

Error-mitigated nonorthogonal quantum eigensolver via shadow tomography

Hang Ren ^{1,2,3} Yipei Zhang,^{1,2,3} Wendy M. Billings,^{1,2,3} Rebecca Tomann ^{1,2,3} Nikolay V. Tkachenko ^{1,2,3}
Mingyu Kang ^{1,2,3} Martin Head-Gordon,^{2,3} and K. Birgitta Whaley ^{1,2,3,*}

¹Berkeley Center for Quantum Information and Computation, Berkeley, California 94720, USA

²Department of Chemistry, University of California, Berkeley, California 94720, USA

³Challenge Institute for Quantum Computation, Berkeley, California 94720, USA



(Received 22 April 2025; accepted 30 December 2025; published 10 March 2026)

We present a shadow-tomography-enhanced nonorthogonal quantum eigensolver (NOQE) for more efficient and accurate electronic structure calculations on near-term quantum devices. By integrating shadow tomography into the NOQE, the measurement cost scales linearly rather than quadratically with the number of reference states, while also reducing the required qubits and circuit depth by half. This approach enables extraction of all matrix elements via randomized measurements and classical postprocessing. We analyze its sample complexity and show that, for small systems, it remains constant in the high-precision regime, while for larger systems, it scales linearly with the system size. We further apply shadow-based error mitigation to suppress noise-induced bias without increasing quantum resources. Demonstrations on the hydrogen molecule in the strongly correlated regime achieve chemical accuracy under realistic noise, showing that our method is both resource-efficient and noise-resilient for practical quantum chemistry simulations in the near term.

DOI: [10.1103/7c5b-3v56](https://doi.org/10.1103/7c5b-3v56)

I. INTRODUCTION

Quantum computing offers the potential for simulating molecular electronic structures that are intractable with classical methods [1]. Accurately determining the ground and excited state energies of molecules is central to understanding chemical reactions, catalysis, and materials design. While variational quantum eigensolvers (VQEs) have received significant interest for these problems, the iterative optimization that lies at the core of these methods often becomes measurement and resource intensive as system size grows [2,3], and may be constrained by the phenomenon of barren plateaux in high dimensions [4]. Nonvariational methods such as quantum phase estimation [5,6] typically require significantly deeper circuits that exceed the capabilities of current quantum hardware.

The nonorthogonal quantum eigensolver (NOQE) [7] presents an alternative approach that has been shown to provide significant quantum advantage for electronic systems with both strong and weak electron correlations. By employing multiple nonorthogonal reference states derived from unrestricted Hartree-Fock (UHF) solutions [8] and a unitary coupled cluster (UCC) ansatz, NOQE solves a generalized eigenvalue problem [9] to obtain molecular energies directly, thereby circumventing costly variational optimization. The use of multiple reference states allows description of strong

correlation, while dressing by UCC-type ansatz allows incorporation of weak correlations. NOQE is both variational and size consistent, two key desirable features found in classical electronic structure algorithms only for the uncorrelated Hartree-Fock solution, and for full configuration interaction, which is intractable for large systems. The classical analog of NOQE, namely, nonorthogonal configuration interaction with exponential correlators, is intractable and the closest classical analogs are exponentially more costly than NOQE [7].

While NOQE avoids the high cost of repeated measurements associated with variational optimization, measuring the matrix elements can still be expensive, given that matrix size scales quadratically $\mathcal{O}(M^2)$ with the number of nonorthogonal reference states M . Additionally, the estimation of individual Hamiltonian matrix elements requires measuring a series of Pauli strings, which may be large for systems with complex Hamiltonians.

A number of protocols have been proposed to simplify measurement strategies of electronic Hamiltonians [10–14]. These have typically focused on diagonal elements of the Hamiltonian matrix, because of the emphasis in the community on VQE methodologies. However, in NOQE we require both diagonal and off-diagonal Hamiltonian elements, as well as off-diagonal overlap elements between nonorthogonal reference states. For diagonal Hamiltonian matrix elements, the most efficient protocols require additional unitary operations to generate premeasurement entanglement [13] and are not directly applicable to measurement of off-diagonal matrix elements.

In this work, we address these issues by integrating shadow tomography (classical shadows for quantum states) [15] into the NOQE framework. Shadow tomography is a powerful measurement protocol that efficiently predicts multiple properties of a quantum state from a small number of randomized

*Contact author: whaley@berkeley.edu

measurements. By applying shadow tomography to each reference state and its phase-shifted superpositions with the vacuum state (auxiliary variants), we eliminate the need to jointly prepare pairs of reference states. Instead, the necessary Hamiltonian and overlap matrix elements are reconstructed classically from classical shadows. This reduces the measurement scaling with respect to the reference states to $O(M)$. It also halves the number of required qubits N and shortens the circuit depth, making the protocol more experimentally feasible.

An additional advantage of our approach is its natural compatibility with shadow-based error mitigation techniques, such as shadow distillation [16]. This enables the suppression of noise-induced bias in the estimated energies without incurring additional quantum overhead.

In the noisy intermediate-scale quantum (NISQ)/pre-fault-tolerant regime, several hybrid eigensolvers share NOQE's paradigm—prepare a compact nonorthogonal basis on hardware, estimate matrix elements, and then solve a classical generalized eigenproblem. Representative examples include quantum-selected configuration interaction [17], which adaptively selects important configurations to expand the subspace; quantum filter diagonalization [18], which generates energy-localized states via short-time dynamics combined with energy filtering; and quantum Krylov subspace diagonalization [19], which constructs a Krylov basis from repeated applications of the Hamiltonian or time-evolution operators. More recently, sample-based quantum diagonalization approaches have been proposed [20,21], further extending this hybrid paradigm toward pre-fault-tolerant algorithms. Our shadow-based NOQE fits into this general class of methods, differing primarily in its use of multiple NOQE-specific reference states and the use of shadow tomography with higher-order U-statistics, which enables seamless shadow-distillation error mitigation.

In this paper, we first analytically demonstrate the effectiveness of the shadow-tomography-enhanced NOQE method by analyzing its sampling complexity and circuit size, and comparing them to those of the standard NOQE protocol using Hadamard measurements. We then validate its practical performance through an application to computing the electronic energies of the hydrogen molecule near the Coulson-Fischer point, a regime characterized by strong electron correlation. In this application, we use noise models based on the state-of-the-art Quantinuum Model H2 hardware system [22].

These demonstrations show that shadow-tomography-enhanced NOQE provides a practical and noise-resilient framework for realizing accurate electronic structure calculations, making it suitable for implementation on near-term quantum devices.

The remainder of the paper is organized as follows. Section II reviews the NOQE framework. Section III introduces the fundamentals of shadow tomography. Building on these concepts, Sec. IV details how to integrate shadow tomography into NOQE, and Sec. V undertakes a sampling complexity analysis that quantifies the resource requirements for measurements using shadow tomography and for using the Hadamard test approach of Ref. [7]. In Sec. VI, we discuss how shadow distillation can be implemented as an error mitigation strategy for NOQE. Section VII demonstrates the

accuracy and efficiency of our approach with calculations for electronic states and energies of the hydrogen molecule, while Sec. VIII confirms the effectiveness of the protocols under realistic noise. Finally, Sec. IX summarizes our findings and outlines potential directions for future research.

II. NOQE FRAMEWORK ANALYSIS

This section reviews the NOQE framework and its measurement requirements. In Sec. II A, we outline the underlying principles of NOQE. In Sec. II B, we summarize the original NOQE circuit that measures Hamiltonian and overlap matrix elements. Finally, in Sec. II C, we quantify the sample complexity of the original protocol, providing a baseline for comparing our shadow tomography-based approach.

A. Working principle of NOQE

Accurate electronic energy estimation is central to quantum chemistry. Traditional algorithms such as the VQE can approximate ground-state energies but often require extensive iterative optimization and very large numbers of measurements [23]. The NOQE offers a different approach, eliminating the need for expensive variational optimization by using a set of nonorthogonal reference states to project the Hamiltonian into a smaller subspace and then directly solve a generalized eigenvalue problem in that subspace.

To construct these reference states, NOQE applies a UCC correlator $e^{\hat{t}_i}$, e.g., the unitary coupled-cluster doubles (UCCD) correlator that includes double excitations, to a UHF state $|\psi_{\text{UHF}}\rangle$ (see Appendix A 1 for details):

$$|\psi_i\rangle = e^{\hat{t}_i}|\psi_{\text{UHF}}\rangle. \quad (1)$$

A key feature of the NOQE approach is that these reference states deriving from distinct UHF states are not orthogonal.

The UHF is chosen over restricted Hartree-Fock for its ability to introduce symmetry-broken references and thus achieve overall greater flexibility in description of strongly correlated systems. The UHF reference offers a mean-field approximation to the wave function of the system and weak (dynamic) correlation is introduced by the UCCD dressing. The strong (static) correlation is captured through the inclusion of multiple symmetry-broken reference states.

The variational parameters of the UCCD ansatz can be approximated through classical preprocessing at polynomial cost using second-order many-body perturbation (MP2) theory [7]. Relatively shallow circuit implementation of the UCCD ansatz can be realized through a low-rank tensor decomposition of the cluster operator and consequent trotterization [24]. We focus here on the UCCD ansatz but other alternatives such as the unitary cluster Jastrow ansatz that can reduce circuit complexity are also possible [7].

The NOQE eigenstates are expressed as a linear combination of dressed nonorthogonal reference states:

$$|\Phi_{\text{NOQE}}^k\rangle = \sum_{i=1}^M c_i^k |\psi_i\rangle. \quad (2)$$

The coefficients c_i^k are found by solving the generalized eigenvalue problem

$$\mathbf{H}\vec{c} = E^k \mathbf{S}\vec{c}, \quad (3)$$

where matrix elements of the Hamiltonian and overlap matrices $\mathbf{H}$ and $\mathbf{S}$ are given by

$$H_{ij} = \langle \psi_i | \hat{H} | \psi_j \rangle, \quad S_{ij} = \langle \psi_i | \psi_j \rangle. \quad (4)$$

The eigenvalue E^k corresponds to the electronic energy of the $|\Phi_{\text{NOQE}}^k\rangle$ state. It is important to note that the operator $e^{\hat{\tau}_i}$ in $|\psi_i\rangle$ is constructed in the single-particle basis of the state $|\psi_{\text{UHF}}\rangle$. Thus, to measure off-diagonal elements, we formally need to transform both correlators to the same single-particle basis using a unitary orbital rotation. In this work, we absorb the unitary orbital rotation operators ($\hat{U}_{i \rightarrow 1}$ and $\hat{U}_{j \rightarrow 1}$) that are required to transform $|\psi_i\rangle$ and $|\psi_j\rangle$ to the same single-particle basis [7], into the corresponding $e^{\hat{\tau}_i}$ and $e^{\hat{\tau}_j}$ circuits. Each reference state $|\psi_i\rangle$ encodes the occupation of N spin orbitals using N qubits, where each qubit represents one spin orbital under the Jordan-Wigner or Bravyi-Kitaev encoding [1]. Full details of these transformations and encodings can be found in the original NOQE paper [7].

Formation of the overlap matrix in the nonorthogonal basis of M reference states requires evaluating $M(M-1)/2$ off-diagonal elements, and the Hamiltonian matrix requires the same number of off-diagonal elements, plus evaluation of M diagonal matrix elements. In general, classical evaluation of matrix elements or operators between mean-field states dressed with the unitary coupled-cluster correlators becomes intractable, because the Baker-Campbell-Hausdorff expansion does not truncate and becomes infinite [25]. In contrast, it was shown in Ref. [23] that quantum computers can efficiently measure these matrix elements using a generalized Hadamard test, formally reducing the complexity scaling to polynomial dependence on the size of the system.

In the following, we use the original NOQE circuit as a concrete example to illustrate how to measure the matrices defined in Eq. (4). Building on this, we then develop a shadow-tomography-based measurement scheme that reduces both circuit width and depth, while improving the measurement efficiency in certain parameter regimes.

B. Original NOQE circuit

The original NOQE protocol employs a quantum circuit for a generalized Hadamard test to estimate both the overlap and Hamiltonian matrix elements. The process prepares two nonorthogonal reference states, $|\psi_i\rangle$ and $|\psi_j\rangle$, and places them into a controlled superposition. Figure 1 illustrates the key steps of this circuit.

Just before measurement, the system is in the state (ignoring the vacuum register)

$$|\phi_{ij}^{\text{end}}\rangle = \frac{1}{\sqrt{2}}(|\psi_i\rangle|+\rangle + e^{i\theta}|\psi_j\rangle|-\rangle). \quad (5)$$

By measuring the last qubit in the computational (Z) basis, one extracts either the real or imaginary part of the desired matrix element, depending on the choice of θ of the phase gate $P(\theta)$.

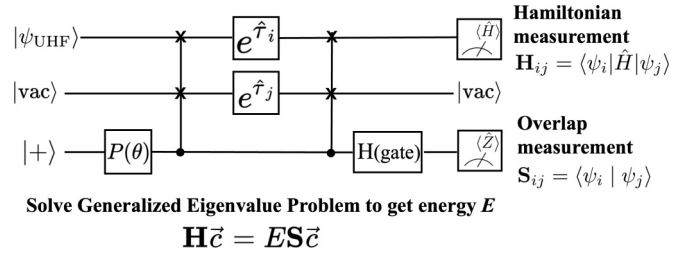


FIG. 1. Circuit diagram of the original NOQE method [7], modified to absorb the basis rotations into the ansatz operators $e^{\hat{\tau}_{i(j)}}$ (see text). Two nonorthogonal reference states, $|\psi_i\rangle$ and $|\psi_j\rangle$, are prepared and combined using controlled operations to enable modified Hadamard tests that measure H_{ij} and S_{ij} . Here, $|\text{vac}\rangle$ denotes a vacuum state (all-zero state in occupancy vector representation). Both qubit registers, $|\psi_{\text{UHF}}\rangle$ and $|\text{vac}\rangle$, consist of N qubits, equal to the number of spin orbitals. After measuring these matrix elements, one solves the generalized eigenvalue problem [Eq. (3)] on a classical computer to obtain the electronic energies.

For $\theta = 0$, measuring the last (ancilla) qubit, yields the real part of the overlap matrix

$$\langle \phi_{ij}^{\text{end}} | \hat{I} \otimes \hat{Z} | \phi_{ij}^{\text{end}} \rangle = \text{Re}(\langle \psi_i | \psi_j \rangle) = \text{Re}(S_{ij}). \quad (6)$$

Similarly, introducing the system Hamiltonian

$$\hat{H} = \sum_{k=1}^{\Omega} w_k \hat{P}_k, \quad (7)$$

where each $\hat{P}_k$ is a Pauli operator specified by Jordan-Wigner transformation of the fermionic Hamiltonian, allows one to measure

$$\langle \phi_{ij}^{\text{end}} | \hat{H} \otimes \hat{Z} | \phi_{ij}^{\text{end}} \rangle = \text{Re}(\langle \psi_i | \hat{H} | \psi_j \rangle) = \text{Re}(H_{ij}). \quad (8)$$

Adjusting θ to $\pi/2$ enables extraction of the corresponding imaginary components. Full details of the measurement procedure are given in Appendix A 2.

Once H_{ij} and S_{ij} are determined for all i, j , one solves the generalized eigenvalue problem of Eq. (3) classically to obtain the target energies.

Note that calculations made for the original NOQE protocol introduced in Ref. [7] were implemented using the unitary operations directly. In the current study, we explicitly design, compile, and optimize the corresponding gate-level circuits. We numerically simulate and validate the performance of the circuits and provide full implementation details in Appendix A.

C. Sample complexity of the original NOQE circuit

To quantify the measurement overhead of different NOQE schemes, here we derive the sample complexity of the original NOQE circuit in Fig. 1. Specifically, we consider the number of circuit repetitions n required to estimate a Hamiltonian matrix element to within an additive error ϵ . Here, ϵ refers to the dimensionless additive error in estimating a properly rescaled version of the observable; see Appendix D 1 for details on the rescaling convention and unit consistency. This definition of ϵ applies throughout the paper.

Estimating the Hamiltonian matrix is generally more challenging than estimating the overlap matrix, so we focus on this harder case. Let $D = 2^N$ be the dimension of the N -qubit Hilbert space. Define B as the upper bound on the squared Frobenius norm of the electronic Hamiltonian $\hat{H}$:

$$\text{Tr}(\hat{H}^2) = \sum_k w_k^2 D \leq B. \quad (9)$$

If there are Ω Pauli terms in $\hat{H}$, the total number of measurements n in order to estimate an individual matrix element H_{ij} to additive error ϵ can be shown to be (see Appendix D 2)

$$n = \mathcal{O}\left(\frac{B\Omega}{D\epsilon^2}\right). \quad (10)$$

Without loss of generality, $\hat{H}$ can always be rescaled such that its spectral norm is bounded by $\|\hat{H}\|_\infty \leq 1$, which implies $B \leq D$. Additionally, it is known in many cases that Ω scales polynomially with the number of spin orbitals N (typically as N^4) [13]. Considering there are M^2 matrix elements $\hat{H}_{ij}$ to evaluate, the total number of measurements required is then

$$n = \mathcal{O}(\text{polylog}(D) M^2 / \epsilon^2). \quad (11)$$

Here, the $\text{polylog}(D)$ factor arises from the polynomial scaling of Ω in N ; using $D = 2^N$, this corresponds to a $(\log D)^2$ dependence. Although Eq. (11) shows favorable scaling with respect to the Hilbert space dimension D , its scaling may become unfavorable when the desired error tolerance ϵ is within chemical accuracy, which often requires $\epsilon \leq 1/D$ for systems of intermediate size. Note that we use the dimensionless definition of ϵ from Appendix D 1, so its numerical value can be directly compared with $1/D$.

While this paper concerns the complexity with respect to ϵ , the error in the matrix elements, this ultimately needs to be translated to $\mathcal{E}$, the error in the energy eigenvalues estimated from solving the generalized eigenvalue problem with the matrices $\mathbf{H}$ and $\mathbf{S}$. In Appendix A 6, we show that the errors $\mathcal{E}$ and ϵ are related by $\mathcal{E} = \mathcal{O}(\epsilon M^{1/2} / s_{\min})$ or $\mathcal{E} = \mathcal{O}(\epsilon M^{3/2} / \sqrt{s_{\min}^2 + E_{\min}^2})$, where $s_{\min}$ ($E_{\min}$) is the minimum eigenvalue of $\mathbf{S}$ ($\mathbf{H}$), depending on whether the Stewart bound [26] or the Mathias-Li bound [27,28] is used, respectively. Note that the latter relation has an improved scaling with respect to M from the previously known results in Refs. [7,29].

Implementing this algorithm on today's noisy quantum computers without error correction necessitates an error mitigation technique that introduces minimal overhead while remaining practically compatible with the NOQE framework. These considerations motivate the integration of shadow tomography into the NOQE framework, as we now discuss in Secs. III and IV.

III. SHADOW TOMOGRAPHY

Shadow tomography provides a powerful and efficient method for extracting information about a quantum state from relatively few measurements [15]. In Sec. III A, we describe the data-acquisition process and the construction of classical shadows. Section III B then explains how to use these classical shadows to estimate observables.

We note that, throughout this paper, operators are denoted with a hat symbol. Scalar estimators are written without a hat, while operator estimators include one. When ambiguity between operators and estimators could occur, a subscript “es” is added to the estimator for clarification.

A. Data acquisition and classical shadows

The first step in shadow tomography is to perform randomized measurements on the state of interest $\hat{\rho}$. For each trial, we (1) sample a unitary $\hat{U}$ at random from a suitable ensemble $\mathcal{U}$, such as the Clifford group, (2) apply $\hat{U}$ to the state ρ and measure in the computational (Z) basis, and (3) record the outcome bitstring b and the chosen unitary $\hat{U}$.

From this measurement, one obtains a *snapshot*

$$\hat{\sigma}_{\hat{U},b} = \hat{U}^\dagger |b\rangle\langle b| \hat{U}. \quad (12)$$

Averaging $\hat{\sigma}_{\hat{U},b}$ over both $\hat{U}$ and b defines a measurement map $\mathcal{M}$ acting on $\hat{\rho}$:

$$\mathbb{E}_{\hat{U},b}[\hat{\sigma}_{\hat{U},b}] = \mathcal{M}(\hat{\rho}). \quad (13)$$

For ensembles like the Clifford group, $\mathcal{M}$ is tomographically complete and invertible. The *classical shadow* of $\hat{\rho}$ is defined as

$$\hat{\rho}_{\hat{U},b} = \mathcal{M}^{-1}(\hat{\sigma}_{\hat{U},b}), \quad (14)$$

where $\mathcal{M}^{-1}$ is the reconstruction map. In particular, for n -qubit Clifford measurements,

$$\mathcal{M}^{-1}(\hat{\sigma}_{\hat{U},b}) = (2^n + 1)\hat{\sigma}_{\hat{U},b} - I. \quad (15)$$

It can be seen easily that the classical shadow is a sample mean estimator of the original state since one can check that

$$\mathbb{E}[\hat{\rho}_{\hat{U},b}] = \hat{\rho}. \quad (16)$$

Since $\mathbb{E}_{\hat{U},b}[\hat{\rho}_{\hat{U},b}] = \hat{\rho}$, averaging many such classical shadows yields an unbiased estimator of the state. With a sufficiently large collection of classical shadows, one can then estimate expectation values of a broad range of observables, as described next.

B. Postprocessing: Predicting properties with classical shadows

Once we have collected a set of classical shadows $\mathcal{D}(n) = \{\hat{\rho}_\alpha\}_{\alpha=1}^n$, we form an unbiased estimator $\hat{\rho}_{\text{cs}} = \frac{1}{n} \sum_{\alpha=1}^n \hat{\rho}_\alpha$ for the underlying quantum state $\hat{\rho}$. Given any observable O , we can then estimate its expectation value as

$$o = \text{Tr}[\hat{\rho}_{\text{cs}} \hat{O}], \quad (17)$$

which satisfies $\mathbb{E}[o] = \text{Tr}[\hat{\rho} \hat{O}]$.

A key advantage of shadow tomography is its efficiency in estimating multiple observables. Specifically, under Clifford group measurements, to estimate K observables $\{\hat{O}_l\}_{l=1}^K$ each within additive error ϵ , a sample complexity of

$$n = \mathcal{O}\left(\frac{\log(K)B}{\epsilon^2}\right) \quad (18)$$

is sufficient, where B is the bound of $B \geq \text{Tr}[\hat{O}_l^2]$ for all l [15]. This scaling improves over full tomography, which in the optimal case would still require $\mathcal{O}(4^N/\epsilon^2)$ samples, an exponential overhead.

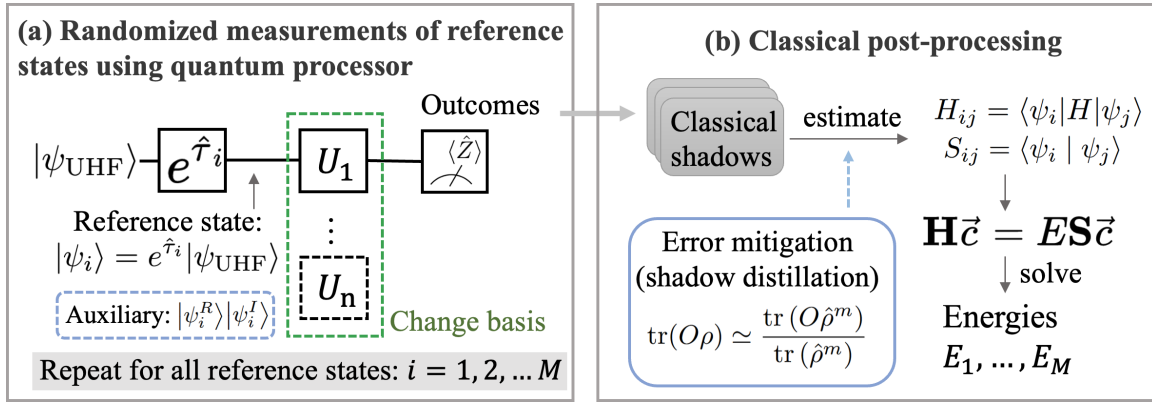


FIG. 2. Overview of the shadow-tomography-enhanced NOQE protocol. (a) The UCC ansatz $e^{\hat{\tau}_i}$ is applied to the unrestricted Hartree-Fock state $|\psi_{\text{UHF}}\rangle$ to prepare each reference state $|\psi_i\rangle$. Randomized measurements are performed on these states and their auxiliary variants by sampling unitaries $\hat{U}_1, \dots, \hat{U}_n$ from the Clifford group. This procedure is repeated for all M reference states. (b) The resulting measurement outcomes are processed into classical shadows, compactly encoding the information about the reference states. From these classical shadows, one reconstructs the Hamiltonian and overlap matrix elements H_{ij} and S_{ij} , then solves the associated generalized eigenvalue problem to obtain the energy spectrum. If needed, shadow-distillation-based error mitigation can be applied to suppress hardware noise-induced biases in the estimated matrix elements.

While Eq. (18) applies to general copies of the same density matrix, in Sec. IV C we show that for NOQE reference states, which are pure, these bounds can be further optimized. Moreover, we will introduce specialized postprocessing techniques tailored for estimating matrix elements H_{ij} and S_{ij} , since these are not simple linear functionals of a single reference state but bilinear forms involving pairs of states.

IV. APPLYING SHADOW TOMOGRAPHY TO NOQE

The original NOQE algorithm employs a modified Hadamard-test-like circuit (Fig. 1) to measure H_{ij} and S_{ij} . Here, we explore the replacement of that procedure with shadow tomography measurements, thereby reducing both the circuit depth and the number of qubits required. This approach also improves the scaling of the number of measurements n with respect to the number of reference states M , reducing it from quadratic to linear. In addition, when high precision is required for systems of moderate size, this method offers improved sample complexity compared to the original protocol. Figure 2 illustrates the pipeline of our shadow-based NOQE algorithm, which consists of two main stages: data acquisition and data postprocessing.

A. Data acquisition: Preparing states and performing randomized measurements

Because the required matrix elements H_{ij} and S_{ij} do not correspond to simple single-state observables, we must collect data from additional states to assist in their estimation. Specifically, for each reference state $|\psi_i\rangle$, we also prepare the auxiliary states

$$|\psi_i^R\rangle = \frac{|0\rangle^{\otimes N} + |\psi_i\rangle}{\sqrt{2}}, \quad |\psi_i^I\rangle = \frac{|0\rangle^{\otimes N} + i|\psi_i\rangle}{\sqrt{2}}. \quad (19)$$

Figure 3 shows an efficient circuit construction for the auxiliary states. The UHF state for n spin orbitals is

$|111 \dots 000\rangle$ (i.e., $N/2$ ones followed by $N/2$ zeros) [1]. Starting from an N -qubit zero state, we first generate an $N/2$ -qubit Greenberger-Horne-Zeilinger (GHZ) state alongside $N/2$ qubits in the zero state. Applying the UCC ansatz $e^{\hat{\tau}_i}$ [Eq. (1)] to this superposition creates

$$\begin{aligned} |\Psi_{\text{end}}^i\rangle &= \frac{1}{\sqrt{2}}(e^{\hat{\tau}_i}|0\rangle^{\otimes N} + e^{\hat{\tau}_i}|1\rangle^{\otimes N/2}|0\rangle^{\otimes N/2}) \\ &= \frac{|0\rangle^{\otimes N} + |\psi_i\rangle}{\sqrt{2}} = |\psi_i^R\rangle. \end{aligned} \quad (20)$$

Here, $|\psi_i\rangle$ arises from acting $e^{\hat{\tau}_i}$ on the $|1\rangle^{\otimes N/2}|0\rangle^{\otimes N/2}$ configuration, while $|0\rangle^{\otimes N}$ remains unchanged due to the particle-conserving nature of the UCC ansatz. To obtain $|\psi_i^I\rangle$, we can add a phase gate S on the first qubit, adding a factor of i to the second term.

The additional overhead for preparing these auxiliary states is minimal (adding n cnot gates for the GHZ state preparation), making this procedure feasible for near-term devices.

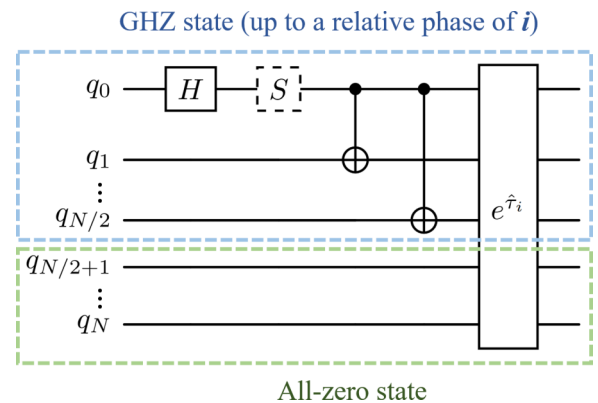


FIG. 3. Quantum circuit for preparing the auxiliary states $|\psi_i^R\rangle$ and $|\psi_i^I\rangle$.

We then perform randomized measurements (as described in Sec. III A) on the set $\{|\psi_i\rangle, |\psi_i^R\rangle, |\psi_i^I\rangle\}$ for each $i = 1, \dots, M$. This yields classical shadow datasets $\mathcal{D}_i(n)$, $\mathcal{D}_i^R(n)$, and $\mathcal{D}_i^I(n)$, each consisting of n classical shadows.

After acquiring shadows for $|\psi_i\rangle$, $|\psi_i^R\rangle$, and $|\psi_i^I\rangle$, we use the resulting datasets to estimate the Hamiltonian and overlap matrix elements, as described in the next subsection.

B. Postprocessing: Estimating matrix elements

After performing shadow tomography on each $|\psi_i\rangle$, $|\psi_i^R\rangle$, and $|\psi_i^I\rangle$, we obtain unbiased estimators of the density matrices of these states $\hat{\rho}_i = |\psi_i\rangle\langle\psi_i|$, $\hat{\rho}_i^R = |\psi_i^R\rangle\langle\psi_i^R|$, and $\hat{\rho}_i^I = |\psi_i^I\rangle\langle\psi_i^I|$ [see Eq. (16)]. These allow us to construct unbiased estimators for the matrix elements H_{ij} and S_{ij} by replacing the corresponding terms $|\psi_i\rangle\langle\psi_i|$, $|\psi_i^R\rangle\langle\psi_i^R|$, $|\psi_i^I\rangle\langle\psi_i^I|$ in the analytical formulas shown below with their shadow tomography estimators $\hat{\rho}_i$, $\hat{\rho}_i^R$, $\hat{\rho}_i^I$.

The overlap matrix elements S_{ij} in Eq. (4) can be expressed in terms of the reference and auxiliary states as

$$|S_{ij}|^2 = \text{Tr}(|\psi_i\rangle\langle\psi_i| |\psi_j\rangle\langle\psi_j|), \quad (21)$$

$$\text{Re}(S_{ij}) = 2 \text{Tr}(|\psi_i^R\rangle\langle\psi_i^R| |\psi_j^R\rangle\langle\psi_j^R|) - \frac{1}{2}(1 + |S_{ij}|^2), \quad (22)$$

$$\text{Im}(S_{ij}) = 2 \text{Tr}(|\psi_i^I\rangle\langle\psi_i^I| |\psi_j^I\rangle\langle\psi_j^I|) - \frac{1}{2}(1 + |S_{ij}|^2). \quad (23)$$

Using these relations, we first estimate $|S_{ij}|^2$ from the reference states, then determine the real and imaginary parts of S_{ij} from the auxiliary states.

Once S_{ij} is known, the Hamiltonian matrix elements follow from

$$H_{ij}S_{ji} = \text{Tr}(|\psi_j\rangle\langle\psi_j| |\psi_i\rangle\langle\psi_i| \hat{H}). \quad (24)$$

Thus, having obtained S_{ij} , we can solve for H_{ij} by calculating the quantity above. The derivations of Eqs. (21)–(24) are detailed in Appendix B.

With all H_{ij} and S_{ij} estimated, we solve the generalized eigenvalue problem in Eq. (3) to find the energies. In essence, shadow tomography transforms NOQE into a hybrid classical-quantum protocol, where the task of estimating the required matrices through quantum circuits is delegated to classical postprocessing of the collected shadows.

A detailed sample complexity analysis comparing this approach to the original NOQE method is presented in Sec. V. Here, we shall continue with quantifying the sample complexity of the protocol.

C. Sample complexity and refinement with U-statistics for pure reference states

The shadow tomography framework described in Sec. III is designed for general density matrices. However, in our NOQE application, the reference states are pure. This allows us to improve the sample complexity by constructing more sophisticated estimators based on U-statistics [30]. The U-statistics helps to produce minimum-variance unbiased estimators of the pure reference states. Note that the sample complexity in Eq. (18) corresponds to the case of U-statistics with $m = 1$.

Using $m = 2$ U-statistics, we define the estimator of $\hat{\rho}$ as

$$\hat{\rho}^{(U_2)} = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} \hat{\rho}_\alpha \hat{\rho}_\beta. \quad (25)$$

Here, n represents the number of samples, and the summation runs over all distinct pairs of samples. For pure states, $\mathbb{E}[\hat{\rho}^{(U_2)}] = \hat{\rho}$, making it an unbiased estimator of $\hat{\rho}$. Reference [31] shows that to estimate K linear observables, i.e., $\text{Tr}(\hat{\rho} \hat{O}_l)$ for $l = 1, \dots, K$ in a D -dimensional Hilbert space to within additive error ϵ using $m = 2$ U-statistics, it suffices to have a sample size of

$$n = \mathcal{O}\left(\log(K) \left\{ \frac{\sqrt{BD}}{\epsilon} + \frac{1}{\epsilon^2} \right\}\right). \quad (26)$$

This represents a notable improvement over the standard shadow tomography bound in Eq. (18). Specifically, when the first term in parentheses dominates (i.e., for large BD and ϵ), the leading term $\frac{\sqrt{BD}}{\epsilon}$ in Eq. (26) exhibits Heisenberg-like scaling with respect to ϵ , in contrast to the standard quadratic scaling. In the regime where the second term dominates (i.e., for small Hilbert space dimension D or very small ϵ), the second-order U-statistics approach still improves upon Eq. (18), as it removes the dependence on the bound B in the numerator, resulting in a more favorable constant sample complexity.

In this work, we find that using $m = 3$ U-statistics can yield even better performance

$$\hat{\rho}^{(U_3)} = \frac{1}{n(n-1)(n-2)} \sum_{\alpha \neq \beta \neq \gamma} \hat{\rho}_\alpha \hat{\rho}_\beta \hat{\rho}_\gamma. \quad (27)$$

For each observable $\hat{O}_l$, we then estimate its expectation value as $o_l = \text{Tr}[\hat{O}_l \hat{\rho}^{(U_3)}]$, which is the situation for estimating diagonal elements of the Hamiltonian H_{ii} . Following a similar derivation to that in Ref. [31], we find that the sample complexity is (see Appendix D 4 for details)

$$n = \mathcal{O}\left(\log(K) \left\{ \frac{D^{2/3} B^{1/3}}{\epsilon^{2/3}} + \frac{1}{\epsilon^2} \right\}\right). \quad (28)$$

This $m = 3$ approach further improves sample complexity for a wide range of precision parameters ϵ . While higher-order U-statistics ($m > 3$) may offer additional improvements, deriving their sample complexities becomes increasingly challenging and we restrict our analysis here to $m = 3$. We note that increasing m also increases the classical postprocessing cost: For n measurement snapshots, computing an m th order U-statistic requires $\mathcal{O}(n^m)$ evaluations of multisnapshot correlators. This scaling represents a trade-off between reduced quantum sample complexity and increased classical computational cost. However, we note that each individual classical shadow can be compactly stored in the stabilizer formalism—as originally proposed in Refs. [15,32]. This allows efficient computation of products between shadows, so the classical overhead remains tractable for moderate values of n .

A schematic illustration of how $m = 2$ U-statistics improves the original shadow estimation, and also how $m = 3$ U-statistics further improves $m = 2$ is shown in Fig. 4. The axis is divided into several ranges of ϵ . In each range, the sample complexity using different orders m (denoted by different

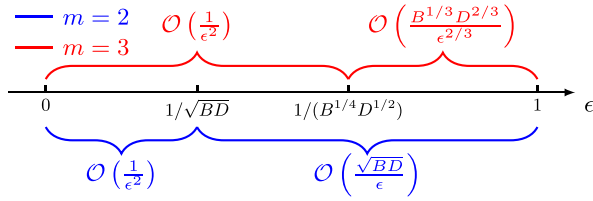


FIG. 4. The sample complexity of linear function shadow estimation can be optimized by using U-statistics of order $m = 2$ [Eq. (26)] and $m = 3$ [Eq. (28)]. Note that the common factor $\log(K)$ is omitted here for simplicity. When ϵ becomes sufficiently small ($\epsilon \leq 1/\sqrt{BD}$), the sample complexity becomes $1/\epsilon^2$ and is independent of B and D . Using higher-order U-statistics allows entry into this regime earlier at larger ϵ . A similar improvement due to U-statistics is observed in the sample complexity of shadow estimation for bilinear functions.

colors) of U-statistics is shown. For example, with $m = 2$, if $\epsilon < 1/\sqrt{BD}$, the $1/\epsilon^2$ term in Eq. (26) dominates and results in a sample complexity of $\mathcal{O}(1/\epsilon^2)$. When $\epsilon \leq \sqrt{B/D}$, employing U-statistics with $m = 2$ [Eq. (26)] or $m = 3$ [Eq. (28)] offers an improvement over the $m = 1$ case [Eq. (18)].

For the shadow estimation of bilinear functions $\text{Tr}[\hat{O}\hat{\rho}_A\hat{\rho}_B]$, such as the off-diagonal matrix elements S_{ij} and H_{ij} , U-statistics can provide improvements similar to the linear function case. With the original bilinear shadow estimation ($m = 1$), the sample complexity for estimating K observables to within an additive error ϵ is given by (see Appendix D 3)

$$n = \mathcal{O}\left(\log(K)\left\{\frac{1}{\epsilon^2} + \frac{\sqrt{DB}}{\epsilon}\right\}\right), \quad (29)$$

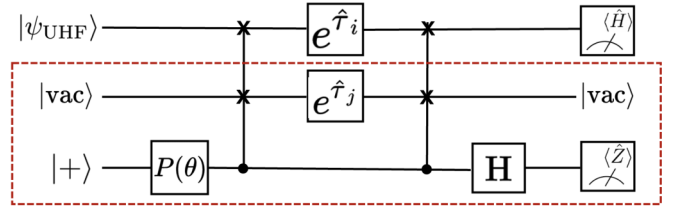
where $B \geq \text{Tr}[\hat{H}^2]$. By employing $m = 3$ U-statistics, we demonstrate in Appendix D 5 that the sample complexity is improved to

$$n = \mathcal{O}\left(\log(K)\left\{\frac{1}{\epsilon^2} + \frac{D^{5/6}B^{1/6}}{\epsilon^{1/3}}\right\}\right). \quad (30)$$

Given the clear advantage of optimizing shadow estimations using U-statistics, we use $m = 3$ U-statistics for estimating matrix elements throughout this work.

We note that the NOQE matrix elements are bilinear functionals of pairs of reference states, whereas standard classical shadows target linear observables. We therefore extend the conventional shadow tomography protocol and tailor it to our requirements here in two ways. First, on the data-acquisition side, we design the auxiliary states $|\psi_i^R\rangle, |\psi_i^I\rangle$ to avoid any joint preparation of $|\psi_i\rangle$ and $|\psi_j\rangle$, which halves both the qubit count and the circuit depth relative to that of circuits for the modified Hadamard test. Second, on the postprocessing side, we estimate the bilinear quantities with higher-order U-statistics for pure states ($m = 3$), yielding the bound in Eq. (30) and linear-in- M measurement scaling. This combination is specific to NOQE and differs from the standard shadow protocols designed for linear properties.

(a) Original NOQE circuit



(b) Shadow-based NOQE circuit $\downarrow$ removed

$$\{|\psi_i\rangle, |\psi_i^R\rangle, |\psi_i^I\rangle\} \rightarrow U \rightarrow \langle Z \rangle$$

FIG. 5. Comparison of the original and shadow-based NOQE measurement schemes, with the original scheme modified to absorb the basis rotations into the ansatz operators (see Fig. 1). (a) The original NOQE circuit employs a modified Hadamard test to measure each matrix element. (b) By using shadow tomography to characterize reference states via randomized measurements and classical postprocessing, the required circuit size and number of qubits are reduced, making the implementation less demanding.

V. CIRCUIT SIMPLIFICATION AND MEASUREMENTS REDUCTION

In this section, we compare the resource requirements and sample complexities of the original and shadow-based NOQE protocols. Table I summarizes the main findings.

For an N -qubit system with M reference states, the original NOQE protocol using the modified Hadamard test (Fig. 1) requires $2N + 1$ qubits to measure $\mathcal{O}(M^2)$ matrix elements. We note that Ref. [7] also proposed a second variant of the NOQE Hadamard test protocol that reduces the qubit count to $N + 1$ and also removes the need for controlled SWAP operations, but requires longer depth circuits. Although removing the controlled-SWAP operations provides a practical reduction in the total number of two-qubit gates, the dominant contribution to the total two-qubit gate count in the circuits remains the implementation of the UCC ansatz, which scale as $\mathcal{O}(N^2)$ with the number of spin orbitals N . Consequently, this alternative can yield a lower constant overhead of gates, without altering the asymptotic scaling of the circuit complexity. Since higher circuit depths are required for the $N + 1$ qubit variant, we focus here on analysis of the $2N + 1$ qubit variant shown in Figs. 1 and 5.

In contrast, the shadow-based NOQE protocol requires only N qubits—halving the qubit—and reduces the necessary reference state preparations from $\mathcal{O}(M^2)$ to $\mathcal{O}(M)$. This improvement arises because matrix elements are reconstructed from postprocessing classical shadows rather than directly measuring pairs of states. Consequently, the circuit is simpler, less noise prone, and more experimentally feasible. See Fig. 5 for illustrations.

In the original NOQE protocol, the two-qubit (cnot) gate count scales as $g \approx \mathcal{O}(N^2)$ [7], with the dominant contribution coming from the UCC ansatz. In the shadow-based NOQE, the random unitaries can be realized with $\mathcal{O}(\log N)$ total gates [33,34], constituting a subleading addition to the $\mathcal{O}(N^2)$ UCC contribution. Overall, the shadow-based approach halves the $\mathcal{O}(N^2)$ UCC-related gate count while introducing only a

TABLE I. Comparison between original and shadow-based NOQE methods. The shadow-based approach reduces the number of qubits, gates, and reference state preparations, resulting in relative resource reduction and improved sample complexity in the regime of high precision ϵ , a large number of reference states M , and a moderate number of qubits N .

	Circuit size		Sample complexity		
	Qubits	Gates	No. of reference states (M)	Dimension: $D = 2^N$	Optimal regime
Original NOQE	$2N + 1$	g	$\mathcal{O}(M^2)$	$\mathcal{O}(\text{polylog}(D)/\epsilon^2)$, Eq. (11)	Larger dimension
Shadow-based NOQE	N	$\approx g/2$	$\mathcal{O}(M)$	$\mathcal{O}(1/\epsilon^2 + D/\epsilon^{1/3})$, Eq. (30)	More reference states, high precision

minimal $\mathcal{O}(\log N)$ overhead for the random Clifford layers, yielding a net reduction in the total number of two-qubit gates.

In addition to improvements in scaling with M , we also examine how performance depends on $D = 2^N$. In the worst-case scenario where $B \sim D$, the sample complexity in Eq. (30) simplifies to

$$n = \mathcal{O}\left(\frac{1}{\epsilon^2} + \frac{D}{\epsilon^{1/3}}\right) \quad (31)$$

for estimating bilinear functions using $m = 3$ U-statistics. Notably, there is no dependence on K , as we estimate a constant number of bilinear functions (S_{ij} and H_{ij}) for each pair of reference states.

Both original and shadow-based NOQE outperform standard quantum state tomography (QST) protocols with better scaling in terms of the dimension D , as summarized in Table II.

We now compare the sample complexity of shadow-based NOQE with the original NOQE. Recall that the sample complexity of the original NOQE is $\mathcal{O}(M^2 \text{polylog}(D)/\epsilon^2)$. In contrast, the shadow-based NOQE with $m = 3$ U-statistics requires $\mathcal{O}(M\{1/\epsilon^2 + D/\epsilon^{1/3}\})$ measurements. The shadow-based NOQE thus outperforms the original NOQE in the small parameter regime. For example, with $M = 30$ and $N = 20$, which are reasonable values for near-term calculations, the shadow-based NOQE outperforms the original NOQE if the desired precision $\epsilon < 0.068$, a regime consistent with chemical-accuracy requirements. The sample complexity behavior is illustrated in Fig. 6, with respect to M and ϵ for fixed N in panel (a), and with respect to N and ϵ for fixed B in panel (b). In Fig. 6, the color scale encodes the logarithm of the ratio of sample complexities for the shadow-based approach to the sample complexity for the original NOQE approach, where each sample complexity is evaluated as the product of

the following two factors from Table I: (1) the scaling with the number of reference states M and (2) the scaling with the Hilbert-space dimension D . The shadow-based bound for (2) in Eq. (30) includes a $\log K$ factor from the randomized measurements, where K is the number of functions of ρ estimated simultaneously. Here, $K = 2$ (overlap and Hamiltonian elements for each pair of reference states), so $\log K = \log 2$ is a constant of $\mathcal{O}(1)$, independent of system size. The complexity ratios were then computed over the ranges of values of M , N , and ϵ shown in the figure, omitting this constant prefactor (which is < 1 and would only further favor the shadow-based protocol for small systems). The NOQE bound for (2) in Eq. (11) is deterministic and does not have any prefactor.

In the high-precision regime, the shadow-based complexity reduces to $\mathcal{O}(M/\epsilon^2)$, independent of D . This is ideal for NOQE, where high accuracy is important for capturing complex electronic correlations. For $\epsilon \sim 10^{-3}$, the advantage of shadow-based NOQE persists up to systems with ~ 20 qubits. We emphasize that our target is the NISQ/pre-fault-tolerant regime. For typical problem instances possible with ~ 20 qubits, the linear-in- D factor in Eq. (31) is not prohibitive, while the goal of achieving chemical accuracy (~ 1.6 mHa) dominates the cost. In this regime, the shadow-based NOQE achieves lower total sample complexity and uses fewer qubits with shallower circuits than the original NOQE. However, we point out that for N values larger than 30, the exponentially scaling term in the sampling complexity of the shadow protocol comes into play and the original NOQE measurement protocol then becomes more efficient, indicating that in the full fault-tolerant regime with $N \gtrsim 50$ qubits, the original NOQE protocol will be preferred.

Overall, the incorporation of shadow tomography into NOQE provides a substantial resource advantage in qubits, gates, and total measurement cost in the case of small to intermediate N for near-term applications. It enables the protocol to scale more favorably with the number of reference states and improves performance under demanding precision requirements.

VI. SHADOW DISTILLATION

While reducing the measurement overhead is crucial to design efficient quantum algorithms, measurement itself is a useful resource in quantum computing [35,36]. In near-term quantum hardware implementations, mitigating the effects of noise is necessary to maintain the full potential of quantum computers [37]. The flexible measurement scheme of shadow tomography makes it well suited for integrating classical error mitigation protocols. Here, we adopt one such technique, known as *shadow distillation* [16].

TABLE II. Sample complexity comparison of various measurement protocols. Here, $D = 2^N$ denotes the dimension of the N -qubit system. Quantum state tomography is shown for both individual (measuring a single copy at a time) and joint (measuring multiple copies simultaneously) measurement scenarios, and both NOQE protocols have better sample complexity than state-of-the-art QST methods.

Protocol	Sample complexity
QST (independent)	D^3/ϵ^2
QST (joint)	D^2/ϵ^2
Original NOQE	$\text{polylog}(D)/\epsilon^2$
Shadow-based NOQE	$1/\epsilon^2 + D/\epsilon^{1/3}$

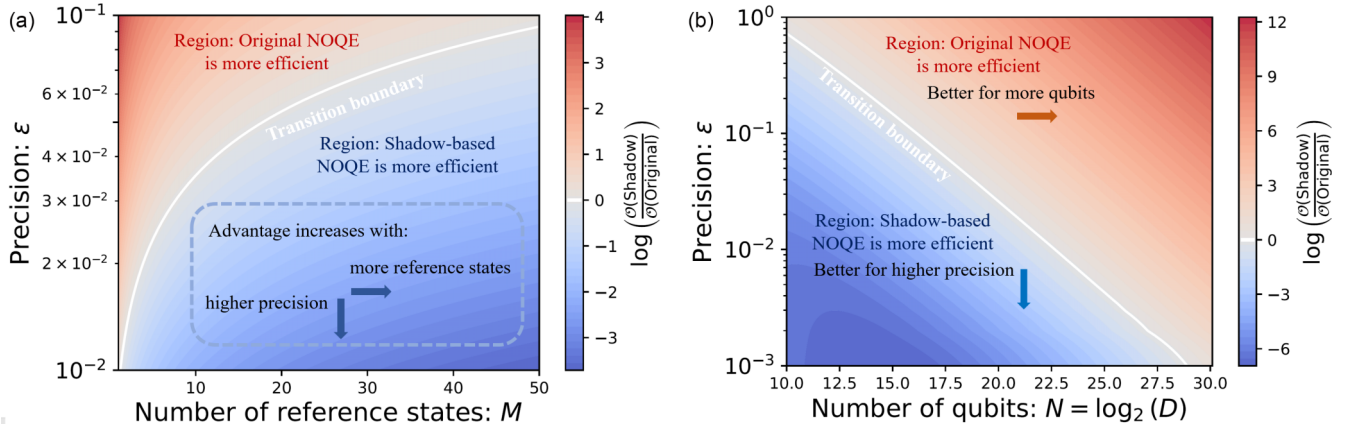


FIG. 6. Sample complexity comparison between the original and shadow-based NOQE protocols. The color scale represents the logarithmic ratio of computational cost, where red regions indicate that the original NOQE is more efficient, while blue regions indicate that the shadow-based NOQE is more efficient. (a) Varying precision ϵ and the number of reference states M , with the number of qubits fixed at $N = 20$. (b) Varying precision ϵ and the number of qubits $N = \log_2(D)$, with the number of reference states fixed at $M = 6$. The white contour represents the transition boundary where both methods have comparable sample complexity. The shadow-based NOQE approach exhibits advantages in the high-precision regime and when the number of reference states is large, and the original NOQE is more efficient for larger numbers of qubits beyond $N \approx 30$.

Consider a target state $|\psi\rangle$ that, due to experimental imperfections such as gate errors, is instead prepared as a noisy density matrix

$$\hat{\rho} = (1 - \epsilon)|\psi\rangle\langle\psi| + \epsilon\hat{\rho}_{\text{error}}, \quad (32)$$

where ϵ characterizes the noise level and $\hat{\rho}_{\text{error}}$ represents the error component that is approximately orthogonal to the target state.

A direct estimate of $\langle\psi|\hat{O}|\psi\rangle$ by computing $\text{Tr}(\hat{O}\hat{\rho})$ is biased. Shadow distillation addresses this by considering

$$\langle\hat{O}\rangle_{(m)} = \frac{\text{Tr}(\hat{O}\hat{\rho}^m)}{\text{Tr}(\hat{\rho}^m)} \quad (33)$$

for a positive integer m . Reference [16] shows that

$$\frac{\text{Tr}(\hat{O}\hat{\rho}^m)}{\text{Tr}(\hat{\rho}^m)} \approx \langle\psi|\hat{O}|\psi\rangle \left[1 + \mathcal{O}\left(\left(\frac{\epsilon}{1-\epsilon}\right)^m\right) \right]. \quad (34)$$

Under weak noise conditions, raising $\hat{\rho}$ to the m th power exponentially suppresses the error component [16,38,39], effectively “distilling” the pure state contribution. However, increasing m also enhances statistical fluctuations from finite sampling, so that there is a trade-off in implementing U-statistics, between achieving bias reduction (which improves exponentially with the moment order m for common noise models) and increasing statistical variance [which grows with m because of the need to estimate higher-order quantities $\text{Tr}(\hat{\rho}^m\hat{O})$ and $\text{Tr}(\hat{\rho}^m)$]. For example, with $m = 4$, the classical postprocessing costs increase to $\mathcal{O}(n^4)$ (see Sec. IV C). We do not attempt a closed-form optimization over m , since the optimal choice depends on the hardware noise channel; instead, we have established empirically that $m = 3$ strikes a good balance between bias suppression and variance inflation, making it a practical choice for our setting.

Implementing shadow distillation with classical shadows is straightforward: We construct $\hat{\rho}^m$ from the U-statistics-

based estimators $\hat{\rho}^{U_m}$ of $\hat{\rho}$. Since we are already employing $m = 3$ U-statistics to estimate $\{\hat{\rho}_i, \hat{\rho}_i^R, \hat{\rho}_i^I\}$, we simply replace these estimators with their normalized versions $\{\hat{\rho}_i/\text{Tr}[\hat{\rho}_i], \hat{\rho}_i^R/\text{Tr}[\hat{\rho}_i^R], \hat{\rho}_i^I/\text{Tr}[\hat{\rho}_i^I]\}$ when evaluating matrix elements in Eqs. (21)–(24). This yields error-mitigated estimates for S_{ij} and H_{ij} .

Because shadow distillation is a purely classical post-processing technique, it requires no additional quantum resources. The classical shadows already contain all necessary state information to apply advanced error mitigation. While we focus on shadow distillation here, other known error mitigation methods [40–44] can be similarly integrated.

Incorporating shadow distillation into our protocol enhances the robustness of the estimated energies against device noise, advancing the practical feasibility of accurate quantum chemistry simulations on near-term quantum hardware. We will evaluate the performance of the shadow distillation developed here in Sec. VIII.

VII. VALIDATING THE CORRECTNESS AND EFFICIENCY OF THE PROTOCOL

We validate the accuracy and efficiency of our protocol by applying it to the hydrogen molecule (H_2) just past the Coulson-Fischer point (interatomic spacing = 1.2 Å), a regime characterized by electronic spin-symmetry breaking and strong correlation [45]. This represents the most challenging case for computation and can be considered a lower bound on the performance of our protocol.

We use the STO-3G basis [46] for a minimal orbital representation, leading to a UHF reference state $|\psi_{\text{UHF}}\rangle = |1100\rangle$ and two reference states $|\psi_1\rangle = e^{\hat{\tau}_1}|\psi_{\text{UHF}}\rangle$ and $|\psi_2\rangle = e^{\hat{\tau}_2}|\psi_{\text{UHF}}\rangle$. Each state encodes four spin orbitals into four qubits, i.e., $N = 4$. Under these conditions, the Hamiltonian

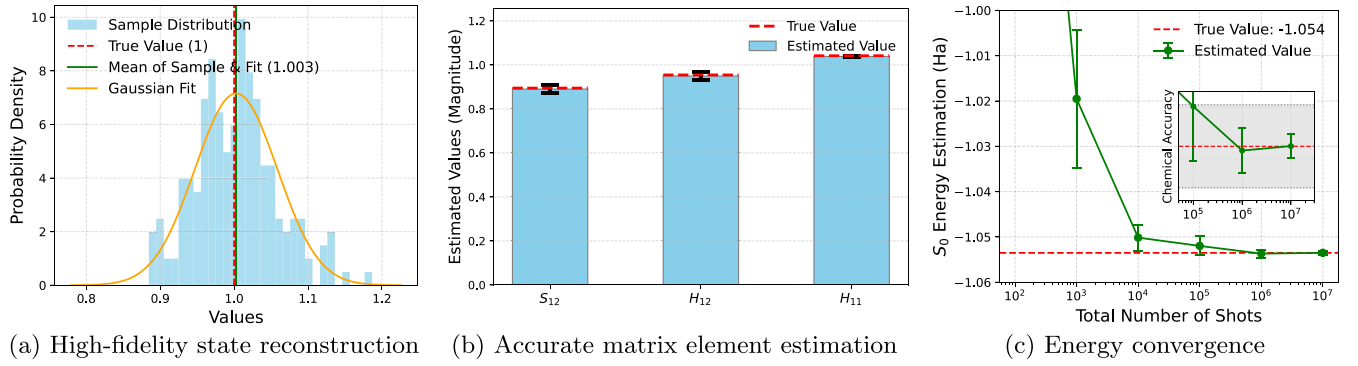


FIG. 7. Performance evaluation of the shadow-based NOQE protocol for the ground electronic state of the hydrogen molecule at $R = 1.2 \text{ \AA}$, in the STO-3G basis. The true value here is the value of the noiseless simulation of NOQE [7]. These calculations include sampling (shot) noise but no hardware noise. (a) Fidelity distribution of the estimated reference state $|\hat{\psi}_1\rangle$ compared to the true state $|\psi_1\rangle$. Values close to unity confirm unbiased state reconstruction. (b) Histograms of estimated Hamiltonian and overlap matrix elements, showing that classical shadow postprocessing yields unbiased, precise results. (c) Convergence of the estimated ground-state energy to the exact value as the number of shots increases, demonstrating the correctness and efficiency of the protocol.

and overlap matrices have the form

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12}^* & h_{22} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 1 & s_{12} \\ s_{12}^* & 1 \end{pmatrix}. \quad (35)$$

Note that in this setting, the values to be estimated are h_{11} , h_{12} , h_{22} , and s_{12} , i.e., the number of observables $K = 4$. Solving the generalized eigenvalue problem in Eq. (3) provides both the ground and excited state energies. Here, we focus on the ground-state energy, a primary target in quantum chemistry. We compare against the noiseless simulation results of the original NOQE protocol as a ground-truth reference (see Appendix A for NOQE implementation details).

In Sec. IV, we established that our protocol yields unbiased energy estimates in the presence of shot noise alone, i.e., no hardware noise. Figure 7 numerically confirms this. In Fig. 7(a), we assess the fidelity of reconstructed reference states. Each fidelity value is estimated from $n = 10\,000$ classical shadows, i.e., 10 000 independent runs of the circuit in Fig. 2. We note that the deviation from perfect fidelity arises from statistical fluctuations, but the results remain close to unity, confirming accurate state reconstruction and reliable observable estimation.

Figure 7(b) shows histograms of the estimated matrix elements. The mean and standard deviations are calculated over 100 independent samples, each employing 10 000 shadows. The estimates closely match the true values, with small error bars, indicating unbiased and precise matrix element determination.

Figure 7(c) demonstrates convergence of the energy estimate to the exact value as the number of shots increases. Around 10^6 shots (i.e., 10^6 classical shadows) suffice to achieve chemical accuracy, defined as an absolute energy error below 1.6×10^{-3} hartree. Given that the exact ground-state energy is -1.054 hartree, this corresponds to a relative error of approximately 0.15%, making the estimate suitable for chemically relevant predictions [47]. This level of performance—achieved with a strongly correlated system known to be challenging—highlights the protocol’s practical effectiveness.

Beyond correctness, we compare circuit complexity. After circuit optimizations, the original NOQE method requires nine qubits, 124 two-qubit gates, and about 268 single-qubit gates. See Appendix A 5 for details. In contrast, our shadow-based approach only needs four qubits, about 54 two-qubit gates, and 110 single-qubit gates. Cutting the number of qubits and gates roughly in half significantly reduces the experimental complexity and the impact of noise.

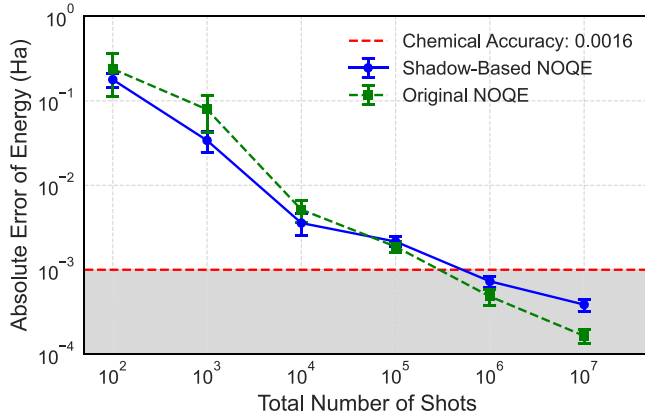
In terms of measurement shots, our shadow-based NOQE method, despite its reduced circuit size, performs equivalently well with the original protocol. Figure 8 plots the absolute error and variance in H_2 energy estimation versus total shots. The error bars are determined based on ten repetitions of the same experiment. Both methods reach chemical accuracy with a similar number of shots, but the shadow-based method is simpler to implement. We remark that achieving chemical accuracy for this strongly correlated four spin-orbitals system with about 10^6 shots is on par with state-of-the-art proposals [13].

Overall, these results confirm that our shadow-tomography-enhanced NOQE protocol reliably produces accurate and precise energy estimates with fewer experimental resources, making it a promising approach for practical quantum chemistry simulations on near-term quantum devices.

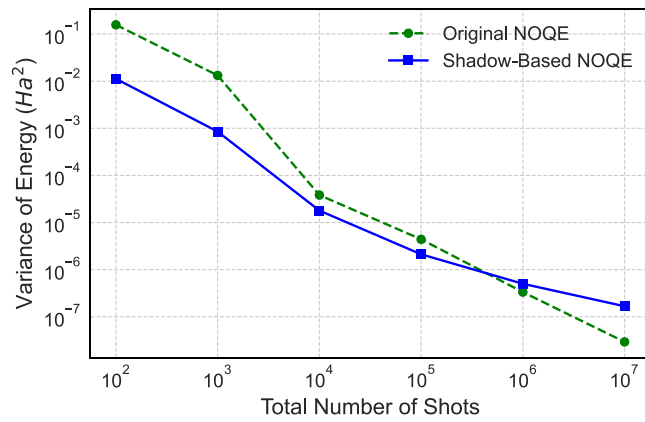
VIII. PERFORMANCE OF THE PROTOCOL UNDER NOISE

In the previous sections, we showed that our estimator is unbiased in the absence of hardware noise, where the only noise present is statistical, which can be reduced by simply increasing the number of measurement shots. Real quantum devices, however, introduce systematic noise through gate errors that cannot be removed merely by taking more samples. Here, we evaluate the protocol’s performance under realistic noise conditions representative of NISQ devices [48].

To simulate hardware noise, we use a noise model informed by the specifications of the Quantinuum System Model H2 quantum processor [22]. The composite noise



(a)

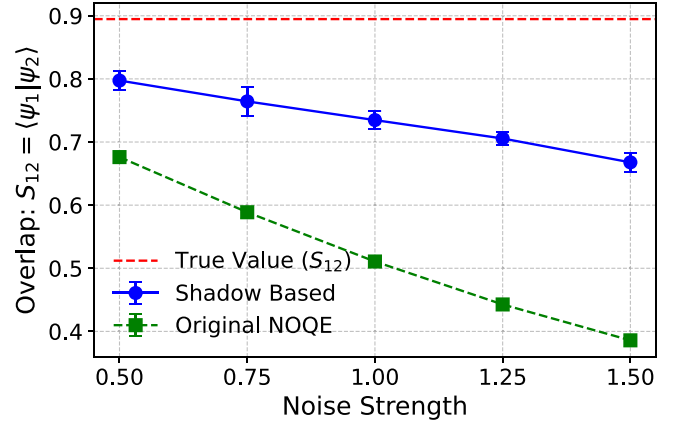


(b)

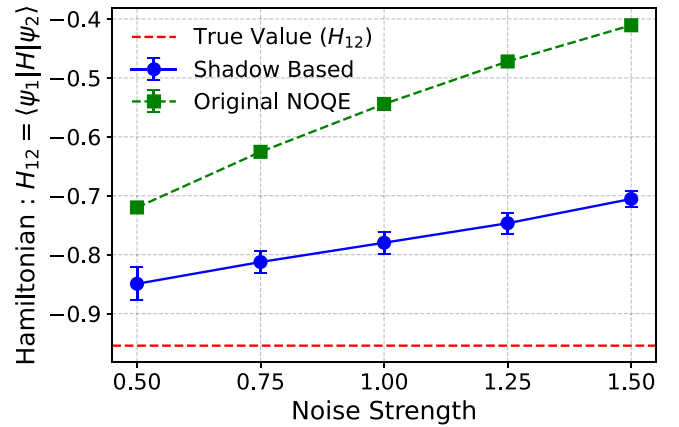
FIG. 8. Comparison of (a) the absolute error and (b) the variance in energy estimation between the shadow-based NOQE protocol with $m = 3$ U-statistics and the original NOQE method, for the ground state of the hydrogen molecule at $R = 1.2 \text{ \AA}$, in the STO-3G basis, including shot noise but no hardware noise. Both approaches achieve chemical accuracy with a similar number of shots, but the shadow-based protocol uses half the number of qubits and gates, reducing experimental complexity and increasing feasibility for near-term implementations.

model we used incorporates depolarizing, amplitude damping, and phase damping channels [49], calibrated to reported single-qubit and two-qubit gate error rates of 3×10^{-5} and 1.5×10^{-3} . We introduce a scaling factor λ to uniformly adjust all noise rates and study the protocol's behavior across varying noise strengths. The unitary operators in the NOQE algorithm were decomposed and compiled into quantum circuits consisting of single-qubit rotations and cnot gates. Noise models, including depolarizing, amplitude damping, and phase damping channels, were applied at the level of one- and two-qubit gates to simulate realistic execution. We summarize the noise model in Tables III and IV.

We now simulate the protocols outlined in Secs. IV and VI under these conditions of both shot noise and hardware noise. As shown in Fig. 9, the shadow-based protocol remains closer to the true values of both overlap and Hamiltonian



(a) Overlap matrix element



(b) Hamiltonian matrix element

FIG. 9. Calculations for the ground state of the hydrogen molecule at $R = 1.2 \text{ \AA}$, in the STO-3G basis with hardware noise in addition to shot noise, showing the robustness of the shadow-based approach with $m = 3$ U-statistics in the presence of hardware noise. Under hardware noise levels scaled by factors of 0.5, 0.75, 1, 1.25, and 1.5 relative to the Quantinuum System Model H2 error rates, we compare estimation of (a) the overlap and (b) the Hamiltonian matrix elements between the shadow-based NOQE protocol and the original NOQE method, both without error mitigation. The shadow-based approach consistently produces estimates closer to the true values, demonstrating improved resilience to hardware noise due to reduced circuit complexity.

off-diagonal matrix elements compared to the original NOQE method, across a range of noise strengths. This enhanced noise robustness arises partly from the simplified circuit structure: By using half as many qubits and gates, the shadow-based approach reduces error accumulation, improving overall fidelity in noisy environments.

While the shadow-based protocol exhibits inherent noise robustness, as seen in Fig. 9, this intrinsic resilience only partially mitigates the impact of noise. To further enhance the accuracy of our protocol, we now apply shadow distillation as described in Sec. VI.

In Fig. 10, we present the results of energy estimation with and without the quantum error mitigation provided by shadow distillation to level $m = 3$. It is evident that the shadow dis-

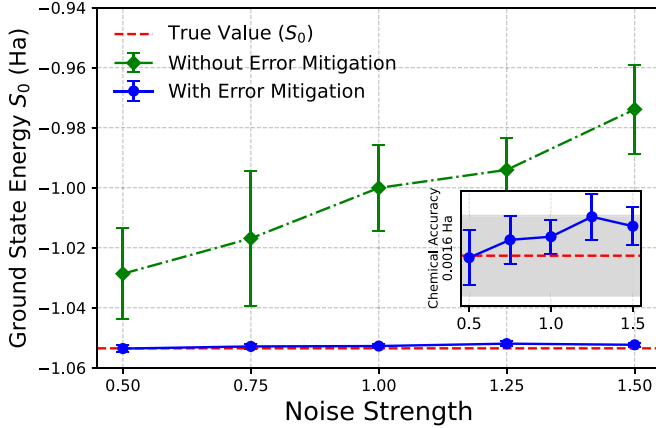


FIG. 10. Energy estimation of the ground state of the hydrogen molecule at $R = 1.2 \text{ \AA}$, in the STO-3G basis, with and without error mitigation using shadow distillation with $m = 3$ U-statistics. Without mitigation (green), noise biases the energy estimates away from the true value (red). With shadow distillation (blue), the noise-induced bias is significantly removed, bringing the estimates back within chemical accuracy = 0.0016 Hartree, gray band in the enlarged view across a broad range of noise strengths.

tilation effectively suppresses noise-induced biases. Without mitigation, energy estimates drift from the true value and exhibit larger variance; with shadow distillation, they remain near the true energy within chemical accuracy. The performance remains stable across a broad range of noise regimes, demonstrating the protocol's robustness and adaptability for quantum processors with varying noise characteristics.

To benchmark our approach against alternative quantum error mitigation approaches, in Fig. 11 we compare its performance against that of zero-noise extrapolation (ZNE) [50], a widely used error mitigation method that requires additional circuits at artificially increased noise levels [51]. Using the

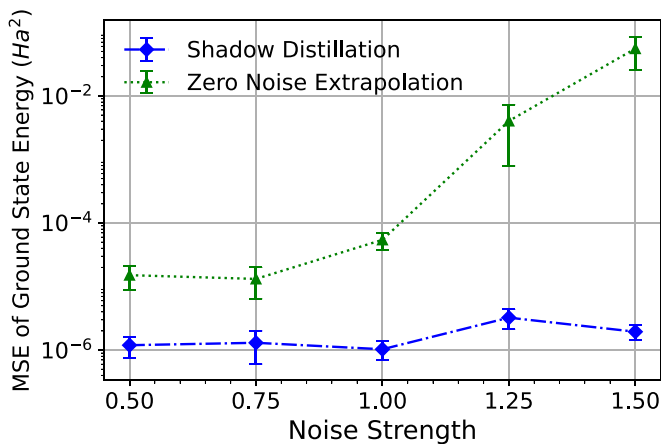


FIG. 11. MSE comparison for the ground state of the hydrogen molecule at $R = 1.2 \text{ \AA}$, in the STO-3G basis. The shadow-based NOQE with shadow distillation and $m = 3$ U-statistics (blue) outperforms the original NOQE method with ZNE (green), achieving lower MSE and indicating better performance in reducing noise-induced bias while also reducing experimental resources.

Mitiq library [52], we implement ZNE on the original NOQE method (see Appendix C for details) and compare its performance to the shadow-based NOQE with shadow distillation using $m = 3$ statistics. We measure performance here via the mean squared error (MSE), i.e., $\sum_i^n (y_i - y_{\text{true}})^2 / n$, which captures both bias and variance. Figure 11 shows that the shadow-based approach with shadow distillation yields lower MSE than the original NOQE with ZNE, indicating more efficient noise suppression together with reduced resource requirements.

In summary, the shadow-based NOQE protocol reduces the circuit complexity and is thereby inherently robust to a portion of the noise. We have found that integrating shadow distillation in the protocol further improves accuracy, effectively removing hardware noise-induced bias. Compared to ZNE, our approach achieves a lower MSE in energy estimates. This highlights the effectiveness of the shadow-based protocol in noisy quantum computing environments.

IX. SUMMARY AND OUTLOOK

We have introduced a shadow-tomography-enhanced NOQE protocol that reduces qubit requirements, circuit depth, and measurement overhead. Our rigorous sample complexity analysis shows a quadratic speedup with respect to the number of reference states, relative to the original NOQE measurement protocol. Furthermore, in the high-precision regime, the complexity no longer depends on the system size, making the protocol particularly well suited for accurate electronic structure energy estimation on moderate-size systems. Numerical simulations of the hydrogen molecule confirm unbiased and highly accurate energy estimations.

In addition to improving sample complexity and resource requirements, the shadow-based protocol exhibits enhanced noise robustness due to its simpler circuit structure. It also naturally supports error mitigation through shadow distillation, effectively suppressing noise-induced biases without additional quantum overhead. Our results show that this error-mitigated approach outperforms common techniques like ZNE, achieving chemical accuracy in estimating the hydrogen molecule's ground-state energy with as few as 10^6 shots per reference state under realistic NISQ conditions.

While we have focused on NOQE, our methods can be extended to other problems that involve estimating overlaps or Hamiltonian matrix elements involving multiple states. If fault-tolerant quantum computing becomes accessible, our protocol could be further optimized by performing joint measurements on entangled reference states and incorporating quantum-gradient estimation algorithms [53,54].

We also note that techniques such as generalized quantum subspace expansions [55] may further enhance error suppression by considering linear combinations of ρ^m at multiple values of m rather than a single m . Exploring this and other advanced error mitigation strategies within the shadow tomography framework could yield even more noise-resilient algorithms for quantum chemistry.

Future investigations into larger and more complex molecular systems, as well as experimental demonstrations on actual quantum hardware, will help further assess the scalability and practicality of the shadow-based approach.

Overall, the noise-resilient, shadow-based NOQE protocol offers a promising path toward practical, high-accuracy quantum algorithms solving electronic structure problems in near-term quantum devices.

ACKNOWLEDGMENTS

The authors thank Unpil Baek, Oskar Leimkuhler, Brandon Schramm, James Shee, Torin Stetina, and William J. Huggins for their valuable discussions. This work was partially supported by a joint development agreement between UC Berkeley and Dow, by the National Science Foundation (NSF) Quantum Leap Challenge Institutes (QLCI) program through Grant No. OMA-2016245, and by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. W.M.B. was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2146752.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY

The data that support the findings of this article are openly available [56].

APPENDIXES

The Appendixes are organized as follows: Appendix A provides a detailed explanation of the working principle and implementation of the NOQE framework, which were too extensive to include in the main text. Appendix B discusses the estimation of matrix using the shadow-based NOQE protocol. Appendix C summarizes the noise model and the implementation of ZNE. Appendix D details the sample complexity derivation that occurs in the paper.

APPENDIX A: NOQE ALGORITHM AND IMPLEMENTATION

The pipeline of implementing the original NOQE algorithm is illustrated in Fig. 12, with details of each step in the subsections below.

1. Decomposition of ansatz

We use the UCCD ansatz to prepare the reference state

$$|\psi_{\text{UCCD}}\rangle = e^{\hat{\tau}} |\psi_{\text{UHF}}\rangle, \quad (\text{A1})$$

$|\psi_{\text{UHF}}\rangle$ is the unrestricted Hartree-Fock state, $\hat{\tau} = \hat{T} - \hat{T}^\dagger$ where

$$\hat{T} = \sum_{pqrs=1}^N t_{ps,qr} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \quad (\text{A2})$$

with $r, s \in \text{occ}; p, q \in \text{virt}$. The standard UCCD amplitudes $t_{ps,qr}$ can be approximated by the MP2 theory. More specifically, we can decompose the UCCD operator $e^{\hat{\tau}}$ into a sum of squares of normal operators using a singular value decomposition:

$$e^{\hat{\tau}} = \exp \left(-i \sum_{l=1}^L \sum_{\mu=1}^m \hat{Y}_{l,\mu}^2 \right). \quad (\text{A3})$$

The normal operators are then trotterized to obtain a low-rank representation [24]

$$e^{\hat{\tau}} \approx \hat{U}_B^{(1,1)\dagger} \prod_{l=1}^L \prod_{\mu=1}^m \exp \left(-i \sum_{pq}^{\rho_l} \lambda_p^{(l,\mu)} \lambda_q^{(l,\mu)} \hat{n}_p \hat{n}_q \right) \tilde{U}_B^{(l,\mu)}, \quad (\text{A4})$$

where $\lambda_p^{(l,\mu)}$ are eigenvalues of the $\hat{Y}_{l,\mu}$ operators. $\tilde{U}_B^{(l,\mu)}$ are sequences of neighboring basis rotations. The circuit implementation of the ansatz is illustrated in Fig. 13. Details of the derivation can be found in the original NOQE paper [7].

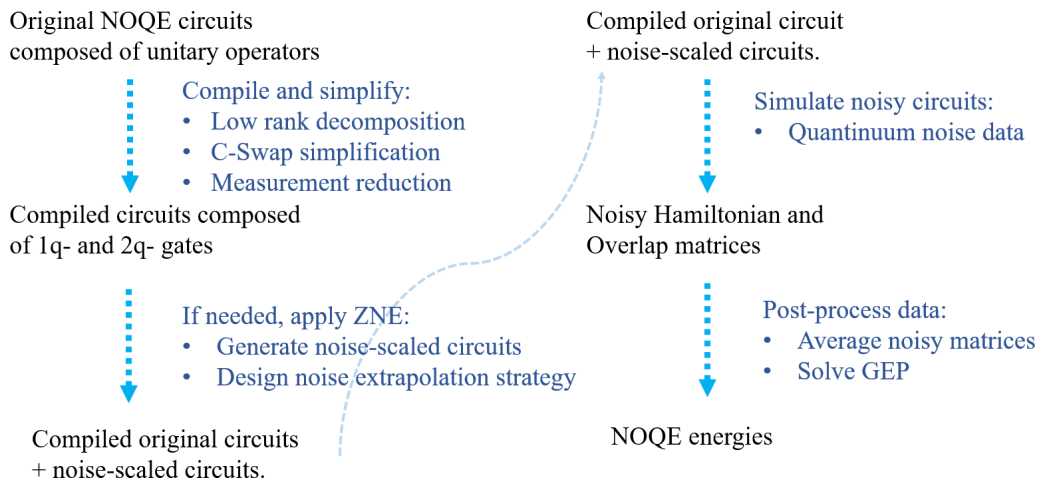


FIG. 12. Working flow of implementing the NOQE algorithm

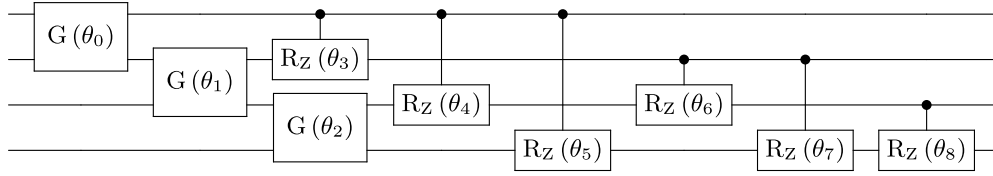


FIG. 13. Basic circuit subroutine for implementation of the UCCD operator $e^{\hat{\tau}}$ for H_2 in the STO-3G basis. $G(\theta)$ are Givens rotations that realize the $\tilde{U}_B^{(l,\mu)}$ basis rotation terms in Eq. (A4). The controlled-Z rotations realize the $\hat{n}_p\hat{n}_q$ number operator terms. Four consecutive subroutines of this form realize a single $e^{\hat{\tau}}$ ansatz for H_2 in the STO-3G basis. Single-qubit gates are omitted here for simplicity.

2. Circuit construction and Hadamard test

The NOQE circuit for evaluation of elements of the Hamiltonian and overlap matrices is given in Fig. 14. Note that in this work, we absorb the unitary orbital rotation operators ($\hat{U}_{i \rightarrow 1}$ and $\hat{U}_{j \rightarrow 1}$) that are required to transform $|\psi_i\rangle$ and $|\psi_j\rangle$ to the same single-particle basis of Ref. [7] into the corresponding $e^{\hat{\tau}_i}$ and $e^{\hat{\tau}_j}$ circuits. Here, we show how to estimate the quantum states and energies of the electronic system of interest from this circuit.

The circuit starts with system qubits in the UHF state, ancilla qubits in the all-zero state, and an extra control qubit in the $|+\rangle$ state. To begin, a phase gate with angle θ is applied to the control qubit, to create state $|\phi_1\rangle$:

$$|\phi_1\rangle = \frac{1}{\sqrt{2}}(|\psi_{\text{UHF}}\rangle|\text{vac}\rangle|0\rangle + e^{i\theta}|\psi_{\text{UHF}}\rangle|\text{vac}\rangle|1\rangle). \quad (\text{A5})$$

An N -qubit controlled SWAP (CSWAP) gate is then applied to transfer the system state to the ancilla register, and then distinct UCCD operators $e^{\hat{\tau}_i}$ and $e^{\hat{\tau}_j}$ are applied to the system and ancilla states. These are followed by application of basis rotations $\hat{U}_{i \rightarrow 1}$ and $\hat{U}_{j \rightarrow 1}$ to system and ancilla registers, respectively, in order to bring their states to a common basis of the UHF reference state $i = 1$. The resulting state $|\phi_2\rangle$ is

$$|\phi_2\rangle = \frac{1}{\sqrt{2}}(|\psi_i\rangle|\text{vac}\rangle|0\rangle + e^{i\theta}|\text{vac}\rangle|\psi_j\rangle|1\rangle), \quad (\text{A6})$$

where the reference states $|\psi_i\rangle$ and $|\psi_j\rangle$ are now expressed in the common basis.

A second CSWAP conditioned on the ancilla qubit is then applied, which brings the system register into a superposition of both reference states now in the same basis, and returns the ancilla register to $|\text{vac}\rangle$. From now on, the ancilla register can be ignored and we will write the final state $|\phi_4\rangle$ in terms of

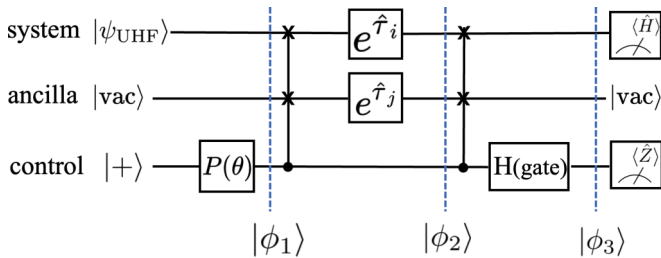


FIG. 14. Original NOQE circuit, modified to absorb the basis rotations into the ansatz operators $e^{\hat{\tau}_{(i)}}$. The quantum states in different stages of the circuits are marked and written in Eqs. (A5)–(A7).

just the system and control qubit:

$$|\phi_3\rangle = \frac{1}{\sqrt{2}}(|\psi_i\rangle|+\rangle + e^{i\theta}|\psi_j\rangle|-\rangle). \quad (\text{A7})$$

We now address the measurement of Hamiltonian and overlap matrix elements between $|\phi_i\rangle$ and $|\phi_j\rangle$. Setting the angle θ of the phase gate to be zero and measuring the control qubit in the Z basis results in the following postmeasurement states in the system register:

$$\begin{aligned} & \frac{|\phi_i\rangle + |\phi_j\rangle}{\sqrt{2 + 2\text{Re}\langle\phi_i|\phi_j\rangle}}, \\ & \text{if } Z = +1, \text{ with probability } P(Z = +1) = \frac{1 + \text{Re}\langle\phi_i|\phi_j\rangle}{2}, \\ & \frac{|\phi_i\rangle - |\phi_j\rangle}{\sqrt{2 - 2\text{Re}\langle\phi_i|\phi_j\rangle}}, \\ & \text{if } Z = -1, \text{ with probability } P(Z = -1) = \frac{1 - \text{Re}\langle\phi_i|\phi_j\rangle}{2}. \end{aligned} \quad (\text{A8})$$

This implies that the measurement statistics of Pauli Z in the control qubit give the real part of the overlap matrix element, i.e., $\langle\hat{I} \otimes \hat{Z}\rangle = \text{Re}\langle\phi_i|\phi_j\rangle$.

To estimate the Hamiltonian matrix element H_{ij} , we further measure the $\hat{H}$ in the system register, after this measurement of the control qubit in the $\hat{Z}$ basis. This is done by decomposing the Hamiltonian into a sum of Pauli terms

$$\hat{H} = \sum_k \omega_k \hat{P}_k \quad (\text{A9})$$

and measuring each Pauli term $\hat{P}_k$ in the system register. Measurements of any of these Pauli terms $\hat{P}_k$ on the postmeasurement states (A8) will generate the expectation values:

$$\begin{aligned} \langle\hat{P}_k\rangle^+ &= \frac{\langle\phi_i|\hat{P}_k|\phi_i\rangle + \langle\phi_j|\hat{P}_k|\phi_j\rangle + 2\text{Re}\langle\phi_i|\hat{P}_k|\phi_j\rangle}{2 + 2\text{Re}\langle\phi_i|\phi_j\rangle}, \\ & \text{if } Z = +1, \text{ with probability } P(Z = +1) = \frac{1 + \text{Re}\langle\phi_i|\phi_j\rangle}{2}, \\ \langle\hat{P}_k\rangle^- &= \frac{\langle\phi_i|\hat{P}_k|\phi_i\rangle + \langle\phi_j|\hat{P}_k|\phi_j\rangle - 2\text{Re}\langle\phi_i|\hat{P}_k|\phi_j\rangle}{2 - 2\text{Re}\langle\phi_i|\phi_j\rangle}, \\ & \text{if } Z = -1, \text{ with probability } P(Z = -1) = \frac{1 - \text{Re}\langle\phi_i|\phi_j\rangle}{2}. \end{aligned} \quad (\text{A10})$$

The total average of these expectation values then gives the real part of the matrix element of $\hat{P}_k$, i.e., $\langle\hat{P}_k\rangle = \text{Re}\langle\phi_i|\hat{P}_k|\phi_j\rangle$.

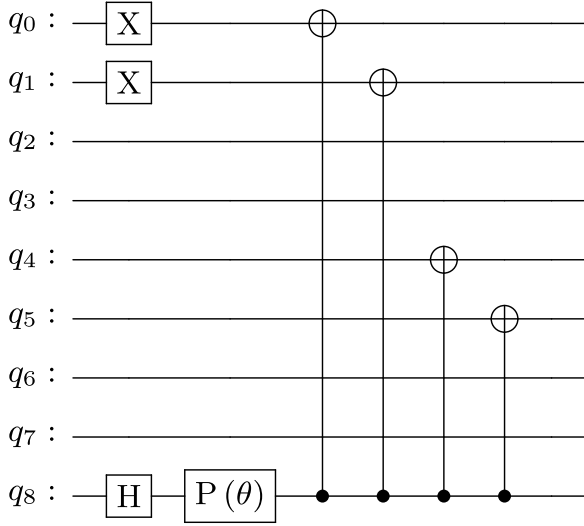


FIG. 15. Simplified implementation of the first CSWAP network for calculations on H_2 in the STO-3G basis.

Adding these $\langle \hat{P}_k \rangle$ according to the Hamiltonian decomposition weights ω_k gives the real part of the Hamiltonian matrix $\sum_k \langle \hat{P}_k \rangle = \text{Re} \langle \phi_i | \hat{H} | \phi_j \rangle$.

Setting θ to $\pi/2$, we can similarly get the imaginary parts of the overlap matrix elements $\text{Im} \langle \phi_i | \phi_j \rangle$ and the imaginary parts of the Hamiltonian matrix elements $\text{Im} \langle \phi_i | \hat{H} | \phi_j \rangle$ by following the same procedure as described above for the real part. Once we have obtained the full $\hat{H}$ and S matrices from measurements on these circuits,

$$H_{ij} = \langle \psi_i | \hat{H} | \psi_j \rangle, \quad S_{ij} = \langle \psi_i | \psi_j \rangle, \quad (\text{A11})$$

we can then solve the generalized eigenvalue problem

$$\mathbf{H}\vec{c} = \mathbf{E}\mathbf{S}\vec{c} \quad (\text{A12})$$

on a classical computer to get the quantum states and energies of the electronic system.

3. Circuit simplification

For NOQE calculations with H_2 in the STO-3G basis, we can reduce the experimental resources of running the NOQE circuit in Fig. 1 by simplifying the CSWAP network. The function of the first CSWAP network is to prepare the state $\frac{1}{\sqrt{2}}(|\psi_{\text{UHF}}\rangle|\text{vac}\rangle|0\rangle + |\text{vac}\rangle|\psi_{\text{UHF}}\rangle|1\rangle)$. Recognizing that the UHF states for H_2 is $|\Phi_{\text{UHF}}\rangle = |1100\rangle$ and $|\text{vac}\rangle = |0000\rangle$, we can use a simple circuit shown in Fig. 15 to prepare the desired state

$$\frac{1}{\sqrt{2}}(|1100\rangle|0000\rangle|0\rangle + |0000\rangle|1100\rangle|1\rangle), \quad (\text{A13})$$

with a significantly smaller two-qubit gate count than is required for a generic CSWAP network (see Appendix A 5).

4. Efficient Hamiltonian measurement

To perform the Hamiltonian measurement, we first apply the Jordan-Wigner transformation on the electronic Hamiltonian to get the corresponding qubit Hamiltonian, then decompose this to the Pauli basis and measure the expectation

value of these Pauli operators, and then evaluate the weighted sum

$$H_{IJ} = \sum_i w_i \langle \phi_I | \hat{P}_i | \phi_J \rangle. \quad (\text{A14})$$

For the H_2 molecule in the STO-3G basis, the Hamiltonian can be decomposed into 27 Pauli terms. Since the UCCD ansatz involves only double excitation, the first reference state will be in a neat form

$$|\phi_I\rangle = a_1|1100\rangle + a_2|0011\rangle \quad (\text{A15})$$

and the second reference state is

$$|\phi_J\rangle = a_3|1100\rangle + a_4|0011\rangle + a_5|0110\rangle + a_6|1001\rangle. \quad (\text{A16})$$

This allows the reduction of some redundant Pauli terms that produce the same expectation value, i.e.,

$$\langle \phi_I | \hat{P}_i | \phi_J \rangle = \pm(i) \langle \phi_I | \hat{P}_j | \phi_J \rangle, \quad \hat{P}_i, \hat{P}_j \in \{\hat{P}_{\text{group}}\}. \quad (\text{A17})$$

The 27 Pauli terms can be categorized into five sets $\{\hat{P}_{\text{group}}\}$:

- (i) $\{IIII, IIZZ, IZIZ, ZIIZ, IZZI, ZIZI, ZZII\}$,
- (ii) $\{ZXZX, IXIX, ZYZY, IYIY, XZXZ, YZYX, XIXI, YIYI\}$,
- (iii) $\{IIIZ, IIZI, IZII, ZIII\}$,
- (iv) $\{YYXX, XYXX, YXXY, XYYX\}$,
- (v) $\{IXZX, IYZY, XZZI, YZYI\}$.

It can be verified that Pauli terms belonging to the same set produce the same result (up to a $\pm$ sign and i), as indicated in Eq. (A17). So measuring expectation values for five Pauli terms, one from each group, is sufficient to reconstruct measurement of the full Hamiltonian.

For the special case of measuring the diagonal matrix elements $\langle \phi_I | \hat{P}_i | \phi_I \rangle$, the first set always produces ± 1 , the second and the last sets always produce 0, so for the diagonal matrix elements it is only necessary to measure two Pauli terms, one from each of the third and fourth sets, to generate the Hamiltonian matrix element.

5. Circuit resource estimation

Here, we count how many gates and circuits are needed to perform the NOQE for H_2 in the STO-3G basis.

In this case, the CSWAP operation in the circuit shown in Fig. 14 represents a set of four controlled-SWAPS applied in parallel to four pairs of qubits. Specifically, qubits (1, 5), (2, 6), (3, 7), and (4, 8) are swapped, conditioned on the value of the control qubit. We refer to this set of four parallel CSWAP operations as a CSWAP network.

For the first (simplified) CSWAP network in Appendix A 3, the two-qubit gate count is 4, as shown in Fig. 15. A generic CSWAP requires at least five arbitrary two-qubit gates [57] (or seven cnot gates [58]). Because realizing arbitrary two-qubit gates is often hard in practice, we count using the cnot metric. Hence, the second, complete CSWAP network costs 4×7 two-qubit gates, and the total two-qubit gate count is

$$N_{\text{CSWAP}}^{(2q)} = 4 + 4 \times 7 = 32. \quad (\text{A18})$$

Implementing one ansatz $e^{\hat{T}}$ needs four consecutive sub-routines of the form shown in Fig. 13. Each of these four

subroutines contains three given rotations and six crz rotations. Compilation of a Givens rotation needs two two-qubit gates, while the crz gate requires one two-qubit gate. Therefore, the total number of two-qubit gates for realizing $e^{\hat{z}}$ is

$$N_{e^{\hat{z}}}^{2q} = N_{\text{Given}}^{2q} + N_{\text{CRZ}}^{2q} = 2 \times 4 \times (3 \times 2 + 6) = 96. \quad (\text{A19})$$

The overall basis rotation $\hat{U}_{I \rightarrow 1}$ requires additional three Givens rotations, leading to a total two-qubit gate count of

$$N_{\hat{U}_{I \rightarrow 1}}^{2q} = 2 \times (3 \times 2) = 12. \quad (\text{A20})$$

We use the Qiskit [59] compiler to further optimize the circuit and count the number of two-qubit native gates (denoted by Rxx) and 1q rotations (denoted by U3). Here, we point out that an arbitrary 1q rotation can be realized by one Ry gate since U3 can be ZYZ decomposed and any Rz can be virtually implemented [60]. Taking this into account leads to the following total two-qubit and single-qubit gate counts:

$$N_{\text{all}}^{2q} = 124, \quad N_{\text{all}}^{1q} = 268. \quad (\text{A21})$$

We now discuss the number of circuits required. To evaluate the off-diagonal element of the overlap matrix, $\langle \hat{I} \otimes \hat{Z} \rangle$ is measured. This requires a single circuit. For the off-diagonal element of the Hamiltonian matrix, we need to measure the expectation values of five Pauli operators in order to construct $\langle \hat{H} \otimes \hat{Z} \rangle$, so five circuits are required. To measure the diagonal element of the Hamiltonian matrix, four qubits are sufficient and only two circuits are needed. (See Appendix A4 for details.)

We note here that we are optimizing the circuit at the gate level here. Given a specific device, it is possible to further undertake circuit synthesis and optimization at the gate level using pulse control, which potentially allows further reduction in circuit running time [61,62].

To summarize, for calculations on H_2 with the STO-3G basis at bond distance $R = 1.2 \text{ \AA}$, for the original NOQE implementation we need to run six nine-qubit circuits with 124 two-qubit gates and two four-qubit circuits with 52 two-qubit gates.

6. Matrix element error and eigenvalue error

This Appendix briefly reviews how the error $\mathcal{E}$ in the energy eigenvalues scales with the error ϵ in the matrix elements and improves on the energy error estimates derived in Ref. [7].

Consider the generalized eigenvalue problem

$$\beta \mathbf{H} \vec{c} = \alpha \mathbf{S} \vec{c}, \quad (\text{A22})$$

where $\mathbf{H}$ and $\mathbf{S}$ are $M \times M$ Hermitian matrices that form a definite pair, i.e., $|\vec{x}^\dagger \mathbf{H} \vec{x}| + |\vec{x}^\dagger \mathbf{S} \vec{x}| \neq 0 \forall \vec{x} \in \mathbb{C}^M$ and $\alpha/\beta = E$ (the superscript for the eigenvalue index k is dropped for brevity). The matrices $\mathbf{H}$ and $\mathbf{S}$ are subject to perturbations $\Delta \mathbf{H}$ and $\Delta \mathbf{S}$, respectively, such that $\mathbf{H}' = \mathbf{H} + \Delta \mathbf{H}$ and $\mathbf{S}' = \mathbf{S} + \Delta \mathbf{S}$. We consider now the generalized eigenvalue problem with the perturbed matrices:

$$\beta' \mathbf{H}' \vec{c}' = \alpha' \mathbf{S}' \vec{c}'. \quad (\text{A23})$$

The perturbation is assumed to be sufficiently small such that $M\chi \leq s_{\min}$, where $\chi := \sqrt{||\Delta \mathbf{H}||^2 + ||\Delta \mathbf{S}||^2}$ ($||\cdot||$ is the matrix spectral norm, defined as the largest singular value of

the matrix) represents the magnitude of the perturbations and $s_{\min}$ is the minimum eigenvalue of $\mathbf{S}$ [28]. The eigenangles are defined as

$$\theta = \tan^{-1} \beta/\alpha, \quad \theta' = \tan^{-1} \beta'/\alpha'. \quad (\text{A24})$$

Note that $\mathcal{E} := |E - E'| = |\cot \theta - \cot \theta'|$.

There are two known bounds on the eigenangle error $|\theta - \theta'|$. First, Stewart proved that [26]

$$|\theta - \theta'| \leq \sin^{-1} \frac{\chi}{c(\mathbf{H}, \mathbf{S})} \leq \sin^{-1} \frac{\chi}{s_{\min}}, \quad (\text{A25})$$

where $c(\mathbf{H}, \mathbf{S}) := \min\{\vec{x}^\dagger (\mathbf{H} + i\mathbf{S}) \vec{x}; \vec{x}^\dagger \vec{x} = 1\}$ is the Crawford number. Subsequently, Mathias and Li proved that [27,28]

$$|\theta - \theta'| \leq \sin^{-1} \frac{M\chi}{d_{\min}} \leq \sin^{-1} \frac{M\chi}{\sqrt{s_{\min}^2 + E_{\min}^2}}, \quad (\text{A26})$$

where $d_{\min} := \min_k |\alpha^k + i\beta^k|$ and $E_{\min}$ is the minimum eigenvalue of $\mathbf{H}$. While the factor of M in the Mathias-Li bound is undesirable compared to the Stewart bound, the Mathias-Li bound can still provide a comparably tight bound when $E_{\min}$ is much larger than $s_{\min}$.

Now, we relate $\chi := \sqrt{||\Delta \mathbf{H}||^2 + ||\Delta \mathbf{S}||^2}$ to the matrix element error ϵ . For the purpose of finding the scaling of χ with respect to ϵ , we may consider a scenario where each and every matrix element of $\mathbf{H}$ and $\mathbf{S}$ is independently measured. Assuming that the matrix element measurements are unbiased (no biasing device noise present), $||\Delta \mathbf{H}||$ and $||\Delta \mathbf{S}||$ are upper bounded by $||\Delta A||$, where ΔA is a $M \times M$ random matrix with entries drawn from a normal i.i.d. with zero mean and variance $\sigma^2 = \mathcal{O}(\epsilon^2)$. It is well known from random matrix theory that the distribution of $||\Delta A||$ is tightly focused around the mean value $2\sqrt{M}\sigma$ [63], which leads to

$$\chi = \mathcal{O}(\sqrt{M}\epsilon). \quad (\text{A27})$$

Combining Eqs. (A25) and (A27), we obtain $\mathcal{E} = \mathcal{O}(\epsilon M^{1/2}/s_{\min})$. Combining Eqs. (A26) and (A27), we obtain $\mathcal{E} = \mathcal{O}(\epsilon M^{3/2}/\sqrt{s_{\min}^2 + E_{\min}^2})$. These two relations provide estimates for the bound on the error $\mathcal{E}$ of the energy eigenvalues due to the error ϵ in the matrix elements.

Reference [7] derived a looser accuracy bound $\mathcal{E} = \mathcal{O}(\epsilon M^2/d_{\min})$, which, when combined with the Mathias-Li bound [Eq. (A26)] led to a loose upper bound estimate on the number of shots required to measure each matrix element [see Eq. (S70) of Ref. [7]]:

$$N_{\text{shots}} = \mathcal{O}\left(\frac{M^4}{d_{\min}^2 \mathcal{E}^2}\right). \quad (\text{A28})$$

This eigenvalue error bound is derived from the extreme constraint that each and every matrix element is measured within error ϵ , which is satisfied with a probability that drops exponentially with M^2 , such that $||\Delta \mathbf{H}||$ and $||\Delta \mathbf{S}||$ can be upper bounded by $M\epsilon$. In the current analysis, we relax this constraint and instead use the spectral norm of a random matrix ΔA with element variance $\sigma^2 = \mathcal{O}(\epsilon^2)$ as the upper bound on $||\Delta \mathbf{H}||$ and $||\Delta \mathbf{S}||$, which leads to Eq. (A27). The resulting bound on the number of shots per matrix element is

then reduced to

$$N_{\text{shots}} = \mathcal{O}\left(\frac{M^3}{d_{\min}^2 \mathcal{E}^2}\right). \quad (\text{A29})$$

A related work [29] on the quantum Krylov subspace diagonalization algorithm [19] takes a different approach and considers the optimal allocation of shots to the matrix elements such that the total number of shots is minimized, rather than imposing a uniform upper bound ϵ to the error in each matrix element. For the case of a non-Toeplitz matrix $\mathbf{H}$ as is relevant for NOQE, the *total* number of shots is then given by

$$N_{\text{shots}}^{\text{tot}} = \mathcal{O}\left(\frac{M^5 \log M}{d_{\min}^2 \mathcal{E}^2}\right), \quad (\text{A30})$$

where the Mathias-Li bound [Eq. (A26)] is again used. Comparing with our estimate in Eq. (A29), Eq. (A30) has an extra factor of $\log M$, which comes from applying the inequality

$$\begin{aligned} \hat{O} &= \hat{I}, \quad \text{with } \hat{\rho}_i = |\psi_i\rangle\langle\psi_i|, \quad \hat{\rho}_i^R = |\psi_i^R\rangle\langle\psi_i^R|, \quad \hat{\rho}_i^I = |\psi_i^I\rangle\langle\psi_i^I|, \\ \hat{\rho}_j &= |\psi_j\rangle\langle\psi_j|, \quad \hat{\rho}_j^R = |\psi_j^R\rangle\langle\psi_j^R|, \quad \hat{\rho}_j^I = |\psi_j^I\rangle\langle\psi_j^I|. \end{aligned} \quad (\text{B1})$$

where the auxiliary states $|\psi_i^R\rangle$ and $|\psi_i^I\rangle$ are defined in Eq. (19) and $|\psi_{i/j}\rangle$ are the reference states. $\text{Re}(S_{ij})$ and $\text{Im}(S_{ij})$ can then be extracted using Eqs. (21)–(23).

To prove Eqs. (22) and (23), we notice that $\langle 0|\psi_i\rangle = 0$ for any reference states $|\psi_i\rangle$. Therefore, we have for $\text{Re}(S_{ij})$

$$\begin{aligned} \text{Tr}(|\psi_i^I\rangle\langle\psi_i^I| |\psi_j^I\rangle\langle\psi_j^I|) &= \frac{1}{4} \text{Tr}[(|0\rangle\langle 0| + |0\rangle\langle\psi_i| + |\psi_i\rangle\langle 0| + |\psi_i\rangle\langle\psi_i|)(|0\rangle\langle 0| + |0\rangle\langle\psi_j| + |\psi_j\rangle\langle 0| + |\psi_j\rangle\langle\psi_j|)] \\ &= \frac{1}{4} \text{Tr}[|0\rangle\langle 0| + |0\rangle\langle\psi_j| + |0\rangle\langle\psi_i| + |\psi_i\rangle\langle 0| + |\psi_i\rangle\langle\psi_j| + |\psi_i\rangle\langle\psi_i| + |\psi_j\rangle\langle 0| + |\psi_j\rangle\langle\psi_i| + |\psi_j\rangle\langle\psi_j|] \\ &= \frac{1}{4}(1 + \langle\psi_i|\psi_j\rangle + \langle\psi_j|\psi_i\rangle + |\langle\psi_i|\psi_j\rangle|^2) = \frac{1}{4}[1 + 2\text{Re}(S_{ij}) + |S_{ij}|^2], \end{aligned} \quad (\text{B2})$$

and for $\text{Im}(S_{ij})$ we have

$$\text{Tr}(|\psi_i^I\rangle\langle\psi_i^I| |\psi_j^R\rangle\langle\psi_j^R|) = \frac{1}{4}[1 - i\langle\psi_i|\psi_j\rangle + i\langle\psi_j|\psi_i\rangle + |\langle\psi_i|\psi_j\rangle|^2] = \frac{1}{4}[1 + 2\text{Im}(S_{ij}) + |S_{ij}|^2]. \quad (\text{B3})$$

To obtain the Hamiltonian matrix elements, the bilinear functions we need to estimate through shadow tomography include

$$\hat{O} = \hat{H}, \quad \text{with } \hat{\rho}_i = |\psi_i\rangle\langle\psi_i| \text{ and } \hat{\rho}_j = |\psi_j\rangle\langle\psi_j|.$$

Then, we can extract H_{ij} using Eq. (24) and S_{ij} .

Finally, in order to improve the sample complexity and also reduce any noise-induced bias via shadow distillation, in all the numerical simulations presented in this work, we have constructed the unbiased estimators $\hat{\rho}_{\text{cs},i}$ and $\hat{\rho}_{\text{cs},j}$ using order-3 U-statistics,

$$\hat{\rho}_{\text{cs},i} = \hat{\rho}_i^{U_3} = \frac{1}{n(n-1)(n-2)} \sum_{\alpha \neq \beta \neq \gamma} \hat{\rho}_{i,\alpha} \hat{\rho}_{i,\beta} \hat{\rho}_{i,\gamma}, \quad (\text{B4})$$

where $\hat{\rho}_{i,x}$ ($x = \alpha, \beta$, or γ) are classical shadows constructed from the state ρ_i and n is the number of classical shadows.

APPENDIX C: NOISE MODEL AND ZERO-NOISE EXTRAPOLATION

In this Appendix, we introduce the noise model used in our simulation and the zero-noise extrapolation techniques.

Characterizing noise in experiments can be challenging [64,65]. To model realistic noise, we use the calibrated noise rates from the Quantinuum System Model H2 hardware in

for the spectral norm of the sum of a Gaussian matrix series. While our bound on N_{shots} is more favorable in an asymptotic sense, future work should consider the optimal allocation of shots, e.g., as in Ref. [29] to find a tighter and nonasymptotic bound on the total number of shots.

APPENDIX B: SHADOW-BASED NOQE ALGORITHM

Here, we give details for the shadow estimation of the matrix elements used in the NOQE algorithm.

The data-acquisition phase of shadow tomography performed on copies of $\hat{\rho}_i$ and $\hat{\rho}_j$ will produce unbiased estimators $\hat{\rho}_{\text{cs},i}$ and $\hat{\rho}_{\text{cs},j}$ for these two states: $\mathbb{E}[\hat{\rho}_{\text{cs},i}] = \hat{\rho}_i$ and $\mathbb{E}[\hat{\rho}_{\text{cs},j}] = \hat{\rho}_j$. These are then used to form unbiased estimators $\text{Tr}(\hat{\rho}_{\text{cs},i} \hat{\rho}_{\text{cs},j} \hat{O})$ for the bilinear functions $\text{Tr}(\hat{\rho}_i \hat{\rho}_j \hat{O})$ of interest.

To estimate the overlap matrix elements S_{ij} , the bilinear functions include

our simulation. The specifications of the Quantinuum System Model H2 are listed in Table III. We incorporate typical single-qubit and two-qubit gate infidelities. To evaluate performance across regimes of different noise magnitudes, we rescale the gate error rates as shown in Table IV.

We now introduce the working principles and describe the implementation of ZNE in our simulation.

TABLE III. Quantinuum System Model H2 specifications. The data is sourced from the Quantinuum product data sheet [22]. In our noise model, we use the typical (typ) calibrated gate infidelity values provided in the data sheet.

General			
Qubits	56		
Connectivity	All to all		
Parallel two-qubit operations	4		
Errors			
	min	typ	max
Single-qubit gate infidelity	1×10^{-5}	3×10^{-5}	2×10^{-4}
Two-qubit gate infidelity	1.3×10^{-3}	1.5×10^{-3}	3×10^{-3}

TABLE IV. Gate error rates at different noise strengths. The single-qubit (1q) and two-qubit (2q) error rates, denoted as p_1 and p_2 , respectively, for base noise strength equal to one are taken directly from the Quantum System Model H2 noise parameters. These error rates are then scaled by multiplying by a noise strength factor λ . The values presented in the table correspond to a composite noise model where, for a given noise strength λ , the depolarizing noise rate, amplitude damping noise rate, and phase damping noise rate are all set equal to the error rates shown here.

Noise strength λ	0.5	0.75	1.0	1.25	1.5
1q gate error rate ($\times 10^{-5}$): p_1	1.5	2.25	3	3.75	6
2q gate error rate ($\times 10^{-3}$): p_2	0.75	1.125	1.5	1.875	2.25

ZNE mitigates noise in quantum circuits by deliberately scaling it through a process known as “circuit folding,” where a unitary operation $\hat{U}$ is transformed into $\hat{U}\hat{U}^\dagger\hat{U}$, as illustrated in Fig. 16(a) below. This folding operation preserves the ideal unitary evolution while increasing the noise, effectively scaling it by a factor (e.g., approximately three times for a single folding). In our simulation, the noise extrapolation factors are set to 0.5, 0.75, 1, 1.25, and 1.5. Note that some noise extrapolation factors are fractional, which we realize through adjusting the noise strength in gate operations, as listed in Table IV. By performing measurements at these different noise levels, we extrapolate the results back to the zero-noise limit using an exponential fit, as shown in Fig. 16(b).

APPENDIX D: MEASUREMENT SAMPLE COMPLEXITY

In this Appendix, we analyze the sample complexity of the NOQE measurement protocols, both in their original form with the Hadamard test and using the shadow tomography protocol outlined in this work. We first establish the necessary preliminaries in Sec. D 1, including key assumptions on the observable $\hat{O}$ and useful trace identities. We then derive

the sample complexity of the original NOQE protocol based on Pauli measurements in Sec. D 2, followed by an analysis of improved shadow-based estimation techniques for bilinear functions in Sec. D 3. We further analyze the performance using higher-order U-statistics for linear and bilinear functions in Secs. D 4 and D 5, respectively. These results highlight the advantages of shadow-based methods over traditional measurement strategies in certain parameter regimes.

1. Preliminaries

In this section, we consider Hermitian observables $\hat{O}$ satisfying the following properties:

$$\text{Tr}(\hat{O}^2) \leq B \quad \text{and} \quad \|\hat{O}\|_\infty = \max_i |\lambda_i(\hat{O})| \in \mathcal{O}(1),$$

where $\lambda_i(\hat{O})$ denotes the eigenvalues of $\hat{O}$. Thus, the spectral norm is bounded by a constant. This assumption is made without loss of generality, since any observable can be rescaled to satisfy it, i.e.,

$$\hat{O} = \frac{\hat{O}_{\text{original}}}{\|\hat{O}_{\text{original}}\|_\infty}. \quad (\text{D1})$$

As a result of this rescaling, any physical units carried by the observable (e.g., energy units for a Hamiltonian) are absorbed into the scaling factor. The estimation error ϵ used throughout this work then refers to the error in estimating the rescaled, unitless observable, and is thus dimensionless. This convention is standard in sample complexity analysis, including in quantum algorithms such as quantum phase estimation. Rescaling in this way helps clearly isolate the dependence on structural parameters like B and ϵ in complexity formulas, without conflating them with physical units. Once a dimensionless estimate is obtained, the corresponding physical quantity (e.g., energy) can be recovered by multiplying the result by $\|\hat{O}_{\text{original}}\|_\infty$.

We also note that this rescaling implies $B \leq D$, where D is the dimension of the Hilbert space on which $\hat{O}$ acts. So one

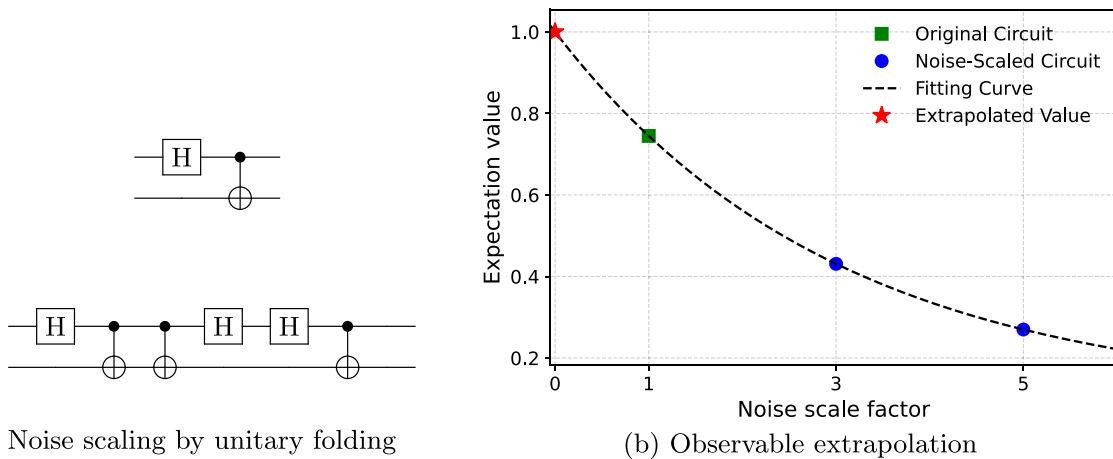


FIG. 16. ZNE error mitigation technique. (a) Generating a noise-scaled circuit by applying unitary folding. The noise level of the original circuit $\hat{U} = (\text{H} \otimes \hat{I})\text{CNOT}$ is intentionally increased by inserting the identity-equivalent operation $\hat{U}^\dagger\hat{U}$, resulting in $\hat{U}\hat{U}^\dagger\hat{U}$. (b) Estimating the zero-noise result by extrapolating noise-scaled results. Unitary folding is performed at different scaling factors (e.g., 3 and 5, as illustrated in the figure). In our implementation, the scaling factors used are 0.5, 0.75, 1.0, 1.25, and 1.5, as listed in Table IV, yielding multiple noise-scaled expectation values for extrapolation. A parameterized curve (e.g., $a + be^{-cx}$) is then fitted to the results, and the intercept at zero noise is taken as the error-mitigated result.

would have $\text{Tr}(\hat{O}\hat{\rho}) \leq \|\hat{O}\|_\infty \in \mathcal{O}(1)$ and also $\text{Tr}(\hat{O}^2\rho) \leq \|\hat{O}\|_\infty^2 \in \mathcal{O}(1)$. Using the Cauchy-Schwarz inequality, one would also have $\text{Tr}(\hat{O})^2 \leq D \text{Tr}(\hat{O}^2)$.

There are several useful equations that we will use in the derivations below:

$$\begin{aligned} & \text{Tr}((\hat{A} \otimes \hat{B} \otimes \hat{C} \otimes \hat{E})(\hat{W}_{(1,2)} \otimes \hat{W}_{(3,4)})(\hat{W}_{(1,3)} \otimes \hat{W}_{(2,4)})) \\ &= \text{Tr}(\hat{A}\hat{E})\text{Tr}(\hat{B}\hat{C}), \end{aligned} \quad (\text{D2})$$

$$\text{Tr}((\hat{A} \otimes \hat{B} \otimes \hat{C} \otimes \hat{E}) \hat{W}_{(3,4)}(\hat{W}_{(1,3)} \otimes \hat{W}_{(2,4)})) = \text{Tr}(\hat{A}\hat{E}\hat{B}\hat{C}), \quad (\text{D3})$$

$$\text{Tr}((\hat{A} \otimes \hat{B} \otimes \hat{C} \otimes \hat{E}) \hat{W}_{(1,2)}(\hat{W}_{(1,3)} \otimes \hat{W}_{(2,4)})) = \text{Tr}(\hat{A}\hat{C}\hat{B}\hat{E}), \quad (\text{D4})$$

where W denotes the SWAP operator.

We also note that in Appendixes B, D3, and D5, we derive only the sample complexity for estimating a single function, since the sample complexity for estimating K functions can be obtained from this by using the standard median-of-mean method [15].

2. Original NOQE measurement

The original NOQE protocol uses the generalized Hadamard test to prepare either $|\phi\rangle = |\psi_i\rangle + |\psi_j\rangle$ or $|\phi\rangle = |\psi_i\rangle - |\psi_j\rangle$, depending on the measurement outcome of the control qubit. Then, measuring the Hamiltonian $\hat{H}$ in the first register gives the expectation value $\langle\phi|\hat{H}|\phi\rangle = \frac{\langle\hat{H}\rangle_i + \langle\hat{H}\rangle_j + 2\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle}{2 + 2\text{Re}\langle\phi_i|\phi_j\rangle}$, thus estimate $\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle$. The imaginary part can be estimated similarly. $\text{Re}\langle\phi_i|\phi_j\rangle$ is given by the probability of getting +1 or -1 in the control qubit measurement. See Appendix A2 for full details.

If one wants to estimate $\text{Re}\langle\phi_i|\phi_j\rangle$ to additive error ϵ , one only needs $\frac{1}{\epsilon^2}$ measurements (equivalent to estimating the parameter of a Bernoulli distribution), which is typically much smaller than the number of measurements needed to estimate $\langle\phi|\hat{H}|\phi\rangle$. Thus, we can safely ignore the statistical error in $\text{Re}\langle\phi_i|\phi_j\rangle$.

To estimate $\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle$ to an additive error ϵ , we then only need to estimate $\langle\phi|\hat{H}|\phi\rangle$ to an additive error ϵ (up to some rescaling because of the presence of $\text{Re}\langle\phi_i|\phi_j\rangle$). The original NOQE protocol measures $\hat{H}$ by decomposing it into $\hat{H} = \sum_i w_i \hat{P}_i$, where $\hat{P}_i$ is some Pauli operator and w_i is some real coefficient, and then measuring individual Pauli terms. Normally, the Frobenius norm of $\hat{H}$, i.e., the square root of the sum of the absolute squares of all matrix elements, is bounded:

$$\text{Tr}(\hat{H}^2) = \sum_i w_i^2 D \leq B, \quad (\text{D5})$$

where D is the dimension of the Hilbert space. If $\tilde{n}$ measurements are performed to measure a single Pauli term $\hat{P}_i$, and the estimator of this is constructed as $p_i = \frac{\sum_{l=1}^{\tilde{n}} X_{li}}{\tilde{n}}$, where $X_{li} \in \{-1, +1\}$ is the outcome of the l th measurement, one can check that $\mathbb{E}[p_i] = \langle\phi|\hat{P}_i|\phi\rangle$. The variance of this is

$$\text{Var}[p_i] = \frac{\sum_{l=1}^{\tilde{n}} \text{Var}[X_{li}]}{\tilde{n}^2} = \frac{\text{Var}[X_i]}{\tilde{n}} \leq \frac{1}{\tilde{n}}, \quad (\text{D6})$$

which makes sense because this is again just estimation over a Bernoulli distribution.

After repeating this for every Pauli $\hat{P}_i$ in the Hamiltonian $\hat{H}$, each with $\tilde{n}$ measurements, the estimator h for $\langle\phi|\hat{H}|\phi\rangle$ is then constructed as $h = \sum_i w_i p_i$. The variance of this is

$$\text{Var}[h] = \sum_i w_i^2 \text{Var}[p_i] \leq \sum_i w_i^2 \frac{1}{\tilde{n}} \leq \frac{B}{D\tilde{n}}. \quad (\text{D7})$$

If one wants to estimate $\langle\phi|\hat{H}|\phi\rangle$ to additive error ϵ , one needs to ensure that $\text{Var}[h] \leq \epsilon^2$ (by Chebyshev's inequality). This implies that $\tilde{n} \geq \frac{B}{D\epsilon^2}$. If there are Ω Pauli terms in $\hat{H}$, then the total number of measurements n is

$$n = \Omega\tilde{n} = \mathcal{O}\left(\frac{B\Omega}{D\epsilon^2}\right). \quad (\text{D8})$$

Similar bounds have been derived previously, e.g., $n = \mathcal{O}(\frac{\sum_i |w_i|^2}{\epsilon^2})$ [13], which is smaller than Eq. (D8). This can be obtained without the last inequality used in Eq. (D7). Here, we express the bound using B and D instead of $\sum_i w_i^2$ in order to be consistent with the later analysis for shadow estimation. A comment on the bounds in Eq. (D8) is in order here, namely, that in the worst case, Ω can be as large as D^2 . In this situation $n = \mathcal{O}(\frac{B D^2}{\epsilon^2})$, which is much worse than the bound one would get by performing global Clifford shadow tomography on $|\phi\rangle$, namely, $n = \mathcal{O}(\frac{B}{\epsilon^2})$ [15,31].

3. Shadow estimation of bilinear functions

Here, we derive the number of measurements required to estimate bilinear function of two pure states $\hat{\rho}$ and $\hat{\sigma}$ of the form

$$o = \text{Tr}(\hat{O}\hat{\rho}\hat{\sigma}) + \text{Tr}(\hat{\sigma}\hat{O}\hat{\rho}) = \text{Tr}(\hat{W}(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I})\hat{\rho} \otimes \hat{\sigma}), \quad (\text{D9})$$

where $\hat{W}$ is the SWAP operator. We perform global Clifford measurement shadow tomography on $\hat{\rho}$ and $\hat{\sigma}$, respectively, obtaining classical shadows $\{\hat{\rho}_i\}_{i=1}^n$ and $\{\hat{\sigma}_i\}_{i=1}^n$, each of n samples. We then construct the unbiased estimator for $\hat{\rho}_{\text{es}}$ and $\hat{\sigma}_{\text{es}}$ as

$$\hat{\rho}_{\text{es}} = \frac{1}{n} \sum_i \hat{\rho}_i, \quad (\text{D10a})$$

$$\hat{\sigma}_{\text{es}} = \frac{1}{n} \sum_i \hat{\sigma}_i. \quad (\text{D10b})$$

The unbiased estimator for the bilinear function o is constructed as

$$\begin{aligned} o_{\text{es}} &= \text{Tr}(\hat{W}(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I})\hat{\rho}_{\text{es}} \otimes \hat{\sigma}_{\text{es}}) \\ &= \frac{1}{n^2} \sum_{i,k} \text{Tr}(\hat{W}(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I})\hat{\rho}_i \otimes \hat{\sigma}_k). \end{aligned} \quad (\text{D11})$$

The accuracy of this estimator is controlled by its variance, which is equal to

$$\begin{aligned}
\text{Var}[o_{\text{es}}] &= \mathbb{E}[o_{\text{es}}^2] - \mathbb{E}[o_{\text{es}}]^2 \\
&= \frac{1}{n^4} \sum_{i,j,k,l} \text{Cov}(o_{ik}, o_{jl}) \\
&= \mathbb{E} \left[\frac{1}{n^4} \sum_{i,j,k,l} \text{Tr}(\hat{W}(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I}) \hat{\rho}_i \otimes \hat{\sigma}_k) \text{Tr}(\hat{W}(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I}) \hat{\rho}_j \otimes \hat{\sigma}_l) \right] - o^2 \\
&= \frac{1}{n^4} \sum_{i,j,k,l} \text{Tr}((\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I}) \otimes (\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I}) \mathbb{E}[\hat{\rho}_i \otimes \hat{\sigma}_k \otimes \hat{\rho}_j \otimes \hat{\sigma}_l] \hat{W}_{(1,2)} \otimes \hat{W}_{(3,4)}) - o^2 \\
&= \frac{1}{n^4} \sum_{i,j,k,l} \text{Tr}\{(\hat{I} \otimes \hat{I} \otimes \hat{O} \otimes \hat{O} + \hat{I} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I} + \hat{O} \otimes \hat{I} \otimes \hat{O} \otimes \hat{I} + \hat{O} \otimes \hat{O} \otimes \hat{I} \otimes \hat{I}) \\
&\quad \times \mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_j \otimes \hat{\sigma}_k \otimes \hat{\sigma}_l] \hat{W}_{(1,3)} \otimes \hat{W}_{(2,4)}\} - o^2, \tag{D12}
\end{aligned}$$

where the last equality is obtained by relabeling the Hilbert spaces of subsystems 2 and 3. We now present a useful lemma that we will repeatedly use for calculations in the remainder of this section.

Lemma D1. (Lemma 34 of Ref. [31]) For random Clifford measurement, the second moment of the classical shadow is

$$\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i] = (I \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}) \left(\frac{D+1}{D+2} \hat{W} - \frac{1}{D+2} \hat{I} \right). \tag{D13}$$

We now evaluate Eq. (D12). The values of the various terms in this depend on whether i and j are equal, as well as whether k and l are equal, i.e., whether they are independent classical shadows or not. We shall analyze the different situations individually.

a. When $i \neq j$ and $k \neq l$

In this case, $\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_j \otimes \hat{\sigma}_k \otimes \hat{\sigma}_l] = \hat{\rho} \otimes \hat{\rho} \otimes \hat{\sigma} \otimes \hat{\sigma}$ and thus $\text{Cov}(o_{ik}, o_{jl}) = 0$.

b. When $i = j$ and $k \neq l$

In this case, we have

$$\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_j \otimes \hat{\sigma}_k \otimes \hat{\sigma}_l] = \mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i] \otimes \hat{\sigma} \otimes \hat{\sigma} = (\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}) \otimes \hat{\sigma} \otimes \hat{\sigma} \left(\frac{D+1}{D+2} \hat{W}_{(1,2)} - \frac{1}{D+2} \hat{I} \right). \tag{D14}$$

So, we now have (omitting $\otimes$ and explicitly indicating the dot product by $\cdot$ or $\times$ when it exists)

$$\begin{aligned}
\text{Cov}(o_{i,k}, o_{j,l}) &= \text{Tr} \left\{ (\hat{I} \hat{I} \hat{O} \hat{O} + \hat{I} \hat{O} \hat{O} \hat{I} + \hat{O} \hat{I} \hat{O} + \hat{O} \hat{O} \hat{I} \hat{I}) \cdot (\hat{I} \hat{\rho} \hat{\sigma} + \hat{I} \hat{\rho} \hat{\sigma} + \hat{\rho} \hat{I} \hat{\sigma}) \right. \\
&\quad \times \left. \left(\frac{D+1}{D+2} \hat{W}_{(1,2)} - \frac{1}{D+2} \hat{I} \right) \cdot \hat{W}_{(1,3)} \hat{W}_{(2,4)} \right\} - o^2 \\
&= \frac{2(D+1)}{D+2} \{ \text{Tr}(\hat{O} \hat{\sigma} \hat{O} \hat{\sigma}) + \text{Tr}(\hat{O} \hat{\sigma} \hat{\rho} \hat{O} \hat{\sigma}) + \text{Tr}(\hat{\sigma} \hat{O}^2 \hat{\sigma}) + \text{Tr}(\hat{\sigma} \hat{O} \hat{\rho} \hat{O} \hat{\sigma}) + \text{Tr}(\hat{\rho} \hat{\sigma} \hat{O}^2 \hat{\sigma}) + \text{Tr}(\hat{O} \hat{\sigma} \hat{O} \hat{\rho} \hat{\sigma}) \} \\
&\quad - \frac{4}{D+2} \{ \text{Tr}(\hat{O} \hat{\sigma})^2 + \text{Tr}(\hat{O} \hat{\sigma}) \text{Tr}(\hat{\rho} \hat{O} \hat{\sigma}) + \text{Tr}(\hat{O} \hat{\sigma}) \text{Tr}(\hat{\sigma} \hat{O} \hat{\rho}) \} - o^2, \tag{D15}
\end{aligned}$$

where we used Eq. (D3). Using the fact that $\hat{\rho}$ and $\hat{\sigma}$ are pure states, so that, for example, $\text{Tr}(\hat{O} \hat{\sigma} \hat{O} \hat{\sigma}) = \text{Tr}(\hat{O} \hat{\sigma})^2$, we can further simplify Eq. (D15) as

$$\begin{aligned}
\text{Cov}(o_{i,k}, o_{j,l}) &= \frac{2D-2}{D+2} [\text{Tr}(\hat{O} \hat{\sigma})^2 + \text{Tr}(\hat{O} \hat{\sigma}) \text{Tr}(\hat{O} \hat{\sigma} \hat{\rho} + \hat{O} \hat{\rho} \hat{\sigma})] + \frac{2D+2}{D+2} [|\langle \sigma | \hat{O} | \rho \rangle|^2 + \text{Tr}(\hat{O}^2 \sigma)(1 + |\langle \rho | \sigma \rangle|^2)] - o^2 \\
&\leq 2\|\hat{O}\|_\infty^2 + 4\|\hat{O}\|_\infty^2 + 2\|\hat{O}\|_\infty \|(\hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{I}) \hat{W}\|_\infty + 4\|\hat{O}\|_\infty^2 - o^2 \\
&= 10\|\hat{O}\|_\infty^2 + 2\|\hat{O}\|_\infty \cdot 2\|\hat{O}\|_\infty - o^2 \\
&= 14\|\hat{O}\|_\infty^2 - o^2 \\
&\leq 14\|\hat{O}\|_\infty^2 \in \mathcal{O}(1). \tag{D16}
\end{aligned}$$

c. When $i \neq j, k = l$

This is similar to the above and we get $\text{Cov}(o_{i,k}, o_{j,l}) \in \mathcal{O}(1)$.

d. When $i = j$, and $k = l$

In this case, we have

$$\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_j \otimes \hat{\sigma}_k \otimes \hat{\sigma}_l] = (\hat{I}\hat{I} + \hat{I}\hat{\rho} + \hat{\rho}\hat{I})(\hat{I}\hat{I} + \hat{I}\hat{\sigma} + \hat{\sigma}\hat{I}) \times \left(\frac{D+1}{D+2}\hat{W}_{(1,2)} - \frac{1}{D+2}\hat{I} \right) \left(\frac{D+1}{D+2}\hat{W}_{(3,4)} - \frac{1}{D+2}\hat{I} \right) \times (\hat{W}_{(1,3)}\hat{W}_{(2,4)}). \quad (\text{D17})$$

Therefore, we have

$$\begin{aligned} \text{Cov}(o_{i,k}, o_{l,j}) &= \text{Tr} \left\{ (\hat{I}\hat{I}\hat{\sigma}\hat{\sigma} + \hat{I}\hat{\sigma}\hat{\sigma}\hat{I} + \hat{\sigma}\hat{I}\hat{I}\hat{\sigma} + \hat{\sigma}\hat{\sigma}\hat{I}\hat{I}) \right. \\ &\quad \times (\hat{I}\hat{I}\hat{I}\hat{I} + \hat{I}\hat{I}\hat{\sigma}\hat{I} + \hat{I}\hat{I}\hat{\sigma}\hat{I} + \hat{I}\hat{\rho}\hat{I}\hat{I} + \hat{I}\hat{\rho}\hat{I}\hat{I} + \hat{I}\hat{\rho}\hat{\sigma}\hat{I} + \hat{\rho}\hat{I}\hat{I}\hat{I} + \hat{\rho}\hat{I}\hat{I}\hat{\sigma} + \hat{\rho}\hat{I}\hat{\sigma}\hat{I}) \\ &\quad \times \left[\frac{(D+1)^2}{(D+2)^2}(\hat{W}_{(1,2)}\hat{W}_{(3,4)}) + \frac{1}{(D+2)^2}\hat{I} - \frac{D+1}{(D+2)^2}(\hat{W}_{(1,3)} + \hat{W}_{(1,2)}) \right] (\hat{W}_{(1,3)}\hat{W}_{(2,4)}) \Big\} - o^2 \\ &= \frac{(D+1)^2}{(D+2)^2} \{ 3\text{Tr}(\hat{\sigma})^2 + 4\text{Tr}(\hat{\sigma}\hat{\sigma} + \hat{\sigma}\hat{\rho})\text{Tr}(\hat{\sigma}) + 4\text{Tr}(\hat{\sigma}\hat{\rho})\text{Tr}(\hat{\sigma}\hat{\sigma}) + 2\text{Tr}(\hat{\sigma})\text{Tr}(\hat{\sigma}\hat{\rho}\hat{\sigma} + \hat{\sigma}\hat{\rho}\hat{\sigma}) \\ &\quad + 5\text{Tr}(\hat{\sigma}^2) + 2D\text{Tr}(\hat{\sigma}^2\hat{\sigma} + \hat{\sigma}^2\hat{\rho}) + 2\text{Tr}(\hat{\sigma}^2\hat{\rho} + \hat{\sigma}^2\hat{\sigma}) + D\text{Tr}(\hat{\sigma}\hat{\rho})\text{Tr}(\hat{\sigma}\hat{\sigma}) + 2\text{Tr}(\hat{\sigma}^2)\text{Tr}(\hat{\rho}\hat{\sigma}) + D\text{Tr}(\hat{\sigma}\hat{\rho}\hat{\sigma}\hat{\sigma}) \} \\ &\quad + \frac{1}{(D+1)^2} \{ 4\text{Tr}(\hat{\sigma}^2) + 8\text{Tr}(\hat{\sigma}\hat{\rho} + \hat{\sigma}\hat{\sigma})\text{Tr}(\hat{\sigma}) + 4\text{Tr}(\hat{\sigma})\text{Tr}(\hat{\sigma}\hat{\rho}\hat{\sigma} + \hat{\sigma}\hat{\rho}\hat{\sigma}) + 8\text{Tr}(\hat{\sigma}\hat{\sigma})\text{Tr}(\hat{\sigma}\hat{\rho}) \} \\ &\quad - \frac{D+1}{(D+2)^2} \{ 8\text{Tr}(\hat{\sigma}^2) + 16\text{Tr}(\hat{\sigma}^2\hat{\sigma} + \hat{\sigma}^2\hat{\rho}) + 16\text{Tr}(\hat{\sigma}\hat{\rho}\hat{\sigma}\hat{\sigma}) + 8\text{Tr}(\hat{\sigma}^2\hat{\rho}\hat{\sigma} + \hat{\sigma}^2\hat{\sigma}\hat{\rho}) \} - o^2. \end{aligned} \quad (\text{D18})$$

The dominant term of the above expression is the $\text{Tr}(\hat{\sigma})^2$ term, which is upper bounded by $D\text{Tr}(\hat{\sigma}^2)$. Therefore, we have

$$\text{Cov}(o_{i,k}, o_{l,j}) \in \mathcal{O}(D\text{Tr}(\hat{\sigma}^2)) \leq \mathcal{O}(DB). \quad (\text{D19})$$

e. Final result

Putting all of these different cases *a.* through *d.* together, we have the bound for the variance of our estimator o_{es} :

$$\text{Var}[o_{\text{es}}] \in \mathcal{O}\left(\frac{n^3 + n^2 DB}{n^4}\right) = \mathcal{O}\left(\frac{1}{n} + \frac{DB}{n^2}\right). \quad (\text{D20})$$

Therefore, in order to estimate the bilinear function o within ϵ with high probability, which means $\text{Var}(o_{\text{es}}) \leq \epsilon^2$, we need the number of samples n to be

$$n = \mathcal{O}\left(\frac{1}{\epsilon^2} + \frac{\sqrt{DB}}{\epsilon}\right). \quad (\text{D21})$$

We now add some comments about this bound. First, recall that when performing random Clifford measurements on $\hat{\rho}$, the sample complexity for estimating $\text{Tr}(\hat{\sigma}\hat{\rho})$ is $n = \mathcal{O}(\frac{B}{\epsilon^2})$. This means that when $\epsilon \leq \sqrt{\frac{B}{D}}$, our bound is smaller than $n = \mathcal{O}(\frac{B}{\epsilon^2})$ and estimating $\text{Tr}(\hat{\sigma}\hat{\sigma}\hat{\rho})$ (with $\hat{\rho}$ and $\hat{\sigma}$ both pure states) is even more accurate than estimating $\text{Tr}(\hat{\sigma}\hat{\rho})$. Second, Eq. (D21) coincides with the sample complexity of estimating linear functions of $\hat{\rho}$ using $m = 2$ U-statistics.

4. Shadow estimation using U-statistics

Now, we turn to derive the measurement bound for estimating linear functions of the form $\text{Tr}(\hat{\sigma}\hat{\rho})$, where $\hat{\rho}$ is a pure state, with shadow tomography using U-statistics. First, we perform shadow tomography (random Clifford measurement) on $\hat{\rho}$, obtaining classical shadows $\{\hat{\rho}_i\}_{i=1}^n$. Now, we use order- m U-statistics to construct an unbiased estimator $\hat{\rho}^{(U_m)}$ of $\hat{\rho}$, which can be taken to be

$$\hat{\rho}^{(U_m)} = \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{i_1 \neq i_2 \neq \cdots \neq i_m} \hat{\rho}_{i_1} \hat{\rho}_{i_2} \cdots \hat{\rho}_{i_m}, \quad (\text{D22})$$

where m is some positive integer. Since $\hat{\rho}$ is a pure state ($\hat{\rho}^m = \hat{\rho}$), one can check that Eq. (D22) is an unbiased estimator of $\hat{\rho}$, i.e.,

$$\mathbb{E}[\hat{\rho}^{(U_m)}] = \hat{\rho}. \quad (\text{D23})$$

The original classical shadow [15] that is equivalent to $m = 1$ has sample complexity

$$n = \mathcal{O}\left(\frac{B}{\epsilon^2}\right), \quad (\text{D24})$$

and Ref. [31] has proven that the sample complexity for $m = 2$ is

$$n = \mathcal{O}\left(\frac{1}{\epsilon^2} + \frac{\sqrt{DB}}{\epsilon}\right). \quad (\text{D25})$$

Now, we derive the sample complexity for $m = 3$, which we have used in this paper.

The estimator of our linear function is now

$$\hat{o} = \text{Tr}(\hat{O}\hat{\rho}^{(U_m)}). \quad (\text{D26})$$

The error is controlled by the variance of our estimator, so the remainder of this section is devoted to calculating this variance. For $m = 3$, the variance is

$$\begin{aligned} \text{Var}[o_{\text{es}}] &= \mathbb{E}[o_{\text{es}}^2] - \mathbb{E}[o_{\text{es}}]^2 \\ &= \frac{1}{[n(n-1)(n-2)]^2} \sum_{i \neq j \neq k} \sum_{\alpha \neq \beta \neq \gamma} \mathbb{E}[\text{Tr}(\hat{O}\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k)\text{Tr}(\hat{O}\hat{\rho}_\alpha\hat{\rho}_\beta\hat{\rho}_\gamma)] - o^2 \\ &= \frac{1}{[n(n-1)(n-2)]^2} \sum_{i \neq j \neq k} \sum_{\alpha \neq \beta \neq \gamma} \text{Tr}\{(\hat{O} \otimes \hat{O})\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_\alpha\hat{\rho}_\beta\hat{\rho}_\gamma)]\} - o^2 \\ &= \frac{1}{[n(n-1)(n-2)]^2} \sum_{i \neq j \neq k} \sum_{\alpha \neq \beta \neq \gamma} \text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}). \end{aligned} \quad (\text{D27})$$

We shall analyze the terms $\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma})$ case by case.

a. When $i, j, k, \alpha, \beta, \gamma$ are all distinct

In this case, all terms are independent of all others, and so $\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) = 0$.

b. When i, j, k and α, β, γ have one pair of equal indices

It is sufficient to consider the cases $i = \alpha, i = \beta, i = \gamma$, or $j = \beta$, since all the other cases are equivalent to one of these cases.

$i = \alpha$: In this case, we need to evaluate

$$\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_i\hat{\rho}_\beta\hat{\rho}_\gamma)] = \mathbb{E}[(\hat{\rho}_i\hat{\rho}) \otimes (\hat{\rho}_i\hat{\rho})] = \mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i] \times (\hat{\rho} \otimes \hat{\rho}). \quad (\text{D28})$$

Using Lemma D1, we have

$$\begin{aligned} \text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \text{Tr}\left\{(\hat{O} \otimes \hat{O})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I})\left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I}\right)(\hat{\rho} \otimes \hat{\rho})\right\} - o^2 \\ &= \text{Tr}\left\{(\hat{\rho}\hat{O} \otimes \hat{\rho}\hat{O} + \hat{\rho}\hat{O} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{\rho}\hat{O})\left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I}\right)\right\} - o^2 \\ &\leq 3\text{Tr}(\hat{\rho}\hat{O}\hat{\rho}\hat{O}) - o^2 \\ &\leq 3\text{Tr}(\hat{\rho}\hat{O})^2 \\ &\leq 3\|\hat{O}\|_\infty^2. \end{aligned} \quad (\text{D29})$$

$i = \beta$: In this case, we need to evaluate

$$\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_i\hat{\rho}_\beta\hat{\rho}_\gamma)] = \mathbb{E}[(\hat{\rho}_i\hat{\rho}) \otimes (\hat{\rho}\hat{\rho}_i\hat{\rho})] = (\hat{I} \otimes \hat{\rho}) \times \mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i] \times (\hat{\rho} \otimes \hat{\rho}). \quad (\text{D30})$$

Using Lemma D1, we have

$$\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) = \text{Tr}\left\{(\hat{O} \otimes \hat{O})(\hat{I} \otimes \hat{\rho})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I})\left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I}\right)(\hat{\rho} \otimes \hat{\rho})\right\} - o^2$$

$$\begin{aligned}
&= \text{Tr} \left\{ (2\hat{\rho}\hat{O} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{\rho}\hat{O}\hat{\rho}) \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right) \right\} - o^2 \\
&\leq 3\text{Tr}(\hat{O}\hat{\rho})^2 - o^2 \\
&\leq 3\|\hat{O}\|_\infty^2.
\end{aligned} \tag{D31}$$

Similar calculations show that for any case where only one pair of indices is equal, the covariance is always of the same order, i.e.,

$$\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) \in \mathcal{O}(\|\hat{O}\|_\infty^2). \tag{D32}$$

c. When i, j, k and α, β, γ have two pairs of equal indices

There are six different situations. These are (1) (i, j) and (α, β) are equal, (2) (i, j) and (α, γ) are equal, (3) (i, j) and (β, γ) are equal, (4) (i, k) and (α, γ) are equal, (5) (i, k) and (β, γ) are equal, and (6) (j, k) and (β, γ) are equal. We need to distinguish between two cases in each of these situations. For example, we need to distinguish $i = \alpha$ while $j = \beta$, from $i = \beta$ while $j = \alpha$: We shall refer to these as cyclic and anticyclic terms, respectively. It turns out that the order of magnitude of these terms depends only on whether they are cyclic or anticyclic, but not which one of the six situations they are in. Therefore, we show here only the most convenient example for each type, which is $i = \alpha$ while $j = \beta$ (cyclic) and $i = \beta$ while $j = \alpha$ (anticyclic), to illustrate the calculations.

Cyclic terms: e.g., $i = \alpha$ while $j = \beta$. In this case, we need to evaluate

$$\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_i\hat{\rho}_\beta\hat{\rho}_\gamma)] = \mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_i\hat{\rho}_j\hat{\rho}_\gamma)] = \mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i] \times \mathbb{E}[\hat{\rho}_j \otimes \hat{\rho}_j] \times (\hat{\rho} \otimes \hat{\rho}). \tag{D33}$$

Using Lemma D1, we have

$$\begin{aligned}
\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \text{Tr} \left\{ (\hat{O} \otimes \hat{O})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I})^2 \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right)^2 (\hat{\rho} \otimes \hat{\rho}) \right\} - o^2 \\
&= \text{Tr} \left\{ (\hat{\rho}\hat{O} \otimes \hat{\rho}\hat{O})(\hat{I} \otimes \hat{I} + 3\hat{I} \otimes \hat{\rho} + 3\hat{\rho} \otimes \hat{I} + 2\hat{\rho} \otimes \hat{\rho}) \left(\frac{D^2 + 2D + 2}{(D+2)^2}\hat{I} - \frac{2(D+1)}{(D+2)^2}\hat{W} \right) \right\} - o^2 \\
&\leq 9\text{Tr}(\hat{O}\hat{\rho})^2 - o^2 \\
&\leq 9\|\hat{O}\|_\infty^2.
\end{aligned} \tag{D34}$$

Anticyclic terms: e.g., $i = \beta$ while $j = \alpha$. In this case, we rearrange the covariance to have

$$\begin{aligned}
\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \mathbb{E}[\text{Tr}\{(\hat{O} \otimes \hat{O})(\hat{\rho}_i \otimes \hat{\rho}_j)(\hat{\rho}_j \otimes \hat{\rho}_i)(\hat{\rho}_k \otimes \hat{\rho}_\gamma)\}] \\
&= \mathbb{E}[\text{Tr}\{(\hat{O} \otimes \hat{O})(\hat{\rho}_i \otimes \hat{\rho}_j)(\hat{\rho}_j \otimes \hat{\rho}_i)(\hat{\rho} \otimes \hat{\rho})\}] \\
&= \mathbb{E}[\text{Tr}\{(\hat{O} \otimes \hat{\rho}_i)(\hat{\rho}_i \otimes \hat{\rho})(\hat{\rho}_j \otimes \hat{O})(\hat{\rho} \otimes \hat{\rho}_j)\}] \\
&= \text{Tr}\{(\hat{O} \otimes \hat{I})(\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i])(\hat{I} \otimes \hat{\rho}\hat{O})(\mathbb{E}[\hat{\rho}_j \otimes \hat{\rho}_j])(\hat{\rho} \otimes \hat{I})\}.
\end{aligned} \tag{D35}$$

Using Lemma D1, and ignoring o^2 we have

$$\begin{aligned}
&\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) \\
&\leq \text{Tr} \left\{ (\hat{\rho}\hat{O} \otimes \hat{I})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}) \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right) (\hat{I} \otimes \hat{\rho}\hat{O})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}) \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right) \right\} \\
&= \text{Tr} \left\{ (\hat{\rho}\hat{O} \otimes \hat{I} + \hat{\rho}\hat{O} \otimes \hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{I}) \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right) (\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O}) \left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I} \right) \right\}.
\end{aligned} \tag{D36}$$

We now look at these terms one by one. The term proportional to $\frac{(D+1)^2}{(D+2)^2}$ is

$$\begin{aligned}
&\text{Tr}\{(\hat{\rho}\hat{O} \otimes \hat{I} + \hat{\rho}\hat{O} \otimes \hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{I})\hat{W}(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O})\hat{W}\} \\
&= \text{Tr}\{(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{\rho} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho})(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O})\} = (4D+5)\text{Tr}(\hat{O}\hat{\rho})^2.
\end{aligned} \tag{D37}$$

The term proportional to $\frac{1}{(D+2)^2}$ is

$$\text{Tr}\{(\hat{\rho}\hat{O} \otimes \hat{I} + \hat{\rho}\hat{O} \otimes \hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{I})(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O})\} = 9\text{Tr}(\hat{O}\hat{\rho})^2. \tag{D38}$$

The term proportional to $\frac{D+1}{(D+2)^2}$ is

$$\begin{aligned} & \text{Tr}\{(\hat{\rho}\hat{O} \otimes \hat{I} + \hat{\rho}\hat{O} \otimes \hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{I})(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O})\hat{W}\} \\ & + \text{Tr}\{(\hat{\rho}\hat{O} \otimes \hat{I} + \hat{\rho}\hat{O} \otimes \hat{\rho} + \hat{\rho}\hat{O}\hat{\rho} \otimes \hat{I})\hat{W}(\hat{I} \otimes \hat{\rho}\hat{O} + \hat{I} \otimes \hat{\rho}\hat{O}\hat{\rho} + \hat{\rho} \otimes \hat{\rho}\hat{O})\} \\ & = 18\text{Tr}(\hat{O}\hat{\rho})^2. \end{aligned} \quad (\text{D39})$$

Summing these together, we have

$$\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) \in \mathcal{O}(D\|\hat{O}\|_\infty^2). \quad (\text{D40})$$

d. When i, j, k and α, β, γ has three pairs of equal indices

In this case, α, β, γ is just a permutation of i, j, k , i.e., $(\alpha, \beta, \gamma) = \pi(i, j, k)$, with $\pi \in S_3$. There are six different permutations in S_3 , and similar to the last subsection, the covariance actually only depends on the parity of π . We shall look at the most convenient examples of each of these two situations.

Even parity, $\text{Sign}(\pi) = +1$ with $i = \alpha, j = \beta$, and $k = \gamma$ as the example: In this case, we have

$$\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) = \text{Tr}\{(\hat{O} \otimes \hat{O})\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k)]\} - \sigma^2 = \text{Tr}\{(\hat{O} \otimes \hat{O})\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i]\mathbb{E}[\hat{\rho}_j \otimes \hat{\rho}_j]\mathbb{E}[\hat{\rho}_k \otimes \hat{\rho}_k]\} - \sigma^2. \quad (\text{D41})$$

Using Lemma D1, this is just

$$\begin{aligned} \text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \text{Tr}\left\{(\hat{O} \otimes \hat{O})(\hat{I} \otimes \hat{I} + \hat{I} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I})^3\left(\frac{D+1}{D+2}\hat{W} - \frac{1}{D+2}\hat{I}\right)^3\right\} - \sigma^2 \\ &= \text{Tr}\left\{(\hat{O} \otimes \hat{O})(\hat{I} \otimes \hat{I} + 7\hat{I} \otimes \hat{\rho} + 7\hat{\rho} \otimes \hat{I} + 12\hat{\rho} \otimes \hat{\rho})\left(\frac{(D+1)(D^2+2D+4)}{(D+2)^3}\hat{W} - \frac{3(D+1)^2+1}{(D+2)^3}\hat{I}\right)\right\} - \sigma^2 \\ &= \frac{(D+1)(D^2+2D+4)}{(D+2)^3}\{\text{Tr}(\hat{O}^2) + 14\text{Tr}(\hat{O}^2\hat{\rho}) + 12\text{Tr}(\hat{O}\hat{\rho})^2\} \\ &\quad - \frac{3(D+1)^2+1}{(D+2)^3}\{\text{Tr}(\hat{O})^2 + 14\text{Tr}(\hat{O})\text{Tr}(\hat{O}\hat{\rho}) + 12\text{Tr}(\hat{O}\hat{\rho})^2\} - \sigma^2 \\ &\leq \text{Tr}(\hat{O}^2) + 14\text{Tr}(\hat{O}^2\hat{\rho}) + 12\text{Tr}(\hat{O}\hat{\rho})^2 + \frac{42}{d+2}|\text{Tr}(\hat{O})\text{Tr}(\hat{O}\hat{\rho})| \\ &\leq \text{Tr}(\hat{O}^2) + 26\|\hat{O}\|_\infty^2 + 42\sqrt{\text{Tr}(\hat{O}^2)/D}\|\hat{O}\|_\infty \\ &\in \hat{\mathcal{O}}(\text{Tr}(\hat{O}^2)). \end{aligned} \quad (\text{D42})$$

Odd parity, $\text{Sign}(\pi) = -1$ with $i = \gamma, j = \beta$, and $k = \alpha$ as the example: In this case, we have

$$\text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) = \text{Tr}\{(\hat{O} \otimes \hat{O})\mathbb{E}[(\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k) \otimes (\hat{\rho}_k\hat{\rho}_j\hat{\rho}_i)]\} - \sigma^2. \quad (\text{D43})$$

To estimate this, we need the following equation:

$$(\hat{A} \otimes \hat{C})(\hat{B} \otimes \hat{E}) = \hat{A}\hat{B} \otimes \hat{C}\hat{E} = \text{Tr}_3((\hat{I} \otimes \hat{A} \otimes \hat{E})(\hat{C} \otimes \hat{B} \otimes \hat{I})\hat{W}_{(1,3)}). \quad (\text{D44})$$

Using this, we then have

$$\begin{aligned} \text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \mathbb{E}[\text{Tr}\{(\hat{\rho}_k \otimes \hat{\rho}_i)(\hat{O}\hat{\rho}_i \otimes \hat{O}\hat{\rho}_k)(\hat{\rho}_j \otimes \hat{\rho}_j)\}] - \sigma^2 \\ &= \mathbb{E}[\text{Tr}\{(\hat{I} \otimes \hat{I} \otimes \hat{O})(\hat{I} \otimes \hat{\rho}_k \otimes \hat{\rho}_k)(\hat{I} \otimes \hat{O} \otimes \hat{I})(\hat{\rho}_i \otimes \hat{\rho}_i \otimes \hat{I})\hat{W}_{(1,3)}(\hat{\rho}_j \otimes \hat{\rho}_j \otimes \hat{I})\}] - \sigma^2 \\ &= \text{Tr}\{(\hat{I} \otimes \hat{I} \otimes \hat{O})\mathbb{E}[\hat{I} \otimes \hat{\rho}_k \otimes \hat{\rho}_k](\hat{I} \otimes \hat{O} \otimes \hat{I})\mathbb{E}[\hat{\rho}_i \otimes \hat{\rho}_i \otimes \hat{I}]\hat{W}_{(1,3)}\mathbb{E}[\hat{\rho}_j \otimes \hat{\rho}_j \otimes \hat{I}]\} - \sigma^2. \end{aligned} \quad (\text{D45})$$

For convenience, we now again omit $\otimes$ and write the dot product explicitly. Using Lemma D1, we have

$$\begin{aligned} \text{Cov}(o_{ijk}, o_{\alpha\beta\gamma}) &= \text{Tr}\left\{\hat{I}\hat{I}\hat{O} \times (\hat{I}\hat{I}\hat{I} + \hat{I}\hat{I}\hat{\rho} + \hat{I}\hat{\rho}\hat{I}) \times \left(\frac{D+1}{D+2}\hat{W}_{(2,3)} - \frac{1}{D+2}\hat{I}\right) \times (\hat{I}\hat{O}\hat{I}) \times (\hat{I}\hat{I}\hat{I} + \hat{I}\hat{\rho}\hat{I} + \hat{\rho}\hat{I}\hat{I})\right. \\ &\quad \left. \times \left(\frac{D+1}{D+2}\hat{W}_{(1,2)} - \frac{1}{D+2}\hat{I}\right) \times \hat{W}_{(1,3)} \times (\hat{I}\hat{I}\hat{I} + \hat{I}\hat{\rho}\hat{I} + \hat{\rho}\hat{I}\hat{I}) \times \left(\frac{D+1}{D+2}\hat{W}_{(1,2)} - \frac{1}{D+2}\hat{I}\right)\right\} - \sigma^2. \end{aligned} \quad (\text{D46})$$

The dominant term of the above expression is the term proportional to $\frac{(D+1)^3}{(D+2)^3}$, which we evaluate to be

$$\frac{(D+1)^3}{(D+2)^3}\text{Tr}\{(\hat{I}\hat{I}\hat{O} + \hat{I}\hat{I}(\hat{O}\hat{\rho}) + \hat{I}\hat{\rho}\hat{O}) \times \hat{W}_{(2,3)} \times (\hat{I}\hat{O}\hat{I} + \hat{I}(\hat{O}\hat{\rho})\hat{I} + \hat{\rho}\hat{O}\hat{I}) \times \hat{W}_{(1,2)} \times \hat{W}_{(1,3)} \times (\hat{I}\hat{I}\hat{I} + \hat{I}\hat{\rho}\hat{I} + \hat{\rho}\hat{I}\hat{I}) \times \hat{W}_{(1,2)}\}$$

$$\begin{aligned}
&= \frac{(D+1)^3}{(D+2)^3} \text{Tr}\{(\hat{I}\hat{O}\hat{I} + \hat{I}(\hat{O}\hat{\rho})\hat{I} + \hat{I}\hat{O}\hat{\rho}) \times (\hat{I}\hat{O}\hat{I} + \hat{I}(\hat{O}\hat{\rho})\hat{I} + \hat{\rho}\hat{O}\hat{I}) \times (\hat{I}\hat{I}\hat{I} + \hat{\rho}\hat{I}\hat{I} + \hat{I}\hat{I}\hat{\rho})\hat{W}_{(1,2)} \times \hat{W}_{(1,3)} \times \hat{W}_{(1,2)} \times \hat{W}_{(2,3)}\} \\
&= \frac{(D+1)^3}{(D+2)^3} \text{Tr}\{(\hat{I}\hat{O}\hat{I} + \hat{I}(\hat{O}\hat{\rho})\hat{I} + \hat{I}\hat{O}\hat{\rho}) \times (\hat{I}\hat{O}\hat{I} + \hat{I}(\hat{O}\hat{\rho})\hat{I} + \hat{\rho}\hat{O}\hat{I}) \times (\hat{I}\hat{I}\hat{I} + \hat{\rho}\hat{I}\hat{I} + \hat{I}\hat{I}\hat{\rho})\} \\
&= \frac{(D+1)^3}{(D+2)^3} \{(D^2 + 6D + 5)\text{Tr}(\hat{O}^2) + (2d^2 + 6d + 5)\text{Tr}(\hat{O}^2\hat{\rho}) + (D^2 + 2D)\text{Tr}(\hat{O}\hat{\rho})^2 + (2D + 1)\text{Tr}(\hat{O}\hat{\rho})\} \\
&\in \hat{O}(D^2\text{Tr}(\hat{O}^2)).
\end{aligned} \tag{D47}$$

e. Final result

Putting all the above contributing terms together, we find that the variance of the order-3 U-statistics estimator is

$$\text{Var}[o_{\text{es}}] \in \mathcal{O}\left(\frac{n^5 + n^4D + n^3D^2B}{n^6}\right) = \mathcal{O}\left(\frac{1}{n} + \frac{D}{n^2} + \frac{D^2B}{n^3}\right) = \mathcal{O}\left(\frac{1}{n} + \frac{D^2B}{n^3}\right). \tag{D48}$$

Therefore, to estimate o within error ϵ , we need to make sure the variance is smaller than ϵ^2 , which gives the overall sample complexity:

$$n = \mathcal{O}\left(\frac{1}{\epsilon^2} + \frac{D^{2/3}B^{1/3}}{\epsilon^{2/3}}\right). \tag{D49}$$

5. Shadow estimation of bilinear functions with U-statistics

Here, we discuss the sample complexity of estimating bilinear functions of the form $o = \text{Tr}(\hat{O}\hat{\rho}\hat{\sigma}) + \text{Tr}(\hat{O}\hat{\sigma}\hat{\rho})$ using the order-3 U-statistics that we used in the main text. The estimator of o is then

$$o_{\text{es}} = \frac{1}{n^2(n-1)^2(n-2)^2} \sum_{i \neq j \neq k} \sum_{p \neq q \neq r} \text{Tr}\{\hat{O}\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k\hat{\sigma}_p\hat{\sigma}_q\hat{\sigma}_r\} + \text{Tr}\{\hat{O}\hat{\sigma}_p\hat{\sigma}_q\hat{\sigma}_r\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k\}. \tag{D50}$$

Again, the sample complexity is determined by the variance

$$\begin{aligned}
&n^4(n-1)^4(n-2)^4 \text{Var}(o_{\text{es}}) \\
&= \sum_{i_1 \neq j_1 \neq k_1} \sum_{i_2 \neq j_2 \neq k_2} \sum_{p_1 \neq q_1 \neq r_1} \sum_{p_2 \neq q_2 \neq r_2} \text{Tr}\{(\hat{I} \otimes \hat{I} \otimes \hat{O} \otimes \hat{O} + \hat{I} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I} + \hat{O} \otimes \hat{I} \otimes \hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{O} \otimes \hat{I} \otimes \hat{I}) \\
&\quad \times \mathbb{E}[\hat{\rho}_{i_1}\hat{\rho}_{j_1}\hat{\rho}_{k_1} \otimes \hat{\rho}_{i_2}\hat{\rho}_{j_2}\hat{\rho}_{k_2} \otimes \hat{\sigma}_{p_1}\hat{\sigma}_{q_1}\hat{\sigma}_{r_1} \otimes \hat{\sigma}_{p_2}\hat{\sigma}_{q_2}\hat{\sigma}_{r_2}]\hat{W}_{(1,3)}\hat{W}_{(2,4)}\} - o^2,
\end{aligned} \tag{D51}$$

and the derivation of the sample complexity is very similar to that in Appendixes D3 and D4 but involves much more calculation. Instead of presenting the full calculations here, we provide only a sketch of the derivation and focus here on the main contributing sources of the variance.

The first main contribution to the variance comes from the terms where only one pair of indices in the sets $i_1, i_2, j_1, j_2, k_1, k_2$ and $p_1, p_2, q_1, q_2, r_1, r_2$ is equal. For example, $i_1 = i_2$ while (j_1, j_2, k_1, k_2) are distinct and $(p_1, p_2, q_1, q_2, r_1, r_2)$ are also distinct. A single such term contributes $\mathcal{O}(\|\hat{O}\|_\infty) = \mathcal{O}(1)$ to the variance. There are in total $\mathcal{O}(n^{11})$ such terms, leading to a total contribution of $\mathcal{O}(n^{11})$ to the variance (ignoring the $\sim 1/n^{12}$ prefactor).

The second main contribution to the variance comes from the terms with six pairs of indices that are equal. We focus on the example where $i_1 = k_2, j_1 = j_2, k_1 = i_2$ and $p_1 = r_2, q_1 = q_2, r_1 = p_2$. In this case, the contribution to the variance (ignoring the $\sim 1/n^{12}$ prefactor) is

$$\begin{aligned}
&\text{Tr}\{(\hat{I} \otimes \hat{I} \otimes \hat{O} \otimes \hat{O} + \hat{I} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I} + \hat{O} \otimes \hat{I} \otimes \hat{I} \otimes \hat{O} + \hat{O} \otimes \hat{O} \otimes \hat{I} \otimes \hat{I})\mathbb{E}[\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k \otimes \hat{\rho}_k\hat{\rho}_j\hat{\rho}_i \otimes \hat{\sigma}_p\hat{\sigma}_q\hat{\sigma}_r \otimes \hat{\sigma}_r\hat{\sigma}_q\hat{\sigma}_p]\hat{W}_{(1,3)}\hat{W}_{(2,4)}\} \\
&= \text{Tr}\{(\hat{I}^{\otimes 6} \otimes \hat{O} \otimes \hat{O} + \hat{I}^{\otimes 3} \otimes \hat{O} \otimes \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I}^{\otimes 4} \otimes \hat{O} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I}^{\otimes 4}) \\
&\quad \times \mathbb{E}[(\hat{\rho}_k \otimes \hat{I} \otimes \hat{I} \otimes \hat{\rho}_k)(\hat{I} \otimes \hat{\rho}_i \otimes \hat{\rho}_i \otimes \hat{I})(\hat{I} \otimes \hat{I} \otimes \hat{\rho}_j \otimes \hat{\rho}_j) \otimes (\hat{\sigma}_r \otimes \hat{I} \otimes \hat{I} \otimes \hat{\sigma}_r)(\hat{I} \otimes \hat{\sigma}_p \otimes \hat{\sigma}_p \otimes \hat{I})(\hat{I} \otimes \hat{I} \otimes \hat{\sigma}_q \otimes \hat{\sigma}_q)] \\
&\quad \times \hat{W}_{(1,3)}\hat{W}_{(2,4)}\hat{W}_{(5,7)}\hat{W}_{(6,8)}\hat{W}_{(3,7)}\hat{W}_{(4,8)}\},
\end{aligned} \tag{D52}$$

where in the equality we have used the fact

$$\hat{A}\hat{B}\hat{C} \otimes \hat{E}\hat{F}\hat{G} = \text{Tr}_{1,2}\{(\hat{C} \otimes \hat{I} \otimes \hat{I} \otimes \hat{E})(\hat{I} \otimes \hat{G} \otimes \hat{A} \otimes \hat{I})(\hat{I} \otimes \hat{I} \otimes \hat{B} \otimes \hat{F})\hat{W}_{(1,3)}\hat{W}_{(2,4)}\}. \tag{D53}$$

By using Lemma D1, we are able to identify the leading order contribution of Eq. (D52) as

$$\begin{aligned}
&\left(\frac{D+1}{D+2}\right)^6 \text{Tr}\{(\hat{I}^{\otimes 6} \otimes \hat{O} \otimes \hat{O} + \hat{I}^{\otimes 3} \otimes \hat{O} \otimes \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I}^{\otimes 4} \otimes \hat{O} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I}^{\otimes 4}) \\
&\quad \times (\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}^{\otimes 3})\hat{W}_{(1,4)}(\hat{I}^{\otimes 4} + \hat{I} \otimes \hat{\rho} \otimes \hat{I}^{\otimes 2} + \hat{I}^{\otimes 2} \otimes \hat{\rho} \otimes \hat{I})\hat{W}_{(2,3)}(\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \otimes \hat{\rho} + \hat{I}^{\otimes 2} \otimes \hat{\rho} \otimes \hat{I})\hat{W}_{(3,4)}\}
\end{aligned}$$

$$\begin{aligned}
& \otimes (\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \otimes \hat{\sigma} + \hat{\sigma} \otimes \hat{I}^{\otimes 3}) \hat{W}_{(5,8)} (\hat{I}^{\otimes 4} + \hat{I} \otimes \hat{\sigma} \otimes \hat{I}^{\otimes 2} + \hat{I}^{\otimes 2} \otimes \hat{\sigma} \otimes \hat{I}) \hat{W}_{(6,7)} (\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \hat{\sigma} + \hat{I}^{\otimes 2} \otimes \hat{\sigma} \otimes \hat{I}) \hat{W}_{(7,8)} \\
& \times \hat{W}_{(1,3)} \hat{W}_{(2,4)} \hat{W}_{(5,7)} \hat{W}_{(6,8)} \hat{W}_{(3,7)} \hat{W}_{(4,8)} \} \\
& \leq \text{Tr} \{ (\hat{I}^{\otimes 6} \otimes \hat{O} \otimes \hat{O} + \hat{I}^{\otimes 3} \otimes \hat{O} \otimes \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{I}^{\otimes 4} \otimes \hat{O} + \hat{I}^{\otimes 2} \otimes \hat{O} \otimes \hat{O} \otimes \hat{I}^{\otimes 4}) \\
& \times (\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \otimes \hat{\rho} + \hat{\rho} \otimes \hat{I}^{\otimes 3}) (\hat{I}^{\otimes 4} + \hat{I} \otimes \hat{\rho} \otimes \hat{I}^{\otimes 2} + \hat{I}^{\otimes 2} \otimes \hat{\rho} \otimes \hat{I}) (\hat{I}^{\otimes 4} + \hat{\rho} \otimes \hat{I}^{\otimes 3} + \hat{I} \otimes \hat{\rho} \otimes \hat{I}^{\otimes 2}) \\
& \otimes (\hat{I}^{\otimes 4} + \hat{I}^{\otimes 3} \otimes \hat{\sigma} + \hat{\sigma} \otimes \hat{I}^{\otimes 3}) (\hat{I}^{\otimes 4} + \hat{I} \otimes \hat{\sigma} \otimes \hat{I}^{\otimes 2} + \hat{I}^{\otimes 2} \otimes \hat{\sigma} \otimes \hat{I}) (\hat{I}^{\otimes 4} + \hat{\sigma} \otimes \hat{I}^{\otimes 3} + \hat{I} \otimes \hat{\sigma} \otimes \hat{I}^{\otimes 2}) \times \hat{W}_{(3,8)} \hat{W}_{(4,7)} \} \\
& \in \hat{O}(D^5 \text{Tr}(\hat{O}^2)).
\end{aligned} \tag{D54}$$

There are $\mathcal{O}(n^6)$ such terms, and therefore they contribute $\mathcal{O}(n^6 D^5 B)$ to the variance.

It can be shown that all other types of terms contribute the same or a smaller order of magnitude to the variance as $\mathcal{O}(n^{11})$ or $\mathcal{O}(n^6 D^5 B)$. Therefore, the total variance is given by

$$\text{Var}[o_{\text{es}}] \in \mathcal{O}\left(\frac{n^{11} + n^6 D^5 B}{n^{12}}\right) = \mathcal{O}\left(\frac{1}{n} + \frac{D^5 B}{n^6}\right). \tag{D55}$$

Finally, to ensure the estimation of the bilinear function o is within an additive error ϵ , we need $\text{Var}[o_{\text{es}}] \leq \epsilon^2$, which in turn gives the sample complexity

$$n = \mathcal{O}\left(\frac{1}{\epsilon^2} + \frac{D^{5/6} B^{1/6}}{\epsilon^{1/3}}\right). \tag{D56}$$

-
- [1] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys.* **92**, 015003 (2020).
- [2] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [3] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).
- [4] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, A review of barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [5] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, [arXiv:quant-ph/9511026](https://arxiv.org/abs/quant-ph/9511026).
- [6] D. S. Abrams and S. Lloyd, Simulation of many-body Fermi systems on a universal quantum computer, *Phys. Rev. Lett.* **79**, 2586 (1997).
- [7] U. Baek, D. Hait, J. Shee, O. Leimkuhler, W. J. Huggins, T. F. Stetina, M. Head-Gordon, and K. B. Whaley, Say no to optimization: A nonorthogonal quantum eigensolver, *PRX Quantum* **4**, 030307 (2023).
- [8] G. W. Pratt Jr., Unrestricted Hartree-Fock method, *Phys. Rev.* **102**, 1303 (1956).
- [9] B. Ghogh, F. Karray, and M. Crowley, Eigenvalue and generalized eigenvalue problems: Tutorial, [arXiv:1903.11240](https://arxiv.org/abs/1903.11240).
- [10] B. O’Gorman, W. J. Huggins, E. G. Rieffel, and K. B. Whaley, Generalized SWAP networks for near-term quantum computing, [arXiv:1905.05118](https://arxiv.org/abs/1905.05118).
- [11] V. Verteletskiy, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* **152**, 124114 (2020).
- [12] N. C. Rubin, R. Babbush, and J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, *New J. Phys.* **20**, 053020 (2018).
- [13] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *npj Quantum Inf.* **7**, 23 (2021).
- [14] W. van Dam, H. Liu, G. H. Low, A. Paetznick, A. Paz, M. Silva, A. Sundaram, K. Svore, and M. Troyer, End-to-end quantum simulation of a chemical system, [arXiv:2409.05835](https://arxiv.org/abs/2409.05835).
- [15] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [16] A. Seif, Z.-P. Cui, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, *PRX Quantum* **4**, 010303 (2023).
- [17] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Quantum-selected configuration interaction: Classical diagonalization of Hamiltonians in subspaces selected by quantum computers, [arXiv:2302.11320](https://arxiv.org/abs/2302.11320).
- [18] R. M. Parrish and P. L. McMahon, Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation, [arXiv:1909.08925](https://arxiv.org/abs/1909.08925).
- [19] C. L. Cortes and S. K. Gray, Quantum Krylov subspace algorithms for ground-and excited-state energy estimation, *Phys. Rev. A* **105**, 022417 (2022).
- [20] S. Piccinelli, A. Baiardi, M. Rossmannek, A. C. Vazquez, F. Tacchino, S. Mensa, E. Altamura, A. Alavi, M. Motta, J. Robledo-Moreno, *et al.*, Quantum chemistry with provable convergence via randomized sample-based quantum diagonalization, [arXiv:2508.02578](https://arxiv.org/abs/2508.02578).

- [21] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, T. Shirakawa, *et al.*, Chemistry beyond the scale of exact diagonalization on a quantum-centric supercomputer, *Sci. Adv.* **11**, eadu9991 (2025).
- [22] Quantinuum, Quantinuum System Model H2, 2024, <https://www.quantinuum.com/products-solutions/quantinuum-systems/system-model-h2>, accessed 21 November 2024.
- [23] W. J. Huggins, J. Lee, U. Baek, B. O’Gorman, and K. B. Whaley, A non-orthogonal variational quantum eigensolver, *New J. Phys.* **22**, 073009 (2020).
- [24] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K. Chan, Low rank representations for quantum simulation of electronic structure, *npj Quantum Inf.* **7**, 83 (2021).
- [25] A. Bonfiglioli and R. Fulci, *Topics in Noncommutative Algebra: The Theorem of Campbell, Baker, Hausdorff and Dynkin*, Lecture Notes in Mathematics (Springer, Heidelberg, Germany, 2011), Vol. 2034.
- [26] G. W. Stewart, Perturbation bounds for the definite generalized eigenvalue problem, *Linear Algebra Appl.* **23**, 69 (1979).
- [27] R. Mathias and C.-K. Li, The definite generalized eigenvalue problem: A new perturbation theory, Manchester Centre Computational Mathematics, NAREP, 2004.
- [28] E. N. Epperly, L. Lin, and Y. Nakatsukasa, A theory of quantum subspace diagonalization, *SIAM J. Matrix Anal. Appl.* **43**, 1263 (2022).
- [29] G. Lee, D. Lee, and J. Huh, Sampling error analysis in quantum Krylov subspace diagonalization, *Quantum* **8**, 1477 (2024).
- [30] S. Kotz and N. L. Johnson, *Breakthroughs in Statistics Volume I: Foundations and Basic Theory*, Springer Series in Statistics (Springer-Verlag, Berlin, Germany, 2012).
- [31] D. Grier, H. Pashayan, and L. Schaeffer, Sample-optimal classical shadows for pure states, *Quantum* **8**, 1373 (2024).
- [32] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
- [33] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, Shallow shadows: Expectation estimation using low-depth random Clifford circuits, *Phys. Rev. Lett.* **133**, 020602 (2024).
- [34] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, *Science* **389**, 92 (2025).
- [35] Y. Zhang, P. Lewalle, and K. B. Whaley, Solving k -SAT problems with generalized quantum measurement, *npj Quantum Inf.* **11**, 170 (2025).
- [36] P. Lewalle, Y. Zhang, and K. B. Whaley, Optimal Zeno dragging for quantum control: A shortcut to Zeno with action-based scheduling optimization, *PRX Quantum* **5**, 020366 (2024).
- [37] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [38] 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).
- [39] B. Koczor, Exponential error suppression for near-term quantum devices, *Phys. Rev. X* **11**, 031057 (2021).
- [40] A. Zhao and A. Miyake, Group-theoretic error mitigation enabled by classical shadows and symmetries, *npj Quantum Inf.* **10**, 57 (2024).
- [41] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, and Z. Wang, Logical shadow tomography: Efficient estimation of error-mitigated observables, [arXiv:2203.07263](https://arxiv.org/abs/2203.07263).
- [42] H. Jnane, J. Steinberg, Z. Cai, H. C. Nguyen, and B. Koczor, Quantum error mitigated classical shadows, *PRX Quantum* **5**, 010324 (2024).
- [43] H. Ren, Y. Zhang, Z. Zheng, C. Ying, L. Xu, M. Rahmani, and K. B. Whaley, Error mitigated metasurface-based randomized measurement schemes, *Phys. Rev. Res.* **6**, 033310 (2024).
- [44] H.-Y. Hu, A. Gu, S. Majumder, H. Ren, Y. Zhang, D. S. Wang, Y.-Z. You, Z. Mineev, S. F. Yelin, and A. Seif, Demonstration of robust and efficient quantum property learning with shallow shadows, *Nat. Commun.* **16**, 2943 (2025).
- [45] C. A. Coulson and I. Fischer, XXXIV. Notes on the molecular orbital treatment of the hydrogen molecule, *London Edinburgh Dublin Philos. Mag. J. Sci.* **40**, 386 (1949).
- [46] W. J. Hehre, R. F. Stewart, and J. A. Pople, Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals, *J. Chem. Phys.* **51**, 2657 (1969).
- [47] Wikipedia, Computational Chemistry, 2024, accessed 24 November 2024.
- [48] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [49] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2010).
- [50] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* **8**, 031027 (2018).
- [51] J. Cheng, Z. Liang, R. Yang, H. Ren, Y. Shi, T. Li, and X. Qian, Fidelity estimator, randomized benchmarking and ZNE for quantum pulses, [arXiv:2305.12597](https://arxiv.org/abs/2305.12597).
- [52] R. LaRose, A. Mari, N. Shammah, P. Karalekas, and W. Zeng, Mitiq: A software package for error mitigation on near-term quantum computers (2020), <https://github.com/unitaryfund/mitiq>.
- [53] W. J. Huggins, K. Wan, J. McClean, T. E. O’Brien, N. Wiebe, and R. Babbush, Nearly optimal quantum algorithm for estimating multiple expectation values, *Phys. Rev. Lett.* **129**, 240501 (2022).
- [54] A. Gilyén, S. Arunachalam, and N. Wiebe, Optimizing quantum optimization algorithms via faster quantum gradient computation, in *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, edited by Timothy M. Chan (SIAM, San Diego, CA, 2019), pp. 1425–1444.
- [55] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, *Phys. Rev. Lett.* **129**, 020502 (2022).
- [56] H. Ren, Data of “Error-mitigated nonorthogonal quantum eigensolver via shadow tomography”, GitHub (2025), <https://github.com/whaley-group-berkeley/Data-EM-NOQE-via-Shadow-Tomography>.
- [57] J. A. Smolin and D. P. DiVincenzo, Five two-bit quantum gates are sufficient to implement the quantum Fredkin gate, *Phys. Rev. A* **53**, 2855 (1996).
- [58] P. M. Q. Cruz and B. Murta, Shallow unitary decompositions of quantum Fredkin and Toffoli gates for connectivity-aware equivalent circuit averaging, *APL Quantum* **1**, 016105 (2024).

- [59] A. Aavadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, [arXiv:2405.08810](#).
- [60] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, *Phys. Rev. A* **96**, 022330 (2017).
- [61] Z. Liang, J. Cheng, Z. Song, H. Ren, R. Yang, K. Liu, P. Kogge, T. Li, Y. Ding, and Y. Shi, Towards advantages of parameterized quantum pulses, [arXiv:2304.09253](#).
- [62] J. Cheng, Y. Zhu, Y. Zhou, H. Ren, Z. Song, and Z. Liang, EPOC: A novel pulse generation framework incorporating advanced synthesis techniques for quantum circuits, [arXiv:2405.03804](#).
- [63] R. Vershynin, Spectral norm of products of random and deterministic matrices, *Probab. Theory Relat. Fields* **150**, 471 (2011).
- [64] H. Ren and Y. Li, Modeling quantum devices and the reconstruction of physics in practical systems, *Phys. Rev. Lett.* **123**, 140405 (2019).
- [65] R. Shaffer, H. Ren, E. Dyrenkova, C. G. Yale, D. S. Lobser, A. D. Burch, M. N. Chow, M. C. Reville, S. M. Clark, and H. Häffner, Sample-efficient verification of continuously-parameterized quantum gates for small quantum processors, *Quantum* **7**, 997 (2023).