Even Shorter Quantum Circuit for Phase Estimation on Early Fault-Tolerant Quantum Computers with Applications to Ground-State Energy Estimation

Zhiyan Ding[1,*] and Lin Lin[1,2,3,†]

[1] *Department of Mathematics, University of California, Berkeley, California 94720, USA*
[2] *Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[3] *Challenge Institute of Quantum Computation, University of California, Berkeley, California 94720, USA*

(Received 6 December 2022; revised 21 February 2023; accepted 24 April 2023; published 26 May 2023)

We develop a phase-estimation method with a distinct feature: its maximal run time (which determines the circuit depth) is δ/ϵ, where ϵ is the target precision, and the preconstant δ can be arbitrarily close to 0 as the initial state approaches the target eigenstate. The total cost of the algorithm satisfies the Heisenberg-limited scaling $\widetilde{\mathcal{O}}(\epsilon^{-1})$. As a result, our algorithm may significantly reduce the circuit depth for performing phase-estimation tasks on early fault-tolerant quantum computers. The key technique is a simple subroutine called quantum complex exponential least squares (QCELS). Our algorithm can be readily applied to reduce the circuit depth for estimating the ground-state energy of a quantum Hamiltonian, when the overlap between the initial state and the ground state is large. If this initial overlap is small, we can combine our method with the Fourier-filtering method developed in [Lin and Tong, PRX Quantum 3, 010318, 2022], and the resulting algorithm provably reduces the circuit depth in the presence of a large relative overlap compared to ϵ. The relative-overlap condition is similar to a spectral-gap assumption but it is aware of the information in the initial state and is therefore applicable to certain Hamiltonians with small spectral gaps. We observe that the circuit depth can be reduced by around 2 orders of magnitude in numerical experiments under various settings.

DOI: 10.1103/PRXQuantum.4.020331

I. INTRODUCTION

Phase estimation is one of the most important quantum primitives. The problem of phase estimation can be equivalently stated as estimating the eigenvalue of a quantum Hamiltonian H, under the assumption that we can query H via the Hamiltonian-evolution operator $U=e^{-i\tau H}$ for some real number τ. There are two important performance metrics for the phase estimation: the maximal run time, denoted by $T_{\max}$, and the total run time T_{total}, which is the sum of the run time multiplied by the number of repetitions from each circuit in the algorithm. $T_{\max}$ and T_{total} approximately measures the circuit depth and the total cost of the algorithm, respectively, in a way that is independent of the details in implementing U. If we are also given an eigenvector $|\psi\rangle$ associated with an eigenvalue $e^{-i\tau\lambda}$, the Hadamard test is arguably the simplest algorithm for estimating the phase $\lambda\in[-\pi/\tau,\pi/\tau)$. It uses only one ancilla qubit and a single query to U controlled by the ancilla qubit, i.e., $T_{\max}=\tau$. This makes the Hadamard test ideally suited for early fault-tolerant quantum devices, which are expected to have a very limited number of logical qubits and may have difficulty in handling circuits beyond a certain maximal depth. The Hadamard test has many drawbacks too: it requires $|\psi\rangle$ to be an exact eigenstate, which is a stringent condition that cannot be satisfied in most scenarios. It also requires $N_s=\mathcal{O}(\epsilon^{-2})$ repetitions to estimate λ to precision ϵ,and hence the total run time is $\mathcal{O}(\epsilon^{-2})$.

Both problems can be addressed by quantum phase estimation (QPE) and its many variants [1–6]. Generically, estimating the phase to ϵ accuracy with a high success probability requires $T_{\max}$ to be at least π/ϵ for QPE [7, Section 5.2.1] [8]. Additionally, the total run time of QPE is $\mathcal{O}(\epsilon^{-1})$ and achieves the Heisenberg-limited scaling [9–11], which is the optimal scaling permitted by quantum mechanics. The standard version of QPE (see, e.g., Ref. [7, Chapter 5]) uses at least $\log_2(\pi(\tau\epsilon)^{-1})$ ancilla qubits and is not suitable for early fault-tolerant devices

*zding.m@berkeley.edu
†linlin@math.berkeley.edu

but the semiclassical version of QPE [1,2,12] can achieve the same task and uses only one ancilla qubit.

To our knowledge, in all existing works on QPE satisfying Heisenberg-limited scaling, the maximal run time π/ϵ is non-negotiable, in the sense that the preconstant in front of ϵ^{-1} cannot be significantly reduced in general. This is because QPE-type methods construct, directly or indirectly, a filtering function that transitions from 1 to 0 on an interval of width ϵ. This can be a severe limitation in practice, since estimating λ to precision 0.001 means that U needs to be coherently queried approximately 3000 times in the quantum circuit (assuming that $\tau = 1$). It is therefore desirable to have a phase-estimation method that satisfies the following properties:

(1) It allows $|\psi\rangle$ to be an inexact eigenstate with one ancilla qubit.
(2) It maintains the Heisenberg-limited scaling: to estimate λ to precision ϵ with probability $1-\eta$, the total cost is $\mathcal{O}(\epsilon^{-1}\,\mathrm{poly}\log(\epsilon^{-1}\eta^{-1}))$.
(3) It reduces the circuit depth: the maximal run time can be (much) lower than π/ϵ, especially when $|\psi\rangle$ is close to be an exact eigenstate of U.

When $|\psi\rangle$ is an exact eigenstate, the maximal run time in QPE-type methods may be reduced by means of a trade-off between the circuit depth and the number of repetitions. However, if the initial state is not an exact eigenstate, this strategy is no longer directly applicable. In this paper, we introduce algorithms that can satisfy properties (2) and (3), without assuming that $|\psi\rangle$ is an exact eigenstate.

A. Main idea

To achieve this, our quantum circuit (Fig. 1) is the same as the circuit used in the Hadamard test but replaces $U = e^{-i\tau H}$ by $U^n = e^{-in\tau H}$ for a sequence of integers n. This is a simple circuit, it uses only one ancilla qubit, and it is suitable for early fault-tolerant quantum computers. The most challenging component may be the implementation of the controlled time evolution. Under additional assumptions, the controlled time evolution for certain unitaries may be replaced by uncontrolled time evolution (see, e.g., Refs. [13–17]).

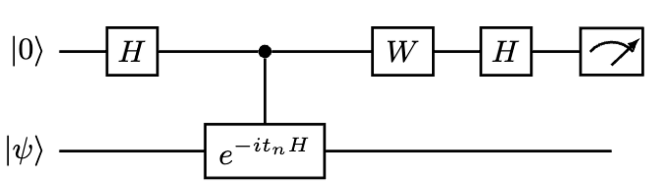

FIG. 1. The quantum circuit used for collecting the input data. H is the Hadamard gate, $t_n = n\tau$. Choosing $W = I$ or $W = S^\dagger$ (S is the phase gate) allows us to estimate the real or the imaginary part of $\langle\psi|\exp(-it_nH)|\psi\rangle$.

Let $t_n = n\tau$ for $n = 0,\ldots,N-1$. Also, for simplicity, we refer to $T_{\max} := N\tau$ as the maximal running time (the actual maximal running time is $(N-1)\tau$). The circuit provides an estimate of the value of $\langle\psi|e^{-it_nH}|\psi\rangle$ by measuring the success probability of the first qubit. Repeated measurements at different n provide a (complex) time series

$$\{(t_n, Z_n)\}_{n=0}^{N-1}, \tag{1}$$

where Z_n is a complex-valued random variable such that $\mathbb{E}(Z_n) = \langle\psi|\exp(-it_nH)|\psi\rangle$. In the intuitive analysis, we may assume $Z_n \approx \langle\psi|\exp(-it_nH)|\psi\rangle$. We give the detailed construction of Z_n in Sec. II A.

Without loss of generality, let us denote the target eigenstate by $|\psi_0\rangle$, the target eigenvalue by λ_0 (in this context, λ_0 does not need to be the smallest eigenvalue of H), and the overlap between the initial state $|\psi\rangle$ and the target eigenstate by $p_0 = |\langle\psi|\psi_0\rangle|^2$. We also assume $\lambda_0 \in [-\pi,\pi)$. If $p_0 = 1$ (i.e., $|\psi\rangle$ is the target eigenstate) and the number of samples for each n is sufficiently large, we have $Z_n \approx e^{-it_n\lambda_0}$, which is an exponential function. If $p_0 < 1$, we may still fit the input data provided by Eq. (1) using a complex exponential $r\exp(-it_n\theta)$, where $r \in \mathbb{C}, \theta \in \mathbb{R}$.

A main step of our method is a subroutine that solves the following nonlinear least-squares problem:

$$(r^*,\theta^*) = \mathrm{argmin}_{r\in\mathbb{C},\theta\in\mathbb{R}} L(r,\theta),$$
$$L(r,\theta) = \frac{1}{N}\sum_{n=0}^{N-1}|Z_n - r\exp(-it_n\theta)|^2, \tag{2}$$

where θ^* gives the approximation to the phase λ_0. Note that once we obtain the data set from the quantum circuit in Fig. 1, minimizing $L(r,\theta)$ only requires classical computation. This subroutine is dubbed *quantum complex exponential least squares* (QCELS). The minimization problem can be efficiently solved on classical computers (see Sec. II B). An illustrative example of QCELS using the spectrum from the transverse-field Ising model (TFIM) model (for details, see Sec. V A) is shown in Fig. 2(a). In the graph, we set the initial overlap $p_0 = 0.8$. The scatter points are the data points (t_n, Z_n) defined in Eq. (1), and the curve represents the fitting function $r^*\exp(-it\theta^*)$, where (r^*,θ^*) is the optimal solution to Eq. (10).

If $p_0 = 1$ and $N = 2$, the behavior of QCELS is very similar to that of the Hadamard test, we can estimate $\theta \approx \lambda_0 \bmod [-\pi/\tau, \pi/\tau)$ to any precision, and the circuit depth is independent of ϵ. However, there are some immediate issues with this approach:

(a) It is not so clear whether Eq. (2) can estimate λ_0 accurately with a short-depth circuit, even if p_0 is close but not equal to 1. Moreover, this method clearly fails if there exists some eigenstate $|\psi_i\rangle$ such that $|\langle\psi|\psi_i\rangle|^2 > p_0$. So p_0 should be larger than some minimal threshold.

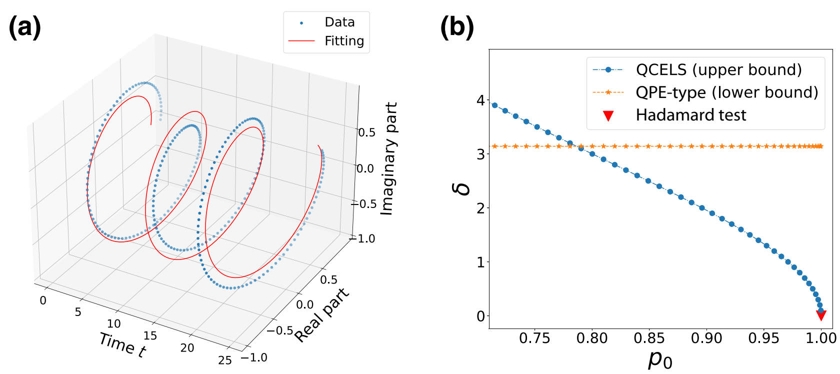

FIG. 2. (a) Fitting the noisy input data with $p_0 = 0.8$ using a complex exponential function. The mismatch between the data and the best fit reflects that the input data are more complex than a single complex exponential function. Despite this mismatch even in the absence of any Monte Carlo sampling error, QCELS is able to accurately estimate the phase under proper conditions. (b) A comparison of the theoretical upper bound of $\delta = T_{\max}\epsilon$ for QCELS ($T_{\max}$ is the maximal run time) with the lower bound of δ for QPE-type methods when $p_0 \geq 0.71$. The Hadamard test is only applicable when $p_0 = 1$ and in this case δ can be chosen to be arbitrarily small.

(b) For each n, the number of measurements is at least 1. If $N = \Theta(\epsilon^{-1})$, the total run time is at least $N(N-1)/2 = \Theta(\epsilon^{-2})$ and the method does not satisfy the Heisenberg-limited scaling.

Our main body of work is to address these issues and to develop an efficient algorithm for postprocessing the input time series generated by quantum computers.

The answer to point (a) is given by Theorem 1. Roughly speaking, when $p_0 > 0.71$, we may choose a proper $\delta > 0$ so that the maximal run time is $T_{\max} = N\tau = \delta/\epsilon$ and the global minima to Eq. (2) can estimate λ_0 to precision $\delta/T_{\max} = \epsilon \pmod{[-\pi/\tau, \pi/\tau)}$. Moreover, when Z_n is sufficiently concentrated around its expectation and as $p_0 \to 1$, δ can be chosen to be arbitrarily small. Therefore, the maximal run time (and the circuit depth) can be continuously reduced as the input state approaches an exact eigenstate and QCELS maintains the desirable behavior of the Hadamard test when $p_0 < 1$.

To address point (b), we can start from a small value of τ, which allows us to estimate λ_0 to precision $\delta/(N\tau)$. If δ and N are fixed, then this estimate can only reach limited precision. Similar to the binary search strategy for refining the estimate of the eigenvalues [13,15,18], we can refine this estimate by increasing the maximal run time. Specifically, we can multiply τ by a constant and repeat the process with fixed δ and N. We only need to repeat the process $J = \log_2(\delta/(N\epsilon))$ times. At the last step, we have $\tau_J = \delta/(N\epsilon)$ and the maximal circuit depth is $T_{\max} = N\tau_J = \delta/\epsilon$. This procedure is called *multilevel QCELS* and is described by Algorithm 1. According to Theorem 2, when $p_0 \approx 1$, we may choose $\delta = \Theta(\sqrt{1-p_0}) \ll 1$ and estimate λ_0 to precision ϵ. The maximal run time is $T_{\max} = N\tau = \delta/\epsilon$ and the total cost is $\widetilde{\mathcal{O}}(\delta^{-1}\epsilon^{-1})$. Both Theorem 2 and the numerical results verify that Algorithm 1 satisfies the desired properties (1), (2), and (3) listed in Sec. I. In particular, the circuit depth can be continuously adjusted by the parameter δ [see a comparison of the theoretical circuit depth of different methods in Fig. 2(b)] and the algorithm satisfies the Heisenberg-limited scaling for all choices of δ within the allowed range determined by p_0 and the noise level due to measurements.

B. Ground-state energy estimation

As an application, we consider the problem of estimating the ground-state energy (the algebraically smallest eigenvalue) of an n-qubit quantum Hamiltonian H. Here, we assume ground-state energy $\lambda_0 \in [-\pi, \pi)$ and $|\psi_0\rangle$ is the associated eigenvector. In the absence of additional assumptions, the task can be quantum Merlin Arthur (QMA) hard [3,19,20]. Hence we assume that an initial quantum state $|\psi\rangle = U_I|0^n\rangle$ can be prepared via a unitary U_I and the overlap $p_0 = |\langle\psi|\psi_0\rangle|^2 > 0$. If $p_0 \geq 0.71$, we can readily apply Theorem 2 to estimate λ_0 using a short-depth circuit.

If p_0 is small, we propose an algorithm combining the multilevel QCELS algorithm with the Fourier-filtering technique developed in Ref. [15] to estimate λ_0. To demonstrate the efficiency of the algorithm, we assume that there is an interval I containing λ_0 and a slightly larger interval $I' \supset I$ with a positive distance D separating I and $(I')^c$ [see Eq. (36)]. We introduce a concept called the *relative overlap* of the initial vector $|\psi\rangle$ with the ground state with respect to the intervals I and I':

$$p_r(I, I') = \frac{|\langle\psi|\psi_0\rangle|^2\, \mathbf{1}_I(\lambda_0)}{\sum_{\lambda_k \in I'} |\langle\psi|\psi_k\rangle|^2}. \tag{3}$$

Here, the denominator is assumed to be nonvanishing and $\mathbf{1}_I(\cdot)$ is the indicator function on I such that $\mathbf{1}_I(\lambda_0) = 1$

if $\lambda_0 \in I$ and $\mathbf{1}_I(\lambda_0) = 0$ if $\lambda_0 \notin I$. The most straightforward scenario is that the system has a spectral gap $\Delta = \lambda_1 - \lambda_0$. We can then choose $I = [-\pi, \lambda_{\text{prior}} + \Delta/4]$, $I' = [-\pi, \lambda_{\text{prior}} + 3\Delta/4]$, and $D = \Delta/2$, where λ_{prior} is a rough estimation of λ_0 such that $|\lambda_{\text{prior}} - \lambda_0| \leq \Delta/4$. The relative overlap in this case will be 1. It should be noted that the preceding discussion considers a worst-case scenario. In a real application, even if the spectral gap is very small, it may be feasible to choose suitable values for I and I' that result in a distance D significantly larger than the spectral gap, while still achieving a large relative overlap $p_r(I, I')$. Theorem 3 states that as long as the relative overlap is larger than 0.71, we can estimate the ground-state energy to precision ϵ, where the maximal run time is $T_{\max} = \widetilde{\Theta}(D^{-1}) + \delta/\epsilon$ and the total run time T_{total} is approximately $\widetilde{\mathcal{O}}(p_0^{-2}\delta^{-2}(D^{-1} + \delta/\epsilon))$. Hence this algorithm is particularly useful when $D \gg \epsilon$. As the relative overlap approaches 1, δ can be chosen to be arbitrarily small.

C. Related works

Based on the generalized uncertainty relation [21], there exists a uniform complexity lower bound for the problem of phase estimation [22], i.e., the square of the error is always $\Omega(N_s^{-1}N^{-2})$ in the expectation sense, where N is the query depth (with $\tau = 1$) and N_s is the number of repetitions. In our case, to estimate the ground-state energy with precision ϵ, we have $N_s T_{\max}^2 = \Omega(\epsilon^{-2})$, where $T_{\max}$ is the maximal run time. From this perspective, the Hadamard test [with $N_s = \mathcal{O}(\epsilon^{-2})$ and $T_{\max} = \mathcal{O}(1)$] and QPE [with $N_s = \mathcal{O}(1)$ and $T_{\max} = \mathcal{O}(\epsilon^{-1})$] are at the two ends of the spectrum. It is possible to achieve a measurement-depth trade-off by setting $N_s = \mathcal{O}(\epsilon^{-2(1-\alpha)}), T_{\max} = \mathcal{O}(\epsilon^{-\alpha})$ for some $0 < \alpha < 1$ [23]. However, the total cost will be at least $N_s T_{\max} = \mathcal{O}(\epsilon^{\alpha-2})$. Hence when $\alpha < 1$, this strategy does not satisfy the Heisenberg-limited scaling.

Our work is related to robust phase estimation (RPE) in the context of quantum metrology for single-qubit systems, which was first proposed in Ref. [24]. RPE satisfies the Heisenberg-limited scaling and allows the input state to be an inexact eigenstate, as long as the overlap with the desired eigenstate is larger than a certain constant [25]. Due to these advantages, RPE has been applied in quantum metrology, as well as other systems that can be viewed as effective single-qubit systems [25–29]. Empirical observations also suggest that the maximal run time of RPE may be smaller than π/ϵ. However, we are not aware of theoretical analysis on this aspect of the algorithm. (Note: After the submission of this manuscript, recent analysis in Ref. [30] shows that RPE can also achieve short maximal runtime.) It is possible to generalize RPE to perform phase estimation of general n-qubit systems. That being said, QCELS and multilevel QCELS can also be applied for parameter estimation in quantum metrology, with a provably short circuit depth.

There are a few other phase-estimation algorithms that also use a single ancilla qubit. The efficiency of the algorithms has so far been demonstrated numerically. Reference [31] develops a postprocessing technique to extract eigenvalues from phase-estimation data based on a classical time-series (or frequency) analysis. Reference [32] proposes a method that estimates $\langle\psi|\exp(-itH)|\psi\rangle$ first and then performs a classical Fourier transform to estimate the eigenvalues. A very different type of algorithm for ground-state energy estimation is the variational quantum eigensolver (VQE) [23,33,34], which constructs a variational ansatz $|\psi(\theta)\rangle$ to approximate the lowest eigenvector $|\psi_0\rangle$ and the parameter θ of the ansatz is adjusted to minimize the energy $\langle\psi(\theta)|H|\psi(\theta)\rangle$. The advantage of the VQE is that the quantum circuit is very simple because short-depth circuits (often without using ancilla qubits) are enough to estimate $\langle\psi(\theta)|H|\psi(\theta)\rangle$. However, the efficiency and accuracy of VQE largely depend on the representation power of the variational ansatz $\psi(\theta)$ and the solver of the nonconvex optimization problem. Similar to VQE, there are also other algorithms that try to perform phase estimation using the quantum states generated in the time evolution, such as the quantum imaginary time evolution (QITE) algorithm [35] and some methods based on the classical Krylov-subspace method, such as the quantum subspace diagonalization [14,36]. However, these methods also lack a provable complexity upper bound and existing theoretical analysis on quantum subspace diagonalization methods [37] has not been able to reveal the advantage of such methods compared to classical QPE methods.

For ground-state energy estimation, a number of quantum algorithms [13,15,38–41] have been developed for ground-state energy estimation using the Hamiltonian-evolution input model. However, the maximal run time of all existing works satisfying the Heisenberg-limited scaling is at least C/ϵ for some constant $C = \Omega(1)$ that is independent from the overlap p_0. Take the method in Ref. [15], for instance, which uses the same quantum circuit as in Fig. 1 to generate the input data and can estimate λ_0 with Heisenberg-limited scaling for any $p_0 > 0$. The method uses a Fourier filter to approximate the shifted sign function. To resolve the ground-state energy to precision ϵ, the shifted sign function defined on $[-\pi, \pi)$ should make a transition from 1 to 0 within a small interval of size $\epsilon/2$. The maximal run time is $\widetilde{\mathcal{O}}(\epsilon^{-1})$ and the preconstant is larger than π (see [15, Appendix A]). A similar mechanism of constructing filtering functions is used in the near-optimal ground-state preparation and ground-state energy estimation algorithm based on the block-encoding input model [18], the quantum eigenvalue transformation of unitary matrices (QETU) using the Hamiltonian-evolution input model [13], and the statistical approach with a randomized implementation of Hamiltonian evolution [40].

More recently, Ref. [41] has introduced a method that uses the Fourier-filtering techniques from Ref. [15] to generate a rough estimation $\widetilde{\lambda}_0$ for λ_0 in the first step. Then, it uses a derivative Gaussian filter around $\widetilde{\lambda}_0$ to refine the estimation of λ_0. The main result [41, Corollary 1.3] is that if the system has a spectral gap Δ, for any $\alpha \in [0,1]$, the maximal run time can be chosen to be $\widetilde{\mathcal{O}}(\epsilon^{-\alpha}\Delta^{-1+\alpha})$, and the total cost is $\widetilde{\mathcal{O}}(\Delta^{1-\alpha}\epsilon^{-2+\alpha})$. When $\alpha = 1$, this reduces to the previous result in Ref. [15], i.e., both the maximal run time and the total cost scale as $\widetilde{\mathcal{O}}(\epsilon^{-1})$. When $\alpha = 0$, the maximal run time becomes $\widetilde{\mathcal{O}}(\Delta^{-1})$, which can be much smaller than $\widetilde{\mathcal{O}}(\epsilon^{-1})$ when $\Delta \gg \epsilon$, and this is compensated by an increase in the total cost to $\widetilde{\mathcal{O}}(\epsilon^{-2}\Delta)$. We show in Corollary 4 that under the same assumption, we may choose $\delta = \epsilon/\Delta$ in Theorem 3 and the maximal run time of our method is also $\widetilde{\mathcal{O}}(\Delta^{-1})$, while the total run time is $\widetilde{\mathcal{O}}(\epsilon^{-2}\Delta)$. While the maximal run time allowed by Ref. [41, Corollary 1.3] should be at least $\widetilde{\mathcal{O}}(\Delta^{-1})$, our Theorem 3 allows an even shorter maximal run time of $\mathcal{O}(1)$ under proper conditions. For example, in many quantum systems, although the spectral gap Δ is very small, the relative contribution of the ground state to the initial state is significant in some large interval $[\lambda_0, \lambda_0 + D)$, where $D = \Omega(1) \gg \Delta$. Applying the result of Theorem 3 to this system, the maximal run time is $\mathcal{O}\left(D^{-1}\right) = \mathcal{O}(1)$. It may still be difficult to estimate this relative overlap in practice. However, unlike the spectral gap, the relative-overlap condition is aware of the information in the initial state and this relaxed condition may significantly increase the applicability range of our algorithm in practice, especially for certain Hamiltonians with a small spectral gap (see numerical examples in Sec. V B).

D. Organization

The rest of the paper is organized as follows. In Sec. II, we introduce QCELS and multilevel QCELS by assuming that p_0 is larger than a certain constant threshold. We also provide an intuitive analysis why QCELS can reduce the maximal run time when p_0 is close to 1. In Sec. III, we analyze the maximal run time and the total cost of our methods. We then extend the method to any $p_0 > 0$ in Sec. IV. The numerical simulation of our method is provided in Sec. V, where we mainly compare our method with QPE, followed by discussions and future directions in Sec. VI. We provide detailed proofs of our theorems in the appendixes. Appendixes B, C, and D contain the proofs for Theorems 1–3, respectively. In Appendixes A and E, we introduce several technical lemmas that are relevant to our main proof.

II. MAIN METHOD

A. Generating input data from quantum circuit

In Fig. 1, we may:

(1) Set $W = I$, measure the ancilla qubit and define a random variable X_n such that $X_n = 1$ if the outcome is 0 and $X_n = -1$ if the outcome is 1. Then,

$$\mathbb{E}(X_n) = \mathrm{Re}\left(\langle\psi|\exp(-in\tau H)|\psi\rangle\right). \tag{4}$$

(2) Set $W = S^\dagger$, measure the ancilla qubit, and define a random variable Y_n such that $Y_n = 1$ if the outcome is 0 and $Y_n = -1$ if the outcome is 1. Then,

$$\mathbb{E}(Y_n) = \mathrm{Im}\left(\langle\psi|\exp(-in\tau H)|\psi\rangle\right). \tag{5}$$

Given two preset parameters $N, N_s > 0$ and time step $\tau > 0$, we use the quantum circuit in Fig. 1 to prepare the following data set:

$$\mathcal{D}_H = \{(n\tau, Z_n)\}_{n=0}^{N-1}, \tag{6}$$

where Z_n is calculated by running the quantum circuit (Fig. 1) N_s times. More specifically,

$$Z_n = \frac{1}{N_s}\sum_{k=1}^{N_s}\left(X_{k,n} + iY_{k,n}\right). \tag{7}$$

Here, $X_{k,n}$ and $Y_{k,n}$ are independently generated by the quantum circuit (Fig. 1) with different W and satisfy Eqs. (4) and (5), respectively. Hence in the limit $N_s \to \infty$, we have

$$Z_n = \langle\psi|\exp(-in\tau H)|\psi\rangle. \tag{8}$$

To prepare the data set in Eq. (6), the maximal simulation time is $T_{\max} = (N-1)\tau$ and the total simulation time is $N(N-1)N_s\tau/2 = NN_sT_{\max}/2$. To reduce the complexity of our algorithm, it suffices to find an efficient way to postprocess the data set given in Eq. (6) so that a good approximation to λ_0 can be constructed with proper choices of N, N_s, and τ.

B. QCELS and its intuitive analysis

Using the data set given in Eq. (6), we define the mean-square error (MSE):

$$L(r,\theta) = \frac{1}{N}\sum_{n=0}^{N-1}\left|Z_n - r\exp(-i\theta n\tau)\right|^2, \tag{9}$$

where $r \in \mathbb{C}$ and $\theta \in \mathbb{R}$. The approximation to λ_0 is constructed by minimizing the loss function $L(r,\theta)$. Let

$$(r^*, \theta^*) = \arg\min_{r\in\mathbb{C},\theta\in\mathbb{R}} L(r,\theta); \tag{10}$$

then θ^* is an approximation to λ_0 and this defines the QCELS algorithm. Note that once we obtain the data set from the quantum circuit, minimizing $L(r,\theta)$ only requires classical computation.

The mean-square error $L(r,\theta)$ is a quadratic function with respect to r. For a fixed value of θ, minimization with respect to r gives

$$r(\theta) = \frac{1}{N}\sum_{n=0}^{N-1} e^{i\theta n\tau} Z_n \tag{11}$$

and

$$\min_{r\in\mathbb{C}} L(r,\theta) = \frac{1}{N}\sum_{n=0}^{N-1} |Z_n|^2 - \frac{1}{N}\left|\sum_{n=0}^{N-1} Z_n e^{i\theta n\tau}\right|^2. \tag{12}$$

Therefore, minimizing $L(r,\theta)$ is equivalent to maximizing

$$f(\theta) = \left|\sum_{n=0}^{N-1} Z_n e^{i\theta n\tau}\right|^2, \tag{13}$$

which is a nonlinear function with respect to θ. According to the definition of Z_n in Eq. (7) (for the rigorous statement, see Appendix A),

$$|Z_n - \langle\psi|\exp(-in\tau H)|\psi\rangle| = \mathcal{O}(N_s^{-1/2}).$$

Thus, when $N_s \gg 1$ and $p_0 \approx 1$, intuitively we have

$$Z_n \approx \langle\psi|\exp(-in\tau H)|\psi\rangle = \exp(i\lambda_0 n\tau), \tag{14}$$

which implies

$$f(\theta) \approx \left|\frac{\exp(i(\lambda_0-\theta)N\tau)-1}{\exp(i(\lambda_0-\theta)\tau)-1}\right|^2 = \left|\frac{\sin((\lambda_0-\theta)N\tau/2)}{\sin((\lambda_0-\theta)\tau/2)}\right|^2. \tag{15}$$

Recall that $T_{\max} = N\tau$. When N is large enough, the maximum of $|(\sin((\lambda_0-\theta)N\tau/2))/(\sin((\lambda_0-\theta)\tau/2))|$ occurs at $\theta = \lambda_0$ and the closest local maximal θ^* satisfies $|\theta^* - \lambda_0| \geq \pi/T_{\max}$. Therefore, to find the maximal value of $f(\theta)$ on the interval $[-\pi,\pi)$, we may choose a uniform grid of size up to $\lceil T_{\max}\rceil$ and perform gradient ascent from each grid point. By maximizing over the values from all the local maxima, we can robustly find the global maxima of f (and hence the global minima of the loss function L). As an illustration, in Fig. 3, we give an example of the landscape of the loss function and compare the optimization results with different initial guesses.

When $N, N_s \gg 1$ and p_0 is sufficiently large, we can show that θ^* is a good approximation to λ_0 with relatively small depth. The rigorous theoretical results are presented in Sec. III. Here, we briefly introduce the intuition and leave more details to Sec. II C.

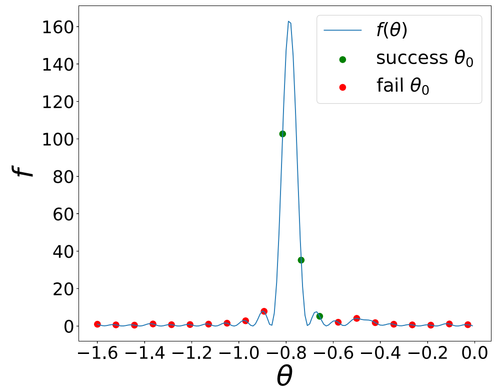

FIG. 3. The landscape of the objective function $f(\theta)$ in Eq. (13) and a number of possible choices of the initial guess θ_0 with $T_{\max} = 80$ and the eight-site TFIM model (see details in Sec. V A). Here, $p_0 = 0.8$ and the landscape for other values of p_0 is similar. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is used to maximize the objective function $f(\theta)$ for different initial guesses θ_0. The error threshold is set as 0.01, meaning that the optimization problem is successfully solved if $|\theta^* - \mathrm{argmax}_\theta f(\theta)| < 0.01$.

Let $\{(\lambda_m, |\psi_m\rangle)\}_{m=0}^{M-1}$ be the set of eigenpairs of the Hamiltonian H and let the distance of θ^* to λ_m be

$$R_m = \left|(\lambda_m - \theta^*)\tau \bmod [-\pi,\pi)\right|. \tag{16}$$

Our goal is to prove that R_0 is small. When $N_s \gg 1$, intuitively we have

$$Z_n \approx \langle\psi|\exp(-in\tau H)|\psi\rangle = p_0\exp(-i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m \exp(-i\lambda_m n\tau), \tag{17}$$

where $p_m = |\langle\psi_m|\psi\rangle|^2$ is the overlap between the initial quantum state and the mth eigenvector. Hence in the limit $N_s \to \infty$,

$$L(r,\theta) = \frac{1}{N}\sum_{n=0}^{N-1}\left|p_0\exp(-i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m\exp(-i\lambda_m n\tau) - r\exp(-i\theta n\tau)\right|^2. \tag{18}$$

Similarly to the above computation, we find that minimizing the right-hand side is equivalent to maximizing the following function:

$$f(\theta) = \left|\sum_{n=0}^{N-1}\left[p_0 \exp(i(\lambda_0-\theta)n\tau) + \sum_{m=1}^{M-1} p_m \exp(i(\lambda_m-\theta)n\tau)\right]\right|^2$$
$$= \left|p_0 \frac{\exp(i(\lambda_0-\theta)N\tau)-1}{\exp(i(\lambda_0-\theta)\tau)-1} + \sum_{m=1}^{M-1} p_m \frac{\exp(i(\lambda_m-\theta)N\tau)-1}{\exp(i(\lambda_m-\theta)\tau)-1}\right|^2. \tag{19}$$

Therefore,

$$f(\lambda_0) = \left|p_0 N + \sum_{m=1}^{M-1} p_m \frac{\exp(i(\lambda_m-\lambda_0)N\tau)-1}{\exp(i(\lambda_m-\lambda_0)\tau)-1}\right|^2. \tag{20}$$

When the overlap between the initial state and $|\psi_0\rangle$ dominates, i.e., $p_0 > \sum_{m=1}^{M-1} p_m$, we can use the first equality in Eq. (19) to obtain

$$\sqrt{f(\lambda_0)} \geq \sum_{n=0}^{N-1}\left(p_0 - \sum_{m=1}^{M-1} p_m\right) = N\left(p_0 - \sum_{m=1}^{M-1} p_m\right). \tag{21}$$

Note that

$$\left|\frac{\exp(i(\lambda-\theta)N\tau)-1}{\exp(i(\lambda-\theta)\tau)-1}\right| = \left|\frac{\sin((\lambda-\theta)\tau N/2)}{\sin((\lambda-\theta)\tau/2)}\right| \leq \frac{\pi}{|(\lambda-\theta)\tau \bmod [-\pi,\pi)|}. \tag{22}$$

The second equality of Eq. (19) together with Eq. (22) gives

$$\sqrt{f(\theta^*)} \leq \frac{\pi\sum_{m=0}^{M-1} p_m}{\min_{m=0}^{M-1}|(\lambda_m-\theta^*)\tau \bmod [-\pi,\pi)|} = \frac{\pi}{\min_m R_m}. \tag{23}$$

As a result,

$$\left(p_0 - \sum_{m=1}^{M-1} p_m\right) N \leq \sqrt{f(\lambda_0)} \leq \sqrt{f(\theta^*)} \leq \frac{\pi}{\min_m R_m}. \tag{24}$$

Equation (24) implies that there must exist some m^* such that $\left(p_0 - \sum_{m=1}^{M-1} p_m\right) R_{m^*} \leq \pi/N$ and that θ^* must be close to one of the eigenvalues. Since $p_0 > \sum_{m=1}^{M-1} p_m$, it is reasonable to expect that this eigenvalue should be λ_0 ($m^*=0$) and we first have

$$|(\lambda_0-\theta^*) \bmod [-\pi/\tau,\pi/\tau)| \leq \frac{\pi}{T_{\max}\left(p_0 - \sum_{m=1}^{M-1} p_m\right)}. \tag{25}$$

When p_0 is very close to 1, we can further improve the bound given in Eq. (25). Since θ^* is the maximal point,

$$\left(p_0 - \sum_{m=1}^{M-1} p_m\right) N \leq \sqrt{f(\lambda_0)} \leq \sqrt{f(\theta^*)} \leq \left|\frac{\sin(NR_0/2)}{\sin(R_0/2)}\right| + (1-p_0)N, \tag{26}$$

where the last inequality comes from the first equality of Eq. (22). This implies that

$$\left|\frac{\sin(NR_0/2)}{\sin(R_0/2)}\right| \geq (3p_0-2)N. \tag{27}$$

If we define $\delta = R_0 N$, we have $\sin(N(\delta/2N))/\sin(\delta/2N) \geq (3p_0-2)N$. When $\delta < \pi$, using the Taylor expansion, we have

$$\frac{\sin(N(\delta/2N))}{\sin(\delta/2N)} \approx N\left(1-\frac{\delta^2}{24}\right) \geq (3p_0-2)N.$$

Combining this with Eq. (27), we have

$$\delta^2 \approx 72(1-p_0). \tag{28}$$

Therefore, as $p_0 \to 1$, $\delta = \Theta(\sqrt{1-p_0}) \to 0$. Note that $\sin(Nx)/\sin(x)$ is monotonically decreasing on $[0,\pi/(2N)]$; if we can prove a loose bound $R_0 \in [0,\pi/N]$, then it can be refined to

$$R_0 \leq \frac{\delta}{N} \tag{29}$$

or

$$|(\lambda_0-\theta^*) \bmod [-\pi/\tau,\pi/\tau)| \leq \frac{\delta}{T_{\max}} = \epsilon. \tag{30}$$

In other words, it suffices to choose the maximal run time $T_{\max} = \delta/\epsilon$.

The above intuitive analysis summarizes the reason why QCELS can estimate λ_0 with a short-depth circuit. The precise statement is given in Theorem 1 and the behavior of the preconstant δ is demonstrated in Fig. 2(b).

C. Multilevel QCELS

Even though QCELS can reduce the maximal run time, it does not satisfy the Heisenberg-limited scaling. To see this, note that $T_{\max} = (N-1)\tau = \mathcal{O}(\epsilon^{-1})$. If τ is a constant, then the total simulation time T_{total} is $\Omega(\epsilon^{-2})$. We may also attempt to choose N to be a constant and let $\tau = \mathcal{O}(\epsilon^{-1})$. However, the loss function is a periodic function in θ with period $2\pi/\tau$. So we can only obtain $|(\lambda_0 - \theta^*) \bmod [-\pi/\tau, \pi/\tau)| = \mathcal{O}(\epsilon)$, which is a meaningless estimate, since $\pi/\tau = \mathcal{O}(\epsilon)$.

In this section, we provide a multilevel QCELS algorithm that maintains the reduced maximal run time and satisfies the Heisenberg-limited scaling. Roughly speaking, we construct a sequence of data sets $\{\mathcal{D}_{H,j}\}$ using an increasing sequence of $\{\tau_j\}$. The maximal simulation time of the algorithm $T_{\max} = N\tau_J$ and the total simulation time of the algorithm $T_{\text{total}} = \sum_{j=1}^{J} N(N-1)N_s\tau_j$. The parameters in the algorithm should be chosen properly. The increasing speed of τ_j should also be chosen properly. If τ_j increases too slowly, we need more iteration steps, which increases the total cost. If τ_j increases too rapidly, there might exist more than one candidate for the estimation interval for λ_0 in each iteration. We propose the choice of $\tau_{j+1} = 2\tau_j$ (for the precise choice of $\{\tau_j\}$, see Eq. (35)), and this procedure is similar to Kitaev's algorithm [3]. Each solution of the optimization problem based on $\{\mathcal{D}_{H,j}\}$ helps us shrink the estimation interval and finally we obtain a small estimation interval for λ_0. The pseudocode of the multilevel QCELS algorithm is given in Algorithm 1.

III. COMPLEXITY ANALYSIS WITH A LARGE OVERLAP

To analyze the complexity of the multilevel QCELS method in Algorithm 1, we need to find an upper bound of the maximal or total simulation time for finding an ϵ approximation to λ_0. In this section, we assume that the initial overlap is large, i.e., $p_0 > 0.71$. The extension to the small p_0 regime is discussed in Sec. IV.

For each $0 \le n \le N-1$, we define

$$E_n = Z_n - \langle\psi|\exp(-in\tau H)|\psi\rangle = Z_n - \left(p_0\exp(-i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m \exp(-i\lambda_m n\tau)\right), \quad (31)$$

which corresponds to the error that occurs in the expectation estimation given in Eq. (7). Note that the $\{E_n\}$ are independent complex random variables with zero expectation and bounded magnitude. Using classical probability theory, we can give a sharp tail bound for E_n with respect to N, N_s, which is important for us in order to derive a choice of N, N_s in our algorithm. The detailed discussion and tail bounds for E_n can be found in Appendix A.

Using a proper tail bound for E_n, we can analyze the performance of QCELS in Theorem 1. This also corresponds to an iteration in Algorithm 1.

Theorem 1 (complexity of QCELS, informal): *Let θ^* be the solution of QCELS in Eq. (10). Given $p_0 > 0.71$, $0 < \eta < 1/2$, $0 < \epsilon < 1/2$, we can choose*

$$\delta = \Theta(\sqrt{1-p_0}) \quad (32)$$

Preparation: Number of data pairs: N; number of samples: N_s; number of iterations: J; sequence of time steps: $\{\tau_j\}_{j=1}^{J}$; Quantum oracle: $\{\exp(-i\tau_j H)\}_{j=1}^{J}$;
Running:
$\lambda_{\min} \leftarrow -\pi$; $\lambda_{\max} \leftarrow \pi$; ▷ $[\lambda_{\min}, \lambda_{\max}]$ contains λ_0
$j \leftarrow 1$;
for $j = 1, \ldots, J$ **do**
  Generate the data set in Eqs. (6) and (7) using the circuit in Fig. 1 with $t_n = n\tau_j$.
  Define loss function $L(r,\theta)$ according to (9).
  Minimizing the loss function (by maximizing Eq. (13)).

$$(r_j^*, \theta_j^*) \leftarrow \underset{r\in\mathbb{C},\theta\in[-\lambda_{\min},\lambda_{\max}]}{\arg\min} L(r,\theta),$$

  $\lambda_{\min} \leftarrow \theta_j^* - \frac{\pi}{2\tau_j}$; $\lambda_{\max} \leftarrow \theta_j^* + \frac{\pi}{2\tau_j}$ ▷ Shrink the search interval by 1/2
end for
Output: θ^*

Algorithm 1. Multilevel quantum complex exponential least squares *and*

$$NN_s = \widetilde{\Omega}\left(\delta^{-(2+o(1))}\right), \quad \min\{N, N_s\} = \Omega(1),$$
$$T = N\tau = \frac{\delta}{\epsilon}, \tag{33}$$

so that

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau)\right| < \epsilon\right) \geq 1 - \eta. \tag{34}$$

The precise statement of Theorem 1 and the proof are given in Appendix B. To show Eq. (34), two parts of the error need to be controlled. First, as discussed before, we should control E_n by increasing the number of samples N_s, so that it does not change the loss function too much. This is particularly important as $p_0 \to 1$. When the condition given in Eq. (33) is satisfied, the probability of $|E_n| = \mathcal{O}(\delta^2) = \mathcal{O}(1 - p_0)$ is at least $1 - \eta$. The second part of the error comes from the pollution from eigenvalues other than λ_0 when $p_0 < 1$. As a result, δ cannot be arbitrarily small and needs to satisfy the relation in Eq. (32) as $p_0 \to 1$.

Although Theorem 1 is a complexity result for one step of QCELS, it also shows that this basic version of QCELS cannot satisfy the Heisenberg-limited scaling. Fix $0 < \delta < 1$; then, if we choose the parameter according to the condition in Eq. (33) and $N = \Theta(1)$, we have $\tau = \Theta(1/\epsilon)$. However, this makes the length of the estimation interval $|[-\pi/\tau, \pi/\tau)| = \Theta(\epsilon)$ and the resulting estimation θ^* is meaningless. To solve this problem, we need to choose $\tau = \Theta(1)$. Then, we should set $N = \Theta(1/\epsilon)$ to satisfy the third condition in Eq. (33); this finally makes the total cost $T_{\text{total}} = N(N-1)\tau/2 = \Omega(1/\epsilon^2)$, which violates the Heisenberg-limit scaling.

Theorem 1 can be used to describe the maximal run time of Algorithm 1. Using Eq. (34), we can obtain many candidates of the estimation interval for λ_0 after solving each minimization problem. On the other hand, we can choose τ_j properly so that only one of this candidate survives in each iteration. After eliminating other candidates, the estimation in Eq. (34) can be directly written as $|\theta^* - \lambda_0| < \delta/T$, which implies that $T_{\max} = \delta/\epsilon$ is enough to obtain ϵ precision in our algorithm. Now, we are ready to introduce the choice of the parameters and the main complexity result of Algorithm 1.

Theorem 2 (complexity of multilevel QCELS, informal): *Let θ^* be the output of Algorithm 1. Given $p_0 > 0.71$, $0 < \eta < 1/2$, $0 < \epsilon < 1/2$, we can choose δ according to Eq. (32),*

$$J = \lceil \log_2(1/\epsilon) \rceil + 1, \quad \tau_j = 2^{j-1-\lceil \log_2(1/\epsilon)\rceil}\frac{\delta}{N\epsilon},$$
$$\forall 1 \leq j \leq J. \tag{35}$$

Choose $NN_s = \widetilde{\Theta}\left(\delta^{-(2+o(1))}\right)$. Then,

$$T_{\max} = N\tau_J = \frac{\delta}{\epsilon},$$
$$T_{\text{total}} = \sum_{j=1}^{J} N(N-1)N_s\tau_j/2 = \widetilde{\Theta}\left(\delta^{-(1+o(1))}\epsilon^{-1}\right)$$

and

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi, \pi)\right| < \epsilon\right) \geq 1 - \eta.$$

The precise statement of Theorem 2 and the proof are given in Appendix C. Theorem 2 shows that as $p_0 \to 1$, the multilevel QCELS algorithm satisfies the Heisenberg-limited scaling and the maximal run time can be much smaller than π/ϵ. On the other hand, there is a trade-off between the maximal simulation time and the total simulation time. In particular, $N_s N = \widetilde{\Theta}\left(\delta^{-2}\right)$ diverges as $\delta \to 0$. This implies that, although T_{total} achieves the Heisenberg-limited scaling, the preconstant may become too large if the circuit depth is forced to be very small.

IV. GROUND-STATE ENERGY ESTIMATION WITH A SMALL INITIAL OVERLAP

When p_0 is smaller than the threshold value of 0.71, our strategy is to find a way to "increase p_0" in the input data. If the system has a spectral gap $\Delta = \lambda_1 - \lambda_0 \gg \epsilon$, we can then use the algorithm from a previous work [15] to construct an eigenvalue filter to effectively filter out the contribution above $\lambda_0 + \Delta/2$ in the initial state, using a circuit with maximal run time $\widetilde{\Theta}(\Delta^{-1})$. The effective value of p_0 in the filtered data can be approximately 1 and the multilevel QCELS algorithm becomes applicable.

The spectral gap is a property of the Hamiltonian. For many quantum systems of interest, the spectral gap Δ can be very small. Since QCELS can accurately estimate the eigenvalues starting from an inexact eigenstate, the filtering step does not need to be perfect either if p_0 is small. Consider an interval I containing λ_0, a larger interval $I' \supset I$, and define the distance

$$D = \text{dist}((I')^c, I) = \min_{x_1 \notin I', x_2 \in I} |x_1 - x_2|. \tag{36}$$

Then, the *relative overlap* of the initial vector $|\psi\rangle$ with the ground state [as defined in Eq. (3)], denoted by $p_r(I, I')$, plays the role of the effective value of p_0. Specifically, if $p_r(I, I') \geq 0.71$, we can effectively filter out the contribution from $(I')^c$ in the initial state using the algorithm in Ref. [15] with maximal run time $\widetilde{\Theta}(D^{-1})$. After the filtering operation, the relative overlap plays the role of p_0 in the previous section and we can apply the multilevel QCELS algorithm to estimate λ_0 with respect to the filtered data.

The concept of the relative overlap may allow us to estimate the ground-state energy for certain small gapped systems with a short-depth circuit, especially when $p_r(I, I') \approx 1$ and D is much larger than the spectral gap Δ. Furthermore, unlike the spectral-gap assumption, which is a property of the Hamiltonian, the relative overlap is a property of the initial state. This introduces flexibility in the initial-state design that may be useful for future explorations.

A. Algorithm

The modified algorithm has three steps:

(1) *Rough estimation of* λ_0. We obtain two rough-estimation intervals $I \subset I' \subset [-\pi, \pi]$ for λ_0, meaning that $\lambda_0 \in I$ and $p_r(I, I') \approx 1$.

(2) *Eigenvalue filtering to remove high-energy contribution.* Define a polynomial $F_q(x) = \sum_{l=-d}^{d} \hat{F}_{l,q} \exp(ilx)$ such that

$$(1)\ |F_q(x) - 1| \le q, \quad \forall x \in I;$$
$$(2)\ |F_q(x)| \le q, \quad \forall x \in [-\pi, \pi] \setminus I'. \tag{37}$$

We use Ref. [15, Lemma 6] to construct F_q that satisfies Eq. (37) with $d = \Theta(D^{-1}\mathrm{polylog}(q^{-1}))$, $\left|\hat{F}_{l,q}\right| = \Theta(|l|^{-1})$, and $\sum_{l=-d}^{d} \left|\hat{F}_{l,q}\right| = \Theta(\log(d))$.

(3) *Refined estimation of* λ_0 *with multilevel QCELS.* We can apply Algorithm 1 with the filtered data set (see detail in Algorithm 2) to obtain an accurate estimation of the ground-state energy.

Define

$$\hat{F}_{l,q} = \left|\hat{F}_{l,q}\right| e^{i\phi_{l,q}}, \quad \beta_l = \left|\hat{F}_{l,q}\right| / \sum_{l=-d}^{d} \left|\hat{F}_{l,q}\right|. \tag{38}$$

The main algorithm is summarized in Algorithm 2. Here, the "DataGenerator" is used to filter out the high-energy contribution. According to the construction of F_q and $Z_{n,q}$, we have

$$\left|Z_{n,q} - p_0 \exp(-i\lambda_0 n\tau)\right| \approx \Bigg|p_0(F_q(\lambda_0) - 1)\exp(-i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m F_q(\lambda_m)\exp(-i\lambda_m n\tau)\Bigg| \le \frac{\left(1 - p_r(I, I')\right)p_0}{p_r(I, I')} + q.$$

This implies that the new data set successfully removes the high-energy contribution when $q \ll 1$ and $p_r(I, I') \approx 1$, and the solution θ^* to the optimization problem in Eq. (39) should be a good approximation to λ_0. We also need a sequence of data sets $\left\{\left(n\tau_j, Z_{n,q}\right)\right\}_{n=0}^{N-1}$ with an increasing sequence of $\{\tau_j\}_{j=1}^{J}$ to shrink and keep the correct estimation interval. Similarly to Algorithm 1, we propose the choice of $\tau_{j+1} = 2\tau_j$ (for the precise choice of τ_j and J, see Theorem 3).

B. Complexity analysis

In this section, we analyze the complexity of Algorithm 2 to show that it can reduce the circuit depth and maintain the Heisenberg-limited scaling. Define the expectation estimation error:

$$E_{n,q} = Z_{n,q} - \langle\psi| F_q(H)\exp(-in\tau H)|\psi\rangle = \frac{1}{N_s}\sum_{k=1}^{N_s} Z_{k,n,q} - \mathbb{E}(Z_{k,n,q}) \tag{39}$$

and

$$G_{n,q} = Z_{n,q} - p_0\exp(-i\lambda_0 n\tau) = E_{n,q} + p_0(F_q(\lambda_0) - 1)\exp(-i\lambda_0 n\tau) + \sum_{k=1}^{M-1} p_k F_q(\lambda_k)\exp(-i\lambda_k n\tau).$$

Using Eq. (37), we obtain

$$|G_{n,q}| \le |E_{n,q}| + q + \frac{\left(1 - p_r(I, I')\right)p_0}{p_r(I, I')}.$$

By choosing a large N_s to reduce the expectation error $|E_{n,q}|$, increasing the quality of the filter to reduce the approximation error q, and assuming $p_r(I, I') \approx 1$, we can effectively reduce the error $G_{n,q}$. These choices will let us find a good approximation of λ_0 by solving the optimization problem. In Algorithm 2, constructing the loss function contains two steps. First, to construct the eigenvalue filter, the required circuit depth is $d = \widetilde{\Theta}(D^{-1})$. We then combine the eigenvalue filter with the Algorithm 1 to construct the filtered data set. This increases the circuit depth to $T_{\max} = d + \delta/\epsilon = \widetilde{\Theta}(D^{-1}) + \delta/\epsilon$. To construct a sufficiently accurate loss function, the number of repetitions is $NN_s = \widetilde{\Theta}\left(p_0^{-2}\delta^{-(2+o(1))}\right)$. This implies the total evolution time $T_{\text{total}} = \widetilde{\Theta}\left(p_0^{-2}\delta^{-(2+o(1))}\left(D^{-1} + \delta/\epsilon\right)\right)$.

The result is summarized in the following theorem.

Theorem 3 (complexity of Algorithm 2, informal): *Given any failure probability* $0 < \eta < 1$, *target precision* $0 < \epsilon < 1/2$, *and knowledge of the relative overlap* $p_r(I, I') \ge 0.71$, *we can set* $\delta = \Theta\left(\sqrt{1 - p_r(I, I')}\right)$, $d = \widetilde{\Theta}(D^{-1})$, $q = \Theta(p_0\delta^2)$, $NN_s = \widetilde{\Omega}\left(p_0^{-2}\delta^{-(2+o(1))}\right)$, min

Preparation: Number of data pairs: N; number of samples: N_s; number of iterations: J; sequence of time steps: $\{\tau_j\}_{j=1}^{J}$;
Prepare a rough estimation:
Generate estimation intervals of λ_0 such that $\lambda_0 \in I \subset I'$.
Running:
$\lambda_{\min} \leftarrow -\pi$; $\lambda_{\max} \leftarrow \pi$; ▷ $[\lambda_{\min}, \lambda_{max}]$ is the estimation interval of λ_0
$j \leftarrow 1$;
while $j \le J$ **do**
  Generate a data set $\{(n\tau_j, Z_{n,q})\}_{n=0}^{N-1}$ using DataGenerator(τ_j, d, N_s).
  Define loss function $L_j(r,\theta)$

$$L_j(r,\theta) = \frac{1}{N}\sum_{n=0}^{N-1} \left|Z_{n,q} - r\exp(-i\theta n\tau_j)\right|^2 . \quad (39)$$

  Minimizing loss function: $(r_j^*, \theta_j^*) \leftarrow \mathrm{argmin}_{r\in\mathbb{C},\theta\in[-\lambda_{\min},\lambda_{\max}]} L_j(r,\theta)$.
  $\lambda_{\min} \leftarrow \theta_j^* - \frac{\pi}{2\tau_j}$; $\lambda_{\max} \leftarrow \theta_j^* + \frac{\pi}{2\tau_j}$ ▷ Shrink the search interval by 1/2.
  $j \leftarrow j+1$
end while
Output: θ^*

function DataGenerator(τ, d, N_s) ▷ Generate the filtered data set
  $k \leftarrow 1$;
  while $k \le N_s$ **do**
    Generate a random variable $r \in [-d, d-1, \ldots, d]$ with the distribution $\mathbb{P}(r = l) = \beta_l$.
    Run the quantum circuit (Figure 1) with $t_n = r + n\tau$ and $W = I$ to obtain $\widetilde{X}_{k,n}$.
    Run the quantum circuit (Figure 1) with $t_n = r + n\tau$ and $W = S^\dagger$ to obtain $\widetilde{Y}_{k,n}$.
    $Z_{k,n,q} \leftarrow \left(\widetilde{X}_{k,n} + i\widetilde{Y}_{k,n}\right)\exp(i\phi_{r,q})\left(\sum_{j=-d}^{d}\left|\hat{F}_{j,q}\right|\right)$.
    $k \leftarrow k+1$
  end while
  $Z_{n,q} \leftarrow \frac{1}{N_s}\sum_{k=1}^{N_s} Z_{k,n,q}$.
end function

Algorithm 2. Multilevel QCELS based ground-state energy estimation with small initial overlap

$\{N, N_s\} = \widetilde{\Omega}(p_0^{-2})$, and J, τ_j according to Theorem 2 with $T = N\tau_J = \delta/\epsilon$. Then,

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi, \pi)\right| < \epsilon\right) \ge 1 - \eta,$$

where θ^* is the output of Algorithm 2. In particular, the maximal evolution time is $T_{\max} = d + \delta/\epsilon = \widetilde{\Theta}(D^{-1}) + \delta/\epsilon$ and the total evolution time is $T_{\text{total}} = \widetilde{\Theta}\left(p_0^{-2}\delta^{-(2+o(1))}\left(D^{-1} + \delta/\epsilon\right)\right)$.

The detailed statement and the proof of this theorem can be found in Appendix D. This theorem is an analog of Theorem 2. As a special case, we assume that the spectral gap $\Delta = \lambda_1 - \lambda_0$ is much larger than the precision ϵ. We can construct $I = [-\pi, \lambda_{\text{prior}} + \Delta/4]$, $I' = [-\pi, \lambda_{\text{prior}} + 3\Delta/4]$, and $D = \Delta/2$, where λ_{prior} is a rough estimation of λ_0 such that $|\lambda_{\text{prior}} - \lambda_0| \le \Delta/4$. Then, Theorem 3 gives the following complexity estimate.

Corollary 4 (Complexity of Algorithm 2 with a spectral gap): *Given any $0 < \delta < 1$, failure probability $0 < \eta < 1$, target precision $0 < \epsilon < 1/2$, and spectral gap $\Delta = \lambda_1 - \lambda_0$, we can set $d = \widetilde{\Theta}(\Delta^{-1})$, $q = \widetilde{\Theta}\left(p_0\delta^2\right)$, $NN_s = \widetilde{\Theta}\left(p_0^{-2}\delta^{-(2+o(1))}\right)$, $\min\{N, N_s\} = \widetilde{\Theta}(p_0^{-2})$, and J, τ_j according to Theorem 2. Then,*

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi, \pi)\right| < \epsilon\right) \ge 1 - \eta,$$

In particular, we have:

(a) The cost of preparing the rough estimation (construct $\lambda_{\text{prior}}, I, I'$): the maximal evolution time $T_{\max,1} = \widetilde{\Theta}(\Delta^{-1})$ and the total evolution time $T_{\text{total},1} = \widetilde{\Theta}(\Delta^{-1}p_0^{-2})$.
(b) The cost of constructing the loss function: the maximal evolution time $T_{\max,2} = d + \delta/\epsilon = \widetilde{\Theta}(\Delta^{-1}) + \delta/\epsilon$ and the total evolution time $T_{\text{total},2} = \widetilde{\Theta}\left(p_0^{-2}\delta^{-(2+\zeta)}\left(\Delta^{-1} + \delta/\epsilon\right)\right)$.

In Corollary 4, if $\epsilon \ll \Delta$, we can choose $\delta = \epsilon/\Delta$; then $T_{\max} = \max\{T_{\max,1}, T_{\max,2}\} \approx \widetilde{\Theta}(\Delta^{-1})$ and $T_{\text{total}} = T_{\text{total},1} + T_{\text{total},2} \approx \widetilde{\Theta}(p_0^{-2}\Delta\epsilon^{-2})$. This recovers the results of [41, Theorem 1.1]. Note that with such a choice of δ, the total cost does not satisfy the Heisenberg-limited scaling.

V. NUMERICAL SIMULATION

In this section, we numerically demonstrate the efficiency of our method using two different models. In Sec. V A, we assume a large initial overlap and compare the performance of Algorithm 1 with QPE for a transverse-field Ising model. In Sec. V B, we assume that the initial overlap is small and compare the performance of Algorithm 2 with QPE for a Hubbard model. The Hamiltonian is constructed using the QuSpin package [42].

In our numerical experiments, we normalize the spectrum of original Hamiltonian H so that the eigenvalues belong to $[-\pi/4, \pi/4]$. Given a Hamiltonian H, we define the normalized Hamiltonian:

$$\widetilde{H} = \frac{\pi H}{4\|H\|_2}. \tag{40}$$

We then use the QCELS-based Algorithm 1 or Algorithm 2, as well as QPE, to estimate the smallest eigenvalue of $\widetilde{H}$ and measure the error accordingly.

A. Ising model

Consider the one-dimensional transverse-field Ising model (TFIM) model defined on L sites with periodic boundary conditions

$$H = -\left(\sum_{i=1}^{L-1} Z_i Z_{i+1} + Z_L Z_1\right) - g\sum_{i=1}^{m} X_i, \tag{41}$$

where g is the coupling coefficient, Z_i, X_i are Pauli operators for the ith site, and the dimension of H is 2^L. We choose $L = 8, g = 4$. We apply Algorithm 1 (referred to as QCELS for simplicity in this subsection) and QPE to estimate λ_0 of the normalized Hamiltonian $\widetilde{H}$ [see Eq. (41)]. In the test, we set $p_0 = 0.6, 0.8$ and implement QCELS (with $N = 5$ and $N_s = 100$) and QPE 10 times to compare the averaged error. The comparison of the results is shown in Fig. 4. The errors of both QPE and QCELS are proportional to the inverse of the maximal evolution time ($T_{\max}$). However, the constant factor $\delta = T\epsilon$ of QCELS is much smaller than that of QPE. Figure 4 shows that QCELS reduces the maximal evolution time by 2 orders of magnitude, even in this case when $p_0 = 0.6$ is smaller than the theoretical threshold 0.71. This suggests that the numerical performance of QCELS can be significantly better than the theoretical prediction in Theorem 2. The error of QPE is observed to scale as $6\pi/T$. Moreover, the total evolution time (T_{total}) of QCELS is also smaller (by nearly an order of magnitude) than that of QPE.

B. Hubbard model

Consider the one-dimensional Hubbard model defined on L spinful sites with open boundary conditions

$$H = -t\sum_{j=1}^{L-1}\sum_{\sigma\in\{\uparrow,\downarrow\}} c_{j,\sigma}^{\dagger} c_{j+1,\sigma} + U\sum_{j=1}^{L}\left(n_{j,\uparrow} - \frac{1}{2}\right)\left(n_{j,\downarrow} - \frac{1}{2}\right).$$

Here, $c_{j,\sigma}$ ($c_{j,\sigma}^{\dagger}$) denotes the fermionic annihilation (creation) operator on site j with spin σ. $\langle\cdot,\cdot\rangle$ denotes sites that are adjacent to each other. $n_{j,\sigma} = c_{j,\sigma}^{\dagger} c_{j,\sigma}$ is the number operator.

We choose $L = 4, 8$, $t = 1$, and $U = 10$. To implement Algorithm 2 (also referred to as QCELS for simplicity in this subsection) and QPE, we normalize H according to

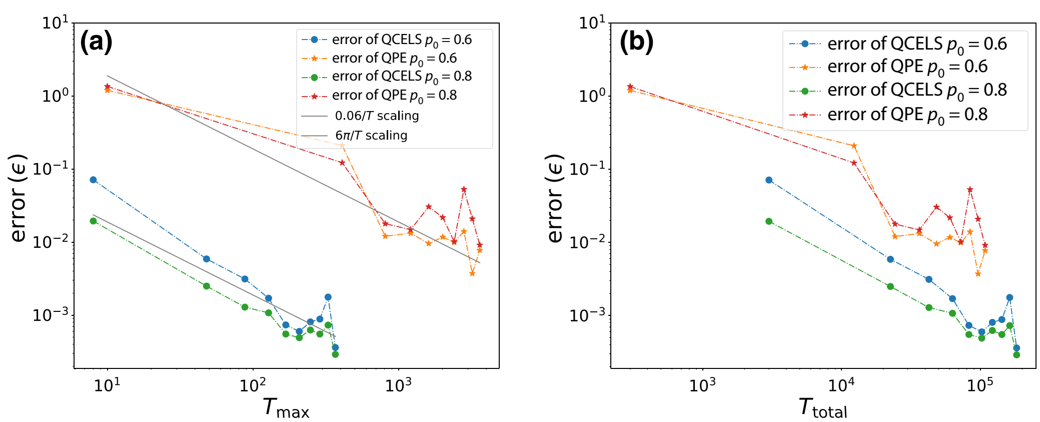

FIG. 4. QPE versus QCELS in the TFIM model with eight sites. The initial overlap is large ($p_0 = 0.6, 0.8$). (a) The depth ($T_{\max}$). (b) The cost (T_{total}). For QCELS, we choose $N = 5$ and $N_s = 100$. J and τ_j are chosen according to Theorem 2. Both methods have the error scaling linearly in $1/T_{\max}$. The constant factor $\delta = T\epsilon$ of QCELS is much smaller than that of QPE.

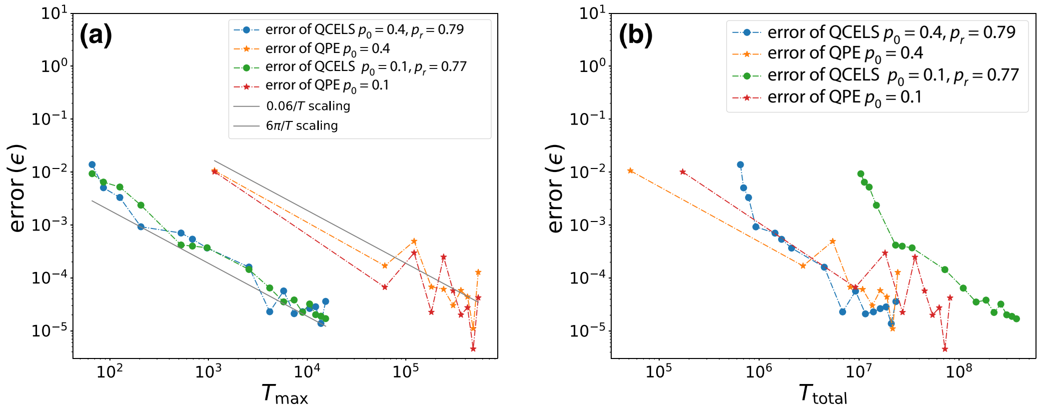

FIG. 5. QPE versus QCELS in the Hubbard model with four sites. The initial overlap is small ($p_0 = 0.1, 0.4$). (a) The depth ($T_{\max}$). (b) The cost (T_{total}). For QCELS, we choose $N = 5$ and $N_s = \lfloor 15p_0^{-2}\log(d)\rfloor$. J, τ_j are chosen according to Corollary 4. Compared with QPE, to achieve the same accuracy, QCELS requires a much smaller circuit depth.

Eq. (41) and choose a small initial overlap ($p_0 = 0.1, 0.4$). Following the method in Sec. IV, we first use the algorithm in Ref. [15] to find a rough estimation, λ_{prior}, of λ_0 such that $|\lambda_{\text{prior}} - \lambda_0| \le D/2$, where D is chosen properly so that the relative overlap $p_r(I, I') > 0.75$ with intervals $I = [-\pi, \lambda_{\text{prior}} + D/2]$ and $I' = [-\pi, \lambda_{\text{prior}} + 3D/2]$. In our test, we set $D = (\lambda_K - \lambda_0)/4$ with $K = \sum_{k=1}^{K} p_k > p_0/3$. We find that the (normalized) relative gap (D) is 0.63 and 0.26 for $L = 4, 8$, respectively. This is significantly larger than the spectral gap, which is 0.018 and 0.005 for $L = 4, 8$, respectively.

After obtaining the rough estimation λ_{prior}, we construct the eigenvalue filtering F_q according to Ref. [15, Lemma 6] to separate I, I'. Noting that $\text{dist}(I, (I')^c) = D$, we set $d = \lfloor 15/D \rfloor$ to ensure a small enough approximation error q. We run QCELS with $N = 5$ and $N_s = \lfloor 15p_0^{-2}\log(d)\rfloor$ and QPE 5 times to compare the averaged error. The results are shown in Fig. 5 (four sites) and Fig. 6 (eight sites). In both figures, it can be seen that the maximal evolution time of QCELS is almost 2 orders of magnitude smaller than that of QPE. The total costs of the two methods are comparable when $p_0 = 0.4$ and the total cost of QCELS is larger than that of QPE when $p_0 = 0.1$, mainly due to the increase of the number of repetitions N_s. We note that, for small p_0, since we first construct the eigenvalue filter F_q, the circuit depth of QCELS is at least $d = \lfloor 15/D \rfloor$. Thus, it is reasonable to choose $\tau_J \ge d$. This directly ensures a relatively small error ($\epsilon \le 10^{-2}$) in our case. In other cases when the gap D is large and only low accuracy is needed, it may be possible to further reduce the circuit depth.

VI. DISCUSSION AND CONCLUSIONS

Due to the relatively transparent circuit structure and the minimal number of required ancilla qubits, the quantum circuit in Fig. 1 is suitable for early fault-tolerant quantum devices and has received significant attention in performing a variety of tasks on quantum computers. Note that all

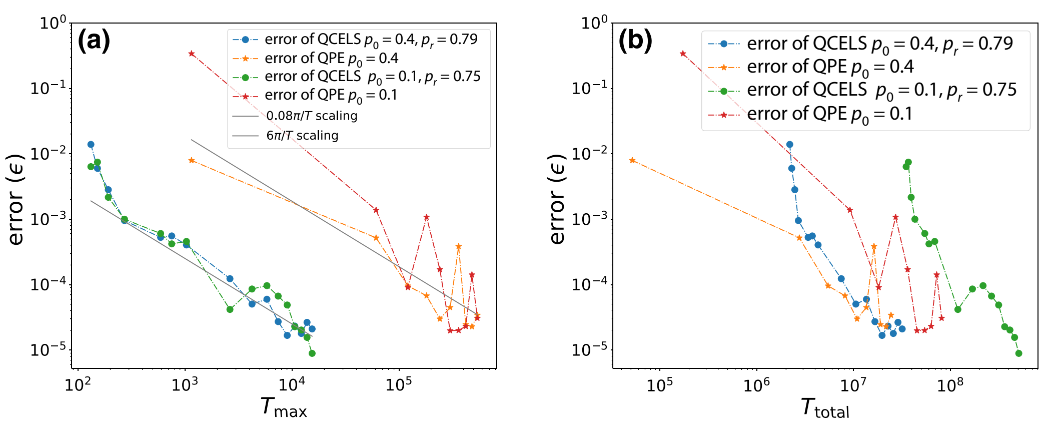

FIG. 6. QPE versus QCELS in the Hubbard model with eight sites. (a) The depth ($T_{\max}$). (b) The cost (T_{total}). For QCELS, we choose $N = 5$ and $N_s = \lfloor 15p_0^{-2}\log(d)\rfloor$. J and τ_j are chosen according to Corollary 4. Compared with QPE, to achieve the same accuracy, QCELS has a much smaller circuit depth.

algorithms in this paper (including the filtering algorithm in Ref. [15]) all use the same circuit and the only difference is in the postprocessing procedure. This paper finds that the circuit in Fig. 1 is even more powerful than previously thought for phase estimation and ground-state energy estimation, especially when the initial overlap p_0, or the relative overlap p_r, is large. The advantage of our method can be theoretically justified when p_0 or p_r approaches 1. The numerical results show that even when p_0 or p_r is away from 1 (e.g., 0.8), our algorithms can still outperform QPE and reduce the maximal run time (and hence the circuit depth) by around 2 orders of magnitude.

Viewed more broadly, the problem of postprocessing the quantum data from the circuit in Fig. 1 is a signal-processing problem using a simple (complex) exponential fitting function. Many methods have been developed in the context of classical signal processing for similar purposes (see, e.g., Refs. [43–46]). We think that at least two features distinguish the quantum setting from the classical counterpart: (1) it is a priority to reduce the maximal run time; and (2) each data point in the signal is inherently noisy and the total number of measurements needs to be carefully controlled. While these classical data-processing methods can be applied to the phase-estimation problem, we are not yet aware of analytical results demonstrating the efficiency of such methods in the quantum setting. Such connections could be an interesting direction to explore in the future.

When the initial overlap p_0 is small, we combine QCELS with the Fourier-filtering algorithm in Ref. [15] to effectively amplify this overlap as shown in Algorithm 2. Another natural choice is to use the quantum eigenvalue transformation of unitary matrices (QETU) [13], which is a more powerful and slightly more complex circuit than that in Fig. 1, to amplify the overlap with the ground state. While we demonstrate applications of QCELS-based algorithms to estimate ground-state energies, such algorithms may be useful in a much wider context, such as estimating excited-state energies and other observables [47]. Simultaneous estimation of multiple eigenvalues using the same circuit is another interesting topic, which may open the door for the development of efficient algorithms for a broader class of quantum systems with small spectral gaps.

Our numerical experiments are available via GitHub [48].

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) Quantum Leap Challenge Institute (QLCI) program under Grant No. OMA-2016245 and by a Google Quantum Research Award. L.L. is a Simons Investigator. We thank Yulong Dong, Ethan Epperly, Daniel Stilck Franca, Peter Johnson, and Yu Tong for discussions.

APPENDIX A: ERROR BOUND FOR THE EXPECTATION ESTIMATION

In this appendix, we give a bound for E_n defined in Eq. (31) with respect to N, N_s. Recall that

$$\begin{aligned} E_n &= Z_n - \langle\psi|\exp(-in\tau H)|\psi\rangle \\ &= Z_n - \left(p_0\exp(-i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m \exp(-i\lambda_m n\tau)\right) \end{aligned}$$

and

$$Z_n = \frac{1}{N_s}\sum_{k=1}^{N_s}\left(X_{k,n} + iY_{k,n}\right),$$

where $X_{k,n}$ and $Y_{k,n}$ are independently generated by the quantum circuit (Fig. 1) with different W and satisfy Eqs. (4) and (5), respectively. We also define the average error term for each θ as

$$\overline{E}_\theta = \frac{1}{N}\sum_{n=0}^{N-1} E_n \exp(i\theta n\tau).$$

In the following part of this appendix, we prove the following bounds for E_n and $\overline{E}_\theta$:

(a) [Eq. (A4) of Lemma 5] Given $0 < \eta < 1/2$, when $\min\{N, N_s\} = \Omega(\log(\eta^{-1}))$,

$$\mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}|E_n| \geq 10^{-3}\right) \leq \eta. \tag{A1}$$

(b) [Eq. (A5) of Lemma 6] Given $0 < \eta < 1/2$ and $0 < \rho, \xi < 10\pi$, when $NN_s = \Omega\left(\xi^{-2}\mathrm{polylog}(\xi^{-1}\eta^{-1})\right)$,

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta\right| \geq \xi\right) \leq \eta. \tag{A2}$$

(c) [Eq. (A6) of Lemma 6] Given $0 < \eta < 1/2$ and $0 < \rho, \xi < 10\pi$, when $NN_s = \Omega\left(\xi^{-2}\mathrm{polylog}(\xi^{-1}\eta^{-1})\right)$,

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta - \overline{E}_{\lambda_0}\right| \geq \rho\xi\right) \leq \eta. \tag{A3}$$

Define $E_{k,n} = X_{k,n} + iY_{k,n} - \langle\psi|\exp(-in\tau H)|\psi\rangle$. Then, we have $\mathbb{E}(E_{k,n}) = 0$ and $|E_{k,n}| \leq 2$. Using the bound and independence of $\{E_{k,n}\}_{k,n}$, we can first show the following lemma.

Lemma 5: *Given* $0 < \eta < 1$, *then*

$$\mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}|E_n| > \frac{2}{\sqrt{N_s}} + \sqrt{\frac{2\ln(1/\eta)}{N}}\right) \leq \eta. \tag{A4}$$

Proof of Lemma 5. First, since $E_n = 1/N_s \sum_{k=1}^{N_s} E_{k,n}$, we have

$$\mathbb{E}(E_n) = 0, \quad \mathbb{E}\left(|E_n|\right) \le \left(\mathbb{E}\left(|E_n|^2\right)\right)^{1/2} \le \frac{2}{\sqrt{N_s}},$$
$$|E_n| \le 2.$$

Since $\{|E_n|\}$ are bounded by 2 and independent of each other, according to Hoeffding's inequality, we have

$$\mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}|E_n| - \frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}(|E_n|) > \delta\right) \le \exp\left(-\frac{N\delta^2}{2}\right).$$

Combining this inequality with $\mathbb{E}\left(|E_n|\right) \le 2/\sqrt{N_s}$ and choosing $\delta = \sqrt{2\ln(1/\eta)/N}$, we prove Eq. (A4). ■

We note that the magnitude bound in Eq. (A4) of Lemma 5 is stronger than what we need in the analysis of the optimization problem given in Eq. (10). Intuitively, for fixed θ, with high probability, we have $\left|\overline{E}_\theta\right| = \mathcal{O}\left(1/\sqrt{NN_s}\right)$, which is a much better bound than Eq. (A4). However, we cannot directly use sub-Gaussian properties in this case since θ is not fixed in the optimization process. On the other hand, we expect that when N, N_s are chosen properly, θ^* should belong to a tiny interval around λ_0, meaning that it is not necessary to give a uniform bound for $\left|\overline{E}_\theta\right|$ for all θ. In particular, to control the effect of the error term, it suffices to bound $\left|\overline{E}_\theta\right|$ when θ is close to λ_0. This bound is stated in the following lemma.

Lemma 6: *Given* $0 < \eta < 1/2$ *and* $\rho > 0$*, then*

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta\right| \ge \left(4\sqrt{2}\log^{1/2}\left(\frac{8\sqrt{N_sN}}{\eta}\right)+\rho\right)\frac{1}{\sqrt{N_sN}}\right) \le \eta \tag{A5}$$

and

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta - \overline{E}_{\lambda_0}\right| \ge \left(4\sqrt{2}\log^{1/2}\left(\frac{8\sqrt{N_sN}}{\eta}\right)+1\right)\frac{\rho}{\sqrt{N_sN}}\right) \le \eta \tag{A6}$$

Proof of Lemma 6. For $\theta_1, \theta_2 \in [\lambda_0 - \rho/T, \lambda_0 + \rho/T]$, we have

$$\begin{aligned}\left|\overline{E}_{\theta_1} - \overline{E}_{\theta_2}\right| &\le \left|\frac{1}{N}\sum_{n=0}^{N-1} E_n\left(\exp(i\theta_1 n\tau) - \exp(i\theta_2 n\tau)\right)\right| \\ &\le \frac{1}{N}\sum_{n=0}^{N-1}|E_n|\,|2\sin((\theta_1-\theta_2)\tau n/2)| \\ &\le |\theta_1 - \theta_2| T,\end{aligned}$$

where we use $|E_n| \le 2$ in the last inequality. This implies that $\overline{E}_\theta$ is a T-Lipschitz function in θ.

Now, we first consider the tail bound of $\overline{E}_\theta$ for fixed $\theta \in [\lambda_0 - \rho/T, \lambda_0 + \rho/T]$. Write

$$\overline{E}_\theta = \frac{1}{N_sN}\sum_{n=0}^{N-1}\sum_{m=1}^{N_s-1} a_{n,m} + ib_{n,m},$$

where

$$a_{n,m} = X_{m,n}\mathrm{Re}\left(\exp(i\theta n\tau)\right) - Y_{m,n}\mathrm{Im}\left(\exp(i\theta n\tau)\right)$$

and

$$b_{n,m} = X_{m,n}\mathrm{Im}\left(\exp(i\theta n\tau)\right) + Y_{m,n}\mathrm{Re}\left(\exp(i\theta n\tau)\right).$$

It is straightforward to see that $\{a_{n,m}\}$ are independent random variables with zero expectation and $|a_{n,m}| \le 2$. Then, according to sub-Gaussian theory, for any $\xi > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{N_sN}\sum_{n=0}^{N-1}\sum_{m=1}^{N_s-1} a_{n,m}\right| \ge \xi\right) \le 2\exp\left(-\frac{NN_s\xi^2}{8}\right).$$

Similar bounds also hold for $\{b_{n,m}\}$. Thus, we obtain that, for any $\xi > 0$,

$$\mathbb{P}\left(\left|\overline{E}_\theta\right| \ge \xi\right) \le 4\exp\left(-\frac{NN_s\xi^2}{32}\right).$$

Given any $\epsilon > 0$, we can find a set of $\lfloor 2\rho/T\epsilon \rfloor$ points $\{\theta_i\}_{i=1}^{\lfloor 2\rho/T\epsilon\rfloor}$ such that for any $\theta \in [\lambda_0 - \rho/T, \lambda_0 + \rho/T]$, there exists i such that $|\theta_i - \theta| \le \epsilon$. Because $\overline{E}_\theta$ is T-Lipschitz, we have

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta\right| \ge \xi + T\epsilon\right) \le \mathbb{P}\left(\sup_{1\le i\le\lfloor\frac{2\rho}{T\epsilon}\rfloor}\left|\overline{E}_{\theta_i}\right| \ge \xi\right) \le \frac{8\rho}{T\epsilon}\exp\left(-\frac{NN_s\xi^2}{32}\right).$$

Choose $\epsilon = \rho/T\sqrt{N_sN}$ and $\xi = 4\sqrt{2}\log^{1/2}\left(8\sqrt{N_sN}/\eta\right) 1/\sqrt{N_sN}$; hence we prove Eq. (A5).

Next, to prove Eq. (A6), we first consider the tail bound of $\overline{E}_\theta - \overline{E}_{\lambda_0}$ for fixed $\theta \in [\lambda_0 - \rho/T, \lambda_0 + \rho/T]$. Write

$$\overline{E}_\theta - \overline{E}_{\lambda_0} = \frac{1}{N_sN}\sum_{n=0}^{N-1}\sum_{m=1}^{N_s-1}\overline{a}_{n,m} + i\overline{b}_{n,m},$$

where

$$\overline{a}_{n,m} = X_{m,n}\mathrm{Re}\left(\exp(i\theta n\tau) - \exp(i\lambda_0 n\tau)\right) - Y_{m,n}\mathrm{Im}\left(\exp(i\theta n\tau) - \exp(i\lambda_0 n\tau)\right)$$

and

$$\overline{b}_{n,m} = X_{m,n}\mathrm{Im}\left(\exp(i\theta n\tau) - \exp(i\lambda_0 n\tau)\right) + Y_{m,n}\mathrm{Re}\left(\exp(i\theta n\tau) - \exp(i\lambda_0 n\tau)\right).$$

Note that $\{a_{n,m}\}$ are independent random variables with zero expectation and $|a_{n,m}| \le 2n\rho/N$. Then, according to Hoeffding's inequality, for any $\xi > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{N_sN}\sum_{n=0}^{N-1}\sum_{m=1}^{N_s-1}\overline{a}_{n,m}\right| \ge \xi\right) \le 2\exp\left(-\frac{NN_s\xi^2}{8\rho^2}\right).$$

Similarly to before, we finally have

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]}\left|\overline{E}_\theta - \overline{E}_{\lambda_0}\right| \ge \xi + T\epsilon\right) \le \mathbb{P}\left(\sup_{1\le i\le\lfloor\frac{2\rho}{T\epsilon}\rfloor}\left|\overline{E}_{\theta_i}\right| \ge \xi\right) \le \frac{8\rho}{T\epsilon}\exp\left(-\frac{NN_s\xi^2}{32\rho^2}\right).$$

Choose $\epsilon = \rho/T\sqrt{N_sN}$ and $\xi = 4\sqrt{2}\log^{1/2}\left(8\sqrt{N_sN}/\eta\right) \rho/\sqrt{N_sN}$; hence we prove Eq. (A6). ■

APPENDIX B: PROOF OF THEOREM 1

For convenience of later discussion, we define a constant

$$\alpha = 1 + \left(\max_{c\in(0,\pi/2]}\frac{\sin(c)}{\pi + c}\right) \approx 1.217. \tag{B1}$$

In this appendix, we prove the following theorem.

Theorem 7 (complexity of QCELS): *Let θ^* be the solution of Eq. (10), let α be defined in Eq. (B1), and let $T = N\tau$ and $p_0 > 0.71$. Given the depth constant $0 < \delta \le 4$ and the failure probability $0 < \eta < 1/2$, if $\delta = \Theta(\sqrt{1-p_0})$ and N, N_s satisfy*

$$NN_s = \Omega\left(\delta^{-(2+o(1))}\mathrm{polylog}\left(\log(\zeta^{-1})\eta^{-1}\right)\right),$$
$$\min\{N, N_s\} = \Omega(\mathrm{polylog}\left(\eta^{-1}\right)), \quad T = \frac{\delta}{\epsilon}, \tag{B2}$$

then

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau)\right| < \frac{\delta}{T}\right) \ge 1 - \eta.$$

We note that Theorem 1 is a direct corollary of Theorem 7. The proof of Theorem 7 contains two steps. We first use the bound of the expectation error in Appendix A to show a proposition that induces a rough-complexity estimation to an iteration in Algorithm 1. Then, we use the rough-complexity result as a prior estimation and study the loss function more carefully to obtain a sharper complexity estimation.

1. Rough estimate

The rough-complexity estimation is stated in the following proposition.

Proposition 8: *Let θ^* be the solution of Eq. (10), let α be defined in Eq. (B1), and let $T = N\tau$ and $p_0 > 0.71$. Given the depth constant $0 < \delta \le 4$ and the failure probability $0 < \eta < 1/2$, if there exists a small enough number $\xi > 0$ such that*

$$\frac{p_0}{(1+\alpha)p_0 - \alpha - \xi} \le \frac{\delta\cos(\delta/10)}{2\sin(\delta/2)} \tag{B3}$$

and N and N_s satisfy

$$NN_s = \Omega\left(\xi^{-2}\mathrm{polylog}\left(\xi^{-1}\eta^{-1}\right)\right),$$
$$\min\{N, N_s\} = \Omega(\mathrm{polylog}\left(\eta^{-1}\right)), \tag{B4}$$

then

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau)\right| < \frac{\delta}{T}\right) \ge 1 - \eta.$$

According to the proposition, we can choose δ, N, and N_s according to the inequality in Eq. (B3). First, we use the parameter ξ in Eq. (B3) to represent an upper bound of "$|E_n|$" [defined in Eq. (31)] and N and N_s are chosen according to the results in Appendix A so that E_n satisfies this upper bound. Second, when δ is very small, to make sure that p_0 satisfies Eq. (B3), we need $p_0 = 1 - \mathcal{O}(\delta^2)$, which implies Eq. (32) in Theorem 1. Finally, the lower bound $p_0 > 0.71$ comes from Eq. (B8), later in the proof of Proposition 8. More specifically, to obtain Eq. (B8), we need $p_0/((1+\alpha)p_0 - \alpha - 10^{-3}) \leq 2$, which implies $p_0 > 0.71$.

We first rewrite the loss function given in Eq. (9). Note that for any fixed θ,

$$
\begin{aligned}
\max_{r\in\mathbb{C}} L(r,\theta) &= \frac{1}{N}\sum_{n=0}^{N-1}\left|p_0\exp(i(\theta-\lambda_0)n\tau) + \sum_{m=1}^{M-1} p_m \exp(i(\theta-\lambda_m)n\tau) + E_n\exp(i\theta n\tau)\right|^2 \\
&\quad - \left|\frac{1}{N}\sum_{n=0}^{N-1} p_0\exp(i(\theta-\lambda_0)n\tau) + \sum_{m=1}^{M-1} p_m \exp(i(\theta-\lambda_m)n\tau) + E_n\exp(i\theta n\tau)\right|^2 \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\left|p_0\exp(i\lambda_0 n\tau) + \sum_{m=1}^{M-1} p_m \exp(i\lambda_m n\tau) + E_n\exp(i\theta n\tau)\right|^2 \\
&\quad - \left|\frac{1}{N}\sum_{n=0}^{N-1}\left[p_0\exp(i(\theta-\lambda_0)n\tau) + \sum_{m=1}^{M-1} p_m \exp(i(\theta-\lambda_m)n\tau) + E_n\exp(i\theta n\tau)\right]\right|^2 .
\end{aligned}
$$

This means that minimizing $L(r,\theta)$ is equivalent to maximizing the magnitude of the following function:

$$
\begin{aligned}
f(\theta) &= \sum_{n=0}^{N-1}\left[\exp(i(\theta-\lambda_0)n\tau) + \sum_{m=1}^{M-1}\frac{p_m}{p_0}\exp(i(\theta-\lambda_m)n\tau) + \frac{E_n}{p_0}\exp(i\theta n\tau)\right] \\
&= \frac{\exp(i(\theta-\lambda_0)N\tau)-1}{\exp(i(\theta-\lambda_0)\tau)-1} + \sum_{m=1}^{M-1}\frac{p_m}{p_0}\frac{\exp(i(\theta-\lambda_m)N\tau)-1}{\exp(i(\theta-\lambda_m)\tau)-1} + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(i\theta n\tau).
\end{aligned} \tag{B5}
$$

Define

$$
\overline{E}_\theta = \frac{1}{N}\sum_{n=0}^{N-1} E_n\exp(i\theta n\tau).
$$

Now, we are ready to prove Proposition 8:

Proof of Proposition 8. Define $R_m = |(\lambda_m - \theta^*)\tau \bmod (-\pi,\pi]|$ for $0 \leq m \leq M-1$. We separate the following proof into three steps. In the first step, we give a lower bound for "$|f(\lambda_0)|$". Then, we give a loose upper bound for R_0, using the fact that $|f(\theta^*)| \geq |f(\lambda_0)|$. Finally, we improve the bound to δ/T.

Step 1: Lower bound for "$|f(\lambda_0)|$"

Using Eq. (E4) in Appendix E 2, we have

$$
\begin{aligned}
\lim_{\theta\to\lambda_0}|f(\theta)| &\geq N - \sum_{k=1}^{M-1}\frac{p_k}{p_0}\left|\frac{\exp(i(\lambda_0-\lambda_k)T)-1}{\exp(i(\lambda_0-\lambda_k)\tau)-1}\right| \\
&\quad - \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}N \\
&\geq \left(1-(\alpha-1)\frac{1-p_0}{p_0} - \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\right)N.
\end{aligned} \tag{B6}
$$

Step 2: Loose upper bound for R_0

We claim that for α in Eq. (B1),

$$
R_0 N \leq \frac{\pi p_0}{((1+\alpha)p_0 - \alpha - (\left|\overline{E}_{\theta^*}\right| + \left|\overline{E}_{\lambda_0}\right|))}. \tag{B7}
$$

If the claim does not hold, note that

$$
\left|\exp(i(\lambda_0-\theta^*)\tau)-1\right| = |2\sin(R_0/2)| \geq \frac{2}{\pi}R_0.
$$

Combining this with Eq. (E1) in Appendix E 1,

$$
\begin{aligned}
|f(\theta^*)| &\le \frac{\pi}{R_0} + \sum_{m=1}^{M-1} \frac{p_m}{p_0}\frac{R_m N}{R_m} + \frac{|\overline{E}_{\theta^*}|}{p_0} N \\
&\le \left(\frac{\pi}{R_0 N} + \frac{1-p_0}{p_0} + \frac{|\overline{E}_{\theta^*}|}{p_0}\right) N \\
&< \left(1 - (\alpha-1)\frac{1-p_0}{p_0} - \frac{|\overline{E}_{\lambda_0}|}{p_0}\right) N \\
&\le \lim_{\theta\to\lambda_0} |f(\theta)|,
\end{aligned}
$$

where we use $R_0 N > \pi p_0 / \left((1+\alpha)p_0 - \alpha - \left(|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}|\right)\right)$ in the second last inequality. This contradicts to the fact that $|f(\theta^*)|$ is the maximum. Thus, we must have Eq. (B7).

Step 3: Improve upper bound to δ/T with probability $1-\eta$

Define $\beta = p_0/((1+\alpha)p_0 - \alpha - \xi)$. First, combining the second inequality of Eq. (B4) with Eq. (A4) of Lemma 5 [or Eq. (A1)], we have

$$
\mathbb{P}\left(\left|\overline{E}_{\theta^*}\right| + \left|\overline{E}_{\lambda_0}\right| > 10^{-3}\right) \le \eta/2.
$$

When $\left|\overline{E}_{\theta^*}\right| + \left|\overline{E}_{\lambda_0}\right| \le 10^{-3}$, according to Eq. (B7), we have $R_0 \le \pi p_0/\left(\left((1+\alpha)p_0 - \alpha - 10^{-3}\right)N\right)$. Also, plugging $\rho = \pi p_0/\left(\left((1+\alpha)p_0 - \alpha - 10^{-3}\right)\right)$ into Eq. (A5) of Lemma 6 [or Eq. (A2)] and using the first inequality of Eq. (B4), we have

$$
\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\rho}{T},\lambda_0+\frac{\rho}{T}]} \left|\overline{E}_\theta\right| \ge \xi/2\right) \le \eta/2,
$$

Combining the above two inequalities, we have

$$
\mathbb{P}\left(|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| \ge \xi\right) \le \eta.
$$

Then, to prove Eq. (34), it suffices to show that $R_0 \le \delta/N$ when $|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| \le \xi$.

When $|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| \le \xi$, using Eq. (B7), $p_0 > 0.71$, and the fact that ξ is small enough ($\xi < 10^{-3}$), we further have

$$
R_0 \le \frac{\pi p_0}{((1+\alpha)p_0 - \alpha - \xi)N} = \frac{\pi\beta}{N} < \frac{2\pi}{N}. \tag{B8}
$$

Since $|(\exp(i\theta N)-1)/(\exp(i\theta)-1)| = |\sin(N\theta/2)/\sin(\theta/2)|$, from Eq. (B5) we also have

$$
|f(\theta^*)| \le \left(\frac{\sin(NR_0/2)}{N\sin(R_0/2)} + \frac{1-p_0}{p_0} + \frac{\xi}{2p_0}\right) N \tag{B9}
$$

and

$$
|f(\theta^*)| \ge \left(1 - (\alpha-1)\frac{1-p_0}{p_0} - \frac{\xi}{2p_0}\right) N \ge \lim_{\theta\to\lambda_0} |f(\theta)|. \tag{B10}
$$

Combining Eqs. (B9) and (B10), we have

$$
\frac{\sin(NR_0/2)}{\sin(R_0/2)} \ge N\frac{(1+\alpha)p_0 - \alpha - \xi}{p_0} = \frac{N}{\beta}. \tag{B11}
$$

Note that

$$
\frac{\sin(\delta/2)}{\sin(\delta/(2N))} \le \frac{2\sin(\delta/2)N}{\cos(\delta/(2N))\delta} \le \frac{N}{\beta},
$$

where we use $\sin(\delta/(2N)) \ge \cos(\delta/(2N))\delta/(2N)$ in the first inequality and $\beta \le \delta\cos(\delta/10)/2\sin(\delta/2)$ in the second inequality. Hence

$$
\frac{\sin(NR_0/2)}{\sin(R_0/2)} \ge \frac{\sin(N\delta/(2N))}{\sin(\delta/(2N))}.
$$

Finally, because $\sin(Nx)/\sin(x)$ is monotonically decreasing in $(0,\pi/N]$, we have

$$
R_0 \le \frac{\delta}{N}. \tag{B12}
$$

This concludes the proof. ■

2. Refined estimate

According to Proposition 8, Eqs. (B3) and (B4), when $\delta \to 0$, we should set $\xi \sim O(\delta^{-2})$ and $NN_s \sim O(\delta^{-4})$ to ensure $T_{\max} = \delta/\epsilon$. Thus, we cannot directly prove Theorem 7 using Proposition 8. We need to reduce the scaling of N and N_s with respect to δ^{-1}.

The main idea is to use a different way to bound the expectation error E_n. To achieve a better bound for the error term, instead of bounding $\overline{E}_{\theta^*}$ and $\overline{E}_{\lambda_0}$ separately, now we can bound the difference of these two error terms using Eq. (A6). Intuitively, when θ^* and λ_0 are close to each other, it is likely that these two error terms will cancel each other when we compare the difference between $|f(\theta^*)|$ and $|f(\lambda_0)|$. This intuition is justified by Eq. (A6) of Lemma 6. Assume that we already know $|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]| < \delta/T$; then $NN_s \ge \widetilde{\mathcal{O}}\left(\delta^2\xi^{-2}\right)$ suffices to guarantee $\left|\overline{E}_{\theta,q} - \overline{E}_{\lambda_0,q}\right| \ge \xi$

with high probability. Formally, comparing this requirement with the first inequality of Eq. (B4), we can reduce the blow-up rate to $\mathcal{O}(\delta^{-2})$ when $\delta \to 0$, which matches the condition in Theorem 7. However, the above calculation assumes that $|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]| < \delta/T$, which is unknown to us in prior. To overcome this difficulty, we need to use an iteration argument to achieve the desired order. We first show the following lemma to start our iteration.

Lemma 9: *Given $0 < \delta \leq 4$ and $0 < \eta < 1/2$ and defining $T = N\tau$, assume that the condition of Proposition 8 is satisfied. If there exists $0 < \xi_1 < \xi$ such that*

$$\frac{p_0}{\left(\sqrt{2}\alpha + 1\right)p_0 - \sqrt{2}\alpha - \sqrt{2}(\xi_1 + \xi\delta/2)} < \frac{\delta\cos(\delta/10)}{2\sin(\delta/2)} \tag{B13}$$

and

$$NN_s = \Omega\left(\delta^2\xi_1^{-2}\mathrm{polylog}(\xi^{-1}\eta^{-1})\right), \tag{B14}$$

then there exists $\delta_{\mathrm{new}} \in (0, \delta)$ such that

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta_{\mathrm{new}}}{T}\right) \geq 1 - \eta,$$

where δ_{new} is the unique solution to the following equation:

$$\frac{p_0}{\left(\sqrt{2}\alpha + 1\right)p_0 - \sqrt{2}\alpha - \sqrt{2}(\xi_1 + \xi\delta_{\mathrm{new}}/2)} = \frac{N\sin(\delta_{\mathrm{new}}/(2N))}{\sin(\delta_{\mathrm{new}}/2)}. \tag{B15}$$

Proof of Lemma 9. Define $R_0 = |(\lambda_0 - \theta^*)\tau \bmod (-\pi, \pi]|$. Similarly to the proof of Proposition 8, we have

$$\mathbb{P}\left(R_0 < \frac{\delta}{N}\right) \geq 1 - \eta/4, \quad \mathbb{P}\left(|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| \geq \xi\right) \leq \eta/4.$$

Combining Eq. (A6) of Lemma 6 (setting $\rho = \delta$) with Eq. (B14) and $p_0 > 0.71$, we have

$$\mathbb{P}\left(\sup_{\theta \in [\lambda_0 - \frac{\delta}{T}, \lambda_0 + \frac{\delta}{T}]} \left|\overline{E}_\theta - \overline{E}_{\lambda_0}\right| \geq \xi_1\right) \leq \eta/2.$$

Thus, with probability $1 - \eta$, we have

$$R_0 < \frac{\delta}{N}, \quad |\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| < \xi, \quad \left|\overline{E}_{\theta^*} - \overline{E}_{\lambda_0}\right| < \xi_1. \tag{B16}$$

From Eq. (B16), it suffices to prove $|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]| < \delta_{\mathrm{new}}/T$.

Because θ^* is the maximal point, using Eq. (B5) and the result in Appendix E 2, we have

$$\begin{aligned}
&\left|\sum_{n=0}^{N-1}\exp(i(\lambda_0 - \theta^*)n\tau) + \sum_{m=1}^{M-1}\frac{p_m}{p_0}\frac{\exp(i(\lambda_m - \theta^*)N\tau) - 1}{\exp(i(\lambda_m - \theta^*)\tau) - 1} + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\theta^* n\tau)\right| \\
&\geq \left|N + \sum_{m=1}^{M-1}\frac{p_m}{p_0}\frac{\exp(i(\lambda_m - \lambda_0)N\tau) - 1}{\exp(i(\lambda_m - \lambda_0)\tau) - 1} + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\lambda_0\tau)\right| \\
&\geq \left|N + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\lambda_0\tau)\right| - (\alpha - 1)\frac{1 - p_0}{p_0}N.
\end{aligned} \tag{B17}$$

Also, using Eq. (B16), we have

$$\begin{aligned}
&\left|\sum_{n=0}^{N-1}\exp(i(\lambda_0 - \theta^*)n\tau) + \sum_{m=1}^{M-1}\frac{p_m}{p_0}\frac{\exp(i(\lambda_m - \theta^*)N\tau) - 1}{\exp(i(\lambda_m - \theta^*)\tau) - 1} + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\theta^* n\tau)\right| \\
&\leq \left|\sum_{n=0}^{N-1}\exp(i(\lambda_0 - \theta^*)n\tau) + \sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\lambda_0 n\tau)\right| + \frac{1 - p_0}{p_0}N + \frac{\xi_1}{p_0}N
\end{aligned}$$

$$
\begin{aligned}
&= \left|\frac{\sin(R_0N/2)}{\sin(R_0/2)}\exp\left(iR_0(N-1)/2\right)+\sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\lambda_0 n\tau)\right|+\frac{1-p_0}{p_0}N+\frac{\xi_1}{p_0}N\\
&\le \left|\frac{\sin(R_0N/2)}{\sin(R_0/2)}+\sum_{n=0}^{N-1}\frac{E_n}{p_0}\exp(-i\lambda_0 n\tau)\right|+\frac{1-p_0}{p_0}N+\frac{\xi_1+\xi\delta/2}{p_0}N\\
&= \left|\frac{\sin(R_0N/2)}{N\sin(R_0/2)}+\frac{\overline{E}_{\lambda_0}}{p_0}\right|N+\frac{1-p_0}{p_0}N+\frac{\xi_1+\xi\delta/2}{p_0}N,
\end{aligned}
\tag{B18}
$$

where we use $\left|\overline{E}_{\theta^*}-\overline{E}_{\lambda_0}\right|<p_0\xi_1$ in the second inequality and $|\exp\left(iR_0(N-1)/2\right)-1|\le\delta/2$ and $\left|\sum_{n=0}^{N-1}E_n \exp(-i\lambda_0 n\tau)\right|\le N\xi$ in the last inequality.

Let

$$
\sum_{n=0}^{N-1}E_n\exp(-i\lambda_0\tau)=N\left|\overline{E}_{\lambda_0}\right|\left(\cos(\theta_E)+i\sin(\theta_E)\right),
$$

where $\overline{E}_{\lambda_0}=1/N\sum_{n=0}^{N-1}E_n\exp(-i\lambda_0 n\tau)$. Then $\left|\overline{E}_{\lambda_0}\right|\le\xi<p_0/\pi<p_0/2N(\sin(R_0N/2))/(\sin(R_0/2))$ according to the second inequality of Eq. (B16) and the estimates in Eqs. (B17) and (B18) can be combined to obtain

$$
\begin{aligned}
&\sqrt{\left(\frac{\sin(R_0N/2)}{N\sin(R_0/2)}+\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2+\left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2}+\frac{\alpha(1-p_0)+(\xi_1+\xi\delta/2)}{p_0}\\
&\ge\sqrt{\left(1+\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2+\left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2}.
\end{aligned}
\tag{B19}
$$

Because $\left(\sqrt{(x+a)^2+b^2}\right)'\ge1/\sqrt{2}$ when $x,a,b>0$ and $x+a>b$, we obtain

$$
\begin{aligned}
&\sqrt{\left(1+\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2+\left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2}\\
&-\sqrt{\left(\frac{\sin(R_0N/2)}{N\sin(R_0/2)}+\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2+\left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2}\\
&\ge\left(1-\frac{\sin(R_0N/2)}{N\sin(R_0/2)}\right)\frac{1}{\sqrt{2}},
\end{aligned}
$$

where we set $a=\left|\overline{E}_{\lambda_0}\right|\cos(\theta_E)$ and $b=\left|\overline{E}_{\lambda_0}\right|\sin(\theta_E)$. Plugging this back into Eq. (B19), we obtain

$$
\frac{\sin(R_0N/2)}{N\sin(R_0/2)}\ge\frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta/2)}{p_0}.
\tag{B20}
$$

According to Eq. (B13), we first have

$$
\frac{\sin(\delta/2)}{N\sin(\delta/(2N))}<\frac{2\sin(\delta/2)}{\delta\cos(\delta/10)}<\frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta/2)}{p_0}.
$$

Thus, there exists $\delta_1<\delta$ such that

$$
\frac{\sin(\delta_1/2)}{N\sin(\delta_1/(2N))}=\frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta/2)}{p_0}.
$$

Using a similar argument as in the proof of Proposition 8, we have $|R_0| \leq \delta_1/N$. Then, similarly to the previous argument, we can improve Eq. (B20), meaning that R_0 should satisfy the following inequality:

$$\frac{\sin(R_0N/2)}{N\sin(R_0/2)} \geq \frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta_1/2)}{p_0}.$$

Because

$$\frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta_1/2)}{p_0} > \frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta/2)}{p_0} = \frac{\sin(\delta_1/2)}{N\sin(\delta_1/(2N))},$$

there exists $\delta_2 < \delta_1$ such that

$$\frac{\sin(\delta_2/2)}{N\sin(\delta_2/(2N))} = \frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta_1/2)}{p_0}$$

and $|R_0| \leq \delta_2/N$. Doing this recurrently, we finally have $|R_0| \leq \delta_{\text{new}}/N$, where

$$\frac{\sin(\delta_{\text{new}}/2)}{N\sin(\delta_{\text{new}}/(2N))} = \frac{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi\delta_{\text{new}}/2)}{p_0}.$$

Finally, because $h_1(x) = \sin(x/2)/N\sin(x/(2N))$ is a concave function ($x \in [0,\pi]$) with $h_1(0)=1, h_1'(0)=0$ and $h_2(x) = (\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_1+\xi x/2))/p_0$ is a linearly decreasing function with $0 < h_2(0) < 1$, δ_{new} is the unique solution satisfying Eq. (B15). ∎

Equation (B14) provides a direction for reducing the scaling of N and N_s with respect to δ. If we ignore Eq. (B3), to guarantee a short depth $\delta_{\text{new}}/\epsilon$, we can choose $\delta \approx \delta_{\text{new}}$, $\xi \sim \delta_{\text{new}}$, and $\xi_1 \sim \delta_{\text{new}}^2$; then, it suffices to choose $NN_s \sim \Theta(\delta_{\text{new}}^{-2})$. This gives us the desired order. However, the previous argument cannot be directly applied because we also need $\xi = \mathcal{O}(\delta^2)$ and $NN_s = \Omega(\xi^{-2})$ according to Eqs. (B3) and (B4). These two requirements would bring back the original δ_{new}^{-4} dependence on NN_s.

Even the previous argument cannot be applied directly; we can still use Lemma 9 to improve the scaling with respect to δ in the following way. According to Lemma 9, for fixed small δ_{new}, according to Eq. (B13), to make the depth smaller than $\delta_{\text{new}}/\epsilon$, we should choose

$$\xi = \mathcal{O}\left(\min\{\delta^2, \delta_{\text{new}}\}\right), \quad \xi_1 = \mathcal{O}\left(\delta_{\text{new}}^2\right),$$

for all $1 \leq i \leq M$. Then, according to the first inequality of Eqs. (B4) and (B14), we set

$$NN_s = \widetilde{\Theta}\left(\max_{0\leq i\leq M-1}\left\{\xi^{-2}, \delta^2\xi_1^{-2}\right\}\right).$$

Minimizing this in δ, ξ, and ξ_1 with fixed δ_1, a proper choice of these parameters should be

$$\xi = \Theta\left(\delta^2\right), \quad \xi_1 = \Theta\left(\delta_{\text{new}}^2\right), \quad \delta = \Theta\left(\delta_{\text{new}}^{\frac{2}{3}}\right).$$

This choice of parameters would reduce the blow-up rate of NN_s to $\widetilde{\Theta}\left(\delta_{\text{new}}^{-8/3}\right)$ when $\delta_{\text{new}} \to 0$. Using a similar argument as before, we can apply Lemma 9 repeatedly. In each iteration, we can slightly reduce the scaling of NN_s and the final scaling can be $\mathcal{O}(\delta^{-2+o(1)})$. This iteration process can be carried out using the following lemma.

Lemma 10: *Given $0 < \delta \leq 4$, $0 < \eta < 1/2$, define $T = N\tau$. Given an integer $M > 1$, a decreasing sequence $\{\xi_i\}_{i=0}^M$ with small enough ξ_0, and a decreasing sequence $\{\delta_i\}_{i=0}^M$, assume that the condition of Proposition 8 is satisfied with $\xi = \xi_0$ and $\delta = \delta_0$. If*

$$NN_s = \Omega\left(\delta_i^2\xi_{i+1}^{-2}\text{polylog}(M\xi_{i+1}^{-1}\eta^{-1})\right) \quad \text{(B21)}$$

and

$$\frac{p_0}{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_{i+1}+\xi_0\delta_{i+1}/2)} \leq \frac{\delta_{i+1}\cos(\delta_{i+1}/10)}{2\sin(\delta_{i+1}/2)}. \quad \text{(B22)}$$

for all $0 \le i \le M-1$. *Then,*

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0)\ \bmod\ [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_M}{T}\right) \ge 1-\eta. \tag{B23}$$

Proof of Lemma 10. First, according to Lemma 9 and Proposition 8, the conditions in Eqs. (B21) and (B22) ensure that

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_1}{T}\right) \ge 1-\eta/2,$$
$$\mathbb{P}\left(|\overline{E}_{\theta^*}|+|\overline{E}_{\lambda_0}| \ge \xi_0\right) \le \eta/8.$$

Combining the second inequality of Eq. (B21) with Eq. (A6) of Lemma 6, we also have

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\delta_1}{T},\lambda_0+\frac{\delta_1}{T}]}\left|\overline{E}_\theta-\overline{E}_{\lambda_0}\right| \ge \xi_2\right) \le \frac{\eta}{4(M-1)}.$$

Thus, with probability $1-(3\eta/4+\eta/(4(M-1)))$, we have

$$R_0 < \frac{\delta_1}{N}, \quad |\overline{E}_{\theta^*}|+|\overline{E}_{\lambda_0}| < \min\{\xi_0,10^{-3}\},$$
$$\left|\overline{E}_{\theta^*}-\overline{E}_{\lambda_0}\right| < \xi_2.$$

Similarly to the proof of Lemma 9, we have

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_{\text{new}}}{T}\right)$$
$$\ge 1-\left(\frac{3\eta}{4}+\frac{\eta}{4(M-1)}\right),$$

where δ_{new} is the unique solution to the following equation:

$$\frac{p_0}{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_2+\xi_0\delta_{\text{new}}/2)} = \frac{N\sin(\delta_{\text{new}}/(2N))}{\sin(\delta_{\text{new}}/2)}.$$

Because

$$\frac{p_0}{\left(\sqrt{2}\alpha+1\right)p_0-\sqrt{2}\alpha-\sqrt{2}(\xi_2+\xi_0\delta_2/2)} = \frac{\delta_2\cos(\delta_2/10)}{2\sin(\delta_2/2)} < \frac{N\sin(\delta_2/(2N))}{\sin(\delta_2/2)},$$

we must have $\delta_2 > \delta_{\text{new}}$ and

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_2}{T}\right)$$
$$\ge 1-\left(\frac{3\eta}{4}+\frac{\eta}{4(M-1)}\right).$$

Combining this with the second inequality of Eq. (B21) ($i=2$), with probability $1-(3\eta/4+\eta/(2(M-1)))$, we have

$$R_0 < \frac{\delta_2}{N}, \quad |\overline{E}_{\theta^*}|+|\overline{E}_{\lambda_0}| < \xi_0, \quad \left|\overline{E}_{\theta^*}-\overline{E}_{\lambda_0}\right| < \xi_3,$$

which implies

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_3}{T}\right)$$
$$\ge 1-\left(\frac{3\eta}{4}+\frac{\eta}{2(M-1)}\right).$$

Doing this repeatedly, we finally obtain Eq. (B23). ■

Now, we are ready to prove Theorem 7.

Proof of Theorem 7. For fixed decreasing sequence $\{\delta_i\}_{i=0}^M$ with small enough δ_0, according to Eqs. (B3) and (B22), to make the depth smaller than δ_M/ϵ, we need

$$(1+\sqrt{2}\alpha)(1-p_0) = \mathcal{O}\left(\min_i\left\{\delta_i^2\right\}\right),$$
$$\xi_0 = \mathcal{O}\left(\min\{\delta_0^2,\delta_M\}\right), \quad \xi_i = \mathcal{O}\left(\delta_i^2\right), \tag{B24}$$

for all $1 \le i \le M$. The first equation implies that the smallest δ_M that we can choose is

$$\delta_M = \Theta(\sqrt{1-p_0}).$$

According to Eqs. (B4) and (B21), the last two equations of Eq. (B24) imply that

$$NN_s = \widetilde{\Omega}\left(\max_{0\le i\le M-1}\left\{p_0^{-2}\xi^{-2}, p_0^{-2}\delta_i^2\xi_{i+1}^{-2}\right\}\right). \tag{B25}$$

Minimizing Eq. (B25) in $\{\delta_i\}_{i=0}^{M-1}, \{\xi_i\}_{i=0}^M$ with fixed δ_M, a proper choice of these parameters should be

$$\xi_i = \Theta\left(\delta_i^2\right), \quad \delta_i = \Theta\left(\delta_M^{\frac{2-(1/2)^i}{2-(1/2)^M}}\right),$$

for $0 \le i \le M$. Thus, to ensure that

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_M}{T}\right) \ge 1-\eta,$$

we can choose $NN_s = \Omega\left(\delta_M^{-\frac{4}{2-(1/2)^M}}\text{polylog}(M\delta_M^{-1}\eta^{-1}p_0^{-1}\Delta^{-1})\right)$. Set $\zeta = \frac{1}{2^M}\frac{1}{2-(1/2)^M}$; hence we conclude the proof. ■

APPENDIX C: PROOF OF THEOREM 2

In this appendix, we prove a rigorous version of Theorem 2 as follows.

Theorem 11 (complexity of multilevel QCELS): *Let θ^* be the output of Algorithm 1. Given $p_0 > 0.71$, $0 < \eta < 1/2$, $0 < \epsilon < 1/2$, and small $\zeta > 0$, we can choose δ according to Eq. (32). Let*

$$J = \lceil \log_2(1/\epsilon) \rceil + 1,$$
$$\tau_j = 2^{j-1-\lceil \log_2(1/\epsilon)\rceil} \frac{\delta}{N\epsilon}, \quad \forall 1 \le j \le J$$

and

$$NN_s = \Theta\left(\delta^{-(2+\zeta)} \mathrm{polylog}\left(\log(\zeta^{-1})\log(\epsilon^{-1})\eta^{-1}\right)\right),$$
$$\min\{N, N_s\} = \Omega(\mathrm{polylog}\left(\log(\epsilon^{-1})\eta^{-1}\right)).$$

Denote θ^* as the output of Algorithm 1; then

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0)\right| < \epsilon\right) \ge 1 - \eta.$$

In particular, we have

$$T_{\max} = N\tau_J = \frac{\delta}{\epsilon},$$
$$T_{\text{total}} = \sum_{j=1}^{J} N(N-1)N_s\tau_j/2$$
$$= \Theta\left(\frac{\mathrm{polylog}\left(\log(\zeta^{-1})\log(\epsilon^{-1})\eta^{-1}\right)}{\delta^{1+\zeta}\epsilon}\right).$$

Proof of Theorem 11. To prove Theorem 2, it suffices to prove that for each fixed j,

$$\mathbb{P}\left(\lambda_0 \in \left[\theta_j^* - \frac{\delta}{N\tau_j}, \theta_j^* + \frac{\delta}{N\tau_j}\right]\right) \le 1 - \frac{j\eta}{J}. \tag{C1}$$

We prove this by induction. First, let $j = 1$. According to Theorem 7,

$$\mathbb{P}\left(\left|(\theta_1^* - \lambda_0) \bmod [-\pi/\tau_1, \pi/\tau_1]\right| < \frac{\delta}{T}\right) \ge 1 - \frac{\eta}{J}.$$

Because $\tau_1 = 2^{-\lceil \log_2(1/\epsilon)\rceil}\delta/N\epsilon \le \delta/N < 1/4$, we have $\pi/\tau_1 > \pi$. Then,

$$\mathbb{P}\left(\left|\theta_1^* - \lambda_0\right| < \frac{\delta}{T}\right) \ge 1 - \frac{\eta}{J},$$

Assume that Eq. (C1) is true for $j = K-1$, meaning that

$$\mathbb{P}\left(\lambda_0 \in \left[\theta_{K-1}^* - \frac{\delta}{N\tau_{K-1}}, \theta_{K-1}^* + \frac{\delta}{N\tau_{K-1}}\right]\right) \ge 1 - \frac{(K-1)\eta}{J}.$$

Using Theorem 7 again, we have

$$\mathbb{P}\left(\lambda_0 \in \left[\theta_{K-1}^* - \frac{\delta}{N\tau_{K-1}}, \theta_{K-1}^* + \frac{\delta}{N\tau_{K-1}}\right] \bigcap C_k\right) \ge 1 - \frac{K\eta}{J}, \tag{C2}$$

where

$$C_k = \bigcup_{c\in\mathbb{Z}} \left[\theta_K^* - \frac{\delta}{N\tau_K} + \frac{2c\pi}{\tau_K}, \theta_K^* + \frac{\delta}{N\tau_K} + \frac{2c\pi}{\tau_K}\right].$$

Noting that $\tau_K = 2\tau_{K-1}$ and $\delta/N < \pi/4$, we have

$$\frac{2\pi}{\tau_K} - \frac{4\delta}{N\tau_K} > \frac{\pi}{\tau_K}.$$

Since $\theta_K^* \in [\theta_{K-1}^* - \pi/\tau_K, \theta_{K-1}^* + \pi/\tau_K]$, we obtain

$$\left[\theta_{K-1}^* - \frac{\delta}{N\tau_{K-1}}, \theta_{K-1}^* + \frac{\delta}{N\tau_{K-1}}\right] \bigcap C_k = \left[\theta_{K-1}^* - \frac{\delta}{N\tau_{K-1}}, \theta_{K-1}^* + \frac{\delta}{N\tau_{K-1}}\right] \bigcap \left[\theta_K^* - \frac{\delta}{N\tau_K}, \theta_K^* + \frac{\delta}{N\tau_K}\right].$$

Combining this with Eq. (C2), we obtain

$$\mathbb{P}\left(\lambda_0 \in \left[\theta_K^* - \frac{\delta}{N\tau_K}, \theta_K^* + \frac{\delta}{N\tau_K}\right]\right) \ge 1 - \frac{K\eta}{J},$$

which concludes the proof. ∎

APPENDIX D: PROOF OF THEOREM 3

In this appendix, we prove the following rigorous version of Theorem 3.

Theorem 12 (complexity of Algorithm 2): *Given small $\zeta > 0$ and the failure probability $0 < \eta < 1$, assume that $\delta = \Theta\left(\sqrt{1 - p_r(I, I')}\right)$ and $p_r(I, I')$ is close enough to 1. Let $d = \Theta(D^{-1}\mathrm{polylog}(p_0^{-1}\delta^{-1}))$, $q = \Theta(p_0\delta^2)$,*

$$J = \lceil \log_2(1/\epsilon)\rceil + 1, \quad \tau_j = 2^{j-1-\lceil \log_2(1/\epsilon)\rceil}\frac{\delta}{N\epsilon},$$
$$\forall 1 \le j \le J,$$

and

$$NN_s = \Omega\left(p_0^{-2}\delta^{-(2+\zeta)}\mathrm{polylog} \times \left(\log(\zeta^{-1})\log(\epsilon^{-1})D^{-1}\eta^{-1}p_0^{-1}\right)\right),$$
$$\min\{N, N_s\} = \Omega(p_0^{-2}\mathrm{polylog}\left(\log(\epsilon^{-1})D^{-1}\eta^{-1}\right)).$$

Then,

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta}{T}\right) \geq 1 - \eta,$$

where θ^* is the output of Algorithm 2. In particular, to construct the loss function,

$$T_{\max} = d + \delta/\epsilon, \quad T_{\text{total}} = \Theta\left(p_0^{-2}\delta^{-(2+\zeta)}\text{polylog}\right. \\ \left.\times\ (\log(\zeta^{-1})\log(\epsilon^{-1})D^{-1}\eta^{-1}p_0^{-1})\ (d + \delta/\epsilon)\right).$$

Define $\mathcal{F} = \sum_{l=-d}^{d}\left|\hat{F}_{l,q}\right|$. Since F_q in Algorithm 2 is chosen according to Ref. [15, Lemma 6], we have $\mathcal{F} = \Theta(\log(D^{-1}\text{polylog}(q^{-1})))$. Then, similarly to the proof of Theorem 11 (in Appendix C), to prove Theorem 12, it suffices to show the following theorem, which gives us the complexity of one step of the iteration with a general choice of I, I', and F_q.

Theorem 13: *Given small $\zeta > 0$ and the failure probability $0 < \eta < 1$, assume that $\delta = \Theta\left(\sqrt{1 - p_r(I, I')}\right)$ and that $p_r(I, I')$ is close enough to 1. Set $q = \Theta(p_0\delta^2)$,*

$$NN_s = \Omega\left(p_0^{-2}\delta^{-(2+\zeta)}\mathcal{F}^2\text{polylog}(\log(\zeta^{-1})\eta^{-1}p_0^{-1}\mathcal{F})\right), \\ \min\{N, N_s\} = \Omega(p_0^{-2}\mathcal{F}^2\text{polylog}\left(\eta^{-1}\mathcal{F}\right)).$$

Then,

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta}{T}\right) \geq 1 - \eta, \quad \text{(D1)}$$

where θ^* is defined in Eq. (39) (with $\tau_j = T$).

Define

$$\overline{E}_\theta = \frac{1}{N}\sum_{n=0}^{N-1} E_{n,q}\exp(i\theta n\tau),$$

where $E_{n,q}$ is defined in Eq. (40). Similarly to the large-p_0 setting, we first give a bound for the expectation error $E_{n,q}$ in the following lemma.

Lemma 14: *Assume that $q < 1$. Given $0 < \eta < 1/2$ and $0 < \rho, \xi < 10\pi$, then:*

(a) When $\min\{N, N_s\} = \Omega(p_0^{-2}\mathcal{F}^2\log(\eta^{-1}\mathcal{F}))$,

$$\mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}|E_{n,q}| > 0.001p_0\right) \leq \eta. \quad \text{(D2)}$$

(b) When $NN_s = \Omega\left(p_0^{-2}\xi^{-2}\mathcal{F}^2\text{polylog}(\xi^{-1}\eta^{-1}p_0^{-1})\mathcal{F}\right)$,

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0 - \frac{\rho}{T}, \lambda_0 + \frac{\rho}{T}]}\left|\overline{E}_{\theta,q}\right| \geq p_0\xi\right) \leq \eta. \quad \text{(D3)}$$

(c) When $NN_s = \Omega\left(\rho^2p_0^{-2}\xi^{-2}\mathcal{F}^2\text{polylog}(\xi^{-1}\eta^{-1}p_0^{-1}\mathcal{F})\right)$,

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0 - \frac{\rho}{T}, \lambda_0 + \frac{\rho}{T}]}\left|\overline{E}_{\theta,q} - \overline{E}_{\lambda_0,q}\right| \geq p_0\xi\right) \leq \eta. \quad \text{(D4)}$$

Proof of Lemma 14. The proof is very similar to that of Lemmas 5 and 6 after noting that $E_{n,q} = 1/N_s\sum_{k=1}^{N_s}Z_{k,n,q} - \mathbb{E}(Z_{k,n,q})$ and $|Z_{k,n,q}| \leq \mathcal{F}$. ■

Before proving Theorem 13, we first prove a result that is similar to Proposition 8.

Lemma 15: *Define $T = N\tau$ and assume that $p_r(I, I')$ is close enough to 1. Set $q = \Theta(p_0\delta^2)$. Given $\Theta\left(\sqrt{1 - p_r(I, I')}\right) \leq \delta \leq 4$, $0 < \eta < 1/2$. If there exists a small enough number $\xi > 0$ such that*

$$\frac{1}{1 - \xi - \delta^2/200} \leq \frac{\delta\cos(\delta/10)}{2\sin(\delta/2)} \quad \text{(D5)}$$

and

$$NN_s = \Omega\left(p_0^{-2}\xi^{-2}\mathcal{F}^2\text{polylog}\left(\xi^{-1}\eta^{-1}p_0^{-1}\mathcal{F}\right)\right), \\ \min\{N, N_s\} = \Omega\left(p_0^{-2}\mathcal{F}^2\text{polylog}\left(\eta^{-1}\mathcal{F}\right)\right), \quad \text{(D6)}$$

then

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta}{T}\right) \geq 1 - \eta, \quad \text{(D7)}$$

where θ^* is defined in Eq. (39).

Proof of Lemma 15. $R_0 = |(\lambda_0 - \theta^*)\tau \bmod (-\pi, \pi]|$. Recall that

$$G_{n,q} = Z_{n,q} - p_0\exp(i\lambda_0 n\tau) \\ = E_{n,q} + p_0(F_q(\lambda_0) - 1)\exp(-i\lambda_0 n\tau) \\ + \sum_{k=1}^{M-1} p_k F_q(\lambda_k)\exp(-i\lambda_k n\tau)$$

and

$$|G_{n,q}| \leq |E_{n,q}| + q + \frac{\left(1 - p_r(I, I')\right)p_0}{p_r(I, I')}. \quad \text{(D8)}$$

Note that

$$\min_{r\in\mathbb{C},\theta\in\mathbb{R}} \frac{1}{N}\sum_{n=0}^{N-1}\left|Z_{n,q}-r\exp(-i\theta n\tau)\right|^2$$
$$=\min_{r\in\mathbb{C},\theta\in\mathbb{R}} \frac{1}{N}\sum_{n=0}^{N-1}\left|p_0\exp(-i\lambda_0 n\tau)+G_{n,q}-r\exp(-i\theta n\tau)\right|^2$$
$$=\min_{r\in\mathbb{C},\theta\in\mathbb{R}} \frac{1}{N}\sum_{n=0}^{N-1}\left|\exp(-i\lambda_0 n\tau)+\frac{G_{n,q}}{p_0}-\frac{r}{p_0}\exp(-i\theta n\tau)\right|^2.$$

Similarly to the proof of Proposition 8, this minimization problem is equivalent to maximizing the magnitude of the following function in θ:

$$f(\theta)=\frac{\exp(i(\theta-\lambda_0)N\tau)-1}{\exp(i(\theta-\lambda_0)\tau)-1}+\sum_{n=1}^{N-1}\frac{G_{n,q}}{p_0}\exp(i\theta n\tau). \tag{D9}$$

First, using Eqs. (D2) and (D6),

$$\mathbb{P}\left(\frac{|E_{n,q}|}{p_0}>5*10^{-4}\right)\le \eta/2.$$

Combining this with Eq. (D8), $q=\Theta(p_0\delta^2)$, and the assumption that $1-p_r(I,I')$ is small enough, we can have

$$\mathbb{P}\left(\frac{|G_{n,q}|}{p_0}>10^{-3}\right)\le \eta/2.$$

When $|G_{n,q}|/p_0\le 10^{-3}$, similarly to the proof of Proposition 8, we have $R_0\le \pi/0.9N$. This implies that, with probability $1-\eta/2$,

$$\frac{|G_{n,q}|}{p_0}\le 10^{-3},\quad R_0\le\frac{\pi}{0.9N}.$$

The second inequality gives us a loose bound for R_0. Combining these two inequalities with Eq. (D3) ($\rho=\pi/0.9$), the first inequality of Eq. (D6), and $\Theta\left(\sqrt{1-p_r(I,I')}\right)\le\delta\le 4$, we further have

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N-1}\frac{G_{n,q}}{p_0}\exp(i\theta^* n\tau)\right|+\left|\frac{1}{N}\sum_{n=1}^{N-1}\frac{G_{n,q}}{p_0}\exp(i\lambda_0 n\tau)\right|\ge\delta^2/200+\xi\right)\le\eta.$$

Finally, similarly to the proof of Proposition 8, because

$$\frac{\sin(\delta/2)}{\sin(\delta/(2N))}\le\frac{2\sin(\delta/2)N}{\delta\cos(\delta/10)}\le N(1-\xi-\delta^2/200),$$

we have $\mathbb{P}(R_0\le\delta/N)\ge 1-\eta$, which implies Eq. (D7). ■

Similarly to the discussion in Appendix B, we cannot directly prove Theorem 13 using Lemma 15. We need to use a different way to bound the expectation error $E_{n,q}$ to reduce the scaling of N, N_s with respect to δ. First, we show the following lemma, which is similar to Lemma 9.

Lemma 16: *Define $T=N\tau$ and assume that $p_r(I,I')$ is close enough to 1. Given $\Theta\left(\sqrt{1-p_r(I,I')}\right)\le\delta\le 4$, $0<\eta<1/2$, and small enough $\xi>0$, assume that Lemma 15, Eqs. (D5) and (D6) hold and there exists $0<\xi_1<\xi$ such that*

$$\frac{1}{1-\sqrt{2}(\xi_1+\xi\delta/2+\delta^2/200)}<\frac{\delta\cos(\delta/10)}{2\sin(\delta/2)}. \tag{D10}$$

Let $\Theta\left(\sqrt{1-p_r(I,I')}\right)\le\delta_{\text{new}}<\delta$ satisfy

$$\frac{1}{1-\sqrt{2}(\xi_1+\xi\delta_{\text{new}}/2+\delta_{\text{new}}^2/200)}\le\frac{\delta_{\text{new}}\cos(\delta_{\text{new}}/10)}{2\sin(\delta_{\text{new}}/2)}.$$

If $q=\Theta(p_0\delta_{\text{new}}^2)$ and

$$NN_s=\Omega\left(p_0^{-2}\delta^2\xi_1^{-2}\mathcal{F}^2\text{polylog}\left(\xi_1^{-1}\eta^{-1}p_0^{-1}\mathcal{F}\right)\right), \tag{D11}$$

then

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0)\bmod[-\pi/\tau,\pi/\tau]\right|<\frac{\delta_{\text{new}}}{T}\right)\ge 1-\eta.$$

Proof of Lemma 16. Define $R_0=|(\lambda_0-\theta^*)\tau \bmod (-\pi,\pi]|$. According to Lemmas 14 and 15, we have $\mathbb{P}(R_0<\delta/N)\ge 1-\eta/4$ and

$$\mathbb{P}\left(\left|\overline{E}_{\theta^*}\right|+\left|\overline{E}_{\lambda_0}\right|\ge p_0\xi\right)\le\eta/4.$$

Combining Eq. (D4) of Lemma 14 with Eq. (D11), we have

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\delta}{T},\lambda_0+\frac{\delta}{T}]}\left|\overline{E}_\theta-\overline{E}_{\lambda_0}\right|\ge p_0\xi_1\right)\le\eta/2.$$

Thus, with probability $1-\eta$, we have

$$R_0<\frac{\delta}{N},\quad \left|\overline{E}_{\theta^*}-\overline{E}_{\lambda_0}\right|<p_0\xi_1 \tag{D12}$$

and

$$\left|\overline{E}_{\theta}\right| + \left|\overline{E}_{\lambda_0}\right| \leq p_0\xi. \tag{D13}$$

It suffices to prove that $|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]| < \delta^\star/T$, assuming Eqs. (D12) and (D13).

Because θ^* is the maximal point of Eq. (D9), we have

$$\left|\sum_{m=0}^{M-1} \frac{p_m F_q(\lambda_m)}{p_0} \frac{\exp(i(\theta^* - \lambda_m)N\tau) - 1}{\exp(i(\theta^* - \lambda_m)\tau) - 1} + \frac{N\overline{E}_{\theta^*}}{p_0}\right| \geq \left|N + \sum_{n=0}^{N-1} \frac{E_n}{p_0}\exp(i\lambda_0 n\tau)\right| - \frac{\delta^2 N}{400},$$

where we use $q \leq p_0\delta^2/1600$ and $(1 - p_r(I, I'))/p_r(I, I') \leq \delta^2/800$ in the last inequality.

Also, using Eq. (D12), we have

$$\begin{aligned}
&\left|\sum_{m=0}^{M-1} \frac{p_m F_q(\lambda_m)}{p_0} \frac{\exp(i(\theta^* - \lambda_m)N\tau) - 1}{\exp(i(\theta^* - \lambda_m)\tau) - 1} + \frac{N\overline{E}_{\theta^*}}{p_0}\right| \leq \left|\sum_{n=0}^{N-1} \exp(i(\theta^* - \lambda_0)n\tau) + \sum_{n=0}^{N-1} \frac{E_n}{p_0}\exp(i\lambda_0 n\tau)\right| + \frac{\delta^2 N}{400} + \xi_1 N \\
&= \left|\frac{\sin(R_0N/2)}{\sin(R_0/2)} \exp\left(iR_0(N-1)/2\right) + \sum_{n=0}^{N-1} \frac{E_n}{p_0}\exp(i\lambda_0 n\tau)\right| + \frac{\delta^2 N}{400} + \xi_1 N \\
&\leq \left|\frac{\sin(R_0N/2)}{\sin(R_0/2)} + \sum_{n=0}^{N-1} \frac{E_n}{p_0}\exp(i\lambda_0 n\tau)\right| + (\xi_1 + \xi\delta/2 + \delta^2/400)N,
\end{aligned}$$

where we use $\left|\overline{E}_{\theta^*} - \overline{E}_{\lambda_0}\right| < p_0\xi_1$ in the first inequality and $|\exp(iR_0(N-1)/2) - 1| \leq \delta/2$ and $\left|\sum_{n=0}^{N-1} \frac{E_n}{p_0} \exp(i\lambda_0 n\tau)\right| \leq N\xi$ in the second inequality.

Assume that

$$\sum_{n=0}^{N-1} E_n \exp(i\lambda_0\tau) = N\left|\overline{E}_{\lambda_0}\right| \left(\cos(\theta_E) + i\sin(\theta_E)\right),$$

where $\overline{E}_{\lambda_0} = 1/N\sum_{n=0}^{N-1} E_n \exp(i\lambda_0 n\tau)$.

Note that $\left|\overline{E}_{\lambda_0}\right|/p_0 \leq \xi < 1/\pi < 1/2N(\sin(R_0N/2))/(\sin(R_0/2))$ according to the second inequality of Eq. (D12). Then,

$$\begin{aligned}
&\sqrt{\left(\frac{\sin(R_0N/2)}{N\sin(R_0/2)} + \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2 + \left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2} \\
&\quad + \xi_1 + \xi\delta/2 + \delta^2/200 \\
&\geq \sqrt{\left(1 + \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2 + \left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2}.
\end{aligned} \tag{D14}$$

Because $\left(\sqrt{(x+a)^2 + b^2}\right)' \geq 1/\sqrt{2}$ when $x, a, b > 0$ and $x + a > b$, we obtain

$$\begin{aligned}
&\sqrt{\left(1 + \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2 + \left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2} \\
&- \sqrt{\left(\frac{\sin(R_0N/2)}{N\sin(R_0/2)} + \frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\cos(\theta_E)\right)^2 + \left(\frac{\left|\overline{E}_{\lambda_0}\right|}{p_0}\sin(\theta_E)\right)^2} \\
&\geq \left(1 - \frac{\sin(R_0N/2)}{N\sin(R_0/2)}\right)\frac{1}{\sqrt{2}},
\end{aligned}$$

where we see that $a = \left|\overline{E}_{\lambda_0}\right|\cos(\theta_E)$ and $b = \left|\overline{E}_{\lambda_0}\right|\sin(\theta_E)$. Plugging this back into Eq. (D14), we obtain

$$\frac{\sin(R_0N/2)}{N\sin(R_0/2)} \geq 1 - \sqrt{2}(\xi_1 + \xi\delta/2 + \delta^2/200). \tag{D15}$$

According to Eq. (D10), we first have

$$\frac{\sin(\delta/2)}{N\sin(\delta/(2N))} < \frac{2\sin(\delta/2)}{\delta\cos(\delta/10)} < 1 - \sqrt{2}(\xi_1 + \xi\delta/2 + \delta^2/200).$$

Thus, there exists $\delta_1 < \delta$ such that

$$\frac{\sin(\delta_1/2)}{N\sin(\delta_1/(2N))} = 1 - \sqrt{2}(\xi_1 + \xi\delta/2 + \delta^2/200).$$

Using a similar argument as in the proof of Proposition 8, we have $|R_0| \leq \delta_1/N$. Then, similarly to the previous argument, we can improve Eq. (D15), meaning that R_0 should satisfy the following inequality:

$$\frac{\sin(R_0 N/2)}{N\sin(R_0/2)} \geq 1-\sqrt{2}(\xi_1+\xi\delta_1/2+\delta_1^2/200).$$

Because

$$1-\sqrt{2}(\xi_1+\xi\delta_1/2+\delta_1^2/200) > 1 - \sqrt{2}(\xi_1+\xi\delta/2+\delta^2/200) = \frac{\sin(\delta_1/2)}{N\sin(\delta_1/(2N))},$$

there exist $\delta_2 < \delta_1$ such that

$$\frac{\sin(\delta_2/2)}{N\sin(\delta_2/(2N))} = 1-\sqrt{2}(\xi_1+\xi\delta_1/2+\delta_1^2/200)$$

and $|R_0| \leq \delta_2/N$. Because $h_1(x) = \sin(x/2)/N\sin(x/(2N)) + \sqrt{2}x^2/200$ is a concave function $(x\in[0,\pi])$ with $h_1(0)=1, h_1'(0)=0$ and $h_2(x) = 1-\sqrt{2}(\xi_1+\xi x/2)$ is a linearly decreasing function that satisfies $0 < h_2(0) < 1$, if δ_{new} satisfies

$$\Theta\left(\sqrt{1-p_r(I,I')}\right) \leq \delta_{\text{new}} < \delta$$

and

$$\frac{\sin(\delta_{\text{new}}/2)}{N\sin(\delta_{\text{new}}/(2N))} < \frac{2\sin(\delta_{\text{new}}/2)}{\delta_{\text{new}}\cos(\delta_{\text{new}}/10)} < 1 - \sqrt{2}(\xi_1+\xi\delta_{\text{new}}/2+\delta_{\text{new}}^2/200), \tag{D16}$$

we must have $|R_0| \leq \delta_{\text{new}}/N$. ■

Similarly to Appendix B, the next step is to do the iteration using the following lemma.

Lemma 17: *Define $T = N\tau$ and assume that $p_r(I,I')$ is close enough to 1. Given $0<\eta<1/2$, an integer $M>1$, a decreasing sequence $\{\xi_i\}_{i=0}^M$ with small enough ξ_0, and a decreasing sequence $\{\delta_i\}_{i=0}^M$ with $\delta_0 \leq 4$ and $\delta_M \geq \Theta\left(\sqrt{1-p_r(I,I')}\right)$, assume that ξ_0, δ_0, N, N_s satisfy the conditions in Lemma 15. If $q = \Theta(p_0\delta_M^2)$,*

$$NN_s = \Omega\left(p_0^{-2}\delta_i^2\xi_{i+1}^{-2}\mathcal{F}^2\text{polylog}\left(M\xi_{i+1}^{-1}\eta^{-1}p_0^{-1}\mathcal{F}\right)\right) \tag{D17}$$

and

$$\frac{1}{1-\sqrt{2}(\xi_{i+1}+\xi_0\delta_{i+1}/2+\delta_{i+1}^2/200)} \leq \frac{\delta_{i+1}\cos(\delta_{i+1}/10)}{2\sin(\delta_{i+1}/2)}, \tag{D18}$$

for all $0\leq i\leq M-1$. Then,

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_M}{T}\right) \geq 1-\eta. \tag{D19}$$

Proof of Lemma 17. First, according to Lemmas 15 and 16, we have

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_1}{T}\right) \geq 1-\eta/2,$$
$$\mathbb{P}\left(|\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| \geq p_0\xi_0\right) \leq \eta/4.$$

Combining Eq. (D17) ($i=1$) with Eq. (D4) of Lemma 14, we obtain

$$\mathbb{P}\left(\sup_{\theta\in[\lambda_0-\frac{\delta_1}{T},\lambda_0+\frac{\delta_1}{T}]}\left|\overline{E}_\theta-\overline{E}_{\lambda_0}\right| \geq p_0\xi_2\right) \leq \frac{\eta}{4(M-1)}.$$

Thus, with probability $1-(3\eta/4+\eta/(4(M-1)))$,

$$R_0 < \frac{\delta_1}{N}, \quad |\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| < p_0\xi_0, \quad \left|\overline{E}_{\theta^*}-\overline{E}_{\lambda_0}\right| < p_0\xi_2.$$

Similarly to the proof of Lemma 16, these inequalities imply that

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_{\text{new}}}{T}\right) \geq 1-\left(\frac{3\eta}{4}+\frac{\eta}{4(M-1)}\right),$$

where δ_{new} is the unique solution to the following equation:

$$\frac{1}{1-\sqrt{2}(\xi_2+\xi_0\delta_{\text{new}}/2+\delta_{\text{new}}^2/200)} = \frac{N\sin(\delta_{\text{new}}/(2N))}{\sin(\delta_{\text{new}}/2)}.$$

Because

$$\frac{1}{1-\sqrt{2}(\xi_2+\xi_0\delta_2/2+\delta_2^2/200)} \leq \frac{\delta_2\cos(\delta_2/10)}{2\sin(\delta_2/2)} < \frac{N\sin(\delta_2/(2N))}{\sin(\delta_2/2)},$$

we must have $\delta_2 > \delta_{\text{new}}$ and

$$\mathbb{P}\left(\left|(\theta^*-\lambda_0) \bmod [-\pi/\tau,\pi/\tau]\right| < \frac{\delta_2}{T}\right) \geq 1-\left(\frac{3\eta}{4}+\frac{\eta}{4(M-1)}\right).$$

Combining this with the second inequality of Eq. (D17) ($i=2$), with probability $1-(3\eta/4+\eta/(2(M-1)))$, we obtain

$$R_0 < \frac{\delta_2}{N}, \quad |\overline{E}_{\theta^*}| + |\overline{E}_{\lambda_0}| < p_0\xi_0, \quad \left|\overline{E}_{\theta^*} - \overline{E}_{\lambda_0}\right| < p_0\xi_3,$$

which implies that

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta_3}{T}\right) \geq 1 - \left(\frac{3\eta}{4} + \frac{\eta}{2(M-1)}\right).$$

Doing this repetitively, we prove Eq. (D19). ■

Now, we are ready to prove Theorem 13.

Proof of Theorem 13. According to the conditions of Lemma 17, the smallest δ_M that we can choose is $\Theta\left(\sqrt{1-p_r(I,I')}\right)$. For a fixed decreasing sequence $\{\delta_i\}_{i=0}^M$, according to Eqs. (D5) and (D18), to make the depth smaller than δ_M/ϵ, we should choose

$$\xi_0 = \mathcal{O}\left(\min\{\delta_0^2, \delta_M\}\right), \quad \xi_i = \mathcal{O}\left(\delta_i^2\right),$$

for all $1 \leq i \leq M$. Then, according to the first inequality of Eq. (D6) and Eq. (D17), we set

$$NN_s = \widetilde{\Omega}\left(\max_{0\leq i\leq M-1}\left\{p_0^{-2}\xi^{-2}, p_0^{-2}\delta_i^2\xi_{i+1}^{-2}\right\}\right). \tag{D20}$$

Minimizing Eq. (D20) in $\{\delta_i\}_{i=0}^{M-1}, \{\xi_i\}_{i=0}^M$ with fixed δ_M, a proper choice of these parameters should be

$$\xi_i = \Theta\left(\delta_i^2\right), \quad \delta_i = \Theta\left(\delta_M^{\frac{2-(1/2)^i}{2-(1/2)^M}}\right),$$

for $0 \leq i \leq M$. Thus, to ensure that

$$\mathbb{P}\left(\left|(\theta^* - \lambda_0) \bmod [-\pi/\tau, \pi/\tau]\right| < \frac{\delta_M}{T}\right) \geq 1 - \eta,$$

we can choose

$$NN_s = \Omega\left(\delta_M^{-\frac{4}{2-(1/2)^M}} \mathrm{polylog}(M\delta_M^{-1}\eta^{-1}p_0^{-1}\mathcal{F})\right).$$

Set $\zeta = \frac{1}{2^M}\frac{1}{2-(1/2)^M}$; hence we conclude the proof. ■

APPENDIX E: ADDITIONAL ESTIMATES

1. Properties of $|(\exp(i\theta N) - 1)/(\exp(i\theta) - 1)|$ when $\theta \in [0, \pi/N]$

In this appendix, we show that $|(\exp(i\theta N) - 1)/(\exp(i\theta) - 1)|$ is a decreasing and concave function when $\theta \in [0, \pi/N]$.

First, we have the bound

$$\left|\frac{\exp(i\theta N) - 1}{\exp(i\theta) - 1}\right| = \left|\sum_{j=0}^{N-1} e^{ij\theta}\right| \leq N. \tag{E1}$$

To study the other properties of this function, we write it as

$$\left|\frac{\exp(i\theta N) - 1}{\exp(i\theta) - 1}\right| = \left|\frac{\sin(\theta N/2)}{\sin(\theta/2)}\right|.$$

Then, it suffices to study the function

$$g(\theta) = \frac{\sin(\theta N)}{\sin(\theta)}, \quad \theta \in [0, \pi/(2N)].$$

When $N = 2$, we obtain $g(\theta) = 2\cos(\theta)$ is decreasing and concave when $\theta \in [0, \pi/4]$. Note that

$$\frac{\sin(\theta N)}{\sin(\theta)} = \cos((N-1)\theta) + \cos(\theta)\frac{\sin((N-1)\theta)}{\sin(\theta)}.$$

Because $\cos((N-1)\theta)$ and $\cos(\theta)$ are decreasing and concave functions when $\theta \in [0, \pi/(2N)]$, using an induction argument, $g(\theta)$ is also a decreasing and concave function when $\theta \in [0, \pi/(2N)]$.

2. Lower bound for $\lim_{\theta\to\lambda_0} |f(\theta)|$ when $N > 2$

First, note that

$$\lim_{\theta\to\lambda_0} \frac{\exp(i(\lambda_0 - \theta)T) - 1}{\exp(i(\lambda_0 - \theta)\tau) - 1} = N.$$

Then,

$$\lim_{\theta\to\lambda_0} \mathrm{Re}(f(\theta)) \geq N + \mathrm{Re}\left(\sum_{k=1}^{M-1} \frac{p_k}{p_0}\frac{\exp(i(\lambda_0 - \lambda_k)T) - 1}{\exp(i(\lambda_0 - \lambda_k)\tau) - 1}\right) - \frac{\overline{E}}{p_0}N,$$

which implies that

$$\begin{aligned}\lim_{\theta\to\lambda_0} |f(\theta)| &\geq \lim_{\theta\to\lambda_0} \mathrm{Re}(f(\theta)) \\ &\geq \left(N + \mathrm{Re}\left(\sum_{k=1}^{M-1} \frac{p_k}{p_0}\frac{\exp(i(\lambda_0 - \lambda_k)T) - 1}{\exp(i(\lambda_0 - \lambda_k)\tau) - 1}\right)\right) \\ &\quad - \frac{\overline{E}}{p_0}N.\end{aligned} \tag{E2}$$

Let $\alpha = 1 + \max_{c\in(0,\pi/2]} \sin(c)/(\pi + c)$ be defined as in Eq. (B1). Now, we prove the claim

$$g(\theta) = \mathrm{Re}\left(\frac{\exp(i\theta N) - 1}{\exp(i\theta) - 1}\right) \geq -(\alpha - 1)N.$$

Assume not, meaning that there exists $\widetilde{\theta}$ such that

$$g\left(\widetilde{\theta}\right) < -(\alpha - 1)N; \quad \text{(E3)}$$

then, similarly to the argument in Appendix E 1, we first have $|\widetilde{\theta}| < 2\pi^2/3N < \pi$. Note that

$$\begin{aligned} g\left(\widetilde{\theta}\right) &= \frac{[\cos(N\widetilde{\theta}) - 1][\cos(\widetilde{\theta}) - 1] + \sin(N\widetilde{\theta})\sin(\widetilde{\theta})}{(\cos(\widetilde{\theta}) - 1)^2 + \sin^2(\widetilde{\theta})} \\ &= \frac{\cos((N-1)\widetilde{\theta}/2)\sin(N\widetilde{\theta}/2)}{\sin(\widetilde{\theta}/2)} \\ &= \cos^2((N-1)\widetilde{\theta}/2) + \frac{1}{2}\frac{\sin((N-1)\widetilde{\theta})\cos(\widetilde{\theta}/2)}{\sin(\widetilde{\theta}/2)} \\ &\geq \frac{1}{2}\frac{\sin((N-1)\widetilde{\theta})\cos(\widetilde{\theta}/2)}{\sin(\widetilde{\theta}/2)}. \end{aligned}$$

Since $g\left(\widetilde{\theta}\right) < 0$, we must have $(\sin((N-1)\widetilde{\theta}))/\sin(\widetilde{\theta}/2) < 0$, which implies that $\left|\widetilde{\theta}\right| > \pi/(N-1)$. Let $c^* = (N-1)|\widetilde{\theta}| - \pi$; then $0 < c < \pi/2$ and

$$\begin{aligned} \left|g\left(\widetilde{\theta}\right)\right| &\leq \frac{1}{2}\frac{\sin(c)}{\tan((c+\pi)/(2(N-1)))} \leq (N-1)\frac{\sin(c^*)}{\pi + c^*} \\ &\leq (N-1)\left(\max_{c\in(0,\pi/2]}\frac{\sin(c)}{\pi + c}\right), \end{aligned}$$

where we use $\tan(\theta) \geq \theta$ when $\theta \in [0, \pi/2]$ in the second inequality. This contradicts the assumption in Eq. (E3). Thus, we must have $g(\theta) \geq -\left(\max_{c\in(0,\pi/2]}\sin(c)/(\pi + c)\right) N = -(\alpha - 1)N$, which finally implies that

$$\lim_{\theta\to\lambda_0}|f(\theta)| \geq \left(1 - (\alpha - 1)\frac{1 - p_0}{p_0} - \frac{\overline{E}}{p_0}\right)N. \quad \text{(E4)}$$

[1] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, How to perform the most accurate possible phase measurements, Phys. Rev. A **80**, 052114 (2009).

[2] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free Heisenberg-limited phase estimation, Nature **450**, 393 (2007).

[3] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Province, RI, 2002).

[4] E. Knill, G. Ortiz, and R. D. Somma, Optimal quantum measurements of expectation values of observables, Phys. Rev. A **75**, 012328 (2007).

[5] D. Nagaj, P. Wocjan, and Y. Zhang, Fast amplification of QMA, Quantum Inf. Comput. **9**, 1053 (2009).

[6] D. Poulin and P. Wocjan, Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer, Phys. Rev. Lett. **103**, 220502 (2009).

[7] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[8] The analysis in Ref. [7, Section 5.2.1] assumes the phase $\lambda \in [0, 1)$. For a general H the eigenvalues of which are contained in $[-\|H\|_2, \|H\|_2)$, we first define a scaled Hamiltonian $\widetilde{H} = \frac{1}{2}(H/\|H\|_2 + I)$, and apply QPE with a forward Hamiltonian evolution (i.e., applying $e^{-it\widetilde{H}}$ with $t > 0$). In order to estimate λ to precision ϵ, we need to estimate $\widetilde{\lambda} = \frac{1}{2}(\lambda/\|H\|_2 + 1)$ to precision $\widetilde{\epsilon} = \epsilon/(2\|H\|_2)$ with high success probability and the scaled maximal evolution time satisfies $\widetilde{T}_{\max} \geq 2\pi/\widetilde{\epsilon}$. Therefore, the unscaled maximal evolution time (with respect to $U = e^{-itH}$) is $T_{\max} = \widetilde{T}_{\max}/(2\|H\|_2) = 2\pi/\epsilon$. This analysis also implies that if we use QPE with both forward and backward Hamiltonian evolution (i.e., applying $U = e^{-itH}$ with $t \in \mathbb{R}$), then the maximal evolution time can be halved and $T_{\max} \geq \pi/\epsilon$.

[9] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photonics **5**, 222 (2011).

[10] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, General Optimality of the Heisenberg Limit for Quantum Metrology, Phys. Rev. Lett. **105**, 180402 (2010).

[11] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Ultimate limits to quantum metrology and the meaning of the Heisenberg limit, Phys. Rev. A **85**, 042112 (2012).

[12] R. B. Griffiths and C.-S. Niu, Semiclassical Fourier Transform for Quantum Computation, Phys. Rev. Lett. **76**, 3228 (1996).

[13] Y. Dong, L. Lin, and Y. Tong, Ground-State Preparation and Energy Estimation on Early Fault-Tolerant Quantum Computers via Quantum Eigenvalue Transformation of Unitary Matrices, PRX Quantum **3**, 040305 (2022).

[14] W. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. Whaley, A non-orthogonal variational quantum eigensolver, New J. Phys. **22**, 07 (2020).

[15] L. Lin and Y. Tong, Heisenberg-Limited Ground State Energy Estimation for Early Fault-Tolerant Quantum Computers, PRX Quantum **3**, 010318 (2022).

[16] S. Lu, M. C. Bañuls, and J. I. Cirac, Algorithms for Quantum Simulation at Finite Energies, PRX Quantum **2**, 020321 (2021).

[17] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error Mitigation via Verified Phase Estimation, PRX Quantum **2**, 020317 (2021).

[18] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).

[19] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe, The power of quantum systems on a line, Commun. Math. Phys. **287**, 41 (2009).

[20] J. Kempe, A. Kitaev, and O. Regev, in *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science* (Springer-Verlag, Berlin, Heidelberg, Chennai, India, 2005), p. 372.

[21] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Generalized uncertainty relations: Theory, examples, and Lorentz invariance, Ann. Phys. **247**, 135 (1996).

[22] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Metrology, Phys. Rev. Lett. **96**, 010401 (2006).

[23] D. Wang, O. Higgott, and S. Brierley, Accelerated Variational Quantum Eigensolver, Phys. Rev. Lett. **122**, 140504 (2019).
[24] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements, New J. Phys. **11**, 073023 (2009).
[25] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92**, 062315 (2015).
[26] F. Belliardo and V. Giovannetti, Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error, Phys. Rev. A **102**, 042613 (2020).
[27] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, Experimental Phase Estimation Enhanced by Machine Learning, Phys. Rev. Appl. **10**, 044033 (2018).
[28] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, Experimental Demonstration of a Cheap and Accurate Phase Estimation, Phys. Rev. Lett. **118**, 190502 (2017).
[29] A. E. Russo, K. M. Rudinger, B. C. A. Morrison, and A. D. Baczewski, Evaluating Energy Differences on a Quantum Computer with Robust Phase Estimation, Phys. Rev. Lett. **126 21**, 210501 (2021).
[30] H. Ni, H. Li, and L. Ying, On Low-Depth Algorithms for Quantum Phase Estimation, arXiv:2302.02454 (2023).
[31] T. O'Brien, B. Tarasinski, and B. Terhal, Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments, New J. Phys. **21**, 02 (2019).
[32] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Simulating physical phenomena by quantum networks, Phys. Rev. A **65**, 042323 (2002).
[33] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).
[34] 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, Nat. Commun. **5**, 4213 (2014).
[35] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).
[36] N. H. Stair, R. Huang, and F. A. Evangelista, A multireference quantum Krylov algorithm for strongly correlated electrons, J. Chem. Theory Comput. **16**, 2236 (2020).
[37] E. N. Epperly, L. Lin, and Y. Nakatsukasa, A theory of quantum subspace diagonalization, SIAM J. Matrix Anal. Appl. **43**, 1263 (2022).
[38] D. S. Abrams and S. Lloyd, Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors, Phys. Rev. Lett. **83**, 5162 (1999).
[39] Y. Ge, J. Tura, and J. I. Cirac, Faster ground state preparation and high-precision ground energy estimation with fewer qubits, J. Math. Phys. **60**, 022202 (2019).
[40] K. Wan, M. Berta, and E. T. Campbell, Randomized Quantum Algorithm for Statistical Phase Estimation, Phys. Rev. Lett. **129**, 030503 (2022).
[41] G. Wang, D. Stilck-Franca, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, arXiv:2209.06811 (2022).
[42] P. Weinberg and M. Bukov, QuSpin: a PYTHON Package for Dynamics and Exact Diagonalisation of Quantum Many Body Systems part I: Spin chains, SciPost Phys. **2**, 003 (2017).
[43] G. Beylkin and L. Monzón, On approximation of functions by exponential sums, Appl. Comput. Harmon. Anal. **19**, 17 (2005).
[44] A. Bunse-Gerstner, in *1997 European Control Conference (ECC)* (Institute of Electrical and Electronics Engineers, Brussels, Belgium, 1997), p. 3747.
[45] A. Damle, G. Beylkin, T. Haut, and L. Monzón, Near optimal rational approximations of large data sets, Appl. Comput. Harmon. Anal. **35**, 251 (2013).
[46] J. Hauer, C. Demeure, and L. Scharf, Initial results in Prony analysis of power system response signals, IEEE Trans. Power Syst. **5**, 80 (1990).
[47] R. Zhang, G. Wang, and P. Johnson, Computing ground state properties with early fault-tolerant quantum computers, Quantum **6**, 761 (2022).
[48] https://github.com/zhiyanding/QCELS