Mitigating algorithmic errors in a Hamiltonian simulation

Suguru Endo,[1,*] Qi Zhao,[2] Ying Li,[3] Simon Benjamin,[1] and Xiao Yuan[1,†]

[1]*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*
[2]*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China*
[3]*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

Quantum computers can efficiently simulate many-body systems. As a widely used Hamiltonian simulation tool, the Trotter-Suzuki scheme splits the evolution into the number of Trotter steps N and approximates the evolution of each step by a product of exponentials of each individual term of the total Hamiltonian. The algorithmic error due to the approximation can be reduced by increasing N, which however requires a longer circuit and hence inevitably introduces more physical errors. In this work, we first study such a trade-off and numerically find the optimal number of Trotter steps N_{opt} given a physical error model in a near-term quantum hardware. Practically, physical errors can be suppressed using recently proposed error mitigation methods. We then extend physical error mitigation methods to suppress the algorithmic error in Hamiltonian simulation. By exploiting the simulation results with different numbers of Trotter steps $N \leqslant N_{opt}$, we can infer the exact simulation result within a higher accuracy and hence mitigate algorithmic errors. We numerically test our scheme with a five-qubit system and show significant improvements in the simulation accuracy by applying both physical and algorithmic error mitigations.

I. INTRODUCTION

It is hard to simulate quantum systems using a classical computer, as the computational cost increases exponentially with the system size. Such a problem can be resolved by quantum simulation, as proposed by Feynman in 1982 [1], saying "let the computer itself be built of quantum mechanical elements which obey quantum mechanical laws." Quantum simulation of many-body systems has become one of the most promising applications of quantum computing. Given the Hamiltonian H of a system, a vital step is to realize the time-evolution operator $U(t) = e^{-iHt}$, which can be used for studying both its dynamic [2] and its static [3] properties.

Several methods have been proposed to efficiently approximate the time evolution operator $U(t)$ [4–9]. Although the latest methods [6,7,9] have been significantly improved, the simulation accuracy is still limited by finite resources, such as short circuit depth, finite system runtime, and large physical errors in the system. To study such a limitation, we focus on the the Trotterization method [10], introduced for quantum simulation by Lloyd [2]. Suppose the system Hamiltonian H can be decomposed into a sum of Hamiltonians, H_k, that only involves few-body interactions, i.e., $H = \sum_k H_k$. Then,

the Trotterization method approximates the time-evolution unitary operator $U(t) = e^{-i \sum_k H_k t}$ by decomposing it into a product form,

$$U(t) = \left(\prod_k e^{-i H_k t/N} \right)^N + O(t^2/N). \tag{1}$$

Here, N is the number of Trotter steps and $O(t^2/N)$ is the algorithmic error due to a finite value of N. As H_k only has local interactions, each term $e^{-iH_k t/N}$ can be efficiently realized on a quantum computer. However, as H_k terms generally do not commute with each other, Trotterization only approximates the time-evolution operator $U(t)$. By increasing the number of Trotter steps N, the algorithmic error $O(t^2/N)$ can be arbitrarily suppressed. However, the circuit depth increases linearly with the number of Trotter steps. A deep circuit introduces more physical errors, which corrupt quantum simulations in noisy intermediate-scale devices [11]. Consequently, we can only use a small number of Trotter steps for systems without quantum error correction [12].

Recently, several error mitigation methods have been introduced to suppress physical errors in shallow circuits [13–20]. One of the methods relies on extrapolation [14–16], which works by deliberately increasing the error rate of the quantum hardware, and using the expectation values of several points with higher error rates to infer the error-free value. The extrapolation method can be applied to suppress general physical errors that can be well controlled in shallow circuits. However, it fails to work when the circuit depth is too long, which is when many errors happens. As such, even with error mitigation, we cannot choose too large a number of Trotter steps. As a result, even if physical errors can be mitigated, the accuracy of the simulations will still be limited by algorithmic errors.

*suguru.endo@oriel.ox.au.uk
†xiao.yuan.ph@gmail.com

In our work, we study algorithmic and physical errors resulting from implementing Trotterization on near-term devices. Following Ref. [12], in Sec. II we first find the optimal number of Trotter steps N_{opt} for error-prone quantum simulations. In Sec. III, we review recently proposed error mitigation methods and show how physical errors can be suppressed. In Sec. IV, we then extend the physical error extrapolation method to suppress the algorithmic error in Trotterization. We use expectation values obtained from different numbers of Trotter steps N_1 and N_2 ($N_{opt} \geqslant N_1, N_2$) and extrapolate the results to estimate a more accurate expectation value than can be obtained using only the optimal number of Trotter steps N_{opt}. In Sec. V, we numerically test our result by considering a five-qubit Hamiltonian. We numerically show the optimal number of Trotter steps N_{opt} under an inhomogeneous Pauli error model and apply both physical and algorithmic error mitigation methods to significantly increase the simulation accuracy. In Sec. VI, we conclude our work and discuss its possible extension in general quantum information processing.

II. THE OPTIMAL NUMBER OF TROTTER STEPS FOR NOISY QUANTUM SIMULATION

In this section, we first review the theoretical analysis [12] about the optimal number of Trotter steps in a Hamiltonian simulation with physical noise. Denote the channel of the Trotter decomposition $e^{-iH_k t/N}$ as $\mathcal{V}_k$ and the physical noise as an extra channel, $\mathcal{E}_k$, then the noisy stroboscopic channel for the ith Trotter step is

$$\mathcal{E}_i^{strobo} = \mathcal{E}_L \circ \mathcal{V}_L \circ \mathcal{E}_{L-1} \circ \mathcal{V}_{L-1} \circ \cdots \circ \mathcal{E}_1 \circ \mathcal{V}_1, \quad (2)$$

where L is the number of the local Hamiltonians. With the number of Trotter steps N, the entire noisy channel of the Trotter decomposition is

$$\mathcal{E}^{noisyTrotter} = \mathcal{E}_N^{strobo} \circ \mathcal{E}_{N-1}^{strobo} \circ \cdots \circ \mathcal{E}_1^{strobo}. \quad (3)$$

The distance between the two channels $\mathcal{E}_1$ and $\mathcal{E}_2$ is defined by the trace distance or distinguishability between the output states of the two channels [21,22],

$$D(\mathcal{E}_1, \mathcal{E}_2) = \max_{\rho} \|\mathcal{E}_1(\rho) - \mathcal{E}_2(\rho)\|, \quad (4)$$

where the maximization is over a properly chosen state set and $\|M\| = \text{Tr}[\sqrt{M^\dagger M}]$ for matrix M. The distance between the ideal channel $\mathcal{U}^{ideal}$ for the evolution e^{-iHt} and the noisy implementation $\mathcal{E}^{noisyTrotter}$ is

$$D(\mathcal{U}^{ideal}, \mathcal{E}^{noisyTrotter}) \leqslant \sum_{i=1}^{N} D(\sqrt[N]{\mathcal{U}^{ideal}}, \mathcal{E}_i^{strobo}),$$

$$\leqslant \sum_{i=1}^{N} D(\sqrt[N]{\mathcal{U}^{ideal}}, \sqrt[N]{\mathcal{V}^{nonoise}})$$

$$+ D(\mathcal{E}_i^{strobo}, \sqrt[N]{\mathcal{V}^{nonoise}}), \quad (5)$$

where the second line follows from the chaining property $D(\mathcal{E}_1 \circ \mathcal{E}_2, \mathcal{E}_1' \circ \mathcal{E}_2') \leqslant D(\mathcal{E}_1, \mathcal{E}_1') + D(\mathcal{E}_2, \mathcal{E}_2')$, the third line follows from the triangle inequality $D(\mathcal{E}_1, \mathcal{E}_2) \leqslant D(\mathcal{E}_1, \mathcal{E}_3) + D(\mathcal{E}_2, \mathcal{E}_3)$, and $\sqrt[N]{\mathcal{V}^{nonoise}}$ is the channel of the noise-free

stroboscopic evolution $\prod_k e^{-iH_k t/N}$ of each Trotter decomposition. Now, we define the algorithmic and physical errors ε_{alg} and ε_{phys} as

$$\varepsilon_{alg} = \sum_{i=1}^{N} D(\sqrt[N]{\mathcal{U}^{ideal}}, \sqrt[N]{\mathcal{V}^{nonoise}}) = \frac{\alpha}{N},$$

$$\varepsilon_{phys} = \sum_{i=1}^{N} D(\mathcal{E}_i^{strobo}, \sqrt[N]{\mathcal{V}^{nonoise}}) = \beta N, \quad (6)$$

where $\alpha = D(\sqrt[N]{\mathcal{U}^{ideal}}, \sqrt[N]{\mathcal{V}^{nonoise}})N^2$ and $\beta = D(\mathcal{E}^{strobo}, \sqrt[N]{\mathcal{V}^{nonoise}})$. Here, for simplicity, we assume that the noise model is the same for each stroboscopic sequence; that is, $\mathcal{E}_i^{strobo}$ is the same for different i. Note that $\varepsilon_{alg} \propto 1/N$ is because the algorithmic error can be linearly suppressed with an increasing number of Trotter steps; while $\varepsilon_{phys} \propto N$ is because errors linearly accumulate with a larger number of Trotter steps and hence longer circuits. By optimizing the distance

$$D(\mathcal{U}^{ideal}, \mathcal{E}^{noisyTrotter}) = \frac{\alpha}{N} + \beta N, \quad (7)$$

we can get the optimized number of Trotter steps as

$$N_{opt} = \sqrt{\alpha/\beta}, \quad (8)$$

with the corresponding trace distance $D = 2\sqrt{\alpha\beta}$. Therefore, due to the existence of physical errors, we cannot choose an infinitely large number of Trotter steps to suppress the algorithmic error. Although it is not easy to analytically calculate α and β for a general physical Hamiltonian and noise models, we numerically show the optimal number of Trotter steps in Sec. V. In the following, we first show how to suppress the physical errors ε_{phys} with the recently proposed physical error mitigation methods. Then, we extend the methods to suppress the algorithmic error ε_{alg}.

III. ERROR MITIGATION AND EXTRAPOLATION TECHNIQUE

Now, we show how to suppress physical errors with the recently proposed error mitigation methods [13–20]. Especially, we focus on the Richardson and exponential extrapolation error mitigation methods [14–16]. Due to imperfections of gate operations, such as decoherence, errors can accumulate in the quantum circuit, so that the noisy output state ρ_{noise} becomes

$$\rho_{noise} = \mathcal{N}_{N_{tot}} \circ \mathcal{G}_{N_{tot}} \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{G}_1(\rho_{init}), \quad (9)$$

where the ideal operation can be expressed as $\mathcal{G}_N \circ \mathcal{G}_{N-1} \cdots \mathcal{G}_1(\rho_{init})$. Here, $\mathcal{N}_k$ is the noise channel accompanying the kth ideal gate operation $\mathcal{G}_k$, N_{tot} is the number of gates, and ρ_{init} is the initial input state for the quantum circuit. The noisy output state can be corrected with fault-tolerant error correction that utilizes extra qubits to detect errors and correct the state. However, fault-tolerant error correction is considerably costly and is hard to realize with current quantum hardware. For noisy intermediate-scale quantum devices with a restricted number of qubits, error mitigation methods require no extra qubit and can suppress errors with simple postprocessing of different runs of the quantum circuits.

Here we focus on the quantum computation tasks of which the final goal is to calculate the expectation value of a certain observable, $\hat{A}$. Suppose the error-free circuit output state ρ and the noiseless expectation value is $\langle \hat{A} \rangle = \mathrm{Tr}(\hat{A}\rho)$. For a noisy circuit with error strength ε, the noisy output state ρ_ε deviates from the ideal noiseless state. For example, for stochastic errors

$$\mathcal{N}_k = (1-p)\mathcal{I} + p\mathcal{N}_k', \tag{10}$$

we can set $\varepsilon \propto p$, where $\mathcal{I}$ is the identity map, $\mathcal{N}_k'$ is a noise map, and $p \in [0,1]$ denotes the strength of the noise. The noisy expectation value $\langle \hat{A} \rangle(\varepsilon)$ can be regarded as a function of ε and it can be expanded according to different orders of ε,

$$\langle \hat{A} \rangle(\varepsilon) = \langle \hat{A} \rangle(0) + \sum_{j=1}^{n} A_j \varepsilon^j + O(\varepsilon^{n+1}), \tag{11}$$

where A_j is the jth derivative of $\langle \hat{A} \rangle(\varepsilon)$ in the Taylor expansion and $\langle \hat{A} \rangle(0)$ is the noiseless expectation value. For a considerably small ε, we have $\langle \hat{A} \rangle(\varepsilon) \approx \langle \hat{A} \rangle(0)$. With several different noisy expectations $\langle \hat{A} \rangle(\varepsilon)$ for different ε, error mitigation is to infer the noiseless expectation value $\langle \hat{A} \rangle(0)$ to a higher accuracy.

A. Richardson extrapolation

Suppose the expectation value of the observable $\langle \hat{A} \rangle(\varepsilon)$ is measured for several rescaled noise rates $a_j \varepsilon$ with $a_0 = 1 < a_1 < a_2 < \cdots < a_n$, which can be achieved by increasing the physical noise [23]. Then we can estimate $\langle \hat{A} \rangle(0)$ by the Richardson extrapolation method [24] discussed in Refs. [14,15]:

$$
\begin{aligned}
\langle \hat{A} \rangle_{\mathrm{est}}(0) &= \sum_{i=0}^{n} \gamma_i \langle \hat{A} \rangle(a_i \varepsilon), \\
&= \sum_{i=0}^{n} \gamma_i \langle \hat{A} \rangle(0) + \sum_{j=1}^{n} A_j \varepsilon^j \sum_{i=0}^{n} \gamma_i a_i^j + O(\varepsilon^{n+1}) \\
&= \langle \hat{A} \rangle(0) + O(\varepsilon^{n+1}),
\end{aligned} \tag{12}
$$

where γ_i is chosen such that $\sum_{i=0}^{n} \gamma_i = 1$ and $\sum_{i=0}^{n} \gamma_i a_i^j = 0$ for $j = 1, 2, \ldots, n$. By extrapolating n points with different error strengths, we can accurately estimate $\langle \hat{A} \rangle(0)$ and suppress the error to $O(\varepsilon^{n+1})$. When $n = 1$, we call it linear or two-point extrapolation; when $n = 2$, we call it second-order Richardson or three-point extrapolation. The variance of the estimation $\langle \hat{A} \rangle_{\mathrm{est}}(0)$ is

$$\mathrm{Var}[\langle \hat{A} \rangle_{\mathrm{est}}(0)] = \sum_{i=0}^{n} \gamma_i^2 \mathrm{Var}[\langle \hat{A} \rangle_{\mathrm{est}}(a_i \varepsilon)]^2. \tag{13}$$

Suppose the variance $\mathrm{Var}[\langle \hat{A} \rangle_{\mathrm{est}}(a_i \varepsilon)]^2$ values are similar for different error strengths, the variance of the estimation $\langle \hat{A} \rangle_{\mathrm{est}}(0)$ is $\sum_{i=1}^{n} \gamma_i^2$ times larger than the variance of each measurement $\hat{A}(\varepsilon)$. Therefore, to achieve the same shot noise of each measurement $\hat{A}(\varepsilon)$, we need to run the circuit $\sum_{i=1}^{n} \gamma_i^2$ more times and we denote

$$\Gamma_{\mathrm{phys}} = \sum_{i=1}^{n} \gamma_i^2 \tag{14}$$

as the cost of physical error mitigation. Note that, Γ_{phys} generally increases exponentially to n [25]. We can thus only choose a small constant value of n in practice in order to avoid such an exponentially increasing cost.

For example, the cost of linear extrapolation with error rates ε and $r\varepsilon$ ($r > 1$) under the assumption $\mathrm{Var}[\langle \hat{A} \rangle(\epsilon)] = \mathrm{Var}[\langle \hat{A} \rangle(r\epsilon)]$ is

$$\Gamma_{\mathrm{phys}} = \frac{1 + r^2}{(1 - r)^2}, \tag{15}$$

which is a monotonically decreasing function of r. The resource cost can be reduced with a larger r, however with an increase of the estimation error. Therefore, one needs to optimize r by taking into account the shot noise due to finite samples and the error due to the extrapolation method.

B. Exponential extrapolation

The extrapolation method in the previous section is equivalent to using a polynomial function to fit the values with different error strengths. The optimal fitting function is not necessarily a polynomial function. In Ref. [16], exponential extrapolation was introduced by fitting an exponential function and was shown to be able to suppress more errors than the Richardson extrapolation. When two error rates, ϵ and $r\epsilon$ ($r > 1$), are used, the estimation value via exponential extrapolation is

$$\langle \hat{A} \rangle_{\mathrm{est}}(0) = \langle \hat{A} \rangle(\epsilon)^{\frac{r}{r-1}} \langle \hat{A} \rangle(r\epsilon)^{\frac{1}{1-r}}. \tag{16}$$

As the performance of two-point exponential extrapolation has been shown to be adequate in Ref. [16] and in our simulation results in Sec. V, we only use two-point exponential extrapolation here. The intuitive reason that the exponential function is a suitable function for extrapolation is as follows. Following the stochastic model in Eq. (10) and considering a circuit with identity gates, the noise process of the quantum circuit can be written as

$$\prod_{k=1}^{N_{\mathrm{tot}}} \mathcal{N}_k = \prod_{k=1}^{N_{\mathrm{tot}}} \{(1-p)[I] + p\mathcal{N}_k'\} \tag{17}$$

$$= \sum_{m=0}^{N_{\mathrm{tot}}} q_m \mathcal{K}_m, \tag{18}$$

where N_{tot} is the number of the gate in the quantum circuit, $q_m = \binom{N_{\mathrm{tot}}}{m}(1-p)^{N_{\mathrm{tot}}-m} p^m$, and $\mathcal{K}_m$ is the average of the terms that have m errors. In the limit of $N_{\mathrm{tot}} \to \infty$ and $N_{\mathrm{tot}} p \to \mathrm{const}$, the binomial distribution q_m can be approximated by the Poisson distribution $q_m \approx e^{-N_{\mathrm{tot}} p}(N_{\mathrm{tot}} p)^m / m!$, and the channel is

$$\prod_{k=1}^{N_{\mathrm{tot}}} \mathcal{N}_k \approx e^{-N_{\mathrm{tot}} p} \sum_{m=0}^{N_{\mathrm{tot}}} \frac{(N_{\mathrm{tot}} p)^m}{m!} \mathcal{K}_m. \tag{19}$$

As the channel is proportional to $e^{-N_{\mathrm{tot}} p}$, the exponential function may be a better function for extrapolation.

By assuming that $\langle \hat{A} \rangle(\epsilon) \propto e^{-N_{\mathrm{tot}} \epsilon}$, and $\mathrm{Var}[\langle \hat{A} \rangle(\epsilon)] = \mathrm{Var}[\langle \hat{A} \rangle(r\epsilon)]$, we can get the variance of the estimation:

$$\mathrm{Var}[\langle \hat{A} \rangle_{\mathrm{est}}(0)] = \frac{(r^2 e^{2N_{\mathrm{tot}} \epsilon} + e^{2N_{\mathrm{tot}} r\epsilon})}{(r-1)^2} \mathrm{Var}[\hat{A}(\epsilon)]^2. \tag{20}$$

Thus the cost for the exponential extrapolation is

$$\Gamma_{\text{phys}} = \frac{(r^2 e^{2N_{\text{tot}}\epsilon} + e^{2N_{\text{tot}}r\epsilon})}{(r-1)^2}. \quad (21)$$

Due to the exponential dependence of r and $N_{\text{tot}}\epsilon$, the sampling cost of exponential extrapolation can be much larger than that of linear extrapolation for large r and $N_{\text{tot}}\epsilon$. However, error mitigation generally works in the regime where $N_{\text{tot}}\epsilon \leqslant 1$, and therefore the cost of linear and exponential extrapolations are similar when $r \approx 1$. In practice, exponential extrapolation has been shown to outperform linear extrapolation in a 19-qubit numerical simulation under inhomogeneous and leakage errors [16].

IV. ERROR MITIGATION FOR ALGORITHMIC ERRORS

In this section, we discuss how the algorithmic error in Trotterization can be regarded as an analogy of the physical error and how to extend the extrapolation error mitigation method to suppress the algorithmic error. Given a Hamiltonian H that has decomposition $H = \sum_k H_k$, the first-order Trotter formula approximates the time-evolution operator $U(t) = e^{-i\sum_k H_k t}$ by decomposing it into a product form,

$$U(t) = \left(\prod_k e^{-iH_k t/N} \right)^N + \sum_{i<j} [H_i, H_j] t^2/2N + \sum_{m=3}^{\infty} E(m), \quad (22)$$

where the higher-order terms $E(m)$ can be upper bounded by $\|E(m)\| \leqslant N \|Ht/N\|^m/m!$, and $\|\cdot\|$ denotes the maximal eigenvalue of the matrix. Denote $\varepsilon_N = 1/N$ and $U_{\varepsilon_N}(t)$ as

$$U_{\varepsilon_N}(t) \equiv \left(\prod_k e^{-iH_k t \varepsilon_N} \right)^{1/\varepsilon_N}, \quad (23)$$

then we can straightforwardly see how $U_{\varepsilon_N}(t)$ approximates $U(t) = e^{-i\sum_k H_k t}$, i.e.,

$$\lim_{\varepsilon_N \to 0^+} U_{\varepsilon_N}(t) = U(t) = e^{-i\sum_k H_k t}. \quad (24)$$

Suppose the time-evolution operator $U_{\varepsilon_N}(t)$ is applied to an initial state $|\psi_0\rangle$, then the output state is $U_{\varepsilon_N}(t)|\psi_0\rangle$. When we measure the observable A of the final output state, the average value is

$$\langle \hat{A}(t) \rangle (\varepsilon_N) = \langle \psi_0| U_{\varepsilon_N}(t)^\dagger \hat{A} U_{\varepsilon_N}(t) |\psi_0\rangle, \quad (25)$$

while the error-free expectation value for the observable is $\langle \hat{A}(t) \rangle (0) = \langle \psi|_0 e^{iHt} \hat{A} e^{-iHt}(t) |\psi\rangle_0$. Regarding $\langle \hat{A}(t) \rangle (\varepsilon_N)$ as a function of ε_N, we can expand $\langle \hat{A}(t) \rangle (\varepsilon_N)$ as a function of ε_N by using the Taylor expansion

$$\langle \hat{A}(t) \rangle (\varepsilon_N) = \langle \hat{A}(t) \rangle (0) + \sum_{j=1}^{n} \varepsilon_N^j \hat{A}(t)_j + O(\varepsilon_N^{j+1}), \quad (26)$$

where $\langle \hat{A}(t) \rangle (0)$ and $\hat{A}(t)_j$ are independent of ε_N.

Therefore, the extrapolation error mitigation method can be applied to suppress the algorithmic error. In the original error mitigation scheme, we need to boost the error rate ε to several different error rates $a_j \varepsilon$, which can be realized by intensionally adding more noise to the circuit. Here, as the

error rate ε_N is $1/N$, we can straightforwardly increase the algorithmic error ε_N with a smaller number of Trotter steps. For example, by taking $N' = N/2$, we can effectively double the algorithmic error $\varepsilon_{N'} = 2\varepsilon_N$. With n different numbers of Trotter steps $N_n \leqslant \cdots \leqslant N_1 \leqslant N_0 = N_{\text{opt}}$ and $N_j = N_0/a_j$, we can therefore suppress the algorithmic error by a linear combination of the results:

$$\langle \hat{A}(t) \rangle_{\text{est}} (0)$$

$$= \sum_{i=0}^{n} \gamma_i' \langle \hat{A}(t) \rangle (\varepsilon_{N_i})$$

$$= \sum_{i=0}^{n} \gamma_i' \langle \hat{A}(t) \rangle (0) + \sum_{j=1}^{n} \hat{A}(t)_j \varepsilon_{N_{\text{opt}}}^j \sum_{i=0}^{n} \gamma_i' a_i^j + O(\varepsilon_{N_{\text{opt}}}^{n+1})$$

$$= \langle \hat{A}(t) \rangle (0) + O(\varepsilon_{N_{\text{opt}}}^{n+1}). \quad (27)$$

Here, γ_i' is chosen such that $\sum_{i=0}^{n} \gamma_i' = 1$ and $\sum_{i=0}^{n} \gamma_i' a_i^j = 0$ for $j = 1, 2, \ldots, n$. Therefore, we can suppress the algorithmic error to an order of $O((\varepsilon_{N_{\text{opt}}})^{n+1}) = O((1/N_{\text{opt}})^{n+1})$. The details of the upper bound for the algorithmic error are shown in the Appendix.

The variance of the estimation $\langle \hat{A} \rangle_{\text{est}} (0)$ is

$$\text{Var}[\langle \hat{A} \rangle_{\text{est}} (0)] = \sum_{i=0}^{n} \gamma_i'^2 \text{Var}[\langle \hat{A} \rangle_{\text{est}} (\varepsilon_{N_i})]^2. \quad (28)$$

Thus, by assuming that the variance is the same for different $\langle \hat{A} \rangle_{\text{est}} (\varepsilon_{N_i})$, the variance of the estimation $\text{Var}[\langle \hat{A} \rangle_{\text{est}} (0)]$ is $\sum_{i=0}^{n} \gamma_i'^2$ times larger than the variance of $\langle \hat{A} \rangle_{\text{est}} (\varepsilon_{N_i})$. We denote the cost of the extrapolation for the algorithmic error by

$$\Gamma_{\text{alg}} \equiv \sum_{i=0}^{n} \gamma_i'^2. \quad (29)$$

Moreover, we can combine the extrapolation for physical errors with the extrapolation for algorithmic errors, and the total cost is

$$\Gamma_{\text{alg+phys}} \equiv \Gamma_{\text{alg}} \Gamma_{\text{phys}}. \quad (30)$$

Note that the only requirement for the extrapolation method is that $U_{\varepsilon_N}(t)$ can be expressed as an explicit function of ε_N, and $\lim_{\varepsilon_N \to 0^+} U_{\varepsilon_N}(t) = U(t)$. Therefore, the same argument can be applied to the higher-order Trotterization [4]. Furthermore, this method can still be applied even if the expectation value cannot be efficiently expanded as a function of $1/N$. The method also works as long as $\langle \hat{A} \rangle$ can be expanded with ε, which is a function $f(N_t)$ of the tunable parameter N_t.

V. NUMERICAL SIMULATION

In this section, we consider a five-qubit system and numerically test our algorithmic error mitigation method for simulating real time evolution. As shown in Fig. 1, the Hamiltonian only has local and near-neighborhood interactions,

$$H = J \sum_{i=1}^{5} Z_3 Z_i + B \sum_{i=1}^{5} X_i, \quad (31)$$

where X_i (Z_i) denotes the spin-$\frac{1}{2}$ Pauli x (z) operator acting on the ith qubit, $J = 3$, and $B = 2$. In our numerical
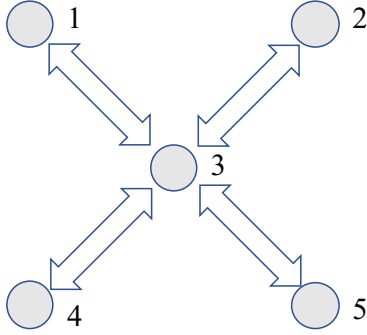
FIG. 1. Schematic of the five-qubit Hamiltonian.

simulation, we start with initial state $|\psi_0\rangle = |0, 0, 0, 0, 0\rangle$, evolve it with time t, and measure $\hat{A} = X_1$. We consider both algorithmic error from the finite number of Trotter steps and physical error from gate noises. We consider inhomogeneous Pauli error for both single- and two-qubit gates,

$$\mathcal{E}(\rho) = (1 - p)\rho + p_x X\rho X + p_y Y\rho Y + p_z Z\rho Z, \quad (32)$$

where $p = p_x + p_y + p_z$. For a single-qubit gate, we set $p_x = p_y = 2.0 \times 10^{-5}$ and $p_z = 6.0 \times 10^{-5}$; for a two-qubit gate, we take the error channel $\mathcal{E}_2 = \mathcal{E} \otimes \mathcal{E}$ and set $p_x = p_y = 1.0 \times 10^{-4}$ and $p_z = 3.0 \times 10^{-4}$ for each $\mathcal{E}$. Note that such a noise rate corresponds to the current state-of-the-art experiment system [26,27].

Without considering gate error and shot noise to the measurement, one should use the infinite large number of Trotter steps to increase the simulation accuracy. With gate errors, the optimal number of Trotter steps is limited as show in Fig. 2(a). Here, we fix the total evolution time as $t = 0.5$ and numerically find the optimal number of Trotter steps to be 25.

To suppress the algorithmic error, we use two or three different numbers of Trotter steps to infer the value corresponding to the infinite number of Trotter steps. For a given number of Trotter steps N, the runtime of the circuit is related to the circuit depth, which is proportional to N, and the circuit repetition time m, which is used to get an accurate estimation of the measurement. We denote the total runtime resource cost M as $M = mN$. To compare different simulation scenarios, we thus consider the same total cost M for a fair comparison. For three- and two-point extrapolation, we divide M equally to different numbers of Trotter steps. For example, for three-Trotter-step extrapolation, $N_1^{(3)}$, $N_2^{(3)}$, and $N_3^{(3)}$, we have $m_i^{(3)} = M/(3N_i^{(3)})$ for $i = 1, 2$, and 3; for two-Trotter-step extrapolation, $N_1^{(2)}$, $N_2^{(2)}$, $m_i^{(2)} = M/(2N_i^{(2)})$ for $i = 1$ and 2; and for one number of Trotter steps $N_1^{(1)}$, $m_1^{(1)} = M/N_1^{(1)}$.
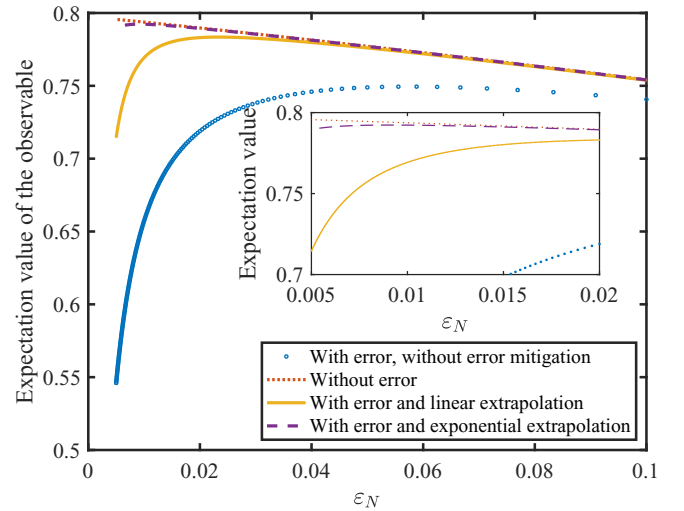
In our simulation, we compare the three cases for suppressing algorithmic errors: no error mitigation, linear extrapolation, and three-point extrapolation. In order to quantify the performance of our simulation method, we evaluate the error of the estimation value $\langle \hat{A}(t) \rangle (0)_{\text{est}}$ by

$$\delta^2 = [\langle \hat{A}(t) \rangle (0) - \langle \hat{A}(t) \rangle (0)_{\text{est}}]^2, \quad (33)$$

where $\langle \hat{A}(t) \rangle (0)$ is the error-free average value.
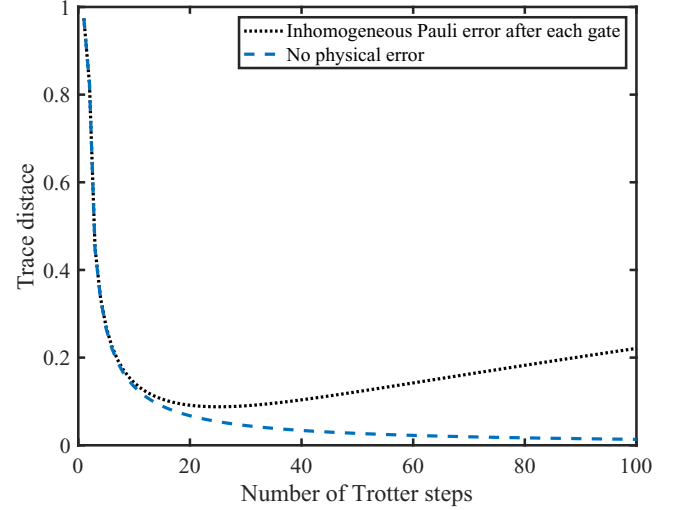
FIG. 2. The optimal number of Trotter steps N and the continuity of the measurement as a function of the inverse of the number of Trotter steps $\varepsilon_N = 1/N$. (a) The trace distance between the ideal state and the simulated state in the presence of noise. The blue curve denotes the trace distance without physical error, which goes to zero with an infinite large number of Trotter steps. The black curve denotes the trace distance with the inhomogeneous Pauli error after each gate and the optimal number of Trotter steps is 25. (b) The expectation value of the observable versus $\varepsilon_N = 1/N$. Here N is the number of Trotter steps. The horizontal axis corresponds to the number of Trotter steps 10–200. From the simulation result, we can confirm the continuity of the function.

Now, we show that physical and algorithmic errors can be suppressed with the error mitigation methods. To begin with, we check the continuity of $\langle \hat{A}(t) \rangle$ as a function of $\varepsilon_N = 1/N$, to show that $\langle \hat{A}(t) \rangle$ can be Taylor expanded by ε_N. We consider the case with and without physical error mitigation. We measure the expectation values with the original error probability p and the twice-boosted error probability $2p$. Then by applying the linear and exponential extrapolation methods, we can suppress the physical errors. As shown in
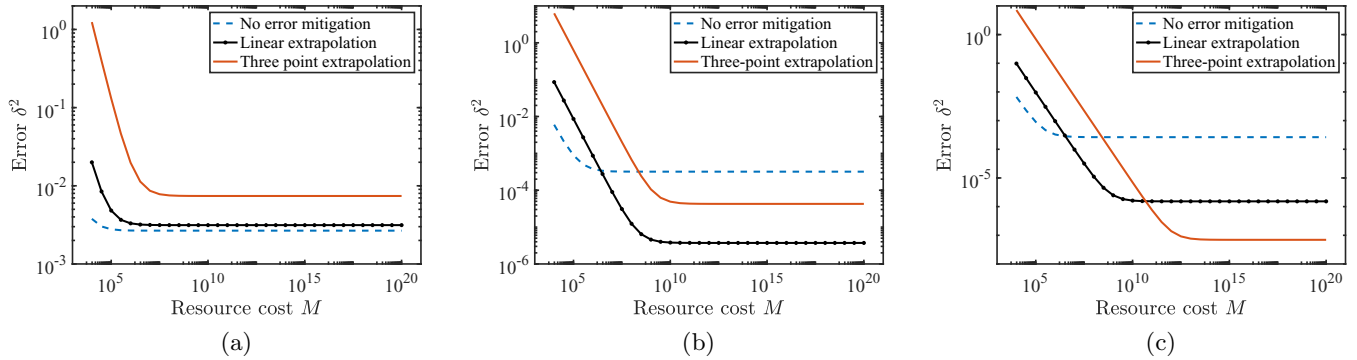
FIG. 3. The error δ^2 against the total runtime resource M for different cases with no algorithmic error mitigation, linear algorithmic error extrapolation, and three-point algorithmic error extrapolation. We consider physical errors to the circuit and set the total evolution time as $t = 0.5$. The number of Trotter steps $N = 25$, 20, and 15; $N = 25$ and 15; and $N = 25$ are used for the three-point, linear extrapolation, and no error mitigation cases, respectively. The blue line denotes the no error mitigation case, the black line denotes the linear algorithmic error extrapolation case, and the red line denotes the three-point algorithmic error extrapolation case. (a) No physical error mitigation is applied. (b) Linear extrapolation is applied to suppress physical errors. (c) Exponential extrapolation is applied to suppress physical errors. Here, we set the ratio r of the two error rates to be 2 for both linear and exponential extrapolation.

Fig. 2(b), we can see that $\langle \hat{A}(t) \rangle$ is indeed a continuous function of ε_N, which confirms the possibility of algorithmic error mitigation. Furthermore, the best accuracies achieved by linear and exponential extrapolation methods are $\delta^2 = 2 \times 10^{-4}$ and 2.79×10^{-5}, with the number of Trotter steps being $N = 43$ and 109, respectively. This clearly shows that we cannot increase the number of Trotter steps infinitely even when the extrapolation method is applied.

To show the effect of algorithmic error mitigation, we also consider three cases for physical errors: no error mitigation, linear extrapolation, and exponential extrapolation. First, we consider the case where no error mitigation is employed for the physical error. In such a case, physical errors still dominate and we find that algorithmic error extrapolation cannot improve the simulation accuracy, as shown in Fig. 3(a). This is because, although we suppress the algorithmic error by linearly combining the results from different numbers of Trotter steps, the large deviation due to physical errors makes the estimation worse. As the total resource M increases, the shot noise is suppressed and the accuracy converges. The accuracy δ^2 at the converged point without algorithmic and physical error mitigations is 2.67×10^{-3}.

Next, we consider the cases where linear extrapolation is applied to suppress physical errors. Subsequently, we apply the algorithmic error extrapolation to suppress algorithmic errors due to Trotterization. As shown in Fig. 3(b), we find that the algorithmic linear extrapolation outperforms the no error mitigation and three-point extrapolation cases for large M. The converged accuracy with sufficiently large $M \gtrsim 10^8$ is $\delta^2 = 3.72 \times 10^{-6}$ under linear extrapolation of algorithmic errors. The improvement of the accuracy is 717 times, compared with the case where no error mitigation is applied for both physical and algorithmic errors.

Finally, in Fig. 3(c), we plot the result for the case where exponential extrapolation is used for suppressing physical errors. It can be seen that the physical error is successfully reduced, and three-point extrapolation works properly, surpassing the performance of linear extrapolation for large $M \gtrsim 10^{13}$. The converged accuracy δ^2 for sufficiently large

M is 6.95×10^{-8}. This accuracy corresponds to 3.8×10^4 times improvement, compared with the case where no error mitigation is applied for both physical and algorithmic errors.

VI. DISCUSSION

In this work, we propose an error mitigation method for suppressing algorithmic errors in Trotterization of Hamiltonian simulation. We first show that the optimal number of Trotter steps is finite due to physical errors. Then, we show how the recently proposed physical error mitigation methods can be extended to suppress algorithmic errors in Trotterization. We numerically test our algorithmic error mitigation method in a five-qubit Hamiltonian and show how it can improve simulation accuracy by combining it with physical error mitigation. Although we only focus on the first-order Trotterization, our scheme can also be extended to other Trotterization schemes, such as higher-order Trotterization [4] and randomization Trotterization [5]. This is because the expectation values obtained from these methods can also be written as a function of the number of Trotter steps. Although, in the presence of physical errors, higher-order Trotter decompositions may not reduce the overall error [14], we leave these extensions in future works. Moreover, the other recently proposed Hamiltonian simulation methods, e.g., Taylor series [6] and quantum signal processing [7,8], offer alternative methods of Hamiltonian simulation instead of Trotterization. These methods divide the simulation time t into N segments and simulate the evolution for each time t/N. Considering $1/N$ as an error, our extrapolation method can also be applied by modifying N and extrapolating the results with different N to improve the estimation accuracy.

The extrapolation method has broad applications in diverse fields, including error mitigation, computational chemistry, linear optics simulation, etc. In classical computational chemistry, extrapolation is widely used in solving molecular structure problems and in Monto Carlo simulation of dynamics [28,29]. In linear optics simulation, the extrapolation method is used to simulate single photons with imperfect

photon sources such as coherent states from lasers [25]. We can also apply this formalism for mitigating the algorithmic error to other schemes, such as variational quantum eigensolver (VQE) [13,17,30–34]. The result obtained from VQE can be regarded as a function of the depth of the quantum circuit, and the extrapolation method can be applied. For example, considering the hardware efficient ansatz [32] with different circuit depths, one can get simulation results with different accuracy. Therefore, by running experiments with different small depths, the extrapolation method can be applied to infer results with a much larger depth. However, when the algorithm gets trapped in local minima, the error mitigation method will fail to work. We leave the discussion about the practical performance to future works. A similar argument can also be applied to other types of variational algorithms, for simulating real or imaginary time evolutions [14,35,36].

APPENDIX: ERROR BOUND FOR RICHARDSON EXTRAPOLATION

Suppose a Hamiltonian H can be decomposed into the sum of local Hamiltonians, H_k, as $H = \sum_k H_k$. Then, by using the first-order Trotter formula, we can approximate the time-evolution unitary operators as

$$U(t) = \exp(-iHt) = \left[\prod_i \exp(-iH_i t/N) \right]^N \quad \text{(A1)}$$

$$+ \sum_{i<j}[H_i, H_j]t^2/2N + \sum_{m=3}^{\infty} E(m), \quad \text{(A2)}$$

where $E(m)$ can be upper bounded by $\|E(m)\| \leqslant N\|Ht/N\|^m/m!$, and $\|\cdot\|$ denotes the maximal eigenvalue. We denote U_N as $[\prod_i \exp(-iH_i t/N)]^N$, $\varepsilon_N = 1/N$, and $U_{\varepsilon_N}(t) \equiv (\prod_k e^{-iH_k \varepsilon_N})^{1/\varepsilon_N}$, which converges to $U(t)$ with $\varepsilon_N \to 0^+$. The difference between $U(t)$ and $U_{\varepsilon_N}(t)$ can be rewritten as

$$U(t) = U_{\varepsilon_N}(t) + a/N + E_N, \quad \text{(A3)}$$

where $a = \sum_{i<j}[H_i, H_j]t^2/2$ and $E_N = \sum_{m=3}^{\infty} E(m)$, which can be bounded by

$$\|E_N\| \leqslant N \sum_{m=3}^{\infty} \|Ht/N\|^m/m! < \frac{\|H\|^3 t^3}{6N^2} e^{\|H\|t/N}. \quad \text{(A4)}$$

Suppose the time-evolution operator $U_{\varepsilon_N}(t)$ is applied to an initial state $|\psi_0\rangle$, then the output state is $U_{\varepsilon_N}(t)|\psi_0\rangle$. When we measure the observable $\hat{A}$ of the final output state, the average value is

$$\langle \hat{A}(t) \rangle (\varepsilon_N) = \langle \psi_0 | U_{\varepsilon_N}(t)^\dagger \hat{A} U_{\varepsilon_N}(t) | \psi_0 \rangle, \quad \text{(A5)}$$

and the average value for evolution $U(t)$ is $\langle \psi_0 | e^{iHt} \hat{A} e^{-iHt} | \psi_0 \rangle$. The relationship between them can be expressed as

$$\begin{aligned}
\langle \hat{A}(t) \rangle (\varepsilon_N) &= \langle \hat{A}(t) \rangle (0) - 1/N \\
&\quad \times [\langle \psi_0 | U^\dagger \hat{A} a | \psi_0 \rangle + \langle \psi_0 | a^\dagger \hat{A} U | \psi_0 \rangle] \\
&\quad + \langle \psi_0 | (a/N + E_N)^\dagger \hat{A}(a/N + E_N) | \psi_0 \rangle \\
&\quad - [\langle \Psi | U^\dagger \hat{A} E_N | \Psi \rangle + \langle \Psi | (E_N)^\dagger \hat{A} U | \Psi \rangle] \\
&= \langle \hat{A}(t) \rangle (0) + b/N + R_N, \quad \text{(A6)}
\end{aligned}$$

where $b = -[\langle \psi_0 | U^\dagger \hat{A} a | \psi_0 \rangle + \langle \psi_0 | a^\dagger \hat{A} U | \psi_0 \rangle]$ is independent of N, and $R_N = \langle \psi_0 | (a/N + E_N)^\dagger \hat{A}(a/N + E_N) | \psi_0 \rangle - [\langle \psi_0 | U^\dagger \hat{A} E_N | \psi_0 \rangle + \langle \psi_0 | (E_N)^\dagger \hat{A} U | \psi_0 \rangle]$ is bound by

$$R_N \leqslant \left[\left(\frac{\|a\|}{N} + \frac{\|H\|^3 t^3}{6N^2} e^{\|H\|t/N} \right)^2 + \frac{\|H\|^3 t^3}{3N^2} e^{\|H\|t/N} \right] \|\hat{A}\|. \quad \text{(A7)}$$

It is not hard to see that $R_N = O(1/N^2)$.

Now we use the linear combination of $\langle \hat{A}(t) \rangle (\varepsilon_{N_1})$ and $\langle \hat{A}(t) \rangle (\varepsilon_{N_2})$ to estimate $\langle \hat{A}(t) \rangle (0)$ by

$$\langle \hat{A}(t) \rangle_{\text{est}} (0) = \beta_1 \langle \hat{A}(t) \rangle (\varepsilon_{N_1}) + \beta_2 \langle \hat{A}(t) \rangle (\varepsilon_{N_2}). \quad \text{(A8)}$$

We choose the parameters β_1 and β_2 satisfying $\beta_1/N_1 + \beta_2/N_2 = 0$ and $\beta_1 + \beta_2 = 1$, so that the difference between $\langle \hat{A}(t) \rangle_{\text{est}} (0)$ and $\langle \hat{A}(t) \rangle (0)$ is

$$| \langle \hat{A}(t) \rangle_{\text{est}} (0) - \langle \hat{A}(t) \rangle (0) | \leqslant R_{N_1} \beta_1 + R_{N_2} \beta_2. \quad \text{(A9)}$$

Based on the two-point extrapolation method, we use the simulation results with Trotter steps N_1 and N_2, which both are $O(N)$ to reduce the algorithmic error from $O(1/N)$ to $O(1/N^2)$. Similarly, by applying extrapolation with m different points, we can suppress the Trotter algorithmic error to an order of $O(1/N^m)$.

[1] R. P. Feynman, Int. J. Theor. Phys. **21**, 467 (1982).

[2] S. Lloyd, Science **273**, 1073 (1996).

[3] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science **309**, 1704 (2005).

[4] M. Suzuki, J. Math. Phys. **32**, 400 (1991).

[5] A. M. Childs, A. Ostrander, and Y. Su, arXiv:1805.08385.

[6] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Phys. Rev. Lett. **114**, 090502 (2015).

[7] G. H. Low and I. L. Chuang, arXiv:1610.06546.

[8] G. H. Low and I. L. Chuang, Phys. Rev. Lett. **118**, 010501 (2017).

[9] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Proc. Natl. Acad. Sci. U.S.A. **115**, 9456 (2018).

[10] M. Suzuki, Commun. Math. Phys. **51**, 183 (1976).

[11] J. Preskill, Quantum **2**, 79 (2018).

[12] G. C. Knee and W. J. Munro, Phys. Rev. A **91**, 052327 (2015).

[13] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Phys. Rev. A **95**, 042308 (2017).

[14] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).

[15] K. Temme, S. Bravyi, and J. M. Gambetta, Phys. Rev. Lett. **119**, 180509 (2017).

[16] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X **8**, 031027 (2018).

[17] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Phys. Rev. X **8**, 011021 (2018).

[18] M. Otten and S. Gray, arXiv:1806.07860.

[19] S. McArdle, X. Yuan, and S. Benjamin, arXiv:1807.02467.

[20] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Phys. Rev. A **98**, 062339 (2018).

[21] C. A. Fuchs, arXiv:quant-ph/9601020.

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

[23] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, arXiv:1805.04492.

[24] L. F. Richardson, B. J Arthur Gaunt *et al.*, Philos. Trans. R. Soc. London, Ser. A **226**, 299 (1927).

[25] X. Yuan, Z. Zhang, N. Lütkenhaus, and X. Ma, Phys. Rev. A **94**, 062305 (2016).

[26] T. P. Harty, D. T. C. Allcock, C. J. Ballance, L. Guidoni, H. A. Janacek, N. M. Linke, D. N. Stacey, and D. M. Lucas, Phys. Rev. Lett. **113**, 220501 (2014).

[27] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, Phys. Rev. Lett. **117**, 060504 (2016).

[28] L. Brualla, K. Sakkos, J. Boronat, and J. Casulleras, J. Chem. Phys. **121**, 636 (2004).

[29] M. Suzuki, Phys. Lett. A **113**, 299 (1985).

[30] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien, Nat. Commun. **5**, 4213 (2014).

[31] P. J. J. O'Malley, Phys. Rev. X **6**, 031007 (2016).

[32] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature (London) **549**, 242 (2017).

[33] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New J. Phys. **18**, 023023 (2016).

[34] D. Wecker, M. B. Hastings, and M. Troyer, Phys. Rev. A **92**, 042303 (2015).

[35] S. McArdle, T. Jones, S. Endo, Y. Li, S. Benjamin, and X. Yuan, arXiv:1804.03023.

[36] S. Endo, T. Jones, S. McArdle, X. Yuan, and S. Benjamin, arXiv:1806.05707.