

# Exponential Distillation of Dominant Eigenproperties


Bence Bakó<sup>1,2,3</sup>, Tenzan Araki<sup>1,4</sup> and Bálint Koczor<sup>1,\*</sup>

<sup>1</sup>Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom

<sup>2</sup>HUN-REN Wigner Research Centre for Physics, Konkoly–Thege Miklós út 29-33, Budapest H-1525, Hungary

<sup>3</sup>Faculty of Informatics, Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest H-1117, Hungary

<sup>4</sup>Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

 (Received 23 June 2025; revised 26 November 2025; accepted 9 December 2025; published 18 February 2026)

Estimating observable expectation values in eigenstates of quantum systems has a broad range of applications and is an area where early fault-tolerant quantum computers may provide practical quantum advantage. We develop a hybrid quantum-classical algorithm that enables the estimation of an arbitrary observable expectation value in an eigenstate, given an initial state is supplied that has dominant overlap with the targeted eigenstate—but may overlap with any other eigenstates. Our approach builds on and is conceptually similar to purification-based error mitigation techniques; however, it achieves exponential suppression of algorithmic errors using only a single copy of the quantum state. The key innovation is that random time evolution is applied in the quantum computer to create an average mixed quantum state, which is then virtually purified with exponential efficacy. We prove rigorous performance guarantees and conclude that the complexity of our approach depends directly on the energy gap in the problem Hamiltonian and remarkably, can be compared to phase estimation combined with amplitude estimation in terms of its scaling with respect to a target precision. We demonstrate in a broad range of numerical simulations the applicability of our framework in near-term and early fault-tolerant settings. Furthermore, we demonstrate in a 100-qubit example that direct classical simulation of our approach enables the prediction of ground and excited state properties of quantum systems using tensor-network techniques, which we recognize as a quantum-inspired classical approach.

DOI: [10.1103/bglh-9snd](https://doi.org/10.1103/bglh-9snd)

## I. INTRODUCTION

Estimating eigenstate properties of quantum many-body systems is a central task in quantum chemistry, material science, high-energy physics, and beyond [1–4]. While quantum computers may offer an exponential advantage over classical techniques when a good initial state is known [5–7], significant challenges remain [8,9]. For example, typical fault-tolerant quantum algorithms require deep circuits beyond the capabilities of early fault-tolerant machines, expected to emerge in the near term [10,11]. In contrast, variational quantum algorithms are suitable for near-term quantum devices as heuristic alternatives to investigate ground [12–16] and excited state properties [17–19]; however, they may suffer from large shot-number

requirements, from getting stuck in local minima, and from barren plateaus [20–22].

To bridge the gap between variational techniques and fully fault-tolerant quantum algorithms, a range of promising hybrid algorithms have been developed that rely on classically postprocessing data obtained from randomly sampled quantum circuits with moderate quantum resource requirements. These include techniques based on classical shadows [18,19,23,24], statistical phase estimation [25–29], and on Monte Carlo (MC) methods [30–34].

In this work, we present a hybrid quantum-classical algorithm (summarized in Fig. 1) that we call distillation of dominant eigenproperties (DDE). Given an input state that may have overlap with a large number of eigenstates, but has dominant overlap with a targeted eigenstate, our approach enables the estimation of observable expected values in the targeted eigenstate—therefore, generalizing statistical phase estimation type approaches that are limited to eigenenergy estimation [25,26,28,29]. Our approach is inspired by purification-based quantum error mitigation (QEM) techniques [35,36] and we show that DDE exponentially suppresses contributions from nondominant

\*Contact author: [balint.koczor@maths.ox.ac.uk](mailto:balint.koczor@maths.ox.ac.uk)

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

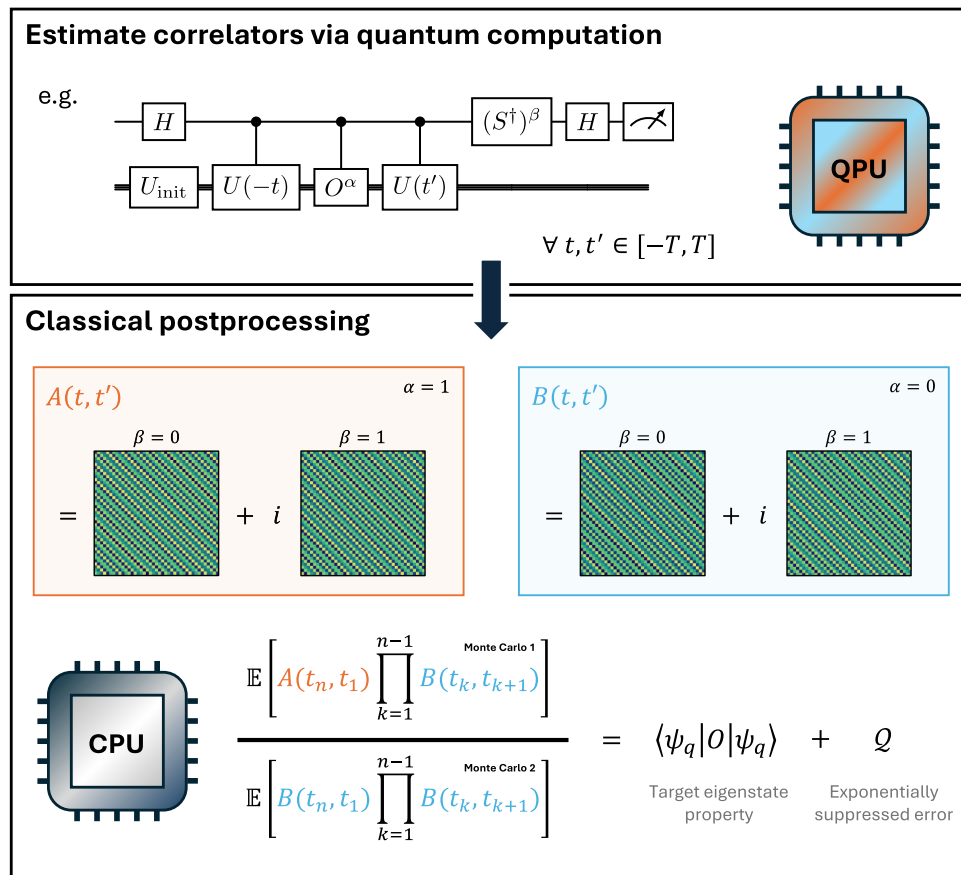


FIG. 1. The distillation of dominant eigenproperties approach proceeds by estimating two-time correlators  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$  using Hadamard test circuits over a 2D time grid in  $t$  and  $t'$ . Then, Monte Carlo sampling is used to evaluate high-dimensional integrals, which lead to the estimation of an observable expectation value in a target eigenstate of a problem Hamiltonian. Theorem 1 guarantees that the error  $\mathcal{Q}$  of the final estimate is suppressed exponentially with respect to chosen hyperparameters.

eigenstates, but without having to prepare multiple copies. It achieves a fundamentally reduced circuit depth at the cost of a reasonably increased number of quantum circuit repetitions and is therefore ideally positioned for early fault-tolerant devices. While related works can estimate expected values using similar quantum resource requirements [31–34], as we detail in Appendix A, our approach is not based on imaginary time evolution; therefore, it is not limited to the ground state and requires no explicit knowledge of the relevant eigenenergies of the problem Hamiltonian. Furthermore, while DDE is more generally applicable than statistical phase estimation algorithms, it achieves optimally shallow circuit depths  $\tilde{O}(\Delta^{-1})$ , comparable to state-of-the-art energy estimation approaches [26,27]. However, we detail in Appendix A that estimating arbitrary expectation values in eigenstates has an increased complexity compared to estimating the energy alone. Remarkably, our approach for estimating arbitrary expectation values achieves a scaling (with respect to the precision) that is comparable to that of

fault-tolerant phase estimation combined with amplitude estimation.

DDE relies on estimating two-time correlations that can be computed using a quantum device and subsequently processed classically. We need only assume the availability of a quantum state with dominant overlap with the targeted eigenstate and a finite energy gap between the target eigenstate and any other eigenstate in the support of the initial state, which is a common assumption across many quantum simulation techniques [37–39]. Obtaining a good initial state is a challenging task in general, based on complexity theoretic results [40]. However, very successful classical approximation methods have been developed for a broad range of important practical problems in quantum chemistry, solid-state physics, and material science, such as matrix product states, Hartree-Fock, and post-Hartree-Fock techniques. The community has identified a range of problems where no known classical approximations are sufficiently accurate to be useful, but can still be used to provide a reasonably good initial state

for quantum computation [41,42]. The primary example here is FeMoco, which is a key target for substantial early practical quantum advantage [37,38]—classical tensor-network approximations have been used to prepare an initial state with above 90% fidelity, but quantum computing is necessary to improve upon this accuracy in order to be useful in practical applications.

Our general framework is compatible with a broad range of initial state preparation and time evolution techniques. We demonstrate in explicit numerical simulations that the initial state can be successfully prepared using Hartree-Fock or matrix product state (MPS) [37–39] approximations, or through quantum heuristics such as the variational quantum eigensolver (VQE) [43]. We also demonstrate that time evolution can be implemented using near-term methods, such as variational quantum simulation [44], or early fault-tolerant techniques, such as product formulas [45,46].

The rest of the paper is organized as follows: In Sec. II, we first present the theoretical basis of DDE and our rigorous performance guarantees. The main results are then presented in Sec. III, in which DDE is formally introduced, and its implementation is detailed. We demonstrate in Sec. IV the utility of DDE through numerical simulations, when combined with existing techniques in both ideal and more realistic hardware implementations. In Sec. V, we detail a novel classical simulation technique for estimating ground and excited state energies in a 100-qubit spin system using tensor-network classical simulation of our DDE approach. We finally conclude in Sec. VI.

## II. THEORETICAL GUARANTEES

The hybrid DDE algorithm proposed in this work is formally related to, and draws inspiration from, the QEM technique virtual distillation (VD) [35,36]. Given  $n$  copies of a mixed state that correspond to the noisy execution of a quantum circuit, VD effectively purifies it to exponentially suppress erroneous contributions in expected values. The crucial difference is that the present DDE algorithm deliberately creates a mixed state out of a set of randomly prepared pure states. In the following, we derive new results to show that sampling random evolution times enables the preparation of a mixed quantum state that is nearly diagonal in the energy eigenbasis of a prespecified problem Hamiltonian. Then, virtually purifying this state as in Refs. [35,36] allows the estimation of eigenstate properties with exponential convergence guarantees.

### A. Random time evolution

Consider a Hamiltonian expressed in its spectral decomposition as  $\mathcal{H} = \sum_k E_k |\psi_k\rangle\langle\psi_k|$ , with eigenenergies  $E_k$  and energy eigenstates  $|\psi_k\rangle$ . An arbitrary initial state can be expressed as  $|\psi(0)\rangle = \sum_k c_k |\psi_k\rangle$ . Time evolving this initial state under  $\mathcal{H}$  leads to  $|\psi(t)\rangle = \sum_k c_k e^{-iE_k t} |\psi_k\rangle$ ,

and we define the corresponding density matrix as  $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ .

We also define the state  $\rho := \sum_k p_k |\psi_k\rangle\langle\psi_k|$ , which is diagonal in the energy eigenbasis with eigenvalues related to the initial state amplitudes as  $p_k = |c_k|^2$ . The following lemma guarantees that if we randomly sample time-evolved states  $\rho(t)$  such that  $t$  is chosen from an appropriate normal distribution, then the average state approximates  $\rho$ .

*Lemma 1.* Averaging the time-evolved states  $\rho(t) = e^{-i\mathcal{H}t} |\psi(0)\rangle\langle\psi(0)| e^{i\mathcal{H}t}$  over evolution times  $t$  that are randomly sampled according to a normal distribution  $\mathcal{N}(0, \sigma)$  leads to a mixed state  $\bar{\rho}$  given by

$$\bar{\rho} = \mathbb{E}_t[\rho(t)] = \int_{-\infty}^{\infty} G(t)\rho(t) dt = \rho + \mathcal{E}, \quad (1)$$

where  $G(t) = (\sqrt{2\pi}\sigma)^{-1} e^{-\frac{t^2}{2\sigma^2}}$  is the probability density function (PDF) of the normal distribution, and  $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$  is a diagonal state with eigenvalues  $p_k = |c_k|^2$  determined by the initial state  $|\psi(0)\rangle = \sum_k c_k |\psi_k\rangle$ . The error term  $\mathcal{E}$  decreases superexponentially as we increase the standard deviation as  $\|\mathcal{E}\|_{\text{HS}} \leq e^{-\frac{\Delta^2\sigma^2}{2}} \sqrt{(1 - e^{-H_2(\mathbf{p})})}$ , where

$$\Delta = \min_{j,k:|c_j|,|c_k|>0} \{|E_j - E_k|\} \quad (2)$$

is the smallest energy gap in the support of the initial state. Furthermore,  $H_2(\mathbf{p})$  is the second Rényi entropy over  $\mathbf{p} = (p_1, p_2, p_3, \dots)$  [see Eq. (3)], and  $\|\cdot\|_{\text{HS}}$  is the usual Hilbert-Schmidt norm. The trace distance is similarly bounded as  $\|\mathcal{E}\|_1 = \|\bar{\rho} - \rho\|_1 \leq e^{-\frac{\Delta^2\sigma^2}{2}} \|\mathbf{c}\|_1$ , where  $\|\mathbf{c}\|_1$  is the  $\ell_1$  norm of the coefficient vector  $\mathbf{c} = (c_1, c_2, c_3, \dots)$  in the eigenbasis of the residual operator  $\mathcal{E}$ .

Please refer to Appendix B for a detailed proof of the Hilbert-Schmidt and the trace distance bounds. The Gaussian window function in the integral enables us to truncate the time grid to  $[-T, T]$  for a sufficiently large  $T$ , e.g.,  $T = 5\sigma$ .

In Fig. 2(left), we numerically verify the above error bounds by simulating a 12-qubit random-field Heisenberg model with periodic boundary conditions—we defer details of the simulation to Appendix E 1. We plot both the Hilbert-Schmidt distance  $\|\mathcal{E}\|_{\text{HS}}$  and the trace distance  $\|\mathcal{E}\|_1$  (dashed lines) for an increasing standard deviation  $\sigma$ , and confirm that they are indeed below their theoretical upper bounds (solid lines).

Most prominent quantum simulation algorithms, such as quantum phase estimation, assume that an initial state that has a dominant overlap with a target eigenstate  $|\psi_q\rangle$  can be prepared efficiently. In these typical applications, the

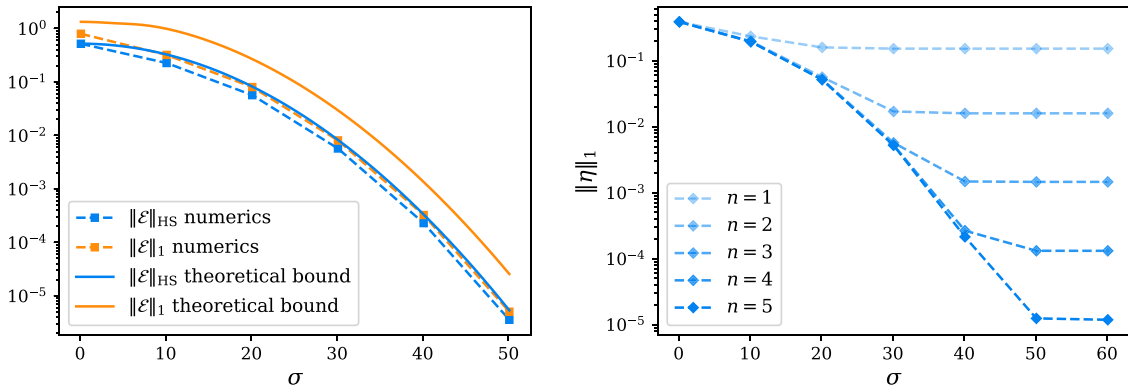


FIG. 2. Numerically verifying our error bounds. Using an initial state with  $p_q \approx 0.85$  and a time grid defined on  $[-5\sigma, 5\sigma]$  with step size  $dt = 1$ , we apply temporal averaging in a 12-qubit random-field Heisenberg model simulation. Error contributions are shown against the standard deviation  $\sigma$ . (Left) Verifying Lemma 1: numerically computed error  $\mathcal{E} = \bar{\rho} - \rho$  norms (dashed lines) are indeed below their corresponding analytical upper bounds (solid lines) and decrease superexponentially as we increase the standard deviation of the normal distribution. (Right) Numerically demonstrating Theorem 1, the error scaling, as measured in the trace norm of  $\eta = \bar{\rho}^n / \text{tr}[\bar{\rho}^n] - |\psi_q\rangle\langle\psi_q|$ , is shown for an increasing number of copies, where  $|\psi_q\rangle$  denotes the target eigenstate. As stated in Theorem 1, this error norm is a sum of two terms: the first term is suppressed exponentially as we increase the number of copies  $n$ , whereas the second term is suppressed superexponentially as we increase  $\sigma$ , as illustrated above.

initial state is low in energy and, therefore, only has support on the few lowest-lying eigenstates; thus  $\Delta$  in the error bounds in Lemma 1 corresponds to the smallest gap among the low-lying eigenstates. In complete generality, the error term  $\mathcal{E}$  can be suppressed efficiently as long as the initial state only overlaps with eigenstates that are well separated in energy. In such instances,  $\rho$  is of low rank, and we obtain a tighter general bound on the trace distance as  $\|\mathcal{E}\|_1 \leq 2\sqrt{\text{rank}(\bar{\rho})}\|\mathcal{E}\|_{\text{HS}} \leq 2\sqrt{\text{rank}(\rho)}\|\mathcal{E}\|_{\text{HS}}$ . Since  $\text{rank}(\rho)$  is equal to the support of  $\mathbf{p}$ , this prefactor expresses how many eigenstates the initial state overlaps with.

Furthermore, the approach becomes more efficient as we increase the overlap of the initial with the target state. Specifically, the Hilbert-Schmidt distance bound in Lemma 1 depends on the Rényi entropy

$$H_a(\mathbf{p}) := \frac{1}{1-a} \ln \left[ \sum_k p_k^a \right] \quad (3)$$

of order  $a = 2$ , which can be upper bounded as

$$H_2(\mathbf{p}) \leq 2H_\infty(\mathbf{p}) = -2 \ln [\max_k p_k]. \quad (4)$$

Substituting the above bound back to Lemma 1 confirms that, indeed,  $\|\mathcal{E}\|_{\text{HS}}$  vanishes in the limit where the initial state is exactly an eigenstate of the Hamiltonian via  $H_2(\mathbf{p}) \rightarrow 0$ . In the rest of this work, we will assume that the initial state has a dominant overlap with a target eigenstate  $|\psi_q\rangle$ , and we will demonstrate a range of practical approaches for the preparation of such initial states in Sec. IV. Our error bound then also scales with this dominant overlap  $p_q = |\langle\psi_q|\psi(0)\rangle|^2$  as  $H_2(\mathbf{p}) \leq -2 \ln [p_q]$ .

## B. Applying virtual distillation

References [35,36] presented an approach for estimating nonlinear functionals of the density matrix of the form  $\text{tr}[\rho^n O]$ , where  $O$  is a Hermitian observable. We now use our result from Lemma 1 to prepare an average mixed state that has dominant overlap with one specific eigenstate  $|\psi_q\rangle$  and apply Theorem 2 of Ref. [35], which we adapt for the present context as the following lemma.

*Lemma 2 (Theorem 2 of Ref. [35]).* Consider the mixed state  $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$  and assume that it has a dominant overlap  $p_q$  with a target eigenstate  $|\psi_q\rangle$ . Applying results of Refs. [35,36] for estimating nonlinear functionals  $\text{tr}[\rho^n O]$  for a Hermitian observable  $O$ , the expectation value  $\langle\psi_q| O |\psi_q\rangle$  can be estimated as

$$\frac{\text{tr}[\rho^n O]}{\text{tr}[\rho^n]} = \langle\psi_q| O |\psi_q\rangle + \mathcal{A}, \quad (5)$$

where the error term  $\mathcal{A}$  is exponentially suppressed via the explicit expression from Ref. [35] as  $|\mathcal{A}| \leq 2(p_q^{-1} - 1)^n \exp\{-(n-1)H_n(\mathbf{p}')\}$ , where  $p'_k = p_k / \sum_{i \neq q} p_i$  for  $k \neq q$ .

Using our results from Lemma 1 and the results of Refs. [35,36], as summarized above in Lemma 2, we can apply VD to the time-averaged state  $\bar{\rho}$  to estimate the target eigenstate property  $\langle\psi_q| O |\psi_q\rangle$  with provably exponential convergence guarantees. The joint error bound is summarized in the following theorem.

*Theorem 1.* Given  $n$  copies of the time-averaged state  $\bar{\rho}$  from Lemma 1, one can apply virtual distillation as

in Lemma 2 to estimate the expectation value of the Hermitian observable  $O$  as

$$\frac{\text{tr}[\bar{\rho}^n O]}{\text{tr}[\bar{\rho}^n]} = \langle \psi_q | O | \psi_q \rangle + \mathcal{Q}, \quad (6)$$

where the error  $\mathcal{Q}$  can be bounded as

$$|\mathcal{Q}| \leq |\mathcal{A}| + 4np_q^{-1} \|O\|_\infty \|\mathcal{E}\|_1 + \mathcal{O}(\|\mathcal{E}\|_1^2). \quad (7)$$

Substituting our bounds on  $\|\mathcal{E}\|_1$  and  $|\mathcal{A}|$  from Lemmas 1 and 2, respectively, we obtain

$$|\mathcal{Q}| \leq 2(p_q^{-1} - 1)^n \exp\{-(n-1)H_n(\mathbf{p}')\} + 4np_q^{-1} e^{-\frac{\Delta^2 \sigma^2}{2}} \|\mathbf{c}\|_1 \|O\|_\infty + \mathcal{O}(\|\mathcal{E}\|_1^2),$$

which shows explicit dependence on the absolute largest observable eigenvalue  $\|O\|_\infty$ , the probability distribution  $p'_k = p_k / \sum_{i \neq q} p_i$  for  $k \neq q$ , and  $\|\mathbf{c}\|_1$ , the  $\ell_1$  norm of the coefficient vector  $\mathbf{c} = (c_1, c_2, c_3, \dots)$  from Lemma 1. Finally, suppressing errors to a desired level of precision  $|\mathcal{Q}|$  requires the following number of copies and time evolution window depth (quantum resources):

$$n = \mathcal{O}(\ln[|\mathcal{Q}|^{-1}]) \quad \text{and} \quad \sigma = \mathcal{O}(\Delta^{-1} \sqrt{\ln[|\mathcal{Q}|^{-1}n]}), \quad (8)$$

ignoring logarithmic factors.

Please refer to Appendix B for a proof. Indeed, our approach matches the  $\tilde{\mathcal{O}}(\Delta^{-1})$  circuit depth scaling of state-of-the-art phase estimation protocols [25,26], which we further detail in Appendix A. The first term in the above error bound can be suppressed exponentially by increasing  $n$ . As we discuss in the next section, the suppression of this first term does not require additional quantum resources, as the number of (virtual) copies only affects classical post-processing in our approach. Therefore, only the suppression of the second error term (along with the higher order terms) requires an increase in quantum resources through increasing the standard deviation  $\sigma$  and consequently the time of the quantum simulation.

We confirm the above bounds in Fig. 2(right) by plotting the trace distance  $\|\bar{\rho}^n / \text{tr}[\bar{\rho}^n] - |\psi_q\rangle\langle\psi_q|\|_1$ , which generally upper bounds the left-hand side of Eq. (7) independently of the observable via

$$\left| \frac{\text{tr}[\bar{\rho}^n O]}{\text{tr}[\bar{\rho}^n]} - \langle \psi_q | O | \psi_q \rangle \right| \leq 2\|O\|_\infty \left\| \frac{\bar{\rho}^n}{\text{tr}[\bar{\rho}^n]} - |\psi_q\rangle\langle\psi_q| \right\|_1.$$

Refer to Ref. [47] for a detailed proof of the above bound. Indeed, Fig. 2(right) nicely illustrates that our error bound in Theorem 1 is a sum of two terms: The first term  $|\mathcal{A}|$  is suppressed exponentially as we increase  $n$ , whereas the second term is increased linearly in  $n$  but gets superexponentially suppressed as we increase the Gaussian window  $\sigma$ .

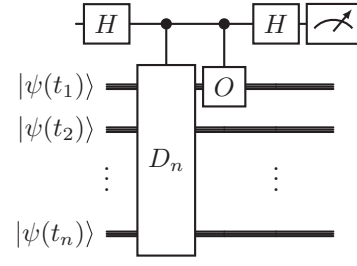


FIG. 3. Quantum circuit that enables estimating eigenstate properties, directly following Lemmas 1 and 2. The initial state  $|\psi(0)\rangle$  is time evolved for a randomly sampled duration in each of the registers before a derangement (denoted  $D_n$ ) and the observable  $O$  are applied, both conditioned on the state of the ancilla qubit. While this implementation requires  $n$  quantum registers, the distillation of dominant eigenproperties approach only requires one register for any  $n$  by using either the circuits shown in Fig. 1 or Fig. 10.

### III. DISTILLATION OF DOMINANT EIGENPROPERTIES

Theorem 1 allows us to apply VD to randomly time-evolved initial states in order to estimate eigenstate properties with exponential convergence guarantees as we increase the number of copies. In this section, we describe our approach for achieving this convergence through the DDE algorithm. In particular, we show that virtual purification of randomly time-evolved states can be performed classically via MC integration, without requiring multiple copies of the quantum state. This way, our approach separates into a quantum computation step and a classical post-processing step. We also provide theoretical guarantees for the variance of the MC estimator.

#### A. Suppressing nondominant contributions by estimating nonlinear functionals

Given  $n$  copies of the time-averaged state  $\bar{\rho}$  from Lemma 1, we can estimate the required nonlinear functionals in Lemma 2 via the circuit in Fig. 3. Specifically, the  $n$  input states are randomly time-evolved pure states  $|\psi(t_i)\rangle$  for  $i \in \{1, 2, \dots, n\}$ ,  $D_n$  is a cyclic permutation or derangement operator [35], and  $O$  is the observable of interest. The controlled  $O$  operation is implemented either by applying individual Pauli terms in the decomposition of  $O$  or via a block encoding using the LCU approach [48]. Estimating the ancilla probability then allows us to estimate the nonlinear functionals  $\text{tr}[\bar{\rho}^n O]$  required in Lemma 2, with its complexity summarized in Lemma 3.

*Lemma 3.* Sampling the ancilla qubit in Fig. 3 allows us to estimate  $\langle \psi_q | O | \psi_q \rangle$  to statistical uncertainty  $\epsilon$  and bias  $\mathcal{Q}$  using time evolutions of depth  $\sigma$ , and number of copies  $n$  from Eq. (8). The required number of shots then scales

as

$$N_s = \tilde{O} \left( p_q^{-2 \ln(|Q|^{-1})} \epsilon^{-2} \right). \quad (9)$$

For a proof, please refer to Appendix B. Therefore, the total time complexity of our approach is obtained as the number of shots multiplied with the circuit depth from Theorem 1—we also compare this complexity with state-of-the-art statistical phase estimation techniques in Appendix A.

This implementation is well suited for modular quantum architectures [49,50], given that the deepest part of the algorithm is the time evolution, which can be performed independently within each module; then a relatively shallow and sparse derangement circuit is used to entangle the copies. However, this implementation does require  $n$  quantum registers to store the  $n$  copies.

In the following, we detail another approach, in which only one register is required, albeit the classical postprocessing cost is substantially increased. Still, the reduction in the required number of qubits is worth the trade-off, especially in an early fault-tolerant setting where the number of logical qubits may be significantly limited. We highlight this trade-off by taking a surface code architecture as an example. In the implementation depicted in Fig. 3, independent time evolutions are performed on  $n$  copies of the initial state before virtual distillation is applied. If we instead use only a single copy and repurpose the physical qubits to increase the size of the surface code logical qubits, then we can increase the code distance by a factor of  $\mathcal{O}(\sqrt{n})$ . This increase in code distance exponentially lowers the logical error rate and therefore enables a polynomially increased time evolution depth. The latter implies that  $\sigma$  can be increased, which leads to an exponential reduction of algorithmic errors via Lemma 1.

## B. Suppressing nondominant contributions by MC integration

We now build on the observation that the nonlinear functional  $\text{tr}[\bar{\rho}^n O]$  of time-averaged states can be related to a high-dimensional integral. This then leads to an approach that does not require  $n$  copies of the average, mixed quantum state, but rather estimates temporal correlators, which are then fed into a classical postprocessing algorithm to approximately evaluate a high-dimensional integral.

*Proposition 1.* The nonlinear functional of time-averaged states  $\bar{\rho}$  is equivalent to an  $n$ -dimensional integral over  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  as

$$\text{tr}[\bar{\rho}^n O] = \int G(\mathbf{t}) F(\mathbf{t}) d\mathbf{t},$$

where  $G(\mathbf{t})$  is the PDF of an  $n$ -variate normal distribution and  $F(\mathbf{t}) = A(t_n, t_1) \prod_{k=1}^{n-1} B(t_k, t_{k+1})$  depends only

on the correlators  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$  for all  $t$  and  $t'$ .

Please refer to Appendix B for a proof. One can apply quadrature integration formulas, which we also discuss in Appendix C, to compute this integral numerically on a truncated discrete time grid. However, as the number of copies increases, this approach becomes highly inefficient as it requires resources that are exponential in the number of copies. Furthermore, certain quadrature techniques are highly susceptible to numerical instabilities in practical applications [51]. Therefore, we focus on approximate integration via MC sampling that in turn provides unbiased estimates for the functionals in question.

*Lemma 4.* Given a quantum state  $|\psi(0)\rangle$ , the integral in Proposition 1 can be replaced by the random variable  $F(\hat{\mathbf{t}})$ , where we choose  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  according to the  $n$ -variate normal distribution with  $G(\mathbf{t})$  as the PDF. We obtain an unbiased estimator as

$$\mathbb{E}[F(\hat{\mathbf{t}})] = \text{tr}[\bar{\rho}^n O],$$

through averaging the MC samples,

$$F(\mathbf{t}) = A(t_n, t_1) \prod_{k=1}^{n-1} B(t_k, t_{k+1}). \quad (10)$$

These samples can be computed classically using the two objects  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$  that are classically stored in a 2D time grid as we detail below.

The denominator in Theorem 1 is obtained as a special case

$$\text{tr}[\bar{\rho}^n] = \int G(\mathbf{t}) J(\mathbf{t}) d\mathbf{t}, \quad (11)$$

where  $J(\mathbf{t}) = B(t_n, t_1) \prod_{k=1}^{n-1} B(t_k, t_{k+1})$ . Evaluating these two integrals then allows us to estimate the expectation value of interest  $\langle \psi_q | O | \psi_q \rangle$  by estimating the ratio of two nonlinear functionals  $\text{tr}[\bar{\rho}^n O]$  and  $\text{tr}[\bar{\rho}^n]$ .

## C. Variance of the estimator

We now use median of means estimators to derive rigorous performance guarantees. In the following, by each MC sample, we refer to one classical evaluation of the object  $F(\hat{\mathbf{t}})$  or  $J(\hat{\mathbf{t}})$ . Using  $2K$  batches of  $N_{\text{batch}}$  samples each, we obtain two estimators for the numerator  $F$  (that estimates

$\text{tr}[\hat{\rho}^n O]$ ) and denominator  $J$  (that estimates  $\text{tr}[\hat{\rho}^n]$ ) as

$$\begin{aligned}\hat{\mu}_{K,b}(F) &:= \text{median}\{\hat{F}_1, \dots, \hat{F}_K\}, \\ \hat{\mu}_{K,b}(J) &:= \text{median}\{\hat{J}_1, \dots, \hat{J}_K\}.\end{aligned}\quad (12)$$

Using these two unbiased estimators, the error of their ratio can be bounded as described in the following theorem.

*Theorem 2.* Two median of means estimators, which use  $2KN_{\text{batch}}$  independent samples in total, enable accurately predicting the ratio  $\mathbb{E}[F]/\mathbb{E}[J]$  with failure probability  $\delta$  and precision  $\epsilon$ . The precision is bounded as

$$\epsilon \leq 2 \frac{|\mathbb{E}[F]| + \mathbb{E}[J]}{\mathbb{E}[J]^2} N_{\text{batch}}^{-1/2} + \mathcal{O}(N_{\text{batch}}^{-1}). \quad (13)$$

In leading terms in  $N_{\text{batch}}$ , we obtain the approximate sample complexity as

$$K = -8 \log \delta \quad \text{and} \quad N_{\text{batch}} \lesssim 4 \frac{(|\mathbb{E}[F]| + \mathbb{E}[J])^2}{\mathbb{E}[J]^4} \epsilon^{-2}. \quad (14)$$

Given a vanishing expectation value  $\mathbb{E}[F] = \text{tr}[\hat{\rho}^n O] \rightarrow 0$ , the approximate sample complexity scales as  $N_{\text{batch}} = \tilde{\mathcal{O}}\left(p_q^{-2 \ln(|Q|^{-1})} \epsilon^{-2}\right)$ .

For a proof, please refer to Appendix B.

While the assumption  $|\mathbb{E}[F]| \rightarrow 0$  can be seen as a limiting factor, we argue that it can be approximately fulfilled using the following technique. We can enforce  $\text{tr}[\hat{\rho}^n O] \approx 0$  by linearly combining the observable with the identity operator as  $O \rightarrow O - c\mathbb{1}$  for a suitably chosen scalar  $c$ , which is equivalent to applying the transformation  $A(t, t') \rightarrow A(t, t') - cB(t, t')$  to our data set. In practice, we find that choosing  $c$  to minimize  $|A(t, t') - cB(t, t')|$  also nearly minimizes the MC variance [52]. We also note that this technique is not specific to median of means estimation and can be applied to other MC estimators as well, including those relying on the sample mean.

### D. Discrete time integral

So far, we assumed that  $t$  and  $t'$  are sampled from a continuous normal distribution in our MC approach. However, as we now detail, the time variables can be discretized given our correlators are Fourier band-limited functions of time. Specifically, the two-time correlators  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$  are linear combinations of the frequency components  $e^{-i(E_j - E_k)t}$ , where  $E_k$  are eigenvalues of the Hamiltonian. Therefore, the largest frequencies that can appear are upper bounded via the spectral radius of the Hamiltonian as  $\|\mathcal{H}\|_{\infty} = \max_{j,k} |E_j - E_k|$ . According to the Nyquist-Shannon sampling theorem [53], sampling the

correlators on a uniform 2D time grid is informationally complete, provided the timestep  $dt$  satisfies the Nyquist condition relative to the highest angular frequency, i.e.,  $dt \propto \|\mathcal{H}\|_{\infty}^{-1}$ . Thus, as long as the continuous time correlators are sampled according to this condition, they can be reconstructed uniquely, e.g., one may use Fourier interpolation methods to reconstruct the correlators to an arbitrarily fine time resolution without additional quantum evaluations, given band-limited functions are fully determined by their Nyquist-rate samples. Furthermore, based on our error bound in Theorem 1, the maximum evolution time  $T$  can be upper bounded by a constant multiple of  $\Delta^{-1}$ . This follows from the rapid decay of the Gaussian distribution, making the truncation error negligible by choosing  $T$  to be a small constant multiple of the standard deviation  $\sigma$  (e.g.,  $5\sigma$ ).

### E. Estimating correlators using a quantum computer

We consider a uniform discretization of  $[-T, T]$  into  $N_T = 2T/dt + 1$  grid points. Each entry of the grid storing the  $B$  object can be computed using a Hadamard test circuit shown in Fig. 1, where the initial state  $|\psi(0)\rangle = U_{\text{init}}|0\rangle$  is prepared using  $U_{\text{init}}$  and then time evolutions with durations  $-t$  and  $t'$  are applied conditioned on the state of the ancilla qubit. Measuring the ancilla qubit in the  $X$  ( $Y$ ) basis gives an unbiased estimator for the real (imaginary) part of  $B(t, t')$ . The circuit to obtain  $A(t, t')$  is then identical to that for  $B(t, t')$ , except that the observable is additionally applied conditioned on the state of the ancilla qubit in between the two controlled time evolutions.

Estimating all entries of the  $B$  correlator matrix requires  $N_s N_T^2$  executions. The  $A$  correlator matrix similarly requires  $N_s N_T^2$  executions if  $O$  is implemented via block encoding and  $N_s N_T^2 \|\mathbf{v}\|_1^2$  executions when the individual Pauli terms in the observable are sampled via importance sampling. Here,  $\|\mathbf{v}\|_1$  is the  $\ell_1$  norm of the Pauli coefficient vector  $\mathbf{v}$  and the number of shots  $N_s \leq \epsilon_s^{-2}$  is determined by the required final precision  $\epsilon_s$ .

These circuits can be optimized as detailed in Appendix D (summarized in Fig. 10), so we refer to them as an unoptimized baseline construction. Specifically, only  $N_T - 1$  entries of the 2D time grid are unique in  $B$ , since its value depends on the time difference  $t' - t$  rather than their unique combinations. The latter property also holds when  $O$  commutes with the problem Hamiltonian in  $A$ , in which case as low as  $N_T$  entries are unique. Thus, the number of circuit executions can be fundamentally reduced to scale linearly in  $N_T$ . Moreover, since  $A$  corresponds to an expectation value rather than a two-time correlator when  $t = t'$ , the diagonal entries of  $A$  do not require a Hadamard test and can instead be evaluated using ancilla-free time evolutions with durations up to  $2T$ .

## F. Summary of the DDE algorithm

Here we summarize the workflow of DDE, which begins with quantum computation, proceeds through classical MC sampling, and finally estimates the desired eigenstate property. The input consists of a Hamiltonian  $\mathcal{H}$ , an initial state  $|\psi(0)\rangle$  that has a dominant overlap with the eigenstate of interest  $|\psi_q\rangle$ , and a target observable  $O$ . The output is an estimate of  $\langle\psi_q|O|\psi_q\rangle$  via the ratio between MC estimators  $\mathbb{E}[F]$  and  $\mathbb{E}[J]$ .

- (1) **Quantum computation:** Choose temporal grid with cutoff  $T$  (consistent with the Gaussian window width  $\sigma$ ) and timestep  $dt$ . The correlators  $A(t, t')$  and  $B(t, t')$  are estimated using a quantum device by evaluating them at discrete  $t$  and  $t'$  points. Details are discussed in Sec. III E and Appendix D.
- (2) **Generate classical MC samples:** Choose hyperparameters for classical postprocessing, namely the number of (virtual) copies  $n$  and the number of MC samples  $N_{\text{MC}}$ . Calculate a discrete probability distribution by evaluating the PDF of the normal distribution  $\mathcal{N}(0, \sigma)$  over the grid  $[-T, T]$ . Generate two sets of random time values  $\hat{t}_F = (\hat{t}_1, \hat{t}_1, \dots, \hat{t}_n)$  and  $\hat{t}_J = (\hat{t}_1, \hat{t}_1, \dots, \hat{t}_n)$  according to the discrete probability distribution  $N_{\text{MC}}$  times. Each MC sample  $F$  and  $J$  is computed as a product of  $A$  and  $B$  values at indices determined by the  $t_F$  and  $t_J$  samples, respectively, as in Eq. (10).
- (3) **Estimate eigenstate property:** The generated MC samples are used to estimate both  $\mathbb{E}[F]$  and  $\mathbb{E}[J]$  in Lemma 4. The desired estimate of the target eigenstate property is obtained by taking the ratio of empirical means  $\mathbb{E}[F]/\mathbb{E}[J]$  or via the ratio of median of means as detailed above.

Please note that the procedure outlined here is a baseline that can be improved in several ways, as we showcase in the following sections with concrete examples.

## IV. DEMONSTRATIONS

In the following, we demonstrate the efficacy of DDE in numerical experiments using different Hamiltonians relevant in condensed matter and high-energy physics applications. We employ several combinations of initial state preparation, time evolution simulation, and classical postprocessing techniques.

### A. Exact simulation

We first verify our DDE approach by performing exact time evolution in a random-field Heisenberg model on 10 qubits with periodic boundary condition. Our initial state has  $p_1 = 0.66$  overlap with the ground state and  $p_k = p_l = 0.17$  with two other eigenstates, and we estimate the expectation value of  $O = Z_1$ .

We evaluate the correlators  $A$  and  $B$  over the time grid defined by  $T = 200$  and  $dt = 0.5$  numerically exactly and use them in our MC approach. In Fig. 4(left), we plot the mean error (solid lines) in our estimate of the expected value  $\Delta\langle O \rangle$ . The result confirms that, indeed, statistical fluctuations are suppressed, and the errors converge to the systematic bias (dashed lines) that we bounded in Theorem 1 as the number of MC samples is increased. This systematic bias can be suppressed exponentially as we increase the number of virtual copies, and we achieve an error  $10^{-2.76}$  with 5 copies in the present demonstration.

While the number of copies could be increased further, one can alternatively use converged values for up to five copies and apply extrapolation, as we demonstrate in Fig. 4(right). Specifically, we approximate our analytical model of the residual error in Lemma 2 using an exponential function in  $n$  given by  $f(n) = a + cb^n$ . We fit the parameters  $a$  and  $b$  and extrapolate to  $n \rightarrow \infty$  to obtain an error  $10^{-3.13}$  in the expectation value, which otherwise would require 6 copies and  $\approx 10^7$  MC samples.

In practice, estimating the correlators  $A$  and  $B$  using a quantum device will introduce shot noise; however, in Appendix F we demonstrate that our MC approach is naturally robust against shot noise.

### B. Early fault-tolerant implementations

Time evolution is one of the most natural applications of a quantum computer and is considered to be the most promising area for achieving early quantum advantage. However, even the simplest systems that are challenging for conventional computers to simulate require quantum circuits potentially beyond the capabilities of non-error-corrected quantum devices [41]. In the following, we illustrate how our approach can be implemented on early fault-tolerant quantum devices.

#### 1. Robustness to Trotter error

It is expected that early fault-tolerant machines will enable simple Hamiltonian simulation algorithms such as Trotterization [45,46]. While advanced variants can achieve zero algorithmic error [54], product formulas in early or partially fault-tolerant devices will likely be limited by circuit depth and therefore will have substantial algorithmic Trotter errors associated with the limited circuit depth. Here, we demonstrate the robustness of DDE to Trotter error and also demonstrate how such errors can be mitigated by applying straightforward extrapolation to the estimated  $A$  and  $B$  correlators.

We consider the Jordan-Wigner transformed Hamiltonian  $\tilde{\mathcal{H}}$  of the Fermi-Hubbard model such that a 4-site system is mapped to 8 qubits, as described in Appendix E 3, and drop the identity term in  $\tilde{\mathcal{H}}$  to obtain  $\mathcal{H} = \sum_j v_j P_j$

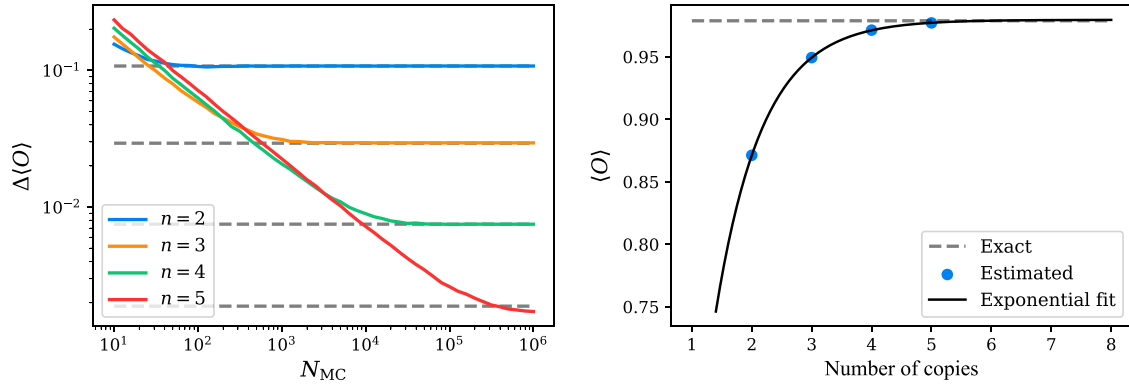


FIG. 4. Verifying the distillation of dominant eigenproperties approach in a 10-qubit random-field Heisenberg exact simulation. (Left) Mean errors  $\Delta(O)$  (estimated from  $10^4$  independent runs) for an increasing number  $n$  of virtual copies (solid lines) converge for an increasing number of Monte Carlo samples to a systematic bias consistent with Theorem 1. (Right) Plotting the systematic bias for an increasing number  $n$  of virtual copies and extrapolating to  $n \rightarrow \infty$  via an exponential model function improves our estimate.

as a linear combination of  $N_P$  Pauli terms  $P_j$  with prefactors  $v_j$ . We prepare the initial state  $\frac{1}{\sqrt{2}}(|10\ 100\ 101\rangle + |01\ 011\ 010\rangle)$ , which has an overlap of  $p_q \approx 0.62$  with the ground state (assuming half filling). Time evolutions generated by  $\mathcal{H}$  are applied on a time grid defined by  $T = 200$  and  $dt = 0.5$ .

First-order Trotterization is implemented by further dividing each timestep into  $M$  uniform intervals such that

$$\exp\{-i\mathcal{H}dt\} \approx \left( \prod_j \exp\{-iv_j P_j dt/M\} \right)^M.$$

We consider a noiseless implementation of each Trotter step, such that the only source of error is the algorithmic Trotter error and choose the Pauli correlation  $O = Z_1 Z_2$  as our observable. Figure 5(left) shows the error in the estimated observable expectation value for an increasing number  $M$  of Trotter steps and for an increasing number  $n$  of virtual copies and confirms that the approach is indeed robust against algorithmic errors as the Trotterized time evolutions rapidly (in  $M$ ) converge to exact ones (horizontal lines).

Given that the correlators  $A$  and  $B$  are estimated expectation values, algorithmic error mitigation techniques can be applied [29,55]. Here, we consider a polynomial extrapolation of the Trotter error whereby we fit a degree-3 polynomial as  $f(x) = ax^3 + bx^2 + c$  to all entries of  $A$  and  $B$  where  $x$  is related to the Trotter step through  $x = 1/M$ . We illustrate this extrapolation on single entries of  $A$  (blue) and  $B$  (orange) in Fig. 5(right, inset). Figure 5(right) verifies that DDE expectation values obtained via extrapolated  $A$  and  $B$  correlators (solid lines) closely approximate those obtained via exact time evolutions (dotted lines). This confirms that DDE is indeed compatible with error mitigation techniques.

## 2. Robustness to gate noise

Early fault-tolerant devices will likely be limited by circuit depth due to non-negligible residual logical errors. Here, we demonstrate that DDE is robust against such errors. Our error model assumes that continuous rotation gates are the dominant source of errors, as these are fundamentally expensive components in fault-tolerant architectures [29,56–58]. For example, when rotations are synthesized using Clifford+T sequences, T gates are typically the most expensive components and more noise-prone than Clifford operations, as we detail in Appendix E 5.

In Fig. 6, we report simulations assuming two error severities such that the average number of errors in the time evolution circuit is 0.1 (dotted lines) and 0.01 (dashed lines) by assuming logical errors after each continuous rotation gate with probability  $\gamma = 1/(100N_G)$ , and  $\gamma = 1/(10N_G)$ , respectively, where  $N_G = N_T N_P M$  is the total number of rotation gates in the Trotter circuits. Indeed, our DDE approach is robust to both error severities given the observable expectation values estimated with noisy circuits (dotted and dashed lines) closely approximate the estimation errors achieved with noise-free circuits (solid lines). Given that, in Fig. 6, we choose the observable to be the Hamiltonian as  $O = \mathcal{H}$ , we can exploit that  $O$  commutes with the time evolution operator and therefore our quantum circuits can be significantly simplified (see Appendix D). Indeed, the optimized circuits yield smaller errors [Fig. 6(right)] than using the unoptimized circuits [Fig. 6(left)] for two reasons. First, reduced circuit depths guarantee less noise accumulation and second, shorter time evolution durations guarantee less Trotter algorithmic errors.

## C. Variational quantum simulation

While fault-tolerant machines are emerging, it is broadly expected that the earliest forms of practical quantum

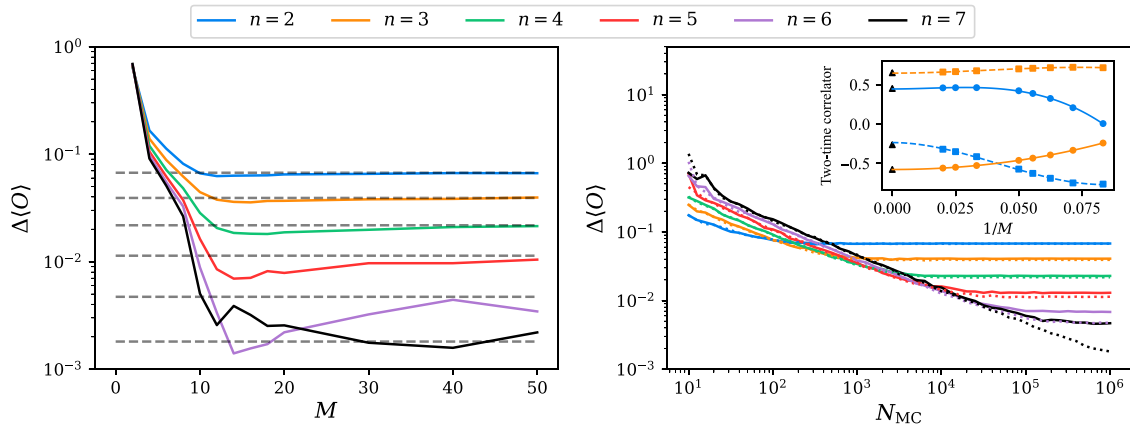


FIG. 5. Trotter simulation compared with exact time evolution in an 8-qubit Fermi-Hubbard model. (Left) Mean errors in the estimated expectation value (using  $10^6$  MC samples and averaging over  $10^3$  runs) for an increasing number  $M$  of Trotter steps and for an increasing number  $n$  of copies. Mean errors using Trotterization (solid lines) rapidly (in  $M$ ) converge to errors using exact time evolution (dashed lines) confirming the robustness of the distillation of dominant eigenproperties approach to algorithmic errors. (Right, inset) Applying polynomial extrapolation to the real (solid line, circles) and imaginary (dashed line, squares) parts of the single matrix entries  $A(169.5, 93.5)$  (blue) and  $B(-31.5, -70.5)$  (orange). The ideal values are also shown for comparison (black triangles). (Right) Mean errors in the expected value estimation using correlators  $A$  and  $B$  estimated using extrapolated entries (solid lines) closely approximate mean errors obtained via exact time evolution (dotted lines) confirming the efficacy of extrapolation as an algorithmic error mitigation technique.

advantage may be achieved with non-error-corrected devices in simulating the time evolution of relatively simple quantum systems, with variational simulation techniques [41] considered to be the most promising approach. Here we consider a simple example from high-energy physics, the lattice Schwinger model [59], as formulated in Ref. [60]. Our goal is to estimate the ground state energy of the corresponding Hamiltonian on 6 qubits, described in Appendix E 2.

First, we use VQE to obtain an initial state by optimizing parameters in an ansatz circuit  $|\psi(\theta)\rangle = U(\theta)|0\rangle$

using gradient descent and deliberately terminate the training before convergence (after 50 steps). Here, we apply 12 layers of a hardware-efficient ansatz with ring topology and show the training curve in Fig. 7(left), as well as populations of the 7 lowest energy eigenstates in the state obtained after 50 training steps. Indeed, a significant disadvantage of VQE is that it is prone to getting stuck in local minima [21], which feature we exploit in the present application: The quantum state we obtain after only 50 training steps has a sufficiently high  $p_1 = 0.53$  overlap with the ground state

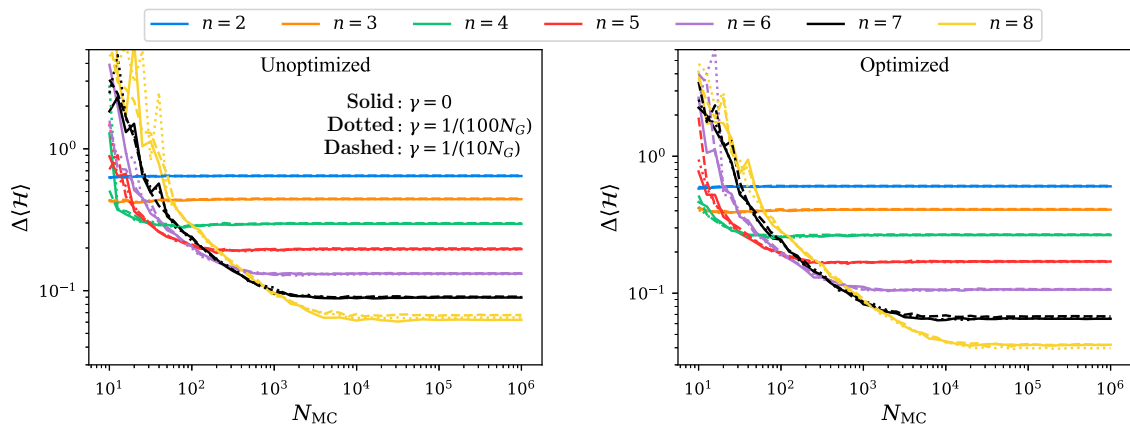


FIG. 6. Noisy Trotter circuit simulation of an 8-qubit Fermi-Hubbard model. The mean error (over  $10^3$  independent runs) of the observable ( $O = \mathcal{H}$ ) expectation value estimate for an increasing number of Monte Carlo samples is shown for increasing logical error rates (dashed vs dotted) for both unoptimized (left) and optimized (right) circuits. The distillation of dominant eigenproperties approach is indeed robust against gate errors, given that the observable estimates using noisy circuits (dotted and dashed) closely approximate the noise-free circuits (solid lines).

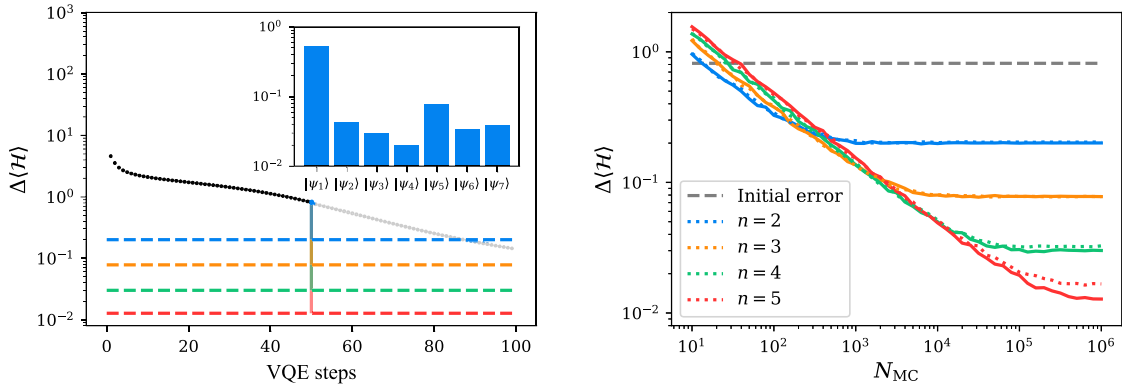


FIG. 7. 6-qubit lattice Schwinger model using variational time evolution. (Left) An initial state is obtained for the distillation of dominant eigenproperties approach by stopping variational quantum eigensolver (black dots) after only 50 gradient descent steps. Applying the distillation of dominant eigenproperties approach (dashed lines) improves exponentially on the estimation error of variational quantum eigensolver as we increase the number of virtual copies. Inset: overlaps  $|\langle \psi_k | \psi(\theta) \rangle|^2$  between the distillation of dominant eigenproperties approach’s initial state (obtained after 50 VQE iterations) and the lowest energy eigenstates  $k \in [1, 7]$ . The initial state has an overlap of  $p_1 = 0.53$  with the ground state and smaller overlaps with excited states. (Right) Mean observable expectation value errors ( $\Delta\langle \mathcal{H} \rangle$ ) for an increasing number of Monte Carlo samples and number of copies. Dotted lines show results obtained from exact, while solid lines show the results from variationally simulated time evolution. The initial deviation from the exact ground state energy is also shown with the dashed horizontal line.

$|\psi_1\rangle$  such that it can be used as an initial state for DDE.

We choose a main time grid defined by  $T = 50$  and  $dt = 0.2$  and apply ansatz-based variational time evolution [44,61] using an integration timestep of  $dt/1000$  to guarantee a sufficiently high fidelity—the fidelity between the variationally time-evolved state and the exactly time-evolved state is above 99.4% for all timesteps (as shown in Fig. 12). For further details on the ansatz-based variational quantum simulation algorithm, we refer the reader to Appendix E 4.

Figure 7(left, dashed lines) shows the observable expectation value errors when DDE is applied on these variationally time-evolved states and confirms that as we increase the number of virtual copies, the VQE energy error can be improved exponentially. Figure 7(right) shows the mean observable expectation value errors for an increasing number of MC samples. Indeed, DDE is robust against algorithmic errors introduced by variational time evolution, given that the observable estimation errors (solid lines) closely approximate the errors obtained using exact time evolution (dotted lines).

DDE in combination with variational techniques seems to be a promising approach for near-term quantum advantage for the following reasons. While VQE is a heuristic approach that may require exponential resources to converge to the ground state, it almost always gets stuck in a local minimum [21], and we observe that these local traps may have sufficient overlap with low-lying eigenstates in numerical simulations. Applying DDE to these locally trapped VQE states is guaranteed to exponentially increase the quality of the observable estimate as long as

efficient time evolution is possible. While variational time evolution can be conveniently combined with VQE without increasing the ansatz circuit depth, these techniques typically require a very large number of shots and fine control of rotation angles [62] to guarantee sufficiently high time evolution fidelity. We also note that while VQE is limited to ground states (or to low-lying eigenstates [17]), our DDE approach can be applied to any eigenstate.

## V. QUANTUM-INSPIRED CLASSICAL SIMULATION

### A. Approach

As detailed in Sec. III E, our approach uses a quantum computer to estimate the correlators  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$ . However, in a broad range of practically relevant scenarios, these correlators can instead be obtained using classical simulation techniques, bypassing the need for a quantum computer. For example, when the initial state can be represented efficiently as an MPS, time evolution can be implemented using the well-established time-evolving block decimation (TEBD) algorithm [63,64], and therefore the desired correlators can be obtained through contractions.

Indeed, tensor networks cannot represent general quantum states efficiently. However, as long as our physical Hamiltonian of interest admits a time evolution of the initial state that can be captured by a family of tensor network states, then DDE can be fully simulated classically. This also means that DDE can improve and broaden the class of classical simulation techniques. For example, while preparing ground states is well-established via, e.g.,

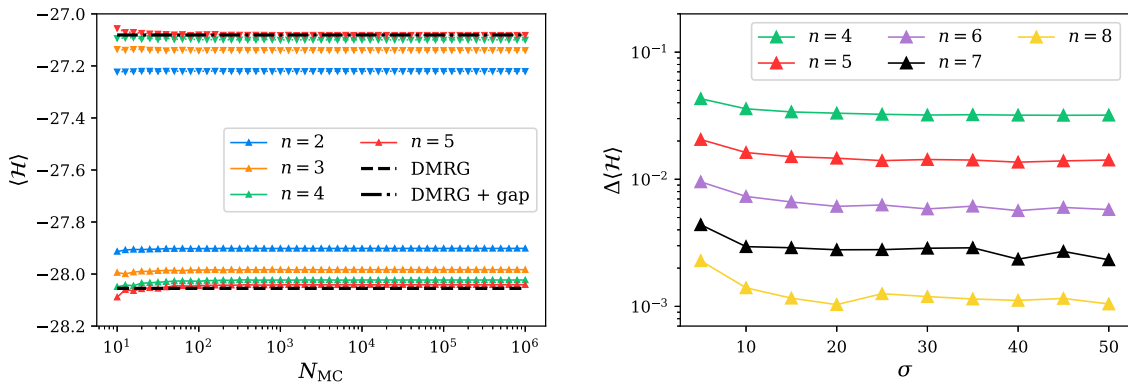


FIG. 8. Matrix product state simulation of a 100-qubit random-field Heisenberg model. (Left) Ground state and excited state energy estimates for an increasing number  $n$  of copies and number of Monte Carlo samples  $N_{MC}$  (triangles) converge to eigenenergy estimates obtained using alternative techniques—density matrix renormalization group and spectroscopy (dashed lines). This confirms the efficacy of the distillation of dominant eigenproperties approach which, however, is not limited to energy estimation and can be applied to the estimation of an arbitrary observable in the excited state—whereas obtaining expected values in excited states is nontrivial using tensor-network techniques. Here we also applied the variance reduction technique detailed in Sec. III C, i.e., the Monte Carlo approach converges for relatively few samples. (Right) Our rigorous error bounds in Theorem 1 guarantee that the estimation error converges superexponentially as we increase the Gaussian time window  $\sigma$ , demonstrated here for an increasing number of copies using a fixed number  $N_{MC} = 10^6$  (hence the nonsmooth curves, and  $N_{MC} = 10^{6.5}$  for  $n = 8$ ). Note that increasing  $\sigma$  increases the depth of time evolution and thus may require an exponential increase in bond-dimension.

the density matrix renormalization group (DMRG) technique [65], preparing excited states and highly excited states is nontrivial using tensor networks.

In contrast, if DDE is supplied with an initial state that has dominant overlap with one of the excited states, then time-evolving this state allows us to estimate observable expectation values in the desired excited state without explicitly preparing it. We demonstrate this application in the following section. Indeed, time evolution may lead to an increase in entanglement of the quantum system and MPS simulation will ultimately be limited to a certain maximal evolution time, which then limits the achievable Gaussian window  $\sigma$ . This leads to a finite approximation error in our bounds in Theorem 1 and further suppressing the simulation error then necessitates a quantum computer—which, however, can efficiently simulate the time evolution of any physical Hamiltonian.

## B. 100-qubit MPS simulation

We demonstrate our quantum-inspired classical DDE algorithm to obtain ground and excited state properties of a 100-qubit Heisenberg chain (with open boundary conditions) in an MPS simulation.

We consider initial states of the form  $|\psi(0)\rangle = c_1 |b_1\rangle + c_2 |b_2\rangle$ , where the amplitudes  $c_1$  and  $c_2$  satisfy  $\{|c_1|^2, |c_2|^2\} = \{0.7, 0.3\}$  and  $\{|c_1|^2, |c_2|^2\} = \{0.3, 0.7\}$  for the ground and excited state demonstrations, respectively.  $|b_1\rangle$  is the dominant bitstring in the DMRG ground state, and  $|b_2\rangle$  is obtained by flipping the single bit in  $|b_1\rangle$  that has the least energy (approximation of a low-lying excited state). We apply TEBD [63,64] to these initial states with a

Trotter error bound  $10^{-4}$  at the maximal evolution time  $T = 200$  using a step size  $dt = 0.5$ . We report the true ground and excited state energies in Fig. 8(left, dashed lines), where the ground state energy was obtained using DMRG [65]. Our estimate of the excited state energy was obtained through spectroscopy [29], whereby we computed the overlaps  $\langle \psi(0) | \psi(t) \rangle$  for all discrete timesteps  $t \in [0, 200]$  whose Fourier transform is the local density of states [66] and thus the frequency of the dominant peak corresponds to the energy of the excited state.

We then use the same time-evolved MPS states to estimate the correlators  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$  and  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$  and apply DDE for estimating the expectation value of the Hamiltonian via  $O = \mathcal{H}$ . Figure 8(left) shows the ground and excited state energies obtained using DDE for an increasing number of copies and for an increasing number of MC samples and confirms that our approach indeed converges to the same energy estimates as DMRG and spectroscopy. Note that DDE is not limited to energy estimation and can be applied to estimate the expected value of any observable  $O$  in the excited state that can be represented efficiently as a matrix product operator—our motivation for estimating the energy is to verify DDE against established techniques.

Furthermore, Fig. 8(left) also nicely demonstrates our MC variance minimization approach detailed in Sec. III C, whereby we applied the transformation  $A(t, t') \rightarrow A(t, t') - cB(t, t')$  via the coefficient  $c = \langle \psi(0) | \mathcal{H} | \psi(0) \rangle$ . This choice nearly minimizes the MC variance and, indeed, the MC estimates in Fig. 8(left) have sufficiently low uncertainty for a relatively low number of samples (please see more details in Appendix F 2).

In Fig. 8(right), we also demonstrate the exponential suppression power for the ground state energy estimation task. This exponential suppression is consistent with our error bound in Theorem 1 and confirms the scalability of our approach to large system sizes well beyond exact diagonalization.

## VI. DISCUSSION AND CONCLUSION

In this work, we introduced a hybrid quantum-classical algorithm, which enables the estimation of eigenstate properties by distilling the dominant contribution of an initial input state. Our DDE approach builds conceptually on purification-based QEM techniques applied to randomly time-evolved quantum states. In contrast to conventional VD, which uses multiple copies, our hybrid algorithm requires only a single quantum register with an ancillary qubit to estimate temporal correlators over a 2D time grid. In the subsequent classical postprocessing part, the exponential suppression power is achieved via a high-dimensional MC integration.

We presented rigorous theoretical guarantees for the algorithm and proved that estimation errors are suppressed exponentially as we increase the number of (virtual) copies and superexponentially as we increase the depth of time evolution (by increasing a Gaussian window width). Furthermore, we formally introduced performance guarantees associated with the MC-based classical postprocessing step using median of means estimators.

We showcased the performance and scalability of our algorithm through an extensive set of numerical experiments in Sec. IV. We studied the robustness of DDE to both algorithmic errors associated with Trotterization and to residual logical errors in quantum gates as relevant in early fault tolerance. Furthermore, we demonstrated the compatibility of DDE with near-term-friendly variational techniques for both initial state preparation and time evolution.

Our approach is naturally compatible with a broad range of initial state preparation and time evolution techniques. For example, while initial state preparation is in general exponentially expensive, for a broad range of practically important problems, good initialization is possible, e.g., by using tensor-network techniques [37,39]. We also demonstrated that VQE could be used to prepare an initial state for DDE. Furthermore, while time evolution is one of the most natural applications of quantum computers, current resource estimates suggest that even relatively simple systems may require deep time evolution circuits. Randomized time evolution formulas can significantly reduce absolute resource requirements for small to medium-scale systems and are indeed compatible with DDE [54,67]. While DDE achieves shallow quantum circuits of depth  $\tilde{O}(\Delta^{-1})$ , its runtime may become impractical when the spectral gap closes, as relevant, e.g., in glassy systems.

This is, however, a general limitation in quantum computing that is consistent with the Quantum Merlin-Arthur (QMA) hardness of eigenstate preparation [40,68]. Still, most candidates for quantum advantage, such as FeMoco, target gapped Hamiltonians and are therefore well suited for our approach. Furthermore, our gap dependence may be reduced significantly by applying quadratic Hamiltonian factorization, which was originally introduced in the context of adiabatic computation, and recently applied to sum of squares simulation of the FeMoco Hamiltonian [69–71].

Another promising direction may be to use analog quantum simulation as part of DDE, i.e., applying analog state preparation with low resource requirements (e.g., using adiabatic evolution) and then implementing analog time evolution. Correlators estimated from the analog quantum simulation experiments would then allow for a better estimate of the expectation value than with the simulation technique alone.

Finally, we also demonstrated that our approach is viable as a quantum-inspired classical simulation technique. We demonstrated that DDE can be effectively combined with state-of-the-art classical tensor-network methods to estimate both ground and excited state properties, thereby complementing the existing toolbox of tensor-network simulation techniques.

## ACKNOWLEDGMENTS

The numerical simulation of quantum circuits made use of the Quantum Exact Simulation Toolkit (QuEST) [72] via the QuESTlink [73] frontend, while tensor network simulations were performed with the quimb Python package [74]. Quantum circuits were illustrated using the Quantikz LaTeX package [75]. B.B. would like to thank the support of the Hungarian National Research, Development and Innovation Office (NKFIH) through the KDP-2023 funding scheme (Grant No. C2245275), and the TKP2021-NVA funding scheme (grant number TKP2021-NVA-29). B.B. also acknowledges the computational resources provided by the Wigner Scientific Computational Laboratory (WSCLAB). T.A. acknowledges support by the Oxford-Uehiro Graduate Scholarship Programme. T.A. also acknowledges the use of the University of Oxford Advanced Research Computing (ARC) facility [76] in carrying out this work. B.K. thanks UKRI for the Future Leaders Fellowship Theory to Enable Practical Quantum Advantage (MR/Y015843/1). The authors also acknowledge funding from the EPSRC projects Robust and Reliable Quantum Computing (RoARQ, EP/W032635/1) and Software Enabling Early Quantum Advantage (SEEQA, EP/Y004655/1). B.K. thanks the University of Oxford for a Glasstone Research Fellowship and Lady Margaret Hall, Oxford for a Research Fellowship. This research was funded in part by UKRI (MR/Y015843/1). For the purpose

of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

### DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

### APPENDIX A: RELATED WORKS

Algorithmic cooling techniques [31,32] rely on quantum computation of the same two-time correlations as in our work, and classical computation of similar integrals through MC sampling to estimate eigenstate properties. While these prior works rely on imaginary time evolution, DDE is based on real-time evolution and provably converges superexponentially to a mixed state to which we apply virtual distillation—which latter step provides a provably exponential error suppression power with respect to the number of virtual copies. These methods also rely on prior knowledge of the target eigenstate energy to alleviate the sampling overhead, similarly to the technique that we discuss in Sec. III C. While we choose to sample random time evolutions from a Gaussian distribution, these works have considered alternative distributions, which may also prove useful in the context of DDE.

Our work is also closely related to quantum virtual cooling [77], which inputs a thermal state and obtains ground state properties via virtual purification. This is similar to applying DDE with the fully quantum implementation of controlled derangement, rather than offloading the computational burden to MC-based classical postprocessing, as discussed in Sec. III A. In our work, the thermal state is replaced by an average of randomly time-evolved states, which leads to the following advantages. First, we reduce the requirement of the number of quantum registers from  $n$  to 1. Second, we enable targeting higher excited states by appropriately choosing the initial state. Third, we provide an additional flexibility in how to perform initial state preparation and classical virtual purification.

Our approach shares similarities with statistical phase estimation techniques, i.e., it similarly requires only a single ancilla qubit and relies on estimating similar objects for subsequent classical postprocessing. While a range of statistical phase estimation techniques are known that make a trade-off between maximal circuit depth and low number of shots, remarkably, the circuit depth (maximum evolution time) of DDE matches the optimal  $\tilde{\mathcal{O}}(\Delta^{-1})$  scaling of the algorithms introduced in Refs. [26,27] for energy estimation.

On the other hand, Refs. [26,27] report a scaling of  $\tilde{\mathcal{O}}(p_q^{-2}\epsilon^{-2})$  with respect to the number of shots but require prior knowledge of a lower bound on the spectral gap  $\Delta$  and the initial overlap  $p_q$ . While DDE can operate under more relaxed assumptions, it requires an increased number of shots  $\tilde{\mathcal{O}}(p_q^{-2\ln(|\mathcal{Q}|^{-1})}\epsilon^{-2})$ . We argue below that (a) the increased number of shots is inevitable, given DDE is fundamentally more general, and (b) that we consider practically important early fault-tolerant settings where  $p_q \approx 1$ , which then results in a reasonable number of shots as demonstrated in our numerical experiments. As a side-note, we also remark that the approach of Ref. [25] reaches Heisenberg scaling in the total runtime  $\tilde{\mathcal{O}}(p_q^{-2}\epsilon^{-1})$ , but it requires deeper circuits given its  $\tilde{\mathcal{O}}(\epsilon^{-1})$  scaling.

Let us now illustrate why estimating an arbitrary expectation value in an eigenstate is a more complex task than estimating eigenenergies given access to an initial state [40]. Estimating eigenenergies with textbook phase estimation has a complexity of  $\mathcal{O}(p_q^{-2}\epsilon^{-1})$  [25, 78,79]. In contrast, using phase estimation to prepare an eigenstate and then sampling to estimate an arbitrary expected value would require a significantly increased runtime  $\mathcal{O}(p_q^{-2}\epsilon^{-3})$ . Remarkably, our sampling-based DDE matches the complexity of phase estimation combined with amplitude estimation [80] in its scaling with respect to the target precision, as the latter has a complexity of  $\mathcal{O}(p_q^{-2}\epsilon^{-2})$ . We have an increased complexity with respect to the overlap in the initial state  $\tilde{\mathcal{O}}(p_q^{-2\ln(|\mathcal{Q}|^{-1})})$ . However, as we detailed in Sec. I, the most promising candidates for achieving early quantum advantages are problems where reasonably good initial states are known with  $p_q \approx 1$ , thanks to decades of classical advancements. In these cases, quantum computing is necessary to improve the precision to a practically useful level, as exemplified by FeMoco ( $p_q \approx 0.9$ ).

### APPENDIX B: PROOFS

#### 1. Proof of Lemma 1

*Proof.* Explicitly, the time-evolved state is

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)| = \sum_{j,k} c_j c_k^* e^{-it(E_j - E_k)} |\psi_j\rangle\langle\psi_k|. \quad (\text{B1})$$

We can substitute this back into Eq. (1),

$$\begin{aligned} \bar{\rho} &= \int_{-\infty}^{\infty} G(t)\rho(t) dt \\ &= \sum_{j,k} c_j c_k^* |\psi_j\rangle\langle\psi_k| \int_{-\infty}^{\infty} G(t)e^{-it(E_j - E_k)} dt. \end{aligned} \quad (\text{B2})$$

The integral above is the Fourier transform of the Gaussian distribution as

$$\int_{-\infty}^{\infty} G(t) e^{-it(E_j - E_k)} dt = e^{-\frac{(E_j - E_k)^2 \sigma^2}{2}}. \quad (\text{B3})$$

Substituting this back yields

$$\begin{aligned} \bar{\rho} &= \sum_{j,k} c_j c_k^* e^{-\frac{(E_j - E_k)^2 \sigma^2}{2}} |\psi_j\rangle \langle \psi_k| \\ &= \underbrace{\sum_k |c_k|^2 |\psi_k\rangle \langle \psi_k|}_{\rho} + \underbrace{\sum_{j \neq k} c_j c_k^* e^{-\frac{(E_j - E_k)^2 \sigma^2}{2}} |\psi_j\rangle \langle \psi_k|}_{\mathcal{E}}. \end{aligned} \quad (\text{B4})$$

Let us now bound the Hilbert-Schmidt distance  $\|\mathcal{E}\|_{\text{HS}} = \|\bar{\rho} - \rho\|_{\text{HS}}$ , exploiting the fact that the squared coefficients  $p_k = |c_k|^2$  form a probability distribution,

$$\begin{aligned} \|\mathcal{E}\|_{\text{HS}}^2 &= \left\| \sum_{j \neq k} c_j c_k^* e^{-\frac{(E_j - E_k)^2 \sigma^2}{2}} |\psi_j\rangle \langle \psi_k| \right\|_{\text{HS}}^2 \\ &= \sum_{j \neq k} |c_j|^2 |c_k|^2 \left( e^{-\frac{(E_j - E_k)^2 \sigma^2}{2}} \right)^2 \\ &\leq e^{-\Delta^2 \sigma^2} \sum_{j \neq k} |c_j|^2 |c_k|^2 \\ &= e^{-\Delta^2 \sigma^2} \left( \sum_{j,k} p_j p_k - \sum_k p_k^2 \right) \\ &= e^{-\Delta^2 \sigma^2} (1 - e^{-H_2(\mathbf{p})}) \\ &\leq e^{-\Delta^2 \sigma^2} \left( 1 - (\max_k p_k)^2 \right). \end{aligned} \quad (\text{B5})$$

Above we introduced the smallest gap in the spectrum of the Hamiltonian with nonzero contribution in the initial state, i.e.,  $\Delta = \min\{|E_j - E_k| \text{ with } j, k : |c_j|, |c_k| > 0\}$ .

Next, we bound  $\|\mathcal{E}\|_1$ , the trace distance between the average state and the diagonal state. This can be written as the sum of vector norms

$$\|\mathcal{E}\|_1 = \sum_l \|\mathcal{E} |x^l\rangle\|_2. \quad (\text{B6})$$

Here  $|x^l\rangle$  is the  $l$ th eigenvector of  $\mathcal{E}$  with eigenvalue  $\lambda_l$ , therefore, due to the eigenvalue equation, we have  $\|\mathcal{E} |x^l\rangle\|_2 = |\lambda_l|$ .

With the notation  $g_{jk} = \exp(-(E_j - E_k)^2 \sigma^2 / 2)$ , each term of the sum in Eq. (B6) can be written as

$$\begin{aligned} \|\mathcal{E} |x^l\rangle\|_2^2 &= \sum_j \left| \sum_{k \neq j} c_j c_k^* g_{jk} x_k^l \right|^2 \\ &= \sum_j |c_j|^2 \left( \sum_{k \neq j} c_k^* g_{jk} x_k^l \right) \left( \sum_{k' \neq j} c_{k'} g_{jk'}^* (x_k^l)^* \right) \\ &= \sum_j |c_j|^2 |\langle x^l | W_j | \mathbf{c} \rangle|^2, \end{aligned} \quad (\text{B7})$$

where  $\mathbf{c} = (c_1, c_2, \dots)$ , and we introduced the diagonal matrices  $W_j = \text{diag}(g_{j1}, g_{j2}, \dots)$ . With this, the trace norm of the error term can be bounded as

$$\begin{aligned} \|\mathcal{E}\|_1 &= \sum_l \sqrt{\sum_j |c_j|^2 |\langle x^l | W_j | \mathbf{c} \rangle|^2} \\ &\leq e^{-\Delta^2 \sigma^2 / 2} \sum_l |\langle x^l | \mathbf{c} \rangle|. \end{aligned} \quad (\text{B8})$$

Similarly to the case of the Hilbert-Schmidt norm, here we also used the fact that  $(W_j)_{kk} = g_{jk} \leq \exp(-\Delta^2 \sigma^2 / 2)$ . The factor  $\sum_l |\langle x^l | \mathbf{c} \rangle| = \|\mathbf{c}\|_1$  is the  $\ell_1$  norm of the coefficient vector  $\mathbf{c}$  in the eigenbasis of the residual operator  $\mathcal{E}$ . The vector norm  $\|\mathbf{c}\|_2 = 1$  is bounded and basis-independent, but the  $\ell_1$  norm is basis-dependent and therefore can in principle be as large as  $\sqrt{2^N}$ . However, our numerical demonstrations point to the fact that it remains within a reasonable scale in practice. This shows that the trace norm of the error term also vanishes superexponentially with the standard deviation of the normal distribution. ■

## 2. Proof of Theorem 1

*Proof.* The quantity of interest is

$$|\mathcal{Q}| = \left| \frac{\text{tr}[\bar{\rho}^n O]}{\text{tr}[\bar{\rho}^n]} - \langle \psi_q | O | \psi_q \rangle \right|, \quad (\text{B9})$$

which can be bounded as

$$|\mathcal{Q}| \leq \left| \frac{\text{tr}[\bar{\rho}^n O]}{\text{tr}[\bar{\rho}^n]} - \frac{\text{tr}[\rho^n O]}{\text{tr}[\rho^n]} \right| + \left| \frac{\text{tr}[\rho^n O]}{\text{tr}[\rho^n]} - \langle \psi_q | O | \psi_q \rangle \right| \quad (\text{B10})$$

by the triangle inequality. The second term is simply the error magnitude  $|\mathcal{A}|$  from Lemma 2. To bound the first term, we use the inequality from Ref. [47] as

$$\left| \frac{\text{tr}[\bar{\rho}^n O]}{\text{tr}[\bar{\rho}^n]} - \frac{\text{tr}[\rho^n O]}{\text{tr}[\rho^n]} \right| \leq 2 \|O\|_{\infty} \left\| \frac{\bar{\rho}^n}{\text{tr}[\bar{\rho}^n]} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1. \quad (\text{B11})$$

Denoting deviations  $\Omega$  and  $\omega$  such that  $\bar{\rho}^n = \rho^n + \Omega$  and  $\text{tr}[\bar{\rho}^n] = \text{tr}[\rho^n] + \omega$ , a further use of the triangle inequality

leads to

$$\begin{aligned} \left\| \frac{\bar{\rho}^n}{\text{tr}[\bar{\rho}^n]} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 &= \left\| \frac{\rho^n + \Omega}{\text{tr}[\rho^n] + \omega} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 \\ &\leq \left\| \frac{\rho^n}{\text{tr}[\rho^n] + \omega} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 + \left\| \frac{\Omega}{\text{tr}[\rho^n] + \omega} \right\|_1. \end{aligned} \quad (\text{B12})$$

Both of these terms can be treated using the Maclaurin series expansion

$$\frac{1}{\text{tr}[\rho^n] + \omega} = \sum_{k=0}^{\infty} (-1)^k \frac{\omega^k}{\text{tr}[\rho^n]^{k+1}}. \quad (\text{B13})$$

For the first term in Eq. (B12), applying the triangle inequality to the expansion gives

$$\begin{aligned} \left\| \frac{\rho^n}{\text{tr}[\rho^n] + \omega} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 &\leq \left\| \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 \sum_{k=1}^{\infty} \left| \frac{\omega}{\text{tr}[\rho^n]} \right|^k \\ &\leq \sum_{k=1}^{\infty} \left( \frac{\|\Omega\|_1}{p_q^n} \right)^k, \end{aligned} \quad (\text{B14})$$

where we also used the fact that  $\omega = \text{tr}[\Omega]$  implies  $|\omega| \leq \|\Omega\|_1$ . The second term in Eq. (B12) can be bounded similarly as

$$\left\| \frac{\Omega}{\text{tr}[\rho^n] + \omega} \right\|_1 \leq \sum_{k=0}^{\infty} \left\| \frac{\omega^k \Omega}{\text{tr}[\rho^n]^{k+1}} \right\|_1 \leq \sum_{k=1}^{\infty} \left( \frac{\|\Omega\|_1}{p_q^n} \right)^k. \quad (\text{B15})$$

Thus, both terms are bounded by the same infinite series with  $\|\Omega\|_1/p_q^n$  as the common ratio. We proceed by expanding  $\bar{\rho}^n = (\rho + \mathcal{E})^n$ , which leads to

$$\begin{aligned} \|\Omega\|_1 &= \left\| \sum_{k=0}^{n-1} \rho^k \mathcal{E} \rho^{n-1-k} + \sum_{\substack{k,l=0 \\ l+k \leq n-2}}^{n-2} \rho^k \mathcal{E} \rho^l \mathcal{E} \rho^{n-2-k-l} \right. \\ &\quad \left. + \dots + \sum_{k=0}^{n-1} \mathcal{E}^k \rho \mathcal{E}^{n-1-k} + \mathcal{E}^n \right\|_1 \\ &\leq \sum_{k=0}^{n-1} \left\| \rho^k \mathcal{E} \rho^{n-1-k} \right\|_1 + \sum_{\substack{k,l=0 \\ l+k \leq n-2}}^{n-2} \left\| \rho^k \mathcal{E} \rho^l \mathcal{E} \rho^{n-2-k-l} \right\|_1 \\ &\quad + \dots + \sum_{k=0}^{n-1} \left\| \mathcal{E}^k \rho \mathcal{E}^{n-1-k} \right\|_1 + \|\mathcal{E}^n\|_1 \\ &\leq \sum_{k=1}^n \binom{n}{k} p_q^{n-k} \|\mathcal{E}\|_1^k. \end{aligned} \quad (\text{B16})$$

The first inequality follows from the triangle inequality. The second inequality applies Hölder's inequality,

the monotonicity of Schatten norms, and the fact that  $\|\rho^k\|_{\infty} = p_q^k$  for all  $k$ . This gives an upper bound on the base of the infinite series in Eq. (B14) and Eq. (B15) as

$$\frac{\|\Omega\|_1}{p_q^n} \leq \sum_{k=1}^n \binom{n}{k} p_q^{-k} \|\mathcal{E}\|_1^k = (1 + \|\mathcal{E}\|_1/p_q)^n - 1, \quad (\text{B17})$$

so that Eq. (B12) implies

$$\begin{aligned} \left\| \frac{\bar{\rho}^n}{\text{tr}[\bar{\rho}^n]} - \frac{\rho^n}{\text{tr}[\rho^n]} \right\|_1 &\leq 2 \sum_{k=1}^{\infty} \left( \frac{\|\Omega\|_1}{p_q^n} \right)^k \\ &\leq 2np_q^{-1} \|\mathcal{E}\|_1 + \mathcal{O}(\|\mathcal{E}\|_1^2). \end{aligned} \quad (\text{B18})$$

Piecing all the bounds together through Eqs. (B10) and (B11), we finally obtain

$$|\mathcal{Q}| \leq |\mathcal{A}| + 4np_q^{-1} \|O\|_{\infty} \|\mathcal{E}\|_1 + \mathcal{O}(\|\mathcal{E}\|_1^2). \quad (\text{B19})$$

Note that the above expressions assume that we are interested in the regime of vanishing  $\|\mathcal{E}\|_1$ , but higher-order terms in the upper bound can also be written out exactly based on Eq. (B17). While these higher-order terms involve binomial coefficients, the bound

$$\sum_{k=1}^n \binom{n}{k} p_q^{-k} \|\mathcal{E}\|_1^k < \sum_{k=1}^n \left( \frac{en}{p_q} \|\mathcal{E}\|_1 \right)^k \quad (\text{B20})$$

makes it clear that  $\|\mathcal{E}\|_1 < p_q/(en)$  serves as a sufficient condition for the first-order term to be the leading term, where  $e$  is the Euler number.

Finally, let us use the bound from Eq. (B19) to asymptotically estimate the resource requirements. Without loss of generality, we assume

$$|\mathcal{Q}|/2 \leq 2(p_q^{-1} - 1)^n e^{-(n-1)H_n(\mathbf{p}')}. \quad (\text{B21})$$

After using the monotonicity of Rényi entropies and rearranging the inequality, we get

$$n \leq \frac{\ln[4|\mathcal{Q}|^{-1}] + H_{\infty}(\mathbf{p}')}{\ln[(p_q^{-1} - 1)^{-1}] + H_{\infty}(\mathbf{p}')}. \quad (\text{B22})$$

From this, we get the scaling  $n = \mathcal{O}(\ln[|\mathcal{Q}|^{-1}])$ .

Similarly, for the second term,

$$|\mathcal{Q}|/2 \lesssim 4np_q^{-1} e^{-\frac{\Delta^2 \sigma^2}{2}} \|c\|_1 \|O\|_{\infty}. \quad (\text{B23})$$

By rearranging this, we get

$$\sigma \lesssim \frac{\sqrt{2 \ln[|\mathcal{Q}|^{-1}n] + 2 \ln[8\|c\|_1 \|O\|_{\infty}]}}{\Delta}. \quad (\text{B24})$$

After Taylor expansion, we get the final scaling  $\sigma = \mathcal{O}(\Delta^{-1} \sqrt{\ln[|\mathcal{Q}|^{-1}n]})$ .  $\blacksquare$

### 3. Proof of sampling overhead

Consider applying DDE via direct estimation of nonlinear functionals as discussed in Sec. III A. The associated sampling overhead scales as  $\mathcal{O}(\text{tr}[\bar{\rho}^n]^{-2})$  [35,36], which is upper bounded as

$$\begin{aligned} \text{tr}[\bar{\rho}^n]^{-2} &= \text{tr}[\rho^n]^{-2} \sum_{k=0}^{\infty} (-1)^k (k+1) \left( \frac{\omega}{\text{tr}[\rho^n]} \right)^k \\ &\leq \text{tr}[\rho^n]^{-2} \sum_{k=0}^{\infty} (k+1) \left( \frac{\|\Omega\|_1}{\text{tr}[\rho^n]} \right)^k \\ &\leq p_q^{-2n} \left( 1 + 2np_q^{-1} \|\mathcal{E}\|_1 + \mathcal{O}(\|\mathcal{E}\|_1^2) \right). \end{aligned} \quad (\text{B25})$$

We used the Maclaurin series expansion for the first equality. The triangle inequality and the fact that  $|\omega| \leq \|\Omega\|_1$  are applied in the second step. The fact that  $\text{tr}[\rho^n] = \sum_k p_k^n \geq p_q^n$  and Eq. (B17) are applied in the third step.

Using the bound on the number of copies from Eq. (B22), in leading terms, the sampling overhead can be written as  $\tilde{\mathcal{O}}(p_q^{-2 \ln(|Q|^{-1})} \epsilon^{-2})$ , where  $\epsilon$  is the standard statistical uncertainty. An important observation here is that the exponent depends only logarithmically on  $|Q|^{-1}$ , meaning that the complexity will be effectively polynomial in the inverse systematic bias.

### 4. Proof of Proposition 1

*Proof.*

$$\begin{aligned} \bar{\rho}^n &= \left( \int_{-\infty}^{\infty} G(t) \rho(t) dt \right)^n \\ &= \int G(t_1) \rho(t_1) G(t_2) \rho(t_2) \cdots G(t_n) \rho(t_n) dt_1 \cdots dt_n \\ &= \int \left( \prod_{k=1}^n G(t_k) \right) \left( \prod_{k=1}^n \rho(t_k) \right) dt_1 \cdots dt_n, \end{aligned} \quad (\text{B26})$$

thus

$$\text{tr}[\bar{\rho}^n O] = \int \left( \prod_{k=1}^n G(t_k) \right) \text{tr}[\rho(t_1) \cdots \rho(t_n) O] dt_1 \cdots dt_n. \quad (\text{B27})$$

Observing that the product in the expression above is a multivariate Gaussian distribution  $\prod_{k=1}^n G(t_k) = G(\mathbf{t})$  and introducing the notation for the trace as  $F(\mathbf{t}) = \text{tr}[\rho(t_1) \rho(t_2) \cdots \rho(t_n) O]$  we can compactly write the integral as

$$\text{tr}[\bar{\rho}^n O] = \int G(\mathbf{t}) F(\mathbf{t}) d\mathbf{t}. \quad (\text{B28})$$

We can significantly simplify  $F(\mathbf{t})$  by exploiting that  $\rho(t_k) = |\psi(t_k)\rangle\langle\psi(t_k)|$  is a projection operator and using

the cyclic property of the trace as

$$\begin{aligned} F(\mathbf{t}) &= \langle\psi(t_n)| O |\psi(t_1)\rangle \prod_{k=1}^{n-1} \langle\psi(t_k)| \psi(t_{k+1})\rangle \\ &= A(t_n, t_1) \prod_{k=1}^{n-1} B(t_k, t_{k+1}). \end{aligned} \quad (\text{B29})$$

■

### 5. Proof of Theorem 2

*Proof.* The variance of the function  $F(\mathbf{t})$  can be bounded as

$$\text{Var}(F) = \mathbb{E}[F^2] - \mathbb{E}[F]^2 \leq \mathbb{E}[F^2] \leq 1, \quad (\text{B30})$$

where in the last inequality we used that  $F^2(\mathbf{t}) \leq 1$  as

$$\int G(\mathbf{t}) F^2(\mathbf{t}) d\mathbf{t} \leq \int G(\mathbf{t}) d\mathbf{t} = 1. \quad (\text{B31})$$

Let  $\hat{F}$  denote the MC empirical mean estimator for the numerator using  $N_{\text{batch}}$  samples. With this, the variance of the estimator can be expressed as

$$\text{Var}(\hat{F}) = \frac{\text{Var}(F)}{N_{\text{batch}}} \leq \frac{1}{N_{\text{batch}}}. \quad (\text{B32})$$

A similar bound can be proven for the denominator, which is contained as a special case.

We can repeat the empirical MC estimation  $K$  times for both  $F$  and  $J$  to obtain  $K$  independent estimates of the numerator and denominator. In practice, this is done by obtaining  $2KN_{\text{batch}}$  independent samples, dividing them into  $2K$  batches and computing the empirical mean for each batch. By taking the median over these means and computing their ratio, we get an estimate with rigorous performance guarantees, as detailed below.

With failure probability  $\delta = e^{-K/8}$ , the median of means is within  $\xi$  distance from the ideal mean, where

$$\xi = 2\sqrt{\text{Var}(\hat{F})} \leq 2N_{\text{batch}}^{-1/2}. \quad (\text{B33})$$

We thus have a median of means estimate for both  $F$  and  $J$  with a guarantee  $(\xi, \delta)$ . Equivalently, for a fixed failure probability  $\delta$  and precision  $\xi$ , we get the required number of batches  $K = -8 \log \delta$  and batch size  $N_{\text{batch}} = 4/\xi^2$ .

In the following, let us use the notation  $f = \mathbb{E}[F]$  and  $j = \mathbb{E}[J]$ . As such, if we now take the ratio of  $F$  and  $J$

with high probability, we guarantee that the error will be

$$\begin{aligned} \epsilon &\leq \left| \frac{f + \xi}{j - \xi} - \frac{f}{j} \right| = \sum_{i=1}^{\infty} \frac{f + j}{j^{i+1}} \xi^i, \quad \text{if } f \geq 0, \\ \epsilon &\leq \left| \frac{f - \xi}{j - \xi} - \frac{f}{j} \right| = \sum_{i=1}^{\infty} \frac{-f + j}{j^{i+1}} \xi^i, \quad \text{if } f < 0. \end{aligned} \quad (\text{B34})$$

Here we used the Taylor expansion of the ratio in the worst case. Since  $j$  is always non-negative, only these two cases need to be considered, which can be written in a single bound as

$$\epsilon \leq \left| \frac{|f| + \xi}{j - \xi} - \frac{|f|}{j} \right| = \sum_{i=1}^{\infty} \frac{|f| + j}{j^{i+1}} \xi^i. \quad (\text{B35})$$

We can now substitute  $\xi$  from Eq. (B33) and obtain a first-order estimate as

$$\epsilon \leq 2 \frac{|f| + j}{j^2} N_{\text{batch}}^{-1/2} + \mathcal{O}(N_{\text{batch}}^{-1}). \quad (\text{B36})$$

Thus, the ratio of the two median estimators has guarantee  $(\epsilon, \delta)$ . Inverting Eq. (B36) in the limit of large  $N_{\text{batch}}$ , as practically relevant in classical postprocessing, we obtain

$$N_{\text{batch}} \lesssim 4 \frac{(|f| + j)^2}{j^4} \epsilon^{-2}. \quad (\text{B37})$$

This gives an estimate that is  $\epsilon$ -close to the ideal ratio with high probability, i.e.,

$$P(|\hat{\mu}_{K,b}(F)/\hat{\mu}_{K,b}(J) - \mu(F/J)| \leq \epsilon) \geq 1 - \delta. \quad (\text{B38})$$

The batch sample size  $N_{\text{batch}}$  depends on the expected value  $|\mathbb{E}[F]| = |\text{tr}[\bar{\rho}^n O]|$  and can be minimized in the special case where  $\text{tr}[\bar{\rho}^n O] \rightarrow 0$ . In this vanishing limit, the bound simplifies to

$$N_{\text{batch}} \lesssim 4 \text{tr}[\bar{\rho}^n]^{-2} \epsilon^{-2}. \quad (\text{B39})$$

In this case, we get the same sampling overhead, as in the Lemma 3, therefore, in leading terms, this sample complexity can be written as  $N_{\text{batch}} = \mathcal{O}(p_q^{-2 \ln(|Q|^{-1})} \epsilon^{-2})$  ■

### APPENDIX C: QUADRATURE INTEGRATION

Quadrature formulas can be used to approximate the integral. While it may seem natural to use Gauss-Hermite quadrature, that provides an exact formula to approximate integrals over the whole real line, it is not reliable in practice [51]. Instead, we can truncate the limits to some sufficiently high number and apply Gauss-Legendre, Clenshaw-Curtis, or Riemann quadrature formulas on the

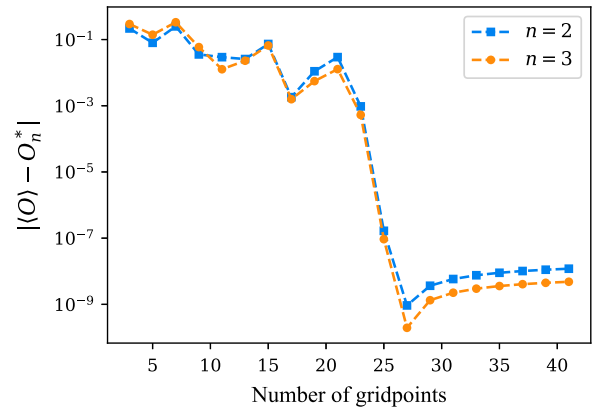


FIG. 9. Deviation from the exactly achievable expectation values  $O_n^*$  for  $n = 2$  and  $n = 3$  copies. The integrals for estimating the expectation value  $\langle O \rangle$  are computed using the simple Riemann rule.

truncated interval. With these quadrature formulas, the multidimensional integral in Lemma 2 is approximated as

$$\begin{aligned} \text{tr}[\bar{\rho}^n O] &= \int_{-\infty}^{\infty} G(\mathbf{t}) F(\mathbf{t}) \, d\mathbf{t} \\ &\approx \int_{-T}^T G(\mathbf{t}) F(\mathbf{t}) \, d\mathbf{t} \\ &\approx \sum_{t_1 \dots t_n} A(t_n, t_1) \prod_{k=1}^{n-1} w_{t_k}^2 G(t_k)^2 B(t_k, t_{k+1}), \end{aligned} \quad (\text{C1})$$

where the sum is taken over all combinations of time values  $t_k$ . Therefore, this computation is exponential in the number of copies, besides potentially introducing numerical instabilities.

In Fig. 9, we show the scaling of the error in the number of gridpoints available, using the simple Riemann rule. Here, the error is defined as the deviation from the analytically achievable estimate with the given number of copies  $O_n^*$ . For this demonstration, we consider the same 10-qubit random-field Heisenberg model as in Sec. IV A and estimate the observable  $O = Z_1$ . The integrals for the numerator and denominator are approximated separately using the Riemann rule, and the estimate  $\langle O \rangle$  is computed by taking their ratio.

### APPENDIX D: OPTIMIZED QUANTUM PROCESSING IN THE HYBRID QUANTUM-CLASSICAL ALGORITHM

An unoptimized method to obtain the  $A$  and  $B$  objects over a 2D grid was presented in Sec. III E, but characteristics of the  $A$  and  $B$  objects enable significant reductions in the computational resources, including the total number of distinct quantum circuit evaluations and the durations of time evolutions involved in each evaluation.

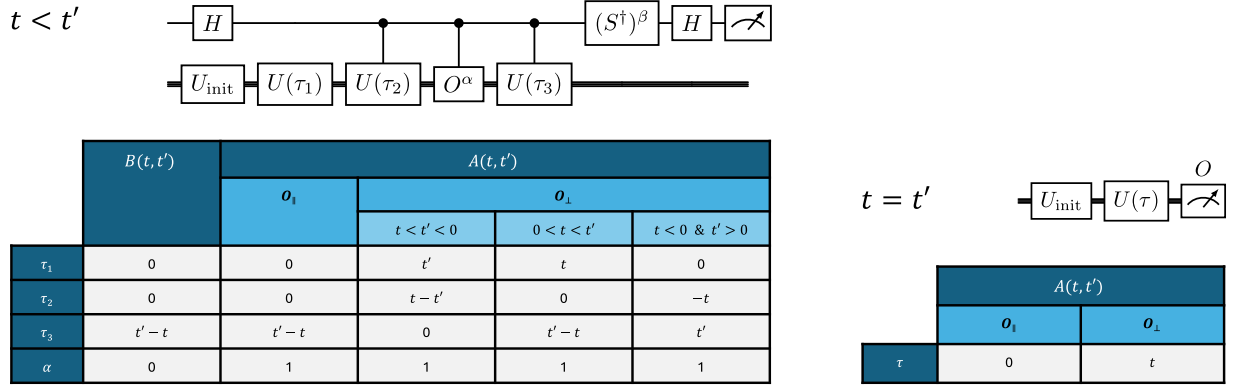


FIG. 10. Quantum circuits involved in the optimized quantum computation step, each accompanied with a table of parameter values for different cases. (Left) A Hadamard test used to obtain all nontrivial  $A(t, t')$  and  $B(t, t')$  objects with  $t < t'$ . The expectation value of the measurement on the ancilla qubit gives the real (imaginary) part of the chosen two-time correlator when the  $S^\dagger$  gate is applied (not applied), corresponding to  $\beta = 1$  ( $\beta = 0$ ). The observable  $O$  can be applied by decomposing it into a term that commutes ( $O_\parallel$ ) or does not commute ( $O_\perp$ ) with the Hamiltonian. (right) A direct time evolution followed by a measurement of  $O$  for the remaining case of  $t = t'$  for  $A(t, t')$ . In practice, the time-evolved state can be measured independently in all of the Pauli-basis terms of  $O$  and classically combined. The duration  $\tau = t$  for  $O_\perp$ , but no time evolution is needed (i.e.,  $\tau = 0$  for all  $t$ ) for  $O_\parallel$ .

Such improvements lead to a lower susceptibility to algorithmic and hardware errors, and we demonstrate this by considering early fault-tolerant devices in Sec. IV B 2. In the following, we describe how such improvements can be made and summarize the resulting circuits in Fig. 10. Note that, while ancilla-free implementations can be applied for problems with symmetries [81–84], including spin models with particle-number conservation considered in this work, here we focus on a general scenario and optimize the Hadamard test.

First, consider obtaining  $B(t, t') = \langle \psi(t) | \psi(t') \rangle$ . Since  $B(t, t) = 1 \ \forall t$ , we do not need to compute the diagonal of the matrix and can assume that  $t \neq t'$ . Furthermore, we only need to compute the case where  $t' > t$  (i.e., the lower triangular part), since the other half can be obtained trivially by taking the complex conjugate as  $B(t', t) = B(t, t')^*$ . Finally,  $\langle \psi(t) | \psi(t') \rangle = \langle \psi(0) | U(t' - t) | \psi(0) \rangle$  implies that we only need to evaluate  $B(t, t')$  once for each fixed  $t' - t$ , and the total duration associated with the Hadamard test is shorter than the unoptimized construction unless  $t' > 0$  and  $t < 0$ , in which case they are both equal to  $|t| + |t'|$ . Utilizing these symmetries, only  $N_T - 1$  entries are sufficient to resolve the whole  $B$  matrix.

Next, consider obtaining  $A(t, t') = \langle \psi(t) | O | \psi(t') \rangle$ . As in the case of  $B(t, t')$ , we only need to obtain the lower half triangle since  $A(t', t) = A(t, t')^*$ . The diagonal,  $A(t, t)$ , can be obtained without introducing ancilla qubits, simply by applying  $U(t)U_{\text{init}}$  to  $|0\rangle$  and measuring the observable as shown in Fig. 10(right). To obtain the remaining entries, suppose we know a decomposition of  $O$  as  $O = O_\parallel + O_\perp$ , where  $[O_\parallel, \mathcal{H}] = 0$  and  $[O_\perp, \mathcal{H}] \neq 0$ . The evaluation of  $A_\parallel(t, t') = \langle \psi(t) | O_\parallel | \psi(t') \rangle$  can be simplified further compared to  $A_\perp(t, t') = \langle \psi(t) | O_\perp | \psi(t') \rangle$ , and the

desired quantity can be obtained by combining them as  $A(t, t') = A_\parallel(t, t') + A_\perp(t, t')$  for all  $t' > t$ . Note that  $O = O_\parallel$  in the important case where  $[O, \mathcal{H}] = 0$  (i.e., when considering symmetries). As before, each of  $O_\parallel$  and  $O_\perp$  can be treated, e.g., by decomposing it into its Pauli basis or by block encoding.

Due to commutativity,  $A_\parallel(t, t') = \langle \psi(0) | O_\parallel U(t' - t) | \psi(0) \rangle$ , so the evaluation of  $A_\parallel(t, t')$  is similar to that of  $B(t, t')$ . Particularly, the same argument as for  $B$  implies that only  $N_T$  entries need to be evaluated. The one additional entry needed in this case corresponds to a diagonal entry, which can be obtained from the ancilla-free circuit shown in Fig. 10(right). Moreover, this ancilla-free circuit does not require any time evolution, since commutativity implies that  $A_\parallel(t, t) = \langle \psi(0) | O_\parallel | \psi(0) \rangle$ . For the remaining case of  $t' > t$ , the quantum circuit used is similar to that of  $B$ , but includes the additional controlled  $O_\parallel$  operation.

The same argument does not hold for  $A_\perp(t, t')$ , and all  $\frac{1}{2}N_T(N_T + 1)$  entries of the matrix must be obtained. However, the total duration of the controlled time evolutions can still be reduced when  $t$  and  $t'$  have the same sign. When  $0 < t < t'$ , for example,  $A_\perp(t, t') = \langle \psi(t) | O_\perp U(t' - t) | \psi(t) \rangle$ , such that a time evolution of duration  $t$  can be performed without being controlled by the ancilla qubit, and the controlled time evolution has a duration given by the time difference  $t' - t$ . The same holds when  $t < t' < 0$ , such that the total duration of the time evolutions in each of these cases is  $\max\{|t|, |t'|\}$  as opposed to  $|t| + |t'|$  required in the unoptimized construction. All of the simplifications discussed so far do not hold for  $A_\perp(t, t')$  with  $t < 0$  and  $t' > 0$ , and we apply the same circuit as the unoptimized construction just for this remaining case.

## APPENDIX E: DETAILS OF NUMERICAL EXPERIMENTS

In all numerical experiments, the standard deviation of the normal distribution is set relative to the maximum evolution time as  $\sigma = T/4$ . We note that while the median of means estimation provides rigorous performance guarantees, it introduces an additional hyperparameter. Therefore, in the present numerical experiments we simply use an empirical mean estimator which has a similar performance in practice. In all numerical experiments we report the performance in terms of the mean absolute error  $\Delta\langle O \rangle = (1/K) \sum_i |\langle O \rangle_i - \langle O \rangle^*|$  over  $K$  independent runs, where  $\langle O \rangle_i$  is the  $i$ th independent estimate of the ratio, and  $\langle O \rangle^*$  is the exact expectation value (in the case of MPS simulation it is not exact, but obtained using alternative techniques).

### 1. Random-field Heisenberg model

We consider the 1D Heisenberg chain with periodic boundary conditions, random local magnetic fields and nearest neighbor interactions as

$$\mathcal{H} = J \sum_{j=1}^N \vec{S}_j \cdot \vec{S}_{j+1} + \sum_j h_j Z_j. \quad (\text{E1})$$

Here  $\vec{S}_j = [X_j, Y_j, Z_j]$  is a Pauli vector on qubit  $j$ ,  $J$  is a coupling constant, and  $h_j \in [-h, h]$  is sampled uniformly at random. In our demonstrations, we use  $J = 0.1$  and  $h = 1$ .

### 2. Lattice Schwinger model

For the demonstrations involving variational methods, we considered a simple toy model in high-energy physics. We use the lattice Schwinger model [59] in its spin Hamiltonian formulation [60]  $\mathcal{H} = \mathcal{H}_{ZZ} + \mathcal{H}_{\pm} + \mathcal{H}_Z$ . These individual Hamiltonian terms are defined as

$$\begin{aligned} \mathcal{H}_{ZZ} &= \frac{J}{2} \sum_{n=2}^{N-1} \sum_{1 \leq k < l \leq n} Z_k Z_l \\ \mathcal{H}_{\pm} &= \frac{J}{2} \sum_{n=1}^{N-1} (w - (-1)^n \frac{m}{2} \sin(\theta)) [X_n X_{n+1} + Y_n Y_{n+1}] \\ \mathcal{H}_Z &= \frac{m \cos \theta}{2} \sum_{n=1}^N (-1)^n Z_n - \frac{J}{2} \sum_{n=1}^{N-1} (n \bmod 2) \sum_{l=1}^n Z_l. \end{aligned} \quad (\text{E2})$$

We set the hyperparameters heuristically to  $J = 1.0$ ,  $m = 0.1$ ,  $\theta = 0.5$ , and  $w = 0.1$ .

### 3. Fermi-Hubbard model

The Fermi-Hubbard model describes the behavior of interacting fermions and is defined by the second-quantized Hamiltonian

$$\tilde{\mathcal{H}} = -t \sum_{\langle i,j \rangle, s} (c_{i,s}^\dagger c_{j,s} + c_{j,s}^\dagger c_{i,s}) + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (\text{E3})$$

The operators  $c_{j,s}^\dagger$ ,  $c_{j,s}$ , and  $n_{j,s} = c_{j,s}^\dagger c_{j,s}$  denote creation, annihilation, and particle-number operators, respectively, at site  $j$  and spin  $s$ . The sum over  $\langle i,j \rangle$  runs over all neighboring lattice sites  $i$  and  $j$ . The first term corresponds to the tunneling of fermions with tunneling amplitude  $t$ , and the second term accounts for on-site interactions with interaction strength  $U$ .

In this work, we consider a 2 by 2 square lattice with  $t = 1$  and  $U = 12$  and impose periodic boundary conditions. After applying the Jordan-Wigner transformation, we obtain a Hamiltonian defined on 8 qubits,

$$\begin{aligned} \tilde{\mathcal{H}} &= 12 - 0.5X_1X_2 - 0.5X_2X_3 - 0.5X_3X_4 - 0.5X_5X_6 \\ &\quad - 0.5X_6X_7 - 0.5X_7X_8 - 0.5Y_1Y_2 - 0.5Y_2Y_3 - 0.5Y_3Y_4 \\ &\quad - 0.5Y_5Y_6 - 0.5Y_6Y_7 - 0.5Y_7Y_8 - 3Z_1 - 3Z_2 \\ &\quad - 3Z_3 - 0.5X_1Z_2Z_3X_4 - 0.5Y_1Z_2Z_3Y_4 - 3Z_4 - 3Z_5 \\ &\quad + 3Z_1Z_5 - 3Z_6 + 3Z_2Z_6 - 3Z_7 + 3Z_3Z_7 \\ &\quad - 0.5X_5Z_6Z_7X_8 - 0.5Y_5Z_6Z_7Y_8 - 3Z_8 + 3Z_4Z_8. \end{aligned} \quad (\text{E4})$$

In the time evolution of the hybrid quantum-classical algorithm, it is important to choose an initial state that has a high overlap with the target eigenstate. By choosing the lowest energy state within the subspace of four spins as the target state, we explicitly diagonalize the Hamiltonian and identify the computational basis states that have the highest overlap with it. Since two computational basis states have equally dominant overlap with the target eigenstate, we choose the initial state to be an equal superposition between them, which is  $\frac{1}{\sqrt{2}}(|10 100 101\rangle + |01 011 010\rangle)$  with an overlap of  $p_q \approx 0.62$ . Note that this is a state that can be easily prepared by applying bit flips onto a Greenberger-Horne-Zeilinger state, as shown in Fig. 11.

### 4. Ansatz-based variational quantum simulation

Variational quantum simulation leverages the variational principle to simulate the dynamics of quantum systems. This method is particularly useful for near-term quantum hardware, where implementing deep quantum circuits may not be feasible.

In our demonstration, we rely on an ansatz-based approach, using the hardware-efficient ansatz with ring topology, as shown in Fig. 13. Using the time grid with  $T = 50$  and  $dt = 0.2$ , we define the step size of the simulation algorithm as  $\delta t = dt/1000$ . Having the optimized

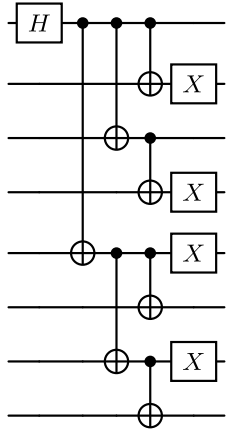


FIG. 11. An example of a quantum circuit that prepares the initial state  $\frac{1}{\sqrt{2}}(|10\ 100\ 101\rangle + |01\ 011\ 010\rangle)$  when applied onto the all zero state  $|00\ 000\ 000\rangle$  (i.e.,  $U_{\text{init}}$ ) considered in the Fermi-Hubbard model. The circuit can be viewed as applying bit flips onto a Greenberger-Horne-Zeilinger state.

parameters at  $t = 0$ , we seek to estimate the parameters corresponding to the time-evolved state at each step. For this, we need to solve, the linear system of equations

$$\sum_j A_{ij}^R \dot{\theta}_j = C_i^I, \quad (\text{E5})$$

where

$$A_{ij} = \frac{\partial \langle \psi(\boldsymbol{\theta}(t)) | \partial \psi(\boldsymbol{\theta}(t)) \rangle}{\partial \theta_i} \quad (\text{E6})$$

is the quantum metric tensor, and

$$C_i = \frac{\partial \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle}{\partial \theta_i} \quad (\text{E7})$$

is the gradient vector. Here  $A_{ij}^R$  and  $C_i^I$  denote the real and imaginary parts respectively. By solving for the time derivatives  $\dot{\boldsymbol{\theta}}$ , we can compute the updated parameters at the next timestep as  $\theta'_i = \theta_i + \dot{\theta}_i \delta t$ . In practical scenarios, the quantum metric tensor can be ill conditioned, therefore instead of solving (E5) we find the time derivatives, that minimize the Tikhonov-regularized system

$$\sum_j M_{ij}^R \dot{\theta}_j = C_i^I, \quad (\text{E8})$$

where  $M = A + \lambda \mathbb{1}$ , and we set  $\lambda = 10^{-4}$ .

It is obvious from this update rule that the fidelity of the states in this iterative process highly depends on the step size. Therefore, we chose  $\delta t$  small enough to maintain high fidelity. The fidelity of the variationally time-evolved states  $|\psi(\boldsymbol{\theta}(t))\rangle$  compared to the exact time evolution  $|\phi(t)\rangle$  is shown in Fig. 12.

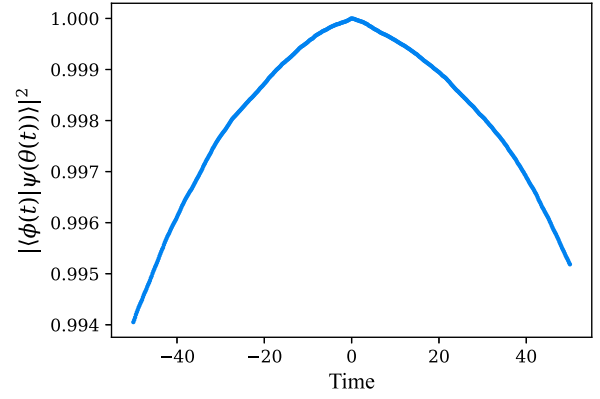


FIG. 12. Fidelity of the variationally time-evolved states.

### 5. Early fault-tolerant quantum devices

For our simulations in Sec. IV B 2, we assume that the dominant error source of an early fault-tolerant device is the approximation of continuous rotation gates by sequences of Clifford+T gates, which noise model is often assumed in the literature [29,56–58]. Since the magic state distillation required to implement T gates with sufficiently suppressed errors is costly, an early fault-tolerant device would have to resort to shallow approximation sequences. This results in deviations from the ideal continuous rotation gate, which we model by applying small depolarizing errors after continuous rotation gates [56,57,85] are applied.

Both the unoptimized construction described in Sec. III E (shown in the quantum computation step of Fig. 1) and the optimized construction given in Appendix D (shown in Fig. 10) involve either direct or controlled time evolutions generated by the Hamiltonian with the term proportional to the identity operator removed. When applying Trotterization as in Sec. IV B 1 for the Fermi-Hubbard Hamiltonian in Eq. (E4), each continuous rotation gate is generated by a Pauli string of the form  $S_i Z_{i+1} \dots Z_{j-1} S_j$  with  $S \in \{X, Y, Z\}$ . Errors for rotations that correspond to

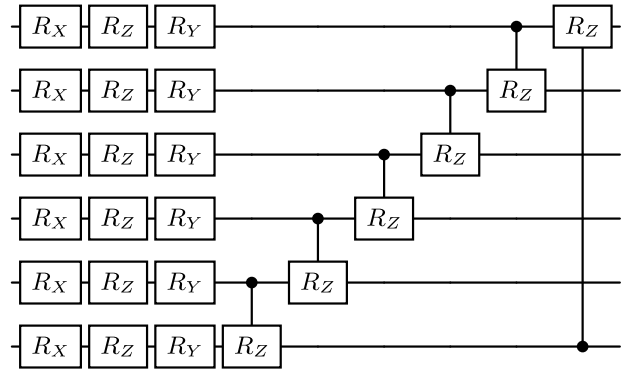


FIG. 13. A single layer of the hardware-efficient ansatz used for variational quantum simulation.

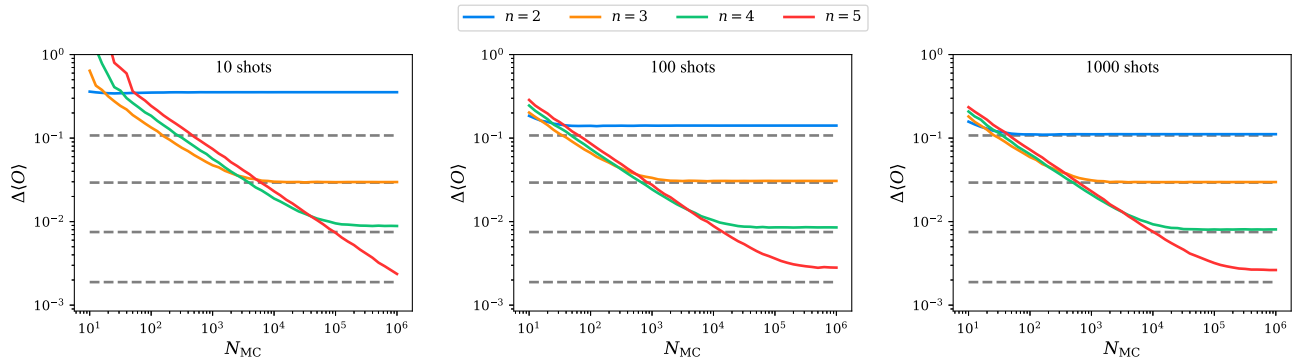


FIG. 14. Exact simulation results for a 10-qubit random-field Heisenberg model, considering shot noise. Here, we used the same settings as in Fig. 4. The effect of having a finite number of shots  $N_s$  was modeled by adding random noise sampled from a Gaussian distribution with standard deviation set to  $1/\sqrt{N_s}$ .

terms with  $i = j$  or  $j = i + 1$  are modeled by single-qubit depolarizing errors occurring with equal probabilities on qubits  $i$  and  $j$  (without double-counting when  $i = j$ ). Since a network of fermionic SWAP gates are able to remove the  $Z$  operators sandwiched between  $S_i$  and  $S_j$  in the rest of the terms [86–88], we model errors for these more general rotation terms similarly by only applying depolarizing errors on qubits  $i$  and  $j$ . When the time evolution is controlled by the state of an ancilla qubit, the ancilla qubit is also subject to a depolarizing error with the same error probability.

## APPENDIX F: ADDITIONAL NUMERICAL RESULTS

### 1. Exact simulation with shot noise

We also consider shot noise arising from the finite number of quantum circuit evaluations for the random-field Heisenberg model, rerunning the MC estimator with noisy data. For a single Pauli string, sampling the quantum state corresponds to sampling from a binomial distribution  $\mathcal{B}(N_s, p)$ , and this can be approximated by a Gaussian

distribution  $\mathcal{N}(N_s p, N_s p(1-p))$  for large number of shots  $N_s$ . Considering the upper bound  $p(1-p) < 1$ , we model the effect of shot noise on the corresponding  $A$  and  $B$  objects by sampling from the Gaussian distribution with standard deviation  $1/\sqrt{N_s}$ . We consider cases where both the real and imaginary parts of each entry of the  $A$  and  $B$  correlators are estimated using 10, 100, or 1000 uncorrelated shots. The corresponding results are shown in Fig. 14.

With a standard deviation corresponding to only 1000 shots, our MC method shows similar performance to the noiseless case. Considering fewer shots, we get irregular behavior that does not necessarily reach the theoretical performance of the noiseless results, but in most cases, it still suppresses the initial error along with the shot noise itself.

### 2. MPS simulations with variance reduction

We compare the effect of the variance reduction method described in Sec. III C for the ground state estimation with MPS simulation. In Fig. 15 we show the error compared to the DMRG energy without (left) and with (right) the variance reduction for different number of copies and

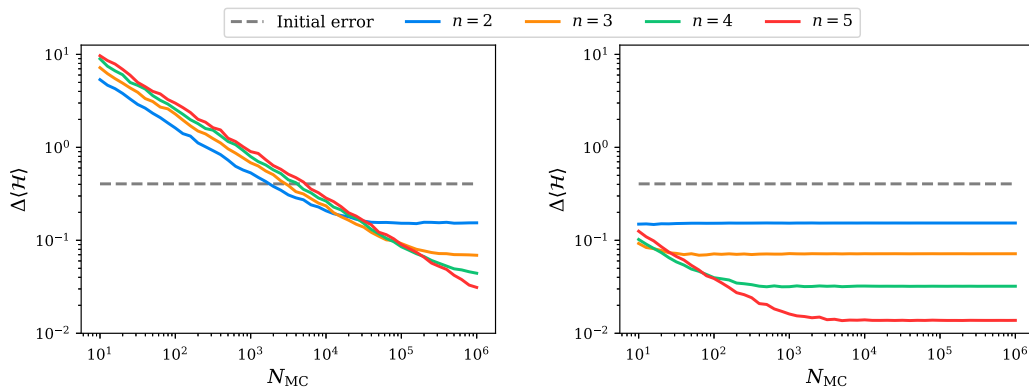


FIG. 15. Matrix product state simulation results without (left) and with (right) the variance reduction technique for the ground state of a 100-qubit random field Heisenberg model.

increasing number of MC samples. It is clear that the initial bias without the reduction technique can be substantial, when it is not yet suppressed by the number of samples.

- 
- [1] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
  - [2] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys.* **92**, 015003 (2020).
  - [3] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, *Chem. Rev.* **120**, 12685 (2020).
  - [4] M. Motta and J. E. Rice, Emerging quantum computing algorithms for quantum chemistry, *WIREs Comput. Mol. Sci.* **12**, e1580 (2022).
  - [5] G. H. Low, T. J. Yoder, and I. L. Chuang, Methodology of resonant equiangular composite quantum gates, *Phys. Rev. X* **6**, 041067 (2016).
  - [6] G. H. Low and I. L. Chuang, Optimal Hamiltonian simulation by quantum signal processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
  - [7] L. Lin and Y. Tong, Near-optimal ground state preparation, *Quantum* **4**, 372 (2020).
  - [8] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, *SIAM J. Comput.* **35**, 1070 (2006). Conference version in *Proceedings of the 24th FSTTCS* (2004), pp. 372–383.
  - [9] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, M. Kastoryano, R. Babbush, J. Preskill, D. R. Reichman, E. T. Campbell, E. F. Valeev, L. Lin, and G. K.-L. Chan, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, *Nat. Commun.* **14**, 1952 (2023).
  - [10] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, *Phys. Rev. X* **8**, 041015 (2018).
  - [11] C. Gidney and M. Ekerå, How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits, *Quantum* **5**, 433 (2021).
  - [12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
  - [13] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
  - [14] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
  - [15] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, *npj Quantum Inf.* **5**, 75 (2019).
  - [16] B. Koczor and S. C. Benjamin, Quantum natural gradient generalized to noisy and nonunitary circuits, *Phys. Rev. A* **106**, 062416 (2022).
  - [17] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, *Quantum* **3**, 156 (2019).
  - [18] G. Boyd and B. Koczor, Training variational quantum circuits with CoVaR: Covariance root finding with classical shadows, *Phys. Rev. X* **12**, 041022 (2022).
  - [19] W. Hwang and B. Koczor, Preparing ground and excited states using adiabatic CoVaR, *New J. Phys.* **27**, 023025 (2025).
  - [20] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, Barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
  - [21] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, *Nat. Commun.* **13**, 7760 (2022).
  - [22] A. Tikku and I. H. Kim, Circuit depth versus energy in topologically ordered systems, [arXiv:2210.06796 \[quant-ph\]](https://arxiv.org/abs/2210.06796).
  - [23] G. Boyd, B. Koczor, and Z. Cai, High-dimensional subspace expansion using classical shadows, *Phys. Rev. A* **111**, 022423 (2025).
  - [24] H. Jnane, J. Steinberg, Z. Cai, H. C. Nguyen, and B. Koczor, Quantum error mitigated classical shadows, *PRX Quantum* **5**, 010324 (2024).
  - [25] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, *PRX Quantum* **3**, 010318 (2022).
  - [26] G. Wang, D. S. França, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, *Quantum* **7**, 1167 (2023).
  - [27] G. Wang, D. S. França, G. Rendon, and P. D. Johnson, Efficient ground-state-energy estimation and certification on early fault-tolerant quantum computers, *Phys. Rev. A* **111**, 012426 (2025).
  - [28] O. Kiss, U. Azad, B. Requena, A. Roggero, D. Wakeham, and J. M. Arrazola, Early fault-tolerant quantum algorithms in practice: Application to ground-state energy estimation, *Quantum* **9**, 1682 (2025).
  - [29] H. H. S. Chan, R. Meister, M. L. Goh, and B. Koczor, Algorithmic shadow spectroscopy, *PRX Quantum* **6**, 010352 (2025).
  - [30] G. Mazzola, Quantum computing for chemistry and physics applications from a Monte Carlo perspective, *J. Chem. Phys.* **160**, 010901 (2024).
  - [31] M. Huo and Y. Li, Error-resilient Monte Carlo quantum simulation of imaginary time, *Quantum* **7**, 916 (2023).
  - [32] P. Zeng, J. Sun, and X. Yuan, Universal quantum algorithmic cooling on a quantum computer, [arXiv:2109.15304 \[quant-ph\]](https://arxiv.org/abs/2109.15304).
  - [33] J. Sun, P. Zeng, T. Gur, and M. S. Kim, High-precision and low-depth eigenstate property estimation: Theory and resource estimation, [arXiv:2406.04307 \[quant-ph\]](https://arxiv.org/abs/2406.04307).
  - [34] Y. Yang, B.-N. Lu, and Y. Li, Accelerated quantum Monte Carlo with mitigated error on noisy quantum computer, *PRX Quantum* **2**, 040361 (2021).
  - [35] B. Koczor, Exponential error suppression for near-term quantum devices, *Phys. Rev. X* **11**, 031057 (2021).

- [36] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, *Phys. Rev. X* **11**, 041036 (2021).
- [37] G. H. Low, R. King, D. W. Berry, Q. Han, A. E. DePrince III, A. White, R. Babbush, R. D. Somma, and N. C. Rubin, Fast quantum simulation of electronic structure by spectrum amplification, *Phys. Rev. X* **15**, 041016 (2025).
- [38] D. W. Berry, Y. Tong, T. Khattar, A. White, T. I. Kim, G. H. Low, S. Boixo, Z. Ding, L. Lin, S. Lee *et al.*, Rapid initial-state preparation for the quantum simulation of strongly correlated molecules, *PRX Quantum* **6**, 020327 (2025).
- [39] M. Mörchen, G. H. Low, T. Weymuth, H. Liu, M. Troyer, and M. Reiher, Classification of electronic structures and state preparation for quantum computation of reaction chemistry, [arXiv:2409.08910](https://arxiv.org/abs/2409.08910).
- [40] A. D. Bookatz, QMA-complete problems, *Quantum Inf. Comput.* **14**, 361 (2014).
- [41] Z. Zimborás, B. Koczor, Z. Holmes, E.-M. Borrelli, A. Gilyén, H.-Y. Huang, Z. Cai, A. Acín, L. Aolita, L. Banchi *et al.*, Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't, [arXiv:2501.05694](https://arxiv.org/abs/2501.05694).
- [42] R. Babbush, R. King, S. Boixo, W. Huggins, T. Khattar, G. H. Low, J. R. McClean, T. O'Brien, and N. C. Rubin, The grand challenge of quantum applications, [arXiv:2511.09124](https://arxiv.org/abs/2511.09124).
- [43] 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).
- [44] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, *Quantum* **3**, 191 (2019).
- [45] H. F. Trotter, On the product of semi-groups of operators, *Proc. Am. Math. Soc.* **10**, 545 (1959).
- [46] M. Suzuki, Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, *Commun. Math. Phys.* **51**, 183 (1976).
- [47] B. Koczor, The dominant eigenvector of a noisy quantum state, *New J. Phys.* **23**, 123047 (2021).
- [48] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, *Quantum Inf. Comput.* **12**, 901 (2012).
- [49] H. Jnane, B. Undseth, Z. Cai, S. C. Benjamin, and B. Koczor, Multicore quantum computing, *Phys. Rev. Appl.* **18**, 044064 (2022).
- [50] T. Araki, J. F. Goodwin, and B. Koczor, Space-time tradeoff in networked virtual distillation, *Phys. Rev. A* **112**, 042619 (2025).
- [51] L. N. Trefethen, Exactness of quadrature formulas, *SIAM Rev.* **64**, 132 (2022).
- [52] Z. Cai, A. Chapman, H. Jnane, and B. Koczor, Biased estimator channels for classical shadows, *Phys. Rev. A* **111**, L030402 (2025).
- [53] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, New York, USA, 2007), 3rd ed.
- [54] C. Kiumi and B. Koczor, TE-PAI: Exact time evolution by sampling random circuits, *Quantum Sci. Technol.* **10**, 045071 (2025).
- [55] G. Rendon, J. Watkins, and N. Wiebe, Improved accuracy for Trotter simulations using Chebyshev interpolation, *Quantum* **8**, 1266 (2024).
- [56] V. Kliuchnikov, K. Lauter, R. Minko, A. Paetznick, and C. Petit, Shorter quantum circuits via single-qubit gate approximation, *Quantum* **7**, 1208 (2023).
- [57] B. Koczor, Sparse probabilistic synthesis of quantum operations, *PRX Quantum* **5**, 040352 (2024).
- [58] M. L. Goh and B. Koczor, Direct estimation of the density of states for Fermionic systems, [arXiv:2407.03414 \[quant-ph\]](https://arxiv.org/abs/2407.03414).
- [59] C. J. Hamer, Z. Weihong, and J. Oitmaa, Series expansions for the massive Schwinger model in Hamiltonian lattice theory, *Phys. Rev. D* **56**, 55 (1997).
- [60] B. Chakraborty, M. Honda, T. Izubuchi, Y. Kikuchi, and A. Tomiya, Classically emulated digital quantum simulation of the Schwinger model with a topological term via adiabatic state preparation, *Phys. Rev. D* **105**, 094503 (2022).
- [61] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [62] B. Koczor, J. J. L. Morton, and S. C. Benjamin, Probabilistic interpolation of quantum rotation angles, *Phys. Rev. Lett.* **132**, 130602 (2024).
- [63] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [64] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix product density operators: Simulation of finite-temperature and dissipative systems, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [65] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [66] N. Ares and D. A. Wisniacki, Loschmidt echo and the local density of states, *Phys. Rev. E* **80**, 046216 (2009).
- [67] E. Campbell, Random compiler for fast Hamiltonian simulation, *Phys. Rev. Lett.* **123**, 070503 (2019).
- [68] D. Aharonov and A. Ta-Shma, Adiabatic Quantum State Generation and Statistical Zero Knowledge, in *Proceedings of the 35th ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2003), pp. 20–29.
- [69] R. D. Somma and S. Boixo, Spectral gap amplification, *SIAM J. Comput.* **42**, 593 (2013).
- [70] G. H. Low and I. L. Chuang, Hamiltonian simulation by uniform spectral amplification, [arXiv:1707.05391](https://arxiv.org/abs/1707.05391).
- [71] A. Zlokapa and R. D. Somma, Hamiltonian simulation for low-energy states with optimal time dependence, *Quantum* **8**, 1449 (2024).
- [72] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, QuEST and high performance simulation of quantum computers, *Sci. Rep.* **9**, 10736 (2019).
- [73] T. Jones and S. Benjamin, QuESTlink—Mathematica embiggened by a hardware-optimised quantum emulator, *Quantum Sci. Technol.* **5**, 034012 (2020).

- [74] J. Gray, quimb: A Python library for quantum information and many-body calculations, *J. Open Source Software* **3**, 819 (2018).
- [75] A. Kay, Tutorial on the Quantikz package, [arXiv:1809.03842 \[quant-ph\]](https://arxiv.org/abs/1809.03842).
- [76] A. Richards, University of Oxford Advanced Research Computing, 2015, <https://doi.org/10.5281/zenodo.22558>
- [77] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, M. Greiner, H. Pichler, and P. Hayden, Quantum virtual cooling, *Phys. Rev. X* **9**, 031013 (2019).
- [78] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Edition (Cambridge University Press, Cambridge, England, 2010).
- [79] 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).
- [80] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, Quantum amplitude amplification and estimation, *Quantum Comput. Quantum Inf.* **305**, 53 (2002).
- [81] X. Xu and Y. Li, Quantum-assisted Monte Carlo algorithms for fermions, *Quantum* **7**, 1072 (2023).
- [82] 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).
- [83] C. L. Cortes and S. K. Gray, Quantum Krylov subspace algorithms for ground- and excited-state energy estimation, *Phys. Rev. A* **105**, 022417 (2022).
- [84] S. Lu, M. C. Bañuls, and J. I. Cirac, Algorithms for quantum simulation at finite energies, *PRX Quantum* **2**, 020321 (2021).
- [85] Z. Cai, X. Xu, and S. C. Benjamin, Mitigating coherent noise using Pauli conjugation, *npj Quantum Inf.* **6**, 17 (2020).
- [86] F. Verstraete, J. I. Cirac, and J. I. Latorre, Quantum circuits for strongly correlated quantum systems, *Phys. Rev. A* **79**, 032316 (2009).
- [87] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, *Phys. Rev. Lett.* **120**, 110501 (2018).
- [88] Z. Cai, Resource estimation for quantum variational simulations of the Hubbard model, *Phys. Rev. Appl.* **14**, 014059 (2020).