Short-depth circuits for efficient expectation-value estimation

A. Roggero and A. Baroni

Phys. Rev. A **101**, 022328 — Published 24 February 2020

DOI: 10.1103/PhysRevA.101.022328

Short-depth circuits for efficient expectation value estimation

A. Roggero*
Institute for Nuclear Theory, University of Washington, Seattle, WA 98195, USA

A. Baroni†
Department of Physics and Astronomy University of South Carolina, 712 Main Street, Columbia, South Carolina 29208, USA

(Dated: January 29, 2020)

The evaluation of expectation values $Tr\left[\rho\mathcal{O}\right]$ for some pure state ρ and Hermitian operator $\mathcal{O}$ is of central importance in a variety of quantum algorithms. Near optimal techniques have been developed in the past and require a number of measurements N approaching the Heisenberg limit $N = O\left(1/\epsilon\right)$ as a function of target accuracy ϵ. The use of Quantum Phase Estimation requires however long circuit depths $C = O\left(1/\epsilon\right)$ making its implementation difficult on near term noisy devices. The more direct strategy of Operator Averaging is usually preferred as it can be performed using $N = O\left(1/\epsilon^2\right)$ measurements and no additional gates besides those needed for the state preparation. In this work we use a simple but realistic model to describe the bound state of a neutron and a proton (the deuteron) to show that the latter strategy can require an overly large number of measurement in order to achieve a prefixed relative target accuracy ϵ_r. We propose to overcome this problem using a single step of QPE and classical post-processing. This approach leads to a circuit depth $C = O\left(\epsilon^{\mu}\right)$ (with $\mu \geq 0$) and to a number of measurements $N = O\left(1/\epsilon^{2+\nu}\right)$ for $0 < \nu \leq 1$ and a much smaller prefactor. We provide detailed descriptions of two implementations of our strategy for $\nu = 1$ and $\nu \approx 0.5$ and derive appropriate conditions that a particular problem instance has to satisfy in order for our method to provide an advantage.

As we are approaching the era of noisy intermediate scale quantum devices (NISQ [1]) the growing interest in practical applications of quantum computing techniques has led to an increased focus on algorithms that are robust against errors and require a limited number of quantum resources (qubits and gates). A central component of many quantum computing algorithms, in particular for applications to quantum simulation [2, 3], is the estimation with error ϵ of the expectation value of an hermitian operator $\mathcal{O}$ on some quantum state described by the density matrix ρ of a system of n qubits:

$$\langle\mathcal{O}\rangle = Tr\left[\mathcal{O}\rho\right] = \langle\Psi|\mathcal{O}|\Psi\rangle\,, \tag{1}$$

where here and in the following we assume the qubits to be in a pure state described by $\rho = |\Psi\rangle\langle\Psi|$. Note that even though this task might not be the origin of a quantum speeed up (but there are cases where it is, see eg. [4]), it is nevertheless an important routine needed in more complicated algorithms achieving a quantum speed up trough other means. A common example is the calculation of the expectation value of the Hamiltonian (the energy) of a many-body system which can be used, together with the variational principle, to guide the preparation of low energy states in algorithms such as the Variational Quantum Eigensolver (VQE) [5–8]. For unitary operators we can evaluate the expectation value in Eq. (1) by simply measuring the state of the device in the appropriate basis. If the change of basis is expensive one can employ the Hadamard Test algorithm from [9] which only requires a single application of $\mathcal{O}$ conditional on an ancilla prepared in a superposition state.

In general however we are interested in evaluating $\langle\mathcal{O}\rangle$ for hermitian operators $\mathcal{O}$ (such as an Hamiltonian) and thus alternative strategies have to be devised. Optimal quantum algorithms have been discovered in the past (see eg. [10]), with efficiencies approaching the Heisenberg limit $N = O(1/\epsilon)$ for the number of times the experiment needs to be repeated in order to achieve some target additive error ϵ. This remarkable result is achieved by making effective use of both the Quantum Phase Estimation [11, 12] and Amplitude Amplification [13] algorithms, along with the observation that we can estimate Eq. (1) from the linear part of

$$\begin{aligned} Tr\left[e^{i\tau\mathcal{O}}\rho\right] &= \langle cos\left(\tau\mathcal{O}\right)\rangle + i\langle sin\left(\tau\mathcal{O}\right)\rangle \\ &= 1 + i\tau\langle\mathcal{O}\rangle + \cdots \end{aligned} \tag{2}$$

up to additive errors that vanish in the limit $\tau \to 0$.

The scheme requires $O(1/\epsilon)$ applications of a controlled version of the unitary operator $U_\tau = e^{i\tau\mathcal{O}}$ and it's inverse (for the QPE part) and $O(1/\epsilon)$ application of the state preparation unitary W (ie. $W|0\rangle = |\Psi\rangle$). Near term quantum devices will be characterized by a substantial level of noise [1] and this will in general prevent us from employing these algorithms even in situations where performing U_τ is simple (eg. for operators $\mathcal{O}$ diagonal in the computational basis) as the need to apply W multiple times will lead to an excessively large gate count.

In general, schemes that are robust to noise are therefore preferable for early applications and this has led to a proliferation of hybrid quantum-classical algorithms [14] for a variety of purposes ranging from quantum simula-

* roggero@uw.edu
† abaro008@odu.edu

tion [15], to approximate optimization [16] and quantum compiling [17]. Our work follows the same philosophy in that we will delegate a substantial computational effort to classical computing resources while at the same time leveraging the available capabilities of current quantum hardware.

In order to simplify the exposition of our work, we anticipate here some of the crucial results obtained in the sections below:

- the sample complexity of the standard Operator Averaging [18] method is in general

$$N_{OA} \sim \frac{1}{\epsilon_r^2}\frac{1}{\rho^2} \tag{3}$$

where ϵ_r is the target **relative** error and the ratio $\rho = |\langle\mathcal{O}\rangle|/\|\mathcal{O}\|$ quantifies the magnitude of the expectation value compared to the norm of the operator (cf. Eq. (15) and Eq. (16) for a tighter bound).

- using the approach based on a polynomial expansion of Eq. (2) up to order K that will be presented in detail in Sec II we can estimate the expectation value, in the special case when $|\Psi\rangle$ is an eigenvector, with sample complexity independent of ρ:

$$N_K \sim \frac{f(K)}{\epsilon_r^{2+1/K}} \tag{4}$$

where $f(K)$ is a quickly decaying function of the polynomial order K (cf. Eq. (28) and Eq. (33) for more details). Using these results our new strategy can prove advantageous for large enough target error resolution

$$\epsilon_r \geq \rho^{2K} f(K)^K \ . \tag{5}$$

Due to the fast decay of the function $f(K)$ we show in Sec. II (in particular in Fig. 1) that even using a low polynomial order $K = 2$ our strategy can be advantageous down to very small errors $\epsilon_r \approx 10^{-9}$

- we discuss also the more general case when $|\Psi\rangle$ is not an eigenvector and provide rigorous bounds on ρ for our scheme to be more efficient than Operator Averaging (see Eq. (31) and Eq. (32) and Sec. V A for more details).

The paper is organized as follows. In Sec. I we review the standard method commonly used to evaluate expectation values of Hermitian operators known as Operator Averaging [18] and in Sec. II we present a detailed discussion of our proposed method based on a single step of phase estimation for performing this task. We then proceed in Sec. III to apply our approach to a simple but challenging nuclear physics problem: the calculation of the deuteron binding energy with a realistic interaction. We have devoted Sec. IV to a more thorough exploration of the effect of noise on both methods and discuss more in depth why an ancilla-based scheme could provide potential benefits on near term noisy devices. As our method is not expected to be competitive in the asymptotically small error limit, in Sec. V we describe how to overcome some of the major implementation challenges in assessing if our scheme could prove advantageous for a particular problem instance and discuss the cost of implementing the necessary time evolution. Finally in Sec. VI we summarize the results and possible future directions of the present work.

I. OPERATOR AVERAGING

We discuss here the standard method commonly used in the literature to evaluate expectation values of Hermitian operators [5, 6] and start setting up the notation used throughout this work. As a first step, and without loss of generality, let us separate out the traceless component $\mathcal{O}_T$ from the observable

$$\mathcal{O} = \mathcal{O}_I + \mathcal{O}_T \quad \mathcal{O}_I \equiv \alpha_0 \mathbb{1} \quad Tr[\mathcal{O}_T] = 0 \tag{6}$$

for some $\alpha_0 \in \mathbb{R}$. The idea behind the Operator Averaging (OA) approach, first proposed in [5], is to exploit a decomposition of $\mathcal{O}_T$ of the following form

$$\mathcal{O}_T = \sum_{k=1}^{L} \alpha_k U_k = \sum_{k=1}^{L} \alpha_k \left[e^{i\theta_k} P_k\right] \quad , \quad a_k > 0 \ \forall k \tag{7}$$

where $P_k \in \{\mathbb{1}, X, Y, Z\}^{\otimes n}$ are tensor products of Pauli matrices and the angles θ_k are introduced in order to keep the coefficient α_k positive. The expectation value of each of the terms in the expansion can be estimated with $O(1)$ circuit depth by directly measuring the corresponding Pauli operator on the quantum hardware to get the finite sample estimators $\widehat{P_k}$ (here and in the following we will use a wide hat to indicate sample estimators) and combine them to form

$$\widehat{O_A} = \alpha_0 + \sum_{k=1}^{L} \alpha_k \widehat{U_k} = \alpha_0 + \sum_{k=1}^{L} \alpha_k e^{i\theta_k} \widehat{P_k} \ , \tag{8}$$

which converges to $\langle\mathcal{O}\rangle$ in the infinite measurement limit. In particular, if we perform M measurement for every one of the L terms in Eq. (8), the variance of this estimator is given simply by

$$Var[\widehat{O_A}] = \sum_{k=1}^{L} \alpha_k^2 Var[\widehat{P_k}] = \sum_{k=1}^{L} \alpha_k^2 \frac{1 - \widehat{P_k}^2}{M} \ , \tag{9}$$

due to the independence of the separate measurements. Note that this implies that in general $Var[\widehat{O_A}] \neq Var[\mathcal{O}]$ and in fact Eq. (9) may be large even for eigenstates of $\mathcal{O}$. Using Eq. (9) we can estimate the total number of measurement $N_{tot} = LM$ required to evaluate $\widehat{\mathcal{O}_A}$ with

precision ϵ to be

$$N_{tot} = \frac{L}{\epsilon^2}\sum_{k=1}^{L}\alpha_k^2\left[1-\widehat{P_k}^2\right] \leq \frac{L\|\overline{\mathcal{O}_T}\|_2^2}{\epsilon^2}\,, \qquad (10)$$

where we defined

$$\|\overline{\mathcal{O}_T}\|_q \equiv \left(\sum_{k=1}^{L}\alpha_k^q\right)^{1/q} \quad \text{for} \quad q\geq 1\,. \qquad (11)$$

Different strategies have been proposed in the literature to reduce the scaling of operator averaging shown in Eq. (10). For instance, in situations like quantum chemistry where a large number of coefficients α_k have possibly a small magnitude, one can use efficient truncation schemes [6, 18] to improve the performance considerably. A complementary approach is to group the L terms into G groups of operators which can be measured together in a single experiment [6, 7, 19], if the newly introduced correlations are not too large this approach can allow again a great reduction in the number of measurements. In general however we still expect the number of groups G to scale with system size (for quantum chemistry applications see eg. [18]). We could also choose the number of measurements to be performed for the k-th term to depend on the magnitude of the expansion coefficients α_k as proposed in [20]:

$$M_k \propto \frac{\alpha_k}{\|\overline{\mathcal{O}_T}\|_1}\,, \qquad (12)$$

this leads in turn to the estimate

$$N'_{tot} = \frac{\|\overline{\mathcal{O}_T}\|_1}{\epsilon^2}\sum_{k=1}^{L}\alpha_k\left[1-\widehat{P_k}^2\right] \qquad (13)$$

and in this way obtaining a better bound for N_{tot} by a factor $\|\overline{\mathcal{O}_T}\|_1^2/(L\|\overline{\mathcal{O}_T}\|_2^2) \leq 1$. For our nuclear physics problem this strategy produced actually a small increase in the cost since the magnitude of the variance of the k-th term does not necessarily correlate with the magnitude of the coefficient α_k. By also using the asymptotic improvement from this adaptive variant, we will consider

$$N_A(\epsilon) = \frac{\|\overline{\mathcal{O}_T}\|_1^2}{\epsilon^2} \lesssim \frac{L\|\overline{\mathcal{O}_T}\|_2^2}{\epsilon^2} \qquad (14)$$

as an estimate for the number of repetitions needed to obtain a precision ϵ with operator averaging.

As expected we find that the estimator of Eq. (8) shows in all cases the usual shot noise behavior for small errors $N_{tot} = O(1/\epsilon^2)$ but its explicit dependence on the operator norm of $\mathcal{O}_T$ (which is a lower bound of $\|\overline{\mathcal{O}_T}\|_1$) can be unfavorable when the expectation value $\langle\mathcal{O}\rangle$ becomes too small. Introducing the ratio

$$R_O = \frac{|\langle\mathcal{O}\rangle|}{\|\overline{\mathcal{O}_T}\|_1} \leq \frac{\|\overline{\mathcal{O}}\|_1}{\|\overline{\mathcal{O}_T}\|_1} \equiv R_O^{max}\,, \qquad (15)$$

we can express the number of shots N_{tot} in terms of the relative error $\epsilon_r = \epsilon/|\langle\mathcal{O}\rangle|$ as

$$N_A(\epsilon_r) = \frac{1}{\epsilon_r^2 R_O^2} \geq \frac{1}{\epsilon_r^2}\left(\frac{\lambda_{max}-\alpha_0}{|\langle\mathcal{O}\rangle|}\right)^2\,, \qquad (16)$$

which makes explicit the quadratic dependence of the classical effort (the number of repetitions) with the ratio between the largest eigenvalue λ_{max} of the target operator $\mathcal{O}$ and its expectation value in the state $|\Psi\rangle$.

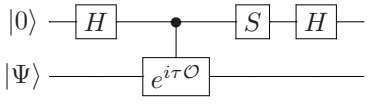
The scheme we propose in the next section can be advantageous whenever the ratio R_O becomes excessively small by providing a scheme with N_{tot} independent of R_O in the special case of eigenvalue estimation and possibly well performing in general (see condition Eq. (31)). This could be important in a large system when $\|\mathcal{O}_T\|$ grows with the number of qubits (indeed for applications in quantum simulation we expect $\|\mathcal{O}_T\| = O\left(poly(n)\right)$, see e.g. [18])) or simply because the expectation value we are after is much smaller than the largest eigenvalue of $\mathcal{O}_T$ like for the ground state of the deuteron with hard-core potentials [21] studied in Sec. III.

II. EXPECTATION VALUES FROM SINGLE STEP PHASE ESTIMATION

The widespread use of the direct algorithm described in the previous section comes from its appealing property of being able to minimize quantum resources since no additional quantum operation is required. This is especially important for NISQ era devices where coherence time and noise will limit the attainable circuit depth [1]. As mentioned in the introduction, algorithms based on full fledged quantum phase estimation, like the one described in [10], will possibly allow to approach Heisenberg limited scaling $\epsilon = O(1/N)$ as a function of the number of measurements N. The price to pay for this is having a circuit depth C that scales as $O(1/\epsilon)$ making its actual implementation challenging on noisy devices (characterized by a circuit depth $C \lesssim 100$ [1]). In this section we show how to effectively use a single step of time evolution to obtain a significant reduction in the total number of measurements respect to the operator averaging method, while enjoying short circuit depths $C_{sQPE} = O(\epsilon^\alpha)$ with $\alpha \geq 0$. Similarly to the approach used in Ref. [10], we will use the small time expansion of the imaginary part of the expectation value of the unitary time evolution $U_\tau = e^{i\tau\mathcal{O}}$ on the state $|\Psi\rangle$ (cf. Eq. (2))

$$\langle sin\left(\tau\mathcal{O}\right)\rangle = \tau\langle\mathcal{O}\rangle - \frac{\tau^3}{6}\langle\mathcal{O}^3\rangle + O\left(\tau^5\|\overline{\mathcal{O}}\|_1^5\right)\,, \qquad (17)$$

to extract $\langle\mathcal{O}\rangle$. In particular for any Hermitian operator $\mathcal{O}$ we consider the following standard circuit (cf. [9, 11]):



(18)

The imaginary part of $Tr\left[e^{i\tau\mathcal{O}}\rho\right]$ can be extracted by measuring the ancilla along the z axis

$$\begin{aligned}\langle Z_a\rangle(\tau) &= p_0(\tau) - p_1(\tau) \\ &= -\langle\Psi|sin\left(\tau\mathcal{O}\right)|\Psi\rangle\,,\end{aligned} \tag{19}$$

where p_0 (p_1) are the probabilities of measuring the ancilla in the $|0\rangle$ ($|1\rangle$) state. In practice we need to estimate the expectation value in Eq. (19) by performing N independent measurements and computing their average $\widehat{Z_a}(\tau)$. For ease of use, from here on we will call the process of extracting $\langle\mathcal{O}\rangle$ from a polynomial fit of a set of estimates of Eq. (19) for different times τ as sQPE (i.e. single step quantum phase estimation).

It is important to note now that, due to the presence of a bias for finite values of τ (coming from the necessity to truncate the series expansion in Eq. (17)), sQPE has a worse asymptotic scaling, given by $N = O\left(\epsilon^{-(2+\kappa)}\right)$ with $\kappa > 0$, compared to the direct method. Despite this loss of efficiency in the asymptotic limit $\epsilon \to 0$, we will show in the next subsections that the constant factors can be very small when the ratio $R_O \ll 1$, provided certain conditions are met (see Eq. (31) and Sec. V A). For instance, in the situation when we prepare an eigenstate of $\mathcal{O}$ and we want to estimate it's eigenvalue λ_Ψ, a bound $R_O \leq 1\%$ is sufficient to guarantee an advantage for the simplest possible version of sQPE (see Sec. II B) up to a respectably small relative error $\epsilon_r = O(10^{-5})$.

We next describe the general high-order sQPE in an idealized setting and in the following two subsection we propose practical implementations for the two lowest order sQPE algorithms where the truncation of Eq. (17) occurs at either the linear (Sec. II B) or the cubic term (Sec. II C). First we will focus our attention on the classical resources (the number of measurements that needs to be performed) and then (Sec. V B) we will discuss the cost associated with the implementation of the (controlled) unitary time evolution $U_\tau = e^{i\tau\mathcal{O}}$ in Eq. (18). We can already anticipate that the gate count will be low since the total evolution time τ has to be kept small enough to minimize the effect of the bias coming from neglecting higher-order terms in the expansion Eq. (17).

A. General scaling

Let's start by considering the idealized case where we know the coefficients $m_k = \langle\mathcal{O}^{2k+1}\rangle$ for $k = 1, \ldots, K$ (in practice the need of estimating these will lead to a sub-optimal algorithm). We can now use the Taylor expansion in τ of Eq. (19) to construct the following biased estimator for the expectation value

$$\mathcal{O}_K(\tau) = -\frac{1}{\tau}\left(\langle Z\rangle_a(\tau) + \sum_{k=1}^{K-1}\tau^{2k+1}\frac{(-1)^k m_k}{(2k+1)!}\right)\,. \tag{20}$$

Here the bias comes from neglecting higher order terms with $k \geq K$ in the expansion and can be written as

$$B_K(\tau) = \mathcal{O}_K(\tau) - \langle\mathcal{O}\rangle = -\sum_{k=K}^{\infty}\tau^{2k}\frac{(-1)^k m_k}{(2k+1)!}\,, \tag{21}$$

while its magnitude can be bounded from above using Lagranges Remainder theorem

$$|B_K(\tau)| \leq \tau^{2K}\frac{|m_K|}{(2K+1)!}\,. \tag{22}$$

Due to the presence of this bias for any finite value of τ we choose to characterize the deviations of our estimator to the exact expectation value using the Mean Squared Error (MSE) defined as

$$\epsilon_M^2(\tau, K) = Var\left[\widehat{\mathcal{O}_K}(\tau)\right] + B_K(\tau)^2\,, \tag{23}$$

where, as before, we denote with $\widehat{\mathcal{O}_K}(\tau)$ a finite population estimator of Eq. (20). The expected total number of measurements required to achieve a final precision (meaning MSE) target ϵ can then be estimated as shown in Sec. I. In particular, for any $\epsilon > B_K(\tau)$ we have

$$N_{tot} = \frac{1}{\tau^2}\frac{1-\widehat{Z_a}(\tau)^2}{\epsilon^2 - B_K(\tau)^2}\,, \tag{24}$$

where we used $Var[\widehat{\mathcal{O}_K}(\tau)] = (1-\widehat{Z_a}(\tau)^2)/(N\tau^2)$ with N the size of the population used to estimate $\widehat{Z_a}(\tau)$. This estimate is minimized with the choice $\tau = \tau_{\rm opt}$ with

$$\tau_{\rm opt} = \left(\frac{(2K+1)!}{\sqrt{2K+1}}\frac{\epsilon}{|m_K|}\right)^{\frac{1}{2K}}\,, \tag{25}$$

which leads to the following estimate for the total number of measurements needed for sQPE at order K

$$N_{sQPE}(K,\epsilon) = \frac{|m_K|^{1/K}}{\epsilon^{2+1/K}}f(K)\,, \tag{26}$$

where we have defined

$$f(K) = \frac{2K+1}{2K}\left(\frac{\sqrt{2K+1}}{(2K+1)!}\right)^{\frac{1}{K}}\,, \tag{27}$$

and as promised the shot noise regime $N_{sQPE} \sim O(1/\epsilon^2)$ is recovered only asymptotically for large values of K. As anticipated before, for a sufficiently small error we expect the Operator Averaging method of Sec. I to outperform the scheme presented here. The advantage of sQPE is in the possibility of having a much weaker dependence of the expectation value to norm ratio R_O. To see this let us first rewrite Eq. (26) using the relative error

$$N_{sQPE}(K,\epsilon_r) = \frac{f(K)}{\epsilon_r^{2+1/K}}\frac{|\langle\mathcal{O}^{2K+1}\rangle|^{1/K}}{|\langle\mathcal{O}\rangle|^{2+1/K}}\,, \tag{28}$$

where we simply used the definition of m_k. Consider now the special case of eigenvalue estimation where $|\Psi\rangle$ is an eigenvector of $\mathcal{O}$ with eigenvalue λ_Ψ that we want to compute. In this limit the ratio of expectation values on the right-hand side of Eq. (28) is just 1 and the total number of measurement required by sQPE is completely independent of the system considered. In the more general case where $|\Psi\rangle$ is not a single eigenvector we will need an additional ingredient to assess the performance of sQPE: a tight upper bound on the bias in Eq. (22) or equivalently on the moment m_K. For this purpose let us first introduce Γ_K as an upperbound of the following ratio

$$\frac{|m_K|}{\|\overline{\mathcal{O}}\|_1^{2K+1}} = \frac{|\langle\mathcal{O}^{2K+1}\rangle|}{\|\overline{\mathcal{O}}\|_1^{2K+1}} \leq \Gamma_K \leq 1 \,, \tag{29}$$

where we remind that $\|\overline{\mathcal{O}}\|_1 = |\alpha_0| + \|\overline{\mathcal{O}_T}\|_1$. We can now use Γ_K to bound the number of measurements as

$$N_{sQPE}(K,\epsilon_r) \leq \frac{f(K)}{\epsilon_r^{2+1/K}} \frac{\Gamma_K^{1/K}}{R_O^{2+1/K}} \left(\frac{\|\overline{\mathcal{O}}\|_1}{\|\overline{\mathcal{O}_T}\|_1}\right)^{2+1/K} \,, \tag{30}$$

and this could be smaller than Eq. (16) when $\Gamma_K \ll 1$. To be more quantitative, the sQPE estimator Eq. (20) will become efficient when $N_{sQPE}(K,\epsilon_r) \leq N_A(\epsilon_r)$ at the desired relative accuracy ϵ_r. Using Eq. (16) and Eq. (28), this condition can be written equivalently as

$$R_O \geq \frac{f(K)^K}{\epsilon_r} \frac{|\langle\mathcal{O}^{2K+1}\rangle|}{\|\overline{\mathcal{O}_T}\|_1^{2K+1}} \,, \tag{31}$$

which can be turned in the following sufficient condition

$$R_O \geq \frac{f(K)^K}{\epsilon_r} \Gamma_K \left(1 + \frac{|\alpha_0|}{\|\overline{\mathcal{O}_T}\|_1}\right)^{2K+1} \,, \tag{32}$$

where we reintroduced the upperbound Γ_K described above and wrote explicitly the dependence of the coefficient α_0 from Eq. (6). We will discuss in detail how to check if the condition in Eq. (31) is satisfied in practical application in Sec. V A, while for now we focus on the simpler application of eigenvalue estimation mentioned above. In this case the right hand side of Eq. (31) takes a simple form and the full inequality can be written as

$$\frac{|\lambda_\Psi|}{\|\overline{\mathcal{O}_T}\|_1} \leq \frac{\epsilon_r^{\frac{1}{2K}}}{\sqrt{f(K)}} \,. \tag{33}$$

We now need only a reasonably tight upperbound λ_u of $|\lambda_\Psi|$ to judge when the condition in Eq. (33) is satisfied. This requirement is rather loose in practice since the left-hand side is bounded from above by $\lambda_{max}/(\lambda_{max}-\alpha_0) \approx O(1)$ and the right hand side grows quickly as a function of K. In order to visualize this effect, we plot in Fig. 1 the minimum relative error ϵ_r that can be achieved for a fixed value of the eigenvalue ratio R_O using the condition from Eq. (33). As this will depend on the chosen order

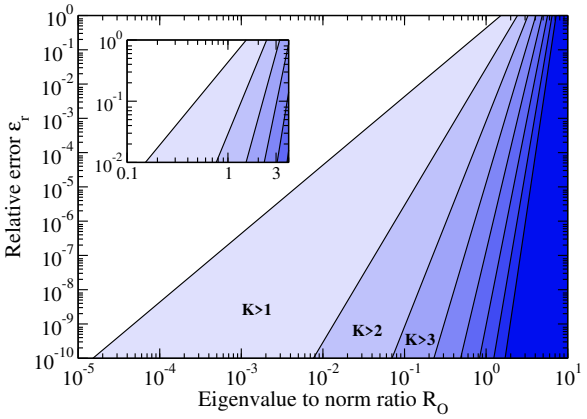


FIG. 1. The solid black curves indicate the set of points where Eq. (33) is satisfied with the equality for different choices of the sQPE order K with an ordering from top left to bottom right: the first solid line refers to $K = 1$ while the second is for $K = 2$ and the shaded region in between indicates the region where the linear method is no longer sufficient and we need to employ sQPE with $K > 1$. The innermost (darkest) region is bordered by the $K = 8$ line. With the inset we are zooming in the top-right corner of the main plot.

K, in the figure we report the boundaries starting from $K = 1$ in the top left corner up to $K = 8$ in the bottom right. In particular this means that to achieve lower error rates than the first solid curve we will need to use sQPE with $K > 1$ in order to ensure the condition in Eq. (33) to be satisfied (i.e. that sQPE could provide an advantage over the Operator Averaging method). Conversely the darkest and innermost region in Fig. 1 is accessible only for $K > 8$.

We can deduce a number of interesting conclusions from this figure. For instance, we can see that if we can place an upperbound on R_O smaller than $\approx 10^{-2}$, then sQPE with $K = 2$ will be more efficient than Operator Averaging down to extremely small relative errors $\epsilon_r \approx 10^{-9}$. This is a major improvement from the limit $\epsilon_r \approx 10^{-4}$ achievable with the linear method of Sec. II B. Furthermore, as we can clearly see in the inset, for target relative error at the 1% level, the linear method can be used effectively up to $R_O \approx 0.1$ while by increasing the order to cubic (i.e. $K = 2$) we can push this up to $R_O \approx 0.8$. Given these observations and the increasing difficulty in implementing sQPE efficiently for large K, it's likely that for many practical situations sQPE with $K = 1, 2$ will be sufficient to achieve a substantial speedup in terms of the number of measurements required to estimate the eigenvalue λ_Ψ. We will provide a similar scaling analysis for the more general problem of expectation value estimation in Sec. V A (see Fig. 11).

It is now time to come back to the problem of performing the polynomial fit needed in Eq. (20) in the realistic case where we do not know the higher-order coefficients m_k. For the last coefficient with $k = K$ we can easily use a reasonably tight upperbound Γ_K to manage the

influence of the bias in Eq. (22) as we did before. This allows for a complete algorithm in the simplest case $K = 1$ achieving the scaling reported in Eq. (30). In higher order algorithms we need to estimate the higher order contributions for $k < K$ by collecting data at different values of the time-step τ and performing a non-linear fit. If no other information is available, we will need at least K values of τ to properly perform the reconstruction (e.g. using the expansion proposed in [10]).

In the following sections we describe an implementation of the simple $K = 1$ (linear) algorithm and a more efficient scheme that adaptively finds the optimal pair (τ_a, τ_b) of time parameters for the $K = 2$ (cubic) case.

B. Linear Algorithm

The simplest case is where we neglect the cubic terms in the expansion of the *sin* so that our estimator in Eq. (20) becomes

$$\mathcal{O}_1(\tau) = \frac{1}{\tau}\langle\Psi|sin\,(\tau H)\,|\Psi\rangle = -\frac{1}{\tau}\langle Z\rangle_a(\delta)\ . \tag{34}$$

In the linear case the optimal time-step of Eq. (25) is

$$\tau_{opt} = \sqrt{\frac{6}{\sqrt{3}}\frac{\epsilon}{|\langle\mathcal{O}^3\rangle|}}\ , \tag{35}$$

and using the result from Eq. (28), we can estimate the number of measurements needed as

$$N_{sQPE}(1,\epsilon_r) = \frac{1}{\epsilon_r^3}\frac{\sqrt{3}}{4}\left|\frac{\langle\mathcal{O}^3\rangle}{\langle\mathcal{O}\rangle^3}\right| \longrightarrow \frac{1}{\epsilon_r^3}\frac{\sqrt{3}}{4}\ , \tag{36}$$

where the limit on the right hand side holds when we prepare $|\Psi\rangle$ in any eigenvector of $\mathcal{O}$. Note that in this limit the resources required for a given target relative accuracy ϵ_r are completely independent on the chosen operator $\mathcal{O}$ or even eigenvector $|\Psi\rangle$. For instance at the 1% level we have

$$N_{sQPE}(1,\epsilon_r = 0.01) = \frac{\sqrt{3}}{4}\times 10^6 \approx 4.3\times 10^5\ . \tag{37}$$

However, in order to get an advantage with sQPE for this situation, the inequality of Eq. (33) has to be satisfied, and therefore the advantage might still depend on the details of the system.

The problem now is that in general situations we will not be able to calculate τ_{opt} without at least an approximate estimate for the wanted expectation value $\langle\mathcal{O}\rangle$ since a good bound Γ_K on the bias is in fact not sufficient. Even in the simpler case of eigenvalue estimation we can only use an upperbound λ_u for the absolute value of the eigenvalue to compute the approximation

$$\widetilde{\tau}_{opt} = \sqrt{\frac{6}{\sqrt{3}}\frac{\epsilon_r}{\lambda_u^2}} \leq \tau_{opt}\ , \tag{38}$$

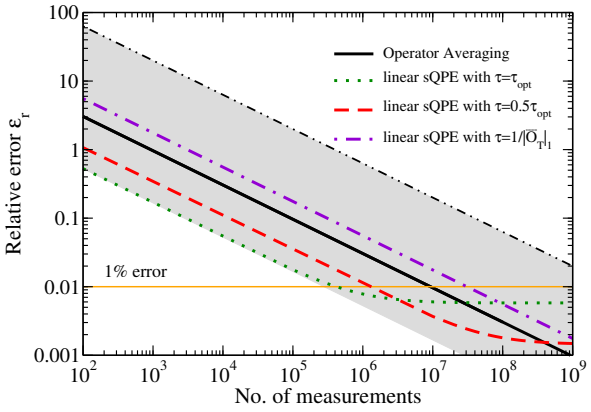


FIG. 2. Results of numerical simulations of the linear algorithm explained in the text together with the operator averaging method of Sec. I. The grey band indicates the expected variation caused by different choices for the time step τ in the linear algorithm. The horizontal orange line indicates an indicative $\epsilon_r = 1\%$ target error threshold.

and this will cause the total number of measurement required to increase as

$$\widetilde{N}_T(1,\epsilon_r) = \left(\frac{\lambda_u}{|\lambda_\Psi|}\right)^{3/2} N_{sQPE}(1,\epsilon_r)\ . \tag{39}$$

This is somewhat acceptable in the eigenvalue estimation case as an upperbound for the eigenvalue is used initially to check if sQPE is at all convenient using the condition in Eq. (33). The relatively strong dependence of $\widetilde{N}_T(1,\epsilon_r)$ on the value of λ_u and the difficulty to obtain the optimal time-step in the general case are the two main problems of the linear algorithm described here and are both strong motivations for developing the self-consistent algorithm described in the next section.

Before moving to the cubic algorithm, we want to further illustrate the sensitivity of the linear method to the particular choice of time step τ and we do so anticipating some results of the calculation of the deuteron binding energy presented in Sec. III. We plot in Fig. 2 the relative error for the estimation of the deuteron ground state energy as a function of the total number of measurements. In particular we show with a black solid line the analytical estimate (cf. Eq. (10)) for the error scaling of the operator averaging method of Sec. I

$$\epsilon = \sqrt{\frac{L}{N_{tot}}\sum_{k=1}^{L}\alpha_k^2\left[1-\widehat{P_k}^2\right]}\ , \tag{40}$$

as a function of the total number of measurements N_{tot}. As mentioned in Sec. I the simple adaptive scheme of Eq. (13) produces a slightly worse performance ($\approx 10\%$ larger constant factor, not shown) than the naive operator averaging method.

For sQPE with $K = 1$ (or linear method) we show

instead the mean squared error

$$\epsilon_M(\tau,1)=\sqrt{\frac{1-\langle Z_a\rangle(\tau)^2}{\tau^2 N}+\tau^4\frac{|\langle\mathcal{O}^3\rangle|^2}{36}}\,, \tag{41}$$

achievable with different time-steps. The green dotted line corresponds to sQPE with $K=1$ using the optimal choice for the time step Eq. (35) while the red dashed curve shows the detrimental effect of using a worse upperbound $\lambda_u=2|\lambda_\Psi|$ resulting in $\tau=\tau_{opt}/2$. In both cases sQPE provides an important speedup over operator averaging but this advantage is very fragile. The purple dot-dashed curve shows results obtained using the smallest time-step which could reasonably provide an advantage: the time step for which the variance in Eq. (23) equals an upperbound of the variance of the operator averaging estimator: $1/\tau=\|\overline{\mathcal{O}_T}\|_1$. The performance of operator averaging can be no worse than that, and indeed we see in Fig. 2 that linear sQPE with this time-step requires ≈ 3 times more measurements than the original scheme. The grey band spans the whole region covered by linear order algorithms with varying time-steps, in particular the upperbound (shown as dot-dot-dashed black line in Fig. 2) corresponds to the worse possible choice for the eigenvalue upperbound $\lambda_u=\lambda_{max}$.

The linear algorithm is simple to implement and can provide already important efficiency gains over Operator Averaging whenever we have the ability to make a good choice for the time-step parameter τ. We will now show how we propose to tackle this issue by using the next order sQPE algorithm corresponding to $K=2$.

C. Cubic algorithm

In order to use the general estimator of Eq. (20) for $K=2$ we need to be able to estimate $m_1=|\langle\mathcal{O}^3\rangle|$. In our implementation we achieve this by computing $\langle Z\rangle_a(\tau)$ for two different values of the time step and use these to extract both $\langle\mathcal{O}\rangle$ and m_1 using a cubic fit. Given a pair of time steps (τ_a,τ_b) the outcome of M independent measurements over of the projector $\Pi_a=|0\rangle\langle 0|$ is described by a pair of binomial random variables $X_a\sim B(M,P_a)$ and $X_b\sim B(M,P_b)$ with probabilities given by

$$\begin{aligned}P_{a/b}&=\frac{1-\langle\Psi|sin(\tau_{a/b}\mathcal{O})|\Psi\rangle}{2}\\&=\frac{1-\tau_{a/b}\langle\Psi|\mathcal{O}|\Psi\rangle+\frac{\tau^3_{a/b}}{6}\langle\Psi|\mathcal{O}^3|\Psi\rangle}{2}+O\left(\tau^5_{a/b}\right)\,.\end{aligned} \tag{42}$$

For small values of the time steps we can approximate these distributions with the 2-parameter family

$$\widetilde{P}_{a/b}(\mu,\eta)=\frac{1-\tau_{a/b}\;\mu+\frac{\tau_{a/b}}{6}\eta}{2}\,, \tag{43}$$

obtained from above by dropping the higher order terms. Estimators for the two parameters μ and η can be obtained by extremizing the likelihood $L(X_a,X_b|\mu,\eta,\tau_a,\tau_b)$ to observe a particular realization (X_A,X_B) given the distribution parameters (μ,η,τ_a,τ_b):

$$\begin{aligned}L(X_a,X_b|\mu,\eta,\tau_a,\tau_b)\propto\widetilde{P}_a^{X_a}(\mu,\eta)\left(1-\widetilde{P}_a(\mu,\eta)\right)^{M-X_a}\\\times\widetilde{P}_b^{X_b}(\mu,\eta)\left(1-\widetilde{P}_b(\mu,\eta)\right)^{M-X_b}\,.\end{aligned} \tag{44}$$

The resulting maximum likelihood estimators are

$$\mu_{mle}=c^\mu_{ab}\left[\frac{\tau_a^2}{\tau_b}\left(1-2\frac{X_b}{M}\right)-\frac{\tau_b^2}{\tau_a}\left(1-2\frac{X_a}{M}\right)\right]\,, \tag{45}$$

$$\eta_{mle}=c^\eta_{ab}\left[\tau_a\left(1-2\frac{X_b}{M}\right)-\tau_b\left(1-2\frac{X_a}{M}\right)\right]\,, \tag{46}$$

where the time-step dependent coefficients are

$$c^\mu_{ab}=\frac{1}{\tau_a^2-\tau_b^2}\quad\text{and}\quad c^\eta_{ab}=\frac{6}{\tau_a\tau_b(\tau_a^2-\tau_b^2)}\,. \tag{47}$$

We can estimate the variance of these estimators by computing the inverse of the Fisher information matrix

$$I(\mu,\eta)_{i,j}=-\mathbb{E}\left[\frac{\partial^2 log\left(L(X_a,X_b|\mu,\eta,\tau_a,\tau_b)\right)}{\partial i\partial j}\right]\,, \tag{48}$$

where the derivatives are taken over $\{i,j\}=\{\mu,\eta\}$ and the parametric dependence of $I(\mu,\eta)$ on the two time-steps $\tau_{a/b}$ has been suppressed for clarity. The results are

$$Var\left[\mu_{mle}\right]=\frac{4}{M}\frac{\tau_a^6P_b\left(1-P_b\right)+\tau_b^6P_a\left(1-P_a\right)}{\tau_a^2\tau_b^2\left(\tau_a^2-\tau_b^2\right)^2}\,, \tag{49}$$

$$Var\left[\eta_{mle}\right]=\frac{144}{M}\frac{\tau_a^2P_b\left(1-P_b\right)+\tau_b^2P_a\left(1-P_a\right)}{\tau_a^2\tau_b^2\left(\tau_a^2-\tau_b^2\right)^2}\,, \tag{50}$$

where the probabilities $P_{a/b}$ will have to be estimated using a finite sample. In this work we use the Bayesian estimators

$$\widehat{P_a}=\frac{X_a+1}{M+2}\qquad\widehat{P_b}=\frac{X_b+1}{M+2}\,, \tag{51}$$

obtained using a slightly informative Beta prior with $\alpha=\beta=1$, but in general one can employ any accurate sample estimator $\widehat{P_{a/b}}$ of the true probabilities $P_{a/b}$.

Note that estimators of the variance obtained in this way are in principle accurate only in the limit of large statistics $M\gg 1$, but in practice we found their use to be reasonable for the application studied in this work. Since the adaptive algorithm we describe below relies on their quality, further work on the construction of better (ie. more robust to noise or more rapidly converging) estimators of the fluctuations of μ and η may prove useful.

In addition to these statistical sources of error we also have a bias coming from the approximation $P_{a/b} \to \widetilde{P}_{a/b}$. This can be estimated to be

$$\begin{aligned} B_\mu(\tau_a,\tau_b) &= \mathbb{E}\left[\mu_{mle}-\mu\right] \\ &= c_{ab}^{\mu}\left[\frac{\tau_a^2}{\tau_b}\langle sin(\tau_b\mathcal{O})\rangle - \frac{\tau_b^2}{\tau_a}\langle sin(\tau_a\mathcal{O})\rangle\right] - \mu \ . \end{aligned} \tag{52}$$

A useful upperbound can be obtained by noticing that $B_\mu(\tau_a,\tau_b)$ is obtained from the remainder of the Taylor expansion in Eq. (42) as

$$B_\mu(\tau_a,\tau_b) = \langle\Psi|R_5|\Psi\rangle \ , \tag{53}$$

where the remainder operator R_5 is defined as

$$\begin{aligned} R_5 = \frac{c_{ab}^{\mu}\mathcal{O}^5}{24}\int_0^1 dt(1-t)^4\Big[&\tau_a^2\tau_b^4 cos(t\tau_b\mathcal{O}) \\ &- \tau_b^2\tau_a^4 cos(t\tau_a\mathcal{O})\Big] \ , \end{aligned} \tag{54}$$

as follows directly from the integral representation of the remainder of a Taylor series. We can now bound the bias in the cubic algorithm using for instance

$$|B_\mu(\tau_a,\tau_b)| \leq \frac{|\langle\Psi|\mathcal{O}^5|\Psi\rangle|}{120}\tau_a^2\tau_b^2\frac{\tau_a^2+\tau_b^2}{|\tau_a^2-\tau_b^2|} \ . \tag{55}$$

Achieving a tight bound for the bias is generally important as it controls the final efficiency of the method. As for the linear method above, for now we will focus on the special case of eigenvalue estimation while leaving the discussion on how to obtain practical upperbounds in more general situations in Sec. V A. In the calculations performed in our work we found a weak dependence of the computational effort with the particular choice of estimator for the bias (see Fig. 7) and will discuss the different options we used in Sec. III.

In the next subsection we present our strategy to determine the time steps for the cubic algorithm.

1. *Optimal determination of the times steps*

As was pointed out at the end of Sec. II B one of the major drawbacks of the linear algorithm is its sensitivity to the choice of the time-step τ. For the cubic algorithm we solve this issue by using ideas from Optimal Design (OD) [22, 23]. OD techniques have been used in a variety of applications to quantum computing ranging from quantum tomography [24–26], to parameter estimation [27, 28], to quantum-gate synthesis [29]. The general underlying idea in OD for parameter estimation is to try to optimize some, possibly unconstrained, hyper-parameters of an experiment (eg. the pair of time steps (τ_a,τ_b) to be used in the cubic sQPE) in order to minimize an estimator for the error in the parameter we want to estimate. In many situations this minimization procedure is translated into the maximization of some measure of the "size" of the Fisher information matrix $I(\mu,\eta|\tau_a,\tau_b)$ of Eq. (48) (these can be eg. one of its norms or its determinant). This procedure can be seen effectively to be a minimization of the Cramer-Rao bound for an unbiased estimator of the target parameter.

Since in our application the maximum likelihood estimator in Eq. (45) has a bias, we will minimize the mean squared error of μ_{mle} instead

$$\epsilon_M(\mu|\tau_a,\tau_b) = Var[\mu_{mle}] + B_\mu^2(\tau_a,\tau_b) \ . \tag{56}$$

One possible adaptive algorithm works by choosing, for any given iteration i, a new pair of time steps $(\tau_a^{i+1},\tau_b^{i+1})$ for the next rounds of M measurements by minimizing Eq. (56) using the estimators $\widehat{\mu}_{mle}$ and $\widehat{\eta}_{mle}$ available at the current iteration i. The initial pair can be chosen randomly provided both time steps are small, in the results shown below we sample one of the two from a uniform distribution $U(0,0.1)$[30] while the second is chosen to minimize the following upper bound for the variance

$$Var\left[\mu_{mle}\right] \leq \frac{1}{M}\frac{\tau_a^6+\tau_b^6}{\tau_a^2\tau_b^2\left(\tau_a^2-\tau_b^2\right)^2} \ , \tag{57}$$

keeping the first fixed. This procedure ensures that the estimator for μ obtained from the new set of measurements has the smallest MSE possible and is thus rather efficient early on. As we collect more data and the variance of our estimator $\widehat{\mu}_{mle}$ gets reduced we should however reduce the contribution of the bias by reducing in magnitude the new pair of time steps. In order to incorporate this effect we should obtain a new set $(\tau_a^{i+1},\tau_b^{i+1})$ of time steps by minimizing the expected variance of $\widehat{\mu}_{mle}$ after the new block of data is collected

$$\epsilon_M^i(\mu|\tau_a,\tau_b) = Var[\mu_{mle}] + (i+1)B_\mu^2(\tau_a,\tau_b) \tag{58}$$

which has the correct shot-noise scaling coming from $Var[\widehat{\mu}] \approx Var[\mu]/i$ as a function of the number of data blocks i collected so far. In practice we cannot evaluate the variance exactly and for the results presented in this work we found sufficient to employ the approximation

$$\widetilde{Var}[\mu_{mle}] = \frac{4}{M}\frac{\tau_a^6\widetilde{P}_b\left(1-\widetilde{P}_b\right)+\tau_b^6\widetilde{P}_a\left(1-\widetilde{P}_a\right)}{\tau_a^2\tau_b^2\left(\tau_a^2-\tau_b^2\right)^2} \tag{59}$$

where we have used

$$\widetilde{P}_{a/b} = \frac{1}{2}\left(1-\tau_{a/b}\widehat{\mu}_{mle}+\frac{\tau_{a/b}^3}{6}\widehat{\eta}_{mle}\right) \ , \tag{60}$$

to estimate the two probabilities $P_{a/b}$ (cf. Eq. (42)). Introducing an upperbound B_μ^u for the bias in Eq. (55), the final cost function we use to find the new set of time steps is therefore

$$\Delta_i\left(\mu|\tau_a,\tau_b\right) = \widetilde{Var}[\mu_{mle}] + (i+1)B_\mu^{u\,2}(\tau_a,\tau_b) \ . \tag{61}$$

The approximation of the variance in Eq. (59) is good in the limit of small time steps $\tau_{a/b}$ but this is not a problem since the presence of the bias term forces the optimal solutions to be numerically small automatically. In order to prevent numerical instabilities early on in the optimization we use a cost function that becomes extremely large whenever $\widetilde{P}_{a/b} \notin [0,1]$.

Before moving to the results in the next section we summarize the cubic algorithm in pseudo code as follows:

sample $\tau_a \sim U(0, \tau_{max})$
compute τ_b as the minimum of Eq. (57)
for iteration number less than N_{iter} **do**
  perform M measurements each to obtain X_a, X_b
  construct current MLE estimators Eq. (45,46)
  evaluate running averages $\widehat{\mu}_{mle}$,$\widehat{\eta}_{mle}$
  use these estimators to define cost function Eq. (61)
  new pair (τ_a, τ_b) are the minimizers of $\Delta_i(\mu|\tau_a, \tau_b)$
end for

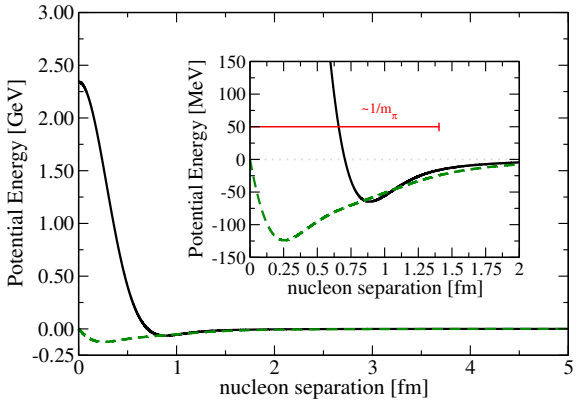


FIG. 3. Central (black solid line) and tensor (green dashed line) contributions to the nuclear potential Argonne V6' used to obtain the deuteron Hamiltonian Eq. (63). In the inset the range of pion-exchange is also shown. The grey dotted line at 0 energy is simply a guide to the eye.

III. THE DEUTERON GROUND STATE

The deuteron is the simplest nucleus present in nature. It is a bound state of a neutron and a proton in a state having total isospin $T = 0$, spin $S = 1$ and angular momentum-parity $J^\pi = 1^+$. It has a small binding energy of approximately 2.2 MeV. The ground state of the deuteron has a non-zero quadrupole moment, originated by the mixing between S- and D-waves generated by pion-exchanges (see eg. [31] for a pedagogical introduction). A simple model for the deuteron is to consider the 2-level system built from an S-wave orbital $|\phi_S\rangle$ and a D-wave orbital $|\phi_D\rangle$:

$$H = \begin{pmatrix} \langle\phi_S|H|\phi_S\rangle & \langle\phi_S|H|\phi_D\rangle \\ \langle\phi_D|H|\phi_S\rangle & \langle\phi_D|H|\phi_D\rangle \end{pmatrix} . \tag{62}$$

Using the Argonne Av6' potential [32] we obtain (approximately [33]) the following Hamiltonian matrix:

$$H = \begin{pmatrix} 5 & -35 \\ -35 & 170 \end{pmatrix} = 87.5\mathbb{1} - 35X + 82.5Z \, . \tag{63}$$

Large cancellations among different contributions produce a ground state energy orders of magnitude smaller than the norm

$$E_{gs} = -2.1174 \quad \|\overline{H_T}\|_1 = 117.5 \quad R_O \approx 0.018 \, ,$$

where, in analogy to the previous sections, we defined H_T to be the traceless part of H while R_O is the ratio defined above in Eq. (15).

This large cancellation is a direct consequence of the hard-core nuclear repulsion that we can see in the central component of the interaction in Fig. 3. The strong repulsion introduces states with very large energies in the many-body Hilbert space, a notorious problem which causes calculations on a finite basis to converge extremely slowly. General strategies to alleviate this issue have been developed in the past (eg. the Similarity Renormalization Group approach [34, 35]) but they are usually accompanied by an increase in the degree on nonlocality of the Hamiltonian (see eg. [36]) which in general will require a (possibly large) increase in the number of terms needed in expansions of the form Eq. (7).

Note however that, even when the detrimental effects of hard-core interactions are mitigated trough an effective theory like the one mentioned above, the requirement of ensuring basis-size convergence by performing multiple calculations with progressively larger basis sets will still lead to a potentially large mismatch between the ground state energy and the Hamiltonian norm. In that case this is due to fact that, as the basis size increases, the ground state energy will decrease at a much slower rate than the maximum eigenvalue (indeed E_{gs} will reach a plateau for large basis while the highest eigenvalue will grow indefinitely). A general strategy to reduce the importance of this problem (like the sQPE scheme presented in this work) is thus welcome more generally.

Let's now start to discuss the performance of the Operator Averaging method of Sec. I on our model deuteron problem Eq. (63). Using the estimate from Eq. (16) we find that the number of measurement required for target relative accuracy ϵ_r is given by

$$N_A(\epsilon_r) = \frac{1}{(R_O \epsilon_r)^2} \approx \frac{3079.4}{\epsilon_r^2} \quad \to \quad 3.1 \times 10^7 \, , \tag{64}$$

where the last limit holds for $\epsilon_r = 1\%$. Even tough this estimate might not be very tight since we neglected the variances in Eq. (10) in order to derive Eq. (16), the fact that we are dealing with a simple one-qubit system and that we haven't considered yet the effect of errors, makes this requirement already alarming. In order to put this number in perspective, the IBM group [7] estimated that $N_A \approx 10^6$ measurements would be sufficient

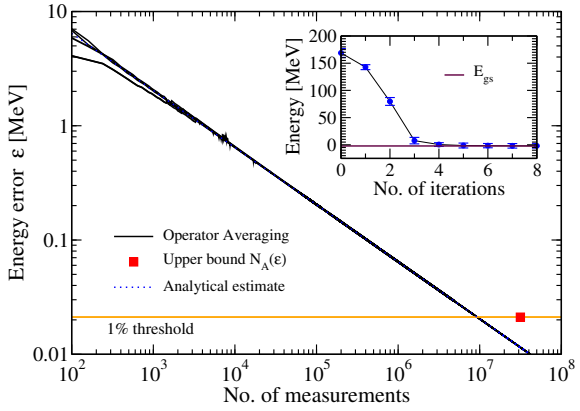


FIG. 4. Results of numerical simulations of the algorithm explained in the text. Results correspond to 5 independent runs with different random number seeds. The blue dotted line corresponds to the expected asymptotic behavior from Eq. (40) while the red dot correspond to the upperbound from Eq.(64). The inset shows the convergence of the VQE optimization using low resolution expectation values with errors $\approx$ 5 MeV (these were obtained using $N_{tot} = 10^3$ measurements per function evaluation). The horizontal line indicates the value of the ground state energy $E_{gs} = -2.1174$ MeV.

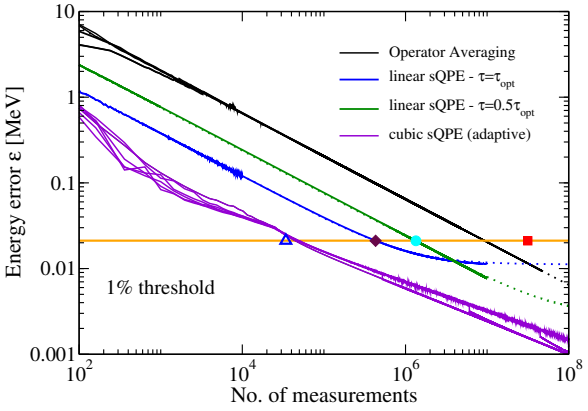
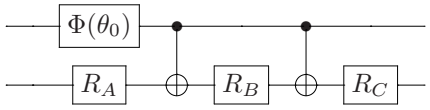


FIG. 5. Final error in the estimator for the ground state energy of the deuteron obtained with the techniques discussed in this work, from top to bottom they are: Operator Averaging (black line), linear sQPE with optimal time-step (blue line), linear sQPE with half of the optimal time-step (green line) and the adaptive cubic sQPE algorithm (purple line) . Dotted lines correspond to the analytical results presented in Fig. 2. See text for the meaning of the marked points.

to reach chemical accuracy for a 6 qubit model of BeH_2 with hundreds of Pauli terms in the Hamiltonian expansion of Eq. (7).

In Fig. 4 we show results for an ideal implementation (no noise apart from statistical fluctuations) of our one qubit model. As we can see the upperbound of Eq. (64) is only a factor of a few larger and we find that $\approx 9.3 \times 10^6$ measurements are needed in this ideal noiseless case. Before moving on to discuss the results we have obtained using sQPE, we want to point out that if our goal was only to optimize a variational state $\left|\Phi(\vec{\theta})\right\rangle$ using the energy expectation value (ie. we want to run VQE [5]) then low accuracy results for the energy could be sufficient to get close to the optimum $\vec{\theta}_{min}$. To illustrate this we show in the inset the deuteron energy expectation value obtained using $N_{tot} = 10^3$ as a function on the iteration in the minimization procedure (for these results we used a simple Nelder-Mead optimizer). Even though the error in the energy is $\gtrsim 200\%$ the angle θ converges towards θ_{min} to within a few percent error in only a small number of iterations (note that for this simple model a single angle is sufficient to prepare the ground state). This striking difference is probably peculiar to simple models like our one-qubit deuteron, since the lack of excited states with low energy in the spectrum of the Hamiltonian gives rise to large gradients in the the variational energy $E(\theta) = \langle\Phi(\theta)|H|\Phi(\theta)\rangle$ and therefore to a relatively easy optimization. Where sQPE could be most useful in this case is for the final estimation of the energy, but in general for more complex systems low order sQPE could be advantageous also in the last stages of optimization where large statistical fluctuations could prevent to reach the minimum.

Let's now turn to discuss the sQPE method starting with the implementation of the controlled time evolution from Eq. (18). In our simple two qubit situation the circuit for the controlled unitary can be constructed using only 2 CNOT gates (cf. [37]):



(65)

where $\Phi(\alpha) = diag\left(1, e^{i\alpha}\right)$ is a phase gate and the R_A, R_B, R_C blocks are formed by appropriate single qubit rotations (see Appendix A for more details).

In Fig. 5 we show the results obtained with both sQPE and operator averaging: the solid curves correspond to empirical results while the dotted lines correspond to the analytical estimates discussed in Sec. II B. As for Fig. 2, the red square marks the location of the upperbound N_A from Eq. (64). For the linear method, the optimal choice Eq. (35) is shown in blue, while in green we present the results obtained using a more conservative value $\tau = \tau_{opt}/2$ (cf. Eq. (38) and the discussion following it). In both cases we see that, in agreement with the results presented in Fig. 2, the linear algorithm requires about an order of magnitude less measurements than operator averaging. In addition to this we see that the estimated number of measurements N_{sQPE} are in very good agreement with the empirical results: the maroon diamond indicates the upperbound $N_{sQPE}(1)$ for linear sQPE Eq. (37) while the cyan circle corresponds to the worse bound obtained trough Eq. (39).

As we discussed in Sec. II B the speedup offered by the linear method is very sensitive to the particular choice of time step used and by employing the adaptive strategy

described in Sec. II C to find the time steps (τ_a, τ_b) the cubic algorithm partially overcomes this problem. In order to implement the algorithm we use the same circuit described for the linear method: we just run it twice for the two time steps separately. The results obtained from 6 different runs are presented in Fig. 5 as purple lines, in all cases we update the time step pair every block of $M = 40$ measurements. In addition, the blue triangle is twice the expected number of measurements $N_{sQPE}(2)$ obtained from the general expression Eq. (26)

$$N_{sQPE}(2, \epsilon_r) = \frac{f(2)}{\epsilon^{5/2}} \sqrt{|\langle H^5 \rangle|} \approx 1.7 \times 10^4 \,, \qquad (66)$$

where the factor of 2 is introduced to account for the fact that in our adaptive scheme we are actually estimating two expectation values: $\langle \mathcal{O} \rangle$ and m_1. The optimal pair of time steps (τ_a, τ_b) obtained during the execution of the algorithm fluctuate around the value $(0.15, 0.3)$ which is not very far from the optimal time step $\tau_{opt} \approx 0.4$ found from Eq. (25). The spread of results at large measurement count is possibly a signature that the optimization of Eq. (61) gets stuck in local minima, we plan to investigate this further in future work.

As explained in Sec. II C, the cubic algorithm needs a good approximation of the bias term $B_\mu(\tau_a, \tau_b)$ in Eq. (55) as this enters directly the cost function Eq. (61) used to determine the optimal time steps. The results presented above where obtained using the following ansatz

$$B^u_\mu(\tau_a, \tau_b) = \frac{|\widehat{\mu}_{mle}\widehat{\eta}_{mle}|}{120} \tau_a^2 \tau_b^2 \frac{\tau_a^2 + \tau_b^2}{|\tau_a^2 - \tau_b^2|} \equiv B_{A1} \,, \qquad (67)$$

where $\widehat{\mu}_{mle}$ and $\widehat{\eta}_{mle}$ are the current best estimators for the distribution parameters (μ, η). This form reduces to the correct one in the eigenvalue estimation limit relevant here where $|\Psi\rangle$ is an eigenstate of $\mathcal{O}$. We want now to present results showing the weak sensitivity of the cubic algorithm to the specific choice of the estimator for the bias, in particular we will use two additional estimators: the exact one from Eq. (52)

$$B_E = \mathbb{E}\left[\mu_{mle} - \mu\right], \qquad (68)$$

which in practical situations we won't have access to, and a different variant of the estimator B_{A1} above defined as

$$B_{A2} = \frac{|\widehat{\mu}_{mle}\widehat{\eta}_{mle}|}{120} \tau_a^2 \tau_b^2 \frac{\max\left[\tau_a^2, \tau_b^2\right]}{|\tau_a^2 - \tau_b^2|} \,. \qquad (69)$$

This estimator is a tighter bound that can be obtained from Eq. (54) using the additional condition

$$\max\left[\tau_a^2, \tau_b^2\right] < \frac{\pi}{\|\overline{\mathcal{O}}\|_1} \,. \qquad (70)$$

In Fig. 6 we show how these estimators evolve as the algorithm proceeds in three different situations. The top panel shows the ideal situation where the optimal pair

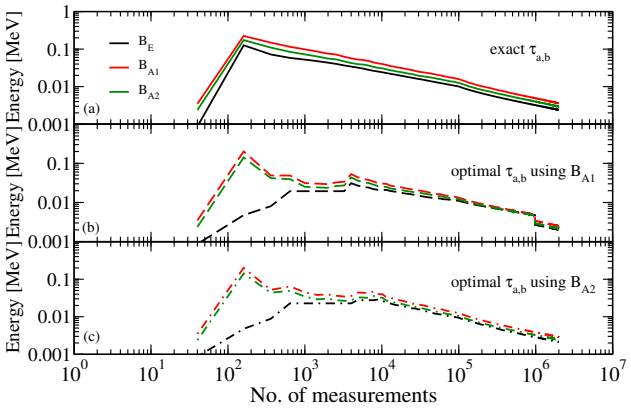


FIG. 6. Different estimators of the energy bias for a cubic sQPE calculation of the deuteron ground state energy, from top to bottom (in all panels): A1 approximation B_{A1}, B_{A2} approximation B_{A2} and exact estimator B_E (see text for the definitions). The top panel shows results for an idealized case while panel (b) and (c) are obtained using approximate optimizations employing the estimators B_{A1} and B_{A2} (cf. Eq. (67) and Eq. (69)).

(τ_a, τ_b) for the next step is obtained using the exact scaled mean squared error from Eq. (58). Even though in practical situation we won't be able to run the cubic algorithm this way, these results provide a ceiling for the performance of approximate algorithms while at the same time show clearly the source of the advantage that is achieved with sQPE: initially the time steps are raised to relatively large values in order to reduce the shot noise limited variance term in the equation above at the expense of a larger bias term. This allows to quickly reduce the error in the expectation value early on when the dominant contribution are statistical fluctuations. As the accuracy increases the importance of the bias term grows and the adaptive algorithm starts to reduce the magnitude of the time steps in order to keep B_μ under control. Furthermore we see that the three bias estimators follow each other rather closely.

The other two panels instead correspond to results obtained using the approximate cost function Δ_i from Eq. (61) with either the B_{A1} ansatz used also in Fig. 5 (central panel) or the B_{A2} ansatz described above (bottom panel). In both situations we recover the same qualitative behavior seen in the ideal case: the bias gets initially increased and then reduced gradually as the accuracy improves. The main difference with the results of the top panel is the lower efficiency obtained in the first stage of this procedure where the actual exact bias remains much smaller than it could have been for the first few hundred measurements (note that as above we update the time-steps every $M = 40$ measurements) but the discrepancy quickly vanishes later when the bias becomes the limiting factor.

The see the impact of these approximations on the final convergence of the expectation value we show in Fig. 7 the results obtained by using the exact pair of time steps

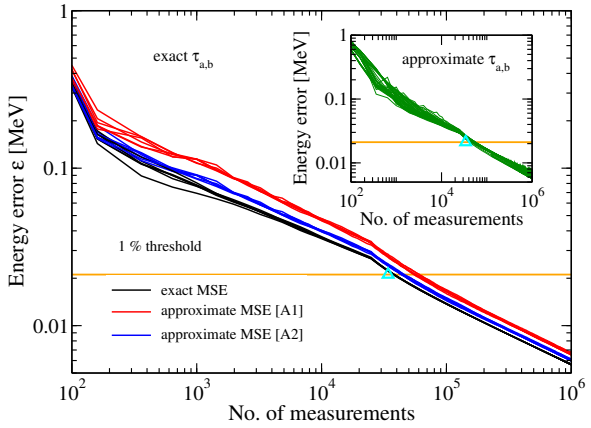


FIG. 7. Different estimators for the final error in the deuteron ground state energy form ideal runs where the time-steps are optimized exactly using Eq. (58), from top to bottom: approximate A1 results, approximate A2 results and exact MSE result . The cyan up triangle is the same as the blue triangle in Fig. 5. The inset show results obtained using the approximate cost function Eq. (61).

$\tau_{a/b}$ obtained as before using Eq. (58) but different approximations to the final mean squared error. The reason this is important is that we need to estimate the bias in order to provide an estimate for the final error of our estimated expectation value. The results in the main panel show the apparent reduced efficiency that is a result of using a bigger bias than the exact one (in this case we used B_{A1} and B_{A2} as above). The cyan up triangle is the same as the blue triangle in Fig. 5 and indicates an optimistic expectation on the efficiency of the cubic algorithm. Somewhat not surprisingly we achieve that estimate only in the ideal exact case shown as black lines in the main panel of Fig. 7 while in the worst case (corresponding to the red curves) we need to perform as much as $\approx 50\%$ more measurements. Interestingly when statistical fluctuations in the estimation of the variance are included these differences mostly disappear and there is no clear preference for different choices of the bias, we can see this from the results shown in the inset of Fig. 7 where the time steps were estimated using the approximate cost function Δ^i from Eq. (61).

These results are encouraging as they show that, even if a tight upperbound for the bias is helpful for the algorithm, an approximate expression can work very well at least for the special case of eigenvalue estimation. Evaluating upperbounds becomes more important in the general case and we leave the discussion for Sec. V A while we now turn to the problem of accounting for the presence of noise in the quantum device.

IV. EFFECT OF NOISE

As we have mentioned in the introduction, noise will be unavoidable for near term quantum devices and it is therefore critical for algorithms to provide robustness against noise if we want to deploy them on a non fault-tolerant quantum computer. Since the methodology we propose goes against the popular trend in that we are trading classical resources (the number of experimental trials) with quantum ones (one more qubit for the ancilla and more gates), we need to provide supporting evidence that our method shows advantages even in the presence of noise and is thus practical.

qubit	Rotation err	Readout err
0	0.0019	0.0865
1	0.0024	0.08
2	0.0024	0.0382
3	0.0027	0.3567
4	0.0036	0.2715

TABLE I. Rotation (U3) and readout errors for IBM's 5-qubit machine 'ibmqx4' on May 8 2019. The error on the CNOT gate on the pairs $[qubit2, qubit1]$ and $[qubit3, qubit2]$ is 4.88% and 6.68% respectively.

The importance of this assessment is critical as there are known cases where the advantage of a quantum algorithm can be drastically reduced by the presence of even small sources of noise (see eg. [38–41]).

We start by showing how, in situations where the condition in Eq. (31) is valid by a large margin (as in the deuteron model discussed here), a substantial increase in classical resources is required to minimize the effect of measurement noise for the Operator Averaging method of Sec. I while with sQPE this problem can be mitigated substantially. In the last part we provide a more general argument in support of our measurement strategy in situations where a clean ancilla is available in the spirit of the D1QC model [42].

A. Measurement noise

Assignment errors in the measurement device used for qubit read out are an important source of bias that needs to be accounted for in order to properly obtain meaningful results. We illustrate the problem showing in Fig. 8 results obtained executing the deuteron problem, described in the previous section, on an emulated version of the IBM 5-qubit machine 'ibmqx4' using the Qiskit software package [43] (see Tab. I for details).

Despite the fact that we are not correcting for any source of errors in these results, the linear sQPE (with optimal time τ) presented on the right seems to provide considerably higher quality results despite the much increased circuit depth. In the following we provide an argument to explain the observed results.

Here we will use an extremely simplified model for these errors that nevertheless captures their essential features. We achieve this by replacing the projectors Π_0, Π_1

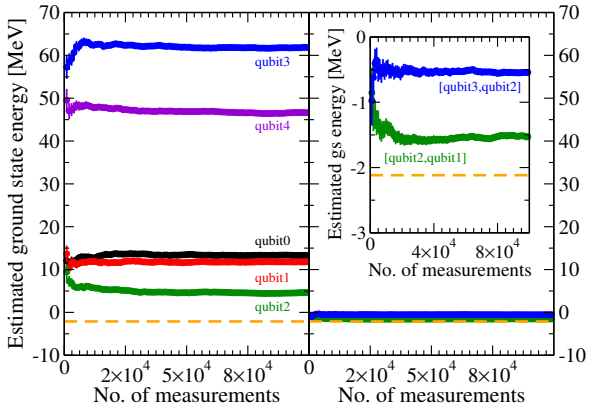


FIG. 8. Results for the deuteron ground state energy obtained using the emulated 'ibmqx4' quantum computer without error mitigation. The left panel shows results obtained using Operator Averaging on each qubit while the right panel shows the performance of the linear sQPE algorithm using pairs $[qubit2, qubit1]$ (green data points) and $[qubit3, qubit2]$ (blue data points) as ancilla and system qubit respectively.

on the states $|0\rangle$, $|1\rangle$ of a qubit with the following ones

$$\begin{aligned}\widetilde{\Pi_0} &= (1-p)\,\Pi_0 + p\Pi_1 \\ \widetilde{\Pi_1} &= (1-p)\,\Pi_1 + p\Pi_0\end{aligned} \tag{71}$$

where $0 < p < 1$. This model can be justified in the limit where assignment errors are both qubit independent and symmetric with respect to the interchange $|0\rangle \leftrightarrow |1\rangle$ and is sufficient for our purpose (see eg. Supplemental Material of [7] for details on a more accurate model). Using Eq. (71) we find that the noisy expectation value of some one-qubit Pauli operator $\langle\widetilde{P_\sigma}\rangle$ is related to the noise-free value by the relation

$$\langle\widetilde{P_\sigma}\rangle = (1-2p)\,\langle P_\sigma\rangle\ , \tag{72}$$

which can be easily inverted to estimate $\langle P_\sigma\rangle$ from $\langle\widetilde{P_\sigma}\rangle$. Despite its simplicity this model is sufficient to completely account for the error afflicting the OA results of Fig. 8, as we can see from the error mitigated results presented in Fig. 9. Note that the linear sQPE energies in the right panel are still biased due mostly to the noise introduced by using the CNOT gates, the mitigation of which is beyond the scope of our discussion here (note however that mitigation techniques [44, 45] will be required also for OA for larger target systems).

Generalizations of Eq. (72) to expectation values of multi-qubit Pauli operators can also be obtained, but we will limit our discussion to the one-qubit case relevant to our deuteron calculations and to sQPE more generally. In fact, an important feature of sQPE is that the output of the algorithm is obtained through measurements on a single qubit thus avoiding the problem of the exponential reduction of signal to noise ratio as a function of the number of qubits involved in the measurement of individual terms in the expansion Eq. (7) (see eg. [7]).

For a generic one-qubit observable described by the general expansion (cf. Eq. (7))

$$\mathcal{O} = \beta_0 + \sum_{i=1}^{3}\beta_i P_i \quad \vec{P} = (X,Y,Z)\ , \tag{73}$$

we can use Eq. (72) to estimate the noise free expectation value:

$$\langle\widehat{\mathcal{O}}\rangle = \beta_0 + \sum_{i=1}^{3}\frac{\beta_i}{1-2\widehat{p}}\langle\widetilde{P}_i\rangle\ , \tag{74}$$

where $\widehat{p}$ is a finite sample estimator ~~estimator~~, with variance δp, of the error probability p. The variance of this estimator can be approximated as

$$Var[\widehat{\mathcal{O}}] = \frac{Var[\widetilde{\mathcal{O}}]}{(1-2\widehat{p})^2} + V_R;\,, \tag{75}$$

with

$$V_R = \frac{4\delta p^2}{(1-2\widehat{p})^2}\sum_{i=1}^{3}\frac{|\beta_i|^2}{(1-2\widehat{p})^2}\langle\widetilde{P}_i\rangle^2\ , \tag{76}$$

where we used a linear expansion to propagate the error (ie. we used $Var[f(x)] \approx f'(x)^2 Var[x]$). The second error term in Eq. (75) comes from the uncertainty in the determination of the error parameter p and provides a noise floor that we need to minimize in order to achieve good accuracies. Assuming the sample estimator $\widehat{p}$ was obtained from N_C calibration measurements, we can bound the contribution of this background as

$$V_R \le 4\frac{\|\overline{\mathcal{O}_T}\|_2^2}{(1-2\widehat{p})^4}\delta p^2 = 4\frac{\|\overline{\mathcal{O}_T}\|_2^2}{(1-2\widehat{p})^4}\frac{\widehat{p}\,(1-\widehat{p})}{N_C}\ . \tag{77}$$

The case of the sQPE algorithm is simpler because we have always to deal with a single qubit to be measured. Using the same correction scheme employed above, assuming again that the higher order coefficients m_i are

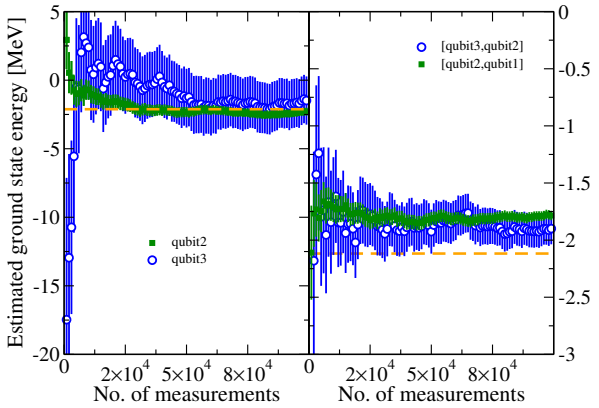


FIG. 9. Same as in Fig. 8 but using the error mitigation strategy described in the text for the measurement noise.

known, we find

$$\langle\widehat{\mathcal{O}_K}(\tau)\rangle = -\frac{1}{\tau}\frac{\langle\widetilde{Z_a}\rangle}{1-2\widehat{p}} - \sum_{k=1}^{K-1}\tau^{2k}\frac{(-1)^k m_k}{(2k+1)!}\,, \tag{78}$$

with variance

$$Var[\widehat{\mathcal{O}_K}(\tau)] = \frac{Var[\widetilde{\mathcal{O}}_K(\tau)]}{(1-2\widehat{p})^2} + V_{RK}\,, \tag{79}$$

and

$$\begin{aligned} V_{RK} &= \frac{4}{\tau^2}\frac{\delta p^2}{(1-2\widehat{p})^2}\frac{\langle\widetilde{Z_a}\rangle^2}{(1-2\widehat{p})^2} \\ &\leq \frac{4}{\tau^2}\frac{1}{(1-2\widehat{p})^4}\frac{\widehat{p}(1-\widehat{p})}{N_C}\,. \end{aligned} \tag{80}$$

It is then clear that the sQPE algorithm will reduce the importance of measurement noise in the same situations where it provides an advantage in the noise free case, namely whenever $\|\overline{\mathcal{O}_T}\| \gg \frac{1}{\tau^2}$ with τ the estimated optimal time step for a particular problem.

In order to assess the practical impact of this error term for the deuteron calculation of Sec. III we have numerically minimized the total number of measurements $N_{tot} = 2M + N_C$ needed to achieve $\epsilon_r = 1\%$ (M for each of the Pauli terms and N_C to estimate $\widehat{p}$) as a function of the error parameter p using directly Eq. (75). In Fig. 10 we show the results of this study: the full black line is the minimal value of N_{tot} needed to reach a target relative error $\epsilon_r = 1\%$, while the green dashed line corresponds to the situation where we have performed a calibration using $N_C^0 = 10^7$ initial measurements to estimate the error ra

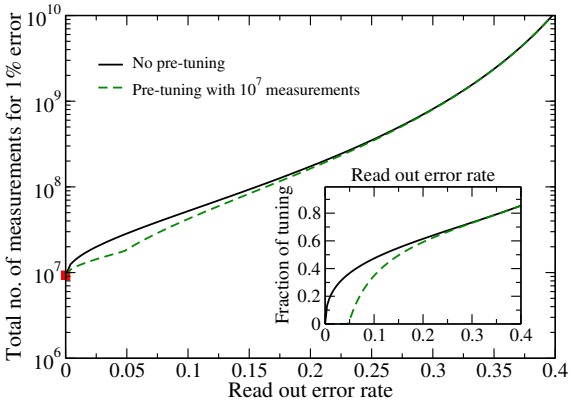


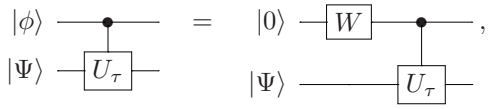
FIG. 10. Estimated number of measurements required to reproduce the ground-state energy of the deuteron to $\epsilon_r = 1\%$ accuracy as a function of the read-out error rate. Continuous lines show are the total number ($N_{tot} = 2M + N_C$) while the dashed lines show the fraction of the total measurements that has to be dedicated to characterize the readout noise.

The inset shows the ratio of tuning measurements needed to achieve the target accuracy as a function of the error rate p in both situations. We see that reasonably small values of the error rate $p \lesssim 5\%$ can be dealt with relative ease using data obtained from previous calibrations, while more substantial efforts are needed for more noisy qubits. Using *qubit*3 with its large error $p \gtrsim 35\%$ would require two orders of magnitude more measurements than in the noiseless case (shown as the red square in Fig. 10) out of which more than $\approx 80\%$ would be needed for characterizing the noise. Fortunately noise levels so high are not common on modern machines, with typical values of a few percent in the superconducting circuit case, but the fact that this very simple source of error is capable of completely swarming the results of a simple single qubit calculation provides another motivation to explore the use of the sQPE scheme on near term architectures.

B. General advantage of ancilla-based schemes

The purpose of this section is to show that the ancilla-based construction of sQPE (cf. Eq. (18)) can be advantageous in general when we want to estimate the value of an expectation value $\langle\mathcal{O}\rangle$ in presence of depolarizing noise (see eg. [46]) in the quantum device. For a realistic advantage to be found we will assume we have either a clean ancilla (meaning completely error free) or at least a qubit subject to a well characterized noise channel and with high measurement fidelity.

Let's start by considering a slight generalization of the sQPE circuit Eq. (18) where we leave unspecified the state of the ancilla before we apply the controlled unitary



(81)

and in the second circuit we made explicit the presence of a new rotation matrix W which prepares $|\phi\rangle$ starting from $|0\rangle$. The action of the time evolution unitary U_τ on the target state $|\Psi\rangle$ can be conveniently expressed as

$$U_\tau|\Psi\rangle = \kappa|\Psi\rangle + \nu\left|\Psi^\perp\right\rangle, \tag{82}$$

where $\kappa, \nu \in \mathbb{C}$ with

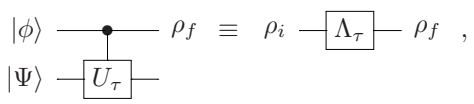
$$\langle\Psi|\Psi^\perp\rangle = 0 \quad \text{and} \quad |\kappa|^2 + |\nu|^2 = 1\,. \tag{83}$$

In other words, the total Hilbert space explored with circuits of type Eq. (81) is only 4-dimensional: we have $\mathbb{C}^2$ for the ancilla and the linear span of $|\Psi\rangle$ and $\left|\Psi^\perp\right\rangle$ for the target system. At this point, a measurement of the y-polarization of the ancilla after the circuit in Eq. (81) will reveal the wanted quantity (cf. Sec. II)

$$\langle Y\rangle_a = \mathcal{I}\left[U_\tau|\Psi\rangle\langle\Psi|\right] = \mathcal{I}\left[\kappa\right] \equiv \kappa_I\,, \tag{84}$$

from which we can extract the expectation value as discussed above (in the expression above $\mathcal{I}$ denotes the imaginary part).

If we trace out the system qubits, the circuit in Eq. (81) can be represented as a quantum channel Λ_τ acting on the ancilla:



(85)

where the output state of the ancilla is indicated here as a density matrix ρ_f. We want now to show how by performing quantum process tomography [47–50] on the ancilla we can extract κ_I.

We can represent the quantum channel Λ_τ using the following Kraus decomposition (see eg. [46])

$$\Lambda_\tau[\rho] = A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger \tag{86}$$

with

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \kappa \end{pmatrix} \qquad A_1 = \begin{pmatrix} 0 & 0 \\ 0 & \nu \end{pmatrix} , \tag{87}$$

but this choice is not unique. A better parametrization of the channels that overcomes this difficulty is to use the Pauli Transfer Matrix [51] defined as

$$R_{ij} = \frac{1}{2} Tr\left[P_i \Lambda_\tau\left[P_j\right]\right] , \tag{88}$$

where the P_i's are the Pauli operators $\{\mathbb{1}, X, Y, Z\}$. For our quantum channel this matrix takes the form

$$R_{U_\tau} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \kappa_R & -\kappa_I & 0 \\ 0 & \kappa_I & \kappa_R & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \tag{89}$$

where $\kappa = \kappa_R + i\kappa_I$. This form makes it apparent that the channel Λ_τ is a composition of a dephasing (or phase damping [46]) channel

$$R_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p_z & 0 & 0 \\ 0 & 0 & 1-p_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \tag{90}$$

with error probability $p_z = |\nu|^2 = 1 - |\kappa|^2$ and a rotation around the z-axis with angle $\theta = tan^{-1}(\kappa_I/\kappa_R)$ described by

$$R_\theta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) & 0 \\ 0 & sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{91}$$

As shown by Wiebe et al. in [52] a Bayesian reconstruction strategy can be effectively employed to learn κ_R and κ_I from measurement of the device even in the presence of substantial depolarizing noise described by the channel

$$\Lambda_D(\rho) = (1 - p_D) + \frac{p_D}{d}\mathbb{1} , \tag{92}$$

where $d = 2^n$ and n is the number of qubits used to encode $|\Psi\rangle$. According to the results in [52], large values $p_D = 50\%$ could be handled with relative ease.

Furthermore, the reconstruction is simplified in our case since the structure of our channel is known beforehand, and one can tailor strategies aimed at estimating matrices of the specific form in Eq. (89). The extent to which these could be leveraged to minimize the negative effect of more realistic noise channels in the system qubits is left for future explorations. Before concluding, we want to point out that the possibility of removing read-out lines from all but the ancilla qubit could also help more generally in reducing the overall noise in the device.

V. IMPLEMENTATION CHALLANGES

During the exposition of our methodology in Sec. II we have only briefely touched upon the practical cost of implementing the core parts of the algorithm on near term quantum devices. This section is dedicated to address these issues. In particular we first present a discussion on how to estimate the potential gain of using sQPE using condition Eq. (31) in practical situations where only partial information on the higher order coefficients m_k is available. We then provide a description of the resources needed to implement the time evolution needed for sQPE using different strategies.

A. Practical bound estimation

Due to the presence of the expectation value $\langle \mathcal{O}^{2K+1} \rangle$, it is difficult in most situations to asses directly if the condition in Eq. (31) holds. In order to obtain a more manageable condition we can rewrite Eq. (31) as

$$\frac{|\langle \mathcal{O} \rangle|}{\|\overline{\mathcal{O}_T}\|_1} \geq \frac{f(K)^K}{\epsilon_r} \frac{\langle \mathcal{O}^{2K} \rangle |\langle \mathcal{O} \rangle| + Cov[\mathcal{O}^{2K}, \mathcal{O}]}{\|\overline{\mathcal{O}_T}\|_1^{2K+1}} . \tag{93}$$

Due to fact that $h(x) = x^K$ with $K \geq 1$ has a bounded first derivative on a finite interval Ω we have

$$|x^K - y^K| \leq \max_{z \in \Omega}[K z^{K-1}]|x - y| , \tag{94}$$

which in turn implies

$$Var[\mathcal{O}^{2K}] \leq 4K^2 \lambda_{max}^{4K-2} Var[\mathcal{O}] \tag{95}$$

with λ_{max} the largest singular value of the operator $\mathcal{O}$. Using the bound

$$|Cov[X, Y]| \leq \sqrt{Var[X] Var[Y]} , \tag{96}$$

which can be obtained by using Jensen's inequality, we arrive at the following, looser, condition

$$\frac{|\langle \mathcal{O} \rangle|}{\|\overline{\mathcal{O}_T}\|_1} \geq \frac{f(K)^K}{\epsilon_r} \frac{\langle \mathcal{O}^{2K} \rangle |\langle \mathcal{O} \rangle| + \|\overline{\mathcal{O}}\|_1^{2K-1} Var[\mathcal{O}]}{\|\overline{\mathcal{O}_T}\|_1^{2K+1}} . \tag{97}$$

Reasonably tight ubberbounds on $Var[\mathcal{O}]$ and $\langle\mathcal{O}^{2K}\rangle$ for small $K = O(1)$ can be obtained in many situations of interest: an important example are many-body calculations of ground-state properties where a variational calculation with classically simulatable trial states can provide such bounds with reasonable efficiency (eg. one could use Quantum Monte Carlo methods [53, 54]). Another situation is when we have some control on the fidelity of the prepared state. For instance consider the case where we are preparing an initial state $|\Psi\rangle$ with large overlap with some eigenvector $|\phi\rangle$ of $\mathcal{O}$

$$|\Psi\rangle = \alpha|\phi\rangle + \beta|\phi^{\perp}\rangle \quad \mathcal{O}|\phi\rangle = \lambda_{\phi}|\phi\rangle \ , \tag{98}$$

with $\langle\phi|\phi^{\perp}\rangle = 0$ and $|\alpha|^2 + |\beta|^2 = 1$. We can obtain a bound on $\langle\mathcal{O}^{2K}\rangle$ if we have a upperbound for the state fidelity

$$F\left[|\Psi\rangle\right] = Tr\left[|\Psi\rangle\langle\Psi|\phi\rangle\langle\phi|\right] = |\langle\Psi|\phi\rangle|^2 < \Delta \ , \tag{99}$$

by using

$$\langle\mathcal{O}^{2K}\rangle \leq \lambda_{\phi}^{2K} + \Delta\|\overline{\mathcal{O}}\|^{2K} \ . \tag{100}$$

When only a bound on the variance is available instead we can use the inequality

$$\langle\mathcal{O}^{2K}\rangle|\langle\mathcal{O}\rangle| \leq \lambda_{max}^{2K+1} \leq \|\overline{\mathcal{O}}\|_1^{2K+1} \ , \tag{101}$$

to obtain the even looser condition

$$\frac{|\langle\mathcal{O}\rangle|}{\|\overline{\mathcal{O}_T}\|_1} \geq \frac{f(K)^K}{\epsilon_r}\frac{\|\overline{\mathcal{O}}\|_1^{2K+1}}{\|\overline{\mathcal{O}_T}\|_1^{2K+1}}\left(1 + \frac{Var[\mathcal{O}]}{\|\overline{\mathcal{O}}\|_1^2}\right) \ . \tag{102}$$

Unfortunately this condition can bee too loose to be of practical value as we can see by looking at the limit $Var[\mathcal{O}] \to 0$: the approximate expression Eq. (97) recovers the correct limit of Eq. (33), while the right hand side of Eq. (102) is always larger than $f(k)^K/\epsilon_r$ and this can produce an overly pessimistic assessment of the efficiency gain achievable with the sQPE schemes of Sec. II.

To get more insight on this problem we plot in Fig. 11 the regions in a two-dimensional $(Var[\mathcal{O}], \epsilon_r)$ space where the inequalities above predict an advantage of our proposed scheme for the estimation of the deuteron's ground state energy. The solid black lines define the maximum relative error ϵ_r achievable for a trial state $|\Psi\rangle$ with a given variance obtained using the condition of Eq. (97). The simple linear algorithm of Sec. II B is predicted to be efficient for situations that stay in the top left corner of parameter space bounded by the first black line. For lower target ϵ_r or larger variance $Var[\mathcal{O}]$ we then progressively need to increase the order K of the algorithm to ensure Eq. (97) is satisfied. From Fig. 11 we also see that for a 1% target error the simple linear scheme with $K = 1$ is predicted to be more efficient up to $Var[\mathcal{O}] \approx E_{GS}^2$ before the growth of the bias term forces us to increase the order in K. This is encouraging since this condition is not necessarily tight in the sense that, due to the use of the upperbound Eq. (96), the inequality of Eq. (97)

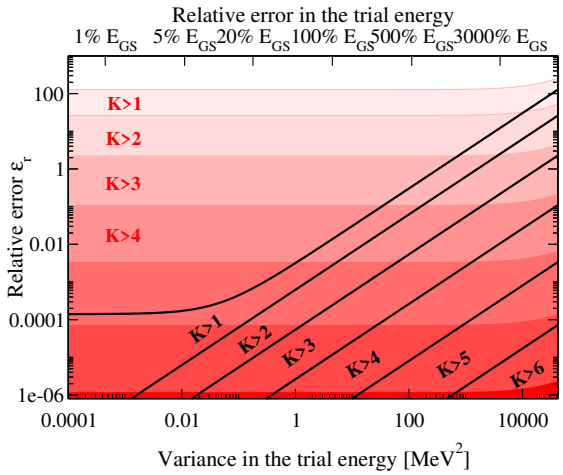


FIG. 11. Estimated parameter regions for problems expected to be efficiently solvable with the scheme of Sec. II as predicted by Eq. (97) (black solid lines) and by the looser condition in Eq. (102) (red contours). See text for more details.

is a sufficient but not necessary condition for Eq. (93) to hold true. The extent to witch one can still find an efficiency gain by using the linear algorithm past this condition will most likely depend on the particular problem instance and is therefore difficult to predict without some prior knowledge on $Cov[\mathcal{O}^{2K}, \mathcal{O}]$.

The red contours in Fig. 11 are obtained instead by using the looser condition Eq. (102) and as we can see they are overly pessimistic: a relatively high order algorithm with $K > 4$ is judged to be needed even in the limiting case when the trial state has zero variance. This observation reinforces the importance of being able to use a tighter condition like Eq. (97) in order to assess meaningfully the possibility of a gain in using the strategy proposed in this work.

B. Time evolution

We finally turn our attention to the problem of estimating the circuit depth required for the implementation of the time-evolution unitary $U_\tau = e^{i\tau\mathcal{O}}$ needed for the sQPE method. In particular we consider the more realistic situation where only an approximation $\widetilde{U_\tau}$ of U_τ with error bounded by δ_τ is available:

$$\|\widetilde{U_\tau} - U_\tau\| \leq \delta_\tau \ . \tag{103}$$

In this situation the induced error on the sQPE estimator Eq. (20) is then $\epsilon_\tau = \delta_\tau/\tau$. A simple way to control the total error ϵ of the calculation is to require that $\epsilon_\tau < \epsilon/2$ and similarly for the MSE ϵ_M in Eq. (23). The latter modification will increase the bound reported in Eq. (26) only by a factor $2^{1+1/K}$. This requirement can be relaxed if one employs optimal algorithms like [55–57] that allow

to implement $\widetilde{U_\tau}$ for $\epsilon_\tau \ll \epsilon$ with only a small increase in gate count.

Due to the relatively short propagation time required by the sQPE algorithm, we can also use simpler strategies based on the Trotter–Suzuki decomposition [58] while still maintaining a short circuit depth. To see how this works let us start by considering a simple first order scheme obtained by dividing the propagation time τ into r segments:

$$T(\tau, r) = e^{i\tau\alpha_0}\left[\prod_{k=1}^{L} e^{i\frac{\tau}{r}\alpha_k U_k}\right]^r , \tag{104}$$

where we used the decomposition in Eq. (8) for the operator $\mathcal{O}$. As shown in [59] the error for this approximation can be bounded by

$$\|T(\tau, r) - U_\tau\| \leq \frac{\left(\tau\|\overline{\mathcal{O}}\|_1\right)^2}{r} e^{\frac{\tau}{r}\|\overline{\mathcal{O}}\|_1} , \tag{105}$$

with the norm $\|\overline{\mathcal{O}}\|_1 = |\alpha_0| + \|\overline{\mathcal{O}_T}\|_1$ defined as in Eq. (11). Note that this is slightly tighter than the result Proposition F.3 obtained in [59] and can be found following the same proof. Following the derivation in [59] we can now provide the following analytic bound for the number of intervals r required to guarantee that $\epsilon_\tau < \epsilon/2$

$$r_1 = \left\lceil \rho_1 \max\left[1, \frac{2e}{\epsilon}\|\overline{\mathcal{O}}\|_1\right]\right\rceil , \tag{106}$$

where we have defined

$$\rho_1 = \tau_{opt}\|\overline{\mathcal{O}}\|_1 = \gamma(K)\frac{\|\overline{\mathcal{O}}\|_1}{|m_K|^{\frac{1}{2K}}}\epsilon^{\frac{1}{2K}} , \tag{107}$$

and rewrote the optimal time step Eq. (25) as

$$\tau_{opt} = \gamma(K)\left(\frac{\epsilon}{|m_K|}\right)^{\frac{1}{2K}} , \tag{108}$$

whith the definition

$$\gamma(K) \equiv \left(\frac{(2K+1)!}{2\sqrt{2K+1}}\right)^{\frac{1}{2K}} . \tag{109}$$

We note that the additional factor of 2 in the denominator is coming from the choice $\epsilon_M = \epsilon/2$ and that and that for reasonably small errors the bound Eq. (106) is maximized with the rightmost expression giving the algorithm an overall depth scaling at best as $r_1 = O\left(1/\sqrt{\epsilon}\right)$.

It is instructive to express these bounds in terms of the relative error ϵ_r and the expectation value ratio R_O from Eq. (15). For instance in the case of eigenvalue estimation we find

$$\rho_1 = \gamma(K)\frac{\|\overline{\mathcal{O}}\|_1}{\|\overline{\mathcal{O}_T}\|_1}\frac{\epsilon_r^{\frac{1}{2K}}}{R_O} , \tag{110}$$

which leads to a generic scaling given by

$$r_1 = O\left(\gamma(K)\frac{\|\overline{\mathcal{O}}\|_1^2}{\|\overline{\mathcal{O}_T}\|_1^2}\frac{\epsilon_r^{\frac{1-2K}{2K}}}{R_O^2}\right) , \tag{111}$$

which in the linear case of Sec. II B simplifies to

$$r_1 = O\left(\frac{\|\overline{\mathcal{O}}\|_1^2}{\|\overline{\mathcal{O}_T}\|_1^2}\frac{1}{R_O^2}\frac{1}{\sqrt{\epsilon_r}}\right) . \tag{112}$$

This provides only a minor advantage over the $O\left(1/\epsilon\right)$ scaling associated with full fledged QPE algorithms [10] which can be easily spoiled with a sufficiently small R_O ratio. It is therefore important to use higher order expansions that are able to achieve a more favourable scaling and in the following we will consider higher order product formulas as an example. If we denote the $(2j)$-th order Trotter-Suzuki formula with r intervals [58, 59] as $S_{2j}(\tau, r)$ we can generalize the error bound Eq. (105) obtained above to

$$\|S_{2j}(\tau, r) - U_\tau\| \leq \frac{\left(2\tau 5^{j-1}\|\overline{\mathcal{O}}\|_1\right)^{2j+1}}{3r^{2j}} e^{2\frac{\tau}{r}5^{j-1}\|\overline{\mathcal{O}}\|_1} , \tag{113}$$

and bound the number of intervals as

$$r_j = \left\lceil \rho_j \max\left[1, \left(\frac{4e}{3\epsilon}5^{j-1}\|\overline{\mathcal{O}}\|_1\right)^{\frac{1}{2j}}\right]\right\rceil , \tag{114}$$

with $\rho_j \equiv 2\rho_1 5^{j-1}$. Again this is slightly tighter than the result obtained by Childs et al. in the Supplemental Material of [59]. As for the linear decomposition the right term dominates for reasonably small errors ϵ and we find the overall scaling

$$r_j = O\left(5^{j+\frac{1}{2j}}\frac{\gamma(K)}{|m_K|^{\frac{1}{2K}}}\|\overline{\mathcal{O}}\|_1^{1+\frac{1}{2j}}\epsilon^{\frac{j-K}{4jK}}\right) . \tag{115}$$

As above we can express this in terms of relative quantities in a compact way for the special case of eigenvalue estimation

$$r_j = O\left(5^{j+\frac{1}{2j}}\gamma(K)\left(\frac{\|\overline{\mathcal{O}}\|_1}{\|\overline{\mathcal{O}_T}\|_1}\right)^{1+\frac{1}{2j}}\frac{\epsilon_r^{\frac{j-K}{4jK}}}{R_O^{1+\frac{1}{2j}}}\right) , \tag{116}$$

and due to the fast growth of the first term in the above expression we might want to keep the order j as small as possible. For instance using $j = K$ will already guarantee a gate count independent on the target precision ϵ and scaling as $O\left(1/R_O^{3/2}\right)$ in terms of the eigenvalue ratio. Notably by simply choosing $j = K + 1$, for the price of a fixed increase in cost of less than a factor of 5 we can achieve a circuit depth that decreases as a function of the target relative error.

Note that these estimates are based on the possibly very pessimistic bounds in Eqs. (105,113) which means that these circuit depths could possibly be greatly reduced in practice (see eg. [59]). Before finishing this section we want to point out that even though the estimates

provided above are for the implementation of U_τ a complete implementation of its controlled version (needed in Eq. (18)) can be obtained with an overall linear increase in depth as a function of the number of qubits in the system register used to represent $|\Psi\rangle$. Tighter bounds will require further knowledge of the particular operator $\mathcal{O}$ whose time evolution we want to simulate, but the results presented here give us good reasons to believe practical implementations could be achievable for interesting systems on near term devices.

VI. SUMMARY AND CONCLUSIONS

In this work we reviewed the standard methodology of Operator Averaging [5, 6, 18] to evaluate expectation values $\langle\mathcal{O}\rangle$ of general Hermitian operators efficiently on quantum computers by minimizing the number of quantum operations needed while maintaining a shot-noise limited number of measurement $N_{tot} = O\left(1/\epsilon^2\right)$. This provides a great advantage on current generation noisy devices where the asymptotically optimal behaviour $N_{tot} = O\left(1/\epsilon\right)$ of methods that employ Quantum Phase Estimation [10, 12] cannot be attained in practice due to the large circuit depths $C_{QPE} = O\left(1/\epsilon\right)$ involved. As we explain in Sec. I however, the Operator Average strategy has a major drawback in that in terms of relative error ϵ_r the total measurement count grows as $N_{tot} = O\left(1/(\epsilon_r R_O)^2\right)$ where R_O defined in Eq. (15) is approximately the ratio between the wanted expectation value and the largest eigenvalue of $\mathcal{O}$.

In this work we propose to use a single step of phase estimation as in the well known Hadamard Test to learn the expectation value $\langle\mathcal{O}\rangle$ by looking at the short time behaviour of $\langle sin(\tau\mathcal{O})\rangle$ instead. This strategy was already discussed in the context of full QPE calculations in [10] and it remains the method of choice for fully error-corrected devices capable of executing accurately very long gate sequences. Our contribution is in showing how, by using circuits implementing only a single Hadamard Test with appropriately chosen time-steps, one can greatly reduce the classical cost (the number of experimental measurements N_{tot}) using much shorter circuits than those needed for QPE. For instance in the important case of eigenvalue estimation we can achieve $N_{tot} = O(1/\epsilon_r^{2+1/K})$ for $K = O(1)$ independent on R_O while keeping the gate count bounded by $C_{sQPE} = O\left(5^K/R_O^{1+1/2K}\right)$ using a very simple general purpose strategy employing the high-order Trotter-Suzuki decomposition [58, 59]. As we argue in Sec. V B the latter requirement can possibly be further reduced by using more advanced simulation strategies [55–57] and we plan to further this possibility along the lines of the study presented in [59] in a future work.

We presented a complete analysis of the first two lowest order sQPE algorithms with $K = 1$ and $K = 2$ in Sec. II together with a self consistent procedure aimed at finding the optimal time-steps to be used in the calculation. As our approach could be extremely helpful in some situations but it is not efficient in an asymptotic scaling sense in general, we provide both strict and easy to estimate conditions to help determine if the use of sQPE con provide a speedup for a particular problem instance (see Sec. II and Sec. V A). As these conditions require the availability of bounds on the expectation value to be computed and some control over the operator spectrum (like bounds on the n-th cumulant $\langle\mathcal{O}^n\rangle$) further work on classically efficient strategies to estimate them (using for instance ideas from [60, 61]) could have a possible big impact on the practicality of our approach. As discussed in Sec. V A classical Quantum Monte Carlo simulations could be employed efficiently in the meantime. Finally in Sec. IV we have shown some evidence on the robustness of our proposed methodology to readout noise on the quantum device and provided arguments to justify the expectation that ancilla based algorithms like sQPE provide in general a much more robust layout to deploy and execute non trivial quantum algorithms on NISQ devices. It will be very interesting in the future to see the impact of adaptive machine-learning techniques as those presented in [52] on the practical feasibility of scaling up quantum computations in the near term.

ACKNOWLEDGMENTS

We wish to thank N. Klco and R. Schiavilla for helpful discussions regarding the subject of this work. We also thank A. Matsuura of Intel Labs for useful conversations and for encouragements. The work of A.R. was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) quantum algorithm teams program under field work proposal number ERKJ333 and by U.S. Department of Energy grant No. DE-FG02-00ER41132. We are also grateful to the T2 group at Los Alamos National Laboratory for the hospitality while working on this project thanks in part to funds from the U.S. Department of Energy, Office of Science, HEP Contract No.DE-KA2401032. A.B. acknowledges support from U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Award No. DE-SC0010300 and DE-SC0019647.

Appendix A: Implementation of controlled time-evolution

We report in this section the implementation of the two-qubit controlled time evolution appearing in Eq. (18) needed for the sQPE algorithm and schematically presented in Eq. (81) in the main text. Throughout this section we assume that the initial state $|\Psi\rangle$ has been prepared with a rotation $R_y(\theta)$ with θ the angle of interest. For a system of two qubits a generic controlled unitary operation associated with a 2×2 unitary matrix U can be represented in the computational basis $|00\rangle$, $|01\rangle$,$|10\rangle$, $|11\rangle$ as

$$C_U = \begin{pmatrix} 1_{2\times2} & 0_{2\times2} \\ 0_{2\times2} & U \end{pmatrix} , \tag{A1}$$

with $1_{2\times2}$ and $0_{2\times2}$ indicating the two-by-two identity matrix null matrix respectively.

Let's now recall the general decomposition of a $U(2)$ unitary

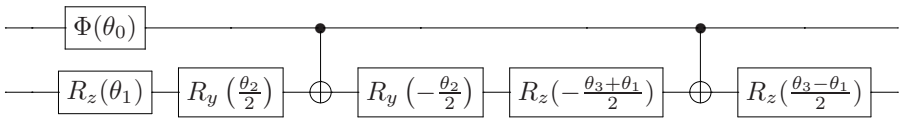
$$U = e^{i\theta_0} R_z(\theta_1) R_y(\theta_2) R_z(\theta_3) \tag{A2}$$

for appropriately chosen angles. The rotation matrices here are defined as

$$R_z(\phi) = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \tag{A3}$$

$$R_y(\theta) = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} . \tag{A4}$$

Using this decomposition, together with the definition Eq. (A1), we can implement the controlled time-evolution operator with the following circuit (see eg. [37])



(A5)

where we have defined the phase gate as

$$\Phi(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} . \tag{A6}$$

Using the following decomposition of the Hamiltonian matrix

$$H = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \equiv \frac{\alpha+\gamma}{2}\mathbb{1} + \beta X + \frac{\alpha-\gamma}{2} Z \tag{A7}$$

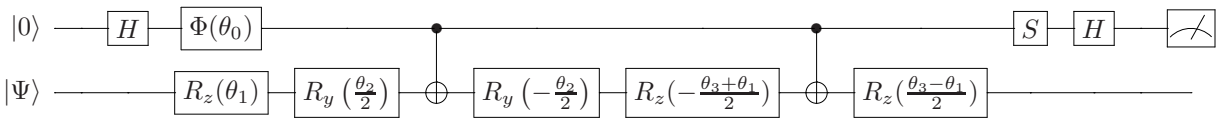
with X, Z Pauli spin matrices, $\mathbb{1}$ the identity matrix and (α, β, γ) real numbers, we can write the exact time-propagator as

$$e^{i\delta H} = e^{i\delta\frac{\alpha+\gamma}{2}} \left[cos(\theta) + i\hat{\theta}\cdot\sigma sin(\theta) \right] \tag{A8}$$

with

$$\vec{\theta} = (\delta\beta, 0, \delta\frac{\alpha-\gamma}{2}) \quad \hat{\theta} = \frac{\vec{\theta}}{\theta} . \tag{A9}$$

From this expression we can easily determine the needed angles $(\theta_0, \ldots, \theta_3)$. The full circuit for sQPE is then



(A10)

and the last rotation can be avoided.

[1] J. Preskill, Quantum **2**, 79 (2018).
[2] R. P. Feynman, International Journal of Theoretical Physics **21**, 467 (1982).
[3] S. Lloyd, Science **273**, 1073 (1996).
[4] P. W. Shor and S. P. Jordan, Quantum Information & Computation **8**, 681 (2008).
[5] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. OBrien, Nature Communications **5**, 4213 (2014).

[6] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).
[7] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, J. M. Brink, Markus Chow, and J. M. Gambetta, Nature **549**, 242 (2017).
[8] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Phys. Rev. A **95**, 020501 (2017).
[9] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Phys. Rev. A **64**, 022319 (2001).
[10] E. Knill, G. Ortiz, and R. D. Somma, Phys. Rev. A **75**, 012328 (2007).
[11] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**, 339 (1998).
[12] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **83**, 5162 (1999).
[13] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, arXiv e-prints , quant-ph/0005055 (2000), arXiv:quant-ph/0005055 [quant-ph].
[14] S. Bravyi, G. Smith, and J. A. Smolin, Phys. Rev. X **6**, 021043 (2016).
[15] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).
[16] E. Farhi, J. Goldstone, and S. Gutmann, arXiv e-prints , arXiv:1411.4028 (2014), arXiv:1411.4028 [quant-ph].
[17] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum **3**, 140 (2019).
[18] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, The Journal of Physical Chemistry Letters **5**, 4368 (2014).
[19] A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, Chem. Sci. **10**, 3746 (2019).
[20] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Quantum Science and Technology **4**, 014008 (2018).
[21] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, Phys. Rev. C **51**, 38 (1995).
[22] H. Chernoff, *Sequential Analysis and Optimal Design* (SIAM Publications, 1972).
[23] T. J. Loredo, AIP Conference Proceedings **707**, 330 (2004).
[24] R. Kosut, I. A. Walmsley, and H. Rabitz, arXiv e-prints , quant-ph/0411093 (2004), arXiv:quant-ph/0411093 [quant-ph].
[25] J. Nunn, B. J. Smith, G. Puentes, I. A. Walmsley, and J. S. Lundeen, Phys. Rev. A **81**, 042109 (2010).
[26] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012).
[27] G. Ballo´ and K. M. Hangos, IFAC Proceedings Volumes **44**, 4344 (2011), 18th IFAC World Congress.
[28] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New Journal of Physics **14**, 103013 (2012).
[29] T. Schulte-Herbrüggen, A. Spörl, N. Khaneja, and S. J. Glaser, Phys. Rev. A **72**, 042331 (2005).
[30] The upperbound on τ here is somewhat arbitrary since the time step will get readjusted anyway.
[31] M. Garcon and J. W. Van Orden, Adv. Nucl. Phys. **26** (2001), 10.1007/0-306-47915-X_4.
[32] R. B. Wiringa and S. C. Pieper, Phys. Rev. Lett. **89**, 182501 (2002).
[33] The exact matrix elements with the Av6' potential are real, the integer representation used here is however a very good approximation and simplifies the discussion.
[34] S. K. Bogner, R. J. Furnstahl, and R. J. Perry, Phys. Rev. C **75**, 061001 (2007).
[35] H. Hergert and R. Roth, Phys. Rev. C **75**, 051001 (2007).
[36] E. R. Anderson, S. K. Bogner, R. J. Furnstahl, and R. J. Perry, Phys. Rev. C **82**, 054001 (2010).
[37] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[38] N. Shenvi, K. R. Brown, and K. B. Whaley, Phys. Rev. A **68**, 052313 (2003).
[39] J. Kołodyński and R. Demkowicz-Dobrzański, Phys. Rev. A **82**, 053804 (2010).
[40] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, Nature Communications **3**, 1063 (2012), arXiv:1201.3940 [quant-ph].
[41] F. Fröwis, P. Sekatski, W. Dür, N. Gisin, and N. Sangouard, Rev. Mod. Phys. **90**, 025004 (2018).
[42] E. Knill and R. Laflamme, Phys. Rev. Lett. **81**, 5672 (1998).
[43] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim, D. Bucher, F. J. Cabrera-Hernádez, J. Carballo-Franquis, A. Chen, C.-F. Chen, J. M. Chow, A. D. Córcoles-Gonzales, A. J. Cross, A. Cross, J. Cruz-Benito, C. Culver, S. D. L. P. González, E. D. L. Torre, D. Ding, E. Dumitrescu, I. Duran, P. Eendebak, M. Everitt, I. F. Sertage, A. Frisch, A. Fuhrer, J. Gambetta, B. G. Gago, J. Gomez-Mosquera, D. Greenberg, I. Hamamura, V. Havlicek, J. Hellmers, Ł. Herok, H. Horii, S. Hu, T. Imamichi, T. Itoko, A. Javadi-Abhari, N. Kanazawa, A. Karazeev, K. Krsulich, P. Liu, Y. Luh, Y. Maeng, M. Marques, F. J. Martín-Fernández, D. T. McClure, D. McKay, S. Meesala, A. Mezzacapo, N. Moll, D. M. Rodríguez, G. Nannicini, P. Nation, P. Ollitrault, L. J. O'Riordan, H. Paik, J. Pérez, A. Phan, M. Pistoia, V. Prutyanov, M. Reuter, J. Rice, A. R. Davila, R. H. P. Rudy, M. Ryu, N. Sathaye, C. Schnabel, E. Schoute, K. Setia, Y. Shi, A. Silva, Y. Siraichi, S. Sivarajah, J. A. Smolin, M. Soeken, H. Takahashi, I. Tavernelli, C. Taylor, P. Taylour, K. Trabing, M. Treinish, W. Turner, D. Vogt-Lee, C. Vuillot, J. A. Wildstrom, J. Wilson, E. Winston, C. Wood, S. Wood, S. Wörner, I. Y. Akhalwaya, and C. Zoufal, "Qiskit: An open-source framework for quantum computing," (2019).
[44] K. Temme, S. Bravyi, and J. M. Gambetta, Phys. Rev. Lett. **119**, 180509 (2017).
[45] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X **8**, 031027 (2018).
[46] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge Univ. Press, 2010).
[47] I. L. Chuang and M. A. Nielsen, Journal of Modern Optics **44**, 2455 (1997).
[48] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White, Phys. Rev. Lett. **90**, 193601 (2003).
[49] M. Mohseni and D. A. Lidar, Phys. Rev. Lett. **97**, 170501 (2006).
[50] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[51] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B.

Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).
[52] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Phys. Rev. A **89**, 042314 (2014).
[53] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[54] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Rev. Mod. Phys. **87**, 1067 (2015).
[55] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Phys. Rev. Lett. **114**, 090502 (2015).
[56] G. Hao Low and I. L. Chuang, ArXiv e-prints (2016), arXiv:1610.06546 [quant-ph].
[57] G. H. Low and I. L. Chuang, Phys. Rev. Lett. **118**, 010501 (2017).
[58] M. Suzuki, Journal of Mathematical Physics **32**, 400 (1991).
[59] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Proceedings of the National Academy of Sciences **115**, 9456 (2018).
[60] T. Baumgratz and M. B. Plenio, New Journal of Physics **14**, 023027 (2012).
[61] A. W. Harrow and A. Montanaro, Quantum **1**, 6 (2017).