Error-mitigated and adaptive depth fermionic classical shadows

Bujiao Wu, Kaiming Bian, and Dax Enshan Koh

Assessing the properties of interacting fermionic systems stands as a fundamental task in modern physics, with wide-ranging applications in quantum chemistry [1], condensed matter physics [2], and materials science [3]. Quantum simulation provides a promising avenue for studying these systems, with analog simulation being a prevalent approach, contingent upon robust readout techniques to estimate properties accurately. The readout step, however, often poses a significant bottleneck in many quantum simulation schemes.

For natural fermionic systems, the focus shifts to estimating k local fermionic observables instead of attempting to learn the full quantum state, which would be impractical due to exponential resource requirements. Notably, the adaptive variational quantum algorithm (VQE) utilizes 2-RDMs up to 4-RDMs to simulate many-body interactions effectively [4–7]. Consequently, methods of readout frequently concentrate on estimating these fermionic reduced density matrices.

While classical shadow algorithms for fermionic systems typically require exponentially many copies in qubit systems, recent advancements have introduced more efficient techniques that do not necessitate encoding of the Hamiltonians [8–10]. However, these approaches do not account for noise, which is prevalent in real physical systems.

In the context of the noisy intermediate-scale quantum (NISQ) era, where current quantum simulators are heavily affected by noise, robust shadow estimation schemes have been developed for qubit systems [11, 12]. Extending these concepts to fermionic systems, we propose an error-mitigated shadow estimation scheme tailored for realistic noise channels. Notably, our method avoids the need for encoding the Hamiltonian into qubit representation.

For n-qubit quantum systems, our algorithm, which employs the easily prepared initial state $|0^n\rangle\langle 0^n|$ assumed to be noiseless, provably efficiently estimates all elements of k-RDMs with $\widetilde{\mathcal{O}}(kn^k)$ scaled copies of quantum states and $\widetilde{\mathcal{O}}(\sqrt{n})$ scaled calibration measurements for the gate-independent time-stationary Markovian (GTM) noise with constant average noise fidelities. We demonstrate the feasibility of our protocol across various common noise channels and determine its effectiveness through numerical experiments. Our findings confirm the potential of our methods in real-world experimental scenarios.

Our error-mitigated fermionic classical shadow technique constitutes an extension of the work by Chen et al. [11], accommodating scenarios where the gate-set lacks (1) 3-design properties [13] and (2) the applicability of the randomized benchmarking scheme developed by Helsen et al. [14].

Zhao et al. [8] have proved that there does not exist subgroup $G \subseteq \mathrm{Cl}_n \cap \mathbb{M}_n$ which has a better fermionic shadow norm, where Cl_n is the n-qubit Clifford group and $\mathbb{M}_n$ is the matchgate group. An intriguing open question remains:

Is there a shallow-depth FCS algorithm that is efficient for some specific set of fermionic observables and maintains the same sample complexity as the original FCS algorithm?

Here we identify the minimum circuit depth required by the FCS protocol for $2k$-local Majorana strings. Beginning with 2-local Majorana strings, we derive an expression that characterizes the dependence of sample complexity on circuit depth. Building on these findings, we investigate the depth required to minimize sample complexity for general k-local Majorana strings.

Specifically, we propose an adaptive-depth fermionic classical shadow (ADFCS) algorithm for optimizing the depth of the brickwork random matchgate circuit. Similar to the FCS algorithm, ADFCS applies a random matchgate circuit $U_{Q_{d^*}}$ and utilizes classical shadows to estimate the expectation value of an observable H. By Ref. [15], constructing a Haar-random matchgate circuit with a brickwork architecture

requires $\Omega(n^2)$ depth. In contrast, ADFCS adaptively selects the depth d^* based on the interaction distance of the observable and constructs the random matchgate circuit $U_{Q_{d^*}}$ accordingly. The depth d^* is linear to $\max\left\{d_{\text{int}}^2/\log n, d_{\text{int}}(S)\right\}$ where d_{int} is the interaction distance of the fermionic observable. For example, we numerically demonstrate that ADFCS with a depth of $d^* = 5$ achieves comparable estimation precision to the FCS method when estimating the expectation value of the Kitaev chain Hamiltonian for $n = 10$ qubits. Additionally, we evaluate the ADFCS method by analyzing the estimation error for several k-Majorana strings with varying interaction distances. The numerical results indicate that adaptively selecting a relatively shallow depth can achieve a sample complexity similar to that of FCS for most k-local Majorana strings.

Error mitigated fermionic classical shadow — Let $\widehat{\mathcal{M}} := \sum_{k=0}^{n} \hat{f}_{2k}\mathcal{P}_{2k}$ be the estimator for the noisy shadow channel $\widetilde{\mathcal{M}} = \sum_{k=0}^{n} f_{2k}\mathcal{P}_{2k}$, where $\hat{f}_{2k} = 2^n \langle\!\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\!\rangle / \binom{n}{k}$ is an unbiased estimator of $f_{2k} \in \mathbb{R}$, by sampling $x \in \{0,1\}^n$ from the noisy shadow protocol obtained by starting from the input state $|\mathbf{0}\rangle\!\rangle$ and applying a noisy quantum circuit $\widetilde{\mathcal{U}}_Q$ followed by a Z-basis measurement, where $\mathcal{U}_Q$ is uniformly randomly picked from the matchgate group, $\mathcal{P}_k$ is the projector onto the subspace $\Gamma_k := \{\gamma_S, |S| = k\}$, γ_S is the product of the Majorana operators indexed by the set S. We give the details of the algorithm in Ref. [16].

Let $\hat{v} := \text{Tr}\left(\widehat{\mathcal{M}}^{-1}\mathcal{U}^\dagger(|x\rangle\langle x|)H\right)$ be an estimation of $\text{Tr}\left(\rho H\right)$ for some observables H in the even subspace $\Gamma_{\text{even}} := \cup_k \Gamma_{2k}$ and quantum states ρ, where x follows the distribution $\text{Tr}\left(|x\rangle\langle x|\,\widehat{\mathcal{U}}(\rho)\right)$ for $x \in \{0,1\}^n$, then we have

$$|\hat{v} - \text{Tr}\left(\rho H\right)| \leq \left|\hat{v} - \text{Tr}\left(\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}\left(\rho\right)H\right)\right| + \left|\text{Tr}\left(\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}\left(\rho\right)H\right) - \text{Tr}\left(\rho H\right)\right| = \varepsilon_e + \varepsilon_c, \tag{1}$$

where $\varepsilon_e := \left|\hat{v} - \text{Tr}\left(\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right)\right|$ is the estimation error and $\varepsilon_c := \left|\text{Tr}\left(\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}\left(\rho\right)H\right) - \text{Tr}\left(\rho H\right)\right|$ is the calibration error. Therefore, by determining the necessary number of samples N_e and K_e to achieve the desired level of estimation error ε_e (as well as N_c and K_c to account for the calibration error ε_c), we can obtain an estimation $\hat{v}$ with an overall error of $\varepsilon = \varepsilon_e + \varepsilon_c$, using $N_e K_e$ copies of the input state ρ.

To calculate the representation of each element $\left\langle {}^{\boldsymbol{k}}\mathbf{D}\middle|{}^{\boldsymbol{k}}\mathbf{D}\right\rangle_{j_1,\ldots,j_k;l_1,\ldots,l_k}$, where j_i, l_i are in the range $[n]$ for $i \in [k]$, we need to calculate $m = \mathcal{O}\left(n^k\right)$ expectations for different $\widetilde{\gamma}_S$, where $|S| = 2k$. With the number of estimation samplings $R_e = \mathcal{O}\left(\frac{kn^k \ln(n/\delta_e)}{\mathcal{B}_k^2 \varepsilon_e^2}\right)$ where $\mathcal{B}_k$ is a parameter associated with noise strength, with its specific definition provided in the main file, and the number of calibration samplings $R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n \ln(1/\delta_c)}{\mathcal{B}_{\min}^2 \varepsilon_c^2}\right)$, where $\mathcal{B}_{\max} = \max_k |\mathcal{B}_k|$, and $\mathcal{B}_{\min} = \min_k |\mathcal{B}_k|$, the estimation error can be bounded to $\varepsilon_e + \varepsilon_c$. The technical version of the work, encompassing all analysis details, is available in Ref. [16].

Adaptive depth fermioinc classical shadow — We utilize U_{Q_d} to denote a d-depth matchgate circuit with brickwork structure, and $\mathbb{U}_{Q_d}$ to denote the set of all d-depth matchgate circuits. We denote $\mathcal{M}_d$ as the d-depth ADFCS shadow channel. Due to the expectation property of the subset $\mathbb{U}_{Q_d}$, the Majorana operator remains invariant under the action of $\mathcal{M}_d$ up to a scaling factor $\alpha_{S,d}$, i.e., $\mathcal{M}_d(\gamma_S) = \alpha_{S,d}\gamma_S$, where $\alpha_{S,d} := \int dU_{Q_d} \left|\langle \mathbf{0}| U_{Q_d}\gamma_S U_{Q_d}^\dagger |\mathbf{0}\rangle\right|^2$. Analogous to the FCS estimator, the ADFCS estimator is an unbiased estimator. For any observable γ_S and an unknown quantum state ρ, the variance of the ADFCS estimator $v = \text{Tr}\left(\hat{\rho}\gamma_S\right)$ is bounded as $1/\alpha_{S,d}$. We show that the quantity $\alpha_{S,d}$ can be represented as a tensor network, which can be efficiently calculated if the depth d is shallow.

To find an explicit representation for $\alpha_{S,d}$ with depth d, we focus on random matchgate circuits characterized by an even number of qubits and an odd circuit depth. For other cases of the random matchgate circuits, the contraction can be analyzed using a similar method. For the random matchgate circuits with this structure, the contraction of $\mathcal{C}$ can be represented within the space of quadratic polynomials with $\frac{n}{2}$

variables. Denote the quadratic polynomial representation space as $\mathcal{P}_{\frac{n}{2}}$, the contraction of $\mathcal{C}$ could be described by the random walk with a simple pattern in space $\mathcal{P}_{\frac{n}{2}}$. We observe that the random walk follows the pattern of a symmetry lazy random walk (SLRW) [17, 18] almost everywhere within the Γ_2 space, and we found that $\alpha_{S,d}$ can be bounded by $\alpha^L_{S,d}$, which is associated with SLRW transition operator $\mathcal{L}_{\Gamma_2}$:

$$\alpha^L_{S,d} = \langle\langle \mathbf{0}, \mathbf{0} | \mathcal{L}^{\lfloor \frac{d}{2} \rfloor}_{\Gamma_2} \mathcal{T}_{\text{init}} | \gamma_S, \gamma_S \rangle\rangle, \tag{2}$$

where $\mathcal{T}_{\text{init}}$ is the tensor of the first layer of the circuit. Hence, the entire tensor network $\mathcal{C}$ represents a random walk starting from γ_S and taking $\lfloor \frac{d}{2} \rfloor$ steps, which is shown in Fig. 4(c). we adaptively select the depth d^* as

$$d^* = \Theta\left(\max\left\{ \frac{d_{\text{int}}(S)^2}{\log(n)}, d_{\text{int}}(S) \right\} \right). \tag{3}$$

Such d^* ensures that the scaling of the overall $\alpha_{S,d}$ is compatible with α_S for the FCS protocol with the theoretical and numerical supports. The details is shown in Ref. [19].

Why QCTiP? — We believe that our work would be a suitable contribution to the QCTiP due to its focus on addressing a fundamental challenge in quantum computation, simulation, and quantum information. It efficiently estimates the expectation values of fermionic Hamiltonians and k-particle reduced density matrices (k-RDMs) of an n-mode fermionic state. These estimations are crucial for simulating a wide range of physical systems relevant to many-body physics, chemistry, and materials science.

We propose an error-mitigated classical shadow (CS) algorithm specifically designed for fermionic systems, taking into account the noise present in near-term quantum devices, particularly gate-independent, time-stationary, and Markovian (GTM) noise. However, implementing FCS requires random matchgate circuits with polynomial depth, which poses a significant challenge for near-term quantum devices due to their limited capabilities. To address this issue, we propose an adaptive-depth fermionic classical shadow (ADFCS) protocol that reduces the depth requirements of FCS while maintaining accuracy. Through theoretical analysis and numerical fitting, we show that the required depth for approximating a Majorana string γ_S—defined as the product of multiple Majorana operators—is proportional to $\max\left\{ \frac{d^2_{\text{int}(S)}}{\log n}, d_{\text{int}}(S) \right\}$, where $d_{\text{int}}(S)$ represents the minimum distance between elements in set S. Our numerical experiments demonstrate the effectiveness of ADFCS in estimating fermionic observables with varying depths of random matchgate circuits. Overall, our work contributes to advancing the field of quantum information science by providing robust and more efficient solutions to a key challenge in quantum algorithm and quantum information, aligning with the goals and interests of the QCTiP audience.

Error-mitigated and adaptive depth fermionic classical shadows

Bujiao Wu, Kaiming Bian and Dax Enshan Koh

Efficiently estimating fermionic Hamiltonian expectation values is vital for simulating various physical systems. Classical shadow (CS) algorithms offer a solution by reducing the number of quantum state copies needed, but noise in quantum devices poses challenges. We propose an error-mitigated CS algorithm assuming gate-independent, time-stationary, and Markovian (GTM) noise. Our algorithm employs the easily prepared initial state $|0^n\rangle\langle 0^n|$ assumed to be noiseless, efficiently estimates k-RDMs with $\widetilde{\mathcal{O}}(kn^k)$ state copies and $\widetilde{\mathcal{O}}(\sqrt{n})$ calibration measurements for GTM noise in n-qubit system. Numerical simulations confirm our algorithm's efficacy in noisy settings.

However, implementing FCS requires polynomial circuit depth, which poses a significant challenge for near-term quantum devices due to their limited capabilities. To address this issue, we propose an adaptive-depth fermionic classical shadow (ADFCS) protocol that reduces the depth requirements of FCS while maintaining accuracy. Through theoretical analysis and numerical fitting, we show that the required depth for approximating a fermionic observable is proportional to the square of its interaction distance. Our numerical experiments demonstrate the effectiveness of ADFCS in estimating fermionic observables with varying depths of random matchgate circuits. Numerical results indicate that ADFCS significantly reduces circuit depth requirements while preserving the resource demands on quantum states.

The details are presented in two separate papers [16, 19], as shown in the following.

Error-mitigated fermionic classical shadows on noisy quantum devices

Bujiao Wu, and Dax Enshan Koh

Efficiently estimating fermionic Hamiltonian expectation values is vital for simulating various physical systems. Classical shadow (CS) algorithms offer a solution by reducing the number of quantum state copies needed, but noise in quantum devices poses challenges. We propose an error-mitigated CS algorithm assuming gate-independent, time-stationary, and Markovian (GTM) noise. For n-qubit systems, our algorithm, which employs the easily prepared initial state $|0^n\rangle\langle 0^n|$ assumed to be noiseless, efficiently estimates k-RDMs with $\widetilde{\mathcal{O}}(kn^k)$ state copies and $\widetilde{\mathcal{O}}(\sqrt{n})$ calibration measurements for GTM noise with constant fidelities. We show that our algorithm is robust against noise types like depolarizing, damping, and X-rotation noise with constant strengths, showing scalings akin to prior CS algorithms for fermions but with better noise resilience. Numerical simulations confirm our algorithm's efficacy in noisy settings, suggesting its viability for near-term quantum devices.

INTRODUCTION

Assessing the properties of interacting fermionic systems constitutes one of the core tasks of modern physics, a task that has a wealth of applications in quantum chemistry [1], condensed matter physics [2], and materials science [3]. Notions of *quantum simulation* offer an alternative route to studying this important class of systems. In *analog simulation*, one prepares the system of interest under highly controlled conditions. However, any such effort makes sense only if one has sufficiently powerful *readout techniques* available that allow one to estimate properties. In fact, the read-out step constitutes a core bottleneck in many schemes for quantum simulation.

Fortunately, for natural fermionic systems, one often does not need to learn the full unknown quantum state; trying to do so regardless would be highly impractical, as the resources required for a full tomographic recovery would scale exponentially with the size of the system. Instead, what is commonly needed are the

so-called *k-particle reduced density matrices*, abbreviated as k-RDMs. These are expectation values of polynomials of fermionic operators of the $2k$-th degree. Naturally, the expectation value of any interaction fermionic Hamiltonian can be estimated using 2-RDMs only [4, 5]. Indeed, the adaptive *variational quantum algorithm* (VQE) [20] also utilizes up to 4-RDMs to simulate many-body interactions in the ground and excited state [6, 7]. That is to say, meaningful methods of read-out often focus on estimating such fermionic reduced density matrices.

On the highest level, several approaches can be pursued when dealing with fermionic operators. One of those—and the one followed here—is to treat the fermionic system basically as a collection of spins. Then given spin Hamiltonians $\{H_i\}_{i=1}^m$ and an unknown quantum state ρ, where $m = \mathcal{O}(\mathrm{poly}(n))$, the *classical shadow* (CS) algorithm or its variants [21–33] in qubit systems are among the most promising ways to calculate the expectations $\mathrm{tr}\{\rho H_i\}$, with the representation of the Hamiltonians $\{H_i\}_{i=1}^m$ in the Pauli basis, which invokes a *fermion-to-spin mapping* such as the *Jordan–Wigner* [34, 35] or *Bravyi-Kitaev encodings* [36, 37]. We define the classical shadow channel as $\mathcal{M}$, which involves operating the unitary channel $\mathcal{U}$ uniformly randomly sampled from the Clifford group before measurements in the Z-basis measurements and classical postprocessing operations on the measurement outcomes. By performing the inverse of the classical shadow channel $\mathcal{M}^{-1}$ on the resulting state after performing $\mathcal{M}$ on the initial state ρ, we obtain the classical shadow representation $\hat{\rho}$ of the quantum state ρ, allowing for the calculation of the expected values of observables $\{H_i\}_{i=1}^m$ with respect to ρ using classical methods.

While the classical shadow algorithm requires exponentially many copies even for some local interacting fermions due to the inefficient representation in the qubit system, recently, several classical shadow algorithms for fermionic systems without encoding of the Hamiltonians have been proposed [8–10]. Zhao, Rubin, and Miyake [8] utilize the generalized CS method [21] for fermionic systems, and proposed an algorithm that requires $\mathcal{O}\left(\binom{n}{k} k^{3/2}(\log n)/\varepsilon^2\right)$ copies for the unknown quantum states to output all the elements of a k-RDM. Low [9] proves that all elements of the k-RDM can be estimated with $\binom{\eta}{k}(1-\frac{\eta-k}{n})^k \frac{1+n}{1+n-k}/\varepsilon^2$ number of copies of the quantum state, where η is the number of particles and n is the number of modes. These fermionic shadow estimation methods (along with the generic classical shadow formalism) do not account for noise in the system, which is an inevitability in real physical systems.

Since we are still in the noisy intermediate-scale quantum (NISQ) era, current quantum simulators are heavily affected by noise; hence, any characterization technique needs to be robust for these simulators to be useful. For qubit systems, robust shadow estimation was developed [11, 12] where Chen et al [11] use techniques from randomized benchmarking to mitigate the effect of gate-independent time-stationary Markovian (GTM) noise channels on the procedure. Jnane et al. [38] proposed error-mitigated classical shadow with probabilistic error cancellation.

Utilizing the robust shadow estimation scheme and taking inspiration from the fermionic shadow estimation of Zhao et al [8, 10], we present an error-mitigated shadow estimation scheme for fermionic systems and demonstrate its feasibility for realistic noise channels. Note that akin to the fermionic CS approaches proposed in Refs. [8, 10], our error-mitigated CS method circumvents the need to encode the Hamiltonian using the qubit representation.

We sample our unitaries $\mathcal{U}_Q$ from the matchgate group [39], a natural choice for our protocol as there is a one-to-one correspondence between two-qubit matchgates and free-fermionic evolution [40, 41]. We therefore design the classical postprocessing operations by leveraging the irreducible representation of the matchgate group. We successfully introduce an unbiased estimator $\widehat{\mathcal{M}}$ for the noisy classical shadow channel $\widetilde{\mathcal{M}}$, where we require an additional calibration protocol to generate the estimator $\widehat{\mathcal{M}}$ with the assumption that the computational basis state $|\mathbf{0}\rangle\langle\mathbf{0}|$ can be prepared noiselessly. Additionally, we demonstrate the efficacy of our protocol under conditions of constant noise strength by evaluating its performance across various common noise channels: depolarizing noise, generalized amplitude damping, X-rotation, and Gaussian unitary noise. The number of samples required for the estimation process of our protocol is in the same order as the noise-free matchgate classical shadow scheme [8, 10].

We determine the effectiveness of our protocol with the above noise models by calculating the expecta-

TABLE 1. Enumeration of the classical shadow protocols in noiseless and noisy settings, for qubit and fermionic systems respectively.

	Clifford-based shadows	Fermionic shadows
Noiseless	Huang et al. [21]	Zhao et al. [8]; Low [9]; Wan et al. [10]
Noisy	Chen et al. [11]; Koh and Grewal [12]	This work

tions of all elements of the k-particle reduced density matrix (k-RDM) when the noise strength is constant. The number of samples required for estimation, in this case, is $\mathcal{O}\left(kn^k \ln(n/\delta_e)/\varepsilon_e^2\right)$ and for calibration is $\mathcal{O}\left(\sqrt{n}\ln n \ln(1/\delta_c)/\varepsilon_c^2\right)$ with error $\varepsilon_e + \varepsilon_c$ and success probability $(1-\delta_e)(1-\delta_c)$.

We have extended the analysis of our error-mitigated fermionic shadow channel estimation to more general physical quantities inspired by the fermionic shadow analysis of Wan et al. [10], with more details in Supplementary Notes 5, 9. We list distinct classical shadow approaches in both noiseless and noisy qubit and fermionic systems in Table 1. Our error-mitigated fermionic classical shadow technique constitutes an extension of the work by Chen et al. [11], accommodating scenarios where the gate-set lacks (1) 3-design properties [13] and (2) the applicability of the randomized benchmarking scheme developed by Helsen et al. [14].

We tested the accuracy and efficacy of our protocol by performing numerical experiments to estimate $\mathrm{tr}\{\rho\widetilde{\gamma}_S\}$ (where $\widetilde{\gamma}_S = U_Q^\dagger \gamma_S U_Q$, where γ_S is the product of $|S|$ Majorana operators and plays a crucial role in computing k-RDMs) on a noisy quantum device subjected to various types of gate noise such as depolarizing, generalized amplitude damping, X-rotation, and Gaussian unitaries. Our numerical investigations confirm the potential of our methods in real-world experimental scenarios.

RESULTS

Basic notations and concepts

Here we give the basic notations and concepts that will be used throughout this work.

Basic notations. The symbols X, Y, and Z denote the Pauli X, Y, and Z operators respectively. The operator $R_X(\theta) = \exp\left(-i\frac{\theta}{2}X\right)$ denotes the rotation operator around the x-axis. A Z-basis measurement is performed with respect to the basis of eigenstates of the Pauli-Z operator. We utilize $\mathbb{I}$ to represent the identity operator on the full system. The set of linear operators on a vector space $\mathcal{H}$ is denoted as $\mathcal{L}(\mathcal{H})$. We utilize the symbol $\widetilde{\mathcal{O}}$ to omit the logarithmic terms.

Superoperator. We denote the superoperator representation of a linear operator $O \in \mathcal{L}(\mathcal{H})$ as $|O\rangle\!\rangle := O/\sqrt{\mathrm{tr}\{OO^\dagger\}}$ and the scaled Hilbert-Schmidt inner product between linear operators as $\langle\!\langle O|R\rangle\!\rangle = \mathrm{tr}\{O^\dagger R\}/\sqrt{\mathrm{tr}\{OO^\dagger\}\,\mathrm{tr}\{RR^\dagger\}}$. The action of a channel $\mathcal{E} \in \mathcal{L}(\mathcal{L}(\mathcal{H}))$ operating on a linear operator $O \in \mathcal{L}(\mathcal{H})$ can hence be written as $\mathcal{E}|O\rangle\!\rangle = \mathcal{E}(O)/\sqrt{\mathrm{tr}\{OO^\dagger\}}$. The channel representation of a measurement with respect to the computational basis can be represented as $\mathcal{X} = \sum_{x\in\{0,1\}^n} |x\rangle\!\rangle\langle\!\langle x|$. We denote the unitary channel corresponding to the unitary operator U as $\mathcal{U}(\cdot) = U(\cdot)U^\dagger$.

Majorana operator. The Majorana operators γ_j for $1 \le j \le 2n$ describes the fermionic system with $\gamma_j = b_{(j+1)/2} + b_{(j+1)/2}^\dagger$ for odd j and $\gamma_j = -i(b_{j/2} - b_{j/2}^\dagger)$ for even j, where b_j and $b_j^\dagger$ are the annihilation and creation operators, respectively, associated with the j-th mode. Let γ_S be the product of the Majorana operators indexed by the subset S, denoted as $\gamma_S = \gamma_{l_1}\cdots\gamma_{l_{|S|}}$ for $|S| > 0$ and $\gamma_\emptyset = \mathbb{I}$, where $S = \{l_1, \ldots, l_{|S|}\}$ and $l_1 < l_2 < \ldots < l_{|S|}$. It can be shown that γ_S forms the complete orthogonal basis for $\mathcal{L}(\mathcal{H})$ for $S \subseteq [2n]$. Let $\Gamma_k := \{\gamma_S | |S| = k\}$ be the subspace of γ_S with cardinality k. We denote the

even subspace as $\Gamma_{\text{even}} := \bigoplus_l \Gamma_{2l}$. Also, we denote $\mathcal{P}_k$ as the projector onto the subspace Γ_k, i.e.

$$\mathcal{P}_k = \sum_{S \in \binom{[2n]}{k}} |\gamma_S\rangle\!\rangle\langle\!\langle \gamma_S|, \tag{4}$$

where we have used the notation that for a set A and an integer k, $\binom{A}{k} = \{T \subseteq A : |T| = k\}$ denotes the set of subsets of A with cardinality k.

Gaussian unitaries. Matchgates are in a one-to-one correspondence with the fermionic Gaussian unitaries and can serve as a qubit representation for these unitaries. We denote $\mathbb{M}_n$ as the matchgate group, and write its elements $U_Q \in \mathbb{M}_n$ in terms of rotation matrices Q belonging to the orthogonal group $\mathrm{Orth}(2n)$ (see Supplementary Note 1 for details) [40, 42]. Following Wan et al.'s study [10], which demonstrated that the continuous matchgate group $\mathbb{M}_n$ and the discrete subgroup $\mathbb{M}_n \cap \mathrm{Cl}_n$ (where Cl_n represents the Clifford group) deliver equivalent performances for fermionic classical shadows, our findings remain applicable to both continuous and discrete matchgate circuits. Since $U_Q^\dagger \gamma_j U_Q = \sum_k Q_{jk}\gamma_k$, the matchgate U_Q transforms the product of Majorana operators γ_S in the $\Gamma_{|S|}$ subspace as $U_Q^\dagger \gamma_S U_Q = \sum_{S' \in \binom{[2n]}{|S|}} \det(Q|_{SS'})\gamma_{S'}$.

k-particle reduced density matrices (k-RDM). We denote a k-RDM as ${}^{k}\mathbf{D}$, which can be obtained by tracing out all but k particles. Here we denote it as a tensor with $2k$ indices,

$${}^{k}\mathbf{D}_{j_1,\ldots,j_k;l_1,\ldots,l_k} = \mathrm{tr}\Big\{\rho b_{j_1}^\dagger \cdots b_{j_k}^\dagger b_{l_1} \cdots b_{l_k}\Big\}, \tag{5}$$

where j_i and l_i are in $[n]$ for $i \in [k]$. The fermionic system can be equivalently described in the Majorana basis, in which case a tensor can be rewritten as the linear combinations of $\mathrm{tr}\{\rho\gamma_S\}$, and $|S| \leq 2k$. Hence all n^{2k} elements of the k-RDM can be obtained by calculating $\mathrm{tr}\{\rho\gamma_S\}$, for the scaling of $\mathcal{O}\left(n^k\right)$ different S with $|S| \leq 2k$ [43].

Pfaffian function. The Pfaffian of a matrix $Q \in \mathbb{R}^{2n\times 2n}$ is defined as

$$\mathrm{pf}(Q) = \frac{2^n}{n!} \sum_{\sigma \in S_{2n}} \mathrm{sgn}\,(\sigma) \prod_{i=1}^{n} Q_{\sigma_{2i-1},\sigma_{2i}}, \tag{6}$$

which can be calculated in $O\left(n^3\right)$ time [44].

Ideal fermionic shadow (Wan et al. [10]) Given an unknown quantum state ρ, the classical shadow method applies a unitary U_Q uniformly randomly sampled from matchgate group $\mathbb{M}_n$, followed by measuring the generated state in the computational basis. With the measurement result $|x\rangle\!\rangle$, we can generate the classical representation $\hat{\rho} = \mathcal{M}^{-1}\mathcal{U}_Q^\dagger |x\rangle\!\rangle$ for the unknown quantum state ρ, where the channel $\mathcal{M}$ describing the overall process is defined as

$$\begin{aligned}\mathcal{M}(\rho) &= \int_Q d\mu(Q) \left[\sum_{x\in\{0,1\}^n} \Big\langle x|U_Q\rho U_Q^\dagger|x\Big|x|U_Q\rho U_Q^\dagger|x\Big\rangle\, U_Q^\dagger\,|x\rangle\langle x|\,U_Q\right] \\ &= \sum_k \frac{\binom{n}{k}}{\binom{2n}{2k}} \mathcal{P}_{2k}(\rho).\end{aligned} \tag{7}$$

Noise assumptions. In this work, we assume that the noise is gate-independent, time-independent, and Markovian (a common assumption in randomized benchmarking (RB) abbreviated as the *GTM noise assumption* [45]) and that the preparation noise for the easily prepared state $|\mathbf{0}\rangle\langle\mathbf{0}|$ is negligible. For the convenience of calculation, we utilize the left-hand side noisy representation for a noisy fermionic unitary $\widetilde{\mathcal{U}}_Q := \Lambda\mathcal{U}_Q$. Here we define *the average fidelity in* Γ_{2k} *subspace* for noise channel Λ as

$$\mathcal{B}_k := \frac{(-i)^k}{2^n\binom{n}{k}} \sum_x \sum_{S\in\binom{[n]}{k}} (-1)^{x_S}\,\mathrm{tr}\big\{|x\rangle\langle x|\,\Lambda(\gamma_{D(S)})\big\}, \tag{8}$$

TABLE 2. Comparison between average noise fidelity F_{ave}, Z-basis average noise fidelity F_Z, and average noise fidelity in Γ_{2k} subspace $\mathcal{B}_k$.

Noise type	Λ_{d}	Λ_{a}	Λ_{r}
Λ_{avg}	$1-p/2$	$2/3-(p_0+p_1)/6+\sqrt{(1-p_0)(1-p_1)}/3$	$(\cos(\theta)+2)/3$
F_Z	$1-p/2$	$1-(p_0+p_1)/2$	$\cos^2(\theta/2)$
$\mathcal{B}_1$	$1-p$	$1-(p_0+p_1)$	$\cos\theta$

where $D(S)=\{2j-1,2j|j\in S\}, 0\leq k\leq n, x_S=\sum_{j\in S}x_j$. It is easy to check that $\mathcal{B}_k=1$ if $k=0$. With some calculations, we have $\mathcal{B}_k=1$ for the noise-free model where noise channel Λ equals the identity. In the following, we give the analysis of the simplified result for $\mathcal{B}_k$ for several common noise models in the qubit system and fermionic system when $k>1$. See more details for the analysis in Supplementary Note 3.

(1) The depolarizing noise with channel representation $\Lambda_{\text{d}}(A)=(1-p)A+p\,\text{tr}\{A\}\frac{\mathbb{I}}{2^n}$ for any n-qubit linear operator A, where the noise strength $p\in[0,1]$, and $\mathcal{B}_k=1-p$.

(2) The generalized amplitude-damping noise with the Kraus representation

$$\Lambda_{\text{a}}(\cdot)=\sum_{\substack{u,v\in\{0,1\}^n\\ u\neq v}}E_{uv}(\cdot)E_{uv}^\dagger+E_0(\cdot)E_0^\dagger, \tag{9}$$

where $E_{uv}=\sqrt{\bar{p}_u}\,|v\rangle\langle u|$ for $u\neq v\in\{0,1\}^n$ and $E_0=\sqrt{\mathbb{I}-\sum_{\substack{u,v\in\{0,1\}^n\\ u\neq v}}E_{uv}^\dagger E_{uv}}$, where the probabilities $\bar{p}_u$ satisfy $(2^n-1)\bar{p}_u\leq 1$ for any $u\in\{0,1\}^n$. The average fidelity $\mathcal{B}_k=1-\sum_{u\in\{0,1\}^n}\bar{p}_u$ if $k\neq 0$. We let $\sum_{u\in\{0,1\}^n}\bar{p}_u$ denote the noise strength.

(3) The X-rotation noise with the channel representation

$$\Lambda_{\text{r}}(\cdot)=R_X(\boldsymbol{\theta})(\cdot)R_X(-\boldsymbol{\theta}) \tag{10}$$

where $R_X(\boldsymbol{\theta})=\exp(-i\sum_{l=1}^n\theta_l X_l/2)$, where the noise strengths θ_l are some real numbers. By some calculations, we have $\mathcal{B}_k=\binom{n}{k}^{-1}\sum_{S\in\binom{[n]}{k}}\prod_{l\in S}\cos\theta_l$. Hence $\min_l\cos^k\theta_l\leq\mathcal{B}_k\leq\max_l\cos^k\theta_l$.

(4) Noise that is a Gaussian unitary [46, 47], where we assume that the noise has no coherence with the environment. This noise channel is denoted as

$$\Lambda_{\text{g}}(\cdot)=U_Q(\cdot)U_Q^\dagger, \tag{11}$$

where U_Q is a Gaussian unitary. By selecting the noise model to be Λ_{g}, we get $\mathcal{B}_k=\binom{n}{k}^{-1}\sum_{S,S'\in\binom{[n]}{k}}\det\big(Q|_{D(S),D(S')}\big)$.

Note that for the noise models defined in (1–3), the average fidelity $\mathcal{B}_k\in[0,1]$ is close to one when the noise strengths are close to zero. For comparison, it is worth noting that the standard average noise fidelity [48] $F_{\text{avg}}=\int_\psi d\psi\,\langle\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi|\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle$ where ψ is drawn from the Haar measure, and the Z-basis average noise fidelity defined in Chen et al. [11], $F_Z=\frac{1}{2^n}\sum_b\langle\langle b|\Lambda|b\rangle\rangle$ are not equivalent to $\mathcal{B}_k$ under the same noise model. We present a comparison of these three quantities for $\Lambda_{\text{d}},\Lambda_{\text{a}},\Lambda_{\text{r}}$ for a single qubit, as depicted in Table 2. They are closely aligned, with $\mathcal{B}_1$ slightly smaller than F_{avg} and F_Z. We give a more detailed analysis in Supplementary material X.

Mitigation algorithm and error analysis

Let $\widehat{\mathcal{M}}:=\sum_{k=0}^n\hat{f}_{2k}\mathcal{P}_{2k}$ be the estimator for the noisy shadow channel $\widetilde{\mathcal{M}}=\sum_{k=0}^n f_{2k}\mathcal{P}_{2k}$. In the Methods Section, we provide an explicit expression and efficiency for the calculation of $\hat{f}_{2k}$. Using the estimated noisy channel $\widehat{\mathcal{M}}$, we can now obtain an estimate for $\{\text{tr}\{\rho H_j\}\}_{j=1}^m$, where ρ represents certain

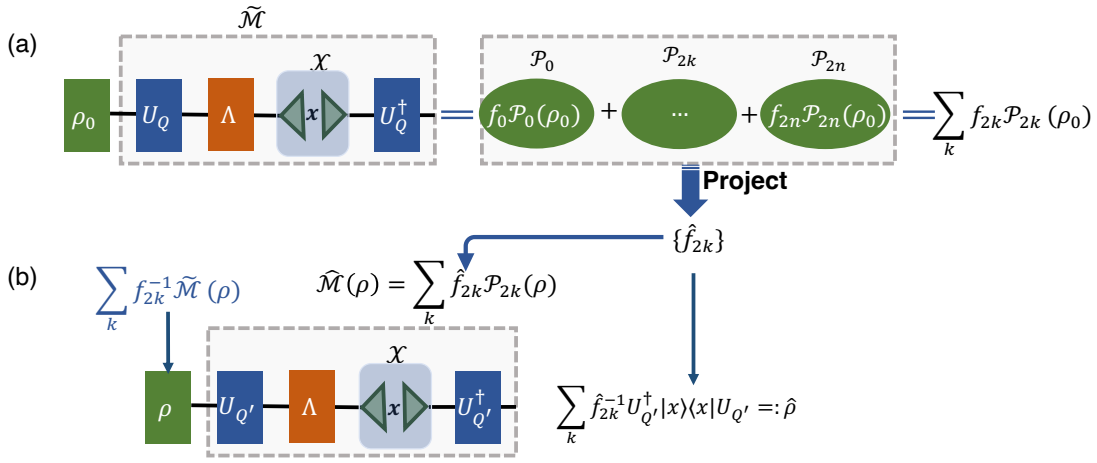

FIG. 1. Schematic diagram of the error-mitigated matchgate classical shadow algorithm. (a) Protocol to learn the noisy classical shadow channel, where $\widetilde{\mathcal{M}} = \mathcal{U}_Q^\dagger \circ \mathcal{X} \circ \Lambda \circ \mathcal{U}_Q(\cdot) = \sum_{k=0}^n f_{2k}\mathcal{P}_{2k}(\cdot)$, where $\mathcal{U}_Q$ is uniformly randomly sampled from the matchgate group. The estimation $\hat{f}_{2k}$ for f_{2k} is obtained by projecting $\widetilde{\mathcal{M}}$ to $\mathcal{P}_{2k}$ subspace with input state $\rho_0 = |\mathbf{0}\rangle\langle\mathbf{0}|$. (b) The shadow estimation $\hat{\rho}$ for input state ρ with the noisy quantum circuit and classical post-processing with the approximated noisy classical shadow channel in (a), where $\mathcal{U}_{Q'}$ is uniformly randomly sampled from the matchgate group.

quantum states and H_j denotes certain observables. If $f_{2k} = 0$, the channel $\widetilde{\mathcal{M}}$ becomes non-invertible. Consequently, the effectiveness of the fermionic CS method diminishes, and we cannot retrieve $\mathrm{tr}(\rho H)$ using it. In this study, we operate under the assumption that the noise is permissible and the fermionic CS channel is consistently invertible. Additionally, we provide a scenario in Supplementary Note 3 where the extreme noise channel occurs, i.e., $f_k = 0$. However, it is anticipated that such an extreme case will rarely occur. Algorithm 1 demonstrates the method for mitigated estimation.

Incorporating the **MedianOfMeans** sub-procedure, as explained in Ref. [21], guarantees that the number of quantum state copies needed relies on the logarithm of the number of observables. We included the **MedianOfMeans** sub-procedure in Supplementary Note 6 to ensure the completeness and consistency of this paper. Our error analysis will involve selecting appropriate values for the number of calibrations and estimation samplings N_c, N_e, K_c, and K_e to estimate the coefficients $\hat{f}_{2k}$ associated with the noisy channel.

Let $\hat{v} := \mathrm{tr}\left\{\widehat{\mathcal{M}}^{-1}\mathcal{U}^\dagger(|x\rangle\langle x|)H\right\}$ be an estimation of $\mathrm{tr}\{\rho H\}$ for some observables H in the even subspace Γ_{even} and quantum states ρ, where x follows the distribution $\mathrm{tr}\left\{|x\rangle\langle x|\widehat{\mathcal{U}}(\rho)\right\}$ for $x \in \{0,1\}^n$, then we have

$$\begin{aligned}|\hat{v} - \mathrm{tr}\{\rho H\}| &\le \left|\hat{v} - \mathrm{tr}\left\{\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\, H\right\}\right| \\ &\quad + \left|\mathrm{tr}\left\{\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\, H\right\} - \mathrm{tr}\{\rho H\}\right| \\ &= \varepsilon_e + \varepsilon_c, \end{aligned} \tag{12}$$

where $\varepsilon_e := \left|\hat{v} - \mathrm{tr}\left\{\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right\}\right|$ is the estimation error and $\varepsilon_c := \left|\mathrm{tr}\left\{\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\, H\right\} - \mathrm{tr}\{\rho H\}\right|$ is the calibration error. Therefore, by determining the necessary number of samples N_e and K_e to achieve the desired level of estimation error ε_e (as well as N_c and K_c to account for the calibration error ε_c), we can obtain an estimation $\hat{v}$ with an overall error of $\varepsilon = \varepsilon_e + \varepsilon_c$, using $N_e K_e$ copies of the input state ρ.

Theorem 1. *Let ρ be an unknown quantum state and $\{H_i\}_{i=1}^m$ be a set of observables in the even subspace*

Algorithm 1 Error-mitigated estimation for noisy fermionic classical shadows

Input Quantum state ρ, observables $H_1, \ldots, H_m$, integers N_c, K_c, N_e, K_e.
Output $\hat{v}_i$ for $i \in [m]$.
$R_c := N_c K_c$;
for $j \leftarrow 1$ to R_c **do**
    Prepare state $\rho_0 = |0^n\rangle$, uniformly sample $Q \in \mathrm{Orth}(2n)$ or $\mathrm{Perm}(2n)$, implement the associated noisy Gaussian unitary $\widehat{U}_Q$ on ρ_0, and measure in the Z-basis with outcomes x;
    Let $\hat{f}_{2k}^{(j)} := 2^n(-1)^k \binom{n}{k}^{-1} \langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle, \forall k \in [n]$;
end for
$\hat{f}_{2k} := \mathbf{MedianOfMeans}\left(\left\{f_{2k}^{(j)}\right\}_{j=1}^{R_c}, N_c, K_c\right) \forall k \in [n]$;
$\widehat{\mathcal{M}} := \sum_{k=1}^n \hat{f}_{2k}\mathcal{P}_{2k}$;
$R_e := N_e K_e$;
for $i \leftarrow 1$ to m **do**
    for $j \leftarrow 1$ to R_e **do**
        Prepare ρ, uniformly sample $Q \in O(2n)$ or $\mathrm{Perm}(2n)$, implement the associated noisy Gaussian unitary $\hat{U}_Q$ on ρ, and measure in the Z-basis with outcomes x;
        Generate estimation $\hat{v}_i^{(j)} := \mathrm{tr}\left\{H_i \widehat{\mathcal{M}}^{-1}\mathcal{U}_Q^\dagger(x)\right\}$;
    end for
    $\hat{v}_i := \mathbf{MedianOfMeans}\left(\left\{\hat{v}_i^{(j)}\right\}_{j=1}^{R_e}, N_e, K_e\right)$;
end for
return $\{\hat{v}_i\}_{i=1}^m$

Γ_{even}. *Consider Algorithm 1 with the number of estimation samplings*

$$R_e = \frac{68(1+\varepsilon_c)^2 \ln(2m/\delta_e)}{2^{2n}\varepsilon_e^2} \sum_{0 \le l_1+l_2+l_3 \le n} g(l_1,l_2,l_3) \sum_{\substack{S_1,S_2,S_3 \text{ disjoint} \\ |S_i|=2l_i \forall i \in [3]}} \mathrm{tr}\{\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\}\,\mathrm{tr}\{\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0\}\,\mathrm{tr}\{\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\},$$

where $g(l_1,l_2,l_3) = \frac{(-1)^{l_1+l_2+l_3}\binom{n}{l_1,l_2,l_3}_p \binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}\mathcal{B}_{l_1+l_3}}{\binom{2n}{2l_1,2l_2,2l_3}_p \binom{n}{l_1+l_2}\binom{n}{l_2+l_3}\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}}$, *and* $H_0 = \max_i(H_i - \mathrm{tr}\{H_i\}\frac{\mathbb{I}}{2^n})$, *and the number of calibration samplings*

$$R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n \ln(1/\delta_c)}{\mathcal{B}_{\min}^2\varepsilon_c^2}\right),$$

where $\mathcal{B}_{\max} = \max_k |\mathcal{B}_k|$ *and* $\mathcal{B}_{\min} = \min_k |\mathcal{B}_k|$. *Then, the outputs* $\{v_i\}_{i=1}^m$ *of the algorithm approximate* $\{\mathrm{tr}\{\rho H_i\}\}_{i=1}^m$ *with error* $\varepsilon_e + \varepsilon_c$ *and success probability* $1 - \delta_e - \delta_c$, *under the assumption that* $\|H_i\|_\infty = \mathcal{O}(1)$, *where* $\|H_i\|_\infty$ *is the spectral norm of* H_i.

We observe that the sampling for estimation we obtained is consistent with Wan et al. [10] in the absence of noise. In the following, we will provide an analysis of the necessary number of measurements to compute $\langle {}^k\mathbf{D}|{}^k\mathbf{D}\rangle$ using Algorithm 1. To calculate the representation of each element $\langle {}^k\mathbf{D}|{}^k\mathbf{D}\rangle_{j_1,\ldots,j_k;l_1,\ldots,l_k}$, where j_i, l_i are in the range $[n]$ for $i \in [k]$, we need to calculate $m = \mathcal{O}\left(n^k\right)$ expectations for different $\widetilde{\gamma}_S$, where $|S| = 2k$. By choosing the observable $H_j = \widetilde{\gamma}_S$ where $|S| = 2k$ and $m = \mathcal{O}\left(n^k\right)$ in Theorem 1, with the number of estimation samplings

$$R_e = \mathcal{O}\left(\frac{kn^k \ln(n/\delta_e)}{\mathcal{B}_k^2\varepsilon_e^2}\right) \tag{13}$$

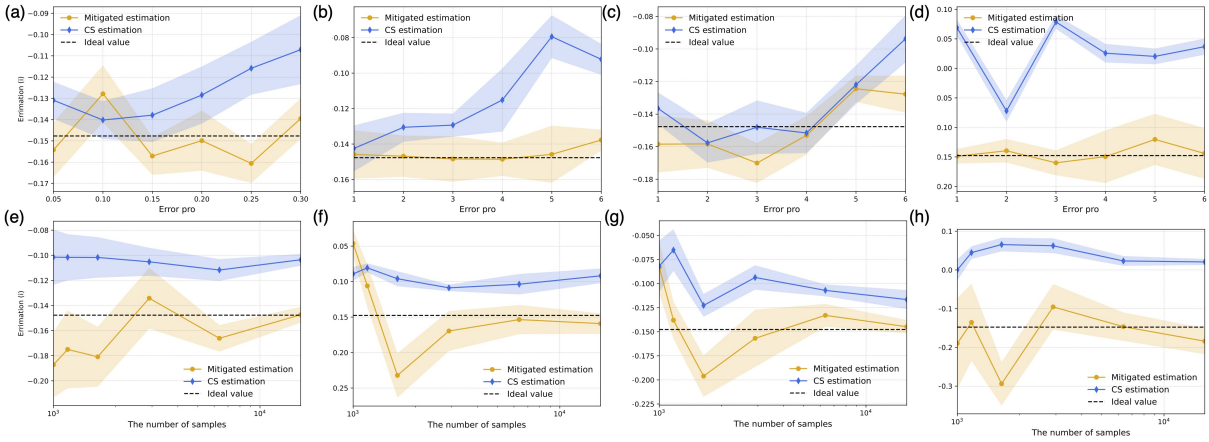

FIG. 2. The estimations of the expectation values of Majorana operators. (a-d) The estimations with the changes of the noise strength and (e-h) the number of samples for fixed noise parameters, where (a) and (e) are associated with depolarizing noise, (b) and (f) are associated with generalized amplitude damping noise, (c) and (g) are associated with X-rotation noise, (d) and (h) are associated with Gaussian unitary noise. The error bar is the estimation of the standard deviation by repeating the procedure for $R = 10$ rounds, and it is estimated to be $\sqrt{\sum_{i=1}^{R}(\hat{v}_i - \bar{v})^2/R}$, where $\bar{v}$ is the expectation of the corresponding estimations. Specially, $\bar{v} = \frac{f_{|S|}}{g}\,\mathrm{tr}\{\rho\gamma_S\}$, $g = \hat{f}_{|S|}$ for mitigated estimation, and $g = \frac{\binom{n}{k/2}}{\binom{2n}{k}}$ for CS estimation. In Fig. 2 (a-d), the x-axis represents the labels for different noise parameters. The dotted horizontal line indicates the value $\mathrm{tr}\{\rho\widetilde{\gamma}_S\}$.

and the number of calibration samplings

$$R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n\ln(1/\delta_c)}{\mathcal{B}_{\min}^2\varepsilon_c^2}\right), \tag{14}$$

the estimation error can be bounded to $\varepsilon_e + \varepsilon_c$. The equations for R_e and R_c can be simplified to $R_e = \mathcal{O}\left(\frac{kn^k\ln(n/\delta_e)}{\varepsilon_e^2}\right)$ and $R_c = \mathcal{O}\left(\frac{\sqrt{n}\ln n\ln(1/\delta_c)}{\varepsilon_c^2}\right)$ for the general noises with constant average fidelity $\mathcal{B}_k$ in subspace Γ_{2k} for any $k \in \{0,\ldots,n\}$. We give more details for the calculations in Supplementary Note 8. However, some types of noise channels, such as certain Gaussian unitary channels present in the related $2n \times 2n$ matrix Q, cannot be mitigated with our mitigation algorithm. In particular, there exists a signed permutation matrix Q for which $f_{2k} = 0$, resulting in complete loss of projection for the observable onto the subspace Γ_{2k}. As a result, it is impossible to calculate $\mathrm{tr}\{\rho\widetilde{\gamma}_S\}$ for any set S containing $2k$ elements. We anticipate that the noise in the quantum device will differ significantly from the U_Q which belongs to the intersection of the matchgate and Clifford groups.

Numerical results

Here we give the numerical results for the mitigated shadow estimation in the fermionic systems. Since the elements of a k-RDM can be expressed in the form $\mathrm{tr}\{\rho\gamma_S\}$, we give the numerical results of the errors of the estimators for the expectation value of local fermionic observables $\mathrm{tr}\{\rho\widetilde{\gamma}_S\}$.

The estimator $\hat{v}$ for $\mathrm{tr}\{\rho\widetilde{\gamma}_S\}$ can be represented as in Eq. (43). Here we choose the number of qubits $n = 4$ and $S = \{1,2\}$, and $\widetilde{\gamma}_S = U_Q^\dagger\gamma_S U_Q$ and Q is uniformly randomly chosen from $\mathrm{Perm}(2n)$. The quantum state ρ is a uniformly randomly generated 4-qubit pure state. As shown in Fig. 2, we depict the estimations of classical shadow estimators [10] and our mitigated Algorithm 1, with the changes of noise

strength (Fig. 2(a-d)) and the changes of the number of samples (Fig. 2(e-h)). Here we numerically test the estimations with respect to depolarizing, amplitude damping, X-rotation, and Gaussian unitaries.

The calibration samples in Algorithm 1 for the numerics are $N_c = 4,000$ and $K_c = 20$ for all noise models. The number of samples for the classical shadow method is set as $N_e = 4,000, K_e = 10$, and for Algorithm 1 are set as $N_e = 4,000/(1-p_{\text{noise}})$ and $K_e = 10$, where p_{noise} is the noise strength varying for different noise settings:

(1) Depolarizing noise $\Lambda_{\text{d}}(\rho) = (1-p)\rho + p\frac{\mathbb{I}}{2^n}$, where ρ is any quantum state and p is the depolarizing noise strength. In Fig. 2 (a), p varies from 0.05 to 0.3 ($p = 0.05j$ for x-axis equals j where $j \in [6]$), and in Fig. 2 (b), $p = 0.2$.

(2) Generalized amplitude damping noise Λ_{a} with representation

$$\Lambda_{\text{a}}(\rho) = \sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} E_{uv}\rho E_{uv}^\dagger + E_0\rho E_0^\dagger, \tag{15}$$

where $E_{uv} = \sqrt{p_{uv}}\,|v\rangle\langle u|$ for $u \neq v \in \{0,1\}^n$ and $E_0 = \sqrt{\mathbb{I} - \sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} E_{uv}^\dagger E_{uv}}$. Note that Eq. (15) is a generalization of Eq. (9), which connects to Eq. (9) by setting $p_{uv} = \bar{p}_u$ for any $u, v \in [n]$. Here we uniformly randomly choose p_{uv} in $\frac{[0,1]+j-1}{6\times 2^{n+1}}$, and labeled it as j in the x-axis of Fig. 2 where $j \in [6]$, and choose the generated damping errors for $j = 5$ case in Fig. 2 (b) as the damping errors for Fig. 2 (f).

(3) X-rotation noise Λ_{r} defined in Eq. (10) with noise parameters $\theta_j = \frac{\pi}{2(8-j)}$ for $j \in [6]$, and the noise strength is chosen as $1 - \cos\theta$ in Fig. 2 (c), and Fig. 2 (g) depicts the errors in the X-rotation noise with noise parameter θ_6. The x-axis of the mitigation results of X-rotation noise in Fig. 2 (c) denotes the label for noise parameters $\cos\theta_j$ for j range from 1 to 6.

(4) The Gaussian unitary noise channel $\mathcal{U}_Q$ is chosen such that Q is sampled from the signed permutation group, ensuring that the coefficient f_1 for the noise channel is non-zero. The associated numerical results are shown in Fig. 2 (d) and (h). The number of estimation samplings $N_e = 8000, K_e = 10$ for Fig. 2 (d). Fig. 2 (h) choose the same noise parameter with the fifth noise parameter in Fig. 2 (d). From the figure, we see that without mitigation, the error is enormous with the CS algorithm.

In Fig. 2 (e-h) the number of samples ranges from $\lfloor 900 + 100\exp(j)\rfloor$ for $j \in \{0,\ldots,5\}$. From Fig. 2, we see that with the increase of the noise strength, the classical shadow method with depolarizing, amplitude damping, and X-rotation noise all gradually diverge to the expected value $\text{tr}\{\rho\widetilde{\gamma}_S\}$, while our error-mitigated estimation protocol in Alg. 1 gives an expected value that is close to the noiseless value. Based on the numerical results depicted in Fig.2 (e-h), it is evident that as the number of samples increases, the estimation outcomes generated by Algorithm 1 approaches the expected value $\text{tr}\{\rho\widetilde{\gamma}_S\}$, while the convergent value for the classical shadow method is far from the expected value with depolarizing, amplitude damping, and X-rotation noises. Conversely, the error bar associated with CS estimations is observed to be smaller compared to mitigated estimations, using the same number of samplings, as illustrated in Fig. 2 (e-h). This is due to the variance of the mitigated estimations associated with the average noise fidelities in Γ_{2k} subspace $\mathcal{B}_k$.

DISCUSSION

We present an error-mitigated classical shadow algorithm for noisy fermionic systems, thereby extending matchgate classical shadows for noiseless systems [8, 10]. With our method, the calibration process requires a number of copies of the classical state $|\mathbf{0}\rangle\langle\mathbf{0}|$ that scales logarithmically with the number of qubits. Assuming a constant average noise fidelity for the noise channel, our algorithm requires the same order of estimation copies as the matchgate classical shadow without error mitigation [10]. Our algorithm is applicable for efficiently calculating all the elements of a given k-RDM.

To provide a clearer demonstration of the average fidelity of common noises, we consider depolarizing, amplitude damping, and X-rotation noises. The average fidelity of depolarizing and amplitude damping noises are given by $(1-p)$ and $(1-\sum_u \bar{p}_u)$ respectively, where $p, \bar{p}_u$ are the noise parameters. For X-rotation noise, the average fidelity lies between $[\min_\theta \cos^k\theta, \max_\theta \cos^k\theta]$ where θ is the rotated angles. To evaluate the effectiveness of our algorithm in mitigating these noises, we compare its performance with the matchgate CS algorithm to calculate the expectations of $\widetilde{\gamma}_S$, where $|S| = 2k$, which is crucial for calculating k-RDMs. Our numerical results show good agreement with the theory, validating the effectiveness of our algorithm.

While our algorithm demonstrates good performance in the presence of common types of noise in near-term quantum devices, further investigations are required to explore its potential limitations and improvements. Some of the open questions that can be addressed in future research include:

(1) Is it possible to extend our algorithm to handle other types of noise, such as time-dependent, non-Markovian and environmental noise [49], or more generally, noise that does not satisfy the GTM assumption? If so, how would these different types of noise impact the performance of our algorithm?

(2) We provide numerical results under the assumption of Gaussian unitary noise, a common noise model in the fermionic platform. An intriguing unanswered question pertains to the performance of our algorithm in the presence of more typical noise channels inherent to fermionic platform.

(3) The number of gates required by the matchgate circuit is $\mathcal{O}(n^2)$ [15]. As a result, the accumulation of noise significantly increases the error mitigation threshold [50, 51], which raises the intriguing question of whether it is feasible to provide an error-mitigated classical shadow using a shallower circuit. This may be compared with Bertoni et al. [52], who propose a shallower classical shadow approach for qubit systems.

(4) In addition, we have included in the Supplementary materials the analysis and numerical results regarding the overlap between a Gaussian state and any quantum states, as well as the inner product between a Slater determinant and any pure state. This prompts the question of whether our algorithm can be utilized to calculate other physical, chemical, or material properties beyond the scope of this paper.

Exploring these questions would enhance our comprehension of the potential and limitations of our algorithm, and could potentially pave the way for advancements in the estimation of fermionic Hamiltonian expectation values with near-term quantum devices.

Note added.— Following the completion of our manuscript, we became aware of recent independent work by Zhao and Miyake [53], who also study ways to counteract noise in the fermionic shadows protocol.

Adaptive-depth randomized measurement for fermionic observables

Kaiming Bian, and Bujiao Wu

The accurate estimation of fermionic observables, such as k-RDMs and molecular Hamiltonians, is essential for advancing quantum physics and chemistry. The fermionic classical shadow (FCS) method offers an efficient framework for estimating these observables without requiring a transformation into Pauli basis. However, the random matchgate circuits in FCS require polynomial-depth circuits with a brickwork structure, which presents significant challenges for near-term quantum devices with limited computational resources. To address this limitation, we introduce an adaptive-depth fermionic classical shadow (ADFCS) protocol, designed to reduce the circuit depth while maintaining the accuracy of the estimations. Through theoretical analysis and numerical fitting, we establish that the required depth for approximating a Majorana string γ_S—defined as the product of Majorana operators—scales as $\max\left\{\frac{d_{\text{int}(S)}^2}{\log n}, d_{\text{int}}(S)\right\}$, where $d_{\text{int}}(S)$ represents the minimum interaction distance between the elements of set S. We demonstrate the effectiveness of the ADFCS protocol through numerical experiments, where it shows similar accuracy to traditional FCS methods while requiring significantly fewer resources. Additionally, we apply ADFCS to compute the expectation value of the Kitaev chain Hamiltonian, further validating its performance in practical scenarios. Our findings suggest that ADFCS can enable more efficient quantum simulations, reducing circuit depth while preserving the fidelity of quantum state estimations, offering a viable solution for near-term quantum devices.

INTRODUCTION

Simulation of strongly correlated fermionic systems is one of the most important applications of quantum computing [54–59], highlighting the necessity for advanced measurement techniques in fermionic systems to capture the intricacies of electron correlations and dynamics. An important example is the k-reduced density matrix (k-RDM) [60–62], which efficiently captures the essential features of many-body fermionic systems by focusing on a subset of particles and involves calculating various expectation values of fermionic observables. It demonstrates the importance of the efficient calculations of the expectation values in fermionic systems.

The classical shadow (CS) algorithm [21] was introduced to give a classical estimator for a quantum state ρ by employing a random Clifford circuit, measurement in the computational basis, and followed by polynomial classical postprocessing. It is particularly suited for estimating linear properties of ρ, such as the expectation values $\{\text{Tr}\left(\rho Q_i\right)\}_{i=1}^m$. The number of measurements required is $\mathcal{O}\left(\max_i \|Q_i\|_{\text{shadow}}^2 \log m/\varepsilon^2\right)$, where ε is the desired estimation error and $\|\cdot\|_{\text{shadow}}$ denotes the shadow norm, which depends on the choice of the unitary ensemble. When the unitary ensemble is the global Clifford group, the shadow norm simplifies to $\|Q\|_{\text{shadow}} = \text{Tr}\left(Q^2\right)$ [21]. This CS algorithm becomes inefficient for general fermionic observables involving k-local Majorana strings, such as k-RDM. To address this limitation, fermionic classical shadow (FCS) algorithms [8–10, 63–65] were developed. These algorithms primarily differ from the CS algorithm by using a Gaussian unitary ensemble rather than the Clifford group for randomness. Both Gaussian unitary and Clifford elements require polynomial-size quantum circuits, which become challenging for near-term quantum devices due to issues like gate noise and limited coherence time [66–71]. In response to this, several approaches have been proposed to design shallow-depth CS protocols [72–75]. However, for FCS side, Zhao et al. [8] have proved that there does not exist subgroup $G \subseteq \text{Cl}_n \cap \mathbb{M}_n$ which has a better fermionic shadow norm, where Cl_n is the n-qubit Clifford group and $\mathbb{M}_n$ is the matchgate group. King et al. [76] also demonstrated that single-copy measurements require $\Omega\left(n^{k/2}/\epsilon^2\right)$ copies of ρ in the any classical shadow based protocol to achieve an ϵ-error estimation for k-local Majorana strings. An intriguing open question remains:

Is there a shallow-depth FCS algorithm that is efficient for some specific set of fermionic observables and maintains the same sample complexity as the original FCS algorithm?

In this work, we identify the minimum circuit depth required by the FCS protocol for k-local Majorana strings. Beginning with 2-local Majorana strings, we derive an expression that characterizes the dependence of sample complexity on circuit depth. Building on these findings, we investigate the depth required to minimize sample complexity for general k-local Majorana strings.

Specifically, we propose an adaptive-depth fermionic classical shadow (ADFCS) algorithm for optimizing the depth of the brickwork random matchgate circuit, as shown in Fig. 3(a-b). Similar to the FCS algorithm, ADFCS applies a random matchgate circuit $U_{Q_{d^*}}$ and utilizes classical shadows to estimate the expectation value of an observable H. By Ref. [15], constructing a Haar-random matchgate circuit with a brickwork architecture requires $\Omega(n^2)$ depth. In contrast, ADFCS adaptively selects the depth d^* based on the interaction distance of the observable and constructs the random matchgate circuit $U_{Q_{d^*}}$ accordingly. The depth d^* is linear to $\max\left\{d_{\text{int}}^2/\log n, d_{\text{int}}(S)\right\}$ where d_{int} is the interaction distance of the fermionic observable. This relationship is obtained by modeling the variance as a random walk in response to variations in d, and selecting an appropriated d such that the summation of the random walk probability on certain sites is large enough. For example, we numerically demonstrate that ADFCS with a depth of $d^* = 5$ achieves comparable estimation precision to the FCS method when estimating the expectation value of the Kitaev chain Hamiltonian for $n = 10$ qubits. Additionally, we evaluate the ADFCS method by analyzing the estimation error for several k-Majorana strings with varying interaction distances. The numerical results indicate that adaptively selecting a relatively shallow depth can achieve a sample complexity similar to that of FCS for most k-local Majorana strings.

PRELIMINARY

Notations. We utilize $\mathcal{H}$ to denote the quantum state Hilbert space, and $\mathcal{L}(\mathcal{H})$ to denote the linear operator space which operates on Hilbert space $\mathcal{H}$. Quantum channel $\Lambda \in \mathcal{L}(\mathcal{L}(\mathcal{H}))$ is a superoperator that maps a linear operator into another. For convenience, we introduce the Pauli-transfer matrix (PTM) representation, which expresses both superoperators and operators in terms of their action on the Pauli basis. A non-zero linear operator $|A\rangle\!\rangle$ can be vectorized as $|A\rangle\!\rangle := \frac{A}{\sqrt{\text{Tr}\left(A^\dagger A\right)}}$ where the normalization ensures that it is properly scaled. The inner product between the vectorized formation of two operators $A, B \in \mathcal{L}(\mathcal{H})$ is defined as $\langle\!\langle A|B\rangle\!\rangle := \frac{\text{Tr}\left(A^\dagger B\right)}{\sqrt{\text{Tr}\left(A^\dagger A\right)\text{Tr}\left(B^\dagger B\right)}}$. Specifically, for the Majorana observable γ_S, the vectorized form is $|\gamma_S\rangle\!\rangle := \frac{\gamma_S}{\sqrt{2^n}}$, where n is the number of qubits. The superoperator Λ can be represented as a matrix in Pauli basis. Its operating on a linear operator $A \in \mathcal{L}(\mathcal{H})$ can be denoted as $\Lambda|A\rangle\!\rangle := \frac{\Lambda(A)}{\sqrt{\text{Tr}\left(A^\dagger A\right)}}$. Further details on the Pauli-transfer matrix representation can be found in Appendix .

Majorana operators and matchgate circuits. The Majorana operators can be defined as

$$\gamma_{2j-1} := a_j + a_j^\dagger, \quad \gamma_{2j} := -\mathsf{i}\left(a_j - a_j^\dagger\right), \tag{16}$$

where $a_j, a_j^\dagger$ are the annihilation and creation operators for the j-th site respectively. Observables in fermionic systems can be expressed as the linear combination of the products of Majorana operators [77]

$$\gamma_S := \gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_{|S|}}, \tag{17}$$

where $i_1 < i_2 < \cdots < i_{|S|}$, and $S = \{i_1, i_2, \cdots, i_{|S|}\}$. Especially, we set $\gamma_\emptyset$ be the identity operator $\mathbb{1}$ when S is the empty set $\emptyset$. A Majorana operator can be represented in Pauli formation as $\gamma_{2j-1} = \left(\prod_{i=1}^{j-1} Z_i\right) X_j, \quad \gamma_{2j} = \left(\prod_{i=1}^{j-1} Z_i\right) Y_j$, for any $j \in [n]$.

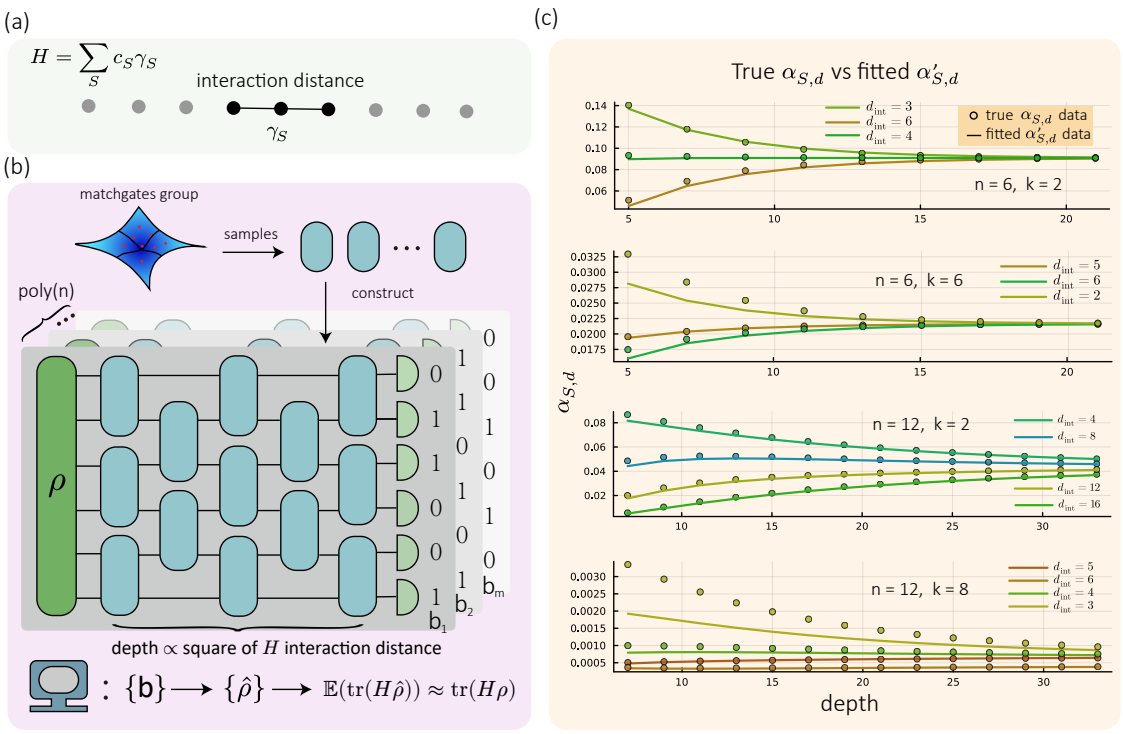

FIG. 3. Illustration of Adaptive-Depth Fermionic Classical Shadow (ADFCS). (a) The interaction distance of the observable H. The Majorana operator γ_S in H can be interpreted as the interaction between particles. The interaction distance d_{int} is defined as the maximum range of interaction of the operators γ_S in H. (b) Sketch of the ADFCS protocol. Unitary gates are independently sampled from the group of two-qubit matchgates within a brickwork architecture. The measured bitstrings b are processed on a classical computer to generate the classical representation of ρ, and $\mathrm{Tr}(H\rho)$ can be estimated using the generated classical estimator $\hat{\rho}$. The depth of the random circuit is determined by the fermionic Hamiltonian $H = \sum c_S \gamma_S$. (c) Illustration of $\alpha'_{S,d}$ and $\alpha_{S,d}$ with the increase of depth d for different number of qubits n and Majorana strings γ_S.

Except for the standard Majorana operators in Eq. (16), there are different bases to express the fermionic Hamiltonian. Different bases could be obtained by the orthogonal transformation of the standard Majorana operators $\tilde{\gamma}_\mu = \sum_{\nu=1}^{2n} Q_{\mu\nu}\gamma_\nu$, where Q is a real orthogonal matrix, $Q \in O(2n)$. The different sets of bases $\{\tilde{\gamma}_\mu\}$ are also called the Majorana operators because they satisfy the same anti-commutation relation $\{\tilde{\gamma}_\nu, \tilde{\gamma}_\mu\} = 2\delta_{\mu\nu}$. The orthogonal transformations of Majorana operators can be represented by unitaries

$$U_Q^\dagger \gamma_\mu U_Q = \sum_{\nu=1}^{2n} Q_{\mu\nu}\gamma_\nu = \tilde{\gamma}_\mu, \tag{18}$$

where the unitaries U_Q are called fermionic Gaussian unitaries. Followed by Eq. (18), the fermionic Gaussian unitary preserves the cardinality $|S| = \#\{i \mid i \in S\}$ of γ_S, which can be expressed as $U_Q^\dagger \gamma_S U_Q = \sum_{S' \in \binom{[2n]}{|S|}} \det\left(Q|_{S,S'}\right) \gamma_{S'}$. The set $\binom{[2n]}{|S|}$ represents the collection of all subsets of $\{1, 2, \ldots, 2n\}$ that contain exactly $|S|$ elements, and the matrix $A|_{S,S'}$ is obtained from the matrix A by selecting the rows indexed by the set S and the columns indexed by the set S'. For simplicity, we refer to a Majorana string γ_S with $|S| = k$ as a *k-local Majorana string* throughout this manuscript.

Matchgate circuits are the quantum circuit representation of fermionic Gaussian unitaries [10]. The collection of all n-qubit matchgate circuits constitutes the matchgate group $\mathbb{M}_n$. This group is generated by rotations of the form $\exp\left(\mathrm{i}\theta X_\mu X_\nu\right), \exp\left(\mathrm{i}\theta Z_\mu\right)$ and X_n, where θ is a real parameter, and μ, ν are indices corresponding to nearest-neighbor qubits in a linear arrangement. Notably, operations within the matchgate

group can be efficiently simulated on a classical computer. This efficiency arises because the action of a matchgate circuit U_Q corresponds to Givens rotations associated with an orthogonal matrix Q, facilitating polynomial-time classical simulation [68, 78].

We present the twirling of the matchgate group in the following lemma, as it is utilized in the proof of our main results.

Lemma 1 (The three moments of uniform distribution in $\mathbb{M}_n$, Ref. [10]). *The k-moment twirling $\mathcal{E}^{(j)}$ is defined by*

$$\mathcal{E}^{(j)}(\cdot) := \int_{U_Q \in \mathbb{M}_n} \mathrm{d}U_Q \; U_Q^{\otimes j}(\cdot)U_Q^{\dagger \otimes j}. \tag{19}$$

The first three moments are

$$\begin{aligned}
\mathcal{E}^{(1)} &= |\mathbb{1}\rangle\!\rangle\langle\!\langle \mathbb{1}| \\
\mathcal{E}^{(2)} &= \sum_{k=0}^{2n} \binom{2n}{k}^{-1} \sum_{\substack{S,S' \subseteq [2n] \\ |S|=|S'|=k}} |\gamma_S\rangle\!\rangle|\gamma_S\rangle\!\rangle\langle\!\langle\gamma_{S'}|\langle\!\langle\gamma_{S'}| \\
\mathcal{E}^{(3)} &= \sum_{\substack{k_1,k_2,k_3 \geq 0 \\ k_1+k_2+k_3 \leq 2n}} |\mathcal{R}\rangle\!\rangle\langle\!\langle\mathcal{R}|,
\end{aligned} \tag{20}$$

where

$$|\mathcal{R}\rangle\!\rangle = \binom{2n}{k_1, k_2, k_3, 2n-k_1-k_2-k_3}^{-1/2} \sum_{\substack{A,B,C \text{disjoint} \\ |A|=k_1,|B|=k_2,|C|=k_3}} |\gamma_A\gamma_B\rangle\!\rangle|\gamma_B\gamma_C\rangle\!\rangle|\gamma_C\gamma_A\rangle\!\rangle. \tag{21}$$

(Fermionic) classical shadow protocol. The classical shadow (CS) process involves randomly applying a unitary $\mathcal{U}$ from some unitary ensemble $\mathbb{U}$ (such as the Clifford group) on a quantum state ρ, followed by measurement in computational bases, yielding outcomes $|b\rangle\!\rangle$ that form a classical description of ρ, known as "*classical shadow*", denoted as $|\hat{\rho}\rangle\!\rangle = \mathcal{M}_{\mathrm{CS}}^{-1}\mathcal{U}^{-1}|b\rangle\!\rangle$, where $\mathcal{M}_{\mathrm{CS}} := \mathbb{E}_{\mathcal{U}\in\mathbb{U}}\left[\mathcal{U}^\dagger \sum_b |b\rangle\!\rangle\langle\!\langle b|\mathcal{U}\right]$. By choosing $\mathbb{U}$ as Clifford group Cl_n, the classical shadow channel can be simplified to $\mathcal{M}_{\mathrm{CS}} = \Pi_0 + \frac{1}{2^n+1}\Pi_1$, where Π_0 is the projector onto the identity subspace and Π_1 is the projector onto the subspace spanned by all Pauli bases except identity [11, 21]. This simple representation allows efficient construction of the classical shadow $\hat{\rho}$.

In the framework of FCS, the shadow channel $\mathcal{M}_{\mathrm{FCS}} := \mathbb{E}_{Q\in O(2n)}\left[\mathcal{U}_Q^\dagger \sum_b |b\rangle\!\rangle\langle\!\langle b|\mathcal{U}_Q\right]$ is not invertible across the entire Hilbert space. Its invertibility is constrained to a subspace spanned by even operators, denoted as $\Gamma_{\text{even}} := \mathrm{span}\{\gamma_S \mid |S| = 2j, j = 0, 1, 2, \cdots, n\}$. For set S with odd cardinality, the channel $\mathcal{M}_{\mathrm{FCS}}(\gamma_S) = 0$, rendering these components inaccessible.

As a result, FCS is limited to recovering expectation values of observables that reside within the Γ_{even} subspace. However, this restriction is sufficient for capturing physical observables, as they are typically even operators due to the conservation of fermionic parity [79]. Unless otherwise specified, all discussions and calculations are assumed to take place within the Γ_{even}.

ADAPTIVE-DEPTH FERMIONIC CLASSICAL SHADOWS

Here we consider the ADFCS by selecting the unitary ensemble as d-depth local matchgates with brick-work architecture, as shown in Fig. 3(a), which is widely applied in superconducting quantum devices [69].

It has been shown that certain observables cannot be addressed by replacing the global $\mathrm{Cl}_n \cap \mathbb{M}_n$ group with shallow-depth local matchgate circuits [8]. However, it remains open when global elements are necessary and how to efficiently generate the expectation values of a given set of fermionic observables using the shallowest depth matchgate circuits while minimizing the number of samplings. Here, we address this question by exploring the relationship between the number of required measurements and the depth of the matchgate circuit for some specific set of fermionic observables using the brickwork architecture. Our approach introduces a d-depth ADFCS channel to estimate $\mathrm{Tr}\left(\rho\gamma_S\right)$ for any quantum state ρ. We then determine the optimized depth order for the expectation values of the specific fermionic observables $\mathrm{Tr}\left(\rho\gamma_S\right)$ with respect to an unknown quantum state ρ prepared by a quantum device.

Shadow channel analysis in ADFCS

For simplicity, we utilize U_{Q_d} to denote a d-depth matchgate circuit with brickwork structure, and $\mathbb{U}_{Q_d}$ to denote the set of all d-depth matchgate circuit throughout this manuscript. We denote $\mathcal{M}_d$ as the d-depth ADFCS shadow channel, which maps a quantum state ρ to

$$\mathcal{M}_d(\rho) := \mathop{\mathbb{E}}_{U\in\mathbb{U}_{Q_d}, b\in\{0,1\}^n} \left[\langle b| U_{Q_d}\rho U_{Q_d}^\dagger |b\rangle U_{Q_d}^\dagger |b\rangle \langle b| U_{Q_d}\right]. \tag{22}$$

Due to the expectation property of the subset $\mathbb{U}_{Q_d}$, the Majorana operator remains invariant under the action of $\mathcal{M}_d$ up to a scaling factor $\alpha_{S,d}$, i.e.,

$$\mathcal{M}_d(\gamma_S) = \alpha_{S,d}\gamma_S, \tag{23}$$

where $\alpha_{S,d} := \int \mathrm{d}U_{Q_d} \left|\langle \mathbf{0}| U_{Q_d}\gamma_S U_{Q_d}^\dagger |\mathbf{0}\rangle\right|^2$. The proof is shown in Appendix .

Eq. (23) provides an explicit expression for the shadow channel, enabling a detailed analysis of the statistical properties of ADFCS. The ADFCS estimator is an unbiased estimator, analogous to the FCS estimator, which follows directly from the linearity of the shadow channel $\mathcal{M}_d$

$$\mathop{\mathbb{E}}_{U_{Q_d},b}(\mathcal{M}_d^{-1}(U_{Q_d}^\dagger |b\rangle \langle b| U_{Q_d})) = \mathcal{M}_d^{-1}(\mathbb{E}(U_{Q_d}^\dagger |b\rangle \langle b| U_{Q_d})) = \rho \tag{24}$$

when $\alpha_{S,d}$ is non-zero.

For any observable γ_S with respect to the average of quantum states ρ, the variance of the ADFCS estimator $v = \mathrm{Tr}\left(\hat{\rho}\gamma_S\right)$ is bounded as follows:

$$\mathrm{Var}\left([]\, v\right] \leq \frac{1}{\alpha_{S,d}}. \tag{25}$$

The detailed derivation of Eq. (25) is provided in Appendix . Using this result and applying Chebyshev's inequality, the estimation error $|v - \mathrm{Tr}\left(\rho\gamma_S\right)|$ can be bouned to ϵ using $\frac{1}{\alpha_{S,d}\epsilon^2}$ copies of the quantum states with high success probability. Numerical results demonstrate that the variance is stable across various quantum states. In the following sections, we will present the calculation for $\alpha_{S,d}$, along with its upper bound.

Tensor network approach to variance bound

By Eq. (25), bounding the variance requires calculating $\alpha_{S,d}$. Here we demonstrate that $\alpha_{S,d}$ can be represented as a tensor network and computed through contraction. Using the Pauli-transfer matrix representation, $\alpha_{S,d}$ can be expressed as

$$\alpha_{S,d} = \langle\!\langle \mathbf{0}, \mathbf{0}| \int \mathrm{d}\mathcal{U}_{Q_d} \mathcal{U}_{Q_d}^{\otimes 2} |\gamma_S, \gamma_S\rangle\!\rangle, \tag{26}$$

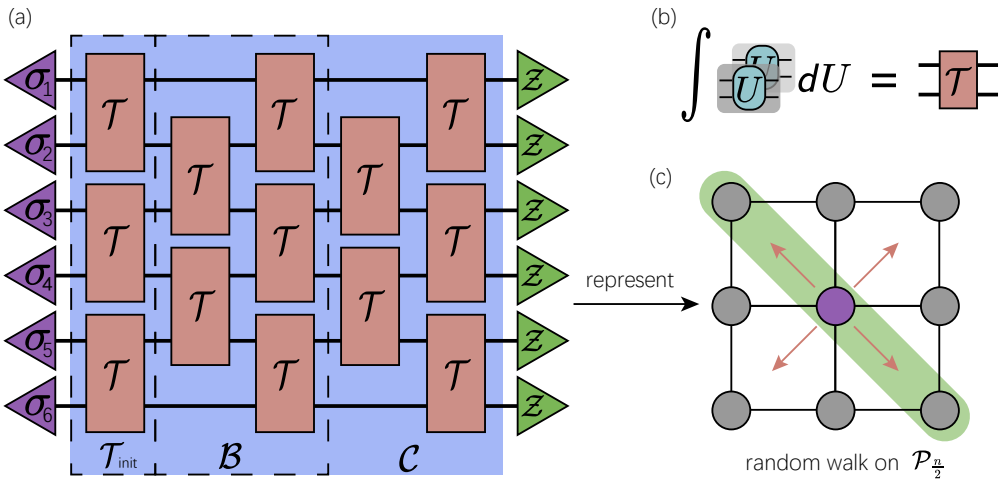

FIG. 4. Representation of the tensor contraction of $\mathcal{C}$ as a random walk on the polynomial space $\mathcal{P}_{\frac{n}{2}}$. (a) The tensor network used to compute $\alpha_{S,d}$. The purple triangles represent Pauli operators corresponding to γ_S via the Jordan-Wigner transformation, while the green triangles, denoted as $\mathcal{Z}$, correspond to the supervector $\frac{1}{2}(|\mathbb{1},\mathbb{1}\rangle\rangle + |Z,Z\rangle\rangle)$. (b) The definition of $\mathcal{T}$ in Eq. (27). (c) The contraction of $\mathcal{C}$ onto the polynomial space $\mathcal{P}_{\frac{n}{2}}$. Transitioning the state $|\gamma_S,\gamma_S\rangle\rangle$ and contracting it with $\mathcal{Z}$ is equivalent to summing the probabilities over the diagonal sites in a 2D random walk.

where $\mathcal{U}_{Q_d}$ is the superoperator of U_{Q_d}, $\mathcal{U}_{Q_d}|\gamma_S\rangle\rangle = U_{Q_d}\gamma_S U_{Q_d}^\dagger$. Consequently, the integration over $\mathcal{U}_{Q_d}$ can be broken down into a product of a series of integrations over the independent 2-qubit matchgates. By Lemma 1, we can represent the integration over each 2-qubit matchgate as a fourth-order tensor $\mathcal{T}$ with indices $\sigma_1,\sigma_2,\sigma_3,\sigma_4$. The explicit form of this tensor is given by

$$\mathcal{T}^{\sigma_1\sigma_2}_{\sigma_3\sigma_4} := \langle\langle\sigma_1,\sigma_2|\langle\langle\sigma_1,\sigma_2|\int_{Q\sim O(4)} \mathrm{d}\mathcal{U}_Q \mathcal{U}_Q^{\otimes 2}|\sigma_3,\sigma_4\rangle\rangle|\sigma_3,\sigma_4\rangle\rangle, \tag{27}$$

where the indices $\sigma_1,\sigma_2,\sigma_3,\sigma_4$ of tensor $\mathcal{T}$ represent the Pauli operators. The tensor $\mathcal{T}$ is the integration of one blue random gate in Fig. 3. By connecting these tensors $\mathcal{T}$ in the same brickwork architecture as U_{Q_d}, the integration over the superoperator in Eq. (26) can be transformed into a tensor network. The transformation details are shown in Appendix .

The value of $\alpha_{S,d}$ can be calculated by contracting the corresponding brickwork tensor network, which is shown in Fig. 4(a). Specifically, the brickwork tensor network, denoted as $\mathcal{C}$, is contracted with $\langle\langle\mathbf{0},\mathbf{0}|$ and $|\gamma_S,\gamma_S\rangle\rangle$ in the Pauli-transfer matrix (PTM) representation to calculate $\alpha_{S,d}$

$$\alpha_{S,d} = \langle\langle\mathbf{0},\mathbf{0}|\mathcal{C}|\gamma_S,\gamma_S\rangle\rangle. \tag{28}$$

Following the proof of Lemma 5 in Ref. [72], the tensor network $\mathcal{C}$ could be represented by a matrix product operator with bond dimension $2^{\mathcal{O}(d)}$. It suggests that calculating $\alpha_{S,d}$ by tensor network contraction is efficient only when the depth d is shallow.

Adaptive depth for two-local Majorana strings

Here we start by determining the minimum depth d for the matchgate subset $\mathbb{U}_{Q_d}$, such that the variance of the 2-local Majorana strings $\{\gamma_S| \ |S|=2\}$ requires the same or fewer measurements compared to the full matchgate group $\mathbb{M}_n$. Notably, for specific sets of Majorana strings $\{\gamma_S\}$, the optimized number of measurements can be significantly improved compared to FCS when using a smaller d. As an instance, suppose we need to estimate the expectation of $\gamma_i\gamma_{i+1}$. The random measurement in FCS actually shuffles

the $\gamma_i\gamma_{i+1}$ to any $\gamma_j\gamma_k$ uniformly. However, a matchgate circuit with $d = 2$ will only slightly perturb the $\gamma_i\gamma_{i+1}$ to the nearby operators $\{\gamma_j\gamma_k \mid i-2 \le j \le i+2, i-1 \le k \le i+3\}$. A smaller sample space allows it to approach its theoretical mean with fewer samples compared to the FCS method.

We will investigate how the sample complexity of the quantum state changes as the depth of the matchgate circuits with brickwork structures increases. To determine the required depth of a random matchgate circuit for efficient measurement of a specific set of 2 local Majorana strings $\{\gamma_S\}$, we derive an expression for $\alpha_{S,d}$ and identify the optimal depth d^* such that $\alpha_{S,d^*} = \frac{1}{\text{poly}(n)}$. It is challenging to give the explicit formulation of the $\alpha_{S,d}$ using the tensor network contraction. However, we could simplify the calculation by restricting the action of the tensor within the subspace spanned by $\Gamma_2 := \text{span}\{\gamma_S \mid |S| = 2\}$. By restricting the input state of the whole tensor network $\mathcal{C}$ to the subspace Γ_2, we can introduce more refined structures by mapping this subspace to a polynomial space. This isomorphism allows us to analyze the system more effectively. We observed that the tensor network contraction of $\mathcal{C}$ in Γ_2 is analogous to a random walk in polynomial space.

We will focus on random matchgate circuits characterized by an even number of qubits and an odd circuit depth. For the random matchgate circuits with this structure, the contraction of $\mathcal{C}$ can be represented within the space of quadratic polynomials with $\frac{n}{2}$ variables. For the circuits with even circuit depth, the contraction can be analyzed using a similar method. Denote the quadratic polynomial representation space as $\mathcal{P}_{\frac{n}{2}}$, the contraction of $\mathcal{C}$ could be described by the random walk with a simple pattern in space $\mathcal{P}_{\frac{n}{2}}$.

To represent this random walk, we use the transition process $\mathcal{B}$ to denote a pair of successive gate layers starting from an odd layer,

$$\mathcal{B}|\gamma_S, \gamma_S\rangle\rangle = \sum_{S':|S'|=|S|} \text{Prob}(\gamma_{S'}|\gamma_S)|\gamma'_S, \gamma'_S\rangle\rangle, \tag{29}$$

which is shown in Fig. 4(a). The contraction of $\mathcal{C}$ can be understood as the repeated application of $\mathcal{B}$ to the state $|\gamma_S, \gamma_S\rangle\rangle$. As previously mentioned, we focus on the circuit with odd depth. Therefore, the entire tensor network can be expressed as

$$\mathcal{C} = \mathcal{B}^{\lfloor d/2 \rfloor}\, \mathcal{T}_{\text{init}}, \tag{30}$$

where $\mathcal{T}_{\text{init}} = \mathcal{T}^{\otimes \frac{n}{2}}$ (and the circuit can be expressed as $\mathcal{C} = \mathcal{B}^{d/2}$ if the depth is an even number). Each $\mathcal{B}$ corresponds to a single transition step in a random walk. Therefore, the entire tensor network $\mathcal{C}$ represents a random walk starting from γ_S and taking $\lfloor \frac{d}{2} \rfloor$ steps, which is shown in Fig. 4(c). The details of mapping the tensor contraction of $\alpha_{S,d}$ to the random walk in $\mathcal{P}_{\frac{n}{2}}$ is shown in Appendix .

We observe that the random walk follows the pattern of a symmetry lazy random walk (SLRW) [17, 18] almost everywhere within the Γ_2 space. SLRW is a stochastic process in which the walker has a probability of staying in the same position at each step and equal probabilities of moving to neighboring positions. Firstly, we consider the random walk transition $\mathcal{B}$ across the subspace Γ_2 adheres to the SLRW

$$\mathcal{L}_{\Gamma_2}|\gamma_S, \gamma_S\rangle\rangle = \sum_{S':|S'|=|S|} \text{Prob}_L(\gamma_{S'}|\gamma_S)|\gamma'_S, \gamma'_S\rangle\rangle, \tag{31}$$

where $\mathcal{L}_{\Gamma_2}$ is the transition operator in SLRW, and the concrete form of $\mathcal{L}_{\Gamma_2}$ is shown in Appendix , Eq. (249). We replace the transition operator $\mathcal{B}$ with the SLRW transition operator $\mathcal{L}_{\Gamma_2}$ in $\mathcal{C}$ and use the transition $\mathcal{L}_{\Gamma_2}$ to calculate the tensor network contraction

$$\alpha^L_{S,d} = \langle\langle \mathbf{0}, \mathbf{0}|\mathcal{L}^{\lfloor \frac{d}{2} \rfloor}_{\Gamma_2}\mathcal{T}_{\text{init}}|\gamma_S, \gamma_S\rangle\rangle. \tag{32}$$

As we have shown in Table 4 of Appendix , the transition result of $\mathcal{L}_{\Gamma_2}$ equals $\mathcal{B}$ for most states in Γ_2, this indicates that the variable $\alpha^L_{S,d}$ likely constitutes the dominant contribution to $\alpha_{S,d}$. Additionally, numerical fitting further confirms that $\alpha_{S,d}$ is primarily governed by $\alpha^L_{S,d}$, as shown in Fig. 3 (b).

There is a known analytical propagation equation for the SLRW [17]. By applying this propagation equation and making certain approximations, the α-value can ultimately be expressed as a Poisson summation

$$\alpha_{S,d}^{L} = \frac{1}{3\sqrt{\pi(d-1)}} \sum_{k=-\infty}^{\infty} \left(e^{-\frac{(kn+a)^2}{d-1}} + e^{-\frac{(kn+b)^2}{d-1}} \right) + \mathcal{O}(e^{-\pi^2 d}). \tag{33}$$

In Eq. (33), we denote the Majorana observable γ_S with $|S| = 2$ as $\gamma_i\gamma_j$, the variable a, b are defined by $a := |\lfloor \frac{i-1}{4} \rfloor - \lfloor \frac{j-1}{4} \rfloor|$, and $b := \lfloor \frac{i-1}{4} \rfloor + \lfloor \frac{j-1}{4} \rfloor + 1$. We leave the detailed proof of Eq. (33) to Appendix .

We aim to estimate the order of $\alpha_{S,d}$ by $\alpha_{S,d}^{L}$, which could be achieved by bounding the ratio $\alpha_{S,d}/\alpha_{S,d}^{L}$ to a constant value. To bound the ratio, we introduce an auxiliary function

$$\Delta(S, S', d) = \langle\langle \gamma_{S'}, \gamma_{S'} | \mathcal{B}^{\lfloor \frac{d}{2} \rfloor} \mathcal{T}_{\text{init}} - \mathcal{L}_{\Gamma_2}^{\lfloor \frac{d}{2} \rfloor} \mathcal{T}_{\text{init}} | \gamma_S, \gamma_S \rangle\rangle, \tag{34}$$

where $S' \in \Gamma_2$. Recall that the operator $\gamma_{S'}$ is defined as $\gamma_{S'} = \gamma_u\gamma_v$ where $S' = \{u, v\}$. In numerical experiments, we observe that the term $\Delta(S, S', d)$ is negative if and only if the element u is close to v for $S' = \{u, v\}$ and any S, d, as shown in Fig. 13 of Appendix . Based on this finding, we show that the ratio $\alpha_{S,d}/\alpha_{S,d}^{L}$ can be bounded by a constant value. The concrete description for bounding the ratio $\alpha_{S,d}/\alpha_{S,d}^{L}$ is provided in Appendix .

The relation between $\alpha_{S,d}^{L}$ and $\alpha_{S,d}$, named $\alpha_{S,d} \geq c\alpha_{S,d}^{L}$ for positive constant c, indicated that the order of $\alpha_{S,d}^{L}$ consistent with the order of $\alpha_{S,d}$. Here, we propose the adaptive depth d^* of a random matchgate circuit by analyzing the order of $\alpha_{S,d}^{L}$. By Eq. (33), for $a = \mathcal{O}(\text{polylog}(n))$ and the measurement depth $d = \text{polylog}(n)$ can yield a polynomially small $\alpha_{S,d} = \Omega(1/\text{poly}(n))$, we postpone the proof in Appendix . Otherwise, if $a = \Omega\left(\text{polylog}(n)\right)$, achieving a polynomially small $\alpha_{S,d}$ (rather than an exponentially small one) requires a measurement circuit with depth of $d = \Omega\left(\frac{(i-j)^2}{\log(n)}\right)$. Thus, we could adaptively select the depth d based on the distance $|i-j|$,

$$d^* = \Theta\left(\max\left\{\frac{(i-j)^2}{\log(n)},\ |i-j|\right\}\right). \tag{35}$$

By Eq. (33), if the depth of the matchgate circuit $d = o\left(d^*\right)$, the value of $\alpha_{S,d}^{L}$ will be exponentially small, indicating that exponential samples are required to obtain a good estimation of the expectation value.

Adaptive depth for general k local Majorana strings

Here we focus on the extension of adaptive depth to general k-local Majorana strings for constant k. Fermionic observables generated by k-local Majorana strings for constant k play a critical role in various quantum models across physics and chemistry. These operators are essential in describing physical observables such as interaction terms, electron correlation, and pairing mechanisms. They appear in models like quantum chemistry Hamiltonians, generalized Hubbard models, spin chain models, etc [54–59].

The required circuit depth depends on the interaction distance d_{int} of γ_S, $d_{\text{int}}(S) := \max\{\, |i_{j+1} - i_j| \mid i_j \in S,\, j \in [2n-1] \,\}$. Recall Eq. (17) that the elements in S are a monotonically increasing sequence. Thus, the d_{int} could be interpreted as the maximum distance between adjacent elements in S. We prove that a random measurement circuit $\mathcal{U}_{Q_d}$ with $d = \Theta(\text{polylog } n)$ is sufficient to make $\alpha_{S,d}$ as large as $\mathcal{O}(1/\text{poly}(n))$ if $d_{\text{int}} = \mathcal{O}(\log n)$ and cardinality k is a constant. From Eq. (28) and Eq. (29), we conclude that the value of $\alpha_{S,d}$ is the sum of probabilities associated with specific sites after the random walk. When the distance of the near neighborhood $i_k, i_{k+1} \in S$ logarithmically small $|i_k - i_{k+1}| = \mathcal{O}(\log(n))$, a logarithmically transition steps can traverse from $|\gamma_S, \gamma_S\rangle\rangle$ to specific sites. Due to Eq. (29), each step

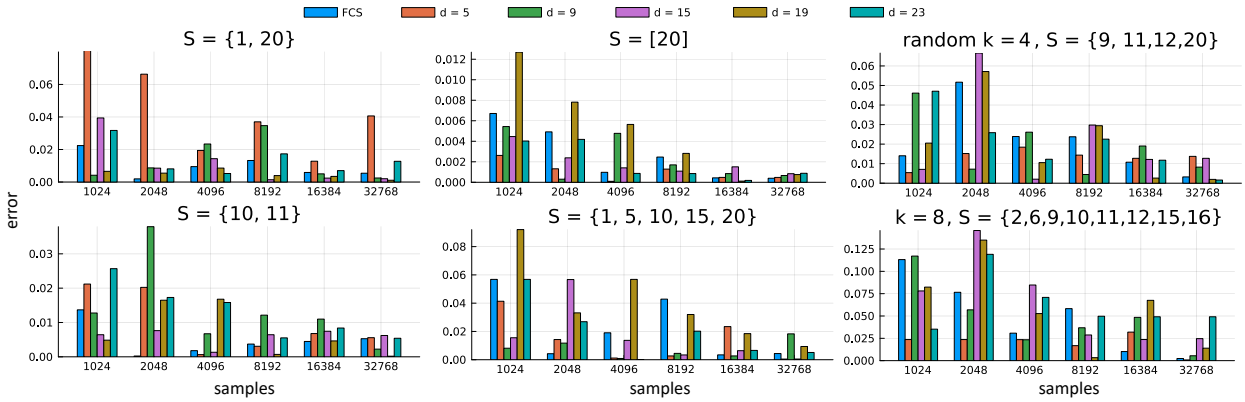

FIG. 5. Estimation errors with increasing numbers of measurements for FCS and ADFCS protocols at different depths d. Each experiment uses a randomly generated 10-qubit quantum state. The set of S for the observable γ_S is randomly generated in each subfigure, and $k = |S|$.

introduces a constant probability factor. Thus, the probability of reaching specific sites is $\mathcal{O}(1/\text{poly}(n))$ after logarithmic steps random walk. We put the proof details in Appendix .

We propose a general formula for approximating $\alpha_{S,d}$. The proposed formula considers the random walk for computing $\alpha_{S,d}$ as the several independent $k = 2$ random walks. Any two elements $i, j \in S$ are treated as an independent random walk starting from $\gamma_i\gamma_j$. After the overall random walk, the probability of $|\gamma_S, \gamma_S\rangle\rangle$ being at a specific site equals the product of the probabilities of each 2-local γ'_S at that site. As a result, the sum of probabilities at these specific sites can be expressed as a production $\prod \alpha_{\{i,j\},d}$. Therefore, we consider the overall random walk as the superposition of all such production

$$\alpha'_{S,d} = \frac{1}{(k-1)!!}\left(\frac{3n}{2}\right)^{\frac{k}{2}} \frac{\binom{n}{k/2}}{\binom{2n}{k}} \sum_{\Lambda\in \text{Par}(S)} \prod_{(i,j)\in\Lambda} \alpha^L_{\{i,j\},d}, \tag{36}$$

where $\text{Par}(S)$ is the set whose elements Λ are sets representing pairwise partitions of S. For example, let $S = \{1, 2, 3, 4\}$, then $\Lambda = \{(1, 2), (3, 4)\}$ is a pairwise partitions of S.

Numerical experiments show that the proposed expression $\alpha'_{S,d}$ fits the true $\alpha_{S,d}$ well. Therefore, it can be concluded that $\alpha'_{S,d}$ provides a good approximation of $\alpha_{S,d}$. Fig. 3(b) show the results of comparison between $\alpha'_{S,d}$ and $\alpha_{S,d}$. The curve of $\alpha'_{S,d}$ is very close to the $\alpha_{S,d}$ when $d > 2\log(n)$, which suggests that the $\alpha'_{S,d}$ is a good approximation of $\alpha_{S,d}$ as well.

We adaptively select the depth based on the interaction distance d_{int} of a fermionic Hamiltonian H, where $d_{\text{int}}(H)$ is the maximum value of $d_{\text{int}}(S)$ for terms γ_S in H. Similar to the case with $|S| = 2$, we adaptively select the depth d^* as

$$d^* = \Theta\left(\max\left\{\frac{d_{\text{int}}(S)^2}{\log(n)}, d_{\text{int}}(S)\right\}\right). \tag{37}$$

Such d^* ensures that each term in Eq. (36) scales polynomially, thereby guaranteeing the overall $\alpha'_{S,d}$ also scales polynomially.

NUMERICAL EXPERIMENT

Here, we provide numerical evidence demonstrating the efficiency of the ADFCS protocol. First, we compare the performance of ADFCS with FCS across various circuit depths and observables γ_S. For each

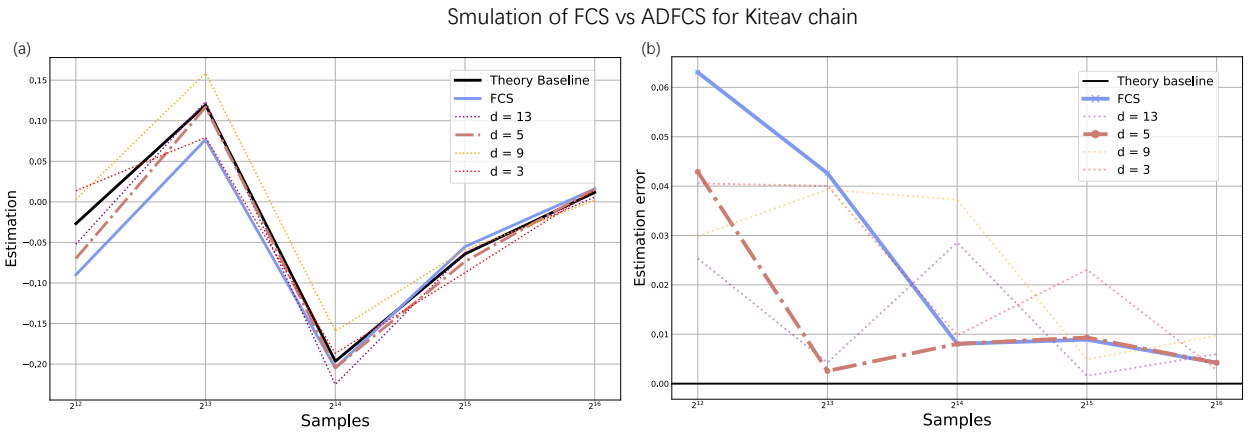

FIG. 6. Application of ADFCS to the Kitaev chain Hamiltonian H_K. (a) The estimation generated from FCS and ADFCS with varying circuit depths. (b) The error between the estimation and the expectation value.

experiment, a random 10-qubit quantum state ρ is chosen, followed by a d-depth random matchgate circuit with the brickwork structure and measurement on the computational basis. We enumerate depth d in $\{5, 9, 15, 19, 23\}$. The estimation error with increasing numbers of measurements for ADFCS with different depth d and FCS is shown in Fig. 5. We estimate the error for $\mathrm{Tr}\left(\gamma_S \rho\right)$ for $k = |S| =\in \{1, 2, 4, 5, 8\}$ and randomly choose set S in each subfigure of Fig. 5. The error is estimated as $\sqrt{\sum_{j=1}^{R} \frac{|\mathrm{Tr}(\gamma_S \hat{\rho}_i) - \mathrm{Tr}(\gamma_S \rho)|^2}{R}}$, where $\hat{\rho}_i$ is generated from the i-th measurement, and R is set to 32. The required depth of the random matchgate circuit for a specific observable γ_S depends on the value of $\alpha_{S,d}$.

When the interaction distance is small, shallower circuits tend to perform better. For instance, when $S = \{10, 11\}$, $S = [20]$, and $S = \{2, 6, 9, 10, 11, 12, 15, 16\}$, random matchgate circuits with depth $d = 5$ provide good estimations of the expectation value. However, when the interaction distance is larger, such as in the case of $S = \{1, 20\}$, circuits with depths ranging from $d = 15$ to $d = 23$ are required to achieve accurate expectation value estimation. If the circuit depth is chosen appropriately, a good estimate can be obtained with a relatively small number of samples.

To further test the performance of ADFCS, we apply the method of estimating the expectation of the Kiteav chain Hamiltonian. The Kitaev chain Hamiltonian is a one-dimensional model that describes a topological superconductor featuring Majorana fermions at its ends [80]. The Kiteav chain Hamiltonian is given by

$$H_K = -\frac{\mathrm{i}\mu}{2} \sum_{j=1}^{n} \gamma_{2j-1}\gamma_{2j} + \frac{\mathrm{i}}{2} \sum_{j=1}^{n-1} \left(\omega_+ \gamma_{2j-1}\gamma_{2j+2} - \omega_- \gamma_{2j}\gamma_{2j+1}\right), \tag{38}$$

where μ is the chemical potential, $\omega_\pm = |\Delta| \pm t$, both Δ and $\mathcal{T}$ are certain energy gaps. We initialize a random state for 12-qubit, set these parameters as $\mu = 2$, $\Delta = 1$, and $t = 0.4$. The results are shown in Fig. 6. Since the maximum interaction distance of H_K is 3, all curves align with the theoretical values due to the short interaction distance in the insect plot. Among these curves, we highlight the curve with $d = 5$, which gives the best estimation.

CONCLUSION

Estimating the expectation value of fermionic observables is a fundamental task in quantum physics and chemistry. The fermionic classical shadow method provides an innovative approach to address this problem

without requiring the conversion of observables into the Pauli-basis representation. This significantly reduces the number of quantum states for certain local fermionic observables. However, the approach requires polynomial-size quantum circuits when sampling matchgate elements from $\mathbb{M}_n$ without considering the information from the fermionic observables. This imposes a considerable challenge for near-term quantum devices.

We propose an adaptive depth fermionic classical shadow (ADFCS) protocol to reduce the heavy circuit depth associated with random matchgate circuits. Using a tensor network approach, we calculate the variance for any d-depth ADFCS protocol, enabling us to determine the required depth d to ensure efficient sampling.

Furthermore, we theoretically analyze the explicit relationship between the variance and the depth d of the random matchgate circuit. Specifically, under certain assumptions, we find that the optimal depth is linear to $\max\left\{d_{\text{int}(S)}^2/\log n, d_{\text{int}(S)}\right\}$ where d_{int} is the interaction distance, for a given Majorana string γ_S. Numerical fitting results are also provided to support our theoretical findings.

We validate the correctness of our algorithm by evaluating the expectation values of several Majorana strings with respect to randomly generated quantum states ρ. The numerical results align with and support our theoretical findings. Additionally, we applied our algorithm to estimate the expectation value of a Kitaev chain Hamiltonian. Comparative numerical analysis demonstrates that our ADFCS algorithm achieves performance comparable to the FCS algorithm, while requiring a significantly shallower circuit depth.

[1] C. J. Cramer. *Essentials of computational chemistry: Theories and models*. John Wiley and Sons, Chichester, 2002.

[2] S. Sachdev. *Quantum Phase Transitions*. Cambridge University Press, Cambridge, 2011.

[3] Efthimios Kaxiras and John D. Joannopoulos. *Quantum Theory of Materials*. Cambridge University Press, 2019.

[4] Christine A Schwerdtfeger, A Eugene DePrince III, and David A Mazziotti. Testing the parametric two-electron reduced-density-matrix method with improved functionals: Application to the conversion of hydrogen peroxide to oxywater. *J. Chem. Phys.*, 134(17):174102, 2011.

[5] Michael R. Peterson and Chetan Nayak. More realistic hamiltonians for the fractional quantum hall regime in gaas and graphene. *Phys. Rev. B*, 87:245129, Jun 2013.

[6] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, and Todd J. Martínez. Quantum computation of electronic transitions using a variational quantum eigensolver. *Phys. Rev. Lett.*, 122:230401, Jun 2019.

[7] Tyler Takeshita, Nicholas C. Rubin, Zhang Jiang, Eunseok Lee, Ryan Babbush, and Jarrod R. McClean. Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources. *Phys. Rev. X*, 10:011004, Jan 2020.

[8] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic partial tomography via classical shadows. *Phys. Rev. Lett.*, 127:110504, Sep 2021.

[9] Guang Hao Low. Classical shadows of fermions with particle number symmetry. *arXiv preprint arXiv:2208.08964*, 2022.

[10] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. Matchgate shadows for fermionic quantum simulation, 2022.

[11] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2:030348, Sep 2021.

[12] Dax Enshan Koh and Sabee Grewal. Classical shadows with noise. *Quantum*, 6:776, aug 2022.

[13] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. The Clifford group fails gracefully to be a unitary 4-design, 2016.

[14] Jonas Helsen, Xiao Xue, Lieven MK Vandersypen, and Stephanie Wehner. A new class of efficient randomized benchmarking protocols. *npj Quantum Information*, 5(1):1–9, 2019.

[15] Zhang Jiang, Kevin J. Sung, Kostyantyn Kechedzhi, Vadim N. Smelyanskiy, and Sergio Boixo. Quantum algorithms to simulate many-body physics of correlated fermions. *Phys. Rev. Appl.*, 9:044036, Apr 2018.

[16] Bujiao Wu and Dax Enshan Koh. Error-mitigated fermionic classical shadows on noisy quantum devices. *npj*

Quantum Information, 10(1):39, 2024.
[17] Luca Giuggioli. Exact spatiotemporal dynamics of confined lattice random walks in arbitrary dimensions: a century after smoluchowski and pólya. *Physical Review X*, 10(2):021045, 2020.
[18] Gregory F Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123. Cambridge University Press, 2010.
[19] Kaiming Bian and Wu Bujiao. Adaptive-depth randomized measurement for fermionic observables. *Upcoming*, 2025.
[20] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625?644, 2021.
[21] 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.
[22] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022.
[23] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Phys. Rev. Lett.*, 127:030503, Jul 2021.
[24] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *Quantum*, 7:896, jan 2023.
[25] Charles Hadfield. Adaptive Pauli shadows for energy estimation, 2021.
[26] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Phys. Rev. Res.*, 5:023027, Apr 2023.
[27] Atithi Acharya, Siddhartha Saha, and Anirvan M. Sengupta. Shadow tomography based on informationally complete positive operator-valued measure. *Phys. Rev. A*, 104:052418, Nov 2021.
[28] Kaifeng Bu, Dax Enshan Koh, Roy J. Garcia, and Arthur Jaffe. Classical shadows with Pauli-invariant unitary ensembles, 2022.
[29] Daniel Grier, Hakop Pashayan, and Luke Schaeffer. Sample-optimal classical shadows for pure states, 2022.
[30] Matteo Ippoliti. Classical shadows based on locally-entangled measurements, 2023.
[31] You Zhou and Qing Liu. Performance analysis of multi-shot shadow estimation. *Quantum*, 7:1044, June 2023.
[32] Roy J. Garcia, You Zhou, and Arthur Jaffe. Quantum scrambling with classical shadows. *Phys. Rev. Res.*, 3:033155, Aug 2021.
[33] You Zhou and Zhenhuan Liu. A hybrid framework for estimating nonlinear functions of quantum states, 2023.
[34] Pascual Jordan and Eugene Paul Wigner. *Über das paulische äquivalenzverbot*. Springer, 1993.
[35] Michael A Nielsen et al. The fermionic canonical commutation relations and the jordan-wigner transform. *School of Physical Sciences The University of Queensland*, 59, 2005.
[36] Sergey B Bravyi and Alexei Yu Kitaev. Fermionic quantum computation. *Annals of Physics*, 298(1):210–226, 2002.
[37] Andrew Tranter, Sarah Sofia, Jake Seeley, Michael Kaicher, Jarrod McClean, Ryan Babbush, Peter V Coveney, Florian Mintert, Frank Wilhelm, and Peter J Love. The b ravyi–k itaev transformation: Properties and applications. *Int. J. Quant. Chem.*, 115(19):1431–1441, 2015.
[38] Hamza Jnane, Jonathan Steinberg, Zhenyu Cai, H Chau Nguyen, and Bálint Koczor. Quantum error mitigated classical shadows. *arXiv preprint arXiv:2305.04956*, 2023.
[39] Leslie G. Valiant. Expressiveness of matchgates. *Theoretical Computer Science*, 289(1):457–471, 2002.
[40] E. Knill. Fermionic linear optics and matchgates, 2001.
[41] Barbara M. Terhal and David P. DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Phys. Rev. A*, 65:032325, Mar 2002.
[42] David P DiVincenzo and Barbara M Terhal. Fermionic linear optics revisited. *Foundations of Physics*, 35:1967–1984, 2005.
[43] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X*, 10:031064, Sep 2020.
[44] Michael Wimmer. Algorithm 923: Efficient numerical computation of the pfaffian for dense and banded skew-symmetric matrices. *ACM Transactions on Mathematical Software (TOMS)*, 38(4):1–17, 2012.
[45] Steven T. Flammia and Joel J. Wallman. Efficient Estimation of Pauli Channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 2020.
[46] Earl T Campbell. Decoherence in open majorana systems. In *10th Conference on the Theory of Quantum Computation, Communication and Cryptography*, page 111, 2015.
[47] Marvellous Onuma-Kalu, Daniel Grimmer, Robert B Mann, and Eduardo Martín-Martínez. A classification

of markovian fermionic gaussian master equations. *Journal of Physics A: Mathematical and Theoretical*, 52(43):435302, 2019.

[48] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011.

[49] Wim C Van Etten. *Introduction to random signals and noise*. John Wiley & Sons, New York, NY, USA, 2006.

[50] Abhinav Deshpande, Pradeep Niroula, Oles Shtanko, Alexey V. Gorshkov, Bill Fefferman, and Michael J. Gullans. Tight bounds on the convergence of noisy random circuits to the uniform distribution. *PRX Quantum*, 3:040329, Dec 2022.

[51] Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. Exponentially tighter bounds on limitations of quantum error mitigation, 2023.

[52] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random Clifford circuits, 2023.

[53] Andrew Zhao and Akimasa Miyake. Group-theoretic error mitigation enabled by classical shadows and symmetries, 2023.

[54] Elliott Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 16(3):407–466, 1961.

[55] Qimiao Si and J Lleweilun Smith. Kosterlitz-thouless transition and short range spatial correlations in an extended hubbard model. *Physical review letters*, 77(16):3391, 1996.

[56] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.

[57] James D Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750, 2011.

[58] Assa Auerbach. *Interacting electrons and quantum magnetism*. Springer Science & Business Media, 2012.

[59] 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.

[60] Rev McWeeny. Some recent advances in density matrix theory. *Reviews of Modern Physics*, 32(2):335, 1960.

[61] Jerzy Cioslowski. *Many-electron densities and reduced density matrices*. Springer Science & Business Media, 2012.

[62] Albert John Coleman and Vyacheslav I Yukalov. *Reduced density matrices: Coulson's challenge*, volume 72. Springer Science & Business Media, 2000.

[63] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Physical Review X*, 10(3):031064, 2020.

[64] Andrew Zhao and Akimasa Miyake. Group-theoretic error mitigation enabled by classical shadows and symmetries. *npj Quantum Information*, 10(1):57, 2024.

[65] Valentin Heyraud, Héloise Chomet, and Jules Tilly. Unified framework for matchgate classical shadows. *arXiv preprint arXiv:2409.03836*, 2024.

[66] Ketan N Patel, Igor L Markov, and John P Hayes. Optimal synthesis of linear reversible circuits. *Quantum Inf. Comput.*, 8(3):282–294, 2008.

[67] Jiaqing Jiang, Xiaoming Sun, Shang-Hua Teng, Bujiao Wu, Kewen Wu, and Jialin Zhang. Optimal space-depth trade-off of cnot circuits in quantum logic synthesis. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 213–229. SIAM, 2020.

[68] Zhang Jiang, Kevin J Sung, Kostyantyn Kechedzhi, Vadim N Smelyanskiy, and Sergio Boixo. Quantum algorithms to simulate many-body physics of correlated fermions. *Physical Review Applied*, 9(4):044036, 2018.

[69] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[70] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, Cheng Guo, Chu Guo, Shaojun Guo, Lianchen Han, Linyin Hong, He-Liang Huang, Yong-Heng Huo, Liping Li, Na Li, Shaowei Li, Yuan Li, Futian Liang, Chun Lin, Jin Lin, Haoran Qian, Dan Qiao, Hao Rong, Hong Su, Lihua Sun, Liangyuan Wang, Shiyu Wang, Dachao Wu, Yu Xu, Kai Yan, Weifeng Yang, Yang Yang, Yangsen Ye, Jianghan Yin, Chong Ying, Jiale Yu, Chen Zha, Cha Zhang, Haibin Zhang, Kaili Zhang, Yiming Zhang, Han Zhao, Youwei Zhao, Liang Zhou, Qingling Zhu, Chao-Yang Lu, Cheng-Zhi Peng, Xiaobo Zhu, and Jian-Wei Pan. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.*, 127:180501, Oct 2021.

[71] Sirui Cao, Bujiao Wu, Fusheng Chen, Ming Gong, Yulin Wu, Yangsen Ye, Chen Zha, Haoran Qian, Chong Ying, Shaojun Guo, et al. Generation of genuine entanglement up to 51 superconducting qubits. *Nature*, 619(7971):738–742, 2023.

[72] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random clifford circuits. *Physical Review Letters*, 133(2):020602, 2024.

[73] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth. *arXiv preprint arXiv:2407.07754*, 2024.

[74] Pierre-Gabriel Rozon, Ning Bao, and Kartiek Agarwal. Optimal twirling depth for classical shadows in the presence of noise. *Phys. Rev. Lett.*, 133:130803, Sep 2024.

[75] Fermi Ma and Hsin-Yuan Huang. How to construct random unitaries. *arXiv preprint arXiv:2410.10116*, 2024.

[76] Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. *arXiv preprint arXiv:2404.19211*, 2024.

[77] Lucas Hackl and Eugenio Bianchi. Bosonic and fermionic gaussian states from kähler structures. *SciPost Physics Core*, 4(3):025, 2021.

[78] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing*, 31(4):1229–1254, 2002.

[79] Ari M. Turner, Frank Pollmann, and Erez Berg. Topological phases of one-dimensional fermions: An entanglement point of view. *Phys. Rev. B*, 83:075102, Feb 2011.

[80] A Yu Kitaev. Unpaired majorana fermions in quantum wires. *Physics-uspekhi*, 44(10S):131, 2001.

[81] Jonas Helsen, Sepehr Nezami, Matthew Reagor, and Michael Walter. Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates. *Quantum*, 6:657, feb 2022.

[82] William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, New York, NY, USA, 2013.

[83] Fionn D Malone, Robert M Parrish, Alicia R Welden, Thomas Fox, Matthias Degroote, Elica Kyoseva, Nikolaj Moll, Raffaele Santagati, and Michael Streif. Towards the simulation of large scale protein–ligand interactions on nisq-era quantum computers. *Chemical Science*, 13(11):3094–3108, 2022.

[84] Jie Liu, Zhenyu Li, and Jinlong Yang. An efficient adaptive variational quantum solver of the schrödinger equation based on reduced density matrices. *J. Chem. Phys.*, 154(24):244112, 2021.

[85] Catherine Overy, George H Booth, NS Blunt, James J Shepherd, Deidre Cleland, and Ali Alavi. Unbiased reduced density matrices and electronic properties from full configuration interaction quantum monte carlo. *J. Chem. Phys.*, 141(24):244117, 2014.

[86] Thomas E O'Brien, Bruno Senjean, Ramiro Sagastizabal, Xavier Bonet-Monroig, Alicja Dutkiewicz, Francesco Buda, Leonardo DiCarlo, and Lucas Visscher. Calculating energy derivatives for quantum chemistry on a quantum computer. *npj Quantum Information*, 5(1):113, 2019.

[87] Nicholas C Rubin, Ryan Babbush, and Jarrod McClean. Application of fermionic marginal constraints to hybrid quantum algorithms. *New Journal of Physics*, 20(5):053020, 2018.

[88] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A*, 95:042308, Apr 2017.

[89] Roe Goodman and Nolan R Wallach. *Symmetry, representations, and invariants*, volume 255. Springer, New York, NY, USA, 2009.

[90] Jing Wu and Quntao Zhuang. Continuous-variable error correction for general gaussian noises. *Phys. Rev. Appl.*, 15:034073, Mar 2021.

[91] Christian Weedbrook, Stefano Pirandola, Raú l García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84(2):621–669, may 2012.

[92] Sergey Bravyi and David Gosset. Complexity of quantum impurity problems. *Communications in Mathematical Physics*, 356:451–500, 2017.

[93] Ronald L Graham, Donald E Knuth, Oren Patashnik, and Stanley Liu. Concrete mathematics: a foundation for computer science. *Computers in Physics*, 3(5):106–107, 1989.

[94] Joseph M Renes, Robin Blume-Kohout, Andrew J Scott, and Carlton M Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

METHODS FOR "ERROR-MITIGATED FERMIONIC CLASSICAL SHADOWS ON NOISY QUANTUM DEVICES"

Noisy fermionic channel representation

Here we present an unbiased estimation approach for the noisy representation of the fermionic shadow channel, which utilizes a protocol similar to the matchgate benchmarking protocol [81]. According to representation theory [82] (see details in Supplementary Note 1), the noisy fermionic shadow channel can be represented as $\widetilde{\mathcal{M}} = \sum_{k=0}^{n} f_{2k}\mathcal{P}_{2k}$. Since $\mathrm{Tr}\left(H\widetilde{\mathcal{M}}(\rho)\right) = \mathrm{Tr}\left(\rho\widetilde{\mathcal{M}}(H)\right)$, with the pre-knowledge of f_{2k} we can calculate $\mathrm{Tr}\left(H\widetilde{\mathcal{M}}(\rho)\right)$ for any observable H in the even subspace. To learn the $2(n+1)$ coefficients, we begin with the easily prepared state $\rho_0 = |\mathbf{0}\rangle\langle\mathbf{0}|$ and apply a noisy unitary channel $\widehat{\mathcal{U}}$ with $\mathcal{U}$ sampled from the matchgate group. We then perform a Z-basis measurement $\mathcal{X}$ with measurement outcomes $x \in \{0,1\}^n$, followed by classically operating the unitary channel $\mathcal{U}^\dagger$ on $|x\rangle\langle x|$. The generated state has expected value $\sum_{k=0}^{n} f_{2k}\mathcal{P}_{2k}(\rho_0)$, and f_{2k} is obtained by projecting the final state to the $\mathcal{P}_{2k}$ subspace with some classical post-processing. We illustrate the learning process of the noisy channel in Fig. 1 (a). The following theorem provides an unbiased estimation of the noisy fermionic classical shadow.

Theorem 2. *The noisy fermionic shadow channel can be represented as $\widetilde{\mathcal{M}} = \sum_k f_{2k}\mathcal{P}_{2k}$, where $\mathcal{P}_{2k}$ is defined in Eq. (4), and $\hat{f}_{2k} = 2^n\langle\langle\mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle/\binom{n}{k}$ is an unbiased estimator of $f_{2k} \in \mathbb{R}$, where $|x\rangle\rangle$ is the measurement outcome from the noisy shadow protocol obtained by starting from the input state $|\mathbf{0}\rangle\rangle$ and applying a noisy quantum circuit $\widetilde{\mathcal{U}}_Q$ followed by a Z-basis measurement, where $\mathcal{U}_Q$ is uniformly randomly picked from the matchgate group.*

The representation of the noisy fermionic channel, denoted by $\widetilde{\mathcal{M}} = \sum_k f_{2k}\mathcal{P}_{2k}$, where $f_{2k} \in \mathbb{C}$, can be obtained by the irreducible representation of the Gaussian unitary. A detailed proof of this theorem is provided in Supplementary Note 4. We claim that the coefficients $\hat{f}_{2k}$ can be efficiently calculated with the following lemma.

Lemma 2. *$\hat{f}_{2k}$ is the coefficient of x^k in the polynomial $p_Q(x)$, where*

$$p_Q(x) = \binom{n}{k}^{-1} \mathrm{pf}\left(C_{|\mathbf{0}\rangle}\right) \mathrm{pf}\left(-C_{|\mathbf{0}\rangle}^{-1} + xQ^T C_{|x\rangle} Q\right), \tag{39}$$

where $C_{|x\rangle} = \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & (-1)^{x_j} \\ (-1)^{x_j+1} & 0 \end{pmatrix}$ is the covariance matrix of $|x\rangle$.

This lemma can be obtained by Proposition 1 in Ref. [10]. For the completeness of this paper, we also give the proof of this lemma in Supplementary Note 4. The coefficients can be calculated with the polynomial interpolation method in polynomial time. With Theorem 2, we can give an unbiased estimation $\widehat{\mathcal{M}}$ for $\widetilde{\mathcal{M}}$.

By the definition of $\hat{f}_{2k}$, and the twirling properties of $\int_Q d\mu(Q)\mathcal{U}_Q^{\otimes 2}$, the expectation value for the estimation $\hat{f}_{2k}$ can be formulated as

$$f_{2k} = \binom{2n}{2k}^{-1}\binom{n}{k}\mathcal{B}_k. \tag{40}$$

We postpone the details of this proof to Supplementary Note 4. It implies that in the noiseless scenario, $\widetilde{M}$ degenerates into $\mathcal{M}$ as defined in Equation (7). Combined with the definition of $\mathcal{B}_k$ in Eq. (8), f_{2k} is close to $\binom{2n}{2k}^{-1}\binom{n}{k}$ if the average noise fidelity in Γ_{2k} subspace is close to one. Recall that $\mathcal{B}_k$ is a constant in the

depolarizing, amplitude-damping, and X-rotation noises with a constant noise strength, which implies that these noises can be efficiently mitigated with our algorithm.

Alternatively, we have a counterexample in Supplementary Note 3 that illustrates the limitations of our mitigation algorithm. Specifically, if the noise follows a Gaussian unitary channel $\mathcal{U}_Q$ where Q is a signed permutation matrix (associated with a discrete Gaussian unitary), then $\mathcal{B}_k$ may become zero. Hence, $f_{2k} = 0$, rendering our estimation approach unsuitable.

Recall that our goal is to estimate $\{\mathrm{Tr}(\rho H_i)\}_{i=1}^m$ using a noisy quantum device and polynomial classical cost, where ρ is an n-qubit quantum state and H_i is an observable in the even subspace Γ_{even}. Here we visualize the estimation process with the guarantee in Theorem 2. We uniformly randomly sample a matchgate U_Q from the matchgate group and apply it to the quantum state ρ, and then measure in the Z-basis to get outcomes x. We define the estimator

$$\hat{v} = \mathrm{Tr}\left(H\widehat{\mathcal{M}}^{-1}(U_Q^\dagger |x\rangle\langle x| U_Q)\right) \tag{41}$$

$$= \sum_{k=0}^{n} \hat{f}_{2k}^{-1}\mathrm{Tr}\left(H\mathcal{P}_{2k}\left(U_Q^\dagger |x\rangle\langle x| U_Q\right)\right). \tag{42}$$

It is easy to show that $\mathrm{tr}(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho))$ is an unbiased estimation of $\mathrm{tr}(\rho H)$ when $\widetilde{\mathcal{M}}$ is invertible, specifically when $f_{2k} \neq 0$ for any k, and H belongs to the even subspace Γ_{even}. Given that $\mathbb{E}[\widehat{\mathcal{M}}] = \widetilde{\mathcal{M}}$ and $\hat{v}$ serves as an unbiased estimator of $\mathrm{tr}(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho))$, it implies that the estimation error $\varepsilon := |\hat{v} - \mathrm{tr}(\rho H)|$ is bounded by $|\hat{v} - \mathrm{tr}(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho))| + |\mathrm{tr}(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)) - \mathrm{tr}(\rho H)|$, which can be minimized with the increasing of the number of samplings for the estimations $\hat{v}$ and $\mathrm{tr}(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho))$, as shown in Eq. (12). Note that the estimator defined in Eq. (42) is not always efficient for all states ρ and observables H. Here we claim that with this estimator, we can efficiently calculate substantial physical quantities such as the expectation value of k-RDM, which not only serves the variational quantum algorithm (VQE) of a fermionic system with up to k particle interactions [83, 84], but also provide supports to the calculations of derivatives of the energy [85, 86] and multipole moments [87]. It is also an indispensable resource for the error mitigation technique [7, 88]. It also serves to calculate the overlap between a Gaussian state and any quantum state, and the inner product between a Slater determinant and any pure state inspired by the fermionic shadow analysis of Wan et al. [10]. We postpone the details to Supplementary Note 5.

Note that all elements of k-RDMs can be derived through $\mathrm{Tr}(\rho\gamma_S)$ for a total of $\mathcal{O}\left(n^k\right)$ sets S with $|S| = 2k$. In an expansion of this concept, we now focus on evaluating $\mathrm{Tr}(\rho\widetilde{\gamma}_S)$ for $\mathcal{O}\left(n^k\right)$ different S with $|S| = 2k$. To calculate the expectation value $\mathrm{Tr}(\rho\widetilde{\gamma}_S)$, we set the input quantum state to be ρ and the observable to be $H = \widetilde{\gamma}_S$ in the estimation formula of Eq. (42), which can then be simplified to

$$\hat{v} = i^k \hat{f}_{2k}^{-1}\mathrm{pf}\left(Q_1 Q^T C_{|x\rangle} Q Q_1^T|_S\right), \tag{43}$$

where $\widetilde{\gamma}_j = \sum_{l\in[2n]} Q_1(j,l)\gamma_l$, $Q_1(j,l)$ is the (j,l)-th element of Q_1, and $C_{|x\rangle} = \bigoplus_{i=1}^n \begin{pmatrix} 0 & (-1)^{x_j} \\ (-1)^{x_j+1} & 0 \end{pmatrix}$ is the covariance matrix of $|x\rangle\langle x|$. Here, $A|_S$ refers to the submatrix obtained by taking the columns and rows of the matrix A that are indexed by S. The simplified quantity can be calculated in polynomial time since $\hat{f}_c$ and the Pfaffian function can be calculated efficiently. We give a detailed proof of the simplification process in Supplementary Note 5.

Supplementary notes for "Error-mitigated fermionic classical shadows on noisy quantum devices"

SUPPLEMENTARY NOTE 1

This section provides an overview of group theory [82, 89] and its application in the context of the matchgate group.

Let $\mathbf{G}$ be a finite group and let $U(V)$ denote the group of unitary linear transformations on a finite complex vector space V. We define a representation ϕ of the group $\mathbf{G}$ on the space V as

$$\phi : \mathbf{G} \to U(V) : \mathbf{G} \to \phi(\mathbf{G}). \tag{44}$$

Maschke's lemma shows that any representation of a group can be uniquely represented as a direct sum of its irreducible representations, i.e.

$$\phi(G) = \bigoplus_{\lambda \in R_G} \phi_\lambda(G)^{\otimes m_\lambda}, \tag{45}$$

for any $G \in \mathbf{G}$, where ϕ_λ is the irreducible representation of $\mathbf{G}$ and m_λ is the multiplicity of ϕ_λ. Here we consider only the multiplicity-free case where $m_\lambda = 1$ for all ϕ_λ. We provide a corollary of Schur's lemma to evaluate twirls over multiplicity-free representations as follows.

Lemma 3 (Rephrase of [82, Lemma 1.7 and Proposition 1.8] and [14, Lemma 1])**.** *Let $\mathbf{G}$ be a finite group, and ϕ be a multiplicity-free representation of $\mathbf{G}$ on a finite complex vector space V with decomposition*

$$\phi(G) \simeq \bigoplus_{\lambda \in R_G} \phi_\lambda G \tag{46}$$

for any $G \in \mathbf{G}$, into inequality irreducible representation ϕ_λ. Then for any linear map $Q : V \to V$, the twirl of Q over ϕ has the form

$$\mathcal{E}_\phi(Q) = \frac{1}{|\mathbf{G}|} \sum_{G \in \mathbf{G}} \phi(G) Q \phi(G)^\dagger = \sum_{\lambda \in R_G} \frac{Tr\left(Q\mathcal{P}_\lambda\right)}{Tr\left(\mathcal{P}_\lambda\right)} \mathcal{P}_\lambda, \tag{47}$$

where $\mathcal{P}_\lambda$ is the projector onto the support of the representation ϕ_λ.

Let $Q^{(u,v)} \in \mathbb{R}^{2n \times 2n}$ be the basic rotation matrix with $Q^{(u,v)} = Q^{(u,v)}_{uv} \oplus \mathbb{I}_{-uv}$, where the submatrix $Q^{(u,v)}_{uv}$ is constructed by the uth and vth rows and columns of $Q^{(u,v)}$, and is equal to

$$Q^{(u,v)}_{uv} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}_{u,v}. \tag{48}$$

The submatrix $\mathbb{I}_{-uv} = \mathbb{I} - |u\rangle\langle u| - |v\rangle\langle v|$ contains all of the rows and columns of identity except the u-th and v-th rows, and columns. We define matchgate $U_{Q^{(u,u+1)}}$corresponding to the rotation matrix $Q^{(u,u+1)}$ of two nearest neighbors as $\exp\left(i\theta X_{u/2} X_{u/2+1}\right)$ for even u, and $\exp\left(i\theta Z_{(u+1)/2}\right)$ for odd u. Let the reflection with angle 0 to the $2k$th row and column, denote as $Q^{(k)} = [-1]_{2k} \oplus \mathbb{I}_{-2k}$, and the matchgate $U_{Q^{(k)}} = X_k$. Hence all of $\left\{Q^{(u,u+1)}, Q^{(k)}\right\}$ generates the orthogonal group $\mathrm{Orth}(2n)$, and

$$U_Q^\dagger \gamma_u U_Q = \sum_{v=1}^{n} Q_{uv} \gamma_v, \tag{49}$$

for any $u \in [2n]$. The matchgate group is denoted as

$$\mathbb{M}_n = \left\{U_Q | Q \in \mathrm{Orth}(2n)\right\}. \tag{50}$$

As an application of Lemma 3, we define the superoperator representation of a given unitary as $\phi(U) := \mathcal{U}$. Then,

$$\phi(U)\phi(V) = \phi(UV). \tag{51}$$

In the following, we review the results of the first three moments of the uniform distribution over the matchgate group.

Lemma 4 (Wan et al. [10])**.** *Let Q be a matrix that is uniformly randomly sampled from the orthogonal group Orth$(2n)$. Then,*

$$\int_Q d\mu(Q)\mathcal{U}_Q = |\mathbb{I}\rangle\!\rangle\langle\!\langle\mathbb{I}| \tag{52}$$

$$\int_Q d\mu(Q)\mathcal{U}_Q^{\otimes 2} = \sum_{k=0}^{2n} |\mathcal{R}_k^{(2)}\rangle\!\rangle\langle\!\langle\mathcal{R}_k^{(2)}| \tag{53}$$

$$\int_Q d\mu(Q)\mathcal{U}_Q^{\otimes 3} = \sum_{\substack{k_1,k_2,k_3\geq 0\\ k_1+k_2+k_3\leq 2n}} |\mathcal{R}_{k_1,k_2,k_3}^{(3)}\rangle\!\rangle\langle\!\langle\mathcal{R}_{k_1,k_2,k_3}^{(3)}|, \tag{54}$$

where μ is the normalised Haar measure on $\mathrm{Orth}(2n)$*, and*

$$|\mathcal{R}_k^{(2)}\rangle\!\rangle = \binom{2n}{k}^{-1/2} \sum_{S\subseteq[2n],|S|=k} |\gamma_S\rangle\!\rangle|\gamma_S\rangle\!\rangle \tag{55}$$

$$|\mathcal{R}_{k_1,k_2,k_3}^{(3)}\rangle\!\rangle = \binom{2n}{k_1,k_2,k_3,2n-k_1-k_2-k_3}^{-1/2} \sum_{\substack{S_1,S_2,S_3\subseteq[2n]\text{ disjoint}\\ |S_j|=k_j\forall j\in[3]}} |\gamma_{S_1}\gamma_{S_2}\rangle\!\rangle|\gamma_{S_2}\gamma_{S_3}\rangle\!\rangle|\gamma_{S_3}\gamma_{S_1}\rangle\!\rangle. \tag{56}$$

It is easy to see that we can further generalize $|\mathcal{R}_k^{(2)}\rangle\!\rangle$, $|\mathcal{R}_{k_1,k_2,k_3}^{(3)}\rangle\!\rangle$ in Eqs. (55) and (56) to $|\widetilde{\mathcal{R}}_k^{(2)}\rangle\!\rangle$, $|\widetilde{\mathcal{R}}_{k_1,k_2,k_3}^{(3)}\rangle\!\rangle$ in the basis of $\widetilde{\gamma}_S$. Specifically,

$$|\widetilde{\mathcal{R}}_k^{(2)}\rangle\!\rangle = \binom{2n}{k}^{-1/2} \sum_{S\subseteq[2n],|S|=k} |\widetilde{\gamma}_S\rangle\!\rangle|\widetilde{\gamma}_S\rangle\!\rangle \tag{57}$$

$$|\widetilde{\mathcal{R}}_{k_1,k_2,k_3}^{(3)}\rangle\!\rangle = \binom{2n}{k_1,k_2,k_3,2n-k_1-k_2-k_3}^{-1/2} \sum_{\substack{S_1,S_2,S_3\subseteq[2n]\text{ disjoint}\\ |S_j|=k_j\forall j\in[3]}} |\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}\rangle\!\rangle|\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}\rangle\!\rangle|\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rangle\!\rangle. \tag{58}$$

SUPPLEMENTARY NOTE 2

Two states that characterize a fermionic system are the Gaussian state and the Slater determinant. Gaussian states are ground or thermal states of fermionic systems, and all information about the state is encoded in its $2N \times 2N$ Hamiltonian. The Slater determinant describes the wave function of a many-particle (fermionic) system. We give formal definitions of them in the following.

Gaussian state. The Gaussian state is defined as

$$\rho_g = \prod_{k=1}^{n} \frac{(\mathbb{I} - i\mu_k\widetilde{\gamma}_{2k-1}\widetilde{\gamma}_{2k})}{2} \tag{59}$$

for some coefficients $\mu_k \in [-1, 1]$, where $\widetilde{\gamma}_k = U_Q^\dagger \gamma_k U_Q$ for some $Q \in \mathrm{Orth}(2n)$. The covariance matrix for a Gaussian state is a $2n \times 2n$ matrix with the (k, l)-th element being $-i\mathrm{Tr}\left([\gamma_k, \gamma_l]\rho_g\right)$. The computational basis $|x\rangle\langle x|$ is a special Gaussian state with the representation

$$|x\rangle\langle x| = \prod_{k=1}^{n} \frac{1}{2}(\mathbb{I} - i(-1)^{x_j}\gamma_{2j-1}\gamma_{2j}), \tag{60}$$

and the covariance matrix representation $C_{|x\rangle}$ is equal to

$$C_{|x\rangle} = \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & (-1)^{x_j} \\ (-1)^{x_j+1} & 0 \end{pmatrix}. \tag{61}$$

The covariance matrix for a general Gaussian state defined in Eq. (59) is

$$C_{\rho_g} = Q \bigoplus_{j=1}^{n} \begin{pmatrix} 0 & \mu_j \\ -\mu_j & 0 \end{pmatrix} Q^T. \tag{62}$$

Slater determinant. The τ-fermion Slater determinant $|\phi_\tau\rangle$ is defined as $|\phi_\tau\rangle = \tilde{b}_1^\dagger \cdots \tilde{b}_\tau^\dagger |\mathbf{0}\rangle$, and $\tilde{b}_k = \sum_{l=1}^{n} U_{kl} b_l$ for some unitary $U \in \mathbb{C}^{n\times n}$.

SUPPLEMENTARY NOTE 3

Recall that the fermionic shadow scheme relies on a twirl of the full space that divides into $2n + 1$ even subspaces Γ_{2k} for $k \in \{0, ..., n\}$. In this section, we will investigate the range of the average fidelity parameters $\mathcal{B}_k$ under different noise scenarios, where $k \geq 1$. To start, we will prove that in (0) noise-free case, the average noise fidelities $\mathcal{B}_k$ equals one for any k.

The noise channels under investigation, which are also discussed in the main text, represent typical noise models applicable to various architectures, containing (1) depolarizing noise, a channel that gives the probability of returning the state itself or the maximally mixed state; (2) amplitude-damping noise, a model of physical processes such as spontaneous emission; (3) X-rotation noise, which describes the rotation offset around the x-axis; and (4) Gaussian unitary noise, which is common in continuous variable systems and has the form of a normal distribution [90, 91].

Since Λ is a completely positive and trace-preserving map, we can write

$$\Lambda(\cdot) = \sum_l E_l(\cdot)E_l^\dagger \tag{63}$$

in its Kraus decomposition such that $\sum_l E_l^\dagger E_l = \mathbb{I}$. In the following, we analyze the range of $\mathcal{B}_k$ for the noise-free model and different common noise models.

(0) By the definition in Eq. (5) in the main file, for a noise-free model where Λ implements the identity transformation $\mathcal{I}$,

$$\mathcal{B}_k = 2^{-n}(-i)^k \binom{n}{k}^{-1} \sum_x \sum_{S\in\binom{[n]}{k}} (-1)^{\sum_{j\in S} x_j} (-i)^k (-1)^{\sum_{j\in S} x_j} \mathrm{Tr}\left(\gamma_{D(S)}^2\right) 2^{-n} \tag{64}$$

$$= 1. \tag{65}$$

(1) For the depolarizing noise with channel representation $\Lambda_{\mathrm{d}}(A) = (1-p)A + p\mathrm{Tr}(A)\frac{\mathbb{I}}{2^n}$ for any linear operator A, where $p \in [0,1]$ is the error probability, we have

$$\mathcal{B}_k = \frac{(-i)^k}{2^n\binom{n}{k}} \sum_x \sum_{S\in\binom{[n]}{k}} (-1)^{\sum_{j\in S} x_j}(1-p)(-i)^k(-1)^{\sum_{j\in S} x_j}(-1)^k \tag{66}$$

$$= 1-p. \tag{67}$$

(2) The general amplitude-damping noise can be denoted as the following equation with Kraus decomposition representation

$$\Lambda_{\mathrm{a}}(A) = \sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} E_{uv}AE_{uv}^\dagger + E_0AE_0^\dagger, \tag{68}$$

where $E_{uv} = \sqrt{p_{uv}}\,|v\rangle\langle u|$, and $E_0 = \sqrt{\mathbb{I} - \sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} E_{uv}^\dagger E_{uv}}$, and the probabilities satisfy $\sum_{v\in\{0,1\}^n\setminus\{u\}} p_{uv} \leq 1$. Here we only analyze the bound of $\mathcal{B}_k$ for the case that all of p_{uv} are the same for $v \in \{0,1\}^n \setminus \{u\}$, and let it be $\bar{p}_u$ since usually p_{uv} are close to each other for a given v. Then the probability that an error happens with output $|u\rangle$ equals $\sum_v p_{vu} = \sum_v \bar{p}_v \in [0,1]$. Typically, the probability of this error is nearly negligible and is bounded by a constant in an actual quantum device. We leave the analysis of the general amplitude-damping noise to interested readers. Note that

$$\sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} E_{uv}^\dagger|x\rangle\langle x|E_{uv} + E_0^\dagger|x\rangle\langle x|E_0 = p_{xx}|x\rangle\langle x| + \sum_{x'\in\{0,1\}^n\setminus\{x\}} p_{x'x}\left|x'\right\rangle\left\langle x'\right| \tag{69}$$

$$= (1-2^n\bar{p}_x)|x\rangle\langle x| + \bar{p}_x\sum_{x'\in\{0,1\}^n}\left|x'\right\rangle\left\langle x'\right|. \tag{70}$$

With the definitions of $\mathcal{B}_k$ and Λ_{a} we have

$$\mathcal{B}_k = \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{x,S\in\binom{[n]}{k}} (-1)^{\sum_{j\in S} x_j}\left(\sum_{\substack{u,v\in\{0,1\}^n \\ u\neq v}} \mathrm{Tr}\left(E_{uv}^\dagger|x\rangle\langle x|E_{uv}\gamma_{D(S)}\right) + \mathrm{Tr}\left(E_0^\dagger|x\rangle\langle x|E_0\gamma_{D(S)}\right)\right) \tag{71}$$

$$= \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{x,S\in\binom{[n]}{k}} (-1)^{\sum_{j\in S} x_j}\left((1-2^n\bar{p}_x)\mathrm{Tr}\left(|x\rangle\langle x|\gamma_{D(S)}\right) + \bar{p}_x\sum_{x'\in\{0,1\}^n}\mathrm{Tr}\left(\left|x'\right\rangle\left\langle x'\right|\gamma_{D(S)}\right)\right) \tag{72}$$

$$= \frac{\sum_x(1-2^n\bar{p}_x)}{2^n} + \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{x,S\in\binom{[n]}{k}} (-1)^{\sum_{j\in S} x_j}\bar{p}_x \sum_{x'\in\{0,1\}^n}(-i)^k(-1)^{\sum_{j\in S} x'_j}(-1)^k \tag{73}$$

$$= 1 - \sum_x \bar{p}_x, \tag{74}$$

if $k \neq 0$, and $\mathcal{B}_k = 1$ if $k = 0$. Eq. (72) holds by the result in Eq. (70). Since usually the noise strength $\sum_x \bar{p}_x$ in the device is close to zero, $\mathcal{B}_k$ is close to one.

(3) For X-rotation noise with channel representation

$$\Lambda_{\mathrm{r}}(\cdot) = R_X(\boldsymbol{\theta})(\cdot)R_X(-\boldsymbol{\theta}) \tag{75}$$

where $R_X(\boldsymbol{\theta}) = \exp(-i\sum_{l=1}^{n}\theta_l X_l/2)$ and $\theta_l \in [0, 2\pi)$, we have

$$\begin{aligned}
\mathcal{B}_k =& \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{S\in\binom{[n]}{k}} \sum_{x} (-1)^{\sum_{l\in S} x_l} \mathrm{Tr}\left(|x\rangle\langle x|\, \Lambda_{\mathrm{r}}(\gamma_{D(S)})\right) && (76)\\
=& \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{S\in\binom{[n]}{k}} \sum_{x} (-1)^{\sum_{l\in S} x_l} \mathrm{Tr}\left(R_X(-\boldsymbol{\theta})\, |x\rangle\langle x|\, R_X(\boldsymbol{\theta})\gamma_{D(S)}\right) && (77)\\
=& \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{S\in\binom{[n]}{k}} \sum_{x} \prod_{l\in S} i(-1)^{x_l}\left(\cos^2\frac{\theta_l}{2} - \sin^2\frac{\theta_l}{2}\right)(-1)^{x_l} && (78)\\
=& \binom{n}{k}^{-1} \sum_{S\in\binom{[n]}{k}} \prod_{l\in S} \cos\theta_l, && (79)
\end{aligned}$$

where Eq. (78) holds since $R_X(-\boldsymbol{\theta})\,|x\rangle\langle x|\,R_X(\boldsymbol{\theta})$ has the expansion form

$$R_X(-\boldsymbol{\theta})\,|x\rangle\langle x|\,R_X(\boldsymbol{\theta}) = \bigotimes_{l=1}^{n} \begin{pmatrix} \cos^2\frac{\theta_l}{2} & -i\sin\frac{\theta_l}{2}\cos\frac{\theta_l}{2} \\ i\sin\frac{\theta_l}{2}\cos\frac{\theta_l}{2} & \sin^2\frac{\theta_l}{2} \end{pmatrix} \tag{80}$$

with basis $\{|x_l\rangle, |1-x_l\rangle\}$, and $\gamma_{D(S)} = \prod_{l\in S}(iZ_l)$, where Z is the Pauli-Z gate.

(4) For the noise with the representation being a Gaussian unitary channel $\mathcal{U}_Q$, we have

$$\begin{aligned}
\mathcal{B}_k &= \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{S\in\binom{[n]}{k}} \sum_{x} (-1)^{\sum_{l\in S} x_l} \mathrm{Tr}\left(|x\rangle\langle x|\, U_Q\gamma_{D(S)}U_Q^\dagger\right) && (81)\\
&= \frac{(-i)^k}{2^n\binom{n}{k}} \sum_{S,S'\in\binom{[n]}{k}} \sum_{x} (-i)^k \det\left(Q|_{D(S'),D(S)}\right)(-1)^k && (82)\\
&= \frac{1}{\binom{n}{k}} \sum_{S,S'\in\binom{[n]}{k}} \det\left(Q|_{D(S'),D(S)}\right). && (83)
\end{aligned}$$

Below, we provide an illustration where selecting Q from a signed permutation matrix demonstrates that $\mathcal{B}_k$ has the potential to become zero. Note that Q is a signed permutation and hence $\det(Q|_{D(S'),D(S)}) \neq 0$ iff $Q|_{D(S'),D(S)}$ maps $D(S')$ to $D(S)$. Let the vector permutation representation of it be σ_Q, where $\sigma_{Qi} = jQ_{ij}$ if $Q_{ij} \neq 0$. Since $D(j) = (2j-1, 2j)$, $\det\left(Q|_{D(S),D(S')}\right) \neq 0$ iff $\sigma_Q|_{D(S)}$ is a signed permutation of $D(S')$. Let $P \subset [k]$ be the largest set such that $\sigma_Q|_{D(P)}$ is a permutation for some $D(P')$ where $\mathrm{size}(P') = \mathrm{size}(P) =: r$. Consequently, $\mathcal{B}_k = 0$ if $r < k < n$. For instance, take $\sigma_Q = (3, 4, 1, 5, 6, 8, 7, 2)$ with $n = 4$. This leads to $\mathcal{B}_k = 0$ when $k \in \{2, 3\}$, since $r = 1$ for both of these cases.

SUPPLEMENTARY NOTE 4

Here we give proofs associated with the properties of the noisy classical fermionic channel.

Proof of Theorem 2 in the main file. The noisy channel $\widetilde{\mathcal{M}}$ can be represented as

$$\widetilde{\mathcal{M}} = \int_Q d\mu(Q)\mathcal{U}_Q^\dagger M_z \Lambda \mathcal{U}_Q = \sum_{k=0}^{n} f_{2k}\mathcal{P}_{2k}, \tag{84}$$

since $\mathcal{P}_{2k}$ forms $n+1$ irreducible representations for the unitary channel $\mathcal{U}_Q$ (See details for the irreducible representations of a group in Supplementary Note 1). Then we have

$$\int_Q d\mu(Q)\mathbb{E}_x\left[\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle\right] = \langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{M}|\mathbf{0}\rangle\rangle = f_{2k}\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle. \tag{85}$$

By using Lemma 5, we can obtain $\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle = 2^{-n}\binom{n}{k}$. Combining this result with the previous findings leads us to

$$\begin{aligned}
\mathbb{E}[\hat{f}_{2k}] &= 2^n\binom{n}{k}^{-1}\int_Q d\mu(Q)\mathbb{E}_x\left[\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle\right] &(86)\\
&= 2^n\binom{n}{k}^{-1} f_{2k}\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle &(87)\\
&= f_{2k}. &(88)
\end{aligned}$$

□

Below, we present proof of Lemma 1 in the main file.

Proof of Lemma 1 in the main file. Recall that $\hat{f}_{2k} = 2^n\binom{n}{k}^{-1}\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle$. Hence, by Lemma 8 (It can also be obtained by Proposition 1 in Ref. [10]), we have $\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle$ is the coefficient of x^k in the polynomial

$$p_Q(x) = \frac{1}{2^n}\mathrm{pf}\left(C_{\mathbf{0}}\right)\mathrm{pf}\left(-C_{\mathbf{0}}^{-1} + x(QC_{|x\rangle}Q^T)\right). \tag{89}$$

Hence $\hat{f}_{2k}$ is the coefficient of x^k in the polynomial $2^n\binom{n}{k}^{-1}p_Q(x) = \binom{n}{k}^{-1}\mathrm{pf}\left(C_{\mathbf{0}}\right)\mathrm{pf}\left(-C_{\mathbf{0}}^{-1} + x(QC_{|x\rangle}Q^T)\right)$. The polynomial in Eq. (89) can be calculated in $\mathcal{O}(r^4)$ time since it has degree at most r, and the Pfaffian function of a $2n\times 2n$ matrix can be calculated in $\mathcal{O}(n^3)$ time. It can be further optimized to $\mathcal{O}(r^3)$ time by Appendix D of Ref. [10], where r is the rank of the covariance matrix. □

The variance can be bounded for the estimation with the bound of f_{2k} and $\mathbb{E}\left[f_{2k}^2\right]$. In the following, we give proofs associated with the calculation of $\hat{f}_{2k}$. Since $\int_Q d\mu(Q)\mathbb{E}_x\left[\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle\right] = f_{2k}\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle$, we give the closed form for $\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle$ with the following lemma.

Lemma 5.

$$\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle = 2^{-n}\binom{n}{k}. \tag{90}$$

Proof. Recall that $|\mathbf{0}\rangle\langle\mathbf{0}| = \prod_{j=1}^{n} \frac{(\mathbb{I}-i\gamma_{2j-1}\gamma_{2j})}{2}$. Hence, by substituting it into Eq. (90), we obtain

$$\langle\langle \mathbf{0}|\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle = \sum_{S\in\binom{[2n]}{2k}} \left|\mathrm{Tr}\left(|\mathbf{0}\rangle\langle\mathbf{0}|\,\gamma_S\right)\right|^2 \tag{91}$$

$$= \sum_{S\in\binom{[2n]}{2k}} \sum_{j=0}^{n} \sum_{T\in\binom{[n]}{k}} \left|\frac{(-i)^k}{4^n}\mathrm{Tr}\left(\gamma_{D(T)}\gamma_S\right)\right|^2 \tag{92}$$

$$= \sum_{T\in\binom{[n]}{k}} \frac{1}{2^n} \tag{93}$$

$$= \frac{\binom{n}{k}}{2^n}, \tag{94}$$

where $D(S) = \{(2j-1, 2j)|j \in S\}$. □

As a generalization of Proposition 2 of Chen et al. [11], we give the bounds for f_{2k} and $\mathbb{E}\left[\hat{f}_{2k}^2\right]$, with prior knowledge of the average noise fidelity $\mathcal{B}_k$ in the following two lemmas.

Lemma 6.

$$f_{2k} = \binom{2n}{2k}^{-1}\binom{n}{k}\mathcal{B}_k. \tag{95}$$

Proof. Since $f_{2k} = \mathbb{E}\left[\hat{f}_{2k}\right]$, to find the value of f_{2k}, we can evaluate the expansion of $\hat{f}_{2k}$ as follows,

$$\mathbb{E}\left[\hat{f}_{2k}\right] = 2^n\binom{n}{k}^{-1}\sum_x \int_Q d\mu(Q)\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle\langle\langle x|\Lambda\mathcal{U}_Q|\mathbf{0}\rangle\rangle \tag{96}$$

$$= 2^n\binom{n}{k}^{-1}\sum_x \int_Q d\mu(Q)\left[\mathrm{Tr}\left(\mathcal{U}_Q^{\otimes 2}|\mathbf{0}\rangle\rangle^{\otimes 2}\langle\langle x|\mathcal{P}_{2k}\langle\langle x|\Lambda\right)\right] \tag{97}$$

$$= 2^n\binom{n}{k}^{-1}\sum_x \sum_{j=0}^{2n} \mathrm{Tr}\left(|\mathcal{R}_j^{(2)}\rangle\rangle\langle\langle \mathcal{R}_j^{(2)}|\cdot|\mathbf{0}\rangle\rangle^{\otimes 2}\langle\langle x|\mathcal{P}_{2k}\langle\langle x|\Lambda\right) \tag{98}$$

$$= 2^n\binom{n}{k}^{-1}\sum_{x,0\le j\le 2n} \langle\langle \mathcal{R}_j^{(2)}||\mathbf{0}\rangle\rangle^{\otimes 2}\left(\langle\langle x|\mathcal{P}_{2k}\otimes\langle\langle x|\Lambda\right)|\mathcal{R}_j^{(2)}\rangle\rangle, \tag{99}$$

where $|\mathcal{R}_j^{(2)}\rangle\rangle = \binom{2n}{j}^{-1/2}\sum_{S\subseteq[2n],|S|=j}|\gamma_S\rangle\rangle|\gamma_S\rangle\rangle$. Eq. (97) holds since $\langle\langle \mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\rangle = \langle\langle x|\mathcal{U}_Q\mathcal{P}_{2k}|\mathbf{0}\rangle\rangle = \langle\langle x|\mathcal{P}_{2k}\mathcal{U}_Q|\mathbf{0}\rangle\rangle$. Eq. (98) holds by Lemma 4. By the definition of $\mathcal{R}_j^{(2)}$, we have

$$\langle\langle \mathcal{R}_j^{(2)}||\mathbf{0}\rangle\rangle^{\otimes 2} = \binom{2n}{j}^{-1/2}\sum_{S\subseteq[2n],|S|=j}\langle\langle \gamma_S|\mathbf{0}\rangle\rangle^2 \tag{100}$$

$$= \binom{2n}{j}^{-1/2}\binom{n}{j/2}[j \text{ even}]\frac{(-1)^{j/2}}{2^n} \tag{101}$$

and

$$(\langle\!\langle x|\mathcal{P}_{2k}\otimes\langle\!\langle x|\Lambda)|\mathcal{R}_j^{(2)}\rangle\!\rangle = \binom{2n}{j}^{-1/2}\sum_{S\subseteq[2n],|S|=j}\langle\!\langle x|\mathcal{P}_{2k}|\gamma_S\rangle\!\rangle\langle\!\langle x|\Lambda|\gamma_S\rangle\!\rangle \tag{102}$$

$$= [j=2k]\binom{2n}{2k}^{-1/2}\sum_{S\in\binom{[n]}{k}}\frac{i^k(-1)^{\sum_{l\in S}x_l}}{\sqrt{2^n}}\langle\!\langle x|\Lambda|\gamma_{D(S)}\rangle\!\rangle \tag{103}$$

where $D(S)=\{2j-1,2j|j\in S\}$. By substituting Eq. (101), and Eq. (103) into Eq. (98), we have

$$f_{2k} = (-i)^k\binom{2n}{2k}^{-1}\sum_{s\in\binom{[n]}{k}}\sum_x(-1)^{\sum_{l\in S}x_l}\mathrm{Tr}\left(|x\rangle\langle x|\,\Lambda(\gamma_{D(S)})\right)2^{-n} \tag{104}$$

$$= \binom{2n}{2k}^{-1}\binom{n}{k}\mathcal{B}_k, \tag{105}$$

where $\mathcal{B}_k$ is defined in Eq. (5) in the main file. □

For the noise-free model, $f_{2k}=\binom{2n}{2k}^{-1}\binom{n}{k}$. We can use a similar technique to show that $\mathbb{E}\left[\hat{f}_{2k}^2\right]$ is also bounded, as demonstrated in the following lemma.

Lemma 7. *Let* $\hat{f}_{2k}=2^n\langle\!\langle\mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\!\rangle/\binom{n}{k}$ *be an estimation of* f_{2k}*, then*

$$\mathbb{E}\left[\left|\hat{f}_{2k}\right|^2\right] = \binom{n}{k}^{-2}\sum_{0\le l\le k}\binom{2n}{2l,2k-2l,2l}_p^{-1}\binom{n}{l,k-l,l}_p^2\mathcal{B}_{2l}. \tag{106}$$

Proof. Firstly, we note that $\langle\!\langle\mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\!\rangle^* = \sum_{S\in\binom{[2n]}{2k}}\langle\!\langle\mathbf{0}|\gamma_S^\dagger\rangle\!\rangle\langle\!\langle\gamma_S^\dagger|\mathcal{U}_Q^\dagger|x\rangle\!\rangle = \langle\!\langle\mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q^\dagger|x\rangle\!\rangle$ is a real number. Hence the estimation $\hat{f}_{2k}$ is a real number, and

$$\mathbb{E}\left[\left|\hat{f}_{2k}\right|^2\right] = 2^{2n}\binom{n}{k}^{-2}\sum_{\boldsymbol{x}}\int_Q d\mu(Q)\left[\langle\!\langle\mathbf{0}|\mathcal{P}_{2k}\mathcal{U}_Q|\boldsymbol{x}\rangle\!\rangle^2\langle\!\langle\boldsymbol{x}|\Lambda\mathcal{U}_Q|\mathbf{0}\rangle\!\rangle\right] \tag{107}$$

$$= 2^{2n}\binom{n}{k}^{-2}\sum_{\boldsymbol{x}}\int_Q d\mu(Q)\left[\mathrm{Tr}\left(\mathcal{U}_Q^{\otimes3}|\mathbf{0}\rangle\!\rangle^{\otimes3}\left(\langle\!\langle\boldsymbol{x}|\mathcal{P}_{2k}\otimes\langle\!\langle\boldsymbol{x}|\mathcal{P}_{2k}\otimes\langle\!\langle\boldsymbol{x}|\Lambda\right)\right)\right] \tag{108}$$

$$= 2^{2n}\binom{n}{k}^{-2}\sum_x\sum_{\substack{k_1,k_2,k_3\ge0\\k_1+k_2+k_3\le2n}}\langle\!\langle\mathcal{R}_{k_1,k_2,k_3}|\mathbf{0}^{\otimes3}\rangle\!\rangle\left(\langle\!\langle x|\mathcal{P}_{2k}\otimes\langle\!\langle x|\mathcal{P}_{2k}\otimes\langle\!\langle x|\Lambda\right)|\mathcal{R}_{k_1,k_2,k_3}\rangle\!\rangle \tag{109}$$

where Eq. (109) holds due to Lemma 4. By the definition of $\mathcal{R}_{k_1,k_2,k_3}$, we have

$$\langle\!\langle\mathcal{R}_{k_1,k_2,k_3}|\mathbf{0}^{\otimes3}\rangle\!\rangle = \binom{2n}{k_1,k_2,k_3}_p^{-1/2}\sum_{\substack{S_1,S_2,S_3\subseteq[2n]\text{ disjoint}\\|S_i|=k_i,i\in[3]}}\langle\!\langle\gamma_{S_1}\gamma_{S_2}|\mathbf{0}\rangle\!\rangle\langle\!\langle\gamma_{S_2}\gamma_{S_3}|\mathbf{0}\rangle\!\rangle\langle\!\langle\gamma_{S_3}\gamma_{S_1}|\mathbf{0}\rangle\!\rangle \tag{110}$$

$$= \binom{2n}{k_1,k_2,k_3}_p^{-1/2}[k_1,k_2,k_3\text{ even}](-1)^{(k_1+k_2+k_3)/2}2^{-3n/2}\binom{n}{\frac{k_1}{2},\frac{k_2}{2},\frac{k_3}{2}}_p \tag{111}$$

where $\binom{n}{k_1,k_2,k_3}_p = \binom{n}{k_1,k_2,k_3,n-k_1-k_2-k_3}$, and

$$
\begin{aligned}
&(\langle\!\langle x|\mathcal{P}_{2k} \otimes \langle\!\langle x|\mathcal{P}_{2k} \otimes \langle\!\langle x|\Lambda)\,|\mathcal{R}_{k_1,k_2,k_3}\rangle\!\rangle \\
&= \binom{2n}{k_1,k_2,k_3}_p^{-1/2} \sum_{\substack{S_1,S_2,S_3\subseteq[2n] \text{ disjoint} \\ |S_i|=k_i, i\in[3]}} \langle\!\langle x|\mathcal{P}_{2k}|\gamma_{S_1}\gamma_{S_2}\rangle\!\rangle \langle\!\langle x|\mathcal{P}_{2k}|\gamma_{S_2}\gamma_{S_3}\rangle\!\rangle \langle\!\langle x|\Lambda|\gamma_{S_3}\gamma_{S_1}\rangle\!\rangle \qquad (112) \\
&= \binom{2n}{k_1,k_2,k_3}_p^{-1/2} [k_1 = k_3 \text{ even}, k_2 = 2k-k_1] \sum_{\substack{S_1,S_2,S_3\subseteq[n] \text{ disjoint} \\ |S_1|=|S_3|=\frac{k_1}{2}, |S_2|=k-\frac{k_1}{2}}} \frac{i^{|S_1|+2|S_2|+|S_3|}(-1)^{\sum_{l\in S_1\cup S_3} x_l} \langle\!\langle x|\Lambda|\gamma_{D(S_3\cup S_1)}\rangle\!\rangle}{2^n} \qquad (113) \\
&= \binom{2n}{2l,2k-2l,2l}_p^{-1/2} [k_1 = k_3 = 2l, k_2 = 2k-2l] \\
&\quad \times \sum_{\substack{S_1,S_3\subseteq[n] \text{ disjoint} \\ |S_1|=|S_3|=l}} \frac{(-1)^k}{2^{3n/2}} \binom{n-2l}{k-l} \mathrm{Tr}\left(\gamma_{D(S_1\cup S_3)}\Lambda(|x\rangle\langle x|)\right)^* (-1)^{\sum_{l\in S_1\cup S_3} x_l} \qquad (114) \\
&= \binom{2n}{2l,2k-2l,2l}_p^{-1/2} [k_1 = k_3 = 2l, k_2 = 2k-2l] \sum_{S\in\binom{[n]}{2l}} \frac{(-1)^k}{2^{3n/2}} \binom{n-2l}{k-l}\binom{2l}{l} \mathrm{Tr}\left(|x\rangle\langle x|\,\Lambda(\gamma_{D(S)})\right)(-1)^{\sum_{l\in S} x_l}. \qquad (115)
\end{aligned}
$$

Substituting Eq. (111) and Eq. (115) into Eq. (109) gives us the following equality

$$
\mathbb{E}\left[\hat{f}_{2k}^2\right] = \binom{n}{k}^{-2} \sum_{0\le l\le \min(k,n-k)} \binom{2n}{2l,2k-2l,2l}_p^{-1} \binom{n}{l,k-l,l}_p^2 \mathcal{B}_{2l}, \qquad (116)
$$

where $\mathcal{B}_l$ is defined in Eq. (5) in the main file. □

In Supplementary Note 6, we will give the upper bounds for the number of samples R_e, R_c to bound estimation and calibration errors $\varepsilon_e, \varepsilon_c$ respectively, based on the results of the above two lemmas.

SUPPLEMENTARY NOTE 5

In this section, we analyze the simplification of the estimations of several physical quantities with error-mitigated matchgate shadow.

Lemma 8. *Given $Q \in \mathrm{Orth}(2n)$, a Majorana operator γ_S and computational basis state $|x\rangle$, where $x \in \{0,1\}^n$, we have*

$$
Tr\left(U_Q^\dagger \gamma_S U_Q\,|x\rangle\langle x|\right) = i^{|S|/2}\mathrm{pf}(QC_{|x\rangle}Q^T|_S), \qquad (117)
$$

if $|S|$ is even, and zero otherwise.

Proof. It is easy to check that if $|S|$ is odd, then $\mathrm{Tr}\left(U_Q^\dagger \gamma_S U_Q\,|x\rangle\langle x|\right) = 0$. In the following, we assume that $|S|$ is even. With the definitions of $C_{|x\rangle}$ and the Pfaffian function, we have $\mathrm{Tr}\left(\gamma_S\,|x\rangle\langle x|\right) = i^{|S|/2}\mathrm{pf}(C_{|x\rangle}|_S)$.

By the properties of the matchgate operation on the Majorana operator, we have

$$\mathrm{Tr}\left(U_Q^\dagger \gamma_S U_Q \left|x\right\rangle\!\left\langle x\right|\right) = \sum_{S' \in \binom{[n]}{|S|}} \det(Q|_{SS'}) \mathrm{Tr}\left(\gamma_{S'} \left|x\right\rangle\!\left\langle x\right|\right) \tag{118}$$

$$= i^{|S|/2} \mathrm{pf}(Q C_{|x\rangle} Q^T|_S), \tag{119}$$

where the last equality holds since $\det(A)\mathrm{pf}(B) = \mathrm{pf}(ABA^T)$ for any $2n \times 2n$ matrix A and a skew-symmetric matrix B (A *skew-symmetric* matrix B is one that satisfies $B^T = -B$) [92]. □

Let $\hat{v}$ be the estimation result obtained by computing $\mathrm{Tr}\left(\widetilde{\gamma}_S \widehat{\mathcal{M}}^{-1}\left(U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q\right)\right)$ where $\widetilde{\gamma}_S = U_{Q_1}^\dagger \gamma_S U_{Q_1}$. We can simplify the estimate of the expectation value of the Majorana operator $\widetilde{\gamma}_S$ as follows:

$$\hat{v} = \mathrm{Tr}\left(\widetilde{\gamma}_S \widehat{\mathcal{M}}^{-1}(U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q)\right) \tag{120}$$

$$= \mathrm{Tr}\left(\gamma_S \widehat{\mathcal{M}}^{-1}(U_{Q_1} U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q U_{Q_1}^\dagger)\right) \tag{121}$$

$$= \frac{1}{\hat{f}_{|S|} 2^n} \sum_{S' \in \binom{[2n]}{|S|}} \mathrm{Tr}\left(\gamma_S \gamma_{S'}\right) \mathrm{Tr}\left(\gamma_{S'} U_{Q_1} U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q U_{Q_1}^\dagger\right) \tag{122}$$

$$= \frac{i^{|S|/2}}{\hat{f}_{|S|}} \mathrm{pf}(Q_1 Q^T C_{|x\rangle} Q Q_1^T|_S), \tag{123}$$

where $\widetilde{\gamma}_j = \sum_{k=1}^{2n} \det(Q_1(j,k))\gamma_k$, and $Q_1(j,k)$ is the (j,k)-th element of matrix Q_1.

Wan et al. [10] introduce techniques to determine the overlap between a Gaussian state and any quantum state, along with a τ-slater determinant and any pure state. Notably, the distinction between the noisy channel and the original CS channel boils down to the varying coefficients f_{2k}. To ensure the integrity of our paper, we outline the equations used to calculate these quantities below. One can find detailed information about the deviation in Ref. [10].

(1) Calculate the overlap between an n-qubit Gaussian state with density matrix ρ_g and any n-qubit quantum state with density matrix ρ, denoted as $\mathrm{Tr}(\rho\rho_g)$. By letting $H = \rho_g$, the estimator can be calculated efficiently since $\mathrm{Tr}\left(\rho_g \mathcal{P}_{2k}\left(U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q\right)\right)$ is the coefficient of z^k in the polynomial $p_{\rho_g,\chi} = \frac{1}{2^n}\mathrm{pf}(C_{\rho_g})\mathrm{pf}(-C_{\rho_g}^{-1} + zC_\chi)$, where $\chi = U_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q$. Hence $\hat{v}$ can be estimated in polynomial time with the polynomial interpolation method.

(2) Calculate the inner product between τ-fermionic Slater determinants $|\phi_\tau\rangle$ and any quantum pure state $|\psi\rangle$, denoted as $\langle\psi|\phi_\tau|\psi|\phi_\tau\rangle$ and since the shadow channel $\widetilde{\mathcal{M}}$ twirls any observable to the even subspace, here we assume $|\psi\rangle$ is in the odd subspace. We require $n+1$ qubits if τ is even (and $n+2$ qubits otherwise) and suppose $|\psi\rangle = U_\psi |\mathbf{0}\rangle$. By performing Hadamard gate on the first qubit, followed by control-U_ψ gate, we generate the initial state $\rho = \frac{(|0^{n+1}\rangle + |1\rangle|\psi\rangle)(\langle 0^{n+1}| + \langle 1|\langle\psi|)}{2}$, and let observable $H = |1\rangle|\phi_\tau\rangle\langle 0^{n+1}|$. It is easy to check $\mathrm{Tr}(\rho H) = \langle\psi|\phi_\tau|\psi|\phi_\tau\rangle/2$. The quantity $\mathrm{Tr}\left(HU_Q^\dagger \left|x\right\rangle\!\left\langle x\right| U_Q\right)$ is the coefficient of z^k in the polynomial $p_{\phi,\chi} = \frac{i^{\tau/2}}{2^{n-\tau/2}}\mathrm{pf}\left(C_{\mathbf{0}} + zQ'^T \widetilde{Q} C_\chi \widetilde{Q}^T Q'|E_\tau\right)$, where $C_{\boldsymbol{\rho}}$ is the covariance matrix for ρ, $Q' = \bigoplus_{j\in[\tau]} \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -i\\ 1 & i\end{pmatrix} \bigoplus_{j=\tau+1}^{n} \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ and $E_\tau = [2n]\backslash\{1,3,\ldots,2\tau-1\}$, and we can get all of the coefficients in $\mathcal{O}\left(n^4\right)$ time, hence the estimator $\hat{v}$ can be obtained efficiently.

SUPPLEMENTARY NOTE 6

To limit the estimation error, we first provide the **MedianOfMeans** method [21], as shown in Algorithm 2. To give a better bound, the number of repetitions is usually set as $K = \mathcal{O}(\mathrm{Var}(\hat{v})/\varepsilon^2)$ and $N = \mathcal{O}(\ln \delta^{-1})$ in the algorithm, where $\hat{v}$ is the estimation in a single round.

Algorithm 2 Median of Means

Input The number of repetitions N, K, and estimations $\{v_j\}_{j=1}^{R}$, where $R = NK$.

Output Estimation $\bar{v}$.

for $l \leftarrow 1$ to K **do**
  $\bar{v}_l := \frac{1}{N}\sum_{j=(l-1)N+1}^{KN} v_j$;
end for
$\bar{v} = \mathrm{median}(v_1, \ldots, v_K)$;
Return $\bar{v}$

Lemma 9 (Theorem S1 in Ref. [21]). *Let $S = \{\hat{v}_1, \ldots, \hat{v}_R\}$ be $R = NK$ identical and independent samples from the same distribution with $N = 34\mathrm{Var}(\hat{v})/\varepsilon^2$ for any $\hat{v} \in S$, and $K = 2\ln\left(2\delta^{-1}\right)$ for some precision parameters $\varepsilon, \delta \in [0,1]$, then the estimation $\bar{v}$ in Algorithm 2 satisfies*

$$\Pr[|\bar{v} - \mathbb{E}[\hat{v}]| \geq \varepsilon] \leq \delta. \tag{124}$$

The required number of samples for accurate estimation can be upper bounded by the variance of estimation $\hat{v}_i^{(j)}$ in a single round of sampling (as shown in Step 14 of Algorithm 1 in the main file) by utilizing **MedianOfMeans** method [21]. We give the variance of a single estimation $\hat{v}$ in the following lemma.

Lemma 10. *For any observable H and an unknown quantum state ρ, the estimation $\hat{v}$ generated in Step 16 of Algorithm 1 in the main file for quantity $Tr\left(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right)$ satisfies*

$$\begin{aligned}\mathrm{Var}(\hat{v}) \leq \frac{(1+\varepsilon_c)^2}{2^{2n}} &\sum_{0\leq l_1+l_2+l_3\leq n} \frac{(-1)^{l_1+l_2+l_3}\binom{n}{l_1,l_2,l_3}_p \binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}\mathcal{B}_{l_1+l_3}}{\binom{2n}{2l_1,2l_2,2l_3}_p \binom{n}{l_1+l_2}\binom{n}{l_2+l_3}\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}} \\ &\times \sum_{\substack{S_1,S_2,S_3 \text{ disjoint} \\ |S_i|=2l_i, i\in[3]}} Tr\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right) Tr\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0^\dagger\right) Tr\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right),\end{aligned} \tag{125}$$

where $H_0 = H - Tr(H)\frac{\mathbb{I}}{2^n}$ is the traceless part of H.

Proof. By the definition of $\hat{v}$, we have

$$\mathrm{Var}(\hat{v}) = \mathbb{E}\left[|\hat{v} - \mathbb{E}[\hat{v}]|^2\right] \tag{126}$$

$$= \mathbb{E}_x \int_Q d\mu(Q)\left[\left|\langle\!\langle H_0^\dagger|\widehat{\mathcal{M}}^{-1}\mathcal{U}_Q^\dagger|x\rangle\!\rangle - \langle\!\langle H_0^\dagger|\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}|\rho\rangle\!\rangle\right|^2\right] \tag{127}$$

$$\leq \sum_x \int_Q d\mu(Q)\left|\langle\!\langle H_0^\dagger|\widehat{\mathcal{M}}^{-1}\mathcal{U}_Q^\dagger|x\rangle\!\rangle\right|^2 \langle\!\langle x|\Lambda\mathcal{U}_Q|\rho\rangle\!\rangle \tag{128}$$

$$= \sum_x \int_Q d\mu(Q)\mathrm{Tr}\left(\mathcal{U}_Q^{\otimes 3}\widehat{\mathcal{M}}^{-1}|H_0\rangle\!\rangle\widehat{\mathcal{M}}^{-1}|H_0^\dagger\rangle\!\rangle|\rho\rangle\!\rangle\langle\!\langle x|^{\otimes 2}\langle\!\langle x|\Lambda\right) \tag{129}$$

$$= \sum_x \sum_{0\leq k_1+k_2+k_3\leq 2n} \langle\!\langle\widetilde{\mathcal{R}}_{k_1k_2k_3}|\widehat{\mathcal{M}}^{-1}|H_0\rangle\!\rangle\widehat{\mathcal{M}}^{-1}|H_0^\dagger\rangle\!\rangle|\rho\rangle\!\rangle\langle\!\langle x|\langle\!\langle x|\Lambda|\widetilde{\mathcal{R}}_{k_1k_2k_3}\rangle\!\rangle, \tag{130}$$

where $H_0 = H - \mathrm{Tr}(H)\frac{\mathbb{I}}{2^n}$ is the traceless part of H. Eq. (129) holds since $\langle\langle H_0^\dagger|\widehat{\mathcal{M}}^{-1}\mathcal{U}_Q^\dagger|x\rangle\rangle = \langle\langle x|\mathcal{U}_Q\widehat{\mathcal{M}}^{-1}|H_0\rangle\rangle$. With the definition of $\widetilde{\mathcal{R}}_{k_1k_2k_3}$ in Eq. (56) and the definition of $\widehat{\mathcal{M}}$, we have

$$\begin{aligned}
&\langle\langle\mathcal{R}_{k_1k_2k_3}|\widehat{\mathcal{M}}^{-1}|H_0\rangle\rangle\widehat{\mathcal{M}}^{-1}|H_0^\dagger\rangle\rangle|\rho\rangle\rangle \\
&= \binom{2n}{k_1,k_2,k_3}_p^{-1/2} \sum_{\substack{S_1,S_2,S_3 \text{ disjoint}\\ |S_i|=k_i, i\in[3]\\ (k_i+k_{i+1}) \text{ even}, i\in[2]}} 2^{-3n/2}\hat{f}_{k_1+k_2}^{-1}\hat{f}_{k_2+k_3}^{-1}\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}^\dagger\widetilde{\gamma}_{S_1}^\dagger H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}^\dagger\widetilde{\gamma}_{S_2}^\dagger H_0^\dagger\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_1}^\dagger\widetilde{\gamma}_{S_3}^\dagger\rho\right) \qquad (131)\\
&\le \frac{2^{-3n/2}}{\binom{2n}{k_1,k_2,k_3}_p^{\frac{1}{2}}} \sum_{\substack{S_1,S_2,S_3 \text{ disjoint}\\ |S_i|=k_i, i\in[3]\\ (k_i+k_{i+1}) \text{ even}, i\in[2]}} \frac{(1+\varepsilon_c)^2}{f_{k_1+k_2}f_{k_2+k_3}}\mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0^\dagger\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right) \qquad (132)\\
&= \frac{2^{-3n/2}\binom{2n}{k_1+k_2}\binom{2n}{k_2+k_3}(1+\varepsilon_c)^2}{\binom{2n}{k_1,k_2,k_3}_p^{\frac{1}{2}}\binom{n}{\frac{k_1+k_2}{2}}\binom{n}{\frac{k_2+k_3}{2}}\mathcal{B}_{\frac{k_1+k_2}{2}}\mathcal{B}_{\frac{k_2+k_3}{2}}} \sum_{\substack{S_1,S_2,S_3 \text{ disjoint}\\ |S_i|=k_i, i\in[3]\\ (k_i+k_{i+1}) \text{ even}, i\in[2]}} \mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0^\dagger\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right), \qquad (133)
\end{aligned}$$

where Eq. (132) holds since $\frac{f_{2k}}{\hat{f}_{2k}} \le 1+\varepsilon_c$ for any $0 \le k \le n$ from Lemma 11, and Eq. (133) holds by Lemma 6. Furthermore,

$$\begin{aligned}
\langle\langle x|^{\otimes 2}\langle\langle x|\Lambda|\mathcal{R}_{k_1k_2k_3}\rangle\rangle &= 2^{-n}\binom{2n}{2l_1,2l_2,2l_3}_p^{-1/2} \sum_{\substack{S_1\in\binom{[n]}{l_1}\\ S_2\in\binom{[n]\setminus S_1}{l_2}\\ S_3\in\binom{[n]\setminus S_1\cup S_2}{l_3}}} (-i)^{l_1+2l_2+l_3}(-1)^{l_1+l_3}(-1)^{\sum_{j\in S_1\cup S_3}x_j}\langle\langle x|\Lambda|\widetilde{\gamma}_{D(S_1\cup S_3)}\rangle\rangle \qquad (134)\\
&= \binom{2n}{2l_1,2l_2,2l_3}_p^{-1/2}\frac{(-i)^{l_1+2l_2+l_3}}{2^{3n/2}}\binom{n-l_1-l_3}{l_2} \sum_{s\in\binom{[n]}{l_1+l_3}}(-1)^{\sum_{j\in S}x_j}\binom{l_1+l_3}{l_1}\mathrm{Tr}\left(|x\rangle\langle x|\,\Lambda(\widetilde{\gamma}_S)\right) \qquad (135)\\
&= \binom{2n}{2l_1,2l_2,2l_3}_p^{-1/2}\frac{(-1)^{l_1+l_2+l_3}}{2^{n/2}}\binom{n}{l_1,l_2,l_3}_p\mathcal{B}_{l_1+l_3}, \qquad (136)
\end{aligned}$$

where $l_i = k_i/2$ for $i \in [3]$. By combining all the elements, we arrive at

$$\begin{aligned}
\mathrm{Var}(\hat{v}) \le \frac{(1+\varepsilon_c)^2}{2^{2n}} &\sum_{0\le l_1+l_2+l_3\le n} \frac{(-1)^{l_1+l_2+l_3}\binom{n}{l_1,l_2,l_3}_p\binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}\mathcal{B}_{l_1+l_3}}{\binom{2n}{2l_1,2l_2,2l_3}_p\binom{n}{l_1+l_2}\binom{n}{l_2+l_3}\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}} \\
&\times \sum_{\substack{S_1,S_2,S_3 \text{ disjoint}\\ |S_i|=2l_i, i\in[3]}} \mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0^\dagger\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right). \qquad (137)
\end{aligned}$$

□

It's worth noting that in the absence of noise, the variance bound is equivalent to the variance described in Ref. [10], where $\mathcal{B}_k = 1$ for any k. Using Chebyshev's inequality, we can ensure that the error ε of $\hat{v}$ is limited to $\frac{\mathrm{Var}(\hat{v})}{\varepsilon^2}$ with a high degree of certainty. To assess the effectiveness of this bound, we provide a specific bound on the variance for estimating various physical quantities:

(0) Calculate $\mathrm{Tr}\left(\rho\widetilde{\gamma}_S\right)$, where $|S|=2k$. If we choose the observable $H=\widetilde{\gamma}_S$, then the variance can be bounded by

$$\frac{(1+\varepsilon_c)^2\binom{2n}{2k}}{\mathcal{B}_k^2\binom{n}{k}}=\mathcal{O}\left(\mathcal{B}_k^{-2}n^k\right). \tag{138}$$

This can be further simplified to $\mathcal{O}\left(n^k\right)$ for noises with constant average fidelity in Γ_{2k} subspace.

(1) Calculate $\mathrm{Tr}\left(\rho\rho_g\right)$, where ρ_g is a Gaussian state. By choosing the observable $H=\rho_g$, the variance is bounded to $\mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}}{\mathcal{B}_{\min}^2}\right)$, where $\mathcal{B}_{\max}=\max_k|\mathcal{B}_k|$ and $\mathcal{B}_{\min}=\min_k|\mathcal{B}_k|$. It can be further simplified to $\mathcal{O}\left(\sqrt{n}\right)$ if the average fidelities of noise $\mathcal{B}_k$ are constants for any $k\in\{0,\ldots,n\}$.

(2) Calculate $\langle\psi|\phi_\tau|\psi|\phi_\tau\rangle$, where $|\phi_\tau\rangle$ is a τ-fermionic Slater determinant. Let the initial input state of the quantum device be $\rho=\frac{(|\mathbf{0}\rangle+|\psi\rangle)(\langle\mathbf{0}|+\langle\psi|)}{2}$, and observable $H=|\phi_\tau\rangle\langle\mathbf{0}|$, the variance is bounded to $\mathcal{O}\left(\frac{\sqrt{n}\ln n\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right)$, which equals $\mathcal{O}\left(\frac{\sqrt{n}\ln n\ln(m/\delta_e)}{\varepsilon_e^2}\right)$ if the average fidelities of noise $\mathcal{B}_k$ are constants for any $k\in\{0,\ldots,n\}$.

We give the details for the simplification of variance in Supplementary Note 8. With this lemma and **MedianOfMeans** method, we can obtain Theorem 3.

Theorem 3. *Given a quantum state ρ and observables $\{H_i\}_{i=1}^m$. With the number of estimation samplings*

$$\begin{aligned}R_e=\frac{68(1+\varepsilon_c)^2\ln(2m/\delta_e)}{\varepsilon_e^2}2^{-2n}&\sum_{0\le l_1+l_2+l_3\le n}\frac{(-1)^{l_1+l_2+l_3}\binom{n}{l_1,l_2,l_3}_p\binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}\mathcal{B}_{l_1+l_3}}{\binom{2n}{2l_1,2l_2,2l_3}_p\binom{n}{l_1+l_2}\binom{n}{l_2+l_3}\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}}\\&\times\sum_{\substack{S_1,S_2,S_3\text{ disjoint}\\|S_i|=2l_i,i\in[3]}}Tr\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right)Tr\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0\right)Tr\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right),\end{aligned} \tag{139}$$

where $H_0=\max_i(H_i-Tr\left(H_i\right)\frac{\mathbb{I}}{2^n})$, by running Algorithm 1 in the main file, we can bound the estimation error $\left|\hat{v}_i-Tr\left(H_i\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right)\right|\le\varepsilon_e$ for any $i\in[m]$ with success probability $1-\delta_e$.

Proof. The theorem can be obtained from Lemma 9 with the number of copies for ρ being $R_e=K_eN_e$, where $K_e=2\ln\left(2m\delta_e^{-1}\right)$, and $N_e=\frac{34\mathrm{Var}(\hat{v})}{\varepsilon_e^2}$. With the bound on the variance in Lemma 10, the value of $|f_{2k}|$ provided in Lemma 6 and the union bound, we see that with the number of estimation samplings

$$R_e=N_eK_e \tag{140}$$

$$=\frac{68\ln(2m/\delta_e)\mathrm{Var}\left(\hat{v}\right)}{\varepsilon_e^2} \tag{141}$$

$$\begin{aligned}\le\frac{68(1+\varepsilon_c)^2\ln(2m/\delta_e)}{\varepsilon_e^2 2^{2n}}&\sum_{0\le l_1+l_2+l_3\le n}\frac{(-1)^{l_1+l_2+l_3}\binom{n}{l_1,l_2,l_3}_p\binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}\mathcal{B}_{l_1+l_2}}{\binom{2n}{2l_1,2l_2,2l_3}_p\binom{n}{l_1+l_2}\binom{n}{l_2+l_3}\mathcal{B}_{l_2+l_3}\mathcal{B}_{l_1+l_3}}\\&\times\sum_{\substack{S_1,S_2,S_3\text{ disjoint}\\|S_i|=2l_i,i\in[3]}}\mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}H_0\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right),\end{aligned} \tag{142}$$

where $\left|\hat{v}_i-\mathrm{Tr}\left(H_i\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right)\right|\le\varepsilon_e$ for any $i\in[m]$, with success probability $1-\delta_e$. □

From this theorem, we can give the estimation of the number of samplings to approximate k-RDM, the overlap of m number of Gaussian state and any quantum state, and m number of τ-fermionic Slater determinant. The number of estimation samplings equals (1) $R_e = \mathcal{O}\left(\frac{k\ln(n/\delta_e)}{\varepsilon_e^2\mathcal{B}_k^2}n^k\right)$ to estimate k-RDMs; (2) $R_e = \mathcal{O}\left(\frac{\sqrt{n}\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right)$ to estimate $\left\{\mathrm{Tr}\left(\rho\rho_{g_j}\right)\right\}_{j=1}^m$; and (3) $R_e = \mathcal{O}\left(\frac{\sqrt{n}\ln n\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right)$ to estimate $\left\{\left\langle\psi|\phi_{\tau_j}\right\rangle\left\langle\phi_{\tau_j}|\psi\right\rangle\right\}_{j=1}^m$. The number of samples in all of these applications is polynomial to the number of qubits n. We give more detailed calculations for the required number of samples in Supplementary Note 8.

SUPPLEMENTARY NOTE 7

As an extension of the calibration bound introduced in Ref. [11], the subsequent lemma illustrates the minimum calibration error resulting from the estimation of f_{2k}.

Lemma 11. *Consider $\widehat{\mathcal{M}}$ as the estimated channel for $\widetilde{\mathcal{M}}$, where $\widehat{\mathcal{M}} = \sum_{k=0}^n \hat{f}_{2k}\mathcal{P}_{2k}$. Let ρ represent a quantum state, and let the observable H belong to the subspace Γ_{even}, then*

$$\left|Tr\left(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right) - Tr\left(H\rho\right)\right| = \|H\|_\infty \max_k(\hat{f}_{2k}^{-1}f_{2k} - 1) \leq \varepsilon_c\|H\|_\infty \tag{143}$$

if $\max_k\left|\hat{f}_{2k} - f_{2k}\right| \leq \frac{|f_{2k}|\varepsilon_c}{1+\varepsilon_c}$.

Proof. For the quantum state in the even subspace Γ_{even}, we have

$$\begin{aligned}
\left|\mathrm{Tr}\left(H\widehat{\mathcal{M}}^{-1}\widetilde{\mathcal{M}}(\rho)\right) - \mathrm{Tr}\left(H\rho\right)\right| &= \left|\mathrm{Tr}\left(\rho\sum_k\left(\hat{f}_{2k}^{-1}f_{2k} - 1\right)\mathcal{P}_{2k}(H)\right)\right| && (144)\\
&\leq \left|\mathrm{Tr}\left(H\rho\right)\right|\max_k\left|\hat{f}_{2k}^{-1}f_{2k} - 1\right| && (145)\\
&\leq \|H\|_\infty\max_k\left|\hat{f}_{2k}^{-1}f_{2k} - 1\right| && (146)\\
&\leq \varepsilon_c\|H\|_\infty. && (147)
\end{aligned}$$

Here $\|\cdot\|_\infty$ denotes the spectral norm. Suppose $\left|\hat{f}_{2k} - f_{2k}\right| \leq \epsilon_{f_{2k}}$, then

$$\begin{aligned}
\left|\hat{f}_{2k}^{-1}f_{2k} - 1\right| &\leq \frac{\left|f_{2k} - \hat{f}_{2k}\right|}{f_{2k} - \left|f_{2k} - \hat{f}_{2k}\right|} && (148)\\
&= \frac{\epsilon_{f_{2k}}}{|f_{2k}| - \epsilon_{f_{2k}}} && (149)\\
&\leq \varepsilon_c. && (150)
\end{aligned}$$

Hence, $\epsilon_{f_{2k}} \leq \frac{|f_{2k}|\varepsilon_c}{1+\varepsilon_c}$. □

With this lemma, we can further give the required number of samplings to bound the error resulting from the estimation of the noisy channel $\widehat{\mathcal{M}}$.

Theorem 4. *For any given unknown quantum state ρ and observables $\{H_i\}_{i=1}^m$, with the number of calibration samplings*

$$R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n\ln(1/\delta_c)}{\mathcal{B}_{\min}^2\varepsilon_c^2}\right), \tag{151}$$

where $\mathcal{B}_{\max} = \max_k |\mathcal{B}_k|$ *and* $\mathcal{B}_{\min} = \min_k |\mathcal{B}_k|$. *we can bound the error resulting in the estimation of channel* $\widetilde{\mathcal{M}}$ *to*

$$\left| Tr\left(H_i \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}}(\rho)\right) - Tr\left(H_i \rho\right) \right| \leq \varepsilon_c \left\| H_0 \right\|_\infty \tag{152}$$

for any $i \in [m]$ *with success probability* $1 - \delta_c$ *by running the calibration process in Algorithm 1 in the main file, where* H_0 *is the noiseless term of the observable* H_i.

By Lemma 11 and error bound of **MedianOfMeans**, we can get

$$R_c = \frac{68(1+\varepsilon_c)^2}{\varepsilon_c} \max_k \frac{\mathbb{E}\left[\hat{f}_{2k}^2\right]}{\left|f_{2k}\right|^2}. \tag{153}$$

By substituting the results of $\mathbb{E}\left[\hat{f}_{2k}^2\right], |f_{2k}|$ in Lemma 6 and Lemma 7 into Eq. (152), and some tedious calculations, we can obtain Theorem 4.

Proof of Theorem 4. By the **MedianOfMeans** method, we have

$$\Pr\left[\left|\hat{f}_k - f_k\right| \geq \epsilon_{f_k}\right] \leq \delta_c \tag{154}$$

with $N_c = 34 \max_k \mathrm{Var}\left(\hat{f}_k\right) / \epsilon_{f_k}^2$ and $K_c = 2\ln\left(2\delta_c^{-1}\right)$ for any error ϵ_{f_k} and failure probability δ_c. With Lemma 11, let $\epsilon_{f_k} = \frac{|f_{2k}|\varepsilon_c}{1+\varepsilon_c}$, then with $R_c = K_c N_c$ calibration samplings we have $|\mathrm{Tr}\left(H_i \widehat{\mathcal{M}}^{-1} \widetilde{\mathcal{M}}(\rho)\right) - \mathrm{Tr}\left(H_i \rho\right)| \leq \varepsilon_c \left\| H_0 \right\|_\infty$ for any $i \in [m]$ with success probability $1 - \delta_c$. Hence, we have

$$N_c = 34 \max_k \mathrm{Var}\left(\hat{f}_k\right) \frac{(1+\varepsilon_c)^2}{\left|f_{2k}\right|^2 \varepsilon_c^2} \tag{155}$$

$$\leq 34 \frac{(1+\varepsilon_c)^2}{\varepsilon_c^2} \max_k \frac{\mathbb{E}\left[\left|\hat{f}_{2k}\right|^2\right]}{\left|f_{2k}\right|^2} \tag{156}$$

$$\leq 34 \frac{(1+\varepsilon_c)^2}{\varepsilon_c^2} \max_k \sum_{0 \leq l \leq \min(k, n-k)} \mathcal{B}_{2l} \frac{\binom{n}{l,k-l,l}_p^2 \binom{2n}{2k}^2}{\binom{2n}{2l,2k-2l,2l}_p \binom{n}{k}^4 |\mathcal{B}_k|^2}. \tag{157}$$

Hence the total number of samplings equals $R_c = K_c N_c = 68 \frac{(1+\varepsilon_c)^2 \ln(2/\delta_c)}{\varepsilon_c^2} \max_k \sum_{0 \leq l \leq \min(k, n-k)} \frac{\mathcal{B}_{2l} \binom{n}{l,k-l,l}_p^2 \binom{2n}{2k}^2}{\binom{2n}{2l,2k-2l,2l}_p \binom{n}{k}^4 |\mathcal{B}_k|^2}$.

Since $k \in \{0, n\}$ will give us the trivial values of $R_c = 68 \frac{(1+\varepsilon_c)^2 \ln(2/\delta_c)}{\varepsilon_c^2} \frac{\mathcal{B}_0}{|\mathcal{B}_k|^2}$ where $k \in \{0, n\}$. In the following, we assume $k \in [n-1]$, and utilize the technique in concrete mathematics [93] to simplify this bound. By the Stirling formula, with the assumption that $n \neq 0$, we have

$$\frac{(2n)!}{(n!)^2} = \Theta\left(\frac{2^{2n}}{\sqrt{\pi n}}\right). \tag{158}$$

Hence as long as $k \notin \{0, n\}$, we have

$$\frac{\binom{2n}{2k}}{\binom{n}{k}^2} = \frac{(2n)!}{(n!)^2} \cdot \frac{(n!)^2}{(2k)!} \cdot \frac{(n-k)!^2}{(2n-2k)!} \tag{159}$$

$$= \Theta\left(\sqrt{k\left(1 - \frac{k}{n}\right)}\right). \tag{160}$$

Similarly, if $l \notin \{0, k, n-k\}$, then we can derive

$$\frac{\binom{n}{l,k-l,l}_p^2}{\binom{2n}{2l,2k-2l,2l}_p} = \Theta\left(\sqrt{\frac{n}{l^2(l^2-nl+k(n-k))}}\right) \tag{161}$$

$$= \Theta\left(\frac{1}{l}\sqrt{\frac{1}{(k-\frac{k^2}{n})-(l-\frac{l^2}{n})}}\right). \tag{162}$$

Let

$$g(l,k) := \frac{\binom{2n}{2k}^2\binom{n}{l,k-l,l}_p^2}{\binom{n}{k}^4\binom{2n}{2l,2k-2l,2l}_p} \tag{163}$$

$$= \frac{c(k-\frac{k^2}{n})}{l\sqrt{(k-\frac{k^2}{n})-(l-\frac{l^2}{n})}} \tag{164}$$

$$= \frac{c\sqrt{n}(k-\frac{k^2}{n})}{l\sqrt{l^2-nl+k(n-k)}} \tag{165}$$

for some constant c, where $l \in [n]\backslash\{k, n-k\}$. Let $h(z) = \frac{c}{l\sqrt{n}}\sqrt{\frac{z^2}{l^2-ln-z}}$, then $h(z)$ is monotonically increasing for $z \in (0, l(n-l))$. Note that if we let z be $k(n-k)$ in the function h, then we get the function $g(l,k)$. Hence for $0 < k < \frac{n}{2}$, the function $g(l,k)$ is monotonically increasing with respect to k, while for $\frac{n}{2} < k < n-1$, the function is monotonically decreasing with respect to k. For simplification, here we assume k is even, the result also holds for odd k with similar calculations. Let $k = \frac{n}{2}$ we have

$$g\left(l, \frac{n}{2}\right) = \frac{c\sqrt{n}\cdot n/4}{l(\frac{n}{2}-l)} \tag{166}$$

$$= \frac{c\sqrt{n}}{2}\left(\frac{1}{l} + \frac{1}{\frac{n}{2}-l}\right). \tag{167}$$

Hence

$$\sum_{0<l<n/2} g(l, n/2) \le c\sqrt{n}\sum_{l\le n/2}\frac{1}{l} \tag{168}$$

$$< c\sqrt{n}\left(\ln(n/2) + \gamma + \frac{1}{n}\right) \tag{169}$$

$$< c'\sqrt{n}\ln n \tag{170}$$

where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant, and c' is a constant.

In contrast, when the input l is either 0, k, or $n-k$, it is straightforward to verify that $g(l,k) = \frac{\binom{2n}{2k}}{\binom{n}{k}^2}$ and it equals $\Theta(\sqrt{k\left(1-\frac{k}{n}\right)}) = \mathcal{O}\left(\sqrt{n}\right)$ if $k \notin \{0, n\}$. When k takes the values 0 or n, the function $g(l,k)$ equals 1. Therefore, in conclusion, we can obtain the upper bound of $\max_k \sum_l g(l,k) = \mathcal{O}\left(\sqrt{n}\ln n\right)$.

Hence the number of calibration samplings

$$R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n\ln(1/\delta_c)}{\mathcal{B}_{\min}^2\varepsilon_c^2}\right). \tag{171}$$

In the assumption that $\mathcal{B}_k$ are constants for any $k \in [n]$, it can be simplified to

$$R_c = \mathcal{O}\left(\frac{\ln(1/\delta_c)\sqrt{n}\ln n}{\varepsilon_c^2}\right). \tag{172}$$

□

Since for physical noise such as depolarizing noise, generalized amplitude damping noise, and X-rotation noise, $\mathcal{B}_k$ are constants for any $k \in [n]$, the number of calibration data equals $\mathcal{O}\left(\frac{\sqrt{n}\ln n\ln(1/\delta_c)}{\varepsilon_c^2}\right)$.

SUPPLEMENTARY NOTE 8

In this section, we will analyze the explicit values for the variances given by Lemma 10 in some specific instances. Note that $|\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right)| \leq |\mathrm{Tr}\left(\rho\right)|\,\|\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\|_2 \leq 1$. Hence for different observables, we have:

(0) Calculate $\mathrm{Tr}\left(\rho\widetilde{\gamma}_S\right)$ for an unknown quantum state ρ and observable $\widetilde{\gamma}_S$ where $|S| = 2k$, c is a constant. By Lemma 10, we have

$$\mathrm{Var}\left(\hat{v}\right) \leq \frac{(1+\varepsilon_c)^2}{2^{2n}} \sum_{0\leq l_1+l_2+l_3\leq n} [l_1=0]\,[l_2=k]\,[l_3=0]\,\frac{(-1)^{l_1+l_2+l_3}\binom{2n}{2l_2}\mathcal{B}_0}{\mathcal{B}_{l_2}^2\binom{n}{l_2}}\mathrm{Tr}\left(\widetilde{\gamma}_S^2\right)\mathrm{Tr}\left(\widetilde{\gamma}_S\widetilde{\gamma}_S^\dagger\right) \tag{173}$$

$$= \frac{(1+\varepsilon_c)^2\binom{2n}{2k}\mathcal{B}_k^{-2}}{\binom{n}{k}} \tag{174}$$

$$= \mathcal{O}(\mathcal{B}_k^{-2}n^k). \tag{175}$$

Combined with the results in Theorem 3 and Theorem 4, we can calculate all of k-RDMs with the number of estimation samplings

$$R_e = \mathcal{O}\left(\frac{k\ln(n/\delta_e)}{\mathcal{B}_k^2\varepsilon_e^2}n^k\right), \tag{176}$$

and the number of calibration samplings

$$R_c = \mathcal{O}\left(\frac{\mathcal{B}_{\max}\sqrt{n}\ln n\ln(1/\delta_c)}{\mathcal{B}_{\min}^2\varepsilon_c^2}\right), \tag{177}$$

the estimation error can be bounded to $\varepsilon_e + \varepsilon_c$.

(1) Calculate the overlap between an n-qubit Gaussian state with density matrices $\{\rho_{g_j}\}_{j=1}^m$ and any n-qubit quantum state with density matrix ρ, denoted as $\{\mathrm{Tr}\left(\rho\rho_{g_j}\right)\}_{j=1}^m$. In the following, we will prove that by choosing the observable $H = \rho_g$, the number of required estimation samplings equals

$$R_e = \mathcal{O}\left(\frac{\sqrt{n}\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right). \tag{178}$$

It can be further simplified to $R_e = \mathcal{O}\left(\frac{\sqrt{n}\ln(m/\delta_e)}{\varepsilon_e^2}\right)$ for the noise with constant average noise fidelities $\{\mathcal{B}_k\}$.

By the definition of Gaussian state ρ_g in Eq. (59), for any given sets $S_1, S_2 \in [2n]$ with cardinalities $|S_1| = 2l_1$, and $|S_2| = 2l_2$, we have

$$\mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}\rho_g\right) = \frac{1}{2^n}(-1)^{l_1+l_2}\mu_S[S \subseteq [n], D(S) = S_1 \cup S_2], \tag{179}$$

$$\leq \frac{1}{2^n}[S \subseteq [n], D(S) = S_1 \cup S_2] \tag{180}$$

where $\mu_S = \prod_{j\in S}\mu_j$. Inequality (180) holds since $\mu_S \in [-1, 1]$. Therefore,

$$\sum_{\substack{S_1,S_2,S_3 \text{ disjoint} \\ |S_i|=2l_i, i\in[3]}} \mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}\rho_g\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_2}\widetilde{\gamma}_{S_3}\rho_g{}^{\dagger}\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_3}\widetilde{\gamma}_{S_1}\rho\right) \leq 1. \tag{181}$$

When combined with Lemma 10, we can bound the variance as follows:

$$\mathrm{Var}\left(\hat{v}\right) \leq \sum_{0\leq l_1+l_2+l_3\leq n} \frac{\binom{n}{l_1,l_2,l_3}_p^2 \binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}|\mathcal{B}_{l_1+l_3}|}{2^{2n}\binom{2n}{2l_1,2l_2,2l_3}_p \binom{n}{l_1+l_2}\binom{n}{l_2+l_3}|\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}|} \tag{182}$$

$$\leq \max_{0\leq l_1+l_2+l_3\leq n} \frac{\binom{n}{l_1,l_2,l_3}_p \binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}|\mathcal{B}_{l_1+l_3}|}{\binom{2n}{2l_1,2l_2,2l_3}_p \binom{n}{l_1+l_2}\binom{n}{l_2+l_3}|\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}|} \sum_{0\leq l_1+l_2+l_3\leq n} \frac{\binom{n}{l_1,l_2,l_3}_p}{2^{2n}} \tag{183}$$

$$\leq c \max_{1\leq l_1+l_2+l_3\leq n-1} \sqrt{\frac{n}{l_1 l_2 l_3 (n-l_1-l_2-l_3)}} \cdot \sqrt{\frac{(l_1+l_2)(n-l_1-l_2)}{n}} \cdot \sqrt{\frac{(l_2+l_3)(n-l_2-l_3)}{n}} \frac{\mathcal{B}_{\max}}{\mathcal{B}_{\min}^2} \tag{184}$$

$$\leq c'\sqrt{n}\frac{\mathcal{B}_{\max}}{\mathcal{B}_{\min}^2}, \tag{185}$$

where $\mathcal{B}_{\max} = \max_k |\mathcal{B}_k|$ and $\mathcal{B}_{\min} = \min_k |\mathcal{B}_k|$, c, c' are constants. Eq. (184) holds since

$$\frac{\binom{2n}{2k}}{\binom{n}{k}^2} = \frac{(2n)!}{(n!)^2} \cdot \frac{(n!)^2}{(2k)!} \cdot \frac{(n-k)!^2}{(2n-2k)!} \tag{186}$$

$$= \Theta\left(\sqrt{k(1-\frac{k}{n})}\right), \tag{187}$$

if $k \notin \{0, n\}$ by the Stirling formula [93]. Eq. (185) holds by noting that the maximum value is obtained by choosing l_1, l_3 in the order of $\Theta(1)$, and $l_2 = \Theta(n)$. Hence the number of required estimation samplings

$$R_e = \mathcal{O}\left(\frac{\sqrt{n}\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right). \tag{188}$$

(2) Calculate the overlap between τ-fermionic Slater determinants $\left\{\left|\phi_{\tau_j}\right\rangle\right\}_{j=1}^m$ and any quantum pure state $|\psi\rangle$, denoted as $\left\{\left\langle\psi|\phi_{\tau_j}\right\rangle\left|\psi|\phi_{\tau_j}\right\rangle\right\}_{j=1}^m$. Let the initial input state of the quantum device be $\rho = \frac{(|\mathbf{0}\rangle+|\psi\rangle)(\langle\mathbf{0}|+\langle\psi|)}{2}$, and observable $H = |\phi_\tau\rangle\langle\mathbf{0}|$, the number of required estimation samplings

$$R_e = \mathcal{O}\left(\frac{\sqrt{n}\ln n\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right), \tag{189}$$

which equals $\mathcal{O}\left(\frac{\sqrt{n}\ln n\ln(m/\delta_e)}{\varepsilon_e^2}\right)$ if the average fidelities of noise $\mathcal{B}_k$ are constants for any $k \in \{0, \ldots, n\}$.

Let

$$L := \max_{0\le l_1+l_2+l_3\le n} \frac{\binom{n}{l_1,l_2,l_3}_p \binom{2n}{2l_1+2l_2}\binom{2n}{2l_2+2l_3}|\mathcal{B}_{l_1+l_3}|}{\binom{2n}{2l_1,2l_2,2l_3}_p \binom{n}{l_1+l_2}\binom{n}{l_2+l_3}|\mathcal{B}_{l_1+l_2}\mathcal{B}_{l_2+l_3}|} \le \frac{c\sqrt{n}\mathcal{B}_{\max}}{\mathcal{B}_{\min}^2}. \tag{190}$$

By Lemma 10 and the properties of Slater determinant, we have

$$\mathrm{Var}\,(\hat{v}) \le \frac{L}{2^{2n}} \sum_{0\le l_1+l_2+l_3\le n} \sum_{\substack{S_1,S_2,S_3 \text{ disjoint} \\ |S_i|=2l_i, i\in[3]}} \mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}\left|\phi_\tau\right\rangle\left\langle\mathbf{0}\right|\right)\mathrm{Tr}\left(\widetilde{\gamma}_{S_1}\widetilde{\gamma}_{S_2}\left|\mathbf{0}\right\rangle\left\langle\phi_\tau\right|\right) \tag{191}$$

$$\le \frac{L}{2^{2n}} \sum_{0\le l_1+l_2+l_3\le n} \sum_{j=0}^{\tau/2} \binom{\tau}{2j} \binom{n-\tau}{l_1-\tau/2+j, l_2-\tau/2+j, l_3-j, n-l_1-l_2-l_3-j} \tag{192}$$

$$= \frac{L}{2^{2n}} \sum_{0\le l_1+l_2+l_3\le n} \sum_{j=J_{\min}}^{J_{\max}} \binom{\tau}{2j} \binom{n-\tau}{l_1-\tau/2+j, l_2-\tau/2+j, l_3-j, n-l_1-l_2-l_3-j} \tag{193}$$

$$\le \frac{L}{2^{2n}} \sum_{j=0}^{\tau/2} \binom{\tau}{2j} \sum_{0\le l_1'+l_2'+l_3'\le n-\tau} \binom{n-\tau}{l_1', l_2', l_3'}_p \tag{194}$$

$$\le \frac{L}{2^{2n}} \sum_{j=0}^{\tau/2} 2^\tau 4^{n-\tau} \tag{195}$$

$$\le \frac{c\sqrt{n}\ln n\mathcal{B}_{\max}}{\mathcal{B}_{\min}^2} \tag{196}$$

where $J_{\min} = \max(\tau/2-l_1, \tau/2-l_2, l_3+\tau-n, l_1+l_2+l_3-n, 0)$, and $J_{\max} = \max(n-\tau/2-l_1, n-\tau/2-l_2, l_3, n-l_1-l_2-l_3, \tau/2)$ for some constant c, where $\mathcal{B}_{\max} = \max_k |\mathcal{B}_k|$ and $\mathcal{B}_{\min} = \min_k |\mathcal{B}_k|$. Eq. (195) holds since

$$\sum_{0\le j_1+j_2+j_3\le n-\tau} \binom{n-\tau}{j_1, j_2, j_3, n-\tau-j_1-j_2-j_3} = 4^{n-\tau} \tag{197}$$

for any $j_1, j_2, j_3 \in \{0, \ldots, n-\tau\}$, and $\sum_{j=0}^{\tau/2}\binom{\tau}{2j} \le 2^\tau$. Hence the number of required estimation samplings

$$R_e = \mathcal{O}\left(\frac{\sqrt{n}\ln n\mathcal{B}_{\max}\ln(m/\delta_e)}{\mathcal{B}_{\min}^2\varepsilon_e^2}\right). \tag{198}$$

SUPPLEMENTARY NOTE 9

In this section, we give the numerical results for the calculation of (1) the overlap between a Gaussian state and any quantum state, and (2) the inner product between a Slater determinant and any pure state.

(1) Mitigated estimation for the fidelities with fermionic Gaussian states Let the observable H be a 4-qubit Gaussian state $\rho_g = \prod_{k=1}^n \frac{(\mathbb{I}-i\mu_k\widetilde{\gamma}_{2k-1}\widetilde{\gamma}_{2k})}{2}$, where $\{\mu_k\}_{k=1}^n$ are uniformly randomly chosen

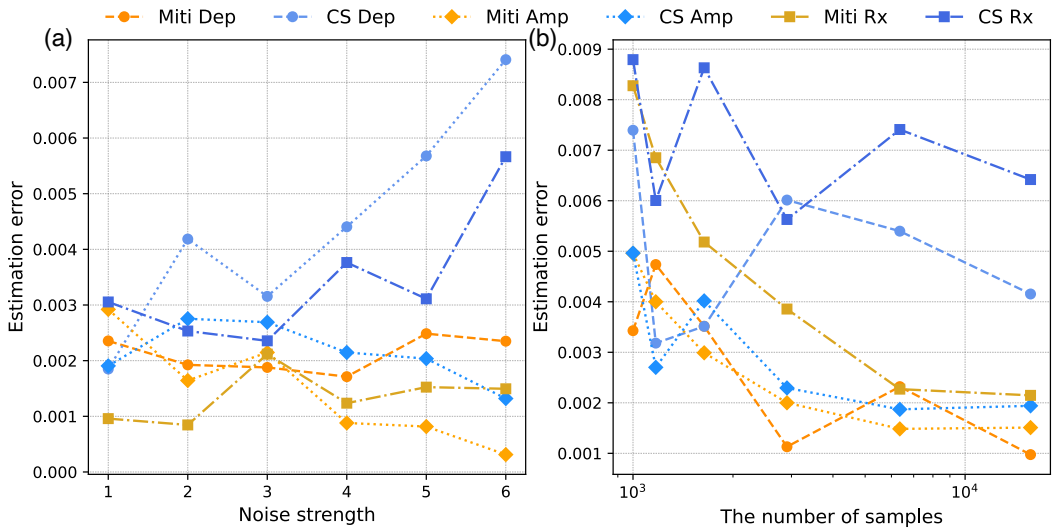

FIG. 7. The estimation errors for the fidelities for the Gaussian state with an unknown quantum state change as (a) the increase of the noise strength and (b) the number of samples increases for a fixed noise parameter. The error $\varepsilon = \sqrt{\sum_{i=1}^{R}(\hat{v}_i - \mathrm{Tr}\left(\rho\widetilde{\gamma}_S\right))^2/R}$ of the estimation is obtained by repeating the procedure $R = 4$ rounds for (a) and $R = 10$ for (b).

from $[0, 1]$. We use the same noise settings for the expectation value of $H = \rho_g$ and the same number of calibration samplings as in the main file. The numerical results are depicted in Fig. 7. The number of estimation samplings is set as $N = N_e K_e = 4000 \times 5$ for matchgate CS, and $N_e = 4000/(1 - p_{\text{noise}})$ for error-mitigated CS, as shown in Fig. 7 (a), where p_{noise} is the noise strength. Fig. 7 (b) depicts the change of the estimation errors with the number of estimation samplings N_e with the fixed noise parameters, where depolarizing noise parameter is as the fourth noise strength of Fig. 7 (a), and the sixth noise strength for both amplitude-damping and X-rotation noises.

(2) Mitigated estimation for overlaps with Slater determinants

To numerically estimate the overlap between a pure state $|\psi\rangle$ and a τ-Slater determinant $|\phi_\tau\rangle$, we uniformly randomly choose a normalized 3-qubit pure state $|\psi\rangle = U_\psi \left|0^3\right\rangle$ and a unitary $U \in \mathbb{C}^{3\times 3}$. τ is chosen as 1. The initial state is generated as $\rho = \frac{(|0^4\rangle + |1\rangle|\psi\rangle)(\langle 0^4| + \langle 1|\langle\psi|)}{2}$ with control-U_ψ operation to the state $\left|0^4\right\rangle$. The observable H is chosen as $|1\rangle|\phi_\tau\rangle\left\langle 0^4\right| = |1\rangle \tilde{b}_1^\dagger \left|0^3\right\rangle\left\langle 0^4\right|$, where $\tilde{b}_1 = \sum_j V_{1j} b_j$ and unitary V is randomly chosen. We depict the numerical results in Fig. 8. The noise settings are the same as in Fig. 7. The number of estimation samplings is set as $N = N_e K_e = 2000 \times 5$ for matchgate CS, and $N_e = 2000/(1 - p_{\text{noise}})$ for error-mitigated CS, as shown in Fig. 8 (a), where p_{noise} is the noise strength. The noise strength for Fig. 8 (b) are the fourth, sixth, and sixth for depolarizing, generalized amplitude damping, and X-rotation noises respectively.

SUPPLEMENTARY NOTE 10

Average noise fidelity is proposed in randomized benchmarking [48], denoted as

$$F_{\text{avg}} = \int_\psi \langle\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi|\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\,. \tag{199}$$

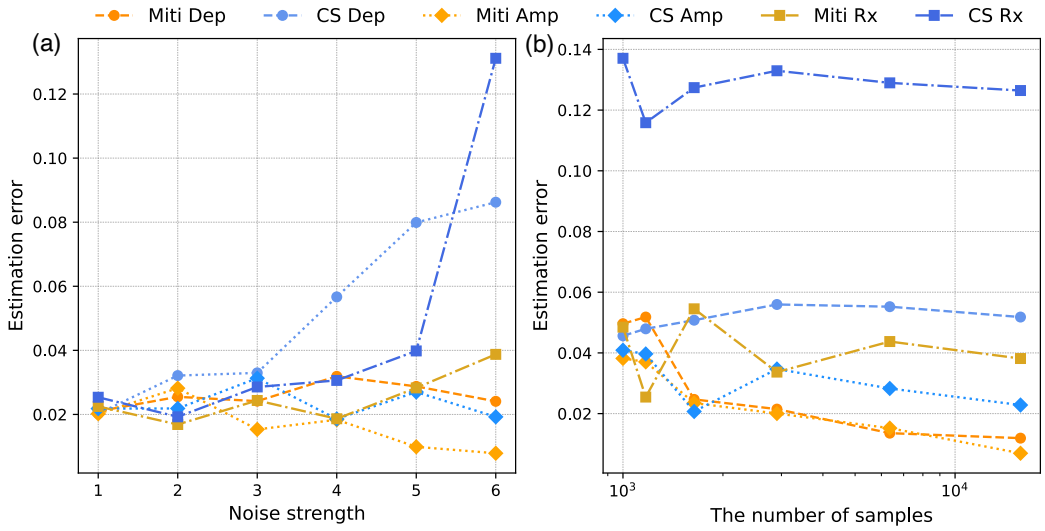

FIG. 8. The estimation errors for the inner product of the τ-Slater determinant and a pure state change as (a) the increase of the noise strength and (b) the number of samples increases for a fixed noise parameter. The error $\varepsilon = \sqrt{\sum_{i=1}^{R}(\hat{v}_i - \mathrm{Tr}\left(\rho\widetilde{\gamma}_S\right))^2/R}$ of the estimation is obtained by repeating the procedure $R = 4$ rounds.

It is easy to check that $F_{\text{avg}} = 1 - (1 - \frac{1}{2^n})p$ for depolarizing noise $\Lambda_{\text{d}}(A) = (1-p)(\cdot) + p\mathrm{Tr}\left(A\right)\frac{\mathbb{I}}{2^n}$. Since the expression of F_{avg} for Λ_{a} and Λ_{r} is more intricate, we solely provide the explicit representation for a single-qubit system here. It has been demonstrated that the set $e = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |-i\rangle\}$ constitutes a 2-design in the single-qubit state space [94]. Therefore, for a single qubit system,

$$F_{\text{avg}} = \frac{1}{6}\sum_{|\psi\rangle\in e}\langle\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi|\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\,. \tag{200}$$

This yields the values for F_{avg} corresponding to the respective noises in Table 2 of the main file. With the expression values in Table 2 of the main file, we see that these quantities are closely aligned, and $\mathcal{B}_1$ is slightly smaller than F_{avg} and F_Z for noise model $\Lambda_{\text{d}}, \Lambda_{\text{a}}$.

We plot the comparison of average fidelity F_{avg}, Z-basis average fidelity [11] F_Z and average fidelity in Γ_2 subspace $\mathcal{B}_1$ for X-rotation noise for θ varies in $[0, 2\pi]$ in Fig. 9. It shows that $\mathcal{B}_1$ is much more subtle to noise.

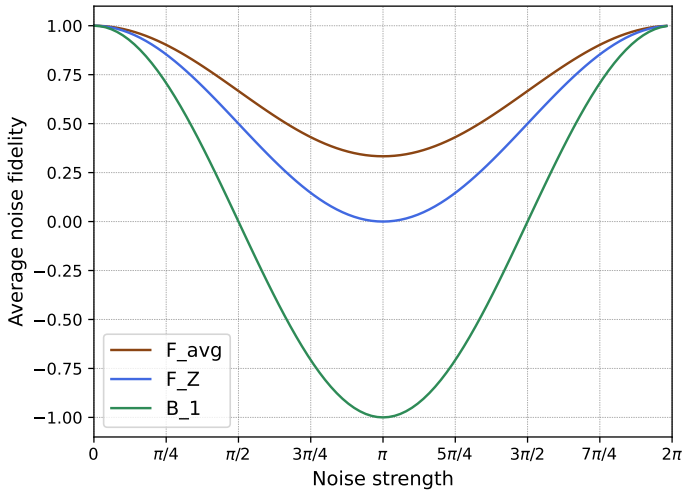

FIG. 9. The average fidelities F_{avg}, F_Z and $\mathcal{B}_1$ for X-rotation noise with noise range in $[0, 2\pi]$.

Appendix for "Adaptive-depth randomized measurement for fermionic observables"

Introduction to Pauli-transfer matrix representation

By employing the Jordan-Wigner transformation, we can express γ_S in terms of Pauli operators. The Pauli-transfer matrix (PTM) representation uses these Pauli operators $\{\sigma_i\}_{i=1}^{2^n}$ as a basis, where $\sigma_i = P_i/\sqrt{2^n}$ is the normalized Pauli operator, allowing us to denote the non-zero linear operators A as a 4^n-length vector $|A\rangle\rangle$ with the j-th entry being

$$A_j = \text{Tr}(A\sigma_j). \tag{201}$$

In this representation, a channel of operators could be expressed as a matrix Λ with the (a, b)-th element being

$$\Lambda(a, b) = \text{Tr}(\sigma_b \Lambda(\sigma_a)). \tag{202}$$

Using this expression, we can represent the second moment of the random matchgate as a fourth-order tensor, which is shown in Appendix .

Majorana operators could diagonalize shadow channel

This section shows the details about diagonalizing shadow channel $\mathcal{M}_d$. Notice that the computational basis $|b\rangle$ are Gaussian states

$$|b\rangle\langle b| = \prod_{j=1}^{n} \frac{1}{2}\left(I - \mathrm{i}(-1)^{b_j}\gamma_{2j-1}\gamma_{2j}\right). \tag{203}$$

Thus, any basis $|b\rangle\langle b|$ can be prepared from the state $|b\rangle\langle b|$ by a Gaussian unitary $U_Q \in \mathbb{M}_n$, which indicates that Pauli-X is in the matchgate group. The state $|b\rangle = |b_1 b_2 \cdots b_n\rangle$ could be denoted as $\prod X_i^{b_i} |0\rangle$. Since

Pauli-X is in the matchgate group, we can absorb the $\prod X_{b_i}$ into the matchgates U_{Q_d} in the expression of shadow channel

$$\mathcal{M}_d(\gamma_S) = \int \mathrm{d}U_{Q_d} \sum_{b\in\{0,1\}^n} \langle b| U_{Q_d}\gamma_S U_{Q_d}^\dagger |b\rangle U_{Q_d}^\dagger |b\rangle \langle b| U_{Q_d} \tag{204}$$

$$= 2^n \int \mathrm{d}U_{Q_d} \langle 0| U_{Q_d}\gamma_S U_{Q_d}^\dagger |0\rangle U_{Q_d}^\dagger |0\rangle \langle 0| U_{Q_d}. \tag{205}$$

If S' is not equal to S and the depth d is not equal to zero, then there exists a permutation matrix Q'_d in one depth matchgate circuit such that

$$[\gamma_S, U_{Q'_d}] = 0, \quad \{\gamma_{S'}, U_{Q'_d}\} = 0. \tag{206}$$

The orthogonal matrix Q'_d could be constructed in one depth because such U'_{Q_d} could be a Pauli operator, which involves one depth matchgate circuit. It implies

$$\frac{1}{2^n}\mathrm{Tr}\left(\gamma_{S'}\mathcal{M}_d(\gamma_S)\right) = \int \mathrm{d}U_{Q_d} \langle 0| U_{Q_d}\gamma_S U_{Q_d}^\dagger |0\rangle \langle 0| U_Q\gamma_{S'} U_{Q_d}^\dagger |0\rangle \tag{207}$$

$$= \int \mathrm{d}U_{Q_d} \langle 0| U_{Q_d}U_{Q'_d}\gamma_S U_{Q'_d}^\dagger U_{Q_d}^\dagger |0\rangle \langle 0| U_{Q_d}U_{Q'_d}\gamma_{S'} U_{Q'_d}^\dagger U_{Q_d}^\dagger |0\rangle \tag{208}$$

$$= -\int \mathrm{d}U_{Q_d} \langle 0| U_{Q_d}\gamma_S U_{Q_d}^\dagger |0\rangle \langle 0| U_{Q_d}\gamma_{S'} U_{Q_d}^\dagger |0\rangle . \tag{209}$$

The result shows that $\mathrm{Tr}\left(\gamma_{S'}\mathcal{M}_d(\gamma_S)\right) = 0$ when S' is not equal to S, thereby $\mathcal{M}_d(\gamma_S) = \alpha_{s,d}\gamma_S$.

Bound the variance of ADFCS estimator with $\alpha_{S,d}$

The variance of the random variable $v = \mathrm{Tr}\left(\hat{\rho}\gamma_S\right)$ is analyzed by deriving an upper bound on its value. First, the variance is bounded by the expected squared magnitude of v, i.e., $\mathrm{Var}(v) \leq \mathbb{E}[|v|^2]$. The expectation is then expressed as an integral over the unitary group U_{Q_d} and averaged over the state ρ

$$\mathbb{E}[|v|^2] = \int \mathrm{d}U_{Q_d}\, \mathbb{E}_\rho \left[\sum_b \langle b|U_{Q_d}\rho U_{Q_d}^\dagger|b\rangle \left|\langle b|U_{Q_d}\mathcal{M}_d^{-1}(\gamma_S)U_{Q_d}^\dagger|b\rangle\right|^2\right]. \tag{210}$$

Now, we start to simplify the expression. Firstly, we average out the ρ, which leads to

$$\mathbb{E}[|v|^2] = 2^{-n}\int \mathrm{d}U_{Q_d} \sum_b \left|\langle b|U_{Q_d}\mathcal{M}_d^{-1}(\gamma_S)U_{Q_d}^\dagger|b\rangle\right|^2. \tag{211}$$

Next, we denote $|b\rangle$ as $\prod_i X_i^{b_i}|0\rangle$ and absorb the Pauli X operators into the matchgate operators. And then uses Eq. (23)

$$\mathbb{E}[|v|^2] = \frac{1}{|\alpha_{S,d}|^2}\int \langle 0|U_{Q_d}\gamma_S U_{Q_d}^\dagger|0\rangle\langle 0|U_{Q_d}\gamma_S^\dagger U_{Q_d}^\dagger|0\rangle \tag{212}$$

$$= \frac{1}{\alpha_{S,d}}. \tag{213}$$

Finally we have $\mathrm{Var}\left([]\, v\right] \leq \frac{1}{\alpha_{S,d}}$. Notice that this inequality can be easily taken equally when $\mathrm{Tr}\left(\rho\gamma_S\right) = 0$.

Details of simplifying $\alpha_{S,d}$ to tensore network

This section shows the details of representing $\alpha_{S,d}$ by the tensor network contraction in PTM representation. According to Lemma 12, the $\alpha_{S,d}$ could be expressed by

$$\alpha_{S,d} = 2^{2n}\langle\langle \mathbf{0},\mathbf{0}| \int \mathrm{d}U_{Q_d}\mathcal{U}_{Q_d}^{\otimes 2}|P_S,P_S\rangle\rangle, \tag{214}$$

where P_S is the Pauli string corresponding to γ_S via Jordan-Wigner transformation.

Since each two-qubit random matchgate is independently sampled, the integral $\int \mathrm{d}U_Q\mathcal{U}_Q^{\otimes 2}$ could be calculated by independently integrating each 2-qubit matchgate. The result of the integral of the 2-qubit matchgates is given by Lemma 1,

$$\begin{aligned}\int_{U_Q\sim\mathbb{M}_2} \mathrm{d}U_Q\mathcal{U}_Q^{\otimes 2} =& |\gamma_\emptyset\rangle\rangle|\gamma_\emptyset\rangle\rangle\langle\langle\gamma_\emptyset|\langle\langle\gamma_\emptyset| + \frac{1}{4}\sum_{i,j}|\gamma_i\rangle\rangle|\gamma_i\rangle\rangle\langle\langle\gamma_j|\langle\langle\gamma_j| \\ &+ \frac{1}{6}\sum_{\substack{i_1\neq i_2\\ j_1\neq j_2}}|\gamma_{i_1}\gamma_{i_2}\rangle\rangle|\gamma_{i_1}\gamma_{i_2}\rangle\rangle\langle\langle\gamma_{j_1}\gamma_{j_2}|\langle\langle\gamma_{j_1}\gamma_{j_2}| \\ &+ \frac{1}{4}\sum_{\substack{i_1\neq i_2, j_1\neq j_2\\ i_1\neq i_3, j_1\neq j_3\\ i_2\neq i_3, j_2\neq j_3}}|\gamma_{i_1}\gamma_{i_2}\gamma_{i_3}\rangle\rangle|\gamma_{i_1}\gamma_{i_2}\gamma_{i_3}\rangle\rangle\langle\langle\gamma_{j_1}\gamma_{j_2}\gamma_{j_3}|\langle\langle\gamma_{j_1}\gamma_{j_2}\gamma_{j_3}| \\ &+ |\gamma_1\gamma_2\gamma_3\gamma_4\rangle\rangle|\gamma_1\gamma_2\gamma_3\gamma_4\rangle\rangle\langle\langle\gamma_1\gamma_2\gamma_3\gamma_4|\langle\langle\gamma_1\gamma_2\gamma_3\gamma_4|,\end{aligned} \tag{215}$$

where i, j are index ranged from 1 to 4. Following the definition Eq. (27), each element of tensor $\mathcal{T}$ could be calculated by Eq. (215). We show these concrete elements in Table 3. The tensor $\mathcal{T}$ presents the average effect of a random two-qubit matchgate.

Here, we represented $|\mathbf{0},\mathbf{0}\rangle\rangle$ in PTM to complete the calculation $\alpha_{S,d} = \langle\langle \mathbf{0},\mathbf{0}|\mathcal{C}|\gamma_S,\gamma_S\rangle\rangle$. Notice the matrix identity

$$|\mathbf{0}\rangle\langle\mathbf{0}| = \frac{1}{2^n}\sum_{\Lambda\subseteq[n]}\prod_{i\in\Lambda}Z_i, \tag{216}$$

where Z_i denotes the application of the Pauli Z operator to the i-th qubit. Especially, when $\Lambda = \emptyset$, let $\prod_{i\in\Lambda}Z_i = \mathbb{1}_n$. Then, the super vector of $|\mathbf{0},\mathbf{0}\rangle\rangle$ could be expressed as

$$|\mathbf{0},\mathbf{0}\rangle\rangle = \frac{1}{2^{2n}}\sum_{\Lambda,\Lambda'\subseteq[n]}|\prod_{i\in\Lambda}Z_i,\prod_{j\in\Lambda'}Z_j\rangle\rangle. \tag{217}$$

Eq. (217) express the supervector $|\mathbf{0},\mathbf{0}\rangle\rangle$ in the PTM representation. Finally, we write the $\mathcal{C}$, $|\mathbf{0},\mathbf{0}\rangle\rangle$, and $|\gamma_S,\gamma_S\rangle\rangle$ in Pauli basis. Thus, the $\alpha_{S,d}$ calculation could be expressed as the tensor network contraction, as illustrated in Fig. 4.

Lemma 12. *The quantity $\alpha_{S,d}$ satisfies the identity $\alpha_{S,d} = 2^{2n}\langle\langle \mathbf{0},\mathbf{0}|\int \mathrm{d}U_{Q_d}\mathcal{U}_{Q_d}^{\otimes 2}|P_S,P_S\rangle\rangle$, where P_S denotes the Pauli string associated with γ_S through the Jordan-Wigner transformation.*

Proof. Follow the definition of $\alpha_{S,d}$, we have

$$\alpha_{S,d} = \int_{Q\sim O_d}\mathrm{d}U_{Q_d}\left|\langle\mathbf{0}|U_{Q_d}\gamma_S U_{Q_d}^\dagger|\mathbf{0}\rangle\right|^2 \tag{218}$$

$$= \int_{Q\sim O_d}\mathrm{d}U_{Q_d}\langle\mathbf{0}|U_{Q_d}\gamma_S U_{Q_d}^\dagger|\mathbf{0}\rangle\langle\mathbf{0}|U_{Q_d}\gamma_S^\dagger U_{Q_d}^\dagger|\mathbf{0}\rangle. \tag{219}$$

$\mathcal{T}$	II	IX	IY	IZ	XI	XX	XY	XZ	YI	YX	YY	YZ	ZI	ZX	ZY	ZZ
II	1															
IX		1/4	1/4					1/4				1/4				
IY		1/4	1/4					1/4				1/4				
IZ				1/6		1/6	1/6			1/6	1/6		1/6			
XI					1/4				1/4					1/4	1/4	
XX				1/6		1/6	1/6			1/6	1/6		1/6			
XY				1/6		1/6	1/6			1/6	1/6		1/6			
XZ		1/4	1/4					1/4				1/4				
YI					1/4				1/4					1/4	1/4	
YX				1/6		1/6	1/6			1/6	1/6		1/6			
YY				1/6		1/6	1/6			1/6	1/6		1/6			
YZ		1/4	1/4					1/4				1/4				
ZI				1/6		1/6	1/6			1/6	1/6		1/6			
ZX					1/4				1/4					1/4	1/4	
ZY					1/4				1/4					1/4	1/4	
ZZ																1

TABLE 3. Values of tensor $\mathcal{T}$. The head of columns represents the input of $\mathcal{T}$ while the head of rows represents the output of $\mathcal{T}$. For example, the values in row 'XY' and column 'YZ' represent the value $\mathcal{T}^{XY}_{YZ}$. The blank space of the table stands for 0. For example, $\mathcal{T}^{XY}_{YZ}$ is equal to $\frac{1}{6}$.

The subscript d denotes the depth of the matchgate circuit. The expression could be simplified by substituting the relationship between $\gamma_S^\dagger$ and γ_S, which is

$$\gamma_S^\dagger = (-1)^{\frac{|S|(|S|-1)}{2}}\gamma_S, \tag{220}$$

due to the anti-commutation relation of Majorana operators $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. Hence $\alpha_{S,d}$ can be expressed as

$$\begin{aligned}
\alpha_{S,d} =&(-1)^{\frac{|S|(|S|-1)}{2}} \int \mathrm{d}U_{Q_d} \langle \mathbf{0}| U_{Q_d}\gamma_S U_{Q_d}^\dagger |\mathbf{0}\rangle^2 && (221)\\
=&(-1)^{\frac{|S|(|S|-1)}{2}} \int \mathrm{d}U_{Q_d} \operatorname{tr}\Big(U_{Q_d}\gamma_S U_{Q_d}^\dagger |\mathbf{0}\rangle \langle \mathbf{0}|\Big)^2 && (222)\\
=&(-1)^{\frac{|S|(|S|-1)}{2}} 2^{2n} \int \mathrm{d}U_{Q_d} \langle\langle \mathbf{0},\mathbf{0}|\mathcal{U}_{Q_d}^{\otimes 2}|\gamma_S,\gamma_S\rangle\rangle && (223)\\
=&(-1)^{\frac{|S|(|S|-1)}{2}} 2^{2n} \langle\langle \mathbf{0},\mathbf{0}| \int \mathrm{d}U_{Q_d} \mathcal{U}_{Q_d}^{\otimes 2}|\gamma_S,\gamma_S\rangle\rangle. && (224)
\end{aligned}$$

In PTM representation, the γ_S corresponds to a Pauli basis with a phase $\pm \mathrm{i}^{\lfloor |S|/2 \rfloor}$. Thus, the super vector $|\gamma_S, \gamma_S\rangle\rangle$ could be represented as

$$|\gamma_S, \gamma_S\rangle\rangle = (-1)^{\left\lfloor \frac{|S|}{2} \right\rfloor} |P_S, P_S\rangle\rangle, \tag{225}$$

where P_S is the Pauli operator that corresponding to the γ_S.

Follows Eq. (224) and Eq. (225), the sign of $\alpha_{S,d}$ is determined by $(-1)^{\frac{|S|(|S|-1)}{2}}(-1)^{\lfloor |S|/2 \rfloor}$. We show that the sign equals to 1 by categorizing the parity of $|S|$.

1. $|S|$ **is an odd number.** Let $|S| = 2q+1$, $q \in \mathbb{N}$, $q \geq 0$. And then

$$(-1)^{\frac{|S|(|S|-1)}{2}}(-1)^{\lfloor |S|/2 \rfloor} = (-1)^{q(2q+1)+q} = 1. \tag{226}$$

2. $|S|$ **is an even number.** Let $|S| = 2q$, $q \in \mathbb{N}$, $q \geq 0$. And then

$$(-1)^{\frac{|S|(|S|-1)}{2}}(-1)^{\lfloor |S|/2 \rfloor} = (-1)^{(2q-1)q+q} = 1. \tag{227}$$

Thus, we conclude that the $\alpha_{S,d}$ could be expressed by

$$\alpha_{S,d} = (-1)^{\frac{|S|(|S|-1)}{2}}(-1)^{\lfloor |S|/2 \rfloor} 2^{2n} \langle\langle \mathbf{0}, \mathbf{0}| \int \mathrm{d}U_{Q_d} \mathcal{U}_{Q_d}^{\otimes 2} |P_S, P_S\rangle\rangle \tag{228}$$

$$= \langle\langle \mathbf{0}, \mathbf{0}| \int \mathrm{d}U_{Q_d} \mathcal{U}_{Q_d}^{\otimes 2} |P_S, P_S\rangle\rangle. \tag{229}$$

□

Mapping the action of tensors to random walk

Here, we give some notations to express the properties of the $\alpha_{S,d}$ tensor network. Let $\mathcal{T}^{(o)}$ represent the odd layer in $\mathcal{C}$, and $\mathcal{T}^{(e)}$ represent the even layer in $\mathcal{C}$, as shown in Fig. 10. Any tensor network $\mathcal{C}$ with brickwork architecture could be expressed by alternately apply $\mathcal{T}^{(o)}$ and $\mathcal{T}^{(e)}$ gates,

$$\mathcal{C}\,(t, b_1, b_2) = \mathcal{T}^{(o)^{b_2}} \mathcal{B}^t \mathcal{T}^{(e)^{b_1}}, \tag{230}$$

where $b_1, b_2 \in \{0, 1\}$, $t + b_1 + b_2$ stands for the depth, and $\mathcal{B} = \mathcal{T}^{(e)}\mathcal{T}^{(o)}$. Let Γ_n' denote the vectorized double supervector space of Γ_n, defined as $\Gamma_n' = \mathrm{span}\{|\gamma_S, \gamma_S\rangle\rangle \mid \gamma_S \in \Gamma_n\}$. Notably, both $\mathcal{T}^{(o)}$ and $\mathcal{T}^{(e)}$ gates are projection operators, satisfying $(\mathcal{T}^{(e)})^2 = \mathcal{T}^{(e)}$ and $(\mathcal{T}^{(o)})^2 = \mathcal{T}^{(o)}$.

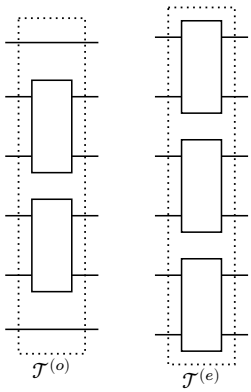

FIG. 10. Illustration for the odd-layer tensor $\mathcal{T}^{(o)}$ and the even-layer tensor $\mathcal{T}^{(e)}$.

In the main text, we frequently use the word "representation" but avoid the rigorous description of it. Here, we use the language of representation theory to give a description of representing $\mathcal{C}$ in mathematical taste. As we mentioned in the main text, the whole tensor $\mathcal{C}$ could be expressed by

$$\mathcal{C} = \begin{cases} \mathcal{B}^t, & d = 2t \\ \mathcal{B}^t \mathcal{T}_{\text{init}}, & d = 2t + 1. \end{cases} \tag{231}$$

For odd depth cases, studying the contraction of $\mathcal{C}$ on Γ_n' is equivalent to analyzing the action of the group $\{\mathcal{B}^t\}$ on the representation space $\mathcal{T}_{\text{init}}(\Gamma_n')$. The situation is similar for the even depth cases when $t \geq$

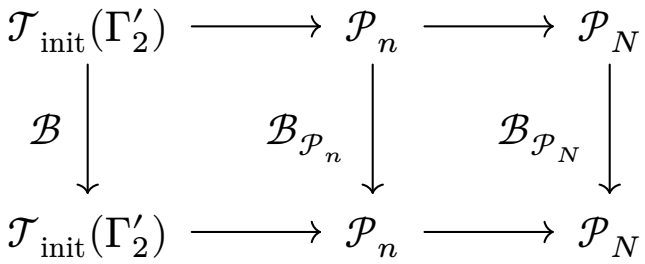

FIG. 11. Illustration for the isometric mapping to the polynomial space. Γ_2' represents the vectorized double super-vector space. Horizontal arrows indicate transitions between spaces, while vertical arrows correspond to the operators $\mathcal{B}$ in different representations. Ultimately, the process is reduced to the N-elementary polynomial within the second-degree polynomial space.

1. The difference between the two cases is the initial state, which is $\mathcal{T}^{(e)}|\gamma_S,\gamma_S\rangle\!\rangle$ for even depth case, and $\mathcal{T}^{(e)}\mathcal{T}^{(o)}|\gamma_S,\gamma_S\rangle\!\rangle$ for odd depth case. The states $\mathcal{T}^{(e)}\mathcal{T}^{(o)}|\gamma_S,\gamma_S\rangle\!\rangle$ are still in the space $\mathcal{T}_{\text{init}}(\Gamma_n')$, thereby we can apply similar method on the even depth cases. Thus, whatever the depth is even or odd, the contraction of $\mathcal{C}$ on Γ_n' could be analyzed by studying the representation of group $\{\mathcal{B}^t\}$. We use the word "the representation of $\mathcal{C}$" for short.

To simplify the analysis, we analyze the random circuit with even qubit number and odd depth d. We will represent $\mathcal{C}$ for the γ_S with $|S|=2$ in polynomial space, as illustrated in Fig. 11. These conditions will be maintained throughout the remainder of this section to ensure consistency and clarity in the discussion.

Reduce the calculation to polynomial space

We represent $\mathcal{C}|\gamma_S,\gamma_S\rangle\!\rangle$ using polynomial space because it provides the essential properties we require, such as multiplication and addition, which will be important for further analysis.

According to Table 3, $\mathcal{T}_{\text{init}}(\Gamma_2')$ could be spanned by the operators with the following form

$$\begin{cases} |\psi_{ij}\rangle\!\rangle = & \frac{1}{4}|X_i\left(\prod_{k=i+1}^{j-1}Z_k\right)Y_j\rangle\!\rangle^{\otimes 2}+\frac{1}{4}|X_i\left(\prod_{k=i+1}^{j-1}Z_k\right)Y_j\rangle\!\rangle^{\otimes 2} \\ & +\frac{1}{4}|X_i\left(\prod_{k=i+1}^{j-1}Z_k\right)Y_j\rangle\!\rangle^{\otimes 2}+\frac{1}{4}|X_i\left(\prod_{k=i+1}^{j-1}Z_k\right)Y_j\rangle\!\rangle^{\otimes 2},\ i<j \\ |\psi_{ii}\rangle\!\rangle = & |Z_i,Z_i\rangle\!\rangle. \end{cases} \tag{232}$$

The isometric map $\phi:\mathcal{T}_{\text{init}}(\Gamma_2')\to\mathcal{P}_n$ could be constructed by $\phi(|\psi_{ij}\rangle\!\rangle):=x_ix_j$, $\phi(|\psi_{ii}\rangle\!\rangle):=x_i^2$. The linear map ϕ is a homomorphism with inverse, which means it is an isometric between $\mathcal{T}_{\text{init}}(\Gamma_2')$ and $\mathcal{P}_n$. The action of $\mathcal{C}$ in $\mathcal{P}_n$ is constructed by

$$\mathcal{C}_{\mathcal{P}_n}(\cdot):=\phi\circ\mathcal{C}\circ\phi^{-1}(\cdot). \tag{233}$$

Thus, the action of d-depth matchgate circuit on the $\gamma_{\{i,j\}}$ is equivalent to the action of $\mathcal{C}_{\mathcal{P}_n}$ on the term x_ix_j. Due to Lemma 13, the action could be further simplified by finding a sub-representation $\mathcal{P}_N$ of representation $\mathcal{P}_n$, where N equals to $\frac{n}{2}$. Recall that n is an even number so N is an integer.

Certain hidden patterns are revealed by reducing the representation to the polynomial space $\mathcal{P}_N$. Most of the main results are been proved in the $\mathcal{P}_N$.

Lemma 13. *The space $\mathcal{P}_N$ isometric to a sub-representation of $\mathcal{P}_n$.*

Proof. Define the map φ as

$$\begin{aligned}\varphi : \mathcal{P}_N &\to \mathcal{P}_n \\ y_i^2 &\mapsto x_{2i-1}^2 + 4x_{2i-1}x_{2i} + x_{2i}^2 \\ y_i y_j &\mapsto (x_{2i-1} + x_{2i})(x_{2j-1} + x_{2j}).\end{aligned} \tag{234}$$

We define the φ as a linear function so that the definition of mapping on a basis induces the mapping on any element of the space

$$\varphi(\sum \xi_{ij} y_i y_j) = \sum \xi_{ij} \varphi(y_i y_j). \tag{235}$$

Thus, the space $\varphi(\mathcal{P}_N)$ is a linear subspace of $\mathcal{P}_n$. Because φ is an injection, there is a map $\varphi' : \mathcal{P}_n \to \mathcal{P}_N$ such that $\varphi' \circ \varphi$ is identity. The action of $\mathcal{C}$ in $\mathcal{P}_N$ is constructed by

$$\mathcal{C}_{\mathcal{P}_N}(y_i y_j) := \varphi' \circ \mathcal{C}_{\mathcal{P}_n} \circ \varphi(y_i y_j). \tag{236}$$

Based on the definition of representation, the space $\mathcal{P}_N$ is a representation of $\mathcal{C}$. And Eq. (236) depicts how to act a tensor on $\mathcal{P}_N$.

The next step is to show that such a representation of $\mathcal{C}$ does not "lose information". Formally, we need to prove that the representation $\mathcal{P}_n$ could be reduced to a sub-representation isometrics to $\mathcal{P}_N$. Naturally, we will consider whether the subspace $\varphi(\mathcal{P}_N)$ will constitute a sub-representation. If it is true, we could reduce the representation to its faithful sub-representation $\varphi(\mathcal{P}_N) \subseteq P_n$, which isometric to the polynomial space $\varphi(\mathcal{P}_N) \simeq \mathcal{P}_N$.

By the definition of sub-representation, we need to prove

$$\mathcal{B}^t(\varphi(\mathcal{P}_N)) \subseteq \varphi(\mathcal{P}_N), \quad \forall t. \tag{237}$$

The proof of statement (237) will be carried out using mathematical induction.

When $t = 0$, $\varphi(\mathcal{P}_N) \subseteq \varphi(\mathcal{P}_N)$ is true. Suppose the statement is true for t^*, then we have

$$(\mathcal{T}^{(e)}_{\mathcal{P}_n}\mathcal{T}^{(o)}_{\mathcal{P}_n})^{t^*}(\varphi(y_i y_j)) = \sum \xi_{lm}\varphi(y_l y_m). \tag{238}$$

And then

$$\mathcal{B}^{t^*+1}_{\mathcal{P}_n}(\varphi(y_i y_j)) = \mathcal{T}^{(e)}_{\mathcal{P}_n}\mathcal{T}^{(o)}_{\mathcal{P}_n}(\sum \xi_{lm}\varphi(y_l y_m)) \tag{239}$$

$$= \sum \xi_{lm}\delta_{lm}\mathcal{T}^{(e)}_{\mathcal{P}_n}\mathcal{T}^{(o)}_{\mathcal{P}_n}(x_{2i-1}^2 + 4x_{2i-1}x_{2i} + x_{2i}^2) \tag{240}$$

$$+ \sum \xi_{lm}(1-\delta_{lm})\mathcal{T}^{(e)}_{\mathcal{P}_n}\mathcal{T}^{(o)}_{\mathcal{P}_n}\left((x_{2i-1} + x_{2i})(x_{2j-1} + x_{2j})\right), \tag{241}$$

where $\mathcal{B}_{\mathcal{P}_n}$, $\mathcal{T}^{(e)}_{\mathcal{P}_n}$ and $\mathcal{T}^{(o)}_{\mathcal{P}_n}$ are defined in a manner analogous to $\mathcal{C}_{\mathcal{P}_n}$. We will prove Eq. (240) is in the space $\varphi(\mathcal{P}_N)$, while the same result of Eq. (241) can be proved by straightforward calculation

$$\begin{aligned}&\mathcal{T}^{(e)}_{\mathcal{P}_n}\mathcal{T}^{(o)}_{\mathcal{P}_n}(x_{2i-1}^2 + 4x_{2i-1}x_{2i} + x_{2i}^2) \\ =&\mathcal{T}^{(e)}_{\mathcal{P}_n}(\frac{1}{6}x_{2i-2}^2 + \frac{2}{3}x_{2i-2}x_{2i-1} + x_{2i-2}x_{2i} + x_{2i-2}x_{2i+1} \\ &+ \frac{1}{6}x_{2i-1}^2 + x_{2i-1}x_{2i} + x_{2i-1}x_{2i+1} + \frac{1}{6}x_{2i}^2 + \frac{2}{3}x_{2i}x_{2i+1} + \frac{1}{6}x_{2i+1}^2) \\ =&\frac{1}{36}x_{2i-3}^2 + \frac{1}{9}x_{2i-3}x_{2i-2} + \frac{5}{12}x_{2i-3}x_{2i-1} + \frac{5}{12}x_{2i-3}x_{2i} + \frac{1}{4}x_{2i-3}x_{2i+1} + \frac{1}{4}x_{2i-3}x_{2i+2} \\ &+ \frac{1}{36}x_{2i-2}^2 + \frac{5}{12}x_{2i-2}x_{2i-1} + \frac{5}{12}x_{2i-2}x_{2i} + \frac{1}{4}x_{2i-2}x_{2i+1} + \frac{1}{4}x_{2i-2}x_{2i+2} \\ &+ \frac{2}{9}x_{2i-1}^2 + \frac{8}{9}x_{2i-1}x_{2i} + \frac{5}{12}x_{2i-1}x_{2i+1} + \frac{5}{12}x_{2i-1}x_{2i+2} \\ &+ \frac{2}{9}x_{2i}^2 + \frac{5}{12}x_{2i}x_{2i+1} + \frac{5}{12}x_{2i}x_{2i+2} + \frac{1}{36}x_{2i+1}^2 + \frac{1}{9}x_{2i+1}x_{2i+2} + \frac{1}{36}x_{2i+2}^2\end{aligned} \tag{242}$$

The result expression is in $\varphi(\mathcal{P}_N)$ because we could find a polynomial y in $\mathcal{P}_N$ such that $\varphi(y)$ equals the result expression,

$$y = \frac{1}{6}y_{i-1}^2 + \frac{5}{3}y_{i-1}y_i + y_{i-1}y_{i+1} + \frac{4}{3}y_i^2 + \frac{5}{3}y_{i-1}y_{i+1} + \frac{1}{6}y_{i+1}^2 \tag{243}$$

$$\varphi(y) = \mathcal{T}_{\mathcal{P}_n}^{(e)}\mathcal{T}_{\mathcal{P}_n}^{(o)}(x_{2i-1}^2 + 4x_{2i-1}x_{2i} + x_{2i}^2) \tag{244}$$

Thus, we get that

$$\mathcal{B}_{\mathcal{P}_n}(x_{2i-1}^2 + 4x_{2i-1}x_{2i} + x_{2i}^2) \subseteq \varphi(\mathcal{P}_N), \tag{245}$$

which complete the proof of statement 237.

We aim to show that $\varphi(\mathcal{P}_N)$ is a faithful sub-representation. Suppose, for the sake of contradiction, that $\varphi(\mathcal{P}_N)$ is not faithful. Then, there exists a non-zero element $y \in \mathcal{P}_N$ such that:

$$\varphi(y) \neq 0 \quad \text{and} \quad \mathcal{B}_{\mathcal{P}_n}^t(\varphi(y)) = 0 \quad \text{for some integer } t \geq 1.$$

Consider applying the operator $\mathcal{B}_{\mathcal{P}_n}^{t'}$ iteratively to both sides of the equation:

$$\mathcal{B}_{\mathcal{P}_n}^{t'}\left[\mathcal{B}_{\mathcal{P}_n}^{t}(\varphi(y))\right] = \mathcal{B}_{\mathcal{P}_n}^{t+t'}(\varphi(y)) = \mathcal{B}_{\mathcal{P}_n}^{t'}(0) = 0.$$

Taking the limit as $t' \to \infty$, we obtain:

$$\lim_{t'\to\infty} \mathcal{B}_{\mathcal{P}_n}^{t+t'}(\varphi(y)) = 0.$$

However, the infinite application of $\mathcal{B}_{\mathcal{P}_n}$ to $\varphi(y)$ yields a non-zero polynomial. This is a contradiction because the same expression cannot simultaneously be zero and non-zero. Therefore, our initial assumption that $\varphi(\mathcal{P}_N)$ is not faithful must be false. Hence, $\varphi(\mathcal{P}_N)$ is indeed a faithful sub-representation.

□

Mapping the spread of polynomials to random walk

We could calculate the output of $\mathcal{B}_{\mathcal{P}_N}(y_iy_j)$ by $\mathcal{B}_{\mathcal{P}_N}(y_iy_j) := \varphi' \circ \phi^{-1} \circ \mathcal{B} \circ \phi \circ \varphi(y_iy_j)$, and the results are shown in Table 4. The definition of function L is

$$L(y_i) = \begin{cases} \frac{3}{4}y_1 + \frac{1}{4}y_2, & i = 1 \\ \frac{1}{4}y_{i-1} + \frac{1}{2}y_i + \frac{1}{4}y_{i+1}, & 1 < i < N \\ \frac{3}{4}y_N + \frac{1}{4}y_{N-1}, & i = N. \end{cases} \tag{246}$$

Notice that for most cases, the subscripts i, j satisfy $1 < i < N-1$ and $i+1 < j \leq N$, then the outcomes of $\mathcal{B}_{\mathcal{P}_N}$ is

$$\mathcal{B}_{\mathcal{P}_N}(y_iy_j) = \left(\frac{1}{4}y_{i-1} + \frac{1}{2}y_i + \frac{1}{4}y_{i-1}\right)\left(\frac{1}{4}y_{j-1} + \frac{1}{2}y_j + \frac{1}{4}y_{j-1}\right). \tag{247}$$

In this case, the action of $\mathcal{B}_{\mathcal{P}_N}$ can be viewed as independently evolving y_i and y_j in a one-dimensional lattice (see Fig. 12)

$$\begin{aligned} y_i &\to \frac{1}{4}y_{i-1} + \frac{1}{2}y_i + \frac{1}{4}y_{i-1} \\ y_j &\to \frac{1}{4}y_{j-1} + \frac{1}{2}y_j + \frac{1}{4}y_{j-1}. \end{aligned} \tag{248}$$

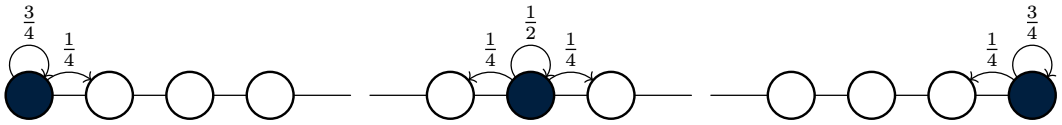

FIG. 12. Illustration of lazy symmetry random walk.

y_iy_j	$\mathcal{B}_{\mathcal{P}_N}(y_iy_j)$
$i=j=1$	$L(y_i)L(y_j)-\frac{5}{144}y_1y_1-\frac{5}{144}y_2y_2+\frac{5}{72}y_1y_2$
$1<i<N, j=i$	$L(y_i)L(y_j)-\frac{5}{144}y_{i-1}y_{i-1}+\frac{1}{36}y_{i-1}y_i+\frac{1}{24}y_{i-1}y_{i+1}-\frac{1}{36}y_iy_i+\frac{1}{36}y_iy_{i+1}-\frac{5}{144}y_{i+1}y_{i+1}$
$1\leq i<N, j=i+1$	$L(y_i)L(y_j)-\frac{1}{48}y_iy_i-\frac{1}{48}y_{i+1}y_{i+1}+\frac{1}{24}y_iy_{i+1}$
$i=j=N$	$L(y_i)L(y_j)-\frac{5}{144}y_Ny_N-\frac{5}{144}y_{N-1}y_{N-1}+\frac{5}{72}y_{N-1}y_N$
other case	$L(y_i)L(y_j)$

TABLE 4. The transition result of $\mathcal{B}_{\mathcal{P}_N}$ with input y_iy_j in different condition. Notice that $y_iy_j=y_jy_i$, the indices of the two factors i and j in term y_iy_j can always be arranged in ascending order $i\leq j$.

Here, we interpreted $y_1,\cdots y_n$ as a finite one-dimensional lattice whose i-th site is labeled by y_i. The action of $\mathcal{B}_{\mathcal{P}_N}$ could be interpreted as a random walk on 2D lattice whose (i,j) site is labeled by y_iy_j.

We observe that the equation $\mathcal{B}_{\mathcal{P}_N}(y_iy_j)=L(y_i)L(y_j)$ holds for most sites. This observation inspire us to firstly analyze the evolving behavior of L, and then using the probability distribution of L to estimate the order of $\alpha_{S,d}$.

We introduce the symmetry lazy random walk (SLRW) in polynomial space, or the one-dimensional lattice, to describe the evolution behavior of L. A SLRW is a type of Markov process shown in Fig. 12. In this process, consider a point located at a site y_i. In the next time interval, this point has a probability of 0.25 moving to one of its neighboring sites y_{i-1} or y_{i+1}, and it has a probability of 0.5 staying in place. If the origin site is on the ends of the lattice, it has a probability of 0.75 staying in place and has a probability of 0.25 moving around. The probability transition relation could be expressed by Eq. (246). We could see that the separate evolution in Eq. (248) fits the form of lazy-symmetry random walk. The transition step $\mathcal{L}_{\Gamma_2}$ in the representation space Γ_2' could be given by

$$\mathcal{L}_{\Gamma_2}|\gamma_S,\gamma_S\rangle\rangle=(L\times L)\circ\varphi'\circ\phi(|\gamma_S,\gamma_S\rangle\rangle), \tag{249}$$

where $(L\times L)(y_iy_j)=L(y_i)L(y_j)$.

In Eq. (247), we showed the result of how y_iy_j transfers in one specific situation. Now, we will show all possible results in any situation in Table 4. The table lists all the possible transition results no matter what inputs it receives. It was created in a similar way to the previous example in Eq. (242).

In Table 4, we can see that in most cases

$$\mathcal{B}_{\mathcal{P}_N}(y_iy_j)=L(y_i)L(y_j) \tag{250}$$

except for the cases when $|i-j|\leq 1$. Moreover, the coefficients of remainder terms $\mathcal{B}_{\mathcal{P}_N}(y_iy_j)-L(y_i)L(y_j)$ are small. Refs. [17] gives the analytical solution of lazy-symmetry random walk,

$$L^t(y_i)=\sum_\mu \mathcal{L}_i(\mu,t)y_\mu, \tag{251}$$

where $L^t(y_i)$ represents the outcome of random walking $\mathcal{T}$ steps from y_i according to the propagation rule L in Eq. (246), and the $\mathcal{L}_i(\mu,t)$ represents the probability of stopping at y_μ after $\mathcal{T}$-steps random walking,

$$\mathcal{L}_i(\mu,t)=\frac{1}{N}+\frac{2}{N}\sum_{k=1}^{N-1}\cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos\left(\left(i-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos^{2t}\left(\frac{\pi k}{2N}\right). \tag{252}$$

Thus, for the evolution that could be separated by $\mathcal{B}_{\mathcal{P}_N}(y_iy_j)=L(y_i)L(y_j)$, we could get the analytical solution results

$$\begin{aligned}\mathcal{C}_{\mathcal{P}_N}(t)(y_iy_j)=&(\mathcal{B}_{\mathcal{P}_N})^t(y_iy_j) &(253)\\ =&L^t(y_i)L^t(y_j)+R(y_iy_j) &(254)\\ =&\sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t)y_\mu y_\nu+\mathscr{R}_{ij}(\mu,\nu,t)y_\mu y_\nu\right), &(255)\end{aligned}$$

where

$$\mathscr{L}_{ij}(\mu,\nu,t):=\mathcal{L}_i(\mu,t)\mathcal{L}_j(\nu,t), \tag{256}$$

and R stands for the reminder terms caused by the near-diagonal terms y_iy_j, $|i-j|\leq 1$ in Table 4.

The calculation of $\alpha_{S,d}$ can be divided into two parts. The first part comes from the lazy-symmetry random walk, and the second part comes from the remainder terms. Recall that $\alpha_{S,d}$ is calculated by the tensor contraction of $\mathcal{C}$, $|P_S\rangle\!\rangle$, and $|\mathbf{0},\mathbf{0}\rangle\!\rangle$. We have mapped both $\mathcal{C}$ and $|P_S\rangle\!\rangle$ into polynomial space. Similarly, $|\mathbf{0},\mathbf{0}\rangle\!\rangle$ is also mapped into polynomial space.

Eq. (217) transforms $|\mathbf{0},\mathbf{0}\rangle\!\rangle$ into the Pauli basis. Notice that $\langle\!\langle Z_i|$ is a linear function that converts a super vector into a number. Specifically, it maps $|Z_i\rangle\!\rangle$ to 1 and maps other Pauli basis vectors to 0. As mentioned earlier, $|Z_i\rangle\!\rangle$ is mapped to x_i^2 in $\mathcal{P}_n$. The derivative operators' space is the polynomial space's dual space. Therefore, we use the space of derivative operators to represent the dual space of $\mathcal{P}_n$

$$\langle\!\langle\psi_{ij}|\to\frac{\partial^2}{\partial x_i\partial x_j},\quad \langle\!\langle\psi_{ii}|\to\frac{1}{2}\frac{\partial^2}{\partial x_i^2}. \tag{257}$$

With similar methods and some algebra, we map the $\langle\!\langle\mathbf{0},\mathbf{0}|$ to $\mathcal{P}_N$ when $|S|=2$

$$\langle\!\langle\mathbf{0},\mathbf{0}|\to\frac{1}{2^{2n}}\frac{1}{6}\sum_{i,j}\frac{\partial^2}{\partial y_i^2}. \tag{258}$$

Thus, the $\alpha_{S,2t+1}$ could be expressed as

$$\alpha_{\{\gamma_i\gamma_j\},2t+1}=\frac{1}{6}\sum_{\mu}\frac{\partial^2}{\partial y_\mu^2}\mathcal{C}_{\mathcal{P}_N}(t)(y_iy_j). \tag{259}$$

Then, we could calculate $\alpha_{S,d}$ by a random walk in $\mathcal{P}_N$ via combining Eq. (255) and Eq. (259)

$$\begin{aligned}\alpha_{\{\gamma_i\gamma_j\},2t+1}=&\frac{1}{3}\sum_{\mu}\mathscr{L}_{ij}(\mu,\mu,t)+\frac{1}{3}\sum_{\mu}\mathscr{R}_{ij}(\mu,\mu,t) &(260)\\ =:&\alpha^L_{\{\gamma_i\gamma_j\},2t+1}+\alpha^R_{\{\gamma_i\gamma_j\},2t+1}, &(261)\end{aligned}$$

where the term $\mathscr{L}_{ij}(\mu,\mu,t)$ is a transition probability.

Estimate the order of $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$

We have separated the calculation of $\alpha_{\{\gamma_i\gamma_j\},2t+1}$ into two parts in Sec. . In this section, we aim to estimate the order of the first part, $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$.

Lemma 14. *The $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ could be estimated by the following formula*

$$\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \frac{1}{3\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} \left(e^{-\frac{(2Nk+a)^2}{2t}} + e^{-\frac{(2Nk+b)^2}{2t}} \right) + \mathcal{O}\left(e^{-\frac{\pi^2}{2}t} \right) \tag{262}$$

where a is defined as $|i-j|$ and b is defined as $i+j-1$.

Proof. Recall that $N = n/2$. By Lemma 16, the expression for $\alpha_{S,2t+1}$ can be simplified to the form given in Eq. (273). Moreover, adding the term corresponding to $k = 0$ does not affect the summation series,

$$3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = -\frac{1}{N} + \frac{1}{N} \sum_{k=0}^{N-1} \left[\cos\left((i-j)\frac{k\pi}{N} \right) + \cos\left((i+j-1)\frac{k\pi}{N} \right) \right] \cos^{4t}\left(\frac{\pi k}{2N} \right). \tag{263}$$

Hence the summation over k spans a complete cycle, enabling us to leverage certain useful properties of trigonometric summations.

Our goal is to use the Poisson summation formula to estimate the order of $\alpha_{S,d}$. To facilitate this, it is more practical to express the formula using exponentials rather than trigonometric functions. Note that e^{-2tx^2} serves as a good approximation for $\cos^{4t}(x)$,

$$\begin{aligned} e^{-2tx^2} - \cos^{4t}(x) &= e^{-2tx^2} - e^{-2tx^2 + O\left(tx^4\right)} \\ &= e^{-2tx^2}\left(1 - e^{O\left(tx^4\right)} \right) \\ &\sim \mathcal{O}\left(tx^4 e^{-2tx^2} \right). \end{aligned} \tag{264}$$

By substituting $\cos^{4t}\left(\frac{\pi k}{2N}\right)$ with $e^{-\frac{k^2\pi^2 t}{2N^2}}$ in Eq. (281), we have

$$3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = -\frac{1}{N} + \frac{1}{N} \sum_{k=0}^{N-1} e^{-\frac{k^2\pi^2 t}{2N^2}} \left[\cos\left((i-j)\frac{k\pi}{N} \right) + \cos\left((i+j-1)\frac{k\pi}{N} \right) \right] + \mathcal{O}\left(e^{-\frac{\pi^2}{2}t} \right) \tag{265}$$

$$= -\frac{1}{N} + \frac{1}{N} \sum_{k=0}^{\infty} e^{-\frac{k^2\pi^2 t}{2N^2}} \left[\cos\left((i-j)\frac{k\pi}{N} \right) + \cos\left((i+j-1)\frac{k\pi}{N} \right) \right] + \mathcal{O}\left(e^{-\frac{\pi^2}{2}t} \right). \tag{266}$$

In Eq. (266), we expand the summation to infinity, and it will not introduce much of errors because

$$\begin{aligned} \sum_{k=N}^{\infty} e^{-\frac{k^2\pi^2 t}{2N^2}} =& e^{-\frac{\pi^2}{2}t} \sum_{k=0}^{\infty} e^{-\frac{k^2\pi^2 t}{2N^2}} \\ \leq& e^{-\frac{\pi^2}{2}t} \sum_{k=0}^{\infty} e^{-\frac{k\pi^2 t}{2N^2}} \\ =& e^{-\frac{\pi^2}{2}t} \frac{e^{\frac{\pi^2 t}{2N^2}}}{e^{\frac{\pi^2 t}{2N^2}} - 1} \\ =& \mathcal{O}\left(e^{-\frac{\pi^2}{2}t} \right) \end{aligned}$$

According to the Poison summation formula,

$$\sum_{k=-\infty}^{\infty} e^{-\frac{k^2\pi^2 t}{2N^2}} \cos\left((i-j)\frac{k\pi}{N}\right) = N\sqrt{\frac{2}{\pi t}} \sum_{k=-\infty}^{\infty} e^{-\frac{(2Nk+i-j)^2}{2t}} \tag{267}$$

By substituting Eq. (267) into Eq. (266), we have

$$\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \frac{1}{3\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} \left(e^{-\frac{(2Nk+i-j)^2}{2t}} + e^{-\frac{(2Nk+i+j-1)^2}{2t}}\right) + \mathcal{O}\left(e^{-\frac{\pi^2}{2}t}\right). \tag{268}$$

□

Lemma 15. *The $\alpha^L_{S,d}$ scales $\mathcal{O}\left(\frac{1}{\mathrm{poly}(n)}\right)$ if $d = \Theta\left(\max\left\{\frac{d_{int}(S)^2}{\log(n)}, d_{int}(S)\right\}\right)$.*

Proof. We start by proving that

$$\sum_{k=q+1}^{\infty} e^{-\frac{(2Nk+i-j)^2}{2t}} = \mathcal{O}\left(e^{-\frac{(2Nq+i-j)^2}{2t}}\right) \tag{269}$$

for positive integer q:

$$\frac{\sum_{k=q+1}^{\infty} e^{-\frac{(2Nk+i-j)^2}{2t}}}{e^{-\frac{(2Nq+i-j)^2}{2t}}} = \sum_{k=q+1}^{\infty} \exp\left[\frac{(2Nq+i-j)^2}{2t} - \frac{(2Nk+i-j)^2}{2t}\right] \tag{270}$$

$$= \sum_{k=q+1}^{\infty} \exp\left[-\frac{2N(k-q)(N(q+k)-(i-j))}{t}\right] \tag{271}$$

$$\leq \sum_{k=q+1}^{\infty} \exp\left[-\frac{2N(k-q)(2Nq-(i-j))}{t}\right], \tag{272}$$

where $\sum_{k=q+1}^{\infty} \exp\left[-\frac{2N(k-q)(2Nq-(i-j))}{t}\right]$ converges to certain constant. Thus, we prove Eq. (269).

Eq. (269) indicates that the order of $\alpha^L_{S,d}$ is determined by $\frac{1}{\sqrt{2\pi t}} e^{-\frac{(i-j)^2}{2t}} + \mathcal{O}\left(e^{-\frac{\pi^2}{2}t}\right)$. The condition $d = 2t+1 = \Omega(d_{\text{int}})$ ensures that $\mathcal{O}\left(e^{-\frac{\pi^2}{2}t}\right) = \mathcal{O}\left(\frac{1}{\mathrm{poly}(n)}\right)$, while the condition $d = \Omega(\frac{d_{\text{int}}(S)^2}{\log(n)})$ ensures that $e^{-\frac{(i-j)^2}{2t}} = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right)$ (notice that $d_{\text{int}}(S) = |i-j|$ for 2-local Majorana γ_S). Thus, we conclude that $\alpha^L_{S,d} = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right)$ when $d = \Theta\left(\max\left\{\frac{d_{\text{int}}(S)^2}{\log(n)}, d_{\text{int}}(S)\right\}\right)$.

□

Lemma 16. *The expression of $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ could be simplified in the following form*

$$3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \frac{1}{N} + \frac{1}{N}\sum_k \left[\cos\left((i-j)\frac{k\pi}{N}\right) + \cos\left((i+j-1)\frac{k\pi}{N}\right)\right] \cos^{4t}\left(\frac{\pi k}{2N}\right). \tag{273}$$

Proof. By the definition of $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and Eq. 256, we have

$$
\begin{aligned}
3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} =& \sum_\mu \left[\frac{1}{N} + \frac{2}{N}\sum_{k=1}^{N-1} \cos\left(\left(i-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos^{2t}\left(\frac{\pi k}{2N}\right)\right] \\
&\times \left[\frac{1}{N} + \frac{2}{N}\sum_{l=1}^{N-1} \cos\left(\left(j-\frac{1}{2}\right)\frac{\pi l}{N}\right)\cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi l}{N}\right)\cos^{2t}\left(\frac{\pi l}{2N}\right)\right] \\
=& 1/N + \frac{2}{N^2}\sum_k \left[\sum_\mu \cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi k}{N}\right)\right]\cos\left(\left(i-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos^{2t}\left(\frac{\pi k}{2N}\right) \qquad (274)\\
&+ \frac{2}{N^2}\sum_l \left[\sum_\mu \cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi l}{N}\right)\right]\cos\left(\left(i-\frac{1}{2}\right)\frac{\pi l}{N}\right)\cos^{2t}\left(\frac{\pi l}{2N}\right) \qquad (275)\\
&+ \frac{4}{N^2}\sum_{k,l=1}^{N-1}\sum_{\mu=1}^{N} \cos\left(\left(i-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos^{2t}\left(\frac{\pi k}{2N}\right) \\
&\times \cos\left(\left(j-\frac{1}{2}\right)\frac{\pi l}{N}\right)\cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi l}{N}\right)\cos^{2t}\left(\frac{\pi l}{2N}\right) \qquad (276)
\end{aligned}
$$

Notice that the summation of cosin function is zero

$$
\begin{aligned}
\sum_{\mu=1}^{N} \cos\left(\left(\mu-\frac{1}{2}\right)\frac{\pi k}{N}\right) &= -\frac{1}{2}\cos\left(\frac{1}{2}\pi(2k+1)\right)\csc\left(\frac{\pi k}{2N}\right) \\
&= \sin(k\pi)\csc\left(\frac{\pi k}{2N}\right) \\
&= 0.
\end{aligned} \qquad (277)
$$

Substitute this identity into Eq. (276), we could eliminate the terms in Eqs. (274) and (275).

Also, notice that

$$
\begin{aligned}
&\sum_{\mu=1}^{N} \cos\left(\frac{\pi\left(\mu-\frac{1}{2}\right)i}{N}\right)\cos\left(\frac{\pi\left(\mu-\frac{1}{2}\right)j}{N}\right) \\
=&\frac{1}{2}\sum_{\mu=1}^{N} \cos\left(\frac{\pi\left(\mu-\frac{1}{2}\right)i}{N} - \frac{\pi\left(\mu-\frac{1}{2}\right)j}{N}\right) + \cos\left(\frac{\pi\left(\mu-\frac{1}{2}\right)i}{N} + \frac{\pi\left(\mu-\frac{1}{2}\right)j}{N}\right) \\
=&\frac{1}{2}\sum_{\mu=1}^{N} \cos\left(\frac{\pi(2\mu-1)(i-j)}{2N}\right) + \cos\left(\frac{\pi(2\mu-1)(i+j)}{2N}\right) \\
=&\frac{1}{4}\left(\sin(\pi(i+j))\csc\left(\frac{\pi(i+j)}{2N}\right) - \sin(\pi(j-i))\csc\left(\frac{\pi(i-j)}{2N}\right)\right).
\end{aligned} \qquad (278)
$$

This result gets value 0 when $\frac{\pi(i+j)}{2N} \neq a\pi$ or $\frac{\pi(i-j)}{2N} \neq b\pi$ for some integer a and b, because $\sin(\pi m) = 0$. The term $\sin(\pi m)\csc\left(\frac{\pi m}{2N}\right)$ gets non-zero only when $\csc\left(\frac{\pi m}{2N}\right)$ gets infinity. Then, we could write down the conditions that i and j satisfy

$$
\begin{cases}
i+j = 2aN \text{ or } |i-j| = 2bN \\
a, b \in \mathbb{Z} \\
1 < i, j < N-1.
\end{cases} \qquad (279)
$$

The equation shows that the result is $j = k$. We can use L'Hôpital's rule to calculate the term

$$\lim_{x\to 0} \sin(\pi x)\csc\left(\frac{\pi x}{2N}\right) = 2N \tag{280}$$

when i and j satisfy the condition $i = j$. Plugin Eq. (280) and Eq. (278) into Eq. (276), we have

$$3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \frac{1}{N} + \frac{2}{N}\sum_k \cos\left(\left(i-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos\left(\left(j-\frac{1}{2}\right)\frac{\pi k}{N}\right)\cos^{4t}\left(\frac{\pi k}{2N}\right). \tag{281}$$

Finally, we use trigonometric identities to expand this equation, thereby completing this proof

$$3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \frac{1}{N} + \frac{1}{N}\sum_k \left[\cos\left((i-j)\frac{k\pi}{N}\right) + \cos\left((i+j-1)\frac{k\pi}{N}\right)\right]\cos^{4t}\left(\frac{\pi k}{2N}\right). \tag{282}$$

□

The relation between $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$

Recall that we have divided the calculation of $\alpha_{\{\gamma_i\gamma_j\},2t+1}$ into two parts. One is the $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and the other is $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$. Lemma 14 gives the order of $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$. In this section, we aim to bound the $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$ by $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$, so that the order of $\alpha_{\{\gamma_i\gamma_j\},2t+1}$ could be given by the $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$.

We begin with the polynomial in Eq. (255)

$$\mathcal{C}_{\mathcal{P}_N}(t)(y_iy_j) = \sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t)y_\mu y_\nu + \mathscr{R}_{ij}(\mu,\nu,t)y_\mu y_\nu\right). \tag{283}$$

Then, we let the polynomial transform one time-interval step, and we get

$$\begin{aligned}
&\mathcal{C}_{\mathcal{P}_N}(t+1)(y_iy_j) && (284)\\
=&\mathcal{B}_{\mathcal{P}_N}\left(\mathcal{C}_{\mathcal{P}_N}(t)(y_iy_j)\right) && (285)\\
=&\sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu)) + \mathscr{R}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu))\right) && (286)\\
=&\sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu)) + \mathscr{R}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu))\right) && (287)\\
=&\sum_{\mu,\nu}\mathscr{L}_{ij}(\mu,\nu,t+1)y_\mu y_\nu + \sum_{\mu,\nu}\mathscr{L}_{ij}(\mu,\nu,t)R(y_\mu,y_\nu)\\
&+\sum_{\mu,\nu}\mathscr{R}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu)). && (288)
\end{aligned}$$

Deduce from Eq. (283), we have

$$\mathcal{C}_{\mathcal{P}_N}(t+1)(y_iy_j) = \sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t+1)y_\mu y_\nu + \mathscr{R}_{ij}(\mu,\nu,t+1)y_\mu y_\nu\right). \tag{289}$$

Compare Eq. (288) and Eq. (289), we have

$$\sum_{\mu,\nu}\mathscr{R}_{ij}(\mu,\nu,t+1)y_\mu y_\nu = \sum_{\mu,\nu}\mathscr{L}_{ij}(\mu,\nu,t)R(y_\mu,y_\nu) + \sum_{\mu,\nu}\mathscr{R}_{ij}(\mu,\nu,t)(L(y_\mu)L(y_\nu)+R(y_\mu,y_\nu)) \tag{290}$$

$$(1+\delta_{l,k})\mathscr{R}_{ij}(l,k,t+1) = \sum_{\mu,\nu}\mathscr{R}_{ij}(\mu,\nu,t)\frac{\partial^2 L(y_\mu)L(y_\nu)}{\partial y_l\partial y_k} + \sum_{\mu,\nu}\left(\mathscr{L}_{ij}(\mu,\nu,t)+\mathscr{R}_{ij}(\mu,\nu,t)\right)\frac{\partial^2 R(y_\mu,y_\nu)}{\partial y_l\partial y_k}. \tag{291}$$

Eq. (291) describes the strict relationship between $\mathscr{R}_{ij}$ and $\mathscr{L}_{ij}$ in a recursive form, thereby giving the relationship between $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$. However, deriving the general term formula from this recursive formula is difficult. Therefore, we hope to use some inequalities to simplify this recursive relationship and thus bound $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$ by $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$.

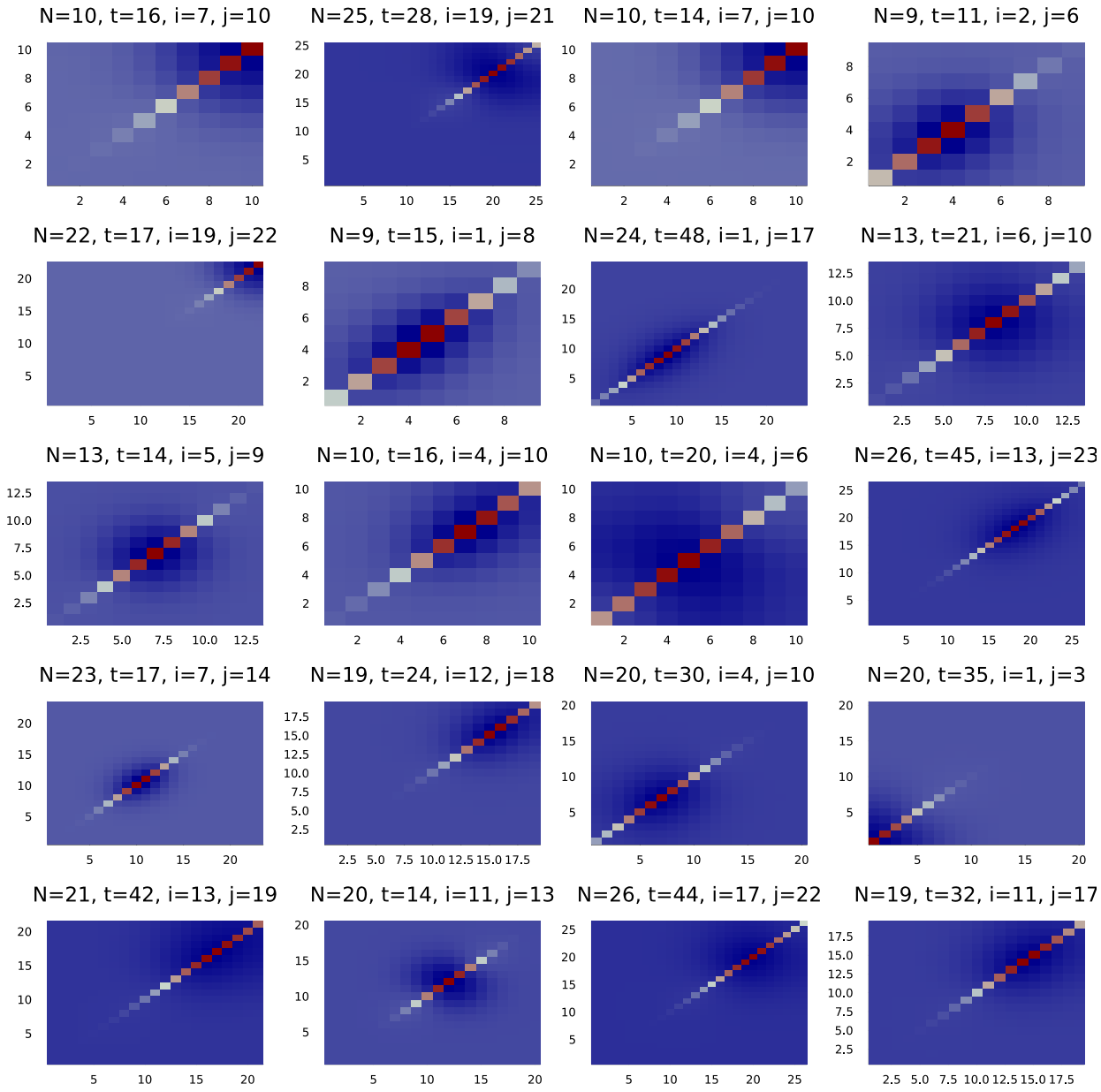

FIG. 13. The value of $\mathscr{R}_{i,j}(\mu,\nu,t)$. We randomly choose N,t,i,j, and calculate the value of $\mathscr{R}_{i,j}(\mu,\mu,t)$ numerically. The pixel is colored red if $\mathscr{R}_{i,j}(\mu,\nu,t)$ smaller than 0, else the pixel is colored blue. The x-axis represents the μ and the y-axis represents the ν. The numerical results show that $\mathscr{R}_{i,j}(\mu,\mu,t)<0$ when $\mu\neq\nu$.

We will first define some auxiliary variables,

$$\zeta_k(t) := \frac{1}{2}\sum_{\mu=1}^{N-k}\left(\mathcal{L}_i(\mu,t)\mathcal{L}_j(\mu+k,t)+\mathcal{L}_i(\mu+k,t)\mathcal{L}_j(\mu,t)\right)$$
$$\beta_k(t) := \frac{1}{2}\sum_{\mu=1}^{N-k}\left(\mathscr{R}_{i,j}(\mu,\mu+k,t)+\mathscr{R}_{i,j}(\mu+k,\mu,t)\right)$$
$$a(t) := \begin{pmatrix}\zeta_0(t)\\ \zeta_1(t)\end{pmatrix}, \quad b(t) := \begin{pmatrix}\beta_0(t)\\ \beta_1(t)\end{pmatrix}.$$

Notice that $3\alpha^L_{\{\gamma_i\gamma_j\},2t+1} = \zeta_0(t)$, $3\alpha^R_{\{\gamma_i\gamma_j\},2t+1} = \beta_0(t)$. Because $\mathscr{L}_{i,j}(\mu,\nu,t)+\mathscr{R}_{i,j}(\mu,\nu,t)$ represents the probability of being in site y_iy_j during a random walk, it satisfies the property that the summation across all sites is 1. Meanwhile, the summation of all $\mathscr{L}_{i,j}(\mu,\nu,t)$ is 1. The two things deduce that

$$\sum_{\mu,\nu}\mathscr{R}_{i,j}(\mu,\nu,t) = 0. \tag{292}$$

Especially, in numerical simulation, we observe that all $\mathscr{R}_{i,j}(\mu,\nu,t)$ are greater than 0 except for $\mu=\nu$. The numerical results are shown in Fig. 13. Under this assumption, we have the following lemma:

Lemma 17. *Assume that* $\forall\mu\neq\nu$, $\mathscr{R}_{i,j}(\mu,\nu,t)\geq 0$, *and* $\forall\mu\neq\nu$, $\mathscr{R}_{i,j}(\mu,\mu,t)\leq 0$. $-\beta_0(t)\leq \frac{25}{72}\max_{k\geq 0}\{\zeta_0(t-k)\}$.

Lemma 17 establishes the mathematical relationship between the two components $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$. By combining this theoretical relationship with the equality described in Eq. (261), we can deduce the order of magnitude of the $\alpha_{\{\gamma_i\gamma_j\},2t+1}$. Combining these pieces allows us to systematically determine the scale or size of $\alpha_{\{\gamma_i\gamma_j\},2t+1}$ based on the other defined quantities.

proof of Lemma 17. From the recursive relationship in Eq. (291) and Table 4, we could write down the recursive relationship of β_k

$$\begin{aligned}\beta_0(t+1) &\geq \frac{6}{16}\beta_0(t)+\frac{8}{16}\beta_1(t)+\frac{2}{16}\beta_2(t)-\frac{14}{144}\left(\beta_0(t)+\zeta_0(t)\right)-\frac{1}{24}\left(\beta_1(t)+\zeta_1(t)\right)\\ &+\frac{4}{16}(\mathscr{R}_{ij}(0,0,t-1)+\mathscr{R}_{ij}(N,N,t-1))-\frac{1}{16}(\mathscr{R}_{ij}(0,1,t-1)+\mathscr{R}_{ij}(1,0,t-1))\\ &-\frac{1}{16}(\mathscr{R}_{ij}(N-1,N,t-1)+\mathscr{R}_{ij}(N,N-1,t-1))\\ &\geq \frac{5}{18}\beta_0(t)+\frac{11}{24}\beta_1(t)-\frac{7}{72}\zeta_0(t)-\frac{1}{12}\zeta_1(t) \qquad (293)\\ &\geq \frac{5}{18}\beta_0(t)+\frac{5}{24}\beta_1(t)-\frac{7}{72}\zeta_0(t)-\frac{1}{12}\zeta_1(t) \qquad (294)\end{aligned}$$

Here, we use the property that $\beta_0(t)+\zeta_0(t)$ is the summation of properties so that it is greater than 0 to inequality deflate the terms at the edges, like y_0y_0 or y_0y_1. Recall that we hold the assumption that $\mathscr{R}_{i,j}(\mu,\mu,t)<0$, and $\forall\mu\neq\nu$, $\mathscr{R}_{i,j}(\mu,\mu,t)>0$, so the edges terms of $\mathscr{R}_{ij}$ could be deflated out as well. Similarly, we have

$$\beta_1(t+1)\geq\frac{5}{9}\beta_0(t)+\frac{5}{12}\beta_1(t)+\frac{1}{18}\zeta_0(t)+\frac{1}{24}\zeta_1(t). \tag{295}$$

Let β_0' and β_1' obtains the above recursive relation

$$\begin{cases}\beta_0'(t+1) = & \frac{5}{18}\beta_0(t)+\frac{5}{24}\beta_1(t)-\frac{7}{72}\zeta_0(t)-\frac{1}{12}\zeta_1(t)\\ \beta_1'(t+1) = & \frac{5}{9}\beta_0(t)+\frac{5}{12}\beta_1(t)+\frac{1}{18}\zeta_0(t)+\frac{1}{24}\zeta_1(t)\end{cases} \tag{296}$$

with the same first term $\beta_0'(0) = \beta_0(0)$ and $\beta_1'(0) = \beta_1(0)$. The β and β' satisfy the relationship

$$\beta_0'(t) \geq \beta_0(t), \quad \beta_1'(t) \geq \beta_1(t). \tag{297}$$

Similarly, we denote b' as $b'(t) := \binom{\beta_0'(t)}{\beta_1'(t)}$,

We rewrite the inequality groups (296) to the matrix form

$$b'(t+1) = C_b b'(t) + C_a a(t), \quad \text{where} \tag{298}$$

$$C_b = \begin{pmatrix} \frac{5}{18} & \frac{5}{24} \\ \frac{5}{9} & \frac{5}{12} \end{pmatrix}, \quad C_a = \begin{pmatrix} -\frac{7}{72} & -\frac{1}{12} \\ \frac{1}{18} & \frac{1}{24} \end{pmatrix}, \tag{299}$$

and the matrices C_b and C_a govern these recursive dynamics. To solve this recursion explicitly, we diagonalize the C_b matrix. This allows us to express the recursion in closed form,

$$b'(t) = C_b^t b'(0) + \sum_{k=0}^{t-1} C_b^k C_a a(t-k-1). \tag{300}$$

The term C_b^t could be calculated by eigenvalue decomposition $C_b = Q\Lambda Q^{-1}$, where $\Lambda = \text{diag}(\frac{25}{36}, 0)$. The eigenvector corresponding to $\frac{25}{36}$ is $(2, \frac{3}{2})^T$, where T denotes the transpose of vector. This allows us to express C_b^k and $C_b^k C_a$ in terms of eigenvalues and eigenvectors

$$C_b^k = \left(\frac{5}{6}\right)^{2k} \begin{pmatrix} 2 & \frac{3}{2} \\ 0 & 0 \end{pmatrix} \tag{301}$$

$$C_b^k C_a = \frac{1}{5}\left(\frac{5}{6}\right)^{2k} \begin{pmatrix} -\frac{1}{9} & -\frac{5}{48} \\ 0 & 0 \end{pmatrix} \tag{302}$$

for any $k > 0$.

The variables we care about are ζ_0 and β_0 because they are directly related to the $\alpha^L_{\{\gamma_i\gamma_j\},2t+1}$ and $\alpha^R_{\{\gamma_i\gamma_j\},2t+1}$. Thus, we mainly consider the first item of $C_b^t b'(0)$ and $C_b^k C_a a(t-k-1)$, which could be expressed in the following form

$$C_b^k C_a a(t-k-1) = \lambda_k \zeta_0(t-k-1) + \eta_k \zeta_1(t-k-1). \tag{303}$$

The λ_k and the η_k are coefficients

$$\begin{aligned} &\lambda_k = -\tfrac{1}{45}\left(\tfrac{5}{6}\right)^{2k}, \ \eta_k = -\tfrac{1}{48}\left(\tfrac{5}{6}\right)^{2k}, \quad \forall k > 0, \\ &\lambda_0 = -\tfrac{7}{72}, \qquad\quad \eta_0 = -\tfrac{1}{12}. \end{aligned} \tag{304}$$

Building on the previous relationships, we can now derive an explicit formula for $\beta_0'(t)$. From the expression for the first element of $C_b^k C_a a(t-k-1)$, we obtain

$$\beta_0'(t) = \sum_{k=0}^{t} \left(\lambda_k \zeta_0(t-k-1) + \eta_k \zeta_1(t-k-1)\right) \tag{305}$$

We also know from the recursive relation of $\mathcal{L}_i$ that the state variables satisfy

$$\zeta_0(t+1) \geq \frac{3}{8}\zeta_0(t) + \frac{1}{2}\zeta_1(t). \tag{306}$$

Leveraging this inequality into Eq. 305 allows us to place an upper bound on $\beta_0'(t)$

$$-\beta_0'(t) \leq -\sum_{k=1}^{t}\left(\lambda_k - \frac{3}{4}\eta_k + 2\eta_{k+1}\right)\zeta_0(t-k-1) + \frac{19}{72}\zeta_0(t) - \frac{1}{16}\zeta_0(t-1) \tag{307}$$

$$\leq \sum_{k=1}^{t}\left(\frac{5}{6}\right)^{2k}\frac{307}{8640}\zeta_0(t-k-1) + \frac{19}{72}\zeta_0(t) - \frac{1}{16}\zeta_0(t-1) \tag{308}$$

$$\leq \frac{25}{72}\max_{k\geq 0}\{\zeta_0(t-k)\} \tag{309}$$

□

Combine Lemma 14 and Lemma 17, we conclude the following theorem:

Theorem 5. *The sample complexity scales $\mathcal{O}(\frac{n}{\epsilon^2})$ up to a log factor for 2-local Majorana strings in the average of ρ when the depth satisfies*

$$d^* = \Theta\left(\max\left\{\frac{d_{int}(S)^2}{\log(n)}, d_{int}(S)\right\}\right) \tag{310}$$

under the following assumption: Compare to the SLRW, the true random walk in polynomial space $\mathcal{P}_N$ (The random walk of $\varphi' \circ \phi(\gamma_S)$ under transition $\mathcal{B}$) has less probability to walks on the diagonal sites $y_i y_j$ but has larger probability to walks on the off-diagonal sites at step t.

Proof. As we have discussed, the sample complexity is decided by $\alpha_{S,d}$, which could be expanded by

$$\alpha_{S,d} = \alpha_{S,d}^L + \alpha_{S,d}^R. \tag{311}$$

The assumption about the random walk means

$$\begin{cases} \mathscr{R}_{i,j}(\mu,\mu,t) \leq 0, \\ \mathscr{R}_{i,j}(\mu,\nu,t) \geq 0, \forall \mu \neq \nu. \end{cases} \tag{312}$$

Following Lemma 17, we have

$$\alpha_{S,d} = \alpha_{S,d}^L + \alpha_{S,d}^R \tag{313}$$

$$\geq \alpha_{S,d}^L - \frac{25}{72}\max_{d'<d}\{\alpha_{S,d'}^L\} \tag{314}$$

$$\geq \frac{47}{72}\max_{d'\leq d}\{\alpha_{S,d'}^L\}. \tag{315}$$

If $a = \mathcal{O}(\log n)$, the prove is complete by Lemma 18 with depth $d^* = \Theta(|i-j|)$. For $a = \omega(\log n)$, we have $\mathcal{O}(e^{-\pi^2 d}) = \mathcal{O}(n^{-\pi^2})$, and $\frac{1}{\sqrt{2\pi t}}e^{-\frac{(i-j)^2}{2t}} = \Omega(n)$ up to a log factor. In this case, we show that $\alpha_{S,d}^L = \Omega(n)$ with a similar method used in Lemma 15. Combine Chebyshev inequality and inequality (25), and we conclude that the sample complexity is $\mathcal{O}(\frac{n}{\epsilon^2})$. □

Efficiency when the distance of set is short

Lemma 18. *The expectation value of $\mathrm{tr}(\rho\gamma_S)$ can be obtained by using $\mathrm{polylog}(n)$-depth matchgate circuit within the Fermionic classical shadows protocol when the distance of S is $\mathcal{O}(\log n)$ and the cardinal number $|S| = 2k$ is a constant.*

Proof. Let the initial tensor be P_S, and apply $\mathcal{C}$ to the tensor P_S. Each $\mathcal{B}$ in $\mathcal{C}$ will transform the P_S to the superposition of a series of Pauli tensors

$$\mathcal{B}(P_S) = \sum_{|S'|=|S|} \xi_{S'} |P_{S'}\rangle\rangle, \tag{316}$$

where $\xi_{S'}$ are real coefficients that satisfy $\sum \xi_{S'} = 1$. Now, we only preserve the branches S' which has smaller distance $d_{\text{int}}(S') < d_{\text{int}}(S)$ unless $d_{\text{int}}(S) = 1$. The condition $d(S) = 1$ means that

$$P_S = \prod_{i\in\Lambda} Z_i \tag{317}$$

via Jordan-Wigner transformation. Only the Pauli basis in the form of $\prod_{i\in\Lambda} Z_i$ have a non-zero inner product with $|\mathbf{0}, \mathbf{0}\rangle\rangle$, which we explain in Sec. .

Table 3 tells us that the summation of the coefficients of remaining branches is greater than $\frac{1}{9^{|S|}}$. We apply $\mathcal{B}$ $d(S)/2$-times, and only the branches with smaller cardinal numbers remain in each step. Afterward, the summation of coefficients of the remaining branches is greater than $\frac{1}{9^{kd(S)}}$. Notice that $|S| = 2k$ is a constant number and $d(S) = \mathcal{O}(\log(n))$, the summation number is

$$\sum_{S''} \xi_{S''} \geq \frac{1}{9^{kd_{\text{int}}(S)}} = \frac{1}{\mathcal{O}(n^k \text{polylog}(n))} \tag{318}$$

where S'' are the remaining branches. □