bar{E}(S) - \langle E \rangle)^2 \rangle}$ instead leads to a perfect overlap between both (not shown).
- [62] Guillermo García-Pérez, Matteo A. C. Rossi, Boris Sokolov, Elsi-Mari Borrelli, and Sabrina Maniscalco, “Pairwise tomography networks for many-body quantum systems,” *Phys. Rev. Res.* **2**, 023393 (2020).
- [63] Jaroslav Řeháček, Zdeněk Hradil, E. Knill, and A. I. Lvovsky, “Diluted maximum-likelihood algorithm for quantum tomography,” *Phys. Rev. A* **75**, 042108 (2007).
- [64] Notice that this definition is consistent with the number of shots S_t imposed at each iteration t : for any state ρ , the expected number of outcomes corresponding to effects $\{\Xi_{(t,m)}\}$ with a particular t , given a total number of shots $S = \sum_t S_t$, is $S \sum_m \text{Tr}[\rho \Xi_{(t,m)}] = S \text{Tr}[\rho \mathbb{I} S_t / S] = S_t$.
- [65] Héctor Abraham, Ismail Yunus Akhalwaya, Gadi Aleksandrowicz, Thomas Alexander, Gadi Alexandrowicz, Eli Arbel, Abraham Asfaw, Carlos Azaustre, Panagiotis Barkoutsos, George Barron, Luciano Bello, Yael Ben-Haim, Lev S. Bishop, Samuel Bosch, David Bucher, et al., “Qiskit: An open-source framework for quantum

computing,” (2019).

[66] The code and data will be made available upon publication.