

Decoherence-free quantum error mitigation by density matrix vectorization

Zhong-Xia Shang^{1,2,3,*}, Zi-Han Chen^{1,2,3,†} and Cai-Sheng Cheng^{4,‡}

¹*Hefei National Research Center for Physical Sciences at the Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei 230026, China*

²*Shanghai Research Center for Quantum Science and CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China*

³*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*

⁴*Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*



(Received 9 December 2024; accepted 3 May 2026; published 26 May 2026)

Fighting against noise is crucial for Noisy Intermediate Scale Quantum (NISQ) devices to demonstrate practical quantum applications. In this work, we give a paradigm of quantum error mitigation based on the vectorization of density matrices for variational quantum algorithms. Different from the ideas of existing quantum error mitigation methods that try to distill noiseless information from noisy quantum states, our proposal directly changes the way of encoding information and maps the density matrices of noisy quantum states to noiseless virtual pure states, which is realized by an NISQ-friendly measurement protocol and a classical postprocessing procedure. Our protocol requires no knowledge of the noise model, no ability to tune the noise strength, and no ancilla qubits for complicated controlled unitaries. Under our encoding, NISQ devices are always preparing virtual pure quantum states that are highly desired resources for variational quantum algorithms to have good performance in many tasks. We show how this protocol can be well fitted into variational quantum algorithms. We give several concrete ansatz constructions that are suitable for our proposal and do theoretical analysis on the sampling complexity, the expressibility, and the trainability. We also give a discussion on how this protocol is influenced by large noise and how it can be well combined with other quantum error mitigation protocols. The effectiveness of our proposal is demonstrated by various numerical experiments.

DOI: [10.1103/kzyj-j3g8](https://doi.org/10.1103/kzyj-j3g8)

I. INTRODUCTION

Quantum computers [1] are expected to take advantage of quantum mechanical phenomena such as quantum superposition and quantum entanglement to give speedup to some classically difficult problems. Such speedup has been theoretically discovered by various quantum algorithms such as the quantum simulation [2], Shor's factoring algorithm [3], and Grover's searching algorithm [4]. To build a quantum computer and show such speedup in reality, the biggest challenge is to fight against noise and protect fragile quantum mechanical behaviors [5]. While there have been significant theoretical developments for fault-tolerant quantum computation [6–13] as well as recent experimental demonstrations [14–21] of quantum error correction and logical operations, in practice, in order to obtain a few logical qubits with highly suppressed logical error rate, simultaneously scaling up the number of

physical qubits and maintaining a low enough physical error rate is still challenging.

Currently, we are in the Noisy Intermediate Scale Quantum (NISQ) era [22], where we lack the ability to build qubits and implement quantum circuits on the physical level with high qualities on a large scale that meet the thresholds for fault-tolerant quantum computing. Thus, NISQ devices are not capable of implementing most of the quantum algorithms that require a deep quantum circuit and a large number of qubits [23]. To make current quantum computers useful, especially under the situation where quantum advantages have already been demonstrated using these devices [24,25], researchers have paid much attention to designing NISQ algorithms [26]. One of the most popular types is the variational quantum algorithms (VQAs) [27] including the Variational Quantum Eigensolver [28] and the Quantum Approximate Optimization Algorithm [29]. The workflow of VQAs is a combination of both quantum and classical computing. We typically use a parametrized quantum circuit to prepare a trial state and measure a cost function corresponding to the expectation value of an observable. The cost function is then fed into a classical optimizer to update parameters in the quantum circuit. This process will repeat until the cost function converges. The parametrized quantum circuit depth is relatively shallower than other algorithms and the optimization procedure has some degree of resilience of the noise. These features make the VQAs more friendly to NISQ quantum computers and

*Contact author: ustcszx@mail.ustc.edu.cn

†Contact author: czh007@mail.ustc.edu.cn

‡Contact author: ccs112202@mail.ustc.edu.cn

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

have been designed for solving problems in various areas [27] including combinatorial optimization [29], quantum chemistry [28], machine learning [30], and simulating quantum dynamics [31].

While VQAs have relatively low requirements for quantum devices, the non-neglectable noise in the devices can still severely reduce their performance in approximating the solutions to the target problems. This can be seen from three perspectives. First, for general settings, the desired target quantum states are pure while NISQ devices can hardly prepare pure states due to the decoherence led by noise, which sets a gap between the reachable VQA results and the true results [32,33]. Second, since the noise grows exponentially as the number of gates grows, NISQ devices can only build shallow quantum circuits with local connectivity to prepare quantum states with limited expressibility [34] and entanglement [35] that may not cover the desired states, which is another gap limiting the accuracy. Third, it has been shown that noise can introduce barren plateau problems [36] to make VQAs untrainable.

All these issues give strong demands on developing methods beyond QECs to suppress the influence of noise. One approach for this purpose is to introduce classical strategies including neural networks [37], Clifford circuits [38], and tensor networks [39] to reduce the circuit depth required on quantum devices. The other approaches [40–47], while realized in different ways, are guided by the philosophy of distilling the noiseless information from noisy quantum states without actually recovering them, which are known as the quantum error mitigation (QEM) methods [48]. Since the mitigation procedures are done after the occurrence of noise, the additional overhead of QEM is exponential [48] rather than polynomial as in QEC; as compensation, QEM has low requirements on hardware, which makes QEM a promising tool for NISQ experiments. However, most of the existing QEM protocols have their own additional unfriendly requirements ranging from prior knowledge of the noise model [40,41] or the symmetry structures of problems [43,44,46] to the ability to tune hardware noise strength [40,42] or do complex indirect collective measurements [45,47].

In this work, we will introduce a QEM paradigm for VQAs based on density matrix vectorization (DMV). Unlike the philosophy of extracting noiseless information from noisy quantum states in existing QEM protocols, our protocol realizes error mitigation by changing the way of encoding information. We encode pure states into linear combinations of vectorized density matrices, which directly leads to decoherence-free virtual pure state preparations from noisy quantum circuits. While the sampling overhead of our protocol to unconditionally eliminate decoherence is still exponential and comparable with other QEM methods such as the probabilistic error cancellation [40,41], the way it mitigates noise might open up possibilities to go beyond recent results on the limitations of QEM [49–52]. By our protocol, the performance gap induced by decoherence can be directly eliminated and the gap led by low expressibility can be well mitigated without introducing additional barren plateau resources. Our protocol requires no additional ancilla qubits and only requires two-qubit collective unitaries before measurements to extract the needed information. In the following

sections, we will show details of our QEM protocol from its basic framework to its VQA applications.

II. DENSITY MATRIX VECTORIZATION

A. QEM by DMV

In NISQ systems, since a pure quantum state can suffer from noise in the environment and thus becomes a mixed state, the wave function description is no longer universal; instead, the density matrix description is adopted to describe pure states and mixed states in a unified language. The basic idea of our protocol is inspired by the mapping

$$\rho \rightarrow |\rho\rangle, \quad (1)$$

where $\rho = \sum_{ij} \rho_{ij} |i\rangle\langle j|$ and $|\rho\rangle$ is defined as $\frac{1}{C_\rho} \sum_{ij} \rho_{ij} |i\rangle|j\rangle$ with the normalization factor $C_\rho = \|\rho\|_F = \sqrt{\sum_{ij} |\rho_{ij}|^2} = \sqrt{\text{Tr}(\rho^2)}$. While DMV has been vastly used as a useful mathematical trick for simplifying many concepts in quantum information science, in this work, we see DMV in a different way. The main concept transition here is that the mapping [Eq. (1)] means that we are treating an n -qubit density matrix ρ as a $2n$ -qubit pure state $|\rho\rangle$, which shares the same idea as in Ref. [53] to demonstrate an exponential quantum speedup. By this encoding, any quantum state either pure or mixed will always be mapped to a pure state. Note that the reverse mapping $|\rho\rangle \rightarrow \rho$ has also been adopted in several works [54,55] for simulating open quantum systems using unitary circuits.

Currently, due to the Hermiticity and the positive semidefiniteness of ρ , we cannot encode all $2n$ -qubit pure states. To solve this problem, we can introduce a generalized mapping f :

$$\begin{aligned} & \{\rho_1, \rho_2, \dots, \rho_K, c_1, c_2, \dots, c_K\} \\ & \rightarrow |\psi\rangle = \frac{1}{C_\psi} \sum_{ij} (c_1 \rho_{1,ij} + c_2 \rho_{2,ij} + \dots + c_K \rho_{K,ij}) |i\rangle|j\rangle, \end{aligned} \quad (2)$$

where we use K density matrices with K complex coefficients to form a pure state $|\psi\rangle$ with the normalization factor $C_\psi := (\sum_{k,l=1}^K c_k^* c_l \text{Tr}(\rho_k \rho_l))^{1/2}$. These density matrix vectorization procedures are visualized in Fig. 1. While the value of K and the value of coefficients can be chosen arbitrarily, it is sufficient to set $K = 4$ with two real and two imaginary coefficients to express any pure state $|\psi\rangle$ (see Appendix D [Eq. (D2)] for a constructive four-term decomposition derived from the Schmidt decomposition). This can be understood from the matrix form $|\psi\rangle \rightarrow \psi$, which can be decomposed first into Hermitian part and i *Hermitian (anti-Hermitian) part; each Hermitian part is indefinite and can be further expressed as a linear combination of two positive semidefinite density matrices.

The mapping [Eq. (2)] currently is only a mathematical transformation whose unnormalized version has been vastly adopted as a convenient representation of quantum elements [56]. To give this mapping a real sense, we need to be able to extract information of $|\psi\rangle$ from $\{\rho_1, \rho_2, \dots, \rho_K, c_1, c_2, \dots, c_K\}$. Interestingly, we find that, given a Hamiltonian H_A , the expectation value with respect

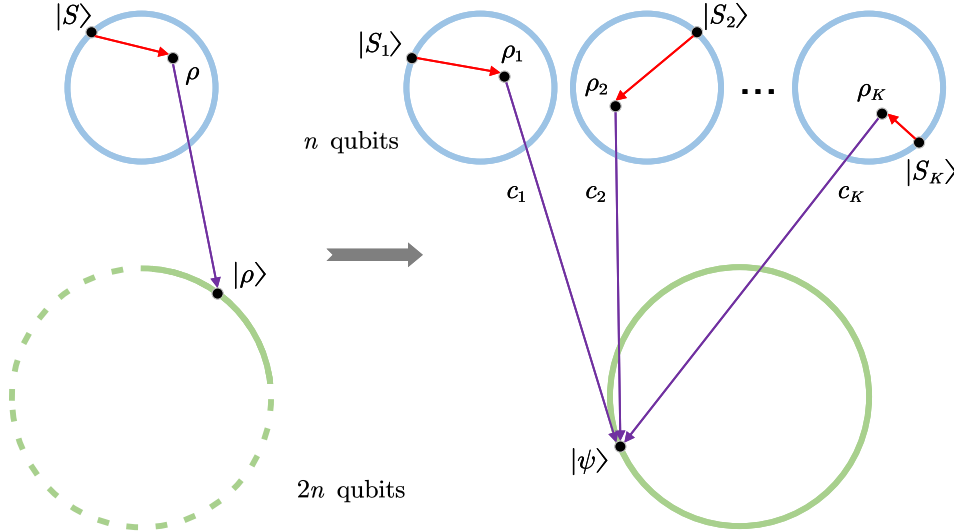


FIG. 1. The basic idea of quantum error mitigation by density matrix vectorization. Ideally, we want to prepare an n -qubit pure state $|S\rangle$ that, due to the noise in NISQ circuits, becomes ρ , which is generally a mixed state. Under the mapping [Eq. (1)], the noisy state ρ is then treated as a $2n$ -qubit pure state $|\rho\rangle$. Thus, in this way, any n -qubit state is unconditionally mapped to a $2n$ -qubit pure state. Due to the Hermiticity and the positive semidefiniteness of density matrices, $|\rho\rangle$ cannot represent arbitrary $2n$ -qubit pure states; thus, the mapping [Eq. (2)] is introduced where any $2n$ -qubit pure state $|\psi\rangle$ can be represented as a linear combination of many density matrices.

to $|\psi\rangle$ can be reexpressed as

$$\langle\psi|H_A|\psi\rangle = \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{k,l=1}^K c_k^* c_l \text{Tr}(\rho_k \rho_l)}, \quad (3)$$

with

$$\langle il|H_B|jk\rangle = \langle ij|H_A|kl\rangle. \quad (4)$$

We call H_B the substitute Hamiltonian of H_A . Equations (3) and (4) mean that by measuring the values of all $\text{Tr}(H_B \rho_i \otimes \rho_j)$ and $\text{Tr}(\rho_i \rho_j)$, one can obtain the value of $\langle\psi|H_A|\psi\rangle$ by classical post combination. We want to mention that if H_A has an odd number of qubits, we can always add an empty qubit (appending identity operators) to make the qubit number even.

We now show an efficient measurement procedure for NISQ hardware to evaluate the expectation value [Eq. (3)]. We can express H_A under the multiqubit Pauli basis as

$$H_A = \sum_{\alpha=1}^m g_{\alpha} P_{\alpha}, \quad (5)$$

where P_{α} are $2n$ -qubit Pauli operators and g_{α} are real numbers. When H_A is transformed into H_B , each P_{α} is transformed into Q_{α} as

$$H_B = \sum_{\alpha=1}^m g_{\alpha} Q_{\alpha}. \quad (6)$$

In the smallest case of $n = 1$, we show the basic substitute operators of all 16 two-qubit Pauli operators in Table I. Note that all 16 basic substitute operators are unitary; among them, the first four (IDs 1–4 in Table I) are also Hermitian. Moreover, the transformation rule in Eq. (4) preserves the tensor product structure. For instance, without loss of generality, consider the Hamiltonian H_A to be of a tensor product form $H_A := H_{A,1} \otimes \cdots \otimes H_{A,t}$. We can construct an operator

$H_B := H_{B,1} \otimes \cdots \otimes H_{B,t}$ with $\langle i_m l_m | H_{B,m} | j_m k_m \rangle = \langle i_m j_m | H_{A,m} | k_m l_m \rangle$ for all $1 \leq m \leq t$ such that operators H_A and H_B satisfy the relation in Eq. (4). Thus, for $n \geq 1$ cases, P_{α} as the tensor product of two-qubit Pauli operators leads to Q_{α} as the tensor product of basic substitute operators, which is also unitary [53] (see Appendix B 2 for a tensor-network proof). In fault-tolerant quantum computers, we can use the Hadamard test circuits [57] and swap test circuits [58] to evaluate $\text{Tr}(H_B \rho_i \otimes \rho_j)$ and $\text{Tr}(\rho_i \rho_j)$.

For NISQ hardware, it is generally difficult to build such controlled circuits with high fidelities. Therefore, we can in-

TABLE I. The 16 two-qubit Pauli operators and their corresponding two-qubit Pauli substitute operators. Explicit circuit diagrams for the unitaries that diagonalize the associated 4×4 transformed operators Q are provided in Appendix H.

ID	P	Q	Spectra of Q
1	II	$0.5II + 0.5XX + 0.5YY + 0.5ZZ$	$\{1, 1, 1, -1\}$
2	XX	$0.5II + 0.5XX - 0.5YY - 0.5ZZ$	$\{1, 1, -1, 1\}$
3	YY	$-0.5II + 0.5XX - 0.5YY + 0.5ZZ$	$\{1, -1, -1, -1\}$
4	ZZ	$0.5II - 0.5XX - 0.5YY + 0.5ZZ$	$\{1, -1, 1, 1\}$
5	IX	$0.5IX + 0.5XI + 0.5iYZ - 0.5iZY$	$\{-1, i, -i, 1\}$
6	XI	$0.5IX + 0.5XI - 0.5iYZ + 0.5iZY$	$\{-1, i, -i, 1\}$
7	YZ	$-0.5iIX + 0.5iXI + 0.5YZ + 0.5ZY$	$\{-1, i, -i, 1\}$
8	ZY	$-0.5iIX + 0.5iXI - 0.5YZ - 0.5ZY$	$\{-1, i, -i, 1\}$
9	IY	$-0.5iY + 0.5iXZ - 0.5YI - 0.5iZX$	$\{-1, i, -i, 1\}$
10	YI	$0.5iY + 0.5iXZ + 0.5YI - 0.5iZX$	$\{-1, i, -i, 1\}$
11	XZ	$0.5iY + 0.5XZ - 0.5iYI + 0.5ZX$	$\{-1, i, -i, 1\}$
12	ZX	$-0.5iY + 0.5XZ + 0.5iYI + 0.5ZX$	$\{-1, i, -i, 1\}$
13	IZ	$0.5IZ + 0.5iXY - 0.5iYX + 0.5ZI$	$\{i, -i, 1, -1\}$
14	ZI	$0.5IZ - 0.5iXY + 0.5iYX + 0.5ZI$	$\{i, -i, 1, -1\}$
15	XY	$0.5iIZ - 0.5XY - 0.5YX - 0.5ZI$	$\{1, -1, -i, i\}$
16	YX	$0.5iIZ + 0.5XY + 0.5YX - 0.5ZI$	$\{1, -1, -i, i\}$

stead adopt the idea of the operator averaging method [59]. For each $\text{Tr}(Q_\alpha \rho_i \otimes \rho_j)$ in $\text{Tr}(H_B \rho_i \otimes \rho_j)$, we can simply do a basis rotation circuit to the diagonal basis of Q_α that requires separate two-qubit gates to rotate each basic two-qubit substitute operators to its diagonal basis and then do repeated measurements to estimate $\text{Tr}(Q_\alpha \rho_i \otimes \rho_j)$. A potential limitation of this approach is the lack of the grouping methods as in traditional Pauli measurements [27]. For $\text{Tr}(\rho_i \rho_j)$, its value equals to $\text{Tr}(S \rho_i \otimes \rho_j)$, with S the tensor product of n two-qubit swap gates that can also be estimated in the diagonal basis of S (sometimes called the destructive SWAP test [60]). After the estimation of values of $\text{Tr}(Q_\alpha \rho_i \otimes \rho_j)$ and $\text{Tr}(S \rho_i \otimes \rho_j)$ by repeated measurements for different $\{Q_\alpha, \rho_i, \rho_j\}$, we can then do classical processing based on Eq. (3) to get an estimation of $\langle \psi | H_A | \psi \rangle$. We give concrete constructions of these rotation circuits in Appendix H.

Being able to evaluate the value of $\langle \psi | H_A | \psi \rangle$ from $\{\rho_1, \rho_2, \dots, \rho_K, c_1, c_2, \dots, c_K\}$ by Eqs. (2)–(6) forms the core of our proposal, which shows that we can “prepare” unconditionally decoherence-free pure quantum states from noisy quantum circuits (Fig. 1) and thus can greatly mitigate the influence of noise. We should emphasize that the philosophy of our error mitigation protocol is not to extract noiseless information from noisy quantum states as in other QEM protocols but to directly encode noiseless quantum states into noisy ones. From the operational level, our protocol requires no knowledge of the noise model, no knowledge of the symmetry structures of problems, no ability to tune the noise strength, and no ancilla qubits for complicated controlled unitaries. Thus, our QEM protocol might be a promising QEM protocol for NISQ quantum applications.

B. Entanglement in terms of DMV

In the mapping [Eq. (1)], an n -qubit density matrix is mapped to a $2n$ -qubit pure state $|\rho\rangle = \frac{1}{\sqrt{C_\rho}} \sum_{ij} \rho_{ij} |i\rangle |j\rangle$, where we will call the n -qubit system denoted by the row index i the row system and the n -qubit system denoted by the column index j the column system. When ρ is a pure state of the form $\rho = |S\rangle\langle S|$, we have $|\rho\rangle = |S\rangle \otimes |S\rangle$ as a product state between the row system and the column system. When ρ is a mixed state of the form $\rho = \sum_i p_i |s_i\rangle\langle s_i|$, the state $|\rho\rangle = \frac{1}{\sqrt{\text{Tr}(\rho^2)}} \sum_i p_i |s_i\rangle \otimes |s_i\rangle$ becomes an entangled state. Note that for the density matrix ρ , we can also define another commonly used pure state for it by purification [61] defined as $|\psi_\rho\rangle = \sum_i \sqrt{p_i} |s_i\rangle$. We proved that the Rényi entropies [62] between these two states have the relation

$$H_\alpha(|\rho\rangle) \leq H_\alpha(|\psi_\rho\rangle). \quad (7)$$

A complete proof of Eq. (7) is provided in Appendix C. Since for $\alpha = 2$, we have $H_\alpha(|\psi_\rho\rangle) = -\ln(\text{Tr}(\rho^2))$, the achievable entanglement of $|\rho\rangle$ is bound by the purity of ρ . When ρ is the maximally mixed state, the state $|\rho\rangle$ becomes the maximally entangled state.

In the mapping [Eq. (2)], the entanglement of the state $|\psi\rangle$ is not only influenced by the purity of density matrices but also the number of linear combinations K . This can be understood as the linear combinations can enlarge the number of nonzero Schmidt coefficients and thus increase the

entanglement. Indeed, we find that when $\{\rho_1, \rho_2, \dots, \rho_K\}$ are prepared randomly (roughly a unitary 1-design [63]), the second-order Rényi entropy can be enlarged by an amount $\ln(K)$ at most.

As we will show later, the sampling complexity of estimating Eq. (3) depends on the purity of density matrices. We require a high average of purity to make the sampling cost tolerable. Since the purity also decides the level of entanglement of pure states from DMV, this high purity restriction can thus limit the expressibility of representing highly entangled states between the row and the column system. To help solve this problem, an observation is that the purity has no restrictions on the entanglement inside the row and the column system, and for most quantum states of physical interest [64], the entanglement will not be saturated between any bipartition of the whole system. Thus, given a $2n$ -qubit pure state, it is possible to find a bipartition with low entanglement and assign the two subsystems under this bipartition as the row and the column systems, respectively, to efficiently represent it. We will show this strategy in the following numerical experiments.

III. VQAS BASED ON DMV

A. Framework

A direct application for QEM is VQA, which is especially suitable for our protocol where error-mitigated pure states do not correspond to the states that circuits aim to prepare when there is no noise. Combined with the classical linear combinations, it is not a trivial task to deterministically prepare a quantum state under our mapping [Eq. (2)] since there may exist many choices of density matrices and coefficients that correspond to the same state. However, for VQAs, this weakness can be well resolved since a classical optimizer will guide us on how to adjust quantum circuits.

Now, given a Hamiltonian H_A , we give a framework to run VQE for its ground states using our QEM protocol. Note that this framework can also be applied to other tasks such as variational quantum simulations [31] and variational quantum machine learning [30]. For each density matrix ρ_i in $\{\rho_1, \rho_2, \dots, \rho_K\}$, we assign a parametrized quantum circuit to prepare it with parameters $\vec{\theta}_i$. If we use an n -qubit noisy circuit to prepare an n -qubit density matrix, then the purity of the density matrix can only be adjusted by uncontrollable noise that is not universal. Thus, instead, we use $2n$ -qubit circuits where we call the first n qubits the upper systems and the last n qubits the lower systems to prepare these density matrices on the upper systems by ignoring (tracing out) the lower systems. For the odd-dimensional case, one can naturally embed the system into an even-dimensional Hilbert space by introducing an empty ancilla qubit. Specifically, if the original system state is ρ , we define its extension as $\rho' = \rho \otimes |0\rangle\langle 0|$ so that ρ' resides in an even-qubit Hilbert space where the DMV protocol applies directly. At the Hamiltonian level, if $H = \sum_j c_j P_j$, $P_j \in \{I, X, Y, Z\}^{\otimes(2n-1)}$, we can extend it as $H' = \sum_j c_j (P_j \otimes I)$, where the ancillary qubit is decoupled from the system and does not affect its physical properties; therefore, the proposed method also applies to the odd-dimensional case. We will show later that $2n$ is an upper bound for circuits,

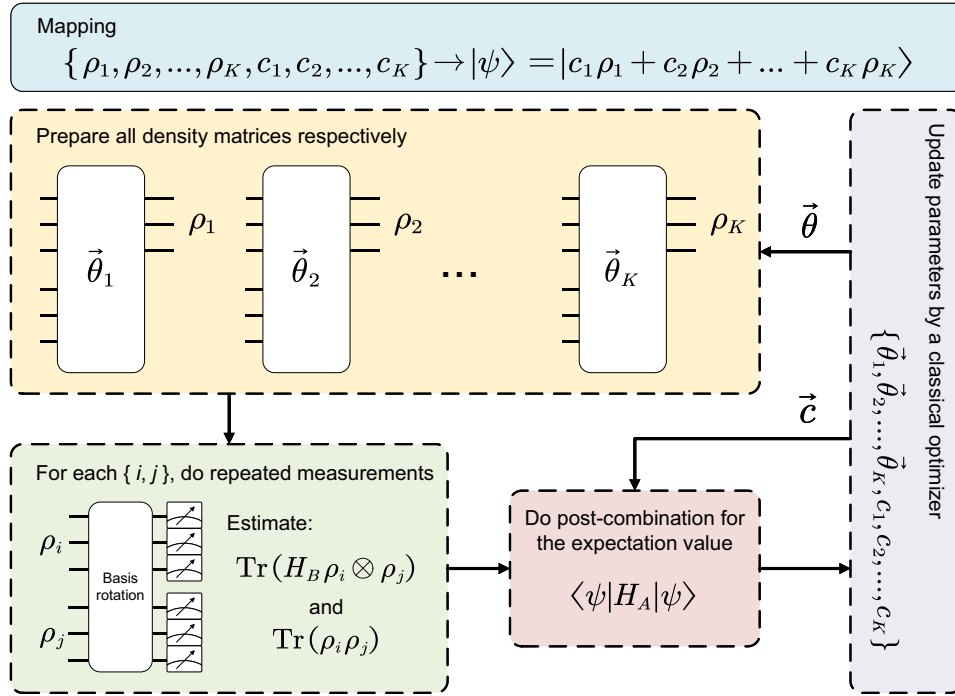


FIG. 2. Running VQAs with QEM by DMV. For each n -qubit density matrices, a $2n$ -qubit parametrized quantum circuit with parameters $\vec{\theta}_i$ is assigned. These density matrices $\{\rho_1, \rho_2, \dots, \rho_K\}$ combined with classical coefficients $\{c_1, c_2, \dots, c_K\}$ form a $2n$ -qubit pure state $|\psi\rangle$ according to the mapping [Eq. (2)]. The expectation value $\langle\psi|H_A|\psi\rangle$ is evaluated by direct measurements for $\text{Tr}(H_B \rho_i \otimes \rho_j)$ and $\text{Tr}(\rho_i \otimes \rho_j)$ with classical processing for linear combinations according to Eq. (3). The basis rotation circuits before measurements are composed of tensor products of two-qubit rotations. The values of $\langle\psi|H_A|\psi\rangle$ as the cost function are determined by coefficients $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K, c_1, c_2, \dots, c_K\}$ and are fed into a classical optimizer to update parameters for the next round.

and in many cases, we might use a circuit of a size $n + L$ with $L \leq n$. Note that the upper and lower systems should not be confused with the row and column systems. Concrete constructions of these circuits will be introduced in the next section.

These density matrices combined with coefficients $\{c_1, c_2, \dots, c_K\}$ are then used to estimate the expectation value $\langle\psi|H_A|\psi\rangle$ by the measurement and the classical combination procedures introduced before. The whole parameter set is defined as $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K, c_1, c_2, \dots, c_K\}$, which is fed into a classical optimizer for updating with $\langle\psi|H_A|\psi\rangle$ as the cost function. When the stopping criteria are met, the virtual pure state from optimized $\{\rho_1, \rho_2, \dots, \rho_K, c_1, c_2, \dots, c_K\}$ by Eq. (2) then corresponds to the approximation of the ground state of H_A . The whole framework is summarized in Fig. 2. Note that since each density matrix is prepared by an at most $2n$ -qubit quantum circuit and during the measurements we only pick two density matrices for collective measurements at a time, we have no need to build K $2n$ -qubit circuits but only need two $2n$ -qubit circuits at most that prepare different density matrices depending on different measurement tasks.

We want to mention that the idea of classical linear combinations of quantum states has been adopted in several proposals [65–67] for variational quantum algorithms to enhance the expressibility of NISQ devices without aggravating the barren plateau problems. One marked difference that sets our work apart from these previous works is that our protocol—also benefiting from classical linear combinations—is inherently robust against noise.

In the following, we will introduce several types of constructions of quantum circuit ansatz for preparing density matrices used for VQAs. We will first introduce an ansatz for general purposes that can universally prepare density matrices with minimal resources and then introduce how chemical-inspired ansatz specially designed for electronic structure problems of molecules can be well fitted into our framework. We will also give numerical examples based on these ansätze to demonstrate the performance of VQAs using our QEM protocol under various noises. Several other strategies for constructing ansatz are discussed at the end of this section.

B. Schmidt ansatz and numerical experiments

In our proposal, we can use a $2n$ -qubit quantum circuit U to prepare an n -qubit density matrix ρ . When there is no noise, the output state $|\phi\rangle$ from U can be expressed in a Schmidt form: $|\phi\rangle = \sum_i \lambda_i |u_i\rangle \otimes |l_i\rangle$, with λ_i the Schmidt coefficients and $|u_i\rangle \otimes |l_i\rangle$ a product state between the upper and the lower system satisfying $\langle u_i | u_j \rangle = \langle l_i | l_j \rangle = \delta_{ij}$. The density matrix ρ is then the reduced density matrix of $|\phi\rangle$ on the upper system with the form $\rho = \sum_i |\lambda_i|^2 |u_i\rangle \langle u_i|$. From the form of ρ , we can find that first, the phase information of Schmidt coefficients is lost, and second, the information of $|l_i\rangle$ is lost. Thus, we can build circuits U whose structure restricts the freedom of redundant information to save resources and reduce the number of useless parameters for VQAs.

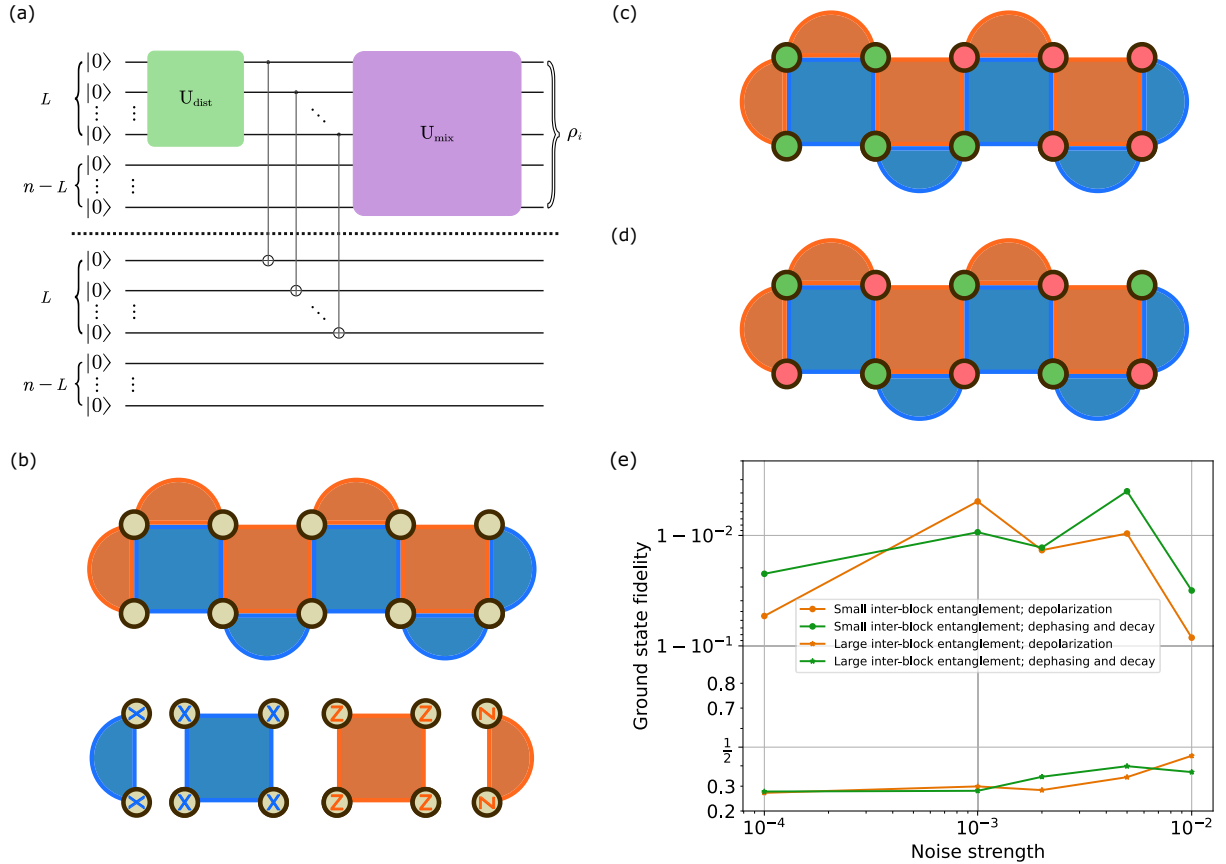


FIG. 3. Schmidt ansatz and numerical experiments. (a) Structure of the Schmidt ansatz with L intrablock CNOT gates. U_{dist} (on L qubits) serves the purpose of adjusting the Schmidt coefficients and can be chosen to be a real orthogonal matrix to reduce the number of single-qubit rotation gates as well as the number of parameters. U_{mix} (on n qubits) is a parametrized unitary circuit that under certain set of parameters is expected to transform all Schmidt components into desired form. (b) Ten-qubit rotated surface code with a unique ground state. (c) Partitioning of the qubits (into a green-qubit block and a red-qubit block) with small intrablock entanglement. (d) Partitioning of the qubits with large intrablock entanglement. (e) Obtained ground-state fidelity under the Schmidt ansatz with $L = 1$ intrablock CNOT gates for both partitions in panels (c) and (d) marked by dots and stars, respectively. The circuit has in total 115 gates that correspond to a circuit fault rate of 0.115 at 10^{-3} noise strength.

The ansatz to achieve this purpose is called the Schmidt ansatz, where we aim to prepare $|\phi\rangle$ with the form $|\phi\rangle = \sum_i \lambda_i |u_i\rangle \otimes |i\rangle$, with λ_i real numbers and $|i\rangle$ the computational basis. The structure of the Schmidt ansatz is shown in Fig. 3(a). First, we implement an L -qubit orthogonal circuit (which is also unitary) on the first L qubits in the upper system to prepare a superposition of computational basis: $\sum_i \lambda_i |i\rangle$ with λ_i real numbers. Next, we do L CNOT gates connecting the first L qubits in the upper system and the first L qubits in the lower system to generate the whole $2n$ -qubit state: $\sum_i \lambda_i |i\rangle |0\rangle^{\otimes(n-L)} \otimes |i\rangle |0\rangle^{\otimes(n-L)}$. The reduced density matrix of this state is of the form $\sum_i |\lambda_i|^2 |i\rangle |0\rangle^{\otimes(n-L)} \langle i| \langle 0|^{\otimes(n-L)}$. This density matrix is then further fed into an n -qubit unitary circuit to adjust states attached to each Schmidt coefficient to get the final density matrix ρ . Thus, we can understand the Schmidt circuit by using the L -qubit orthogonal circuit to adjust $|\lambda_i|^2$ followed by using the n -qubit unitary circuit to adjust $|u_i\rangle$ without giving useless freedom to phases of Schmidt coefficients and $|i\rangle$.

When constructing the Schmidt ansatz in real experiments, we suggest a hardware structure of the shape of a ladder (shown in the Appendix I) such that the CNOT gates connecting

the upper and the lower systems can be built on local physical qubits.

The reason that we set a number L for Schmidt coefficients generations is to restrict the purity of ρ with a lower bound 2^{-L} , which we will show has a direct influence on the sampling complexity. This also means that the size of the Schmidt circuit is $n + L$ smaller than $2n$. When there is no noise, to restrict the sampling complexity into a quasipolynomial scaling with n , we require L to be of order $\mathcal{O}(\text{poly}(\ln(n)))$. As shown before, this will reduce the reachable entanglements between the row system and the column system. However, since the purity does not influence the entanglement inside the row system and the column system, it is possible to make the row and the column systems match a bipartition under which the ground state of H_A has low entanglement between the two systems and meets the requirement on the purity for the sampling cost. We want to mention here that, due to the low entanglement between the row system and the column system, one may expect to use circuit cutting methods [68,69] to achieve the same thing. However, those methods cannot be free of decoherence as in our method.

As an example, we use a ten-qubit rotated surface code with a single ground state [Fig. 3(b)]. Under the partition (into a block of green qubits and another of red qubits) in Fig. 3(c), the dimension of the Schmidt decomposition of the ground state is 2. Thus, in this case, we only need one intrablock CNOT gate in Fig. 3(a). On the other hand, the dimension of the Schmidt decomposition of the ground state is 2^4 under the partition in Fig. 3(d). In our numerical experiment, we use the Schmidt ansatz with $L = 1$ for both partitions [Fig. 3(e)]. We observe that for the partition in Fig. 3(c) with small intrablock entanglement, the Schmidt ansatz performs well for the full range of noise strength from 10^{-4} to 10^{-2} . However, for the partition in Fig. 3(d) with large intrablock entanglement, the performance of the Schmidt ansatz (with $L = 1$) is poor for all noise strengths. Thus, generally speaking, by choosing a partition with small intrablock entanglement, we can use the Schmidt ansatz with a small enough L to potentially achieve simultaneously good performance (in terms of final state fidelity) as well as low enough sample complexity. However, we want to mention that finding a bipartition with small entanglement can be classically difficult in general since it is related to the so-called state hidden subgroup problem, which, interestingly, has been shown to have efficient quantum algorithm to solve [70].

C. Chemical-inspired ansatz and numerical experiments

When the Schmidt ansatz is randomly chosen, we may inevitably face the barren plateau problems [71]. Thus, a major goal in VQAs is to develop ansätze inspired by target problems that have high biases from random quantum circuits (that form a unitary 2-design) [71]. We observe that for certain problems, our method admits physically motivated ansätze, which can be expected to achieve this goal. For instance, the electronic ground state of a paramagnetic molecule [72], in addition to particle number conservation, admits the total spin symmetry; that is, the ground state is composed of the (fermionic) occupation number states with an equal number of spin-up electrons and spin-down electrons. Consider a paramagnetic molecule with $2k$ electrons and $2n$ spin orbits (the first n orbits for spin-up electrons and the second n orbits for spin-down electrons); we apply the Jordan-Wigner transformation [73] to transform all creation and annihilation operators into Pauli-string operators such that the $|1\rangle$ state ($|0\rangle$ state) on the i th qubit in the first half block of all qubits indicates the presence (absence) of an electron on the i th spin-up orbital and similarly for the second half block, which is reserved for spin-down orbitals.

For traditional VQE, we can use a spin-symmetric version of the q-Unitary Coupled-Cluster Single and Double (UCCSD) ansatz [73]:

$$U(\theta) := U_{\uparrow}(\vec{\theta}_{\uparrow})U_{\downarrow}(\vec{\theta}_{\downarrow})U_{\uparrow\downarrow}(\vec{\theta}_{\uparrow\downarrow})U_{\uparrow\uparrow}(\vec{\theta}_{\uparrow\uparrow})U_{\downarrow\downarrow}(\vec{\theta}_{\downarrow\downarrow}), \quad (8)$$

where each term preserves the number of spin-up electrons and the number of spin-down electrons [see Fig. 4(a) for the structure of the ansatz]. Then, for a noiseless circuit, we can guarantee the output state to respect the spin symmetry by

setting the input state to be

$$| \underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k}, \underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k} \rangle, \quad (9)$$

with exactly k electrons in each of the upper and lower systems. Interestingly, the same ansatz and the input state can also guarantee the resulting states in our protocol also satisfy the spin symmetry. We numerically test this ansatz for both traditional VQE and our protocol on the H_4 molecule under circuit-level depolarization noise [Fig. 4(b)]. While the total number of qubits in this numerical experiment is 8, the ansatz circuit is deep and has 5528 gates (and thus 5528 faulty locations) in total. For noise probability ranging from 10^{-5} to 10^{-3} , our protocol manages to achieve close to 1 ground-state fidelity and is about 3 orders of magnitude better than traditional VQE in terms of obtained ground-state fidelity [Fig. 4(b)]. At sufficiently high noise strengths, the DMV-based VQE can underperform standard VQE. The reason is a concentration phenomenon of noisy circuit outputs: for depolarizing (unital) noise, the DMV-encoded outputs concentrate toward the maximally entangled state $2^{-n/2} \sum_i |i\rangle|i\rangle$, which reduces expressibility and induces noise-driven barren plateaus; in contrast, in standard VQE the outputs concentrate toward the maximally mixed state, which maintains nonzero overlap with arbitrary targets. Consequently, when the noise is large enough, DMV may lose its advantage, as reflected in Fig. 4(b). See Sec. IV D for a detailed discussion and references.

Under a more experimentally relevant depolarization noise $p = 0.5 \times 10^{-3}$, our protocol preserves its significant advantage over traditional VQE for all bond lengths in Fig. 4(c). Furthermore, we notice that the purity of the density matrices ρ_i for $i \in \{1, 2, 3, 4\}$ throughout an optimization process generally decreases [Fig 4(d)] by nearly an order of magnitude as the noise increases from the noiseless regime to the highly noisy regime ($\sim 10^{-2}$ noise strength). This can be understood as the upper system getting entangled with the environment. An interesting thing about the purity, in this case, is that in the noiseless regime, the average is much higher than the lower bound set by the number of CNOT gates connecting the lower and the upper systems; thus, a case-by-case analysis is required for the real performance of our protocol.

It is worth noticing that the structure of the ansatz [Eq. (8)] can be further simplified and adjusted for our protocol. The reason we keep it in its original form is to make a fair comparison between traditional VQE and ours. For example, the block $U_{\downarrow}(\theta_{\downarrow})$ can be ignored since it will not change the output density matrices. At the same time, after compiling Eq. (8) into digital quantum circuits, the CNOT gates connecting the upper and the lower systems can be flexibly adjusted to control the sampling complexity. Furthermore, we can use a locking strategy as shown in Ref. [55] to pair up the parameters controlling the upper and the lower blocks (e.g., locking $\vec{\theta}_{\uparrow}$ and $\vec{\theta}_{\downarrow}$) and thus reduce nearly half of the number of parameters.

D. Other ansatz strategy

There are also other strategies for constructing suitable ansatz for our proposal. One is to follow the

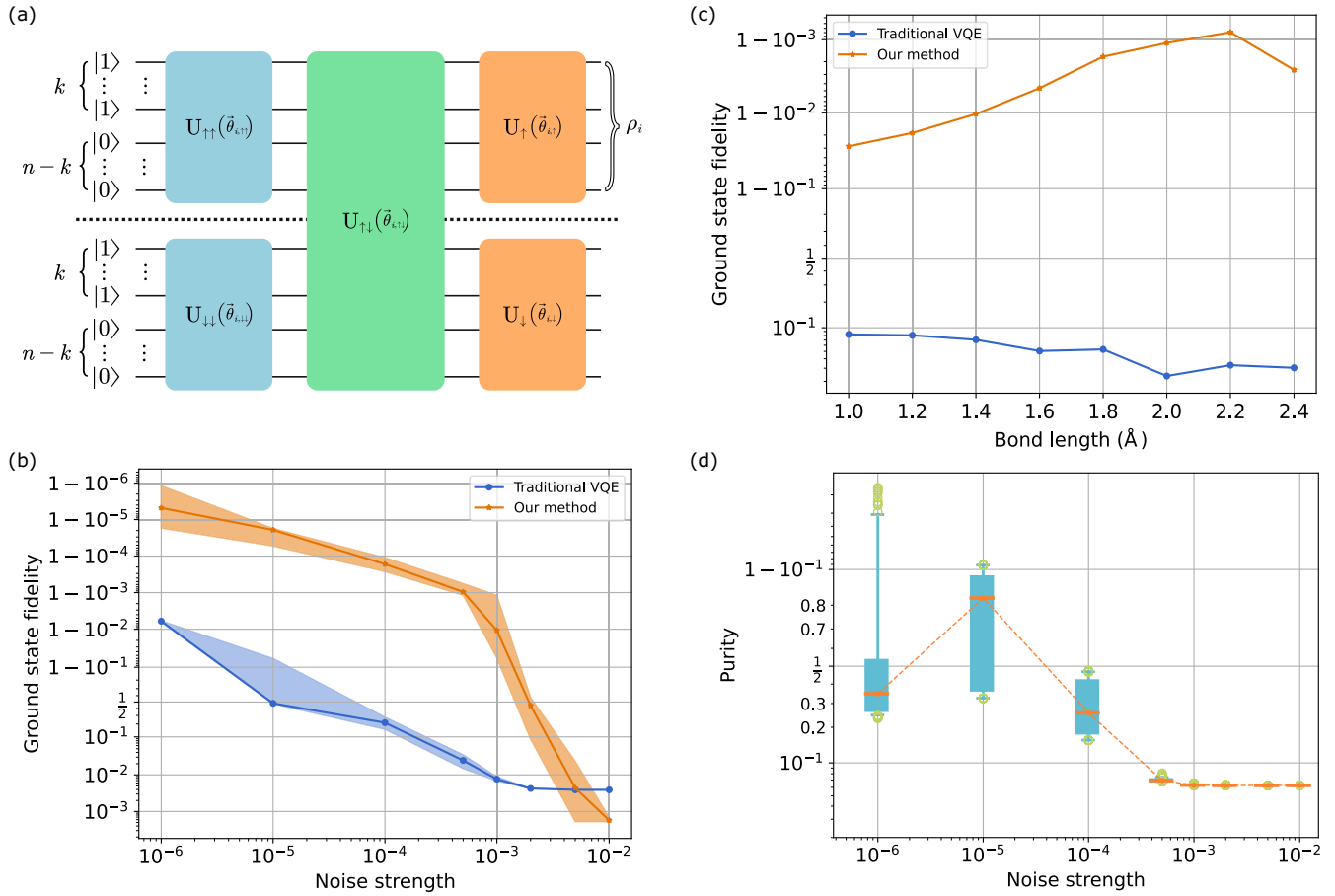


FIG. 4. Solving for electronic ground states of the H_4 molecule under a spin-symmetric ansatz. (a) Spin-symmetric ansatz for both traditional VQE and our method. For traditional VQE, the output state is taken over all $2n$ qubits. For our protocol, the output state is the reduced density matrix of the first block of n qubits. The circuit used for our numerical experiment has 5528 gates, corresponding to a circuit fault rate of 2.764 at 0.5×10^{-3} noise strength. (b) Obtained ground-state fidelity at bond length $2.0 \text{ \AA}$ under circuit-level depolarization noise for various noise strengths. The unit of energy is Hartree. (c) Obtained ground-state fidelity for various bond lengths under noise strength $p = 0.5 \times 10^{-3}$. (d) Box plots of the purity values of ρ_i for all $i \in \{1, 2, 3, 4\}$ throughout an optimization process under different noise strengths. At a specific noise strength, the corresponding orange line and the orange dot mark the median value of all data points, the blue box marks the range from the first quartile to the third quartile and the blue whisker (error bar) captures the range from the 1st percentile to the 99th percentile, and all data points outside of the whisker are marked by green circles.

Hermitian-preserving ansatz introduced in Ref. [55], which aims to prepare a pure state that is Hermitian when turning it back to a matrix. This is done by pairing up the parameters in the upper and the lower systems, which can also help to reduce the freedom of redundant information. However, for this type of ansatz, it is not easy to control the purity of density matrices.

The other is to utilize the idea of mixing [74] where the mixedness of density matrices is generated by tuning parameters in quantum circuits based on a classical probabilistic distribution. Note that this can be understood as using a large number of linear combinations to take the function of CNOT gates connecting the upper and the lower systems. In this way, it is possible to use an n -qubit circuit to generate density matrices with programmable mixedness. The problem with this approach is perhaps a trade-off between the number of parameters of optimizations and the expressibility.

IV. PROPERTIES

A. Sampling complexity

The sampling complexity of estimating Eq. (3) mainly depends on the number K of classical combinations and the purity of $\{\rho_1, \rho_2, \dots, \rho_K\}$. When these density matrices are prepared by the ansatz introduced above, the purities are decided by the number of CNOT gates L connecting the upper and the lower systems (in the Schmidt ansatz and the chemical ansatz) and the circuit fault rate ζ . The number of CNOT gates will set a lower bound 2^{-L} for the purity in the noiseless case. The circuit fault rate ζ is defined as the sum of individual fault probabilities at all possible locations in the ansatz circuit $\zeta = \sum_i p_f$. When $\zeta \lesssim 1$ such that a Poisson distribution can be used to model the probability distribution of the number of faults that occur in the circuit, the fault-free probability is around $e^{-\zeta}$, which will set another lower bound $e^{-2\zeta}$ for the purity.

The true lower bound is thus a competition between 2^{-L} and $e^{-\zeta}$, which results in the following sampling complexity:

$$N \geq \max[e^{4\zeta}, 2^{2L}] \frac{3K^2}{\epsilon^2} \left(\frac{m\|H_A\|_F^2}{2^{2n}} + \|H_A\|_2^2 \right), \quad (10)$$

with $\frac{\|H_A\|_F^2}{2^{2n}} = \sum_{\alpha=1}^m g_{\alpha}^2$. To get Eq. (10), we require the final n -qubit unitaries in the upper system in the Schmidt ansatz and the unitaries in the upper system between two CNOT gates connecting the upper and the lower systems in the chemical ansatz form at least a unitary 1-design [63] such that the values of $\text{Tr}(\rho_i \rho_j)$ for $i \neq j$ will concentrate on 2^{-n} , which can be ignored when n is large. This also means that for large n cases, we have no need to measure values of $\text{Tr}(\rho_i \rho_j)$ for estimating $\langle \psi | H_A | \psi \rangle$ to save the sampling cost. Since random Pauli circuits can already form a 1-design [75], this requirement can be easily satisfied. Equation (10) tells the number of samples needed to estimate $\langle \psi | H_A | \psi \rangle$ within an accuracy ϵ . The polynomial scaling to K indicates that a large number of classical linear combinations is allowed. Note that since the purity has a lower bound 2^{-n} , the sampling complexity is also upper bounded by

$$N \leq 2^{2n} \frac{3K^2}{\epsilon^2} \left(\frac{m\|H_A\|_F^2}{2^{2n}} + \|H_A\|_2^2 \right). \quad (11)$$

As mentioned many times, the basic idea of our QEM protocol is fundamentally different from others. However, in terms of the sampling complexity, the scaling with respect to the circuit fault rate ζ of our protocol is $e^{4\zeta}$, which interestingly coincides with the general exponential scaling behaviors of the sampling overhead in other QME protocols [48]. Indeed, our protocol solves the decoherence by directly changing the way of encoding quantum states; however, to retrieve information under this encoding, an $e^{4\zeta}$ scaling has to be respected because of the purity restriction. Compared with other protocols, our sampling complexity shares the same scaling of ζ as using the probabilistic error cancellation method [40,41] to totally eliminate errors and as using the two-copy version of the virtual state distillation method [45]. When the task is to obtain information on pure states, our method is superior to the virtual state distillation since it needs an M -copy version with M large to capture the pure state behaviors well, which, however, will result in an $e^{2M\zeta}$ scaling on the sampling complexity. Still, we want to emphasize that our advantage is actually not a free lunch since currently it is only compatible with variational approaches while the previous QEM methods are capable of general quantum computation tasks.

B. Expressibility

As discussed earlier, when $K = 1$, states $|\rho\rangle$ from n -qubit density matrices are not able to represent arbitrary $2n$ -qubit pure states due to the Hermiticity and the positive semidefiniteness of density matrices. Thus, to enhance the expressive power, we introduce the classical linear combinations and we have shown that it is sufficient to set $K = 4$ with two real and two imaginary coefficients to express any pure $2n$ -qubit states $|\psi\rangle$. As discussed in Sec. II B, the purity of density matrices and classical linear combinations can both influence

the entanglement between the row and the column systems and thus influence the expressibility. The classical linear combinations can be very useful in saving resources to “prepare” some nontrivial states, for example, the preparation of the generalized $2n$ -qubit Greenberger-Horne-Zeilinger state $\frac{1}{\sqrt{2}}(|0\rangle^{\otimes 2n} + |1\rangle^{\otimes 2n})$. While its entanglement entropy under any bipartition is very low, its preparation requires a $\mathcal{O}(n)$ -depth local unitary circuit [76], which, however, can be seen as a classical linear combination of the two easiest product states $|0\rangle^{\otimes 2n}$ and $|1\rangle^{\otimes 2n}$.

To quantitatively evaluate the expressibility of ansatz we introduced above under DMV encoding, we adopt the tools of the covering number from the statistical learning theory [77], which have been used for measuring the expressibility of ansatz in standard VQAs [78]. When we do not use DMV, the upper bound for the covering number of standard VQAs is given by

$$\mathcal{N}(\mathcal{H}, \epsilon, |\cdot|) \leq \left(\frac{7N_{\text{gt}}\|H_A\|}{\epsilon} \right)^{d^{2k}N_{\text{gt}}}, \quad (12)$$

where $\mathcal{H}$ is the hypothesis space for standard VQAs $\mathcal{H} = \{\text{Tr}(\hat{U}(\theta)^\dagger H \hat{U}(\theta) \rho) \mid \theta \in \Theta\}$, N_{gt} represents the number of trainable gates in $U(\theta)$, d is the dimension of the system, which is 2^{2n} for our cases, ϵ is a very small positive hyperparameter, and each gate $\hat{u}_i(\theta)$ acts on at most k qudits.

For VQAs based on DMV as formulated before, the upper bound for the covering number is

$$\mathcal{N}(\tilde{\mathcal{H}}, \epsilon, |\cdot|) \leq 2^L \mathcal{C} \left(\frac{7N_{\text{GT}}\|H_B\|}{\epsilon} \right)^{d^{2k}N_{\text{GT}}}, \quad (13)$$

where $\tilde{\mathcal{H}}$ is the hypothesis space for VQAs by DMV, L is the number of CNOT gates connecting the upper and lower halves, $\mathcal{C}$ is a constant greater than 1 defined in the Appendix, N_{GT} is the sum of trainable and also meaningful gates in each linear combination circuit, i.e., $N_{\text{GT}} = \sum_{i,j=1}^K N_{\text{gt}}(ij)$, and the other parameters are the same as in the standard case (see Appendix F).

Taking the same observable operator, since we have $\|H_B\| = \|H_A\|$. For analytical convenience, we assume that the number of trainable gates in each combination circuit is the same as in a single standard VQA circuit, i.e., $N_{\text{GT}} = KN_{\text{gt}}$. Since this term appears in the exponential part, the upper bound for the covering number increases exponentially with the number of combinations. Additionally, the constant term $2^L \mathcal{C}$ is greater than 1, further amplifying the upper bound for the covering number. It can be seen that VQAs based on DMV have higher expressibility compared to standard VQAs.

C. Trainability

Here, we give a discussion on the barren plateau issues in our proposal. The parameters used in our proposal for training VQAs are composed of $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K, c_1, c_2, \dots, c_K\}$. The trainability for classical parameters $\{c_1, c_2, \dots, c_K\}$ and for quantum parameters $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K\}$ needs to be discussed separately.

For classical parameters $\{c_1, c_2, \dots, c_K\}$, since they appear as the classical coefficients for combining expectation values as shown in Eq. (3), they will not directly deal with the

exponentially large Hilbert space and thus have no barren plateau issues, as discussed in Refs. [66,67]. Thus, the classical linear combinations can not only enhance the expressibility but also avoid worsening the barren plateau problems, which in some sense can overcome the fundamental trade-off between the expressibility and the trainability [79] in variational quantum algorithms.

For quantum parameters $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K\}$, we know that in standard VQAs, when the parametrized quantum circuits are randomly enough to form a unitary 2-design, the derivatives of cost functions with respect to parameters will concentrate on exponentially small averages [71]. In our case, while the form of the cost function [Eq. (3)] looks very different from those in standard VQAs, we should expect the existence of barren plateaus in highly random Schmidt ansatz. Since we may require $L \leq n$, the barren plateau problems should be understood from the randomness of the n -qubit circuit in the upper systems but not the whole $2n$ -qubit systems. Indeed, if an n -qubit parametrized circuit forms a 2-design, the average variance η^2 of derivatives will be of order $\mathcal{O}(2^{-n})$ in standard VQAs, which will be inherited in our algorithmic framework as

$$\left| \frac{\partial \langle \psi | H_A | \psi \rangle}{\partial \theta} \right| \leq 2^{L+2} K^2 \sum_{\alpha} |g_{\alpha}| \eta \quad (14)$$

when, for example, the n -qubit unitary circuit in the Schmidt ansatz forms a 2-design. Equation (14) shows that when $L \ll n$, barren plateaus occur. On the other hand, when L is comparable with n , Eq. (14) should be ignored as it is a rather loose bound. Instead, we can then analyze barren plateau problems by checking whether the whole $2n$ -qubit circuit forms a 2-design.

The existence of barren plateau problems in randomized ansatz is the main motivation to let us give a construction of a chemical-inspired ansatz in the previous section. Since problem-inspired ansatz can have large derivations from unitary 2-design and thus, can potentially avoid barren plateaus with enough expressibility for the target problems at the same time [79].

D. Noise influence under DMV

The numerical results shown in the last section give us evidence of the effectiveness of our proposal against various types of noise. However, as we can find in Figs. 4(b) and 4(c), as the noise strength increases, our proposal inevitably loses its abilities against noise. This can be understood as the concentration effect of the output states of noisy quantum circuits [32,33,36]. For unital depolarizing noise models, the output states will concentrate on the maximally mixed state [49–51,80]. For nonunital models, under the assumption that the noiseless quantum circuit forms a unitary 2-design, the output states will also concentrate on some fixed states of noisy quantum circuits [52].

Under the DMV encoding, while all the output states are pure states without decoherence, the concentration effect will nevertheless limit the expressibility and trainability of our proposal for VQAs. For example, under the depolarizing noise and DMV, the output states will concentrate on the maximally entangled state $\frac{1}{\sqrt{2^n}} \sum_i |i\rangle |i\rangle$, which thus reduces the

expressibility. This also explains the reason that our proposal shows weaker performance compared with traditional VQE for very large noise, as shown in Figs. 4(b) and 4(c), since, unlike the maximally entangled state, the maximally mixed state can always have nonzero overlaps with arbitrary states. The concentration effect also means that the value of the cost functions will concentrate on the value corresponding to the maximally entangled state such that the derivatives with respect to the parameters $\{\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_K, c_1, c_2, \dots, c_K\}$ will be exponentially small in general and the noise-induced barren plateaus occur [36]. Another thing we want to mention is that while by DMV we can remove the noise-induced decoherence in the state preparation phase, the noise in the measurement phase may inevitably influence the accuracy for estimating the cost functions; thus, we recommend to combine with the measurement error mitigation methods [48,81].

We want to emphasize here that since under our encoding there is no decoherence, the expressibility and trainability we are discussing are restricted to pure states; thus, the inaccuracy induced by the gap between pure states and mixed states can be completely removed.

E. Combinations with other QEM protocols

Since the way of mitigating noise in our proposal is fundamentally different from other QEM protocols, instead of making comparisons between ours and others, it is more attractive to see whether the combinations between them can give further improvements. The answer is strongly positive. Note that all good performance of our proposal in numerical results we have shown before comes directly from changing the way of encoding information [Eq. (2)] that has not tried to fight against the concentration effect. In contrast, in other protocols [48], the basic idea is to extract desired information from concentrated output states. Thus, these are two different philosophies to mitigate noise and the combinations of them should be expected to have a better improvement compared with combinations within existing protocols that share the same philosophy. Also, as we talked about before, measurement error mitigation methods can also be introduced to help fight against measurement errors that DMV cannot mitigate.

The most interesting point is that the combinations of our protocol with others are free lunches and will not introduce additional sampling complexities. Recalling Eq. (10), when there is no noise in the quantum circuit, the sampling complexity is only determined by the number of CNOT gates connecting the upper and the lower systems. Suppose an output density matrix in the noiseless and noisy cases, respectively, if we directly use our QEM estimation, while the value of the cost function still corresponds to some pure state, it will have the tendency to concentrate on the value corresponding to the maximally entangled state and the noise in the circuit will inevitably increase the sampling complexity, as shown in Eq. (10). Instead, we can first use some other QEM protocols to help extract the noiseless expectations shown in the numerator and denominator of the expression of the cost function in Eq. (3). This has two benefits. First, this will help fight against the concentration effect and thus enlarge the expressibility and the trainability. Second, while in this phase we require an exponential number of samples required in other

QEM protocols [48], in the next phase of using Eq. (3) for the cost function estimation the expectation values in Eq. (3) are already from noiseless density matrices with bounded purity 2^{-L} and thus the exponential dependence on the circuit fault rate η of the sampling complexity of DMV encoding in Eq. (10) disappears. Thus, the combinations between our protocols with others will not only improve the overall performance but also avoid the amplification of the sampling complexity.

V. SUMMARY AND OUTLOOK

In summary, we propose a QEM protocol with a philosophy different from existing ones. By using the idea of DMV, we can eliminate the decoherence of quantum states in NISQ devices from the fundamental encoding level. We give a universal and NISQ-friendly measurement technique for measuring the expectation values of operators with respect to the linear combinations of vectorized density matrices. Based on this technique, we can now run virtual noiseless quantum experiments on noisy quantum circuits. In the example of using our protocol for VQAs, we show different ways of constructions of ansatz to extract density matrices and give various numerical results to show the effectiveness of our protocol. We give a detailed analysis of various properties of our proposal, ranging from the sampling complexity, expressibility, and trainability to the noise influence in terms of DMV and its combinations with other QEM protocols.

There are several future directions of our work. One is to find more efficient measurement strategies such as the classical shadows and grouping methods to reduce the sampling complexity. Another direction is to find applications of our QEM protocols beyond the VQAs, which can also be generalized to find more possibilities of DMV beyond this work and Ref. [53]. Also, as discussed in the Schmidt ansatz, it would be important to develop methods for finding bipartitions with small entanglement. One such possibility is to use quantum algorithms [70]. Finally, future potential experimental demonstrations in larger and more challenging problems would be demanded to assess its scalability and ability to contribute toward achieving some form of quantum advantages. We hope this work will open up possibilities for quantum applications.

ACKNOWLEDGMENTS

The authors would like to thank Chao-Yang Lu, Ming-Cheng Chen, Cheng-Cheng Yu, Yu-Xuan Yan, Zhen-Huan Liu, Zhen-Yu Cai, and Zhen-Yu Du for fruitful discussions and suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 124B1017).

DATA AVAILABILITY

We use the QISKIT package [82] for parts of our simulations. An example code can be found online [83].

APPENDIX A: PRELIMINARY: UNITARY DESIGN

The Haar measure is a uniform probability measure defined on the unitary group $U(d)$ satisfying

$$\int_{\text{Haar}} f(UV)dU = \int_{\text{Haar}} f(VU)dU = \int_{\text{Haar}} f(U)dU. \quad (\text{A1})$$

Here, V denotes an arbitrary fixed unitary on the Hilbert space under consideration. Given the Haar measure, the t th moment operator is defined as

$$\mathcal{M}_k(O) = \int_{\text{Haar}} U^{\otimes t} O U^{\dagger \otimes t} dU. \quad (\text{A2})$$

For arbitrary O , the moment operator $\mathcal{M}_k(O)$ lives in the space spanned by the permutation operators $\{V_d(\pi) | \pi \in S_t\}$ with S_t the t th order symmetric group and thus can be expressed as

$$\mathcal{M}_k(O) = \sum_{\pi \in S_t} c_\pi V_d(\pi), \quad (\text{A3})$$

where coefficients c_π can be obtained by solving the linear equations:

$$\text{Tr}(V_d(\kappa)^\dagger O) = \sum_{\pi \in S_t} c_\pi \text{Tr}(V_d(\kappa)^\dagger V_d(\pi)), \quad \text{for all } \kappa \in S_t. \quad (\text{A4})$$

For $t = 1$, the only permutation operator is I , and we have

$$\mathcal{M}_1(O) = \int_{\text{Haar}} U O U^\dagger dU = \frac{\text{Tr}(O)}{d} I. \quad (\text{A5})$$

For $t = 2$, there are two permutation operators: I and S (the swap operator), and we have

$$\begin{aligned} \mathcal{M}_2(O) &= \int_{\text{Haar}} U \otimes U O U^\dagger \otimes U^\dagger dU \\ &= \frac{\text{Tr}(O) - d^{-1} \text{Tr}(SO)}{d^2 - 1} I + \frac{\text{Tr}(SO) - d^{-1} \text{Tr}(O)}{d^2 - 1} S. \end{aligned} \quad (\text{A6})$$

Generating Haar random unitaries on a quantum computer generally needs an exponentially deep random quantum circuit that is unrealistic. However, if only behaviors of t th order moments are needed, we can instead generate unitaries from a distribution ν that forms unitary t -designs defined as

$$\int_{\nu} U^{\otimes t} O U^{\dagger \otimes t} dU = \int_{\text{Haar}} U^{\otimes t} O U^{\dagger \otimes t} dU \quad (\text{A7})$$

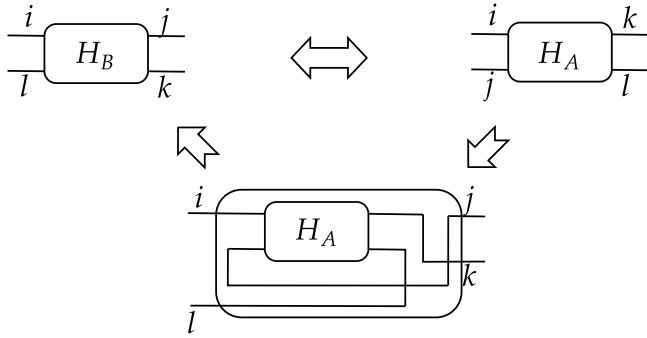
for arbitrary O . Specifically, if ν is a 1-design, Eq. (A5) is satisfied, and if ν is a 2-design, Eqs. (A5) and (A6) are satisfied.

Assuming U and V are drawn from a 1-design distribution ν , then, according to Eq. (A5), we have

$$\int_{\nu} \int_{\nu} \text{Tr}(U \rho_0 U^\dagger V \sigma_0 V^\dagger) dU dV = \frac{\text{Tr}(\rho_0) \text{Tr}(\sigma_0)}{d} = \frac{1}{d}. \quad (\text{A8})$$

Since $\text{Tr}(U \rho_0 U^\dagger V \sigma_0 V^\dagger) \geq 0$, we can use the Markov's inequality $\text{Prob}[X \geq \varepsilon] \leq \frac{E[X]}{\varepsilon}$ to show the concentration of around the average:

$$\text{Prob}[\text{Tr}(U \rho_0 U^\dagger V \sigma_0 V^\dagger) \geq \varepsilon] \leq \frac{1}{d\varepsilon}. \quad (\text{A9})$$

FIG. 5. The graphical representation of $\langle il|H_B|jk\rangle = \langle ij|H_A|kl\rangle$.

When d is large, Eqs. (A8) and (A9) indicate that $\text{Tr}(U\rho_0 U^\dagger V\sigma_0 V^\dagger)$ will concentrate on $\frac{1}{d}$ with high probability. This is a property that will frequently be used in the following sections.

APPENDIX B: SUBSTITUTE OPERATORS

1. Proof of the key equations by tensor networks

In our paper, the most crucial equation is

$$\langle \psi | H_A | \psi \rangle = \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{k,l=1}^K c_k^* c_l \text{Tr}(\rho_k \rho_l)}, \quad (\text{B1})$$

with

$$\langle il | H_B | jk \rangle = \langle ij | H_A | kl \rangle. \quad (\text{B2})$$

For clarity, in what follows we consider the simplified single-term case of the linear combination, i.e., $K = 1$. In this case, $|\psi\rangle \propto \text{vec}(\rho)$ and the normalization reduces to $C_\psi = C_\rho = \sqrt{\text{Tr}(\rho^2)}$.

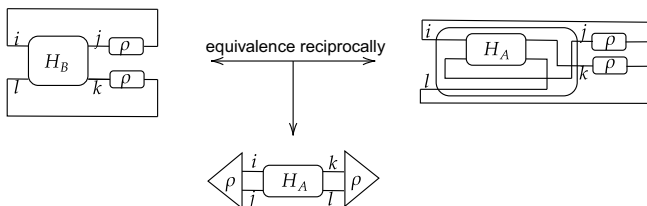
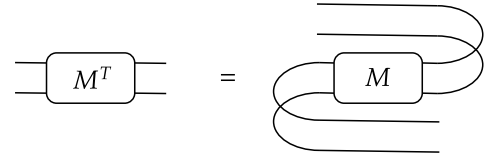
$$\langle \rho | H_A | \rho \rangle = \frac{\text{Tr}(H_B \rho \otimes \rho)}{\text{Tr}(\rho^2)}. \quad (\text{B3})$$

Due to Eq. (B2), we can represent the graphical equivalent of Fig. 5.

Using this relationship, we can easily prove $\langle \rho | H_A | \rho \rangle = \frac{\text{Tr}(H_B \rho \otimes \rho)}{\text{Tr}(\rho^2)}$, as shown in Fig. 6.

2. Proof of whether Q_α is a unitary operator by tensor networks

Next, we will use tensor networks to prove that Q_α is unitary. Here, let us first look at this relationship: Assuming M is a matrix, its conjugate transpose can be graphically represented as shown in Fig. 7.

FIG. 6. The graphical representation of $\langle \rho | H_A | \rho \rangle = \frac{\text{Tr}(H_B \rho \otimes \rho)}{\text{Tr}(\rho^2)}$.FIG. 7. The graphical representation of M^T .

For the case of the replacement operators for the two-qubit Pauli operators as given in the main text, based on the proof in Fig. 6, we can represent the replacement operators graphically as shown in Fig. 8.

Since the criterion for a unitary operator is whether $Q_\alpha Q_\alpha^\dagger = I$, according to Fig. 7, we can provide the graphical representation of $Q_\alpha^\dagger$ as shown in Fig. 9.

Based on Figs. 8 and 9, we can provide a graphical proof of whether Q_α is unitary, as shown in Fig. 10.

APPENDIX C: ENTANGLEMENT RÉNYI ENTROPY OF PURE STATES FROM DMV

Given a $2n$ -qubit pure state written in the Schmidt form

$$|\varphi\rangle = \sum_i \lambda_i |a_i\rangle \otimes |b_i\rangle, \quad (\text{C1})$$

the entanglement Rényi entropy of order α of $|\varphi\rangle$ is then defined as

$$H_\alpha(|\varphi\rangle) = \frac{1}{1-\alpha} \ln_2 \left(\sum_i p_i^\alpha \right), \quad (\text{C2})$$

with $p_i = |\lambda_i|^2$. It can be proved that $H_\alpha(|\varphi\rangle)$ is nonincreasing in α for any $|\varphi\rangle$:

$$H_0(|\varphi\rangle) \geq H_1(|\varphi\rangle) \geq H_2(|\varphi\rangle) \geq \dots \geq H_\infty(|\varphi\rangle). \quad (\text{C3})$$

Given an n -qubit density matrix $\rho = \sum_i p_i |s_i\rangle \langle s_i|$, its corresponding $2n$ -qubit pure state $|\rho\rangle$ by DMV is defined as

$$|\rho\rangle = \frac{1}{\sqrt{\sum_i p_i^2}} \sum_i p_i |s_i\rangle \otimes |s_i\rangle, \quad (\text{C4})$$

and its corresponding $2n$ -qubit pure state $|\psi_\rho\rangle$ by purification is defined as

$$|\psi_\rho\rangle = \sum_i \sqrt{p_i} |s_i\rangle \otimes |s_i\rangle. \quad (\text{C5})$$

We will show in the following that the order- α Rényi entropy of $|\rho\rangle$ is always smaller than that of $|\psi_\rho\rangle$. First, consider the case where $\alpha > 1$. Based on the relation [Eq. (C3)], we

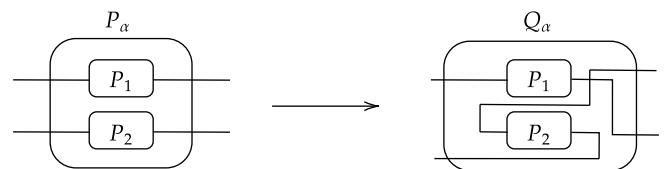
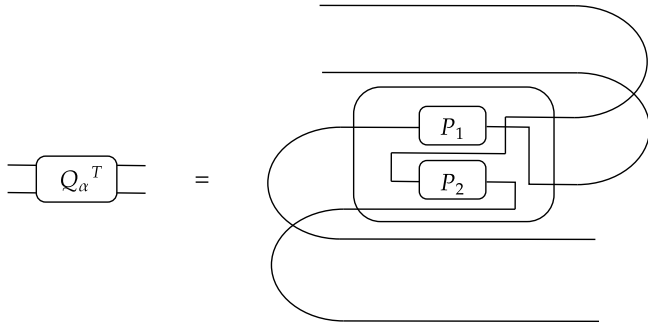


FIG. 8. Graphical representation of the replacement operators for two-qubit Pauli operators.

FIG. 9. The graphical representation of $Q_\alpha^\dagger$.

can show that

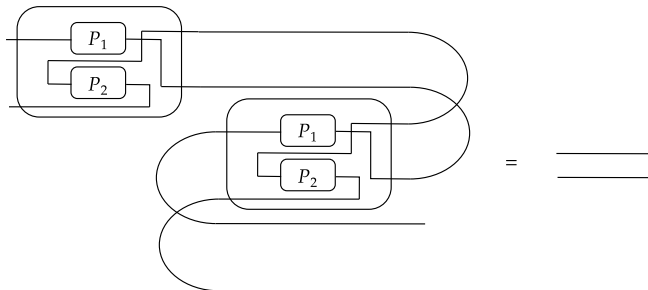
$$\begin{aligned}
 & H_\alpha(|\rho\rangle) - H_\alpha(|\psi_\rho\rangle) \\
 &= \frac{1}{1-\alpha} \ln_2 \left(\sum_i \left(\frac{p_i^2}{\sum_i p_i^2} \right)^\alpha \right) - H_\alpha(|\psi_\rho\rangle) \\
 &= \frac{1}{1-\alpha} \ln_2 \left(\sum_i p_i^{2\alpha} \right) - \frac{\alpha}{1-\alpha} \ln_2 \left(\sum_i p_i^2 \right) \\
 &\quad - H_\alpha(|\psi_\rho\rangle) \\
 &= \frac{1-2\alpha}{1-\alpha} H_{2\alpha}(|\psi_\rho\rangle) + \frac{\alpha}{1-\alpha} H_2(|\psi_\rho\rangle) - H_\alpha(|\psi_\rho\rangle) \\
 &\leq \frac{1-2\alpha}{1-\alpha} H_{2\alpha}(|\psi_\rho\rangle) + \frac{\alpha}{1-\alpha} H_\alpha(|\psi_\rho\rangle) - H_\alpha(|\psi_\rho\rangle) \\
 &= \frac{1-2\alpha}{1-\alpha} (H_{2\alpha}(|\psi_\rho\rangle) - H_\alpha(|\psi_\rho\rangle)) \leq 0. \quad (C6)
 \end{aligned}$$

For the $\alpha < 1$ case, since

$$\begin{aligned}
 & H_\alpha(|\rho\rangle) - H_\alpha(|\psi_\rho\rangle) \\
 &= \frac{1}{1-\alpha} \left[\ln_2 \left(\sum_i p_i^{2\alpha} \right) - \alpha \ln_2 \left(\sum_i p_i^2 \right) \right. \\
 &\quad \left. - \ln_2 \left(\sum_i p_i^\alpha \right) \right],
 \end{aligned}$$

we need to show that

$$\ln_2 \left(\sum_i p_i^{2\alpha} \right) - \ln_2 \left(\sum_i p_i^\alpha \right) \leq \alpha \ln_2 \left(\sum_i p_i^2 \right). \quad (C7)$$

FIG. 10. The graphical representation of $Q_\alpha Q_\alpha^\dagger = I$.

Define function $f(x) := \ln_2(\sum_i p_i^x)$. Equation (C7) can be rewritten as

$$f(2\alpha) - f(\alpha) \leq \alpha f(2), \quad (C8)$$

with $\alpha < 1$. Function f has the following two properties:

- (1) $f(1) = 0$, and
- (2) f is a convex function. We can verify that

$$f''(x) = \frac{1}{\ln 2} \frac{\sum_i p_i^x (\ln p_i - \mathbb{E}[\ln p_i])^2}{\sum_i p_i^x} \geq 0,$$

where $\mathbb{E}[\ln p_i] := \frac{\sum_i p_i^x \ln p_i}{\sum_i p_i^x}$. Thus, $f(x)$ is a convex function.

For the left-hand side of Eq. (C8),

$$\begin{aligned}
 f(2\alpha) - f(\alpha) &= \int_\alpha^{2\alpha} f'(x) dx \leq \int_1^{1+\alpha} f'(x) dx \\
 &= f(1+\alpha) - f(1). \quad (C9)
 \end{aligned}$$

According to the mean value theorem, there exist s and t with $1 \leq s \leq 1+\alpha \leq t \leq 2$ such that $f(1+\alpha) - f(1) = \alpha f'(s)$ and $f(2) - f(1+\alpha) = (1-\alpha)f'(t)$. Then,

$$\begin{aligned}
 f(2) &= f(2) - f(1+\alpha) + f(1+\alpha) - f(1) \\
 &= (1-\alpha)f'(t) + \alpha f'(s) \geq f'(s) \\
 &= \frac{1}{\alpha} (f(1+\alpha) - f(1)). \quad (C10)
 \end{aligned}$$

Combining Eqs. (C9) and (C10), we can prove Eq. (C8). Finally, the same conclusion holds for the special case $\alpha = 1$, as the limit of $\alpha > 1$ or $\alpha < 1$.

In Fig. 11, we show the comparison of Rényi entropy of order 1, 2, and ∞ between pure states from two-dimensional density matrices by DMV and by purification.

The entanglement entropy can be enlarged by classical linear combinations. Consider the mapping

$$\begin{aligned}
 & \{\rho_1, \rho_2, \dots, \rho_K, c_1, c_2, \dots, c_K\} \rightarrow |\psi\rangle \\
 &= \frac{1}{C_\psi} \sum_{ij} (c_1 \rho_{1,ij} + c_2 \rho_{2,ij} + \dots + c_K \rho_{K,ij}) |i\rangle |j\rangle \quad (C11)
 \end{aligned}$$

as in the main text. The reduced density matrices on the row system of $|\psi\rangle$ is

$$\begin{aligned}
 \text{Tr}_c(|\psi\rangle\langle\psi|) &= \frac{\sum_{\alpha\beta jk} c_\alpha c_\beta^* \rho_{\alpha,ij} \rho_{\beta,jk} |i\rangle\langle k|}{C_\psi^2} \\
 &= \frac{\sum_{\alpha\beta} c_\alpha c_\beta^* \rho_\alpha \rho_\beta}{C_\psi^2}. \quad (C12)
 \end{aligned}$$

Here, we only consider the second-order Rényi entropy of $|\psi\rangle$, which is

$$\begin{aligned}
 H_2(|\psi\rangle) &= -\ln \left(\frac{\sum_{abcd} c_a c_b^* c_c c_d^* \text{Tr}(\rho_a \rho_b \rho_c \rho_d)}{C_\psi^4} \right) \\
 &\approx -\ln \left(\frac{\sum_a |c_a|^4 \text{Tr}(\rho_a^4)}{(\sum_a |c_a|^2 \text{Tr}(\rho_a^2))^2} \right), \quad (C13)
 \end{aligned}$$

where the approximation is due to Eqs. (A8) and (A9). For n -qubit density matrices ρ_i and ρ_j , when they are generated randomly and uncorrelatedly by circuits such as the Schmidt

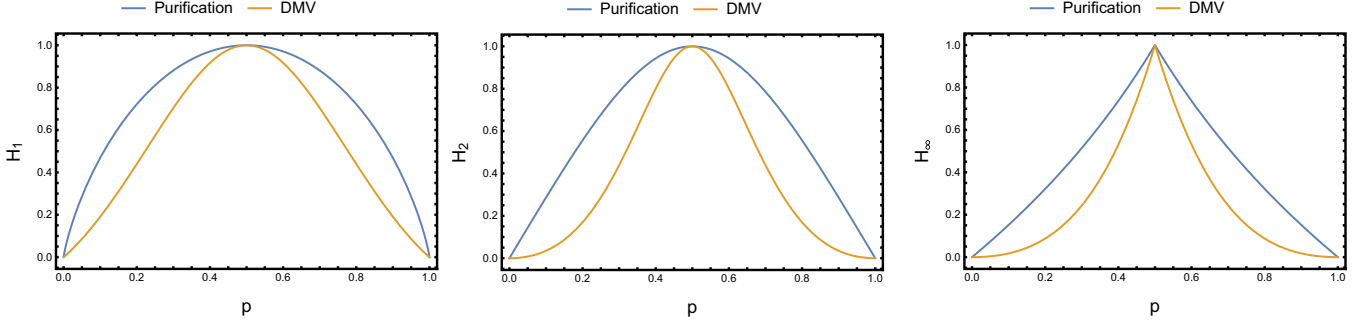


FIG. 11. Left: Rényi entropy of order 1 for two-dimensional density matrices. Middle: Rényi entropy of order 2 for two-dimensional density matrices. Right: Rényi entropy of order ∞ for two-dimensional density matrices.

ansatz with the n -qubit unitary in the upper state forming a 1-design, we have the value of $\text{Tr}(\rho_i \rho_j)$ concentrating on $\frac{1}{2^n}$, which can be ignored when n is large. For values like $\text{Tr}(\rho_a \rho_b \rho_c \rho_d)$, according to Eq. (A5), they will concentrate on $\frac{1}{2^{3n}}$, which can of course also be ignored. To make a comparison with the $K = 1$ case, we can simply assume $\text{Tr}(\rho_i^4) = \text{Tr}(\rho^4)$ and $\text{Tr}(\rho_i^2) = \text{Tr}(\rho^2)$ for all i , and then we have

$$\begin{aligned}
 H_2(|\psi\rangle) &\approx -\ln\left(\frac{\sum_a |c_a|^4 \text{Tr}(\rho^4)}{(\sum_a |c_a|^2 \text{Tr}(\rho^2))^2}\right) \\
 &= -\ln\left(\frac{\text{Tr}(\rho^4)}{\text{Tr}(\rho^2)^2}\right) - \ln\left(\frac{\sum_a |c_a|^4}{(\sum_a |c_a|^2)^2}\right) \\
 &\leq H_2(|\rho\rangle) - \ln\left(\frac{\sum_a |c_a|^4}{(\sum_a |c_a|^2)^2 K}\right) \\
 &= H_2(|\rho\rangle) + \ln(K).
 \end{aligned} \tag{C14}$$

The inequality originated from Cauchy-Schwarz inequality. Thus, the classical linear combination can give an enhancement no more than $\ln(K)$ for entanglement entropy H_2 .

APPENDIX D: ANSATZ

1. Schmidt ansatz

Based on the fact that only the reduced density matrices ρ_i supported on the first n -qubit block are used in our proposal, we developed an ansatz, Schmidt ansatz, in the main text to achieve a better control on the purity of ρ_i as well as to reduce the number of parameters. Here, we demonstrate that as long as the dimension χ of the Schmidt decomposition of the desired ground state $|\Psi_{\text{gs}}\rangle$ is upper bounded by 2^L (assuming that the system is partitioned into two blocks that contain the first n qubits and the second n qubits, respectively), then it suffices to use L CNOT gates between the two blocks in the Schmidt ansatz.

Under the Schmidt decomposition, $|\Psi_{\text{gs}}\rangle$ can be represented as

$$|\Psi_{\text{gs}}\rangle = \sum_{i=1}^{\chi} \lambda_i |\phi_i\rangle |\psi_i\rangle, \tag{D1}$$

where $\lambda_i > 0$ for $\forall i \in \{1, \dots, \chi\}$, and $\{|\phi_i\rangle | i \in \{1, \dots, \chi\}\}$ and $\{|\psi_i\rangle | i \in \{1, \dots, \chi\}\}$ are both sets of orthonormal n -qubit states. We observe that Eq. (D1) can be rewritten as

$$\begin{aligned}
 |\Psi_{\text{gs}}\rangle &= \frac{1}{4} \sum_{i=1}^{\chi} \lambda_i (|\phi_i\rangle + |\psi_i^*\rangle)(|\phi_i^*\rangle + |\psi_i\rangle) \\
 &\quad - \frac{1}{4} \sum_{i=1}^{\chi} \lambda_i (|\phi_i\rangle - |\psi_i^*\rangle)(|\phi_i^*\rangle - |\psi_i\rangle) \\
 &\quad + \frac{i}{4} \sum_{i=1}^{\chi} \lambda_i (-i|\phi_i\rangle + |\psi_i^*\rangle)(i|\phi_i^*\rangle + |\psi_i\rangle) \\
 &\quad - \frac{i}{4} \sum_{i=1}^{\chi} \lambda_i (-i|\phi_i\rangle - |\psi_i^*\rangle)(i|\phi_i^*\rangle - |\psi_i\rangle),
 \end{aligned} \tag{D2}$$

such that we can find four density matrices $\rho_1, \dots, \rho_4$ with $|\Psi_{\text{gs}}\rangle = c_1|\rho_1\rangle - c_2|\rho_2\rangle + ic_3|\rho_3\rangle - ic_4|\rho_4\rangle$, where $c_i \geq 0$ and $|\rho_i\rangle$ is proportional to the i th row in the right-hand side of Eq. (D2) for $\forall i \in \{1, 2, 3, 4\}$. It is straightforward to see that $\text{rank } \rho_i \leq \chi$ for $\forall i \in \{1, 2, 3, 4\}$. Then, as long as $\chi \leq 2^L$, each ρ_i can be prepared via the Schmidt ansatz with L CNOT gates between the upper block and the lower block.

2. Chemical-inspired ansatz

The electric ground state of a paramagnetic molecule has an equal number of spin-up electrons and spin-down electrons. Consider a paramagnetic molecule with $2k$ electrons and $2n$ spin orbits (the first n orbits for spin-up electrons and the second n orbits for spin-down electrons), we use $a_{i,\uparrow}^\dagger$ ($a_{i,\uparrow}$) to denote the creation (annihilation) operator for the i th spin-up orbit and similarly, $a_{i,\downarrow}^\dagger$ ($a_{i,\downarrow}$) the creation (annihilation) operator for the i th spin-down orbit. To account for its spin symmetry, we use the following ansatz circuit based on the q-UCCSD ansatz [73]:

$$U(\theta) := U_\uparrow(\theta)U_\downarrow(\theta)U_{\uparrow\downarrow}(\theta)U_{\uparrow\uparrow}(\theta)U_{\downarrow\downarrow}(\theta), \tag{D3}$$

with

$$\begin{aligned}
U_{\uparrow}(\theta) &:= \prod_{i>j} \exp(\theta_{ij}^{\uparrow}(a_{i,\uparrow}^{\dagger}a_{j,\uparrow} - a_{j,\uparrow}^{\dagger}a_{i,\uparrow})), \\
U_{\downarrow}(\theta) &:= \prod_{i>j} \exp(\theta_{ij}^{\downarrow}(a_{i,\downarrow}^{\dagger}a_{j,\downarrow} - a_{j,\downarrow}^{\dagger}a_{i,\downarrow})), \\
U_{\uparrow\downarrow}(\theta) &:= \prod_{i>l, j>k} \exp(\theta_{ijkl}^{\uparrow\downarrow}(a_{i,\uparrow}^{\dagger}a_{j,\downarrow}^{\dagger}a_{k,\downarrow}a_{l,\uparrow} - a_{l,\uparrow}^{\dagger}a_{k,\downarrow}^{\dagger}a_{j,\downarrow}a_{i,\uparrow})), \\
U_{\uparrow\uparrow}(\theta) &:= \prod_{i>j>k>l} \exp(\theta_{ijkl}^{\uparrow\uparrow}(a_{i,\uparrow}^{\dagger}a_{j,\uparrow}^{\dagger}a_{k,\uparrow}a_{l,\uparrow} - a_{l,\uparrow}^{\dagger}a_{k,\uparrow}^{\dagger}a_{j,\uparrow}a_{i,\uparrow})), \\
U_{\downarrow\downarrow}(\theta) &:= \prod_{i>j>k>l} \exp(\theta_{ijkl}^{\downarrow\downarrow}(a_{i,\downarrow}^{\dagger}a_{j,\downarrow}^{\dagger}a_{k,\downarrow}a_{l,\downarrow} - a_{l,\downarrow}^{\dagger}a_{k,\downarrow}^{\dagger}a_{j,\downarrow}a_{i,\downarrow})), \tag{D4}
\end{aligned}$$

where each term above conserves the number of spin-up electrons and the number of spin-down electrons. We can use a qubit system to simulate this fermionic system by using the Jordan-Wigner (JW) transformation,

$$\begin{aligned}
a_{i,\uparrow}^{\dagger} &= \sigma_{i,\uparrow}^{-} \prod_{r=1}^{i-1} Z_{r,\uparrow} = \frac{1}{2}(X_{i,\uparrow} - iY_{i,\uparrow}) \prod_{r=1}^{i-1} Z_{r,\uparrow}, \\
a_{i,\uparrow} &= \sigma_{i,\uparrow}^{+} \prod_{r=1}^{i-1} Z_{r,\uparrow} = \frac{1}{2}(X_{i,\uparrow} + iY_{i,\uparrow}) \prod_{r=1}^{i-1} Z_{r,\uparrow}, \\
a_{i,\downarrow}^{\dagger} &= \sigma_{i,\downarrow}^{-} \prod_{r=1}^{i-1} Z_{r,\downarrow} \prod_{m=1}^n Z_{m,\uparrow} = \frac{1}{2}(X_{i,\downarrow} - iY_{i,\downarrow}) \prod_{r=1}^{i-1} Z_{r,\downarrow} \prod_{m=1}^n Z_{m,\uparrow}, \\
a_{i,\downarrow} &= \sigma_{i,\downarrow}^{+} \prod_{r=1}^{i-1} Z_{r,\downarrow} \prod_{m=1}^n Z_{m,\uparrow} = \frac{1}{2}(X_{i,\downarrow} + iY_{i,\downarrow}) \prod_{r=1}^{i-1} Z_{r,\downarrow} \prod_{m=1}^n Z_{m,\uparrow}, \tag{D5}
\end{aligned}$$

where we use $2n$ qubits and a Pauli operator P on the i th qubit in the first (second) block of n qubits is denoted by $P_{i,\uparrow}$ ($P_{i,\downarrow}$). In this way, the $|1\rangle$ state on the i th qubit in the first (second) block indicates the presence of an electron on the i th spin-up (spin-down) orbital. Thus, for traditional VQE (on a qubit system), to restrict output states to respect the spin symmetry in the noiseless setting, we simply need to start with the initial state $|1, 1, \dots, 1, 0, \dots, 0, 1, 1, \dots, 1, 0, \dots, 0\rangle$ with k 1s in the first and second block, respectively, and run the corresponding parametrized circuit of Eq. (D3). [See Eqs. (D6) and (D7) for explicit forms of the unitaries used in the parametrized circuit.]

As for our algorithm, we obtain the output state $|\rho(\theta)\rangle$ as a combination of four states $|\rho(\theta)\rangle = \sum_{i=1}^4 c_i |\rho_i(\theta)\rangle$, where each $\rho_i(\theta)$ is the reduced density matrix on the first half of all output qubits. Denote the output state corresponding

to $\rho_i(\theta)$ as $|\psi_i(\theta)\rangle$. Suppose the Schmidt decomposition of $|\psi_i(\theta)\rangle$ reads $|\psi_i(\theta)\rangle = \sum_{\ell} \lambda_{\ell} |\psi_{\ell,\uparrow}\rangle |\psi_{\ell,\downarrow}\rangle$ with $\lambda_{\ell} \in \mathbb{R}^+$, $\{|\psi_{\ell,\uparrow}\rangle\}$ supported on the first n qubits and $\{|\psi_{\ell,\downarrow}\rangle\}$ supported on the rest of the qubits, we obtain $|\rho_i(\theta)\rangle \propto \sum_i \lambda_i^2 |\psi_{\ell,\uparrow}(\theta)\rangle |\psi_{\ell,\uparrow}(\theta)\rangle$. Then, as long as each $|\psi_{\ell,\uparrow}\rangle$ is only composed of states with k spin-up electrons, each $|\rho_i(\theta)\rangle$ satisfies the spin symmetry and the final state $|\rho(\theta)\rangle$ also satisfies the spin symmetry. Thus, we can see that we may simply use the same strategy as the traditional VQE, namely, the same parametrized circuit [Eq. (D3)] and the same initialization to guarantee, in the noiseless circuit setting, the final state $|\rho(\theta)\rangle$ possesses the spin symmetry.

In the following, we describe explicitly the parametrized circuit ansatz that realizes $U(\theta)$ with (parametrized) single-qubit gates and CNOT gates. After the JW transformation, the five unitary operators on the right-hand side of Eq. (D4) become

$$\begin{aligned}
U_{\uparrow}(\theta) &= \prod_{i>j} \exp\left(\frac{i\theta_{ij}^{\uparrow}}{2} \prod_{r=j+1}^{i-1} Z_{r,\uparrow}(X_i Y_j)\right) \times \exp\left(-\frac{i\theta_{ij}^{\uparrow}}{2} \prod_{r=j+1}^{i-1} Z_{r,\uparrow}(Y_i X_j)\right), \\
U_{\uparrow\uparrow}(\theta) &= \prod_{i>j>k>l} \exp\left(\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow}(X_{i,\uparrow} Y_{j,\uparrow} X_{k,\uparrow} X_{l,\uparrow})\right) \exp\left(\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow}(Y_{i,\uparrow} X_{j,\uparrow} X_{k,\uparrow} X_{l,\uparrow})\right)
\end{aligned}$$

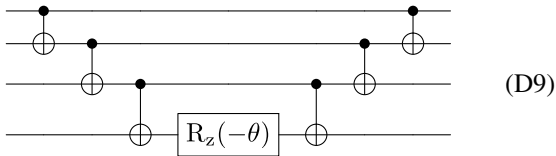
$$\begin{aligned}
& \times \exp \left(\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (Y_{i,\uparrow} Y_{j,\uparrow} X_{k,\uparrow} Y_{l,\uparrow}) \right) \exp \left(\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (Y_{i,\uparrow} Y_{j,\uparrow} Y_{k,\uparrow} X_{l,\uparrow}) \right) \\
& \times \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (X_{i,\uparrow} X_{j,\uparrow} X_{k,\uparrow} Y_{l,\uparrow}) \right) \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (X_{i,\uparrow} X_{j,\uparrow} Y_{k,\uparrow} X_{l,\uparrow}) \right) \\
& \times \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (X_{i,\uparrow} Y_{j,\uparrow} Y_{k,\uparrow} Y_{l,\uparrow}) \right) \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\uparrow}}{8} \prod_{r=j+1}^{i-1} Z_{r,\uparrow} \prod_{r'=l+1}^{k-1} Z_{r',\uparrow} (Y_{i,\uparrow} X_{j,\uparrow} Y_{k,\uparrow} Y_{l,\uparrow}) \right), \quad (D6)
\end{aligned}$$

$$\begin{aligned}
U_{\uparrow\downarrow}(\theta) = & \prod_{i>j>k>l} \exp \left(\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (X_{i,\uparrow} Y_{l,\uparrow} X_{j,\downarrow} X_{k,\downarrow}) \right) \exp \left(\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (X_{i,\uparrow} X_{l,\uparrow} X_{j,\downarrow} Y_{k,\downarrow}) \right) \\
& \times \exp \left(\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (X_{i,\uparrow} Y_{l,\uparrow} Y_{j,\downarrow} Y_{k,\downarrow}) \right) \exp \left(\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (Y_{i,\uparrow} Y_{l,\uparrow} X_{j,\downarrow} Y_{k,\downarrow}) \right) \\
& \times \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (Y_{i,\uparrow} X_{l,\uparrow} X_{j,\downarrow} X_{k,\downarrow}) \right) \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (X_{i,\uparrow} X_{l,\uparrow} Y_{j,\downarrow} Y_{k,\downarrow}) \right) \\
& \times \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\uparrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (Y_{i,\uparrow} X_{l,\uparrow} Y_{j,\downarrow} Y_{k,\downarrow}) \right) \exp \left(-\frac{i\theta_{ijkl}^{\uparrow\downarrow}}{8} \prod_{r=l+1}^{i-1} Z_{r,\downarrow} \prod_{r'=k+1}^{j-1} Z_{r',\downarrow} (Y_{i,\uparrow} Y_{l,\uparrow} Y_{j,\downarrow} X_{k,\downarrow}) \right), \quad (D7)
\end{aligned}$$

and $U_{\downarrow}(\theta)$ and $U_{\downarrow\downarrow}(\theta)$ are similar to $U_{\uparrow}(\theta)$ and $U_{\uparrow\uparrow}(\theta)$, respectively, and are not displayed here. We observe that $U(\theta)$ comprises Pauli-string rotation operators of the form $\exp(i\frac{\theta}{2}\mathbf{P})$, where $\mathbf{P} := \prod_{r \in \text{supp}} P_r$ is a product of single-qubit Pauli operators. For single-qubit Pauli rotation gates, we can transform $R_x(\theta) := \exp(-i\frac{\theta}{2}X)$ to $R_z(\theta) := \exp(-i\frac{\theta}{2}Z)$ and $R_y(\theta) := \exp(-i\frac{\theta}{2}Y)$ to $R_z(\theta)$ using the following two equalities:

$$HR_x(\theta)H = R_z(\theta) \quad \text{and} \quad R_x(\pi/2)R_y(\theta)R_x(-\pi/2) = R_z(\theta), \quad (D8)$$

which are the result of these two facts: $HXH = Z$ and $R_x(\pi/2)YR_x(-\pi/2) = Z$. By generalizing the above argument, we can use single-qubit rotations to transform $\exp(i\frac{\theta}{2}\mathbf{P})$ to $\exp(i\frac{\theta}{2}\prod_{r \in \text{supp}} Z_r)$, which then can be realized by sandwiching a single R_z gate between two staircases of CNOT gates. For instance, $\exp(i\frac{\theta}{2}Z_1 \otimes \cdots \otimes Z_4)$ can be realized by the circuit in Eq. (D9).



APPENDIX E: SAMPLING COMPLEXITY

1. Ratio estimator

Given two random variables X and Y with mean values μ_x and μ_y , the ratio $\frac{X}{Y}$ can be expanded around the point (μ_x, μ_y)

by the Taylor series. The approximation of the expectation value $E[\frac{X}{Y}]$ up to the second-order Taylor expansion is

$$\begin{aligned}
E\left[\frac{X}{Y}\right] & \approx E\left[\frac{\mu_x}{\mu_y} + \frac{1}{\mu_y}(X - \mu_x) - \frac{\mu_x}{\mu_y^2}(Y - \mu_y) + 0 \right. \\
& \quad \left. - \frac{1}{\mu_y^2}(X - \mu_x)(Y - \mu_y) + \frac{\mu_x}{\mu_y^3}(Y - \mu_y)^2 \right] \\
& = \frac{\mu_x}{\mu_y} - \frac{1}{\mu_y^2}\text{Cov}[X, Y] + \frac{\mu_x}{\mu_y^3}\text{Var}[Y]. \quad (E1)
\end{aligned}$$

For the variance $\text{Var}[\frac{X}{Y}] = E[(\frac{X}{Y} - \frac{\mu_x}{\mu_y})^2]$, we can similarly obtain its approximation up to the first-order Taylor expansion:

$$\begin{aligned}
\text{Var}\left[\frac{X}{Y}\right] & \approx E\left[\left(\frac{1}{\mu_y}(X - \mu_x) - \frac{\mu_x}{\mu_y^2}(Y - \mu_y)\right)^2\right] \\
& = E\left[\frac{1}{\mu_y^2}(X - \mu_x)^2 - 2\frac{\mu_x}{\mu_y^3}(X - \mu_x)(Y - \mu_y) \right. \\
& \quad \left. + \frac{\mu_x^2}{\mu_y^4}(Y - \mu_y)^2\right] \\
& = \frac{1}{\mu_y^2}\text{Var}[X] - 2\frac{\mu_x}{\mu_y^3}\text{Cov}[X, Y] + \frac{\mu_x^2}{\mu_y^4}\text{Var}[Y]. \quad (E2)
\end{aligned}$$

The approximations in Eqs. (E1) and (E2) are valid as long as the variances $\text{Var}[X]$ and $\text{Var}[Y]$ and the covariance $\text{Cov}[X, Y]$ are small enough. Equations (E1) and (E2) indicate the following estimator:

(1) Estimator for $\frac{\mu_x}{\mu_y}$: Assuming X and Y are two independent random variables, the value of the ratio of their means $\frac{\mu_x}{\mu_y}$ can be estimated by an asymptotically unbiased estimator $\frac{\bar{X}}{\bar{Y}}$, where $\bar{X}$ and $\bar{Y}$ are the averages of X and Y from sampling. The expectation and the variance of this estimator are

$$E\left[\frac{\bar{X}}{\bar{Y}}\right] = \frac{\mu_x}{\mu_y} + \text{Bias}\left[\frac{\bar{X}}{\bar{Y}}\right] \approx \frac{\mu_x}{\mu_y} + \frac{\mu_x}{\mu_y^3} \text{Var}[\bar{Y}], \quad (\text{E3})$$

$$\text{Var}\left[\frac{\bar{X}}{\bar{Y}}\right] \approx \frac{1}{\mu_y^2} \text{Var}[\bar{X}] + \frac{\mu_x^2}{\mu_y^4} \text{Var}[\bar{Y}]. \quad (\text{E4})$$

The variance and the bias give the mean squared error (MSE):

$$\begin{aligned} \text{MSE}\left[\frac{\bar{X}}{\bar{Y}}\right] &= \text{Bias}\left[\frac{\bar{X}}{\bar{Y}}\right]^2 + \text{Var}\left[\frac{\bar{X}}{\bar{Y}}\right] \\ &\approx \text{Var}\left[\frac{\bar{X}}{\bar{Y}}\right] \approx \frac{1}{\mu_y^2} \text{Var}[\bar{X}] + \frac{\mu_x^2}{\mu_y^4} \text{Var}[\bar{Y}], \end{aligned} \quad (\text{E5})$$

where the first approximation is valid when $\text{Var}[\bar{Y}]^2 \ll \text{Var}[\bar{Y}]$.

2. Overall analysis

We now analyze the sampling complexity of estimating

$$\begin{aligned} \langle \psi | H_A | \psi \rangle &= \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{C_\psi^2} \\ &= \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}, \end{aligned} \quad (\text{E6})$$

where $H_A = \sum_{\alpha=1}^m g_\alpha P_\alpha$ and $H_B = \sum_{\alpha=1}^m g_\alpha Q_\alpha$, where P_α and Q_α are $2n$ -qubit Pauli operators and their substitute operators, respectively. Since $\langle \psi | H_A | \psi \rangle$ is real, for each $g_\alpha c_i^* c_j \text{Tr}(Q_\alpha \rho_i \otimes \rho_j)$, we can separate it into $g_\alpha \text{Re}[c_i^* c_j] \text{Tr}(\text{Re}[Q_\alpha] \rho_i \otimes \rho_j)$ and $-g_\alpha \text{Im}[c_i^* c_j] \text{Tr}(\text{Im}[Q_\alpha] \rho_i \otimes \rho_j)$. Since the eigenvalues of all Pauli substitute operators belong to $\{1, -1, i, -i\}$, during the measurements, each $\text{Re}[Q_\alpha]$ (and $\text{Im}[Q_\alpha]$) can be seen as a random variable: X with an expectation $E[X] = p_0 * 0 + p_1 * 1 + p_{-1} * (-1)$ and a bounded variance $\text{Var}[X] = p_1(1 - p_1) + p_{-1}(1 - p_{-1}) + 2p_1 p_{-1} \leq 1$. For each of such a part, we can assign $N_n/(2K^2m)$ measurements to estimate its value, which means there are N_n measurements in total for estimating the numerator. (We add the zero imaginary parts when $i = j$ for simplicity.) For the denominator, since $\text{Tr}(S \rho_k \otimes \rho_l)$ is real, we only need to estimate $\text{Re}[c_k^* c_l] \text{Tr}(S \rho_k \otimes \rho_l)$. During the measurements, $\text{Tr}(S \rho_k \otimes \rho_l)$ can be seen as a random variable: Z_d with an expectation $E[Z_d] = p_1 * 1 + p_{-1} * (-1)$ and a bounded variance $\text{Var}[Z_d] = 4p_1(1 - p_1) \leq 1$. For each of such a part, we can assign $N_d/(K^2)$ measurements to estimate its value, which means there are N_d measurements in total for estimating the denominator. Following these settings, the MSE of estimating $\langle \psi | H_A | \psi \rangle$ using Eq. (E5) is

$$\begin{aligned} \text{MSE}[\langle \psi | H_A | \psi \rangle] &\approx \frac{1}{C_\psi^4} \sum_{ij\alpha} (\text{Re}[c_i^* c_j]^2 g_\alpha^2 \text{Var}[\text{Tr}(\text{Re}[Q_\alpha] \rho_i \otimes \rho_j)] + \text{Im}[c_i^* c_j]^2 g_\alpha^2 \text{Var}[\text{Tr}(\text{Im}[Q_\alpha] \rho_i \otimes \rho_j)]) \\ &\quad + \frac{\langle \psi | H_A | \psi \rangle^2}{C_\psi^4} \sum_{ij} \text{Re}[c_i^* c_j]^2 \text{Var}[\text{Tr}(S \rho_i \otimes \rho_j)] \\ &\leq \frac{2K^2m}{C_\psi^4 N_n} \sum_{ij\alpha} g_\alpha^2 |c_i|^2 |c_j|^2 + \frac{\langle \psi | H_A | \psi \rangle^2 K^2}{C_\psi^4 N_d} \sum_{ij} |c_i|^2 |c_j|^2 \\ &= \frac{\sum_{ij} |c_i|^2 |c_j|^2}{C_\psi^4} \left(\frac{2K^2m}{N_n} \sum_{\alpha} g_\alpha^2 + \frac{K^2}{N_d} \langle \psi | H_A | \psi \rangle^2 \right) \\ &\leq \frac{\sum_{ij} |c_i|^2 |c_j|^2}{C_\psi^4} \left(\frac{2K^2m \|H_A\|_F^2}{2^{2n} N_n} + \frac{K^2 \|H_A\|_2^2}{N_d} \right) = \chi \left(\frac{2K^2m \|H_A\|_F^2}{2^{2n} N_n} + \frac{K^2 \|H_A\|_2^2}{N_d} \right), \end{aligned} \quad (\text{E7})$$

where $\|H_A\|_2$ and $\|H_A\|_F$ are the 2-norm and the Frobenius norm of H_A . To further estimate the value of $\chi = \frac{\sum_{ij} |c_i|^2 |c_j|^2}{C_\psi^4}$, we need to separate the noiseless and noisy cases. In the following, we will take the case of the Schmidt ansatz as an example, which can be easily generalized to other types of ansatz.

3. Noiseless case

Starting from $|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes n}$, the random Schmidt ansatz will give an output state

$$U_S |0\rangle^{\otimes n} \otimes |0\rangle^{\otimes n} = \sum_{i=0}^{2^L-1} \lambda_i (U(n)|i\rangle) \otimes |i\rangle, \quad (\text{E8})$$

with λ_i generated by an L -qubit random orthogonal circuit and U a random n -qubit unitary circuit. According to our encoding, only the reduced density matrix $\rho = \sum_{i=0}^{2^L-1} \lambda_i^2 U|i\rangle\langle i|U^\dagger$ is needed. In the expression of χ , we need to evaluate the values like $\text{Tr}(\rho\sigma)$ with $\sigma = \sum_{i=0}^{2^L-1} \gamma_i^2 V|i\rangle\langle i|U^\dagger$. If U and V are drawn from a 1-design ν , according to Eqs. (A8) and (A9), $\text{Tr}(\rho_i\rho_j)$ will concentrate on an exponentially small average:

$$\text{Prob}[\text{Tr}(\rho\sigma) \geq \varepsilon] \leq \frac{1}{2^{n\varepsilon}}, \quad (\text{E9})$$

which means that the probability of $\text{Tr}(\rho\sigma)$ greater than a polynomially small ε is exponentially small.

On the other hand, the form of the Schmidt circuit gives a lower bound on the purity: $\text{Tr}(\rho^2) \geq \frac{1}{2^L}$ since each CNOT gate connecting the upper and lower systems can be decomposed into two local tensors with a bond dimension equal to 2: $\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$. Note that these discussions hold for the other two ansätze as long as there are L two-qubit entangled gates connecting the upper and lower systems in total and the local circuits between two such gates form a 1-design, which can be easily satisfied since random Pauli circuits can already form a 1-design.

Using $\text{Tr}(\rho^2) \geq \frac{1}{2^L}$ and $\text{Tr}(\rho_i\rho_j) \approx 0$ when n is large, we have

$$\begin{aligned} \chi &= \frac{\sum_{ij} |c_i|^2 |c_j|^2}{\left(\sum_{kl} c_k^* c_l \text{Tr}(\mathcal{S}\rho_k \otimes \rho_l)\right)^2} \approx \frac{\sum_{ij} |c_i|^2 |c_j|^2}{\left(\sum_k |c_k|^2 \text{Tr}(\rho_k^2)\right)^2} \\ &\leq 2^{2L} \frac{\sum_{ij} |c_i|^2 |c_j|^2}{\left(\sum_k |c_k|^2\right)^2} = 2^{2L}. \end{aligned} \quad (\text{E10})$$

By putting this bound into Eq. (E7), we get

$$\text{MSE}[\langle \psi | A | \psi \rangle] \leq 2^{2L} K^2 \left(\frac{2m \|H_A\|_F^2}{2^{2n} N_n} + \frac{\|H_A\|_2^2}{N_d} \right). \quad (\text{E11})$$

Suppose $N_n = \frac{2N}{3}$ and $N_d = \frac{N}{3}$; to achieve an accuracy ϵ , the required number of measurements N should be

$$N \geq \frac{2^{2L} 3K^2}{\epsilon^2} \left(\frac{m \|H_A\|_F^2}{2^{2n}} + \|H_A\|_2^2 \right). \quad (\text{E12})$$

4. Noisy case

For the noisy case, we can think the original density matrix $\rho = \sum_{i=0}^{2^L-1} \lambda_i^2 U|i\rangle\langle i|U^\dagger$ suffers from a noise channel:

$$\mathcal{N}(\rho) = \sum_{\alpha} \sum_{i=0}^{2^L-1} \lambda_i^2 K_{\alpha} U|i\rangle\langle i|U^\dagger K_{\alpha}^{\dagger}, \quad (\text{E13})$$

with $\sum_{\alpha} K_{\alpha}^{\dagger} K_{\alpha} = I$. Under this noise channel, the average of $\text{Tr}(\mathcal{N}(\rho)\mathcal{N}(\sigma))$ can be estimated assuming that U and V form a 1-design:

$$\begin{aligned} \int_{\nu} \int_{\nu} \text{Tr}(\mathcal{N}(\rho)\mathcal{N}(\sigma)) dU dV &= \sum_{ij\alpha\beta} \lambda_i^2 \gamma_j^2 \text{Tr} \left(\int_{\nu} K_{\alpha} U|i\rangle\langle i|U^\dagger K_{\alpha}^{\dagger} dU \int_{\nu} K_{\beta} V|j\rangle\langle j|V^\dagger K_{\beta}^{\dagger} dV \right) \\ &= \sum_{ij\alpha\beta} \lambda_i^2 \gamma_j^2 \text{Tr} \left(\int_{\nu} U|i\rangle\langle i|U^\dagger dU K_{\alpha}^{\dagger} K_{\beta} \int_{\nu} V|j\rangle\langle j|V^\dagger dV K_{\beta}^{\dagger} K_{\alpha} \right) \\ &= \sum_{ij\alpha\beta} \lambda_i^2 \gamma_j^2 \frac{1}{2^{2n}} \text{Tr}(K_{\alpha} K_{\alpha}^{\dagger} K_{\beta} K_{\beta}^{\dagger}). \end{aligned} \quad (\text{E14})$$

Now suppose that the noise channel is unital, i.e., $\mathcal{N}(I) = I$, which means $\sum_{\alpha} K_{\alpha} K_{\alpha}^{\dagger} = I$, then we have $E(\text{Tr}(\mathcal{N}(\rho)\mathcal{N}(\sigma))) = \frac{1}{2^n}$, which coincides with the noiseless case. For nonunital noise, we need to require the channel to be close to the identity map to get the same conclusion. Note that if the channel is implemented before the unitary U , according to Eqs. (A8) and (A9), we can find that $\frac{1}{2^n}$ is the average no matter whether the channel is unital or not:

$$\sum_{ij} \lambda_i^2 \gamma_j^2 \text{Tr} \left(\int_{\nu} U \sum_{\alpha} K_{\alpha} |i\rangle\langle i| K_{\alpha}^{\dagger} U^\dagger dU \int_{\nu} V \sum_{\beta} K_{\beta} |j\rangle\langle j| K_{\beta}^{\dagger} V^\dagger dV \right) = \sum_{ij} \lambda_i^2 \gamma_j^2 \frac{1}{2^{2n}} \text{Tr}(I) = \frac{1}{2^n} \sum_{ij} \lambda_i^2 \gamma_j^2 = \frac{1}{2^n}. \quad (\text{E15})$$

This means that the unital requirement is only for the noise channel at the end of the circuit. These discussions also work for the alternate gate-noise models.

When the quantum circuit is noisy, the purity of the reduced density matrices in the upper system prepared by the circuit can be smaller than $\frac{1}{2^L}$ since the system can have entanglements with the environment. In this case, we can model noise as discrete, probabilistic faults that can happen at the various locations in the circuit including gates, idling steps, and measurements. Suppose at a location a , the probability

corresponding to no faults is $(1 - p_a)$, then the fault-free probability of the whole circuit is defined as $P_0 = \prod_a (1 - p_a)$ and the purity will be greater than P_0^2 . If $\zeta = \sum_a p_a \sim 1$ and p_a at different locations are the same, then the probability that l faults occur in the circuit is given by the Poisson distribution $P_l = e^{-\zeta} \frac{\zeta^l}{l!}$ due to the Le Cam's theorem with

$$P_0 = e^{-\zeta}. \quad (\text{E16})$$

Thus, we have

$$\chi \leq \max e^{4\zeta}, 2^{2L} \quad (\text{E17})$$

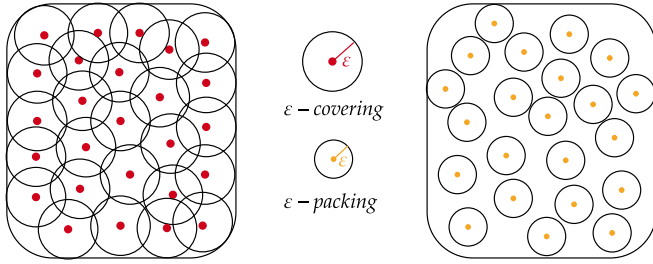


FIG. 12. In the left figure, an ε -cover $\mathcal{Q}$ of the metric space $(\mathcal{M}, d)$ consists of points in $\mathcal{M}$ such that every point $z \in \mathcal{M}$ is no more than ε away from some point $x \in \mathcal{Q}$. The size of the smallest ε -cover, denoted by $\mathcal{N}(\mathcal{M}, d, \varepsilon)$, is termed the ε -covering number of $(\mathcal{M}, d)$. In the right figure, an ε -packing $\mathcal{Q}$ is a subset of $\mathcal{M}$, where the distance between any pair of distinct points x and y in $\mathcal{Q}$ exceeds ε .

and

$$\text{MSE}[\langle \psi | A | \psi \rangle] \leq \max[e^{4\zeta}, 2^{2L}] K^2 \left(\frac{2m \|H_A\|_F^2}{2^{2n} N_n} + \frac{\|H_A\|_2^2}{N_d} \right), \quad (\text{E18})$$

$$N \geq \max[e^{4\zeta}, 2^{2L}] \frac{3K^2}{\varepsilon^2} \left(\frac{m \|H_A\|_F^2}{2^{2n}} + \|H_A\|_2^2 \right). \quad (\text{E19})$$

APPENDIX F: EXPRESSIBILITY BY COVERING NUMBER

1. Covering numbers and packing number

Figure 12 provides a geometric illustration of covering and packing numbers. A subset $\mathcal{Q} \subset \mathcal{M}$ in the set $\mathcal{M}$ is defined as an ε -covering set if for every point $z \in \mathcal{M}$, there exists an $x \in \mathcal{Q}$ such that $d(x, z) \leq \varepsilon$. Conversely, if the distance between any two distinct points x and y in $\mathcal{Q}$ is greater than ε , then this subset $\mathcal{Q} \subset \mathcal{M}$ is termed ε -packing set.

Consider the following problem in the metric space $(\mathcal{M}, d)$. We cannot guarantee the simultaneous existence of an ε -covering set $\mathcal{Q}$ with cardinality N and a 2ε -packing set $\mathcal{Q}'$ with cardinality $\tilde{N} > N$. If such sets exist, then within $\mathcal{Q}'$, there must be at least two elements y_1 and y_2 that lie within an ε -ball centered at some $x \in \mathcal{Q}$, leading to a distance $d(y_1, y_2) \leq 2\varepsilon$. This contradiction implies that covering and packing numbers satisfy the following inequality [77]:

$$\overline{\mathcal{N}}(\mathcal{M}, d, 2\varepsilon) \leq \mathcal{N}(\mathcal{M}, d, \varepsilon) \leq \overline{\mathcal{N}}(\mathcal{M}, d, \varepsilon). \quad (\text{F1})$$

And it is important to highlight that ε is a predefined hyperparameter, specifically a small constant within the interval $(0, 1)$, independent of any other factors [77]. Consequently, we can select ε to be arbitrarily small to meet our needs in the following proof.

We will use two lemmas from Ref. [84], which we include here for the reader's convenience.

Lemma F1 (Lemma 5 of Ref. [84]). Let $f : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ be a bi-Lipschitz mapping between the metric spaces $(\mathcal{M}_1, d_1)$ and $(\mathcal{M}_2, d_2)$, satisfying $f(\mathcal{M}_1) = \mathcal{M}_2$ and the following conditions:

$$\begin{aligned} d_2(f(x), f(y)) &\leq k_1 d_1(x, y), \quad \forall x, y \in \mathcal{M}_1, \\ d_2(f(x), f(y)) &\geq k_2 d_1(x, y), \quad \forall x, y \in \mathcal{M}_1 \quad \text{when} \\ d_1(x, y) &\leq r. \end{aligned} \quad (\text{F2})$$

Then, their covering numbers can be derived from each other and satisfy the following relationship:

$$\mathcal{N}\left(\mathcal{M}_1, d_1, \frac{2\varepsilon}{k_2}\right) \leq \mathcal{N}(\mathcal{M}_2, d_2, \varepsilon) \leq \mathcal{N}\left(\mathcal{M}_1, d_1, \frac{\varepsilon}{k_1}\right). \quad (\text{F3})$$

And the first inequality requires $\varepsilon \leq \frac{k_2 r}{2}$.

Lemma F2 (Lemma 1 of Ref. [84]). For $0 < \tilde{\varepsilon} \leq 1/10$, the $\tilde{\varepsilon}$ -covering number of the unitary group $U(n)$ with respect to the operator norm distance satisfies the following inequality:

$$\left(\frac{3}{4\tilde{\varepsilon}}\right)^{n^2} \leq \mathcal{N}(U(n), \|\cdot\|, \tilde{\varepsilon}) \leq \left(\frac{7}{\tilde{\varepsilon}}\right)^{n^2}. \quad (\text{F4})$$

Lemma F2 provided an expression for the upper bound of the covering number of the unitary group. Now, by leveraging Lemma F1, which gives the proportion k_1 relating the metric of density matrix vectorization to the metric of the unitary group, we derive the upper bound for the covering number of DMV in the following subsection.

2. The expressiveness of VQAs by DMV

This part will present the expressiveness of VQE under DMV. Let us start by reviewing some knowledge of matrix theory and some inequalities we will use.

Discrete form of Holder's inequality: if $p, q > 1$, and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\left| \sum_{k=1}^n a_k b_k \right| \leq \left(\sum_{k=1}^n |a_k|^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^n |b_k|^q \right)^{\frac{1}{q}}. \quad (\text{F5})$$

When $p = q = 2$, we obtain the Cauchy-Schwarz inequality:

$$\sum |x_i y_i| \leq \left(\sum |x_i|^2 \right)^{1/2} \left(\sum |y_i|^2 \right)^{1/2}. \quad (\text{F6})$$

In the content of this section, the norm is defined as follows: Suppose A is an $n \times n$ matrix. The operator norm $\|A\|$ is defined as

$$\|A\| = \sup_{\|x\|_2=1, x \in \mathbb{C}^n} \|Ax\|.$$

Alternatively, $\|A\| = \sqrt{\lambda_1(AA^\dagger)}$, where $\lambda_i(AA^\dagger)$ is the i th largest eigenvalue of the matrix $AA^\dagger$.

When using unconditionally decoherence-free quantum error mitigation, our output can be written as

$$\begin{aligned} \langle \psi | H_A | \psi \rangle &= \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{C_\psi^2} \\ &= \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}. \end{aligned} \quad (\text{F7})$$

Please note that here

$$\begin{aligned} &\text{Tr}(H_B \rho_i \otimes \rho_j) \\ &= \text{Tr}(H_B^{1,2} \otimes I^{3,4} \rho_i^{1,3} \otimes \rho_j^{2,4}) \\ &= \text{Tr}(H_B^{1,2} \otimes I^{3,4} U(\theta_I) \rho_{i0}^{1,3} U(\theta_I)^\dagger \otimes U(\theta_J) \rho_{j0}^{2,4} U(\theta_J)^\dagger). \end{aligned} \quad (\text{F8})$$

In all the equations below, we follow the equation relations given above, namely, $H_B = H_B^{1,2} \otimes I^{3,4}$ and $\rho_i \otimes \rho_j = U(\theta_i)\rho_{i0}^{1,3}U(\theta_i)^\dagger \otimes U(\theta_j)\rho_{j0}^{2,4}U(\theta_j)^\dagger$. So, following the form of Lemma F1, we have

$$\begin{aligned} d_2 & \left(\frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}, \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_{k\varepsilon} \otimes \rho_{l\varepsilon})} \right) \\ &= \left| \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)} - \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_{k\varepsilon} \otimes \rho_{l\varepsilon})} \right|. \end{aligned} \quad (\text{F9})$$

In Appendix A, we provide Eq. (A8), so if U and V are drawn from a 1-design v , $\text{Tr}(\rho_i \rho_j)$ will be exponentially small. In this case, we can simplify the denominator of the expression

$$\begin{aligned} d_2 & \left(\frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}, \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_{k\varepsilon} \otimes \rho_{l\varepsilon})} \right) \\ & \approx \left| \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_k |c_k|^2 \text{Tr}(\rho_k^2)} - \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_k |c_k|^2 \text{Tr}(\rho_{k\varepsilon}^2)} \right|. \end{aligned} \quad (\text{F10})$$

In Appendix A, we know that ε is a predefined hyperparameter, i.e., a very small constant, satisfying $\varepsilon \in (0, 1)$, and is independent of any factor. To satisfy Lemma F2, it is sufficient that $0 < \varepsilon \leq 1/10$. Therefore, for the sake of convenience in derivation, let us consider ε approaching 0, so that

$$\text{Tr}(\rho_k^2) = \text{Tr}(\rho_{k\varepsilon}^2). \quad (\text{F11})$$

In this case, our equation can be further simplified to

$$\begin{aligned} d_2 & \left(\frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}, \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_{k\varepsilon} \otimes \rho_{l\varepsilon})} \right) \\ & \approx \left| \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_k |c_k|^2 \text{Tr}(\rho_k^2)} - \frac{\sum_{i,j=1}^K c_i^* c_j \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})}{\sum_k |c_k|^2 \text{Tr}(\rho_k^2)} \right| \\ &= \left| \frac{\sum_{i,j=1}^K c_i^* c_j (\text{Tr}(H_B \rho_i \otimes \rho_j) - \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon}))}{\sum_k |c_k|^2 \text{Tr}(\rho_k^2)} \right| \\ & \leq \left| \frac{(\sum_{i,j=1}^K |c_i^* c_j|^2)^{1/2} (\sum_{i,j=1}^K |\text{Tr}(H_B \rho_i \otimes \rho_j) - \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})|^2)^{1/2}}{\sum_k |c_k|^2 \text{Tr}(\rho_k^2)} \right| \\ & \leq 2^L \left(\sum_{i,j=1}^K |\text{Tr}(H_B \rho_i \otimes \rho_j) - \text{Tr}(H_B \rho_{i\varepsilon} \otimes \rho_{j\varepsilon})|^2 \right)^{1/2} \\ &= 2^L \left(\sum_{i,j=1}^K |\text{Tr}(\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) \rho_{i0} \otimes \rho_{j0}) - \text{Tr}(\hat{U}_\varepsilon(\theta_{ij})^\dagger H_B \hat{U}_\varepsilon(\theta_{ij}) \rho_{i0} \otimes \rho_{j0})|^2 \right)^{1/2} \\ & \leq 2^L \left(\sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\varepsilon(\theta_{ij})^\dagger H_B \hat{U}_\varepsilon(\theta_{ij})\|^2 \text{Tr}(\rho_{i0} \otimes \rho_{j0}) \right)^{1/2} \\ &= 2^L \left(\sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\varepsilon(\theta_{ij})^\dagger H_B \hat{U}_\varepsilon(\theta_{ij})\|^2 \right)^{1/2} \\ & \leq 2^L \left(\left(\sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\varepsilon(\theta_{ij})^\dagger H_B \hat{U}_\varepsilon(\theta_{ij})\| \right)^2 \right)^{1/2} \\ &= 2^L \sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\varepsilon(\theta_{ij})^\dagger H_B \hat{U}_\varepsilon(\theta_{ij})\| \end{aligned}$$

$$\begin{aligned}
&= 2^L \mathcal{C} \left\| \sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \sum_{i,j=1}^K \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij}) \right\| \\
&= 2^L \mathcal{C} d_1 \left(\sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}), \sum_{i,j=1}^K \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij}) \right). \tag{F12}
\end{aligned}$$

The first inequality stems from the Cauchy-Schwarz inequality (F6), and the second inequality follows from Appendix E3 [Eq. (E10)]. Since each CNOT gate connecting the upper and lower systems can be decomposed into two local tensors with a bond dimension equal to 2: $\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$, we have

$$\frac{\sum_{ij} |c_i|^2 |c_j|^2}{(\sum_k |c_k|^2 \text{Tr}(\rho_k^2))^2} \leq 2^{2L} \frac{\sum_{ij} |c_i|^2 |c_j|^2}{(\sum_k |c_k|^2)^2} = 2^{2L}. \tag{F13}$$

Taking the square root yields our second inequality. The third inequality arises from the triangle inequality, while the fourth equality is due to $\|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij})\| \geq 0$. The last equality arises from the triangle inequality of matrix norms: $\|A + B\| \leq \|A\| + \|B\|$. Hence, we have

$$\begin{aligned}
&\sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij})\| \\
&\geq \left\| \sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \sum_{i,j=1}^K \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij}) \right\|. \tag{F14}
\end{aligned}$$

Here, we introduce a constant such that

$$\begin{aligned}
&\sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij})\| \\
&= \mathcal{C} \left\| \sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \sum_{i,j=1}^K \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij}) \right\|. \tag{F15}
\end{aligned}$$

Here, $\mathcal{C} \geq 1$. Thus, we obtain the last equality.

From the previous inequalities, we obtain $k_1 = 2^L \mathcal{C}$, which is a global property. Now, let us examine the covering number of $\sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij})$, where the trainable unitary operator $\hat{U}(\theta) = \prod_{h=1}^{N_g} \hat{u}_h(\theta_h)$ contains $N_{gt(ij)}$ trainable gates and $N_{fix(ij)} = N_{g(ij)} - N_{gt(ij)}$ fixed gates. (For the case of the Schmidt ansatz as given in the main text, please note that when there are additional quantum gates to the right of the CNOT in the lower half, we consider these gates redundant as they do not enhance our expressiveness. Therefore, we should remove these gates, and let the number of such gates be denoted as $N_{n(ij)}$. In this case, the number of trainable gates that affect our expressive power is $N_{gt(ij)} = N_{g(ij)} - N_{fix(ij)} - N_{n(ij)}$.) To achieve this goal, we consider a fixed ϵ -covering set $\mathcal{S}$ of the

set $\mathcal{N}(U(d^k), \epsilon, \|\cdot\|)$, and define the set as

$$\tilde{\mathcal{S}} := \left\{ \sum_{i,j=1}^K \prod_{h \in \{N_{gt}\}} \hat{u}_h(\theta_h) \prod_{k \in \{N_g - N_{gt}\}} \hat{u}_k \mid \hat{u}_h(\theta_h) \in \mathcal{S} \right\}, \tag{F16}$$

where $\hat{u}_h(\theta_h)$ and $\hat{u}_k$, respectively, specify the trainable and fixed quantum gates used in the ansatz. Please note that for any circuit $\hat{U}(\theta) = \prod_{i=1}^{N_g} \hat{u}_i(\theta_i)$, we can always find a $\hat{U}_\epsilon(\theta) \in \tilde{\mathcal{S}}$, where each trainable gate $\hat{u}_i(\theta_i)$ is replaced by the closest element in the covering set $\mathcal{S}$, and the difference satisfies

$$\begin{aligned}
&\left\| \sum_{i,j=1}^K \hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \sum_{i,j=1}^K \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij}) \right\| \\
&\leq \sum_{i,j=1}^K \|\hat{U}(\theta_{ij})^\dagger H_B \hat{U}(\theta_{ij}) - \hat{U}_\epsilon(\theta_{ij})^\dagger H_B \hat{U}_\epsilon(\theta_{ij})\| \\
&\leq \sum_{i,j=1}^K \|\hat{U}(\theta_{ij}) - \hat{U}_\epsilon(\theta_{ij})\| \|H_B\| \\
&\leq \sum_{i,j=1}^K N_{gt(ij)} \|H_B\| \epsilon = N_{GT} \|H_B\| \epsilon, \tag{F17}
\end{aligned}$$

where $N_{GT} = \sum_{i,j=1}^K N_{gt(ij)}$. The first two inequalities utilize the triangle inequality, and the third inequality stems from $\|\hat{U} - \hat{U}_\epsilon\| \leq N_{gt} \epsilon$. Therefore, by definition, we know that $\tilde{\mathcal{S}}$ is an $N_{GT} \|H_B\| \epsilon$ -covering set for $\mathcal{H}_{\text{circ}}$.

Recalling the upper bound given in Lemma F1, $|\mathcal{S}| \leq (\frac{7}{\epsilon})^{d^{2k}}$. Since there are $|\mathcal{S}|^{N_{GT}}$ possible topological combinations of gates in $\tilde{\mathcal{S}}$, we have $|\tilde{\mathcal{S}}| \leq (\frac{7}{\epsilon})^{d^{2k} N_{GT}}$, and the covering number of $\mathcal{H}_{\text{circ}}$ satisfies

$$\mathcal{N}(\mathcal{H}_{\text{circ}}, N_{GT} \|H_B\| \epsilon, \|\cdot\|) \leq \left(\frac{7}{\epsilon} \right)^{d^{2k} N_{GT}}. \tag{F18}$$

An equivalent expression of the above inequalities is

$$\mathcal{N}(\mathcal{H}_{\text{circ}}, \epsilon, \|\cdot\|) \leq \left(\frac{7 N_{GT} \|H_B\|}{\epsilon} \right)^{d^{2k} N_{GT}}. \tag{F19}$$

Therefore, by Lemma F1, we can conclude that the upper bound on the covering number under our QEM algorithm is

$$\mathcal{N}(\tilde{\mathcal{H}}, \epsilon, \|\cdot\|) \leq 2^L \mathcal{C} \left(\frac{7 N_{GT} \|H_B\|}{\epsilon} \right)^{d^{2k} N_{GT}}. \tag{F20}$$

In contrast, the cover number of traditional VQAs is [78]

$$\mathcal{N}(\mathcal{H}, \epsilon, \|\cdot\|) \leq \left(\frac{7 N_{gt} \|H_A\|}{\epsilon} \right)^{d^{2k} N_{gt}}. \tag{F21}$$

APPENDIX G: TRAINABILITY

The cost function used in this work is parametrized by $\{\bar{\theta}, \bar{c}\}$, with $\bar{\theta}$ the parameters that decide $\{\rho_1, \rho_2, \dots, \rho_k\}$. The derivatives with respect to $\bar{c}$ are trivial since they will not directly deal with the exponentially large Hilbert space and

thus have no barren plateau issues. For $\bar{\theta}$, since we have

$$\left| \frac{\partial \langle \psi | H_A | \psi \rangle}{\partial \theta} \right| \leq \sum_{i,j=1}^K \left| \frac{\partial C_{ij}}{\partial \theta} \right|, \quad (\text{G1})$$

with $C_{ij} = c_i^* c_j \frac{\text{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}$, and for each C_{ij} , we have

$$\begin{aligned} \left| \frac{\partial C_{ij}}{\partial \theta} \right| &\leq \left| \frac{c_i^* c_j}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)} \right| \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + \left| \frac{C_{ij}}{\sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)} \right| \left| \frac{\partial \sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\approx \left| \frac{c_i^* c_j}{\sum_k |c_k|^2 \text{Tr}(S \rho_k \otimes \rho_k)} \right| \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + \left| \frac{C_{ij}}{\sum_k |c_k|^2 \text{Tr}(S \rho_k \otimes \rho_k)} \right| \left| \frac{\partial \sum_{kl} c_k^* c_l \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq 2^L \left| \frac{c_i^* c_j}{|c_i|^2 + |c_j|^2} \right| \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^L |C_{ij}| \sum_{kl} \left| \frac{c_k^* c_l}{|c_k|^2 + |c_l|^2} \right| \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^{L+1} |C_{ij}| \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^{2L+1} \left| \frac{c_i^* c_j}{|c_i|^2 + |c_j|^2} \right| |\text{Tr}(H_B \rho_i \otimes \rho_j)| \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^{2L+2} |\text{Tr}(H_B \rho_i \otimes \rho_j)| \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\approx 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^{2L+2-n} |\langle \rho_i | H_A | \rho_j \rangle| \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| + 2^{2L+2-n} \|H_A\|_2 \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right|, \end{aligned} \quad (\text{G2})$$

where $|\rho\rangle$ is defined as $|\rho\rangle = \frac{1}{\sqrt{\sum_{ij} |\rho_{ij}|^2}} \sum_{ij} \rho_{ij} |i\rangle |j\rangle$. During the derivation, we assume the noiseless case where there are L CNOT gates connecting the upper and the lower system and use $\text{Tr}(\rho_i \rho_j) \approx \frac{1}{2^n}$. For noisy cases, one can use $e^{-2\zeta}$ to replace 2^L . In our settings, each θ only controls a single density matrix; thus, suppose θ is a parameter for ρ ; to further estimate the gradients, we need to consider two types of derivatives: $\left| \frac{\partial \text{Tr}(O \rho \otimes \sigma)}{\partial \theta} \right|$ and $\left| \frac{\partial \text{Tr}(O \rho \otimes \rho)}{\partial \theta} \right|$. If we have $O = \sum_{\alpha} o_{\alpha} P_{\alpha}$, with $P_{\alpha} = P_{\alpha, \text{up}} \otimes P_{\alpha, \text{low}}$ $2n$ -qubit Pauli operators, then we have

$$\begin{aligned} \left| \frac{\partial \text{Tr}(O \rho \otimes \sigma)}{\partial \theta} \right| &\leq \sum_{\alpha} |o_{\alpha}| \left| \frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right| |\text{Tr}(P_{\alpha, \text{low}} \sigma)| \\ &\leq \sum_{\alpha} |o_{\alpha}| \left| \frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right|, \end{aligned} \quad (\text{G3})$$

$$\begin{aligned} \left| \frac{\partial \text{Tr}(O \rho \otimes \rho)}{\partial \theta} \right| &\leq \sum_{\alpha} |o_{\alpha}| \left(\left| \frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right| |\text{Tr}(P_{\alpha, \text{low}} \rho)| \right. \\ &\quad \left. + \left| \frac{\partial \text{Tr}(P_{\alpha, \text{low}} \rho)}{\partial \theta} \right| |\text{Tr}(P_{\alpha, \text{up}} \rho)| \right) \\ &\leq \sum_{\alpha} |o_{\alpha}| \left(\left| \frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right| + \left| \frac{\partial \text{Tr}(P_{\alpha, \text{low}} \rho)}{\partial \theta} \right| \right), \end{aligned} \quad (\text{G4})$$

where during the derivation, we use the fact that $|\text{Tr}(P_{\alpha, \text{low}} \sigma)| \leq 1$ for arbitrary density matrix σ . For ρ , we can model it as

$$\begin{aligned} \rho &= \sum_{\alpha\beta} K_{\alpha} U(\theta) L_{\beta} |0\rangle \langle 0|^{\otimes n} L_{\beta}^{\dagger} U(\theta)^{\dagger} K_{\alpha}^{\dagger} \\ &= \sum_{\alpha} K_{\alpha} U(\theta) \rho_0 U(\theta)^{\dagger} K_{\alpha}^{\dagger}, \end{aligned} \quad (\text{G5})$$

with $\sum_{\alpha} K_{\alpha}^{\dagger} K_{\alpha} = \sum_{\beta} L_{\beta}^{\dagger} L_{\beta} = I$ and $\rho_L = \sum_{\beta} L_{\beta} |0\rangle \langle 0|^{\otimes n} L_{\beta}^{\dagger}$. This form can be understood as $U(\theta)$ sandwiched in between two CNOT gates that connect the upper and lower systems. The influence of these CNOT gates and the rest of the circuits can be formulated as quantum channels $\{K_{\alpha}\}$ and $\{L_{\beta}\}$. $\frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta}$ can have barren plateau issues from various resources. For example, if $U(\theta)$ forms a 2-design, then $E\left(\left(\frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta}\right)^2\right) \leq \eta^2$ with η^2 of order $\mathcal{O}(2^{-n})$. By the Markov's inequality, we have

$$\begin{aligned} \text{Prob} \left[\left| \frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right| \geq \varepsilon \right] &= \text{Prob} \left[\left(\frac{\partial \text{Tr}(P_{\alpha, \text{up}} \rho)}{\partial \theta} \right)^2 \geq \varepsilon^2 \right] \leq \frac{\eta^2}{\varepsilon^2}. \end{aligned} \quad (\text{G6})$$

Thus, $\frac{\partial \text{Tr}(P_{\alpha, \text{up(low)}} \rho)}{\partial \theta}$ will be concentrated to an exponentially small value, which leads to

$$\left| \frac{\partial \text{Tr}(O \rho \otimes \sigma)}{\partial \theta} \right| \lesssim \sum_{\alpha} |o_{\alpha}| \eta = \frac{\eta \|O\|_{1,1}}{2^n}, \quad (\text{G7})$$

$$\left| \frac{\partial \text{Tr}(O \rho \otimes \rho)}{\partial \theta} \right| \lesssim \sum_{\alpha} 2 |o_{\alpha}| \eta = \frac{2\eta \|O\|_{1,1}}{2^n}, \quad (\text{G8})$$

where $\|\cdot\|_{1,1}$ denotes the $L_{1,1}$ norm. For $2n$ -qubit SWAP operator S , we have $\|S\|_{1,1} = 2^n$ since the two-qubit SWAP operator has the decomposition $\frac{1}{2}(II + XX + YY + ZZ)$. For $H_B = \sum_{\alpha=1}^m g_{\alpha} Q_{\alpha}$, note that each Q_{α} is a tensor product of n basic two-qubit Pauli substitute operators that form four groups:

Group 1: $0.5II + 0.5XX + 0.5YY + 0.5ZZ$

$$\begin{aligned} &0.5II + 0.5XX - 0.5YY - 0.5ZZ \\ &-0.5II + 0.5XX - 0.5YY + 0.5ZZ \\ &0.5II - 0.5XX - 0.5YY + 0.5ZZ, \end{aligned}$$

Group 2: $0.5IX + 0.5XI + 0.5iYZ - 0.5iZY$

$$\begin{aligned} &0.5IX + 0.5XI - 0.5iYZ + 0.5iZY \\ &-0.5iIX + 0.5iXI + 0.5YZ + 0.5ZY \\ &-0.5iIX + 0.5iXI - 0.5YZ - 0.5ZY, \end{aligned}$$

Group 3: $-0.5iY + 0.5iXZ - 0.5YI - 0.5iZX$

$$\begin{aligned} &0.5iY + 0.5iXZ + 0.5YI - 0.5iZX \\ &0.5iY + 0.5XZ - 0.5iYI + 0.5ZX \\ &-0.5iY + 0.5XZ + 0.5iYI + 0.5ZX, \end{aligned}$$

Group 4: $0.5IZ + 0.5iXY - 0.5iYX + 0.5ZI$

$$\begin{aligned} &0.5IZ - 0.5iXY + 0.5iYX + 0.5ZI \\ &0.5iIZ - 0.5XY - 0.5YX - 0.5iZI \\ &0.5iIZ + 0.5XY + 0.5YX - 0.5iZI. \end{aligned} \quad (\text{G9})$$

For two such two-qubit operators B_1 and B_2 in different groups, since they share no common Pauli operators, we have $\text{Tr}(B_1^{\dagger} B_2) = 0$. When they are in the same group, the

coefficients before the Pauli operators assure the same result $\text{Tr}(B_1^{\dagger} B_2) = 0$. Since $\text{Tr}(Q_{\alpha}^{\dagger} Q_{\alpha}) = 4^n$, we have

$$\|H_B\|_{1,1} = 2^n \sum_{\alpha} |g_{\alpha}| = \|H_A\|_{1,1}. \quad (\text{G10})$$

Based on these discussions, Eq. (G2) can be further reduced to

$$\begin{aligned} \left| \frac{\partial C_{ij}}{\partial \theta} \right| &\leq 2^{L+1} \left| \frac{\partial \text{Tr}(H_B \rho_i \otimes \rho_j)}{\partial \theta} \right| \\ &+ 2^{2L+2-n} \|H_A\|_2 \sum_{kl} \left| \frac{\partial \text{Tr}(S \rho_k \otimes \rho_l)}{\partial \theta} \right| \\ &\leq \frac{2^{L+2} \eta \|H_A\|_{1,1}}{2^n} + \sum_{kl} \frac{2^{2L+3} \|H_A\|_2 \eta \|S\|_{1,1}}{2^{2n}} \\ &= \left(\frac{2^{L+2} \|H_A\|_{1,1}}{2^n} + \frac{2^{2L+3} K^2 \|H_A\|_2}{2^n} \right) \eta. \end{aligned} \quad (\text{G11})$$

By noticing $\|H_A\|_2 \leq \sum_{\alpha} |g_{\alpha}|$ and Eq. (G10), we have

$$\left| \frac{\partial C_{ij}}{\partial \theta} \right| \leq \left(2^{L+2} + \frac{2^{2L+3} K^2}{2^n} \right) \sum_{\alpha} |g_{\alpha}| \eta \approx 2^{L+2} \sum_{\alpha} |g_{\alpha}| \eta, \quad (\text{G12})$$

where in the last approximation, we assume $2L \ll n$ to limit the sampling complexity. Finally, the whole derivative is bounded by

$$\left| \frac{\partial \langle \psi | H_A | \psi \rangle}{\partial \theta} \right| \leq \sum_{i,j=1}^K \left| \frac{\partial C_{ij}}{\partial \theta} \right| \leq 2^{L+2} K^2 \sum_{\alpha} |g_{\alpha}| \eta. \quad (\text{G13})$$

Note that, generally, we have $\sum_{\alpha} |g_{\alpha}|$ of order $\mathcal{O}(1)$ and η^2 of order $\mathcal{O}(2^{-n})$; thus, when we require $2L \ll n$ and $U(\theta)$ forms a 2-design, the derivative will be exponentially small and thus barren plateau problems occur. Since Eq. (G13) is a rather loose bound, it cannot be used to estimate cases where L is comparable with n and cases where the existence of noise leads to exponentially small purities. For L is comparable with n , it is possible that the whole $2n$ -qubit unitary circuit forms a 2-design, which will then also lead to barren plateau problems. For other resources of barren plateaus, we refer to Refs. [36,85,86].

APPENDIX H: ROTATION UNITARIES FOR MEASUREMENT

In Tables II–V, we give detailed descriptions of all 16 two-qubit Pauli operators and their corresponding substitute operators. We also give concrete constructions of rotation circuits to make these substitute operators diagonal for direct measurements.

TABLE II. Set 1.

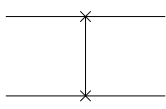
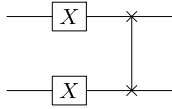
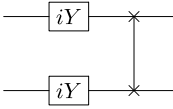
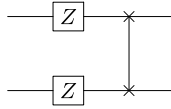
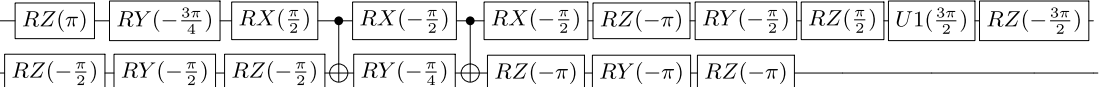
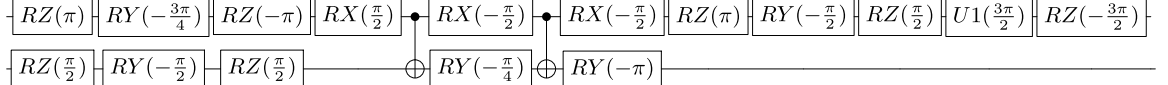
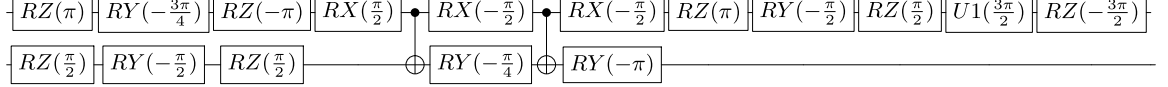
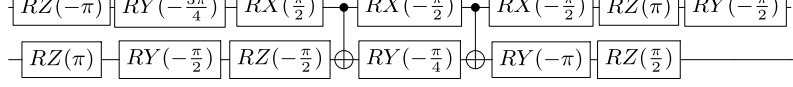
ID	P (Hermite)	Q (Unitary)	Eigenvalues of Q
1	II	$0.5II + 0.5XX + 0.5YY + 0.5ZZ$	$\text{diag}(1,1,1,-1)$
2	XX	$0.5II + 0.5XX - 0.5YY - 0.5ZZ$	$\text{diag}(1,1,-1,1)$
3	YY	$-0.5II + 0.5XX - 0.5YY + 0.5ZZ$	$\text{diag}(1,-1,-1,-1)$
4	ZZ	$0.5II - 0.5XX - 0.5YY + 0.5ZZ$	$\text{diag}(1,-1,1,1)$
Q matrix			
1	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$		
2	$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$		
3	$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$		
4	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$		
Q circuit			
1			
2			
3			
4			
Diagonalizing matrix			
1	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$		
2	$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \end{pmatrix}$		
3	$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \end{pmatrix}$		
4	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$		
Diagonalizing matrix circuit			
1			
2			
3			
4			

TABLE III. Set 2.

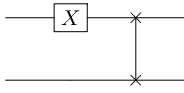
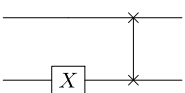
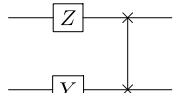
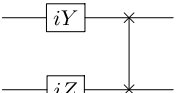
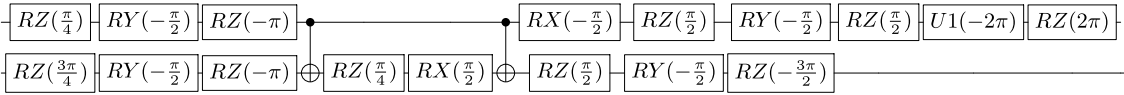
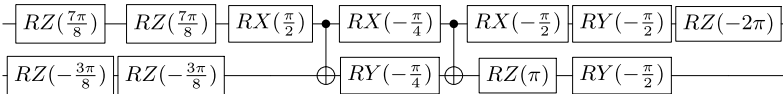
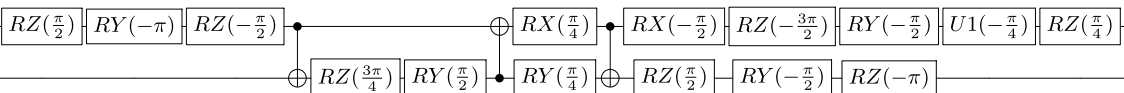
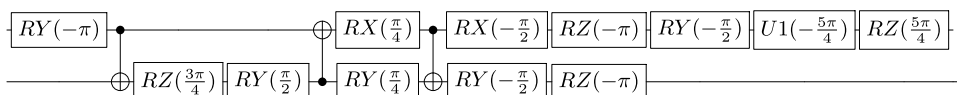
ID	P (Hermite)	Q (Unitary)	Eigenvalues of Q
5	IX	$0.5IX + 0.5XI + 0.5iYZ - 0.5iZY$	$\{-1, i, -i, 1\}$
6	XI	$0.5IX + 0.5XI - 0.5iYZ + 0.5iZY$	$\{-1, i, -i, 1\}$
7	YZ	$-0.5iIX + 0.5iXI + 0.5YZ + 0.5ZY$	$\{-1, i, -i, 1\}$
8	ZY	$-0.5iIX + 0.5iXI - 0.5YZ - 0.5ZY$	$\{-1, i, -i, 1\}$
Q matrix			
5	6		8
$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & -i & 0 & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & i & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & i & 0 & 0 \end{pmatrix}$
Q circuit			
5	6		8
			
Diagonalizing matrix			
5	6		8
$\begin{pmatrix} 0.5 & -0.5i & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5i & 0.5 \\ -0.5 & 0.5 & -0.5i & 0.5 \\ 0.5 & 0.5i & -0.5 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 & -0.5i & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5i & 0.5 \\ -0.5 & -0.5 & 0.5i & 0.5 \\ 0.5 & 0.5i & -0.5 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 & -0.5 & 0.5 & 0.5 \\ -0.5i & 0.5 & 0.5 & 0.5i \\ -0.5i & -0.5 & -0.5 & 0.5i \\ 0.5 & 0.5 & -0.5 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5i & -0.5 & 0.5 & -0.5i \\ 0.5i & 0.5 & -0.5 & -0.5i \\ 0.5 & -0.5 & -0.5 & 0.5 \end{pmatrix}$
Diagonalizing matrix circuit			
5			
			
6			
			
7			
			
8			
			

TABLE IV. Set 3.

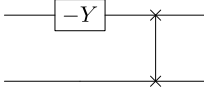
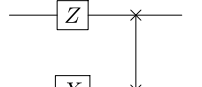
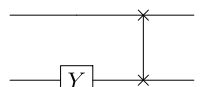
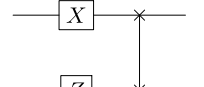
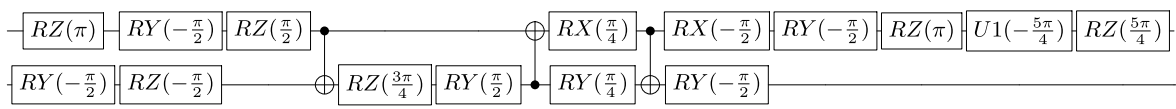
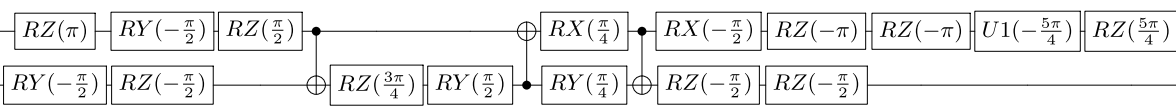
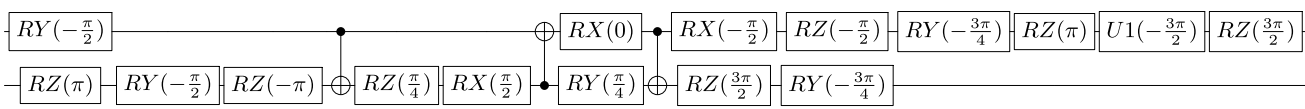
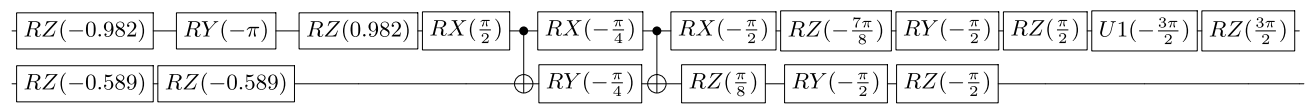
ID	P (Hermite)	Q (Unitary)	Eigenvalues of Q		
9	IY	$-0.5IY + 0.5iXZ - 0.5YI - 0.5iZX$	$\{-1, i, -i, 1\}$		
10	YI	$0.5IY + 0.5iXZ + 0.5YI - 0.5iZX$	$\{-1, i, -i, 1\}$		
11	XZ	$0.5iIY + 0.5XZ - 0.5iYI + 0.5ZX$	$\{-1, i, -i, 1\}$		
12	ZX	$-0.5iIY + 0.5XZ + 0.5iYI + 0.5ZX$	$\{-1, i, -i, 1\}$		
Q matrix					
9	$\begin{pmatrix} 0 & 0 & i & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & -i & 0 & 0 \end{pmatrix}$		12	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \end{pmatrix}$	
10	$\begin{pmatrix} 0 & -i & 0 & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & 0 & i & 0 \end{pmatrix}$		11	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}$	
Q circuit					
9			11		
10			12		
Diagonalizing matrix					
9	$\begin{pmatrix} 0.5 & 0.5 & 0.5 & -0.5 \\ 0.5i & -0.5 & 0.5 & 0.5i \\ 0.5i & 0.5 & -0.5 & 0.5i \\ -0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$		12	$\begin{pmatrix} 0.5 & -0.5i & 0.5 & -0.5 \\ -0.5 & -0.5 & 0.5i & -0.5 \\ -0.5 & 0.5 & -0.5i & -0.5 \\ -0.5 & -0.5i & 0.5 & 0.5 \end{pmatrix}$	
10	$\begin{pmatrix} 0.5 & -0.5 & 0.5 & -0.5 \\ -0.5i & 0.5 & 0.5 & -0.5i \\ -0.5i & -0.5 & -0.5 & -0.5i \\ -0.5 & -0.5 & 0.5 & 0.5 \end{pmatrix}$		11	$\begin{pmatrix} 0.5 & -0.5i & 0.5 & -0.5 \\ -0.5 & 0.5 & -0.5i & -0.5 \\ -0.5 & -0.5 & 0.5i & -0.5 \\ -0.5 & -0.5i & 0.5 & 0.5 \end{pmatrix}$	
Diagonalizing matrix circuit					
9					
10					
11					
12					

TABLE V. Set 4.

ID	P (Hermite)	Q (Unitary)	Eigenvalues of Q
13	IZ	$0.5IZ + 0.5iXY - 0.5iYX + 0.5ZI$	$\{i, -i, 1, -1\}$
14	ZI	$0.5IZ - 0.5iXY + 0.5iYX + 0.5ZI$	$\{i, -i, 1, -1\}$
15	XY	$0.5iIZ - 0.5XY - 0.5YX - 0.5iZI$	$\{1, -1, -i, i\}$
16	YX	$0.5iIZ + 0.5XY + 0.5YX - 0.5iZI$	$\{1, -1, -i, i\}$
Q matrix			
13	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$	14	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$
15	$\begin{pmatrix} 0 & 0 & 0 & i \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}$	16	$\begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ i & 0 & 0 & 0 \end{pmatrix}$
Q circuit			
13		14	
15		16	
Diagonalizing matrix			
13	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	14	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
15	$\begin{pmatrix} \frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}i & 0 & 0 \end{pmatrix}$	16	$\begin{pmatrix} -\frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}i & 0 & 0 \end{pmatrix}$
Diagonalizing matrix circuit			
13			
14			
15			
16			

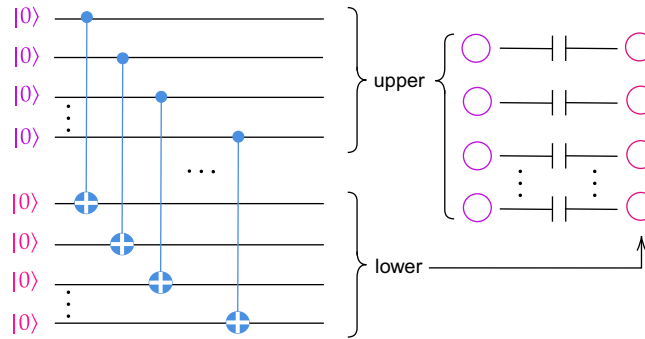


FIG. 13. Suitable hardware structure.

APPENDIX I: WELCOMED HARDWARE STRUCTURE

Since the circuit ansatz we use contains the qubits in the upper system and the qubits in the lower system, and we may always need CNOT gates connecting the upper and the lower system, one-dimensional quantum circuits will make these CNOT gates nonlocal. Thus, a more suitable architecture would be those with a ladder shape such that these CNOT gates are turned locally, as shown in Fig. 13.

-
- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
 - [2] R. P. Feynman, Simulating physics with computers, in *Feynman and Computation*, edited by A. J. G. Hey (CRC Press, Boca Raton, FL, 2018), pp. 133–153.
 - [3] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM Rev.* **41**, 303 (1999).
 - [4] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 1996), pp. 212–219.
 - [5] W. G. Unruh, Maintaining coherence in quantum computers, *Phys. Rev. A* **51**, 992 (1995).
 - [6] P. W. Shor, Fault-tolerant quantum computation, in *Proceedings of the 37th Conference on Foundations of Computer Science* (IEEE, 1996), pp. 56–65.
 - [7] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 1997), pp. 176–188.
 - [8] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
 - [9] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
 - [10] H. Bombin and M. A. Martin-Delgado, Topological quantum distillation, *Phys. Rev. Lett.* **97**, 180501 (2006).
 - [11] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
 - [12] D. Gottesman, Fault-tolerant quantum computation with constant overhead, *Quantum Inf. Comput.* **14**, 1338 (2014).
 - [13] D. Litinski, A game of surface codes: Large-scale quantum computing with lattice surgery, *Quantum* **3**, 128 (2019).
 - [14] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, *et al.*, Exponential suppression of bit or phase flip errors with repetitive error correction, *Nature (London)* **595**, 383 (2021).
 - [15] A. Erhard, H. P. Nautrup, M. Meth, L. Postler, R. Stricker, M. Stadler, V. Negnevitsky, M. Ringbauer, P. Schindler, H. J. Briegel, *et al.*, Entangling logical qubits with lattice surgery, *Nature (London)* **589**, 220 (2021).
 - [16] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, *Phys. Rev. X* **11**, 041058 (2021).
 - [17] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, *et al.*, Realizing repeated quantum error correction in a distance-three surface code, *Nature (London)* **605**, 669 (2022).
 - [18] Y. Zhao, Y. Ye, H.-L. Huang, Y. Zhang, D. Wu, H. Guan, Q. Zhu, Z. Wei, T. He, S. Cao, *et al.*, Realization of an error-correcting surface code with superconducting qubits, *Phys. Rev. Lett.* **129**, 030501 (2022).
 - [19] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, *Nature (London)* **614**, 676 (2023).
 - [20] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature (London)* **626**, 58 (2024).
 - [21] M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz, J. P. Gaebler, T. M. Gatterman, D. Hayes, N. Hewitt, *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates, *arXiv:2404.02280*.
 - [22] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).

- [23] A. Montanaro, Quantum algorithms: An overview, [npj Quantum Inf. 2, 15023 \(2016\)](#).
- [24] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. GSL Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, [Nature \(London\) 574, 505 \(2019\)](#).
- [25] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, Quantum computational advantage using photons, [Science 370, 1460 \(2020\)](#).
- [26] K. Bharti, A. Cervera-Lierta, T. Ha Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, Noisy intermediate-scale quantum algorithms, [Rev. Mod. Phys. 94, 015004 \(2022\)](#).
- [27] 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 \(2021\)](#).
- [28] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, [Nat. Commun. 5, 4213 \(2014\)](#).
- [29] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](#).
- [30] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, [Phys. Rev. Lett. 122, 040504 \(2019\)](#).
- [31] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, [Quantum 3, 191 \(2019\)](#).
- [32] D. S. França and R. Garcia-Patron, Limitations of optimization algorithms on noisy quantum devices, [Nat. Phys. 17, 1221 \(2021\)](#).
- [33] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, Limitations of variational quantum algorithms: A quantum optimal transport approach, [PRX Quantum 4, 010309 \(2023\)](#).
- [34] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, [Adv. Quantum Technol. 2, 1900070 \(2019\)](#).
- [35] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb-Robinson bounds and the generation of correlations and topological quantum order, [Phys. Rev. Lett. 97, 050401 \(2006\)](#).
- [36] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, [Nat. Commun. 12, 6961 \(2021\)](#).
- [37] S.-X. Zhang, Z.-Q. Wan, C.-K. Lee, C.-Y. Hsieh, S. Zhang, and H. Yao, Variational quantum-neural hybrid eigensolver, [Phys. Rev. Lett. 128, 120502 \(2022\)](#).
- [38] Z.-X. Shang, M.-C. Chen, X. Yuan, C.-Y. Lu, and J.-W. Pan, Schrödinger-Heisenberg variational quantum algorithms, [Phys. Rev. Lett. 131, 060406 \(2023\)](#).
- [39] J. Huang, W. He, Y. Zhang, Y. Wu, B. Wu, and X. Yuan, Tensor-network-assisted variational quantum algorithm, [Phys. Rev. A 108, 052407 \(2023\)](#).
- [40] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, [Phys. Rev. Lett. 119, 180509 \(2017\)](#).
- [41] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, [Phys. Rev. X 8, 031027 \(2018\)](#).
- [42] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, [Phys. Rev. X 7, 021050 \(2017\)](#).
- [43] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, [Phys. Rev. Lett. 122, 180501 \(2019\)](#).
- [44] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. De Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, [Phys. Rev. A 95, 042308 \(2017\)](#).
- [45] 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 \(2021\)](#).
- [46] Z. Cai, Quantum error mitigation using symmetry expansion, [Quantum 5, 548 \(2021\)](#).
- [47] Z. Liu, X. Zhang, Y.-Y. Fei, and Z. Cai, Virtual channel purification, [PRX Quantum 6, 020325 \(2025\)](#).
- [48] 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 \(2023\)](#).
- [49] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, [npj Quantum Inf. 8, 114 \(2022\)](#).
- [50] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, [Phys. Rev. Lett. 131, 210602 \(2023\)](#).
- [51] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, [Phys. Rev. Lett. 131, 210601 \(2023\)](#).
- [52] Y. Quek, D. S. França, S. Khatir, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, [Nat. Phys. 20, 1648 \(2024\)](#).
- [53] Z.-X. Shang, Z.-H. Chen, M.-C. Chen, C.-Y. Lu, and J.-W. Pan, A polynomial-time quantum algorithm for solving the ground states of a class of classically hard Hamiltonians, [arXiv:2401.13946](#).
- [54] N. Yoshioka, Y. O. Nakagawa, K. Mitarai, and K. Fujii, Variational quantum algorithm for nonequilibrium steady states, [Phys. Rev. Res. 2, 043289 \(2020\)](#).
- [55] Z.-X. Shang, Hermitian-preserving ansatz and variational open quantum eigensolver, [Phys. Rev. A 109, 062608 \(2024\)](#).
- [56] A. Gilchrist, D. R. Terno, and C. J. Wood, Vectorization of quantum operations and its use, [arXiv:0911.2539](#).
- [57] D. Aharonov, V. Jones, and Z. Landau, A polynomial quantum algorithm for approximating the Jones polynomial, in [Proceedings of the 38th Annual ACM Symposium on Theory of Computing](#) (Association for Computing Machinery, New York, NY, 2006), pp. 427–436.
- [58] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf, Quantum fingerprinting, [Phys. Rev. Lett. 87, 167902 \(2001\)](#).
- [59] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, [New J. Phys. 18, 023023 \(2016\)](#).
- [60] J. C. Garcia-Escartin and P. Chamorro-Posada, Swap test and Hong-Ou-Mandel effect are equivalent, [Phys. Rev. A 87, 052330 \(2013\)](#).
- [61] J. Preskill, [Lecture Notes for Physics 229: Quantum Information and Computation](#) (California Institute of Technology, Pasadena, CA, 1998), Vol. 16, pp. 1–8.
- [62] M. Müller-Lennert, F. Dupuis, O. Szechr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, [J. Math. Phys. 54, 122203 \(2013\)](#).

- [63] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, [Quantum](#) **8**, 1340 (2024).
- [64] B. Zeng, X. Chen, D.-L. Zhou, and X.-G. Wen, *Quantum Information Meets Quantum Matter: From Quantum Entanglement to Topological Phases of Many-Body Systems* (Springer, New York, 2019).
- [65] 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).
- [66] K. Bharti and T. Haug, Iterative quantum-assisted eigensolver, [Phys. Rev. A](#) **104**, L050401 (2021).
- [67] K. Bharti and T. Haug, Quantum-assisted simulator, [Phys. Rev. A](#) **104**, 042418 (2021).
- [68] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating large quantum circuits on a small quantum computer, [Phys. Rev. Lett.](#) **125**, 150504 (2020).
- [69] A. W. Harrow and A. Lowe, Optimal quantum circuit cuts with application to clustered Hamiltonian simulation, [PRX Quantum](#) **6**, 010316 (2025).
- [70] A. Bouland, T. Giurgică-Tiron, and J. Wright, The state hidden subgroup problem and an efficient algorithm for locating unentanglement, in *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 2025), pp. 463–470.
- [71] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, [Nat. Commun.](#) **9**, 4812 (2018).
- [72] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, England, 2013).
- [73] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, [Rev. Mod. Phys.](#) **92**, 015003 (2020).
- [74] J.-G. Liu, L. Mao, P. Zhang, and L. Wang, Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits, [Mach. Learn.: Sci. Technol.](#) **2**, 025011 (2021).
- [75] Z. Webb, The Clifford group forms a unitary 3-design, [Quantum Inf. Comput.](#) **16**, 1379 (2016).
- [76] V. Balasubramanian, M. DeCross, A. Kar, and O. Parrikar, Binding complexity and multiparty entanglement, [J. High Energy Phys.](#) **02** (2019) 069.
- [77] A. N. Kolmogorov and V. M. Tikhomirov, ε -entropy and ε -capacity of sets in function spaces, [Usp. Mat. Nauk](#) **14**, 3 (1959).
- [78] Y. Du, Z. Tu, X. Yuan, and D. Tao, Efficient measure for the expressivity of variational quantum algorithms, [Phys. Rev. Lett.](#) **128**, 080506 (2022).
- [79] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, [PRX Quantum](#) **3**, 010313 (2022).
- [80] Y. Yan, Z. Du, J. Chen, and X. Ma, Limitations of noisy quantum devices in computational and entangling power, [npj Quantum Inf.](#) **11**, 188 (2025).
- [81] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, [Phys. Rev. A](#) **103**, 042605 (2021).
- [82] A. Cross, The IBM Q experience and Qiskit open-source quantum computing software, in *APS March Meeting Abstracts*, Vol. 2018 (American Physical Society, College Park, MD, 2018), pp. L58–003.
- [83] Z. Chen, DMVQE-h, 4-molecules, GitHub, 2024, https://github.com/Zihan-Chen-PhMA/dmvqe_H4_molecule.
- [84] T. Barthel and J. Lu, Fundamental limitations for measurements in quantum many-body systems, [Phys. Rev. Lett.](#) **121**, 080406 (2018).
- [85] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, [Nat. Commun.](#) **12**, 1791 (2021).
- [86] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A unified theory of barren plateaus for deep parametrized quantum circuits, [Nat. Commun.](#) **15**, 7172 (2024).