Energy Extrapolation in Quantum Optimization Algorithms

Chenfeng Cao,1,∗ Yunlong Yu,2,3 Zipeng Wu,1 Nic Shannon,4 Bei Zeng,1,† and Robert Joynt5,3,‡

1 Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
2 State Key Laboratory of Low Dimensional Quantum Physics, Department of Physics, Tsinghua University, Beijing 100084, China
3 Kavli Institute for Theoretical Sciences, University of Chinese Academy of Sciences, Beijing, China
4 Theory of Quantum Matter Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0412, Japan
5 Department of Physics, University of Wisconsin–Madison, 1150 University Avenue, Madison, Wisconsin 53706, USA

(Dated: December 28, 2021)

Quantum optimization algorithms are promising to find the ground state of a target Hamiltonian on near-term quantum devices. However, it is necessary to limit the evolution time or the circuit depth as much as possible since otherwise decoherence will degrade the computation. Even when this is done, there always exists a non-negligible estimation error in the ground state energy. Here we show a scalable extrapolation approach to mitigate this error. With an appropriate regression, we can significantly improve the estimation accuracy for quantum annealing, variational quantum eigensolver and quantum imaginary time evolution for fixed evolution time or circuit depth. The inference is achieved by extrapolating the annealing time to infinity or extrapolating the variance to zero. For Hamiltonians that only involve few-body interactions, the additional overhead is an increase in the number of measurements by a constant factor. We verified the validity of our method via numerical simulation and experiments on an IBM quantum computer. The method is robust to noise, and the techniques are applicable to other physics problems. Analytic derivations for the quadratic convergence feature of the residual energy in quantum annealing and the linear convergence feature of energy variance are given. Our work paves a promising new way for enhancing near-term quantum computing with classical post-processing.

I. INTRODUCTION

Obtaining a precise estimate of the ground state energy of a many-body system Hamiltonian is a central task in modern quantum physics and chemistry [1]. Classical algorithms have had considerable success by using, for example, tensor network methods, but for large systems with generic long-range entanglement the task remains challenging. There is hope that eventually fault-tolerant quantum computers will be able to deal with such problems more accurately and efficiently, but before that, we probably have to stay in the noisy-intermediate-scale-quantum (NISQ) era for a long time [2, 3]. Researchers have developed several quantum algorithms for ground state preparation that can be implemented on existing NISQ processors, such as quantum annealing (QA) [4–6], variational quantum eigensolver (VQE) [7–9], and quantum imaginary time evolution (QITE) [10–13]. Each of these methods shows promise, but is limited by the resources available to support them.

QA is a heuristic approach to find ground states of complex quantum systems inspired by the adiabatic theorem. It starts from the ground state of a simple Hamiltonian $H_{\text{init}}$, then slowly evolves the system to the ground state of $H_f$, the problem Hamiltonian. The required annealing time grows rapidly with the inverse of the minimum spectral gap during the evolution and may become very long, which can render accurate QA calculations very difficult on NISQ annealers.

Another class of NISQ algorithms that attracted particular attention in recent years is the variational quantum algorithms [7–9, 14–22]. Unlike QA, these algorithms also require a classical computer to do an optimization that complements the quantum processing. VQE is the prime example of this class for ground state preparation. We prepare an initial product state, evolve the state with a hardware-efficient parameterized quantum circuit, then iteratively update the parameters to minimize the Hamiltonian expectation value. If the parameterized circuit is capable of preparing the exact ground state, the minimized value equals the ground state energy $E_{gs}$.

VQE has achieved great success in electronic structure calculations [23]. However, its applications to large systems are severely hampered by two daunting obstacles. One is the “barren plateaus” phenomenon. The gradient of the cost function vanishes exponentially with the increase of system size for a deep random quantum circuit [24, 25]. The other is the hardware noise, which may accumulate exponentially and also induce another kind of “barren plateau” [26]. The community has developed error mitigation techniques to mitigate the effect of hardware noise [27–30], but the overhead scales exponentially with circuit depth [31]. Consequently, VQE is only reliable for shallow quantum circuits. Schemes to enhance shallow-depth VQE are highly desired.

∗ chenheng.ca@connect.ust.hk
† zengb@ust.hk
‡ rjoynt@wisc.edu
QITE is another NISQ algorithm proposed recently [10, 11]. It was designed to simulate the imaginary time evolution and thus prepare ground states on a quantum computer. QITE splits the imaginary time process into many steps via the Trotter-Suzuki decomposition and approximates each non-unitary step by unitary evolution. It has also been used to calculate other system properties such as the correlation functions [12]. Unfortunately, compared with VQE, QITE is even less robust to quantum noise. Errors that occurred in a previous step severely affect the accuracy of a latter one. We can only reliably run QITE on real devices with very few Trotter steps.

One question that naturally arises is whether we can infer $E_{gs}$ with shorter evolution times for QA or lower depth for VQE/QITE? We take a clue from a general picture of the state probabilities on an energy level diagram as the quantum resources increase. This is shown in Fig. 1. With the increase of resources, the state obtained in the calculation has a larger overlap with the ground state and smaller overlaps with the excited states. Accordingly, the energy converges to $E_{gs}$, and the variance converges to 0. This pattern suggests two possible extrapolations to the limit of infinite resources when the probabilities condense entirely to the ground state. For QA, we can extrapolate the total energy of the system to an asymptotic value by running QA repeatedly with different annealing times, which we call the “time method”. Alternatively, for all quantum optimization algorithms, we can measure the energy and the variance repeatedly, plot the variance as a function of the energy, and extrapolate to zero variance. This is the “variance method”.

Similar extrapolation techniques have been widely used in computational and nuclear physics [32–34]. Physicists performed extrapolation of energies to zero variance for an eigenstate in quantum Monte Carlo methods [35, 36]. In quantum computing, researchers developed extrapolation-based schemes to mitigate the effect of hardware noise [27, 28] and Trotter error [37]. We note that the technique to enhance VQE with variance extrapolation was independently mentioned in Ref. [38], but without a discussion on circuit depth, noise effect, and experiment.

Quantum noise is the central challenge in the NISQ era. Both time and variance methods are robust to measurement errors. For variance extrapolation, when the gate error rates are within an appropriate range, the extrapolated energy is even more accurate than that in the noiseless case. Due to the efficiency and noise robustness, our methods have great potential in the NISQ era.

The paper is structured as follows. In Sec. II, we introduce our methods and analyze the noise effect. In Secs. III, IV, V, we numerically and experimentally demonstrate the accuracy improvement in QA, VQE, and QITE. The conclusions and future directions are summarized and discussed in Sec. VI. Analytic justification for the quadratic convergence feature of the residual energy in QA is given in Appendix A. Analytic deviation for the linear convergence feature of energy variance is given in Appendix B.

II. METHODS

A. Time Method

In the QA framework, we start from the ground state of a simple initial Hamiltonian $H_{init}$ and evolve the system according to the time-dependent Hamiltonian

$$H(t) = f(t)H_{init} + g(t)H_f,$$

where $H_f$ is the problem Hamiltonian. $f(t)$ and $g(t)$ are continuous functions that satisfy $f(0) = 1, f(t_a) = 0, g(0) = 0, g(t_a) = 1$, where $t_a$ is the final time. According to the adiabatic theorem, if there is no symmetry-related level crossing between the ground state and the first excited state during the evolution and the annealing time is long enough, we can exactly prepare the ground state of $H_f$.

An actual measurement at finite $t_a$ gives the measured energy $E = \langle \psi_f | H_f | \psi_f \rangle$ in the final state $| \psi_f \rangle$. The total probability of finding the system in an excited state of $H_f$ in the measurement is of order $O(1/t_a^2)$ (see Appendix A). If we denote the actual ground state energy $H_f$ by $E_{gs}$,
then the residual energy converges quadratically [39, 40],
\[ E_{\text{res}} = E - E_{\text{gs}} = O(1/t_a^2). \]  (2)

Based on this relation, we use the following protocol. We repeat the experiment for an increasing sequence of \( t_a \)'s and plot the measured \( E \)'s versus \( 1/t_a \). Then we infer the ground state energy \( E_{\text{gs}} \) by quadratic regression to the point \( 1/t_a \to 0 \). The maximum \( t_a \) would be determined by hardware noise on an actual quantum machine.

Theoretically, we can do a similar thing to VQE and QITE, i.e., running VQE and QITE repeatedly with different circuit depths (Trotter steps). However, the residual energy usually decays exponentially with circuit depth (Trotter steps). Compared with quadratic regression, exponential regression is less stable and robust. In this paper, we therefore concentrate on what can be accomplished through the polynomial approach, and leave exponential regression for future investigation.

### B. Variance Method

The variance method is more general in that it can handle evolutions that are characterized by a whole set of parameters (like \( \Theta \)), not just a single parameter (like \( t_a \)). We measure and record the variance
\[ \Delta_{\text{var}} = \langle \psi_f | H_f^2 | \psi_f \rangle - |\langle \psi_f | H_f | \psi_f \rangle|^2 \]  (3)
and the energy \( E \) for short annealing times in the QA or low-depth circuits in the VQE, plot \( E(\Delta_{\text{var}}) \), then linearly extrapolate the variance to 0.

To see why this works, note that the final state \( |\psi_f\rangle \) can be decomposed as
\[ |\psi_f\rangle = F |\phi_0\rangle + e^{i\varphi} \sqrt{1 - F^2} |\psi_{\text{ex}}\rangle, \]
where \( |\phi_0\rangle \) is the exact ground state and \( |\psi_{\text{ex}}\rangle \) is a superposition of the exact excited states \( |\phi_n\rangle \), i.e.,
\[ |\psi_{\text{ex}}\rangle = \sum_{j=1}^{N-1} c_j |\phi_j\rangle. \]  (5)

with \( \sum_{j=1}^{N-1} |c_j|^2 = 1 \), \( F \) is the fidelity between \( |\psi_f\rangle \) and \( |\phi_0\rangle \). When \( F^2 \geq 1/2 \), \( E_{\text{res}} \) is upper bounded by \( \sqrt{\Delta_{\text{var}}} \) [41]. The variance method is based on the assumption that with the increase of quantum resources, the coefficient \( F \) becomes larger, but \( |\psi_{\text{ex}}\rangle \) remains essentially unchanged. The condition is expected to hold in the QA when the energy is close to the ground state energy, since the “leakage” to individual excited states that have avoided level crossings should also be proportional to \( 1/t_a^2 \). In the VQE/QITE, it is more difficult to justify this assumption since the parameter set is model-dependent. However, it can be checked simply by confirming the goodness of the linear fit. Under this condition, Refs. [33, 42] illustrate that
\[ \lim_{E_{\text{res}} \to 0} \Delta_{\text{var}} = \frac{D_2}{D_1} E_{\text{res}}, \]
where \( D_1 \) and \( D_2 \) are constant defined by
\[ D_1 = \sum_{j=1}^{N-1} |c_j|^2 (E_j - E_{\text{gs}}), \quad D_2 = \sum_{j=1}^{N-1} |c_j|^2 (E_j - E_{\text{gs}})^2. \]  (7)

\( E_j \) is the \( j \)-th excited state eigenenergy of \( H_f \).

Suppose we have several energy/variance points near \((E, \Delta_{\text{var}})\) and use them to do linear extrapolation. The estimated ground state energy is
\[ E_{\text{extrp}} = E_{\text{gs}} + \frac{E_{\text{res}}^2}{D_2/D_1 - 2E_{\text{res}}}, \]  (8)

compared with the original data, we reduce the estimation error by
\[ E_{\text{res}}(1 - E_{\text{res}}/(D_2/D_1 - 2E_{\text{res}})), \]  (9)
which is very close to \( E_{\text{res}} \) if \( E_{\text{res}} \ll D_2/D_1 \). The second term is positive, \( E_{\text{extrp}} \) is thus still higher than \( E_{\text{gs}} \).

If the quantum evolution is noisy, we approximately describe the final state as
\[ \rho_f = (1 - \epsilon') |\psi_f\rangle\langle\psi_f| + \epsilon' \frac{I}{N}, \]  (10)
where \( \epsilon' \) is the global depolarizing noise rate. Then, the variance is \( \Delta'_{\text{var}} = \text{tr}(\rho_f H_f^2) - \epsilon' \text{tr}^2(\rho_f H_f^2) \), the energy is \( E' = \text{tr}(\rho_f H_f^2) \), the residual energy is \( E_{\text{res}}' = E' - E_{\text{gs}} \). Suppose \( H_f \) is traceless, we have
\[ \lim_{E_{\text{res}} \to 0} \Delta'_{\text{var}} = \frac{D_2}{D_1} E_{\text{res}}' + O(\epsilon'), \]  (11)
if \( E_{\text{res}}' \ll D_2/D_1 \) and \( \epsilon' \) is relatively small, the linear extrapolation still works. Denote
\[ C' = (D_2/D_1) E_{\text{gs}} + E_{\text{gs}}^2 + \text{tr}(H^2)/N, \]  (12)
the estimated ground state energy via extrapolation is
\[ E_{\text{extrp}}' = E' - \Delta'_{\text{var}} \frac{\partial E'}{\partial \Delta_{\text{var}}} \]
\[ = E_{\text{gs}} + \frac{E_{\text{res}}^2 C' - \epsilon' C'}{D_2/D_1 - 2E_{\text{res}}'}. \]  (13)

The interesting thing is that when \( \epsilon' \) is an appropriate non-zero small value, \( E_{\text{res}}^2 \approx \epsilon' C' \), \( E_{\text{extrp}}' \) is a more precise estimate of \( E_{\text{gs}} \) than the noiseless \( E_{\text{extrp}} \). Further details of the derivation along with an analysis of the effects of noise can be found in Appendix B.

Suppose the system has \( n \) qubits and the Hamiltonian consists of \( m \) Pauli observables whose weights are no greater than \( k \),
\[ H_f = \sum_{j=1}^{m} h[j]. \]  (14)
Figure 2. Extrapolation results obtained in quantum annealing (QA), showing how very accurate estimates of the ground state energy can be obtained in finite time, even for systems of increasing size. (a) Extrapolating \( t_a \) to infinity via quadratic fitting with \( t_a = 15, 16, 17, 18, 19, 20, n = 14 \). The dashed line shows the exact ground state energy. (b) Extrapolating \( \Delta_{\text{var}} \) to 0 via linear fitting with \( t_a = 15, 16, 17, 18, 19, 20, n = 14 \). (c) Energy estimation error versus annealing time \( t_a \) with \( n = 14 \). (d) Error reduction ratio (ERR), defined in Eq. 19, versus system size \( n \) with time extrapolation, showing that this method works for relatively large \( t_a \). (e) ERR versus system size \( n \) with variance extrapolation, showing that this method works for all \( t_a \).

Estimating \( \Delta_{\text{var}} \) introduces additional measurement overhead since we need to measure the expectation values on \( m^2 2k \)-local terms \{\( h[i] h[j] \}\}_{i,j=1,2,...,m} to estimate

\[
\text{tr}(\rho f H_f^2) = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{tr}(\rho_f h[i] h[j]).
\]

The “classical shadow” technique can reduce this cost efficiently [43, 44]. If we construct shadows via random Pauli measurements and use them to predict the local expectation values, the number of required measurements are of order \( O(3^k \log(m)) \) for energy terms, and of order \( O(3^{2k} \log(m^2)) \) for variance terms. When \( k \) is fixed (\( k \sim O(1) \)), the additional measurement overhead is a constant that does not increase with \( n \).

Note that neither the time method nor the variance method depends on the QA schedule \( f(t) \) and \( g(t) \). Our protocols can be applied to QA on arbitrary paths, e.g., QA with a midanneal pausing [45].

### III. EXTRAPOLATED QUANTUM ANNEALING

The workflow of our extrapolation algorithm for quantum annealing is illustrated in Algorithm 1. Numerical stability could be ensured by using only 6 points for the regression.

We apply our method to predict the ground state energy of the \( n = 14 \) one-dimensional transverse-field Ising model (TFIM) with a periodic boundary condition. The initial Hamiltonian is governed by the transverse field only,

#### Algorithm 1 Adaptive energy extrapolation in QA.

**Require:** The target Hamiltonian \( H_f \), the update interval \( \eta \), the convergence tolerance \( \epsilon \).

**Ensure:** A precise estimate of the ground state energy of \( H \).

Set \( T \leftarrow 1 \).

Set \( E_{\text{extrp}} \leftarrow 0 \).

**while** \( E_{\text{extrp}} \) has not converged with tolerance \( \epsilon \) **do**

- Implement QA with 6 evenly spaced annealing times over the interval \([3T/4, T]\).
- Measure the final state and obtain the energies.
- Implement the polynomial regression
  \[
  E = \alpha_0 + \alpha_1 t_a^2.
  \]
  (Or measure the variances \{\( \Delta_{\text{var}} \}\} and implement the linear regression \( E = \alpha_0 + \alpha_1 \Delta_{\text{var}} \))
- Set \( E_{\text{extrp}} \leftarrow \alpha_0 \).
- Set \( T \leftarrow T + \eta \).

**end while**

Output \( E_{\text{extrp}} \).
\[ H_{\text{init}} = - \sum_j X_j. \] (16)

The final Hamiltonian \( H_f \) is the TFIM Hamiltonian,
\[ H_{\text{TFIM}} = -J \sum_j Z_j Z_{j+1} - h \sum_j X_j, \] (17)
with \( J = h = 1 \). \( H_{\text{TFIM}} \) is traceless. The system is gapless in the thermodynamic limit. We choose the linear evolution schedule
\[ f(t) = 1 - t/t_a, \quad g(t) = t/t_a. \] (18)

The estimation errors with or without extrapolation for different \( t_a \) are shown in Fig. 2(c), two detailed extrapolation curves based on time and variance are shown in Fig. 2(a)(b). Both extrapolation schemes can reduce the error by approximately one order of magnitude, i.e., we need much shorter annealing times to estimate the ground state energy to a fixed precision. Considerable speedups can be achieved.

We define the error reduction ratio (ERR) as
\[ \text{ERR} = 1 - \frac{|E_{\text{extrp}} - E_{\text{gs}}|}{|E_{\text{std}} - E_{\text{gs}}|}, \] (19)
where \( E_{\text{gs}} \) is the ground state energy of the target Hamiltonian, \( E_{\text{std}} \) is the lowest energy obtained by standard QA, \( E_{\text{extrp}} \) is the energy obtained via extrapolation. As a goodness criterion, this ratio quantifies to what extent we can mitigate the estimation error. In the ideal case, \( \text{ERR} = 1 \), i.e., we mitigate all the estimation error and obtain the exact ground state energy. ERR as a function of different annealing times and system sizes is shown in Fig. 2(d,e). The scaling with system size is favorable for both schemes. The variance-based scheme can mitigate most errors for all annealing times, and the performance is more stable. The time-based scheme works well for annealing times larger than a critical time.

Now we consider a more realistic case where the measurement of energy/variance is not exact. This is inevitable on real devices due to readout errors and the limited number of measurement samples. The time control is also not precise. For the 14-qubit TFIM, we run QA with \( t_a = 15, 16, 17, 18, 19, 20 \) Suppose \( E, \Delta_{\text{var}}, t_a \) are accompanied by uncorrelated, unbiased, normally distributed precision errors with standard deviations \( \sigma(E), \sigma(\Delta_{\text{var}}), \sigma(t_a) \) respectively. We plot the average ERR for different \( \sigma(E), \sigma(\Delta_{\text{var}}) \) and \( \sigma(t_a) \) in Fig. 3, each pixel is averaged by 10 samples. (If \( |E_{\text{extrp}} - E_{\text{gs}}| > |E_{\text{std}} - E_{\text{gs}}| \), we manually set \( \text{ERR} = 0 \).) When the standard deviation is smaller than the corresponding quantity by one order of magnitude, the extrapolation method still works and ERR is close to 1. This indicates that the extrapolation method is stable under measurement errors as long as they are within reasonable bounds.

IV. EXTRAPOLATED VARIATIONAL QUANTUM EIGENSOLVER

VQE is a leading-edge algorithm for ground state preparation using near-term quantum computers. We run parameterized quantum circuits on a quantum computer and use a classical computer to optimize the parameters. There exist different choices of circuits. A well-studied one is the Hamiltonian variational ansatz \([46, 47]\), where the parameterized gates are determined by the target Hamiltonian. Taking the 1D TFIM as an example, we start from the initial state
\[ |\psi_{\text{init}}\rangle = (|0\rangle + |1\rangle)^{\otimes n}/\sqrt{2^n} \] (20)
and implement the unitary evolution
\[ U(\theta) = \prod_{l=1}^{p} \exp [-i\theta_{2l} \sum_j X_j] \exp [-i\theta_{2l-1} \sum_j Z_j Z_{j+1}]. \] (21)

We then iteratively update the \( 2p \) parameters \( \theta \) to minimize the expectation value of the energy of the final state. Previous research indicates that Hamiltonian variational ansatz with depth \( p = n/2 \) suffices to prepare the ground state of TFIM \([48, 49]\). The quantum circuit is illustrated in Fig. 4(a).

The variance extrapolation method for the VQE algorithm works as follows. During the optimization, the measured \( E \) and the corresponding variance \( \Delta_{\text{var}} \) decrease. We record both \( E \) and \( \Delta_{\text{var}} \) when \( E \) is decreased.

![Figure 3. Average Error Reduction Ratio (ERR), Eq. 19, in case of measurement errors as a function of standard deviations \( \sigma(E), \sigma(\Delta_{\text{var}}) \), and \( \sigma(t_a) \) for (a) the time method and (b) the variance method, \( t_a, E_{\text{res}}, \) and \( \Delta_{\text{var}} \) are the averaged quantities of the fitting data. For both methods there is a broad domain with \( \text{ERR} \approx 1 \), showing the robustness of the extrapolation scheme.](image-url)
Figure 4. Circuit structure and extrapolation results obtained in variational quantum eigensolver (VQE), showing how very accurate estimates of the ground state energy can be obtained with finite circuit depths. (a) The circuit structure of the Hamiltonian variational ansatz for 1D transverse-field Ising model (TFIM). (b) Variance extrapolation for different circuit depths with $n = 14$, $\epsilon = 0$, $\sigma = 0$. (c) Variance extrapolation for different error rates with $n = 8$. (d) Error Reduction Ratio (ERR) versus depth $p$ for different system sizes, $\epsilon = 0$, $\sigma = 0$. (e) Average ERR as a function of the 2-qubit gate error rate $\epsilon$ and measurement deviation $\sigma$.

by 0.01. After several samples of the initial values and optimization, we select the smallest variance $\Delta_{\text{var}}^{\min}$ and the smallest energy $E^{\min}$, keep only the energy-variance data that satisfy

$$
\Delta_{\text{var}} < 2\Delta_{\text{var}}^{\min}, \quad \text{and} \quad E < E^{\min} + 0.5.
$$

(22)

For these data, we implement linear regression to zero for $\Delta_{\text{var}}$ and infer the ground state energy by noting that when the variance vanishes, the energy will be the ground state energy. (In the presence of noise, we modify the range to $\Delta_{\text{var}} < \Delta_{\text{var}}^{\min} + 1$ and $E < E^{\min} + 0.5$.) The large-variance or large-energy data are abandoned for two reasons: the first one is that the relation between $E$ and $\Delta_{\text{var}}$ is approximately linear only when $\Delta_{\text{var}}$ is close to zero and the ground state is the dominating state; the second reason is that for some complicated Hamiltonians, the optimization may be stuck to a “frozen” state whose energy might be low but the variance is quite high, or stuck to an excited state whose variance might be low but the energy is high, we must avoid the influence of these states. Note that the circuit depth is fixed during the whole process.

We apply this extrapolation-assisted VQE to the TFIM with a periodic boundary condition and consider the ideal situation first. We sample 10 initial points and optimize the parameters with the BFGS algorithm. The extrapolation results for $n = 14$, $p = 3, 4, 5, 6$ are shown in Fig. 4(b). Our method can reduce the estimation error by 53%, 63%, 72%, 93%, respectively. Depth-3 VQE with extrapolation outperforms depth-5 VQE, depth-4 VQE with extrapolation outperforms depth-6 VQE. Since training a shallower variational quantum circuit requires fewer training epochs, extrapolation-assisted VQE also possesses the advantage of time-saving. ERR as a function of $n$ and $p$ is illustrated in Fig. 4(d). Unsurprisingly, ERR increases with circuit depth $p$ since the relation between $E^{\text{res}}$ and $\Delta_{\text{var}}$ is only approximately linear for small $\Delta_{\text{var}}$. Fig. 4(d) suggests that our method still works for larger system sizes.

Now we illustrate that our method is still efficient in the advent of noise. Suppose every 2-qubit $R_{ZZ}$ gate $\exp \left[-i\theta Z_i Z_{i+1}\right]$ in Eq. 21 is accompanied by a local depolarizing channel with noise rate $\epsilon$. And for the output state, we estimate $E$ and $\Delta_{\text{var}}$ with normally distributed errors (without loss of generality, we suppose the estimating standard deviations of $E$ and $\Delta_{\text{var}}$ are both $\sigma$). For $n = 8$, $p = 3$, we implement variance extrapolation for noisy VQE, some results are shown in Fig. 4(c). As expected, the minimum energy grows rapidly with increasing error rate. Both algorithmic and hardware
deficiencies cause the estimation error of $E_{gs}$ here. Encouragingly, variance extrapolation can always reduce the majority of the estimation error and gives an accurate estimate of $E_{gs}$. For $\epsilon = \sigma = 0.001, 0.005, 0.01$, our method can reduce the estimation error by 79%, 83%, 94%, respectively. In the last case, the lowest energy achieved by VQE, the extrapolated energy, and the ground state energy are

$$E_{\text{min}} = -9.30, \quad E_{\text{extrp}} = -10.19, \quad E_{gs} = -10.25. \quad (23)$$

This indicates that one can directly use the extrapolation technique for NISQ devices since the two-qubit gate noise rate on a state-of-the-art quantum computer is approximately 0.01 [50]. Within this range, the average ERR even increases with the noise rate. This phenomenon agrees with our analysis in Sec. II B. Further, we plot the average ERR for different $\epsilon$ and $\sigma$ in Fig. 4(e). Our method is pretty robust to noise, and it can serve as a tool for algorithmic and hardware error mitigation.

It is worth noting that the extrapolation method works for VQE with different ansatzes. To better demonstrate the practicability of our approach, we implement VQE with a hardware-efficient circuit on a real IBM quantum device. The machine we use is ibmq_jakarta [51], a superconducting quantum computer with seven qubits. The CNOT error rates range from $6.6 \times 10^{-3}$ to $1.1 \times 10^{-2}$. The connectivity of ibmq_jakarta is given in Fig. 5(a). Four qubits ($Q_0, Q_1, Q_3, Q_5$) are used to run this experiment. Here we consider the one-dimensional anisotropic XYZ Heisenberg chain,

$$H_{\text{XYZ}} = -\sum_j (J_x X_j X_{j+1} + J_y Y_j Y_{j+1} + J_z Z_j Z_{j+1}) \quad (24)$$

with $J_x = 2, J_y = 1, J_z = 1/2$. Fig. 5(b) shows the variational circuit with eight parameters and 3 CNOT gates (different single-qubit gates no longer share the parameters). This circuit is not capable of preparing the exact ground state of $H_{\text{XYZ}}$ even without gate noise.

The initial state here is $|\psi_{\text{init}}\rangle = |\rangle^{\otimes n}$. We randomly sample the initial parameters from $[0, 2\pi]$, and iteratively update them with the simultaneous perturbation stochastic approximation optimizer. During the optimization, we additionally measure some variances and use these data for regression. The lowest energy without extrapolation, the extrapolated energy, and the ground state energy are

$$E_{\text{min}} = -5.44, \quad E_{\text{extrp}} = -6.23, \quad E_{gs} = -6.29. \quad (25)$$

$E_{VQE}$ is much higher than the ground state energy. Extrapolation can reduce the estimation error by 93% and give a precise estimation of the ground state energy. These results are consistent with our numerical simulations, and confirm the validity of the extrapolation approach in a real-world application of a NISQ device.

$$|\psi_f\rangle = \lim_{\beta \to \infty} e^{-\beta H_f} |\psi_{\text{init}}\rangle \quad (26)$$

falls into the ground subspace of $H_f$. This process can
be simulated by the \( p \)-step Trotter decomposition

\[
e^{-\beta H_f} = \left(e^{-\Delta \tau h[1]}e^{-\Delta \tau h[2]} \ldots e^{-\Delta \tau h[m]}\right)^p + O(\Delta \tau^2),
\]

where

\[
\Delta \tau = \beta/p.
\]

is the step interval. Since quantum computers can only implement unitary evolutions, the main idea of QITE is to replace each non-unitary evolution \( e^{-\Delta \tau h[j]} \) in with a \( k \)-local unitary evolution \( e^{-i\Delta \tau A[j]} \) such that the states after these two processes are very close,

\[
|\psi\rangle \approx e^{-i\Delta \tau A[j]}|\psi\rangle,
\]

\( c = ((|\psi\rangle e^{-2\Delta \tau h[j]}|\psi\rangle)^{1/2} \) is the normalization factor. The optimal \( A[j] \) can be determined via measurements on \(|\psi\rangle\).

There are three challenges for QITE. The first one is the Trotter error induced by Trotterization. On NISQ devices, we need to choose a small \( p \) to obtain reliable results. For these small value of \( p \), the step interval \( \Delta \tau \) (Eq. 28) is large, and the corresponding Trotter error is non-negligible. The second one is the local approximation error, we have to use a small-size local operator \( e^{-i\Delta \tau A[j]} \) to approximate each non-unitary \( e^{-\Delta \tau h[j]} \) (we set \( k = 2 \) in the following numerical tests), relation 29 is not accurate. The third one is the hardware noise. Preceding errors may prevent us from finding the correct \( A[j] \) in a later step. All these errors accumulate with circuit depth. As a result, QITE is highly sensitive to errors.

We again consider the transverse-field Ising model (TFIM) and apply the variance method to enhance QITE. The initial state is

\[
|\psi_{init}\rangle = (|0\rangle + |1\rangle)^{\otimes n}/\sqrt{2^n}.
\]

Specifically, we record both the energy and the variance after each Trotter step. After several steps, we implement extrapolation with the recorded data. Fig. 6(b) shows some noiseless extrapolation results with only 3 Trotter steps. The variance method can reduce 78%, 84%, 85% of the estimation error for \( n = 4, 8, 12 \) respectively. Even very shallow circuits can yield reasonable estimates. The mitigating effect is scalable. Now suppose each 2-qubit operator \( e^{-i\Delta \tau A[j]} \) is accompanied by depolarizing noise with the rate \( \epsilon \), and the estimating standard deviations of \( E \) and \( \Delta \text{var} \) are both \( \sigma \). Fig. 6(c) plots a detailed example for \( \epsilon = \sigma = 0.01 \), where the energy decreases first, then increases because the errors become dominant. Here, our method reduces the estimation error by 89%. The lowest energy achieved by QITE, the extrapolated energy, and the ground state energy are

\[
E_{\text{min}} = -8.41, \quad E_{\text{extrp}} = -10.05, \quad E_{\text{gs}} = -10.25.
\]

Figure 6. Schematic and extrapolation results obtained in quantum imaginary time evolution (QITE), showing how better estimates of the ground state energy can be obtained with finite Trotter steps. (a) Schematic of QITE for the 1-D transverse-field Ising model (TFIM). (b) Variance extrapolation for different system sizes with 3 Trotter steps, \( \epsilon = \sigma = 0.01 \). (c) Variance extrapolation for noisy QITE with \( n = 8, \epsilon = \sigma = 0 \). (d) Average error reduction ratio as a function of the 2-qubit gate error rate \( \epsilon \) and measurement deviation \( \sigma \).
Fig. 6(d) shows the average ERR for different $\epsilon$ and $\sigma$. As we analyzed in Sec. II B, the extrapolation method performs especially well for a small non-zero $\epsilon$. Coincidentally, that value here is $\sim 0.01$, which matches the noise rates of current NISQ computers.

VI. CONCLUSIONS

We have proposed extrapolation methods to obtain more precise estimates of the ground state energy with quantum annealing, the variational quantum eigensolver, and quantum imaginary time evolution. Our methods can predict a much more accurate ground state energy with the same annealing time or circuit depth. These methods have great potential in the NISQ era since they are robust to measurement and gate errors. We demonstrated the accuracy enhancement in VQE through an experiment on an IBM quantum computer.

It is conceivable that other NISQ algorithms like quantum approximate optimization algorithm (QAOA) can similarly benefit from this method. One may use the extrapolation technique to precisely estimate excited-state energies in excited-state-preparing algorithms such as subspace-search VQE [52]. In addition, since the adiabatic model and the circuit model are polynomially equivalent [53], any quantum circuit can be translated to the ground state problem via Feynman-Kitaev circuit-to-Hamiltonian construction [54], the time and variance-based extrapolation may benefit other areas of quantum computing as well, e.g., digital quantum simulation. Energy extrapolation is a scalable technique that pushes NISQ algorithms to more practical applications. How to make the best use of it in quantum computing is worth studying further.

ACKNOWLEDGMENTS

We thank Rico Pohle, Xiang-Bin Wang, Jinfeng Zeng for helpful suggestions. We acknowledge the use of IBM Quantum services for this work.

Appendix A: Quadratic convergence of the residual energy

Here we give the justification for the statement that

$$E_{res} = \frac{A}{t_a^2} + O(t_a^{-3}),$$

find an expression for the constant $A$, and give a criterion for when the $1/t_a^4$ term can be neglected.

We have an $N$-dimensional Hilbert space and a time-dependent Hamiltonian with instantaneous eigenstates $|\phi(t)\rangle$:

$$H(t) |\phi_j(t)\rangle = \varepsilon_j(t) |\phi_j(t)\rangle$$

where $j = 0, 1, \ldots, N - 1$. We assume that there are no symmetry-related level crossings so that the ordering of the states is preserved in the evolution. A general state $|\psi(t)\rangle$ that satisfies the Schrödinger equation may be expanded in the diabatic basis as

$$|\psi(t)\rangle = \sum_{j=0}^{N-1} a_j(t) \exp\left[-i \int_0^t \varepsilon_j(t') dt'\right] |\phi_j(t)\rangle$$

where the coefficients satisfy

$$\partial_t a_k(t) = - \sum_{j=0}^{N-1} a_j(t) \exp\left[-i \int_0^t [\varepsilon_j(t') - \varepsilon_k(t')] dt'\right]$$

$$\langle \phi_k(t)|\partial_t H|\phi_j(t)\rangle [\varepsilon_j(t') - \varepsilon_k(t')]^{-1}.$$ (A4)

If we start the system in the ground state $|\phi_0(t = 0)\rangle$ and neglect multiple scattering we can set $a_j(t) = \delta_{01}$ on the right-hand side of this equation and integrate over time to obtain

$$a_k(t) = \int_0^t dt' \exp\left\{i \int_0^{t'} [\varepsilon_k(t'') - \varepsilon_0(t'')] dt''\right\}$$

$$\langle \phi_k(t')|\partial_t H|\phi_0(t')\rangle [\varepsilon_k(t') - \varepsilon_0(t')]^{-1}.$$ (A5)

By differentiating Eq. A2 and substituting the result we obtain the equivalent and slightly simpler form

$$a_k(t) = - \int_0^t dt' \exp\left\{i \int_0^{t'} [\varepsilon_k(t'') - \varepsilon_0(t'')] dt''\right\}$$

$$\langle \phi_k(t')|\partial_t \phi_0(t')\rangle.$$ (A6)

Eqs. A5 and A6 are standard results [55].

To isolate the dependence on $t_a$ it’s now convenient to change variables to $s = t/t_a$, $\varepsilon_0(s = 1) = E_{gs}$ is the final ground state energy. The final state of the evolution is $|\psi(s = 1)\rangle$ while the actual final ground state is $|\phi_0(s = 1)\rangle$. The probability to find the system in the $j$th excited state at the end is $p_j = |a_j(s = 1)|^2$, so the probability of failure of the algorithm is the sum of the $|a_j(s = 1)|^2$ from $j = 1$ to $j = N - 1$.

The $a_j(s = 1)$ are given by

$$a_j(s = 1) = - \int_0^1 ds' \exp\left\{it_a \int_0^{s'} [\varepsilon_j(s'') - \varepsilon_0(s'')] ds''\right\}$$

$$\langle \phi_j(s')|\partial_{s'} \phi_0(s')\rangle.$$ (A7)

This may be written as

$$a_j(s = 1) = - \int_0^1 ds' \exp\left\{it_a F_j(s')\right\} G_j(s'),$$ (A8)

where

$$F_j(s) = \int_0^s [\varepsilon_j(s'') - \varepsilon_0(s'')] ds'.$$ (A9)
and
\[
G_j(s) = \langle \phi_j(s)|\partial_s \phi_0(s)\rangle = -\langle \phi_j(s)|\partial_s H|\phi_0(s)\rangle \tag{A10}
\]
and since \(F_j\) and \(G_j\) are continuous the Riemann-Lebesgue Lemma guarantees that this expression vanishes as \(t_a \to \infty\) for all \(j\). Take a simple random-field Ising Hamiltonian as an example, the probabilities are shown in Fig. 7, numerical results and analytical results match well when \(t_a > 10\).

![Figure 7. The probability to find the system in the \(j\)-th excited state at the end of quantum annealing. The markers represent the numerical results, the lines represent the analytical results of equation A7.](image)

Since \(F_j(s) > 0\) for all \(s\) we can use integration by parts to develop an asymptotic expansion in \(1/t_a\) [56]:

\[
a_j(s = 1) = -\frac{1}{it_a} \left[ e^{it_a F_j(s)} \frac{d}{ds} \left( \frac{G_j(s)}{F_j(s)} \right) \right]_0^1 + \frac{1}{(it_a)^2} \left( e^{it_a F_j(s)} \frac{d}{ds} \left( \frac{1}{F_j(s)} \frac{d}{ds} \left( \frac{G_j(s)}{F_j(s)} \right) \right) \right)_0^1 + O\left( \frac{1}{t_a^3} \right) \tag{A11}
\]

The residual energy in this approximation is

\[
E_{res} = \sum_{j=1}^{N-1} |a_j(s = 1)|^2 [\varepsilon_j(s = 1) - \varepsilon_0(s = 1)], \tag{A12}
\]

The leading term of \(a_j(s = 1)\) is \(O(1/t_a^3)\), so the leading term of \(E_{res}\) is \(O(1/t_a^3)\). The coefficient \(A\) in Eq. A1 is

\[
A = \sum_{j=1}^{N-1} \left[ e^{it_a F_j(s)} \frac{d}{ds} \left( \frac{G_j(s)}{F_j(s)} \right) \right]_0^1 [\varepsilon_j(s = 1) - \varepsilon_0(s = 1)] \tag{A13}
\]

In an asymptotic expansion, the first term dominates when the second term is much smaller, which leads to the criterion

\[
t_a \gg \left| \frac{e^{it_a F_j(s)} d}{ds} \left( \frac{G_j(s)}{F_j(s)} \right) \right|_0^1 \left| \frac{e^{it_a F_j(s)} d}{ds} \left( \frac{G_j(s)}{F_j(s)} \right) \right|_0^1 \tag{A14}
\]

This must hold for all \(j\).

Since \(F_j(s) = \varepsilon_j(s) - \varepsilon_0(s)\) we see that the criterion does depend on the behavior of the gap between the ground state energy and the excited state energies, albeit only at the endpoints of the evolution. The criterion becomes more stringent as the gaps decrease, as is expected. Since this also applies to \(s = 0\), this would suggest using an initial Hamiltonian with a large gap. We leave this problem for future investigation.

**Appendix B: Linear convergence of the energy variance**

In this section, we present the criterion for the linearity of the variance and use it to justify the extrapolation procedure. There is a set of exact eigenstates \(|\phi_j\rangle\) of \(H_f\):

\[
H_f|\phi_j\rangle = E_j|\phi_j\rangle. \tag{B1}
\]

Our goal is to determine \(E_0 = E_{gs}\) by a sequence of variational calculations. The variational wavefunction \(|\psi_f\rangle\) at any stage of the sequence can be written as

\[
|\psi_f\rangle = F|\phi_0\rangle + e^{i\varphi} \sqrt{1 - F^2} |\psi_{ex}\rangle, \tag{B2}
\]

where \(|\phi_0\rangle\) is the exact ground state and \(|\psi_{ex}\rangle\) is a sum of the exact excited eigenstates \(|\phi_j\rangle\) of \(H_f\):

\[
|\psi_{ex}\rangle = \sum_{j=1}^{N} c_j|\phi_j\rangle \tag{B3}
\]

with \(\sum_{j=1}^{N} |c_j|^2 = 1\). \(F\) is the fidelity. We follow Ref. [33] and define two energy moments

\[
D_1 = \sum_{j=1}^{N} |c_j|^2 (E_j - E_{gs}) \tag{B4}
\]

and

\[
D_2 = \sum_{j=1}^{N} |c_j|^2 (E_j - E_{gs})^2. \tag{B5}
\]

Then some manipulation leads to the equations

\[
E - E_{gs} = E_{res} = (1 - F^2)D_1 \tag{B6}
\]

and

\[
\Delta_{var}(s = 1) = (1 - F^2)D_2 - (1 - F^2)^2 D_1^2. \tag{B7}
\]
In the variance extrapolation we perform a sequence of calculations in which $F$ increases but is not measured while $E(F)$ and $\Delta_{\text{var}}(F)$ are measured. Thus we wish to eliminate $F$ from these equations. When this is done we find

$$\Delta_{\text{var}} = \frac{D_2}{D_1} E_{\text{res}} - E_{\text{res}}^2 = \frac{D_2}{D_1} (E - E_{gs}) - (E - E_{gs})^2.$$  \hfill (B8)

Now we assume that $|c_j|^2$, the relative weights of the excited states, and hence also the $D_{1,2}$, do not vary significantly with $F$ near $F = 1$. Then we see that the dependence of $\Delta_{\text{var}}$ on $E_{\text{res}}$ is parabolic and linear when $E - E_{gs} \approx 0$. This criterion for the validity of the linear approximation used in the main text is that

$$E_{\text{res}} = E - E_{gs} \ll D_2/D_1.$$  \hfill (B9)

One cannot easily calculate $D_{1,2}$, since they depend on the structure of $H_f$ as well as on the details of the variational procedure. In practice, however, this is not necessary. One plots $\Delta_{\text{var}}$ against $E$. If $E_{\text{res}}$ satisfies the inequality B9 we can confidently extrapolate to the point where the curve crosses the $E$-axis to find $E_{gs}$. Suppose we have several data points near $(E, \Delta_{\text{var}})$ and use them for extrapolation, the slope is

$$\frac{\partial E}{\partial \Delta_{\text{var}}} = 1/(D_2/D_1 - 2E_{\text{res}}),$$  \hfill (B10)

the estimated energy via extrapolation is

$$E_{\text{extrp}} = E - \Delta_{\text{var}} \frac{\partial E}{\partial \Delta_{\text{var}}} = E - \frac{(D_2/D_1)E_{\text{res}} - E_{\text{res}}^2}{D_2/D_1 - 2E_{\text{res}}} = E_{gs} + \frac{E_{\text{res}}^2}{D_2/D_1 - 2E_{\text{res}}}.$$  \hfill (B11)

In our numerical simulation, $E_{\text{extrp}}$ is always slightly higher than $E_{gs}$ since $D_2/D_1 > 2E_{\text{res}}$ and the second term is positive.

If the quantum circuit is deep and noisy, we can approximately describe the effects of noise by a global depolarizing noise channel, i.e., the final state is

$$\rho_f = (1 - \epsilon') |\psi_f\rangle \langle \psi_f| + \epsilon' I/N,$$  \hfill (B12)

where $\epsilon'$ is the noise rate. In this case, the variance $\Delta'_{\text{var}} = \text{tr}(\rho_f H^2) - \text{tr}^2(\rho_f H)$, the energy $E' = \text{tr}(\rho_f H)$, the residual energy $E'_{\text{res}} = E' - E_{gs}$. Suppose $H$ is traceless, together with equality B8, we obtain

$$\Delta'_{\text{var}} = \frac{D_2}{D_1} E'_{\text{res}} - E'_{\text{res}}^2 + \epsilon' (\frac{D_2}{D_1} E'_{gs} + E_{gs}^2 + \text{tr}(H^2)/N),$$  \hfill (B13)

if $E'_{\text{res}} \ll D_2/D_1$ and $\epsilon'$ is relatively small, the first term $(D_2/D_1)E'_{\text{res}}$ is dominant, the linear extrapolation still works. Denote $(D_2/D_1)E_{gs} + E_{gs}^2 + \text{tr}(H^2)/N$ as $C'$, the estimated ground state energy via extrapolation is

$$E'_{\text{extrp}} = E' - \Delta'_{\text{var}} \frac{\partial E'}{\partial \Delta'_{\text{var}}} = E_{gs} + \frac{E'_{\text{res}}^2 - \epsilon' C'}{D_2/D_1 - 2E'_{\text{res}}}.$$  \hfill (B14)

[1] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin,.T. Zhou, J. Phys. Chem. Lett. 1, 015003 (2020).
[2] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum 2, 79 (2018).
[3] I. H. Deutsch, Harnessing the power of the second quantum revolution, PRX Quantum 1, 020101 (2020).
[4] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Quantum annealing: A new method for minimizing multidimensional functions, Chemical Physics Letters 219, 343 (1994).
[5] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse ising model, Phys. Rev. E 58, 5355 (1998).
[6] T. Albash and D. A. Lidar, Adiabatic quantum computation, Rev. Mod. Phys. 90, 015002 (2018).
[7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, Nature Communications 5, 4213 (2014).
[8] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature 549, 242 (2017).
[9] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, Phys. Rev. Lett. 122, 230401 (2019).
[10] M. Motta, C. Sun, A. T. K. Tan, M. J. O’Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan,
Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nature Physics 16, 205 (2020).

[11] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, npj Quantum Information 5, 75 (2019).

[12] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. Tan, G. K.-L. Chan, and A. J. Minnich, Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution, PRX Quantum 2, 010317 (2021).

[13] C. Cao, Z. An, S.-Y. Hou, D. Zhou, and B. Zeng, Quantum imaginary time evolution steered by reinforcement learning (2021), arXiv:2105.08696.

[14] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nature Reviews Physics 3, 625 (2021).

[15] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, et al., Noisy intermediate-scale quantum (nisq) algorithms (2021), arXiv:2101.08448.

[16] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Science and Technology 2, 045001 (2017).

[17] C. Cao and X. Wang, Noise-assisted quantum autoencoder, Phys. Rev. Applied 15, 054012 (2021).

[18] J. Zeng, C. Cao, C. Zhang, P. Xu, and B. Zeng, A variational quantum algorithm for hamiltonian diagonalization, Quantum Science and Technology 6, 045009 (2021).

[19] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), arXiv:1411.0428.

[20] Y. Yu, C. Cao, C. Dewey, X.-B. Wang, N. Shannon, and R. Joynt, Quantum approximate optimization algorithm with adaptive bias fields (2021), arXiv:2105.11946.

[21] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Variational Quantum Fidelity Estimation, Quantum 4, 248 (2020).

[22] R. Chen, Z. Song, X. Zhao, and X. Wang, Variational quantum algorithms for trace distance and fidelity estimation (2020), arXiv:2012.05768.

[23] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habbegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, J. Neeley, C. Neill, H. Neven, M. Y. Niu, T. E. O’Brien, E. Ostby, A. Petukhov, H. Pumperman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, M. Szalay, T. Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. J. Yao, P. Yeh, and A. Zalcman, Hartree-fock on a superconducting qubit quantum computer, Science 369, 1084 (2020).

[24] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications 9, 4812 (2018).

[25] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications 12, 1791 (2021).

[26] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nature Communications 12, 6961 (2021).

[27] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X 7, 021050 (2017).

[28] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X 8, 031027 (2018).

[29] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. 119, 180509 (2017).

[30] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature 567, 491 (2019).

[31] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation (2021), arXiv:2109.04457.

[32] N. Shimizu, Y. Utsuno, T. Mizusaki, T. Otsuka, T. Abe, and M. Honma, Novel extrapolation method in the monte carlo shell model, Phys. Rev. C 82, 061305 (2010).

[33] N. Shimizu, Y. Utsuno, T. Mizusaki, M. Honma, Y. Tsuchida, and T. Otsuka, Variational procedure for nuclear shell-model calculations and energy-variance extrapolation, Phys. Rev. C 85, 054301 (2012).

[34] Y. Iqbal, F. Becca, S. Sorella, and D. Poilblanc, Gapless spin-liquid phase in the kagome spin-\(\frac{1}{2}\) heisenberg antiferromagnet, Phys. Rev. B 87, 060405 (2013).

[35] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study, Phys. Rev. B 58, 6800 (1998).

[36] S. Sorella, Generalized lanczos algorithm for variational quantum monte carlo, Phys. Rev. B 64, 024512 (2001).

[37] S. Endo, Q. Zhao, Y. Li, S. Benjamin, and X. Yuan, Mitigating algorithmic errors in a hamiltonian simulation, Phys. Rev. A 99, 012334 (2019).

[38] M. Kreshchuk, Quantum Simulation of Quantum Field Theory in the Front Form, Ph.D. thesis, Tufts University (2021).

[39] A. Messiah, Quantum Mechanics, Vol. 2 (North-Holland, Amsterdam, 1962).

[40] S. Suzuki and M. Okada, Residual energies after slow quantum annealing, Journal of the Physical Society of Japan 74, 1649 (2005).

[41] T. Imoto, Y. Seki, Y. Matsuzaki, and S. Kawabata, Improving the accuracy of the energy estimation by combining quantum annealing with classical computation (2021), arXiv:2102.05323.

[42] T. Mizusaki and M. Imada, Precise estimation of shell model energy by second-order extrapolation method, Phys. Rev. C 67, 041301 (2003).

[43] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measure-
ments, *Nature Physics* **16**, 1050 (2020).

[44] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of pauli observables by derandomization, *Phys. Rev. Lett.* **127**, 030503 (2021).

[45] H. Chen and D. A. Lidar, Why and when pausing is beneficial in quantum annealing, *Phys. Rev. Applied* **14**, 014100 (2020).

[46] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).

[47] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the hamiltonian variational ansatz, *PRX Quantum* **1**, 020319 (2020).

[48] D. Wierichs, C. Gogolin, and M. Kastoryano, Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer, *Phys. Rev. Research* **2**, 043246 (2020).

[49] W. W. Ho and T. H. Hsieh, Efficient variational simulation of non-trivial quantum states, *SciPost Phys.* **6**, 29 (2019).

[50] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, S. Hong, C. Jones, A. Petukhov, D. Kafri, S. Demura, B. Burkett, C. Gidney, A. G. Fowler, A. Paler, H. Putterman, I. Aleiner, P. Arute, K. Arya, R. Babbush, J. C. Bardin, A. Bentsen, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, N. Bushnell, B. Chiaro, R. Collins, W. Courtney, A. R. Derk, D. Eppens, C. Erickson, E. Farhi, B. Foxen, M. Giustina, A. Greene, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, T. Huang, W. J. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, K. Kechedzhi, S. Kim, A. Kitaev, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, T. McCourt, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neely, C. Neill, M. Newman, M. Y. Niu, T. E. O’Brien, A. Opreanecak, E. Ostby, B. Pató, N. Redd, P. Roushan, N. C. Rubin, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. White, Z. J. Yao, P. Yeh, J. Yoo, A. Zalcman, H. Neven, S. Boixo, V. Smelyanskiy, Y. Chen, A. Megrant, J. Kelly, and G. Q. A. Exponential suppression of bit or phase errors with cyclic error correction, *Nature* **595**, 383 (2021).

[51] 7-qubit backend: IBM Q team, “IBM Q 7 Jakarta backend specification V1.0.23,” (2021). Retrieved from https://quantum-computing.ibm.com.

[52] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Research* **1**, 033062 (2019).

[53] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, Adiabatic quantum computation is equivalent to standard quantum computation, *SIAM Review* **50**, 755 (2008).

[54] A. Y. Kitaev, A. Shen, M. N. Vyalyi, and M. N. Vyalyi, *Classical and quantum computation*, 47 (American Mathematical Soc., 2002).

[55] A. B. Migdal, *Qualitative Methods in Quantum Theory* (Taylor and Francis, 1977).

[56] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, 1978).