

Expectation value estimation with parametrized quantum circuits

Bujiao Wu,^{1,*} Lingyu Kong,^{1,2} Yuxuan Yan,³ Fuchuan Wei,⁴ and Zhenhuan Liu^{3,†}

¹*International Quantum Academy, Shenzhen 518048, China*

²*Shenzhen Institute for Quantum Science and Engineering,
Southern University of Science and Technology, Shenzhen 518055, China*

³*Center for Quantum Information, Institute for Interdisciplinary
Information Sciences, Tsinghua University, Beijing 100084, China*

⁴*Yau Mathematical Sciences Center and Department of Mathematics, Tsinghua University, Beijing 100084, China*

Estimating properties of quantum states, such as fidelities, molecular energies, and correlation functions, is a fundamental task in quantum information science. Due to the limitation of practical quantum devices, including limited circuit depth and connectivity, estimating even linear properties encounters high sample complexity. To address this inefficiency, we propose a framework that optimizes sample complexity for estimating the expectation value of any observable using a shallow parameterized quantum circuit. Within this framework, we introduce two decomposition algorithms, a tensor network approach and a greedy projection approach that decompose the target observable into a linear combination of multiple observables, each of which can be diagonalized with the shallow circuit. Using this decomposition, we then apply an importance sampling algorithm to estimate the expectation value of the target observable. We numerically demonstrate the performance of our algorithm by estimating the expectation values of some specific Hamiltonians and inner product of a Slater determinant with a pure state, highlighting advantages compared to some conventional methods. Additionally, we derive the fundamental lower bound for the sample complexity required to estimate a target observable using a given shallow quantum circuit, thereby enhancing our understanding of the capabilities of shallow circuits in quantum learning tasks.

* bujiaowu@gmail.com

† qubithuan@gmail.com

I. INTRODUCTION

Learning the properties of an unknown quantum state is crucial across various fields, such as many-body physics, quantum chemistry, and quantum information science [1–6]. Among these properties, the linear properties $\text{Tr}(\rho H)$, where H represents an observable, are of special importance since quantum mechanics is inherently linear. Therefore, various protocols have been developed to facilitate the estimation of linear properties, ranging from conventional tomography-based protocols [7–9] and Pauli decomposition methods [10–20] to quantum memory-assisted protocols [21, 22]. Moreover, understanding the fundamental challenges in estimating linear properties can deepen our comprehension of quantum learning problems [22–25].

Among all the estimation protocols, randomized measurement [26] has gradually become a prevailing toolbox due to its experimental accessibility [27–31] and efficiency in estimating certain state properties [32–39]. Notably, Huang et al. introduced the classical shadow (CS) protocol [40], which aims to provide a classical description of an unknown state ρ and use it to calculate arbitrary state properties. Conducted in a “measure first, ask questions later” manner [26], the CS protocol does not require the information of observables during the quantum experiment and can efficiently estimate a large number of independent observables simultaneously.

Depending on the choice of unitary ensemble, the original classical shadow (CS) protocol can be categorized into two types: local CS and global CS, which sample unitaries from the product of single-qubit Clifford groups and the global n -qubit Clifford group, respectively. The local CS protocol is efficient for estimating k -local observables, where k scales logarithmically with the number of qubits n . For general k -local observables, the sample complexity is $\mathcal{O}(4^k)$, and for k -fold tensor products of single-qubit observables, it improves to $\mathcal{O}(3^k)$. In contrast, the global CS protocol achieves $\mathcal{O}(1)$ sample complexity even for certain nonlocal observables, but requires implementing quantum circuits of logarithmic depth [41], which is still challenging for near-term devices. To make the CS protocol more suitable for noisy intermediate-scale quantum devices, a random shallow circuit-based CS protocol was developed [30, 41–44]. With a relatively low circuit depth, the shallow CS protocol outperforms the local CS protocols for certain nonlocal observables [41, 43, 45]. However, all these CS protocols are designed to be measurement-agnostic; they do not exploit prior knowledge about the target observables. As a result, they become suboptimal in scenarios where only a small number of observables need to be estimated. This limitation motivates approaches that incorporate knowledge of the observables to improve efficiency.

In addition to the CS protocol, several methodologies have been developed to improve the efficiency of estimating the expectation values of specific observables with respect to certain quantum states [10–20]. Given the Pauli decomposition $H = \sum_{i=1}^m \alpha_i Q_i$, these methods aim to efficiently estimate the expectation values for all $\{Q_i\}_{i=1}^m$. This is achieved by performing Clifford circuits sampled from a specific set \mathcal{S} with certain probabilities, followed by measurement in the computational basis and classical post-processing. The gate set \mathcal{S} and these probabilities are carefully designed and vary depending on the specific algorithms. While these methods improve performance compared to the original CS protocol by utilizing information about target observables, they do not consider the practical constraints of quantum circuits. Firstly, implementing some target Clifford gates may require deep circuits [10, 20]. Additionally, since the observable decomposition is restricted to Pauli observables, these methods may still require high sample complexity even with deep circuits.

In this work, we propose a framework for estimating linear properties that unifies the logic of the methodologies mentioned earlier. Specifically, given an unknown quantum state and a variational circuit whose structure is determined by experimental settings, our framework aims to use the information of the target observable to optimize the usage of the parameterized circuit and achieve low sample complexity. Our approach is inspired by the observation that, given an observable H , if we can perform any unitary operation on the quantum state ρ , $\mathcal{O}(\|H\|_2^2 \epsilon^{-2})$ copies of ρ are sufficient to approximate $\text{Tr}(\rho H)$ with the additive error of ϵ and high success probability, where $\|\cdot\|_2$ denotes the spectral norm. This requirement arises from the eigen-decomposition of H , given by $H = U^\dagger \Lambda U$. Thus, we can rotate the quantum state using the unitary U and measure it in the computational basis to estimate $\text{Tr}(\rho H)$. In practical scenarios, where we are limited to L -depth parameterized unitaries $U_L(\theta)$ which are tomography-complete, we can decompose the target observable as $H \approx \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$, with Λ_k being real diagonal matrices. Each term in the summation is an observable that can be diagonalized with the parameterized circuit to efficiently estimate $\text{Tr}(\Lambda_k U_L(\theta^{(k)}) \rho U_L(\theta^{(k)})^\dagger)$.

Based on these observations, our framework can be divided into classical and quantum phases, as illustrated in Fig. 1. In the classical phase, we use a greedy projection or tensor network optimization algorithm to decompose the target observable. In the quantum phase, we design an efficient method to estimate the expectation value of the target observable based on the given decomposition. Many existing Pauli decomposition protocols, such as qubit-commuting measurement approaches [12, 14, 17] and commuting measurement approaches [10, 20], fall within the scope of our framework. The required number of copies of ρ is $\mathcal{O}\left(\frac{\|\Lambda\|_1^2 \log(1/\delta)}{\epsilon^2}\right)$ with additive error ϵ and success probability $1 - \delta$, where Λ is a K -dimensional vector with the k -th entry being $\Lambda(k) = \|\Lambda_k\|_2$, and $\|\Lambda\|_1$ denotes the sum of the absolute

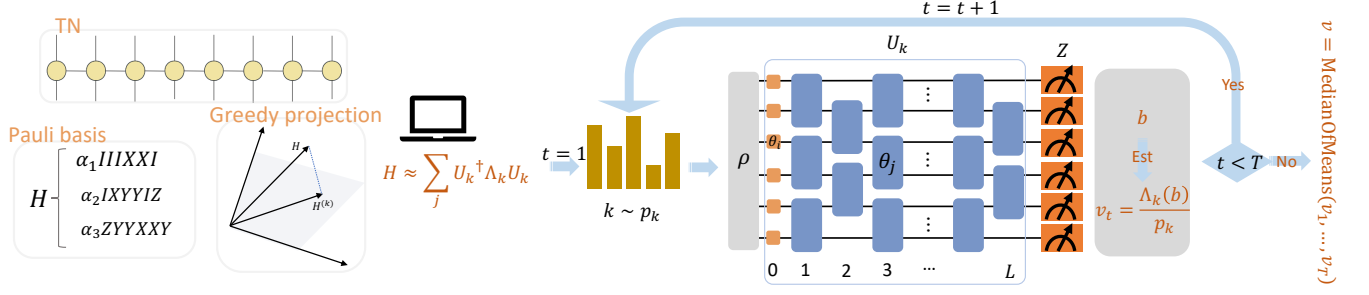


FIG. 1: Overview of the expectation value estimation with parameterized circuit. We utilize a classical algorithm (greedy projection or tensor network method) to generate a decomposition of $H \approx \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$, followed by iteratively repeating T rounds: sampling k with probability $p_k \propto \|\Lambda_k\|_2$, performing $U_L(\theta^{(k)})$ on the prepared quantum state ρ , as illustrated by the L -depth circuit consisting of two-qubit parametrized gates. Afterward, we measure in the computational basis to obtain a result b , which is then used to compute the estimator v by applying the Median-of-Means method. Here, $\Lambda_k(b)$ denotes the b -th diagonal entry of the diagonal matrix Λ_k and b ranges from 0 to $2^n - 1$.

values of all the entries in Λ .

To verify the convergence of our observable decomposition algorithm, we numerically evaluate the distance between the approximate observable $\sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$ and the target observable across various circuit depths L and the number of decomposition terms K . The numerical results indicate that for a fixed number of decomposition terms, this distance decreases as the depth of the ansatz circuits increases, which aligns with our intuition. Moreover, the Frobenius distance decays exponentially as the number of decomposition terms increases when fixing the circuit depth. We further demonstrate the practical efficiency of our algorithms by applying them to estimate the ground energy of a sparse Hamiltonian and the inner product of two pure quantum states, comparing the results with conventional methods.

In addition to providing a practical measurement framework, we also aim to address the fundamental question: *What is the limitation of a shallow quantum circuit in estimating an observable?* For a given variational quantum circuit $U_L(\theta)$, we establish a lower bound on the number of copies of the quantum state required to accurately estimate $\text{Tr}(\rho H)$. This bound is given by $\Omega\left(\frac{\text{Tr}(H_0^2)^2}{\epsilon^2 \delta(H_0)^{4n}}\right)$, where H_0 represents the traceless part of H , and $\delta(H_0)$ is the square of the maximum expectation value of H_0 with respect to the quantum states prepared by applying $U_L(\theta)^\dagger$ to a product state. We also analyze the tightness of this lower bound in specific cases.

II. EXPECTATION VALUE ESTIMATION FRAMEWORK

Our framework can be roughly divided into two phases. The classical phase relies on a specific decomposition algorithm to decompose the target observable into several terms. In the quantum phase, we estimate the expectation value based on the decomposition of the target observable and some probability distribution. We summarize the framework using Algorithm 1, in which $\|\cdot\|_2$ denotes the spectral norm, $\|A\|_2 = \max_j \sigma_j$ with σ_j being the singular value of matrix A , $\|A\|_F = \sqrt{\text{Tr}(AA^\dagger)}$ denotes the Frobenius norm of matrix A , Λ represents a K -dimensional vector with the k -th entry being $\Lambda(k) = \|\Lambda_k\|_2$, and $\|\Lambda\|_1$ denotes the sum of the absolute values of all the entries in Λ . The b -th diagonal entry of Λ_k is denoted by $\Lambda_k(b)$, where $b \in \{0, 1\}^n$ is interpreted as an integer index in the computational basis.

Proposition 1. Let $\hat{H} = \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$ be an ϵ_1 approximation of H in spectral norm. Then $T = \mathcal{O}\left(\frac{\|\Lambda\|_1^2 \log(1/\delta)}{\epsilon_2^2}\right)$ copies of ρ are required to estimate $\text{Tr}(\rho H)$ to additive error $\epsilon = \epsilon_1 + \epsilon_2$ with success probability of at least $1 - \delta$. Here, Λ is a K -dimensional vector with the k -th entry being $\|\Lambda_k\|_2$.

Proof. Suppose there exists a set $\{\theta^{(k)}, \Lambda_k\}_{k=1}^K$ such that $\|H - \hat{H}\|_2 \leq \epsilon_1$ with $\hat{H} = \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$, then

$$\left| \text{Tr}(\rho \hat{H}) - \text{Tr}(\rho H) \right| \leq \left\| \hat{H} - H \right\|_2 \leq \epsilon_1, \quad (1)$$

where the first inequality holds since $\text{Tr}(\rho A) \leq \|A\|_2$ for any matrix A .

Algorithm 1: Framework for Estimating Linear Properties of Quantum States.

input : Quantum state ρ , observable H , error ϵ_1 , the number of samplings T .
output: Estimator v of $\text{Tr}(\rho H)$ with error $\epsilon = \epsilon_1 + \epsilon_2$ and success probability $1 - \delta$.

- 1 *Generate unitaries $\{U_k\}_{k=1}^K$ and real diagonal matrices $\{\Lambda_k\}_{k=1}^K$, where the b -th diagonal entry of Λ_k is denoted by $\Lambda_k(b)$, such that $\left\|H - \sum_{k=1}^K U_k^\dagger \Lambda_k U_k\right\|_2 \leq \epsilon_1$;*
- 2 *Set probabilities $p_k \propto \|\Lambda_k\|_2$;*
- 3 **for** $t \leftarrow 1$ **to** T **do**
- 4 *Sample an index k with probability p_k ;*
- 5 *Perform U_k to the prepared state ρ , then measure in computational basis to obtain the n -bit measurement result b_t ;*
- 6 *Let the estimator $v_t = \frac{\Lambda_k(b_t)}{p_k}$;*
- 7 **end**
- 8 **return** $v := \text{MedianOfMeans}(v_1, \dots, v_t)$

It remains to show that the estimation v generated by Algorithm 1 in combination with Algorithm 2 and important sampling is an unbiased estimation of $\text{Tr}(\rho \hat{H})$ and satisfies

$$\Pr \left[\left| v - \text{Tr}(\rho \hat{H}) \right| \leq \epsilon_2 \right] \geq 1 - \delta \quad (2)$$

where δ represents the failure probability. Combining inequalities (1) and (2), we obtain

$$|v - \text{Tr}(\rho H)| \leq \left| v - \text{Tr}(\rho \hat{H}) \right| + \left| \text{Tr}(\rho \hat{H}) - \text{Tr}(\rho H) \right| \leq \epsilon_1 + \epsilon_2, \quad (3)$$

with a success probability of $1 - \delta$. Let $v_t = \frac{\Lambda_k(b_t)}{p_k}$ be an estimation generated from a single sample, as shown in Step (6) of Algorithm 1. The expectation of v_t can then be expressed as

$$\mathbb{E}[v_t] = \sum_{k=1}^K \frac{1}{p_k} p_k \mathbb{E}[\Lambda_k(b_t)] \quad (4)$$

$$= \sum_{k=1}^K \Pr[b] \Lambda_k(b) \quad (5)$$

$$= \sum_{k=1}^K \text{Tr} \left(U_L(\theta^{(k)}) \rho U_L(\theta^{(k)})^\dagger \Lambda_k \right) \quad (6)$$

$$= \text{Tr}(\rho \hat{H}). \quad (7)$$

Then $\mathbb{E}[v] = \mathbb{E}[v_t] = \text{Tr}(\rho \hat{H})$. Next, we will bound the variance of v_t by noting that

$$\text{Var}(v_t) \leq \mathbb{E}[v_t^2] \quad (8)$$

$$= \sum_{k=1}^K \frac{p_k}{p_k^2} \text{Tr} \left(\rho U_L(\theta^{(k)})^\dagger \Lambda_k^2 U_L(\theta^{(k)}) \right) \quad (9)$$

$$\leq \sum_{k=1}^K \frac{1}{p_k} \|\Lambda_k\|_2^2 \quad (10)$$

$$= \sum_{k=1}^K \frac{\sum_{j=1}^K \|\Lambda_j\|_2}{\|\Lambda_k\|_2} \|\Lambda_k\|_2^2 \quad (11)$$

$$= \|\mathbf{\Lambda}\|_1^2 \quad (12)$$

where $\mathbf{\Lambda}$ is a K -dimensional vector with the k -th entry being $\mathbf{\Lambda}(k) = \|\Lambda_k\|_2$, and $\Lambda_k(b)$ denotes the b -th diagonal entry of the diagonal matrix Λ_k . According to the median-of-means method in Ref. [46], with $T = \mathcal{O} \left(\frac{\|\mathbf{\Lambda}\|_1^2 \log(1/\delta)}{\epsilon_2^2} \right)$ samplings, the estimation error for v can be bounded to $\epsilon_1 + \epsilon_2$ in total with a success probability $1 - \delta$. \square

The central process of the classical stage is generating the decomposition $\{U_k, \Lambda_k\}_k$ of the target observable. In fact, many Pauli decomposition-based quantum measurement protocols offer methodologies for generating such decompositions. Below, we will illustrate two of these methods: one based on qubit-commuting Pauli decomposition and the other on commuting Pauli decomposition.

- (1) **Qubit-commuting decomposition.** We call Pauli operators P and Q qubit-commuting if, for any l -th qubit, $[P_l, Q_l] = 0$. Since the 4^n Pauli operators form a basis for the space of $2^n \times 2^n$ Hermitian matrices, any observable H can be expressed as $H = \sum_{i=1}^m \alpha_i Q_i$, where $\{\alpha_i\}_i$ are real coefficients and $\{Q_i\}_i$ are Pauli operators. Since qubit-commuting Pauli observables can be estimated simultaneously, we can group the Pauli operators Q_1, \dots, Q_m into K qubit-commuting (overlapping) sets e_1, e_2, \dots, e_K , where each set contains Pauli operators that commute with each other. Various grouping methods have been proposed, such as the largest degree first algorithm [11] and the overlapping grouping algorithm [17]. Given a grouping algorithm, we can express the target observable as $H = \sum_{k=1}^K U_k^\dagger \Lambda_k U_k$, where each U_k is the tensor product of local Clifford gates associated with the eigenstates of Pauli operators in the set e_k , and the diagonal matrix Λ_k contains information about the eigenvalues of the Pauli operators in e_k . Further details on the explicit representations of U_k and Λ_k are provided in Appendix B.
- (2) **Commuting decomposition.** The qubit-commuting sets can be relaxed to the fully commuting sets $\{g_k\}_k$, where any two Pauli operators in g_k commute with each other. Consequently, the target observable can be expressed as $H = \sum_{k=1}^K H_k$, with $H_k = \sum_{Q_j \in g_k} \alpha_j Q_j$. Similarly, there exists a set of Clifford gates $\{U_k\}_k$, which may require a polynomial-size quantum circuit to implement, and diagonal matrices $\{\Lambda_k\}_k$ such that $H_k = U_k^\dagger \Lambda_k U_k$ [10] for all k . Details on how to generate $\{U_k, \Lambda_k\}_k$ are provided in Appendix C.

We note that commuting decomposition may require polynomial-depth quantum circuits, making it unsuitable for near-term quantum devices. Nevertheless, its inclusion highlights the generality of our framework, which accommodates a wide range of measurement schemes. There are also alternative methods that leverage a more general basis. For instance, low-rank approximations are often applied to simplify electron-electron repulsion terms, reducing the required number of decomposition terms. Techniques such as density matrix embedding theory and density matrix renormalization group approximate the Hamiltonian by isolating essential degrees of freedom, thereby minimizing the number of terms necessary for an accurate representation [47, 48]. In the following sections, we present two approaches for Hamiltonian approximation: (1) greedy projected decomposition (GPD) and (2) tensor network decomposition (TND).

The central objective of the quantum stage of our framework is to determine a probability distribution $\{p_k\}_k$ used to sample the estimated observables. In Pauli decomposition-based measurement protocols, this probability distribution corresponds to the one used for sampling different Pauli groups. Wu et al. [17] demonstrated that all existing qubit-commuting protocols can be unified through two central processes: Pauli grouping and assigning each group a probability, which fits precisely within our framework. Similarly, we show in Appendix C that existing commuting measurement schemes [10, 20] also fit within this framework. It is important to note that Pauli decomposition-based protocols are limited to Clifford unitary evolution, which cannot achieve the optimal performance for a general observable even with arbitrary Clifford unitaries. Therefore, our goal is to design new algorithms that are compatible with general quantum circuits.

III. GREEDY PROJECTION DECOMPOSITION OF THE HAMILTONIAN

Based on the introduced framework, we design specific algorithms to generate the decomposition of the target observable and estimate the expectation value, as illustrated in Fig. 1. We consider a practical scenario, where the depth and structure of the quantum circuit are given according to the capabilities of the available quantum devices and can go beyond Clifford circuits. Without loss of generality, we hereafter take a typical circuit structure: an L -depth parameterized quantum circuit $U_L(\theta)$ with the brick-like arrangement of 2-qubit parametrized gates, to demonstrate our algorithms. The 0-layer of the ansatz circuit comprises a layer of parametrized single-qubit gates. This parameterized circuit is tomography-complete for any $L \geq 0$ as it can estimate all Pauli observables, even when $L = 0$. Our conclusions can be easily generalized to other circuit structures.

We propose a greedy algorithm to decompose the target observable H using the L -depth parameterized circuit $U_L(\theta)$. In each step, we aim to find an observable, $U_L(\theta)^\dagger \Lambda U_L(\theta)$, that approximates H . This involves finding a unitary $U_L(\theta)$ and a real diagonal matrix Λ that minimize the distance between $U_L(\theta) H U_L(\theta)^\dagger$ and Λ . When using the Frobenius norm to quantify the distance, the optimal choice of Λ is $\Lambda = \text{diag}[U_L(\theta) H U_L(\theta)^\dagger]$, where $\text{diag}(X)$ denotes a diagonal matrix whose elements are the diagonal elements of X . As a comparison, the distance based on

the 2-norm brings us an additional optimization step to determine the optimal Λ . This iterative process updates the Hamiltonian H as outlined in Algorithm 2.

Algorithm 2: Greedy Projected Decomposition (GPD) Algorithm.

input : Observable H , L -depth parameterized quantum circuit $U_L(\theta)$, accuracy ϵ .
output: Set $\{\theta^{(k)}, \Lambda_k\}_k$ such that $\|H - \sum_k U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)\|_2 \leq \epsilon$.

- 1 $H^{(0)} = H$;
- 2 $k = 0$;
- 3 **while** $\|H^{(k)}\|_2 \geq \epsilon$ **do**
- 4 $\theta^{(k)} = \arg \min_{\theta} \|U_L(\theta) H^{(k)} U_L^\dagger(\theta) - \text{diag}[U_L(\theta) H^{(k)} U_L^\dagger(\theta)]\|_F$;
- 5 $\Lambda_k = \text{diag}[U_L(\theta^{(k)}) H^{(k)} U_L^\dagger(\theta^{(k)})]$;
- 6 $H^{(k+1)} = H^{(k)} - U_L^\dagger(\theta^{(k)}) \Lambda_k U_L(\theta^{(k)})$;
- 7 $k = k + 1$;
- 8 **end**
- 9 **return** $\{\theta^{(k)}, \Lambda_k\}$

Using the decomposition generated by Algorithm 2, we claim that Algorithm 1 can produce an estimator of $\text{Tr}(\rho H)$ with bounded error.

Proposition 2. *Given that L -depth unitary set $\{U_L(\theta)\}$ is tomography-complete and integer $L \geq 0, \epsilon \in (0, 1)$, then any Hermitian matrix H can be approximated as $\hat{H} = \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$ such that $\|\hat{H} - H\| \leq \epsilon$, for some integer K and diagonal matrix Λ_k .*

Proof. Below, we will prove that Algorithm 2 can be conclusively terminated within a finite number of steps, denoted as K . Considering one step of parameterized searching, the goal is to minimize the distance $\|H - U_L(\theta)^\dagger \Lambda U_L(\theta)\|_F = \|U_L(\theta) H U_L(\theta)^\dagger - \Lambda\|_F$. Since Λ is a real diagonal matrix, the optimal choice is $\Lambda = \text{diag}[U_L(\theta) H U_L(\theta)^\dagger]$. This choice is optimal because the Frobenius norm quantifies the sum of the squares of all the matrix elements, and choosing Λ in this way eliminates all the diagonal terms of $U_L(\theta) H U_L(\theta)^\dagger$. Furthermore, with this choice of Λ , one can prove that $\|H - U_L(\theta)^\dagger \Lambda U_L(\theta)\|_F = \|U_L(\theta) H U_L(\theta)^\dagger - \Lambda\|_F \leq \|U_L(\theta) H U_L(\theta)^\dagger\|_F = \|H\|_F$. This means that when the optimization algorithm functions ideally, a better approximation of H can always be achieved by increasing the iteration steps.

Furthermore, we can prove that it is always possible to find a θ such that $\|H - U_L(\theta)^\dagger \Lambda U_L(\theta)\|_F < \|H\|_F$, provided our parameterized circuit is tomography-complete. If this is not the case, then $\text{diag}[U_L(\theta) H U_L(\theta)^\dagger] = 0$ must hold for all choices of θ . If $H \neq 0$, this implies the existence of a non-zero observable that cannot be estimated using our parameterized circuits, which contradicts our assumption. We can further prove that if θ is the optimal value we can find, then $\|H\|_F - \|H - U_L(\theta)^\dagger \Lambda U_L(\theta)\|_F \geq \frac{1}{2 \times 4^n} \|H\|_F$. This is because we assume that $U_{L=0}(\theta)$ can realize the tensor product of arbitrary single-qubit gates, thereby enabling the estimation of arbitrary Pauli observables. Given the Pauli decomposition of the observable $H = \sum_i \alpha_i P_i$, one can at least choose $U_L(\theta)^\dagger \Lambda U_L(\theta) = \alpha_j P_j$, where α_j is the largest coefficient. In this case, $\|H\|_F - \|H - \alpha_j P_j\|_F \geq \frac{1}{2 \times 4^n} \|H\|_F$. Considering that the greedy algorithm can achieve better performance than simply picking the largest Pauli term, we can conclude that our algorithm can terminate in a finite number of steps. \square

In Step 4 of Algorithm 2, the classical processing time required is $\mathcal{O}(\text{poly}(n) \cdot 2^{\omega n})$, where $\omega = 2.371552 \dots$ represents the matrix multiplication exponent [49]. This complexity is an improvement over the direct eigen-decomposition time of $\mathcal{O}(2^{3n})$ [50]. Additionally, the matrix multiplication process can be further optimized with a quantum-inspired algorithm [51], allowing for a $\text{poly}(n)$ -rank approximation to be performed in $\text{poly}(n)$ time.

Here, the total error ϵ depends on both the approximation error of H and the estimation error of $\text{Tr}(\rho \hat{H})$, where $\hat{H} = \sum_{k=1}^K U_L(\theta^{(k)})^\dagger \Lambda_k U_L(\theta^{(k)})$. To minimize the resource consumption, we aim to minimize the value of K while keeping a small approximation error. Due to the greedy nature of Algorithm 2, theoretical analysis of the number of decomposed terms K is challenging. However, numerical results indicate that the number of decomposed terms

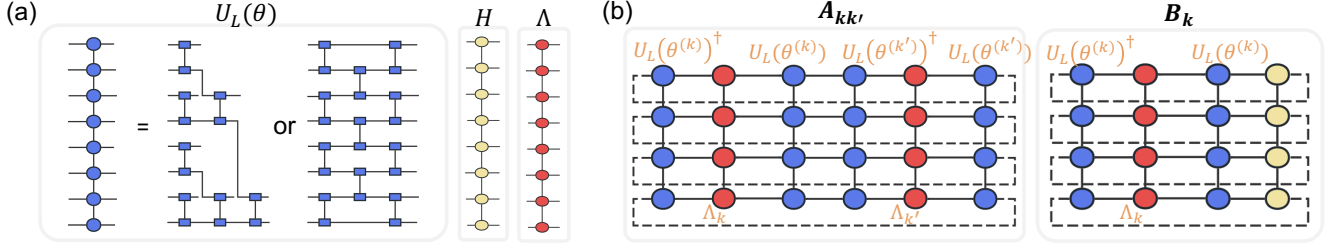


FIG. 2: Illustration of tensor network decomposition for a given Hamiltonian. (a) Tensor network representation of the parametrized unitary $U(\theta_k)$, the diagonal matrix Λ_k , and the Hamiltonian H . (b) Tensor network representation of $A_{kk'}$ and B_k , which are associated with the loss function calculations.

generated by our algorithm decreases as the depth L increases. In the numerical section, we demonstrate that our algorithm, based on parameterized circuits, requires significantly fewer terms (K) to approximate a sparse and dense Hamiltonian within a given error threshold, compared to the number of terms needed in the standard qubit-wise Pauli decomposition approach.

IV. TENSOR NETWORK DECOMPOSITION OF THE HAMILTONIAN

Although the greedy decomposition method works generally for all parameterized circuits, it can be further optimized in some practical cases. When the depth of the parameterized circuit is low, a tensor network-based approach can be employed to approximate the Hamiltonian as $\hat{H} = \sum_{k=1}^K U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)$. Each $U_L(\theta_k)$ corresponds to a unitary tensor network of depth L , such as a unitary matrix product operator or a unitary tree tensor network, constructed following [52–54], as illustrated in Fig. 2(a). The bond dimension is $O(nr_U)$. The diagonal matrices Λ_k are represented as matrix product operators (MPOs), with component-wise form given by $\Lambda_{i_1, \dots, i_n, l_1, \dots, l_n} = \bigotimes_{j=1}^n \Lambda_{j, i_j, l_j, u_{j-1}, u_j}$, where $u_j \in [\beta]$ for $1 \leq j \leq n-1$, and the boundary conditions are $u_0 = 1$, $u_n = 1$. The indices $i_j, l_j \in \{0, 1\}$ denote the input and output basis states, respectively, as depicted in Fig. 2(a). The target Hamiltonian H is also expressed in MPO form.

To approximate the Hamiltonian H , we need to minimize the following loss function:

$$\mathcal{L} = \|H - \sum_k U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)\|_F^2 \quad (13)$$

$$= \text{Tr}(H^2) + \sum_{k, k'} \text{Tr}(U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)^\dagger U_L(\theta_{k'}) \Lambda_{k'} U_L(\theta_{k'})) - 2 \text{Re} \sum_k \text{Tr}(U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k) H) \quad (14)$$

$$=: \sum_{k, k'} A_{kk'} - 2 \text{Re} \sum_k B_k + c, \quad (15)$$

where we define $A_{kk'} := \text{Tr}(U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)^\dagger U_L(\theta_{k'}) \Lambda_{k'} U_L(\theta_{k'}))$, $B_k := \text{Tr}(U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k) H)$, and $c := \text{Tr}(H^2)$ is a constant independent of the parameters $\{\theta^{(k)}, \Lambda_k\}$. The loss function can also be formulated as a tensor network contraction, where the first term corresponds to the tensor $A_{kk'}$, and the second term to the tensor B_k , both illustrated in Fig. 2(b). The quantities $A_{kk'}$ and B_k can be efficiently computed by contracting the corresponding tensor networks depicted in the two diagrams, where the diagrams are connected in a head-to-tail fashion. The overall contraction cost remains polynomial in system size provided that $K = \text{poly}(n)$, the circuit depth $L = O(\log(n))$, and both Λ_k and H admit efficient MPO representations.

Once the contraction is performed, the loss function \mathcal{L} can be minimized using standard gradient descent optimization techniques. By keeping the parameters of all other terms fixed, we can iteratively optimize the parameters $\{\theta^{(k)}, \Lambda_k\}$ for each term k , thereby constructing an approximation of the target Hamiltonian.

V. NUMERICAL RESULTS

Estimating expectation values for Hamiltonians is one of the most crucial tasks in state learning, and numerous algorithms have been developed for this purpose [12, 17, 19, 40]. In our numerical experiments, each two-qubit parameterized gate consists of an iSWAP gate followed by a tensor product of two arbitrary single-qubit gates.

We assess the performance of the GPD algorithm through three representative examples. The first and second examples focus on estimating the expectation values of randomly generated sparse and dense Hamiltonians, respectively. The third example involves estimating the inner product between a Slater determinant $|\phi_\tau\rangle$ and a given pure state $|\psi\rangle$, denoted as $\langle\phi_\tau|\psi\rangle$. The convergence behavior for all three cases is provided in Appendix F.

We demonstrate the effectiveness of the tensor network decomposition algorithm by showing its significant performance advantages over existing qubit-wise commuting algorithms and the classical shadow method.

The numerical simulations were conducted mainly using MindSpore Quantum [55].

VI. GROUND ENERGY ESTIMATION

Sparse Hamiltonians play important roles in quantum computation, chemistry, and simulation [56–58]. Here, we demonstrate the performance of the GPD algorithm with parameterized quantum circuits by estimating the ground state energy of a randomly generated sparse Hamiltonian. We compare our results with state-of-the-art qubit-commuting methods [12, 14, 17–19] and the standard global CS algorithm [40].

TABLE I: Estimation errors for the ground-state energy of a sparse Hamiltonian across different algorithms. The system consists of $n = 8$ qubits, and the GPD algorithm uses a quantum parameterized circuit with depth $L = 4$ and employs $K = 80$ decomposition terms.

Copies	CS [40]	Derand [12]	C-LBCS [19]	SG [18]	Random [40]	Adaptive [14]	OGM [17]	GPD
45	1.095	0.616	1.140	1.247	1.415	1.052	0.738	0.674
160	0.584	0.556	0.708	0.628	1.328	0.831	0.522	0.312
572	0.260	0.361	0.357	0.273	0.637	0.505	0.416	0.203
2038	0.145	0.232	0.184	0.171	0.275	0.248	0.191	0.104
7256	0.066	0.137	0.142	0.082	0.161	0.115	0.122	0.046
25848	0.050	0.069	0.059	0.053	0.084	0.065	0.097	0.030

The numerical results are shown in Table I. Here, we randomly generate an 8-qubit sparse Hamiltonian H with $64 = 8^2$ non-zero matrix elements and the ground energy of approximately -1.536 . In our implementation, we set the number of decomposition terms to $K = 80$ and the depth of the quantum ansatz to $L = 4$, which is significantly smaller than the total number of Pauli terms, 4^8 . Detailed results on how the approximation error between \hat{H} and H depends on the number of decomposition terms—using unitary ansätze with depths ranging from 0, 1, 2, 3—are provided in Appendix F. These results demonstrate that the error decays exponentially with the number of decomposition terms. We further compare our method with existing algorithms, showing that the estimation error achieved by our approach is substantially lower than those of prior methods, illustrating the advantage of our algorithm in estimating the expectation value of sparse Hamiltonians. Specifically, “CS” refers to the global classical shadows protocol [40], which applies a global Clifford operation before measurement, while “Random” denotes the local classical shadows protocol introduced in the same reference.

TABLE II: Estimation errors for the expectation value of a randomly generated dense Hamiltonian across different algorithms. The number of qubits is fixed at $n = 4$, and the GPD algorithm uses a quantum parameterized circuit with depth $L = 4$ and employs $K = 20$ decomposition terms.

Copies	CS [40]	Derand [12]	C-LBCS [19]	SG [18]	Random [40]	Adaptive [14]	OGM [17]	GPD
45	1.702	1.197	1.062	1.737	1.586	1.876	1.152	0.922
160	0.728	0.536	0.742	0.881	1.150	1.261	0.556	0.503
572	0.414	0.296	0.381	0.539	0.707	0.593	0.293	0.253
2038	0.243	0.190	0.206	0.262	0.353	0.313	0.157	0.147
7259	0.103	0.081	0.095	0.124	0.228	0.159	0.085	0.061
25848	0.077	0.064	0.243	0.059	0.096	0.079	0.053	0.046

A. Expectation value estimation for a dense Hamiltonian

To evaluate the effectiveness of the GPD algorithm with parameterized circuits in a general setting, we consider the task of estimating the expectation value of a randomly generated dense Hamiltonian. Specifically, we construct a random matrix A with entries independently sampled from the uniform distribution over $[0, 1]$, and define the

Hamiltonian as $H := (A + A^T)/2$ to ensure Hermiticity. The input quantum state is also randomly generated with amplitudes drawn uniformly from $[0, 1]$. The decomposition is performed using $K = 20$ terms with a parameterized quantum ansatz of depth $L = 4$. We compare the estimation performance of our method against existing algorithms as the number of copies of the quantum state increases. As shown in Table II, the GPD algorithm consistently yields significantly lower estimation errors, demonstrating its effectiveness in general-purpose Hamiltonian estimation tasks. We also report the average estimation error with the number of samplings in the expectation values for several randomly selected Hamiltonians (both sparse and dense) in Appendix F.

TABLE III: Estimation errors for the inner product between a 3-qubit Slater determinant and a 3-qubit pure state across different algorithms. The GPD algorithm uses a quantum parameterized circuit with depth $L = 4$ and employs $K = 20$ decomposition terms.

Copies	CS [40]	FCS [59]	Derand [12]	C-LBCS [19]	SG [18]	Random [40]	Adaptive [14]	OGM [17]	GPD
45	0.366	0.273	0.295	0.245	0.309	0.355	0.311	0.278	0.222
160	0.195	0.160	0.150	0.200	0.165	0.260	0.176	0.159	0.112
572	0.108	0.082	0.072	0.128	0.082	0.126	0.080	0.078	0.056
2038	0.056	0.041	0.041	0.067	0.043	0.072	0.043	0.046	0.030
7256	0.027	0.025	0.023	0.039	0.024	0.044	0.024	0.023	0.015
25848	0.015	0.014	0.017	0.086	0.012	0.023	0.012	0.012	0.009

Inner product estimation—GPD algorithm with parameterized circuits can be easily generalized to estimate non-Hermitian operators by allowing $\{\Lambda_k\}_k$ to be complex-valued diagonal matrices. Hereby, we apply our algorithm to estimate the expectation value of a non-Hermitian operator $O = |1\rangle |\phi_\tau\rangle \langle 0^{n+1}|$. This operator enables the calculation of the inner product between a Slater determinant and a pure quantum state, where $|\phi_\tau\rangle$ represents a τ -Slater determinant. This has applications in algorithms such as the QC-AFQMC algorithm [39]. Here, τ -Slater determinant [59] is defined as $|\phi_\tau\rangle = \tilde{b}_1^\dagger \tilde{b}_2^\dagger \dots \tilde{b}_\tau^\dagger |0^n\rangle$, where $\tilde{b}_k = \sum_{j=1}^n U_{kj} b_j$, U is an $n \times n$ unitary operator, and $\{b_j\}_j$ are annihilation operators.

Provided with the circuit information for preparing $|\psi\rangle$, the inner product of a τ -Slater determinant $|\phi_\tau\rangle$ and a pure state $|\psi\rangle$ can be estimated with our algorithms. We start by preparing the state $|\psi'\rangle = \frac{|0^{n+1}\rangle + |\psi\rangle}{\sqrt{2}}$ using an ancillary qubit and use it as the input state for our algorithms [17]. We then choose the operator $O = |1\rangle |\phi_\tau\rangle \langle 0^{n+1}|$, and hence $\langle \psi | \phi_\tau \rangle = 2\text{Tr}(|\psi'\rangle \langle \psi' | O)$. To demonstrate our algorithms, we set $\tau = 1$, $n = 3$, and use a 3×3 random unitary U . We randomly generate a 3-qubit quantum state $|\psi\rangle$ with $\langle \psi | \phi_\tau \rangle \approx -0.0446 + 0.0552i$. We postpone the distance of the estimated \hat{O} with the ideal O into Appendix F. We give the estimation error for $\langle \psi | \phi_\tau \rangle$ of our algorithm and existing algorithms in Table III. For comparison, we also estimate the errors using existing state-of-the-art qubit-wise commuting measurements [12, 14, 17–19], global/local CS algorithm [40] and fermionic CS algorithm [59]. The numerical result illustrates the advantage of our algorithm in estimating the inner product of a Slater determinant and a quantum pure state. Detailed explanations of the error estimation for these existing algorithms are provided in Appendix E.

B. Expectation value calculation with TND

Here we consider the estimation of the expectation value for a Hamiltonian that admits a matrix product operator (MPO) representation. Specifically, we study the Hamiltonian

$$W = h_0 X_1 \dots X_n + \sum_{j=1}^{n-1} h_j X_1 \dots X_{j-1} Z_j Z_{j+1}. \quad (16)$$

We express W in MPO form as $W = JW^{[1]}W^{[2]}\dots W^{[n]}$, where the boundary tensors are given by $W^{[1]} = (\mathbb{I} \ -Z_1 \ gX_1)$ and $W^{[n]} = (gX_n \ -Z_n \ \mathbb{I}_n)^T$. The bulk tensors for $i \in \{2, \dots, n-1\}$ are defined as

$$W^{[i]} = \begin{pmatrix} \mathbb{I}_i & -Z_i & gX_i \\ 0 & 0 & Z_i \\ 0 & 0 & \mathbb{I}_i \end{pmatrix}. \quad (17)$$

In our numerical experiments, we set $J = 1$ and $g = 1$. The optimal parameters for the unitary ansatz are obtained by minimizing the loss function defined in Eq. (15).

To evaluate the effectiveness of the TND method, we generate a random 8-qubit quantum state ρ and compare the estimation error with that of existing algorithms, as shown in Table IV. In our algorithm, the Hamiltonian is decomposed into three terms, and the depth of the variational quantum circuit is set to 1, with the circuit structure matching that of the GPD algorithm. The results demonstrate that our approach outperforms traditional Pauli-based measurement schemes. We give the approximation error of the Hamiltonian W in Appendix F. To demonstrate the robustness of the evaluation with respect to different quantum input states, we also present the average errors for five randomly selected input states in Appendix F, further supporting the advantages of our TND algorithm with parameterized quantum circuits.

TABLE IV: Estimation errors for the low bond dimension Hamiltonian across different algorithms. The system consists of $n = 8$ qubits. The TND algorithm uses a quantum parameterized circuit with depth $L = 1$ and employs $K = 3$ decomposition terms.

Nsteps	CS [40]	Derand [12]	C-LBCS [19]	SG [18]	Random [40]	Adaptive [14]	OGM [17]	TND
1000	4.616	0.638	0.535	0.369	5.892	0.445	0.448	0.213
3000	3.029	0.700	0.250	0.227	5.932	0.217	0.250	0.103
9000	1.859	0.605	0.156	0.117	6.165	0.121	0.178	0.079
18000	2.292	0.615	0.105	0.108	4.475	0.080	0.094	0.050

VII. LOWER BOUND FOR SAMPLE COMPLEXITY

This work aims to harness the power of parameterized circuits for observable estimation. The numerical results presented in the previous section show that performance is closely tied to circuit depth. This naturally raises the question of whether this dependence stems from our algorithms or if there are inherent limitations to the performance of shallow circuits. In this section, we aim to conduct a protocol-independent analysis of shallow circuit performance. Such an analysis will not only help us evaluate our algorithms but also deepen our understanding of the capabilities of shallow quantum circuits in state learning tasks.

Here, we establish a fundamental lower bound on the number of copies of the quantum state ρ required to accurately estimate the value of $\text{Tr}(\rho H)$. This lower bound holds for all single-copy adaptive strategies that use the given parameterized circuit $U_L(\theta)$ without ancillary systems. Our approach builds on the techniques introduced by Chen et al. [60]. While their work does not cover scenarios where a limited quantum ansatz is utilized.

Proposition 3 (Sample Complexity Lower Bound). *Suppose $H \in \mathbb{C}^{2^n \times 2^n}$ is a Hermitian operator we aim to estimate. Denote $\mathcal{S} := \{U_L(\theta)^\dagger |b\rangle \mid \forall \theta, b \in \{0, 1\}^n\}$. Then for any single-copy adaptive strategy relying solely on the parameterized circuit $U_L(\theta)$ executed on ρ , the sample complexity T for measuring $\text{Tr}(\rho H)$ to accuracy ϵ , for arbitrary unknown state ρ , with success probability of at least $2/3$, is lower bounded by*

$$T = \Omega \left(\frac{\text{Tr}(H_0^2)}{\epsilon^2 \delta(H_0) 4^n} \right), \quad (18)$$

where $H_0 = H - \frac{\text{Tr}(H)}{2^n} \mathbb{I}_{2^n}$ denotes the traceless part of H and $\delta(H_0) := \sup_{|\psi\rangle \in \mathcal{S}} \langle \psi | H_0 | \psi \rangle^2$.

A learning algorithm that uses T copies of measurements to learn a quantum state ρ without quantum memory can be modeled as a rooted tree \mathcal{T} with depth T . The central idea behind proving the proposition is to demonstrate that for a Hamiltonian H , there exist two quantum state sets $\{\rho_x\}_x$ and $\{\frac{\mathbb{I}}{2^n}\}$, such that distinguishing between the expected values $\text{Tr}(\rho_x H)$ and $\text{Tr}(\frac{H}{2^n})$ using the quantum learning algorithm requires at least T samples. This process is carried out with the condition that L -depth quantum circuits are applied before measurement in the computational basis. The problem then reduces to bounding the total variation distance between two distributions generated at the leaves of the rooted tree \mathcal{T} , which quantifies the distinguishability of the distributions produced by the quantum algorithm with T copies of the quantum states. Further details of the proof appear in Appendix D. This bound provides the desired performance guarantee for the algorithm in terms of the required number of samples.

Here, the dependence of this lower bound on the capability of the circuit is mainly reflected on $\delta(H_0)$, which increases with the capability of the quantum devices. When the quantum device can prepare all the eigenstates of H , $\delta(H_0)$ reaches its maximum value. For example, when $H = U_L(\theta)^\dagger P U_L(\theta)$, where P is a Pauli operator, we have $\delta(H_0) = 1$. In this case, our lower bound becomes $\Omega(\frac{1}{\epsilon^2})$, given that $\text{Tr}(H_0^2) = 4^n$. This lower bound is tight in this case because our parameterized circuit can directly diagonalize H . In conclusion, a Hamiltonian $H = U A U^\dagger$

with a shallow-depth unitary U is likely to achieve the lower bound in efficiency, making it particularly useful in near-term quantum applications. The 1D nearest-neighbor Heisenberg Hamiltonian, with its sparsely connected terms, is a strong example of this class. Other examples include Hamiltonians with commuting local terms, such as those in the transverse field Ising model, where only nearest neighbors interact, and hence a shallow-depth U can be constructed. Additionally, matchgate Hamiltonians and certain fermionic Hamiltonians, when decomposed in the Gaussian fermionic basis, permit shallow-depth unitary transformations, which simplify simulation and measurement tasks. Each of these Hamiltonians exemplifies the suitability of our algorithm and further underscores the tightness of the lower bound on circuit depth.

VIII. CONCLUSION AND DISCUSSION

We propose a framework for estimating arbitrary linear properties of a quantum state using available quantum circuits. Our framework takes into account both the limitations of quantum devices and the information of the estimated properties. It includes some existing Pauli decomposition protocols [10–20].

Based on this framework, we propose two classical optimization algorithms that find an optimized decomposition of the observable H with L -depth quantum parameterized circuits such that $H \approx \sum_{k=1}^K U_L(\theta_k)^\dagger \Lambda_k U_L(\theta_k)$. The tensor network approach is well-suited for Hamiltonians that admit compact tensor network representations, while the greedy projection algorithm is designed for general Hamiltonians, particularly in small-sized systems. We then sample the decomposition terms using the importance sampling technique. Consequently, the expectation value of $\text{Tr}(\rho H)$ can be estimated by performing an L -depth quantum parameterized circuit sampled from the set $\{U_L(\theta_k)\}_{k=1}^K$, followed by collecting the corresponding diagonal matrix Λ_k with computational basis measurements. Our algorithms can be easily generalized to other circuit structures and non-Hermitian operators. We numerically verify our algorithm by applying it to calculate the ground energy of a given sparse Hermitian and the inner product of a Slater determinant and a pure state.

The performance of our algorithm heavily depends on the decomposition of the target observable. While our greedy decomposition algorithm has been numerically demonstrated to be efficient compared to some conventional methods, it is not guaranteed to be optimal. Therefore, a key future direction is to develop more efficient classical algorithms for optimizing the decomposition process, such as those based on machine learning. Additionally, our current sample complexity lower bound is loose for low-rank observables, such as fidelity. To address this, we could modify the state discrimination tasks, which form the basis of the proof, to establish a tighter bound. It is also worth exploring the extension of our framework to estimate nonlinear properties and incorporate quantum memories.

In addition to minimizing the sample complexity, we aim to reduce the number of required hardware configurations to estimate a target observable. In practical scenarios, repeating experiments with a fixed configuration is relatively straightforward across various platforms, while changing the configuration often requires additional resources. This goal is equivalent to minimizing the number K in the decomposition $H = \sum_{k=1}^K U_k^\dagger \Lambda_k U_k$, where the $\{U_k\}_k$ are realizable by the hardware. Thus, an important future direction is to explore the relationship between the number of decomposition terms K and the capability of quantum circuits.

We note that Ref. [61] proposed an approach for expectation value estimation that combines informationally (over)complete measurements with tensor network representations of both the Hamiltonian and the measurement operators. A notable distinction of our tensor network approach lies in its treatment of measurements: by bypassing the explicit construction of POVM measurements, our method avoids the additional quantum measurement overhead associated with estimating the probabilities of all POVM basis elements.

ACKNOWLEDGEMENT

We would like to acknowledge Zijian Zhang for the valuable discussions regarding the code from their work. We would like to thank Xiao Yuan, Zhenyu Du, Jens Eisert, and Hong-Ye Hu for the helpful discussions. We acknowledge the MindSpore Quantum framework as the platform for our numerics. BW is supported by the National Natural Science Foundation of China (12405014). ZL and XY acknowledge the support from the National Natural Science Foundation of China Grant No. 12174216 and the Innovation Program for Quantum Science and Technology Grant No. 2021ZD0300804 and No. 2021ZD0300702. FW is supported by BMSTC and ACZSP Grant No. Z221100002722017.

Appendix A: Review of Qubit-Commuting Random Measurements

Let the Pauli operator expansion of a Hermitian operator $H = \sum_{i=1}^m \alpha_i Q_i$. We use the symbol $P \triangleright Q$ to indicate that Q covers P , meaning that each i -th qubit of P , denoted as P_i , is either equal to the identity or to Q_i . We group all m Pauli operators $\{Q_i\}$ into overlapped sets e_1, e_2, \dots, e_K , where all Pauli operators in each e_i commute to each other. There exists a Pauli operator $P^{(i)}$ such that all Pauli operators Q_j in e_i satisfy $Q_j \triangleright P^{(i)}$ for any $i \in [K]$. For instance, when $e = \{Q_1, Q_2\}$, $Q_1 = Z_1 X_2 I_3$ and $Q_2 = Z_1 I_2 Y_3$, we have $P = Z_1 X_2 Y_3$.

In the OGM algorithm [17] and other grouping measurement algorithms [62, 63], sets e_1, \dots, e_K and measurements $P^{(1)}, \dots, P^{(K)}$ are explicitly generated with associated probabilities p_1, \dots, p_K . In other random measurement algorithms, while the groups are not explicitly generated, the analysis in Ref. [17] shows that they implicitly use the same idea. Some algorithms, such as the Derandomized algorithm proposed by Huang et al. [12], generate the counts of $P^{(1)}, \dots, P^{(K)}$ when the total number of measurements T is specified, essentially creating a derandomized version of the probabilistic sampling.

Since the Clifford group stabilizes the Pauli group, any Pauli operator can be decomposed into $P = U^\dagger \Lambda U$, where U is the tensor product of local Clifford gates. Measuring a quantum state ρ with a Pauli operator P is then equivalent to performing a local Clifford gate U , followed by measurement in the computational basis to collect the associated eigenvalues of P [40]. Consequently, qubit-commuting random measurement schemes involve only local Clifford circuits.

Appendix B: Qubit-commuting decomposition

We use the symbol $Q \triangleright P$ to indicate that P covers Q , meaning that each i -th qubit of Q , denoted as Q_i , is either equal to the identity or to P_i . Let the operator $H = \sum_{i=1}^m \alpha_i Q_i$, where $\{Q_i\}_i$ are Pauli operators and $\{\alpha_i\}_i$ are real coefficients. Let e_1, \dots, e_K be (overlapped) qubit-commuting sets such that $e_i \subseteq \{Q_1, \dots, Q_m\}$. We claim that there exists a Pauli operator P_i to cover all elements of the qubit-commuting set e_i for $i \in \{1, \dots, K\}$.

The eigendecomposition for $P^{(i)}$ can be represented as $P^{(i)} = U_i^\dagger O_i U_i$ where U_i is a unitary and O_i is diagonal. For n -qubit tensor product unitary A_j , denote the l -th qubit representation as A_{jl} , i.e., $A_j = \otimes_{l=1}^n A_{jl}$. Then we have

$$H = \sum_{i=1}^m \alpha_i Q_i \quad (\text{B1})$$

$$= \sum_{i=1}^K \sum_{Q_j \in e_i} \alpha_j Q_j \quad (\text{B2})$$

$$= \sum_{i=1}^K \sum_{Q_j \in e_i} \alpha_j \bigotimes_{l=1}^n Q_{jl} \quad (\text{B3})$$

$$= \sum_{i=1}^K \sum_{Q_j \in e_i} \alpha_j U_i^\dagger \left(\bigotimes_{l=1}^n D_{jl} \right) U_i \quad (\text{B4})$$

$$= \sum_{i=1}^K U_i^\dagger \left(\sum_{Q_j \in e_i} \alpha_j D_j \right) U_i \quad (\text{B5})$$

$$= \sum_{i=1}^K U_i^\dagger \Lambda_i U_i \quad (\text{B6})$$

where D_{jl} equals the single-qubit identity \mathbb{I}_2 if $Q_{jl} = \mathbb{I}_2$ and $D_{jl} = O_{jl}$ otherwise, Eq. (B4) holds since $Q_j \triangleright P^{(i)}$ for any $Q_j \in e_i$, and Eq. (B6) holds by setting $\Lambda_i = \sum_{Q_j \in e_i} \alpha_j D_j$.

Appendix C: Review of Commuting Measurements

Ref. [10] and Ref. [20] propose commuting measurement schemes based on grouping or overlapped grouping methods. In this section, we discuss how to decompose $H = \sum_{j=1}^m \alpha_j Q_j$ into $\sum_{k=1}^M U_k^\dagger \Lambda_k U_k$ where Q_j are Pauli operators, U_k are Clifford circuits and M is the number of commuting groups of $\{Q_j\}_{j=1}^m$ [10].

Let $H = \sum_{k=1}^K H_k$ where $H_k = \sum_{Q_j \in g_k} \alpha_j Q_j$ where any two Pauli operators P_i, P_j in set g_k are commute to each other. To generate the U_k and Λ_k for H_k , we start by finding sets $\mathcal{G} = \{\vartheta_1, \dots, \vartheta_n\}$ and $\mathcal{Q} = \{\sigma_1, \dots, \sigma_n\}$ where ϑ_i are Pauli operators and σ_j are single-qubit Pauli operators such that

- $[\vartheta_i, \vartheta_j] = 0$ for any $i, j \in [n]$;
- $[\vartheta_i, Q_l] = 0$ for any $i \in [n], Q_l \in g_k$;
- $\{\vartheta_i, \sigma_j\} = 0$ for any $i \neq j \in [n]$;
- $[\vartheta_i, \sigma_i] = 0$ for any $i \in [n]$;
- $[\sigma_i, \sigma_j] = 0$ for any $i, j \in [n]$.

According to Ref. [64], we can give the set \mathcal{G} , which serves as the generator set for Pauli operators in g_k , meaning that all Q_j can be expressed as a product of some elements ϑ_i in \mathcal{G} . It can also be easily proved that such set \mathcal{Q} also exists. Let $U'_k = \prod_{l=1}^n \frac{1}{\sqrt{2}} (g_l + \sigma_l)$. Suppose $\sigma_l = C_l^\dagger D_l C_l$. Let $L_k = \prod_{l=1}^n C_l$ and $U_k = L_k U'_k$. Then we have $U'_k g_l U_k^\dagger = \sigma_l$, hence $U_k g_l U_k^\dagger = L_k \sigma_l L_k^\dagger = D_l$, and

$$U_k H_k U_k^\dagger = \sum_{Q_j \in g_k} \alpha_j \prod_{\vartheta_i \in S_j} D_i,$$

where $Q_j = \prod_{\vartheta_i \in S_j} g_i$. Such S_j exists by the definition of $\{\vartheta_i\}_i$. Let $\Lambda_k = \sum_{Q_j \in g_k} \alpha_j \prod_{\vartheta_i \in S} D_i$, then $H = \sum_{k=1}^K U_k^\dagger \Lambda_k U_k$.

Appendix D: Proof of Proposition 3

Definition. For probability distributions $p = \{p_1, \dots, p_N\}$ and $q = \{q_1, \dots, q_N\}$, we define the total variation between them by $\text{TV}(p, q) = \frac{1}{2} \sum_{i=1}^N |p_i - q_i|$.

Lemma 1 (Le Cam's two-point method, Lemma 1 in [65] and Lemma 5.3 in [60]). For a learning algorithm without quantum memory being described as a rooted tree \mathcal{T} , let p_ρ be a probability distribution on the leaves, such that $p_{\rho_x}(l)$ equals the probability of arriving at the leaf l . The probability that the learning algorithm that can distinguish quantum state sets $\{\rho_x\}_x$ and $\frac{\mathbb{I}}{2^n}$ correctly is upper bounded by $\frac{1}{2} + \frac{1}{2} \text{TV}(\mathbb{E}_x [p_{\rho_x}], p_{\mathbb{I}/2^n})$.

We derive our result by extending the proof of Theorem 5.5 from Ref. [60]. The central idea is to show that for a Hamiltonian H , there exist two sets of quantum states $\{\rho_x\}_x$ and $\{\frac{\mathbb{I}}{2^n}\}$, such that distinguishing between the expectations $\text{Tr}(\rho_x H)$ or $\text{Tr}(\frac{H}{2^n})$ requires at least T samplings. This task is performed under the condition that L -depth quantum circuits are applied before measurement in the computational basis. By Lemma 1, the task of quantifying how distinguishable the distributions generated by the quantum algorithm using T copies of the quantum states is can be simplified to estimating the total variation distance $\text{TV}(\mathbb{E}_x [p_{\rho_x}], p_{\mathbb{I}/2^n})$ between two distributions, which provides the lower bound of sample complexity $T \geq \Omega\left(\frac{1}{a^2 \delta(H)}\right)$ for the observable measurement task. We will give details later on how to generate this lower bound in the proof part.

We also need to relate the parameter a to the estimation error ϵ . Without loss of generality, assume $\text{Tr}(H) = 0$ for the Hamiltonian under consideration. Denote $H = H_1 = -H_2$. We now define two state sets:

- Set 1: $\{\mathbb{I}/2^n\}$,
- Set 2: $\{\rho_x\}_{x=\{1,2\}}$, each with equal probability,

where $\rho_x = (\mathbb{I} + 3xH_1)/2^n$, and $a \leq \frac{1}{3\|H\|_2}$ to ensure that the matrices in Set 2 are valid density matrices, where $\|H\|_2$ denotes the spectral norm (two-norm) of H .

Consider the state discrimination task involving these two sets. A referee randomly selects one of the sets, and then chooses a state ρ from the selected set according to the corresponding distribution. The referee provides T copies of ρ , and our task is to determine which set ρ comes from, using these T copies. The quantum device available for this task is a n -qubit quantum device of limited capability, whose all possible output states when acting on $|0^n\rangle$ form a set \mathcal{S} .

Note that $\text{Tr}(H \frac{\mathbb{I}}{2^n}) = 0$, $\text{Tr}(H \frac{\mathbb{I}+3aH_1}{2^n}) = \frac{3a\text{Tr}(H^2)}{2^n}$ and $\text{Tr}(H \frac{\mathbb{I}+3aH_2}{2^n}) = -\frac{3a\text{Tr}(H^2)}{2^n}$. One possible way to accomplish the state discrimination task is to measure H to precision $\epsilon = \frac{a\text{Tr}(H^2)}{2^n}$ and check whether the approximate expectation value lies in $[-\frac{a\text{Tr}(H^2)}{2^n}, \frac{a\text{Tr}(H^2)}{2^n}]$. If it holds, we assign ρ to Set 1, otherwise we assign ρ to Set 2.

Thus if we can use this quantum device to measure $\text{Tr}(H\rho)$ with an accuracy of $\epsilon = \frac{a\text{Tr}(H^2)}{2^n}$ and a success probability of at least $2/3$ for any given ρ , then we can accomplish the state discrimination task with the same success probability of at least $2/3$ using this quantum device.

In the following, we show the sample complexity T for accomplishing the state discrimination task by this quantum device with a success probability of at least $2/3$, is lower bounded by $T \geq \Omega\left(\frac{1}{a^2\delta(H)}\right) = \Omega\left(\frac{\text{Tr}(H^2)^2/4^n}{\epsilon^2\delta(H)}\right)$. Consequently, the sample complexity for observable measurement is also lower-bounded by this quantity.

Proof of Proposition 3 in the main text. When the quantum device is capable of performing arbitrary POVM measurements, any state discrimination protocol can be represented by a decision tree, as shown in Figure 4 of [66]. However, since our quantum device is limited to performing parameterized unitary operations $U_L(\theta)$ followed by measurements in the computational basis, the experiment at each node of the decision tree is modified accordingly. Specifically, instead of performing an arbitrary POVM, the measurement at each node is represented by the set of outcomes $\{\alpha_i(\theta)\}_{i=1}^{2^n}$, where $\alpha_i(\theta)$ is the i -th column of $U_L(\theta)^\dagger$.

Given a state discrimination protocol using this quantum device, and the ρ selected by the referee, we label the leaves of the decision tree by l . By Lemma 1, to successfully accomplish the state discrimination task with a probability of at least $2/3$, we need to ensure the depth T of the decision tree (i.e., the sample complexity) is large enough to satisfy the condition in

$$\text{TV}(\mathbb{E}_x[p_{\rho_x}], p_{\mathbb{I}/2^n}) \geq \frac{1}{3}, \quad (\text{D1})$$

where $\rho_x = \frac{\mathbb{I}+3aH_x}{2^n}$ for $x \in \{1, 2\}$ denotes the quantum state in set 2. For each leaf l in the decision tree, there is a unique path from the root node n_0 to the leaf node $n_T = l$. Let the sequence of nodes along this path be denoted as n_0, \dots, n_T and the edges between them as e_1, \dots, e_T . For each $t = 1, \dots, T$, we associate the projector on edge e_t by $|\alpha_{e_t}(\theta_{n_{t-1}})\rangle\langle\alpha_{e_t}(\theta_{n_{t-1}})|$. To give an upper bound of the total variation $\text{TV}(\mathbb{E}_x[p_{\rho_x}], p_{\mathbb{I}/2^n})$, we proceed by calculating the following quantity:

$$\begin{aligned} \frac{\mathbb{E}_x[p_{\rho_x}(l)]}{p_{\mathbb{I}/2^n}(l)} &= \mathbb{E}_x \prod_{t=1}^T \left(\frac{1/2^n + 1/2^n 3a \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H_x | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle}{1/2^n} \right) \\ &= \mathbb{E}_x \prod_{t=1}^T \left(1 + 3a \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H_x | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle \right) \\ &= \mathbb{E}_x \exp \left[\sum_{t=1}^T \log \left(1 + 3a \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H_x | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle \right) \right] \\ &\geq \exp \left[\sum_{t=1}^T \mathbb{E}_x \log \left(1 + 3a \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H_x | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle \right) \right] \\ &= \exp \left[\sum_{t=1}^T \frac{1}{2} \log \left(1 - 9a^2 \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle^2 \right) \right] \\ &\geq \exp \left[-9a^2 \sum_{t=1}^T \langle \alpha_{e_t}(\theta_{n_{t-1}}) | H | \alpha_{e_t}(\theta_{n_{t-1}}) \rangle^2 \right] \\ &\geq \exp[-9a^2 T \delta(H)] \\ &\geq 1 - 9Ta^2 \delta(H) \end{aligned} \quad (\text{D2})$$

where the second inequality holds since $\log(1-u) \geq -2u$ for all $u \in [0, 0.79]$. Using this, we can estimate the total variation as:

$$\text{TV}(\mathbb{E}_x[p_{\rho_x}], p_{\mathbb{I}/2^n}) = \sum_l p_{\mathbb{I}/2^n}(l) \max \left\{ 0, 1 - \frac{\mathbb{E}_x[p_{\rho_x}(l)]}{p_{\mathbb{I}/2^n}(l)} \right\} \leq 9Ta^2 \delta(H). \quad (\text{D3})$$

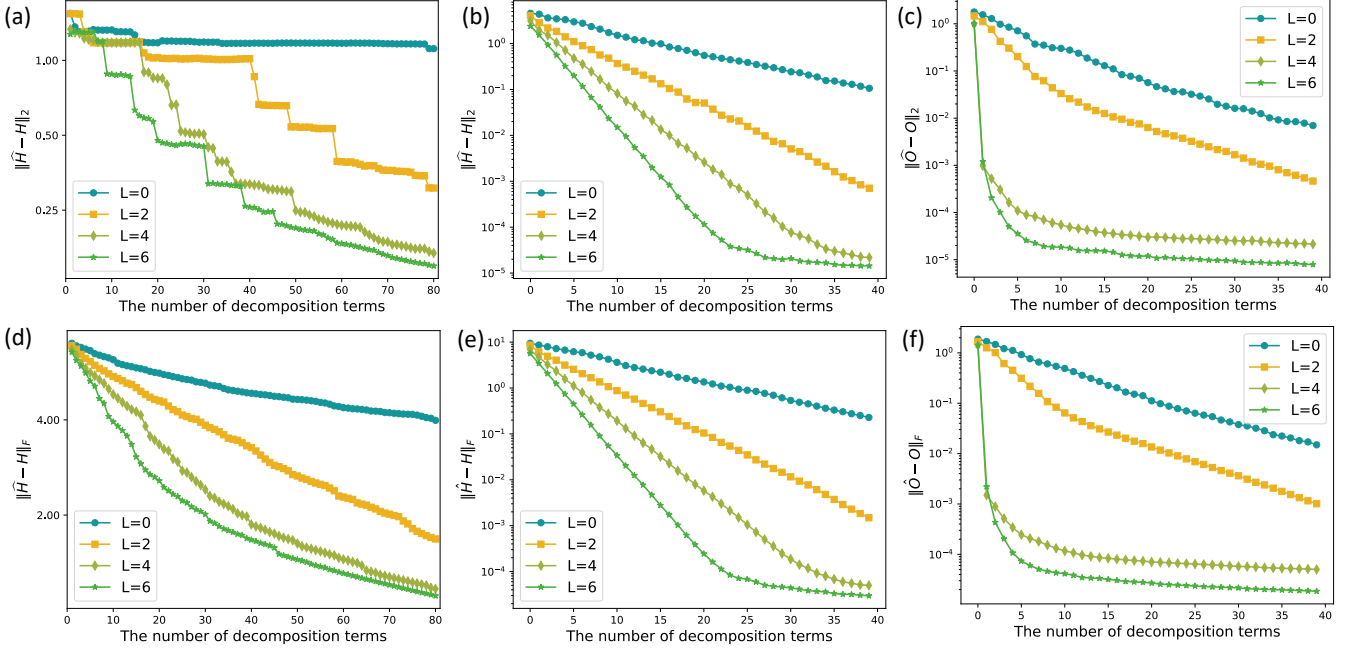


FIG. 3: The variations in Frobenius distance as the number of decomposition terms increases for (a) a sparse Hermitian H and (b) a random generated observable, and (c) a non-Hermitian operator O associated with a Slater determinant.

Combined with Eq. (D1), we need to ensure

$$\frac{1}{3} \leq \text{TV} \left(\mathbb{E}_x [p_{\rho_x}], p_{\mathbb{I}/2^n} \right) \leq 9Ta^2\delta(H), \quad (\text{D4})$$

which gives the bound

$$T \geq \Omega \left(\frac{1}{a^2\delta(H)} \right) = \Omega \left(\frac{\text{Tr}(H^2)^2/4^n}{\epsilon^2\delta(H)} \right). \quad (\text{D5})$$

□

Appendix E: Preliminary to inner product calculation

Given $|\psi\rangle := U_\psi |0\rangle$ and an ancilla initialized to $|0\rangle_a$, by performing Hadamard gate on $|0\rangle_a$, followed by Control- U_ψ gate, we obtain $|\psi'\rangle = \frac{|0^n\rangle + |1\rangle|\psi\rangle}{\sqrt{2}}$.

Since the observable H is not Hermitian and $\text{Tr}(\rho H)$ cannot be directly approximated with existing random measurement algorithms. When executing the existing algorithm to calculate the inner product, we decompose the observable H into two Hermitians $H_1 := (H + H^\dagger)/2$ and $H_2 := -i(H - H^\dagger)/2$, then $H = H_1 + iH_2$, and it can then be approximated in two parts. The total error $\varepsilon = |\hat{v} - \text{Tr}(\rho H)| = \sqrt{(\hat{v}_1 - \text{Tr}(\rho H_1))^2 + (\hat{v}_2 - \text{Tr}(\rho H_2))^2}$ where \hat{v}_j is the estimation of $\text{Tr}(\rho H_j)$ for $j \in \{1, 2\}$.

Appendix F: Supplementary numerics

Here, we present the distances based on 2-norm and F -norm of $H^{(k)}$ and H as k increases for sparse Hamiltonian (a),(d), dense Hamiltonian (b),(e) and operator associated with Slater-determinant (c), (f). As illustrated in Fig. 3, the Frobenius distance between $H^{(k)}$ and H decays exponentially.

Here we give the performance of the tensor-network algorithm and other existing algorithms in the average of five group experiments, concerning the random input states, for the same Hamiltonian W defined in the main text. The comparison with existing qubit-wise commuting algorithms is illustrated in Fig. 4.

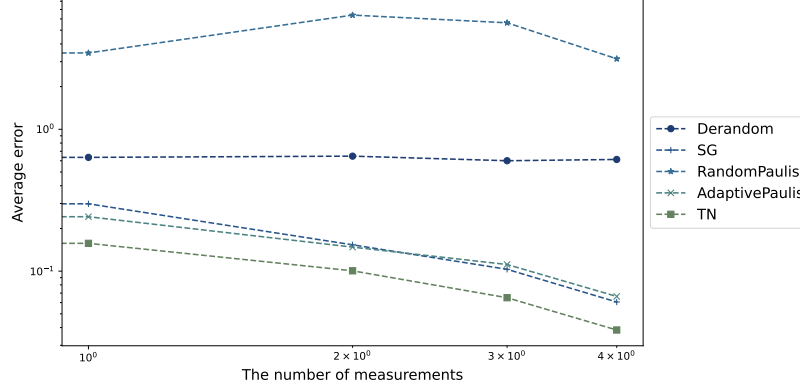


FIG. 4: Comparison of the TN algorithm with Derandom [12], SG [18], RandomPauli [40], and AdaptivePaulis [16] in terms of the average estimation error as a function of the number of measurements. The evaluation is performed over five experiments, each using a different randomly generated quantum input state. The system has 8 qubits. The TN algorithm uses a quantum parameterized circuit with depth $L = 1$ and employs $K = 3$ decomposition terms.

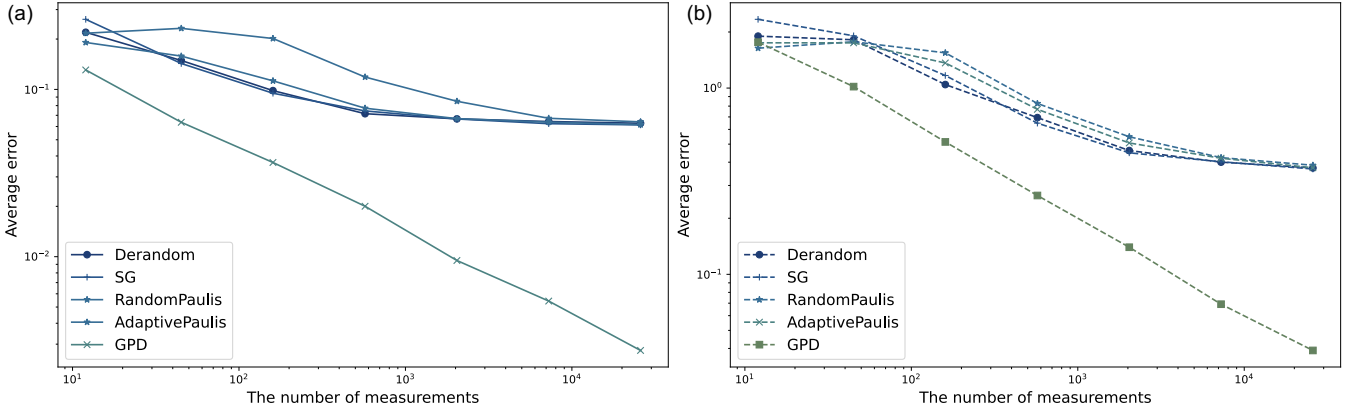


FIG. 5: Comparison of the GPD algorithm with Derandom [12], SG [18], RandomPauli [40], and AdaptivePaulis [16] in terms of average estimation error versus the number of measurements, evaluated over five randomly generated datasets. (a) Random sparse Hamiltonians and (b) Random dense (general) Hamiltonians. The system has 4 qubits. The GPD algorithm uses a quantum parameterized circuit with depth $L = 4$ and employs $K = 20$ decomposition terms.

To further demonstrate the average-case performance of greedy projection Decomposition (GPD) algorithms, we present the average estimation errors over five randomly generated Hamiltonians, including both sparse and dense cases. The number of qubits is fixed at $n = 4$. For each Hamiltonian, we construct a $2^n \times 2^n$ matrix A . In the sparse case, we randomly select n non-zero elements from the interval $[0, 1]$ to populate A , while in the dense case, all elements of A are independently sampled from $[0, 1]$. To ensure Hermiticity, we define the Hamiltonian as $H = (A + A^\dagger)/2$. The input state is the same as that used for the random dense Hamiltonian in the main text and is also randomly generated. The GPD algorithm uses a quantum parameterized circuit with depth $L = 4$ and employs $K = 20$ decomposition terms. The comparison in Fig. 5 demonstrates that, on average, our algorithm significantly outperforms existing qubit-wise commuting methods. As most of these methods exhibit similar performance, we include only a few representative algorithms in the figure to enhance clarity and readability.

Fig. 6(a-b) illustrates the approximation error of the Hamiltonian W as a function of the number of decomposition terms, using ansatz circuits of depths 0, 1, and 2, which shows the estimation error for the Hamiltonian declines with the decomposition terms, and one layer of quantum parameterized circuit is enough for the estimation.

[1] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

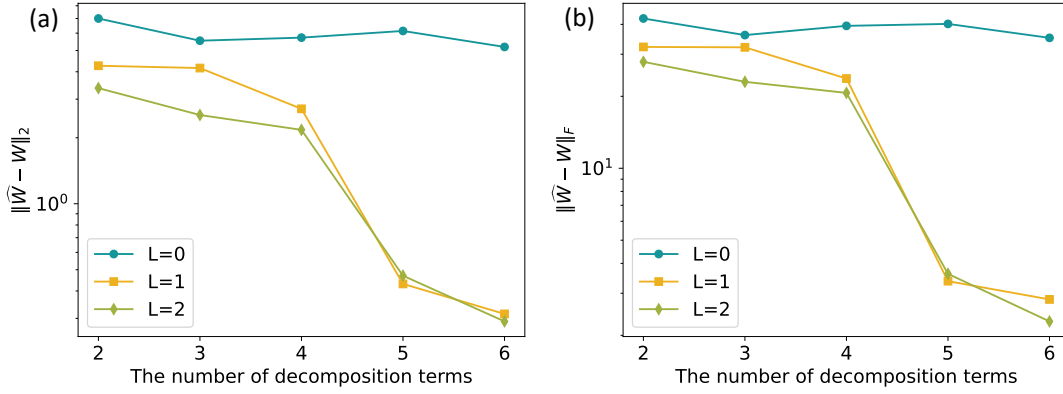


FIG. 6: The variations in Frobenius distance for the estimated Hamiltonian \hat{W} and the ideal low bond dimension Hamiltonian W as the number of decomposition terms increases with the TN method.

- [2] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. *Chemical reviews*, 119(19):10856–10915, 2019.
- [3] Artur F Izmaylov, Tzu-Ching Yen, and Ilya G Ryabinkin. Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators? *Chemical science*, 10(13):3746–3755, 2019.
- [4] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, Jul 2020.
- [5] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Rev. Mod. Phys.*, 92:015003, Mar 2020.
- [6] Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Phys. Rev. Lett.*, 124:020503, Jan 2020.
- [7] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.
- [8] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC ’16, page 913–925, New York, NY, USA, 2016. Association for Computing Machinery.
- [9] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, May 2018.
- [10] Tzu-Ching Yen, Vladyslav Vertelelskyi, and Artur F Izmaylov. Measuring all compatible operators in one series of single-qubit measurements using unitary transformations. *Journal of chemical theory and computation*, 16(4):2400–2409, 2020.
- [11] Vladyslav Vertelelskyi, Tzu-Ching Yen, and Artur F Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of chemical physics*, 152(12), 2020.
- [12] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Phys. Rev. Lett.*, 127:030503, Jul 2021.
- [13] Stefan Hillmich, Charles Hadfield, Rudy Raymond, Antonio Mezzacapo, and Robert Wille. Decision diagrams for quantum measurements with shallow circuits. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 24–34. IEEE, 2021.
- [14] Charles Hadfield. Adaptive pauli shadows for energy estimation. *arXiv preprint arXiv:2105.12207*, 2021.
- [15] Daniel Miller, Laurin E Fischer, Igor O Sokolov, Panagiotis Kl Barkoutsos, and Ivano Tavernelli. Hardware-tailored diagonalization circuits. *arXiv preprint arXiv:2203.03646*, 2022.
- [16] Ariel Shlosberg, Andrew J. Jena, Priyanka Mukhopadhyay, Jan F. Haase, Felix Leditzky, and Luca Dellantonio. Adaptive estimation of quantum observables. *Quantum*, 7:906, January 2023.
- [17] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *Quantum*, 7:896, 2023.
- [18] Alexander Gresch and Martin Kliesch. Guaranteed efficient energy estimation of quantum many-body hamiltonians using shadowgrouping. *Nature communications*, 16(1):689, 2025.
- [19] Zi-Jian Zhang, Kouhei Nakaji, Matthew Choi, and Alán Aspuru-Guzik. A composite measurement scheme for efficient quantum observable estimation. *arXiv preprint arXiv:2305.02439*, 2023.
- [20] Tzu-Ching Yen, Aadithya Ganeshram, and Artur F Izmaylov. Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates. *npj Quantum Information*, 9(1):14, 2023.
- [21] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 325–338, New York, NY, USA, 2018. Association for Computing Machinery.

- [22] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, June 2022.
- [23] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.*, 126:190505, May 2021.
- [24] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585, 2022.
- [25] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 391–404, 2023.
- [26] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, Jan 2023.
- [27] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos. Probing rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, 2019.
- [28] G.I. Struchalin, Ya. A. Zagorovskii, E.V. Kovlakov, S.S. Straupe, and S.P. Kulik. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum*, 2:010307, Jan 2021.
- [29] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. Experimental quantum state measurement with classical shadows. *Phys. Rev. Lett.*, 127:200501, Nov 2021.
- [30] Hong-Ye Hu, Andi Gu, Swarnadeep Majumder, Hang Ren, Yipei Zhang, Derek S Wang, Yi-Zhuang You, Zlatko Mineev, Susanne F Yelin, and Alireza Seif. Demonstration of robust and efficient quantum property learning with shallow shadows. *arXiv preprint arXiv:2402.17911*, 2024.
- [31] Renato Farias, Raghavendra D Peddinti, Ingo Roth, and Leandro Aolita. Robust shallow shadows. *arXiv preprint arXiv:2405.06022*, 2024.
- [32] B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller. Probing scrambling using statistical correlations between randomized measurements. *Phys. Rev. X*, 9:021061, Jun 2019.
- [33] Andreas Elben, Richard Kueng, Hsin-Yuan (Robert) Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, and Benoît Vermersch. Mixed-state entanglement from local randomized measurements. *Phys. Rev. Lett.*, 125:200501, Nov 2020.
- [34] You Zhou, Pei Zeng, and Zhenhuan Liu. Single-copies estimation of entanglement negativity. *Phys. Rev. Lett.*, 125:200502, Nov 2020.
- [35] Andreas Elben, Benoît Vermersch, Rick van Bijnen, Christian Kokail, Tiff Brydges, Christine Maier, Manoj K. Joshi, Rainer Blatt, Christian F. Roos, and Peter Zoller. Cross-platform verification of intermediate scale quantum devices. *Phys. Rev. Lett.*, 124:010504, Jan 2020.
- [36] Ze-Pei Ciani, Hossein Dehghani, Andreas Elben, Benoît Vermersch, Guanyu Zhu, Maissam Barkeshli, Peter Zoller, and Mohammad Hafezi. Many-body chern number from statistical correlations of randomized measurements. *Phys. Rev. Lett.*, 126:050501, Feb 2021.
- [37] Lata Kh Joshi, Andreas Elben, Amit Vikram, Benoît Vermersch, Victor Galitski, and Peter Zoller. Probing many-body quantum chaos with quantum simulators. *Phys. Rev. X*, 12:011018, Jan 2022.
- [38] Zhenhuan Liu, Yifan Tang, Hao Dai, Pengyu Liu, Shu Chen, and Xiongfeng Ma. Detecting entanglement in quantum many-body systems via permutation moments. *Phys. Rev. Lett.*, 129:260501, Dec 2022.
- [39] William J Huggins, Bryan A O’Gorman, Nicholas C Rubin, David R Reichman, Ryan Babbush, and Joonho Lee. Unbiasing fermionic quantum monte carlo with a quantum computer. *Nature*, 603(7901):416–420, 2022.
- [40] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.
- [41] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth. *arXiv:2407.07754*, 2024.
- [42] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Phys. Rev. Res.*, 5:023027, Apr 2023.
- [43] Christian Berton, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random clifford circuits. *Phys. Rev. Lett.*, 133:020602, Jul 2024.
- [44] Ahmed A. Akhtar, Hong-Ye Hu, and Yi-Zhuang You. Scalable and Flexible Classical Shadow Tomography with Tensor Networks. *Quantum*, 7:1026, June 2023.
- [45] Matteo Ippoliti, Yaodong Li, Tibor Rakovszky, and Vedika Khemani. Operator relaxation and the optimal depth of classical shadows. *Phys. Rev. Lett.*, 130:230403, Jun 2023.
- [46] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.
- [47] Gerald Knizia and Garnet Kin-Lic Chan. Density matrix embedding: A strong-coupling quantum embedding theory. *Journal of Chemical Theory and Computation*, 9(3):1428–1432, 2013. PMID: 26587604.
- [48] Alberto Baiardi and Markus Reiher. The density matrix renormalization group in chemistry and molecular physics: Recent developments and new challenges. *The Journal of Chemical Physics*, 152(4):040903, 01 2020.
- [49] Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. *New Bounds for Matrix Multiplication: from Alpha to Omega*, pages 3792–3835.
- [50] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

- [51] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 217–228, New York, NY, USA, 2019. Association for Computing Machinery.
- [52] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac. Matrix product density operators: Simulation of finite-temperature and dissipative systems. *Physical Review Letters*, 93(20), November 2004.
- [53] Y.-Y. Shi, L.-M. Duan, and G. Vidal. Classical simulation of quantum many-body systems with a tree tensor network. *Phys. Rev. A*, 74:022320, Aug 2006.
- [54] Junxiang Huang, Wenhao He, Yukun Zhang, Yusen Wu, Bujiao Wu, and Xiao Yuan. Tensor-network-assisted variational quantum algorithm. *Phys. Rev. A*, 108:052407, Nov 2023.
- [55] Xusheng Xu, Jiangyu Cui, Zidong Cui, Runhong He, Qingyu Li, Xiaowei Li, Yanling Lin, Jiale Liu, Wuxin Liu, Jiale Lu, et al. Mindspore quantum: A user-friendly, high-performance, and ai-compatible quantum computing framework, 2024.
- [56] Dominic W Berry, Graeme Ahokas, Richard Cleve, and Barry C Sanders. Efficient quantum algorithms for simulating sparse hamiltonians. *Communications in Mathematical Physics*, 270:359–371, 2007.
- [57] Aram W. Harrow, Avinandan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, Oct 2009.
- [58] Hitesh J Changlani, Huihuo Zheng, and Lucas K Wagner. Density-matrix based determination of low-energy model hamiltonians from ab initio wavefunctions. *The Journal of chemical physics*, 143(10), 2015.
- [59] Kianna Wan, William J Huggins, Joonho Lee, and Ryan Babbush. Matchgate shadows for fermionic quantum simulation. *Communications in Mathematical Physics*, 404(2):629–700, 2023.
- [60] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585, 2022.
- [61] Stefano Mangini and Daniel Cavalcanti. Low-variance observable estimation with informationally-complete measurements and tensor networks. *arXiv preprint arXiv:2407.02923*, 2024.
- [62] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X*, 8:031022, Jul 2018.
- [63] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method. *Journal of chemical theory and computation*, 16(1):190–195, 2019.
- [64] Sergey Bravyi, Jay M Gambetta, Antonio Mezzacapo, and Kristan Temme. Tapering off qubits to simulate fermionic hamiltonians. *arXiv preprint arXiv:1701.08213*, 2017.
- [65] Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer, 1997.
- [66] Zhenhuan Liu and Fuchuan Wei. Separation between entanglement criteria and entanglement detection protocols. *arXiv preprint arXiv:2403.01664*, 2024.