

# Variational preparation of quantum many-body scars with shallow circuits

Eoin Carolan<sup>1,2</sup>, Nathan Keenan<sup>3</sup>, Gabriele Cenedese<sup>3</sup>, and Giuliano Benenti<sup>1,2</sup>

<sup>1</sup>*Center for Nonlinear and Complex Systems, Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, via Valleggio 11, 22100 Como, Italy*

<sup>2</sup>*Istituto Nazionale di Fisica Nucleare, Sezione di Milano, via Celoria 16, 20133 Milano, Italy*

<sup>3</sup>*Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), UIB-CSIC UIB Campus, E-07122 Palma de Mallorca, Spain*



(Received 8 March 2026; accepted 17 June 2026; published 20 July 2026)

We present and benchmark a type of variational quantum eigensolver (VQE) which we denote  $\sigma$ -VQE. It is designed to target midspectrum eigenstates and prepare quantum many-body scar states. The approach leverages the fact that noisy intermediate-scale quantum devices are limited in their ability to generate generic highly entangled states. This modified VQE pairs a low-depth circuit with an energy-selective objective that explicitly penalizes energy variance around a chosen target energy. The cost function exploits the limited expressibility of the shallow circuit as atypical low-entanglement eigenstates such as scar states are preferentially selected. We validate this mechanism across two complementary families of models that contain many-body scar states: the Shiraishi-Mori embedding approach and a matrix product state parent Hamiltonian construction. We define an unbiased estimation scheme for the nonlinear cost function that is compatible with qubit-wise commuting grouping and bitstring reuse. A proof-of-principle demonstration using a small-system instance was performed on IBM Fez in December 2025 (Heron r2 QPU). These results motivate its use as a practical algorithm for detecting quantum many-body scars and variationally generating states with appreciable scar state overlap.

DOI: [10.1103/518p-42fv](https://doi.org/10.1103/518p-42fv)

## I. INTRODUCTION

Variational methods for approximating eigenstates date back more than a century [1]. The same principles underpin a modern hybrid classical-quantum counterpart—the variational quantum eigensolver (VQE) [2]. The VQE has become a standard paradigm for noisy intermediate-scale quantum (NISQ) hardware [3,4], with applications ranging from quantum chemistry [5,6] to materials science [7,8] and quantum machine learning [9]. In an iteration of the VQE algorithm, a parametrized circuit prepares a trial state, a cost (or objective) function is estimated by measurements and classical postprocessing, and a classical optimizer updates the circuit parameters to ideally lower the cost function for the next iteration. It is one example of a hybrid quantum-classical algorithm [10,11]. While textbook VQEs focus on ground states by minimizing  $\langle H \rangle$  with respect to the circuit state, many research questions that probe excited-state structure, finite-temperature physics, and nonequilibrium initial-state engineering demand methods that can target interior eigenstates without requiring deep circuits or full diagonalization.

Quantum many-body scars (QMBS) are atypical eigenstates that violate the eigenstate thermalization hypothesis (ETH) in a spectrum that otherwise follows it [12–14]. The ETH predicts that individual eigenstates at fixed energy

density in generic interacting systems reproduce microcanonical thermodynamics for local observables. Scarred eigenstates are counterexamples [15,16]. They exhibit anomalously low entanglement compared to nearby states, which enables long-lived coherent dynamics and revivals from simple product-state initial conditions. Although towers of scar states and strong revivals are especially prominent in parts of the scar literature, they are not necessary for the terminology used here, which follows the standard usage that also encompasses exact embedded Shiraishi-Mori-type constructions [17]. These features make QMBSs relevant testbeds for exploring how nonergodic structure can persist without integrability or disorder. They have potential practical applications due to their ability to stabilize coherent dynamics at energies where typical states are highly entangled and rapidly thermalize. For recent reviews on QMBSs, see Refs. [18,19].

In this work we modify the standard VQE to target QMBSs by using an eigenstate-targeting objective that is naturally suited to midspectrum search [20]. We denote this protocol the  $\sigma$ -VQE. The central modification relative to the “textbook” implementation of the VQE is a nonlinear cost function that depends on the first and second moments of the Hamiltonian and explicitly penalizes energy variance at a chosen target energy  $E_{\text{tar}}$ . By construction, the objective is minimized by eigenstates with energy closest to  $E_{\text{tar}}$ , while coherent superpositions of eigenstates incur a variance penalty. This principle is closely related to a wider family of excited-state VQE strategies explored in other contexts [21–26], but here it is deployed alongside shallow quantum circuits specifically as a diagnostic and preparation primitive for QMBSs. The same cost function was introduced in Ref. [20] under the name

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

VQE-S and studied in ideal state-vector simulations. Here, we develop its finite-shot unbiased implementation, extend the benchmarking to two complementary scar constructions including a tunable entanglement parent Hamiltonian setting, and provide a proof-of-principle hardware demonstration on IBM Fez.

A second key design choice is to combine the  $\sigma$ -VQE cost function with a deliberately shallow circuit ansatz [27]. Shallow circuits are often treated as a limitation of NISQ devices, while here they become an asset. Generic midspectrum eigenstates typically exhibit volume-law entanglement, whereas scarred eigenstates atypically display low entanglement. A low-depth circuit strongly constrains the entanglement accessible to prepared states. In the benchmark settings studied here, that representational restriction disfavors generic highly entangled midspectrum eigenstates while leaving the anomalous low-entanglement target states accessible. We stress, however, that this is a statement about representational bias rather than a general claim that shallow circuits are always easier to optimize. The low-depth ansatz has the additional benefit of exposing the system to less noise than would be seen for deeper circuits.

A practical challenge shared by VQE-style algorithms is circuit overhead. Estimating Hamiltonian-based objective functions can involve measuring many Pauli strings and thus preparing the same circuit many times over. This motivates the use of estimators and measurement strategies that (a) exploit operator structure and grouping, (b) remain unbiased at finite shot count, and (c) make maximal use of each bitstring collected by the measurement of the system at the end of the circuit. In the  $\sigma$ -VQE implementation presented in this work we address these requirements by combining a qubit-wise commuting (QWC) measurement strategy with an estimator designed to reuse each sampled bitstring across many strings within a commuting group, together with an unbiased construction for the nonlinear terms. Further improvement of the algorithm can naturally be found in the broader landscape of measurement-frugal estimation schemes (including grouping and derandomization) that aim to reduce the number of circuit executions required for a fixed estimation accuracy [28–30]. We demonstrate the feasibility of  $\sigma$ -VQE through a proof-of-principle realization on superconducting hardware using an IBM Heron r2 processor. This processor has a heavy hex lattice topology and was accessed in December 2025.

The structure of the paper is as follows. In Sec. II, we introduce the  $\sigma$ -VQE algorithm. We specify the shallow entangling ansatz, the cost function, and the two classical optimizers we consider. We use a gradient-based optimizer [adaptive moment estimation (ADAM)] in the noiseless setting, where gradients (needed to update circuit parameters iteration to iteration) can be estimated using parameter-shift rules [31,32], and a gradient-free optimizer [simultaneous perturbation stochastic approximation (SPSA)] when contending with device noise and significant shot budget limitations. We then present our QWC-compatible measurement scheme and construct an unbiased finite-shot estimator for the nonlinear objective. In Sec. III we outline two families of benchmark Hamiltonians that contain scarred eigenstates. The first is the Shiraishi-Mori projector-embedding construction [17] to embed a product eigenstate at finite energy density into an

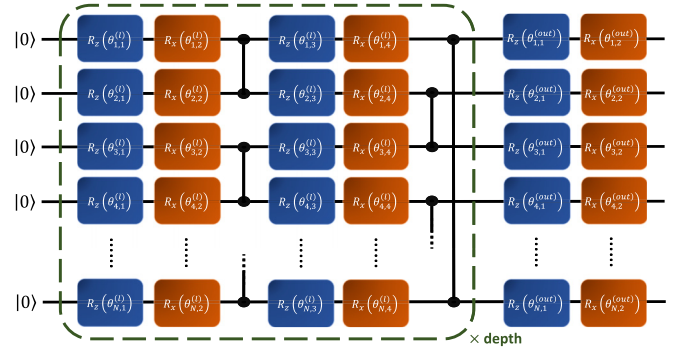


FIG. 1. Hardware-efficient ansatz used throughout this work.  $N$  is the number of qubits in the circuit, and the index ( $l$ ) runs from 1 to the total depth of the ansatz. Each layer applies single-qubit rotations on all qubits, followed by a pattern of two-qubit entangling gates (cz) on a periodic nearest-neighbor chain. The circuit depth denotes the number of such layers.

otherwise chaotic spectrum. The second uses a parent Hamiltonian embedding of random matrix product states (MPSs) [33], enabling controlled tuning of scar state entanglement. Finally, in Sec. IV we benchmark  $\sigma$ -VQE performance in idealized simulation and under realistic finite-shot and hardware noise, including a proof-of-principle execution on superconducting quantum hardware.

## II. $\sigma$ -VQE

The  $\sigma$ -VQE is a hybrid quantum-classical variational algorithm. In the ideal case a noiseless parametrized quantum circuit  $U(\theta)$  prepares the state  $|\psi(\theta)\rangle$ . Expectation values for observables relevant for a cost objective (typically the Hamiltonian) are estimated on that state, and a classical optimizer uses this information to update the circuit parameters  $\theta$  to find a parameter configuration that lowers the cost function. This process repeats until convergence to an optimal solution (within an acceptable tolerance) or resource limits are reached. The key differences with  $\sigma$ -VQE relative to textbook ground-state VQE concern the cost function and the way in which we estimate it. The cost is chosen to select an eigenstate of our Hamiltonian of interest with energy close to a specified target energy,  $E_{\text{tar}}$ . The estimator is built to remain unbiased under finite-shot sampling and to be compatible with QWC grouping. The optimizer can be either a gradient-based routine such as ADAM or a gradient-free stochastic routine such as SPSA. We summarize all these elements below.

### A. Circuit ansatz and cost function

The circuit ansatz is a key element of VQEs that must balance expressibility with efficiency for the learning task [34]. We choose a hardware-efficient ansatz built from  $d$  layers of single-qubit rotations followed by nearest-neighbor cz entangling gates, as shown in Fig. 1. The variational parameters  $\theta$  are the collection of rotation angles in the one-qubit gates, while the cz gates are fixed and parameter free. The circuit is initialized in the all-zero  $N$ -qubit computational basis state,  $|0\rangle^{\otimes N}$ . The hardware-efficient ansatz is typical in NISQ-based VQE chemistry and materials applications [35].

It is designed to be relatively agnostic to the specific Hamiltonian one wishes to implement, and instead be relatively easy to implement on a device using native operations. We intentionally restrict the circuit depth so that we preferentially target low-entanglement states. This expressive bias is central to using  $\sigma$ -VQE as a scar-hunting primitive, since generic midspectrum eigenstates are typically volume-law entangled while scars are not [20,36]. The cost function that defines the  $\sigma$ -VQE [20] is designed to penalize both deviation from a target energy and noneigenstates. Let

$$\langle H \rangle_\theta = \langle \psi(\theta) | H | \psi(\theta) \rangle, \quad \langle H^2 \rangle_\theta = \langle \psi(\theta) | H^2 | \psi(\theta) \rangle. \quad (1)$$

We then define the cost function as

$$C(\theta) = a \langle (H - E_{\text{tar}})^2 \rangle_\theta + b (\langle H^2 \rangle_\theta - \langle H \rangle_\theta^2), \quad (2)$$

where  $\theta$  denotes the vector of variational parameters,  $E_{\text{tar}}$  is the target energy from which deviations in the energy of the circuit state are penalized, and  $a$  and  $b$  are real positive weights where we take  $a + b = 1$  without loss of generality. The term weighted by  $a$  forces the mean energy toward the target value  $E_{\text{tar}}$ . The term weighted by  $b$  is the energy variance. If  $|\psi(\theta)\rangle$  is an exact eigenstate of  $H$  with eigenvalue  $\lambda$ , then  $\langle H \rangle_\theta^2 = \langle H^2 \rangle_\theta = \lambda^2$ , and

$$C(\theta) = a(\lambda - E_{\text{tar}})^2. \quad (3)$$

This highlights the importance of the squaring of the first term to minimize  $C(\theta)$  for eigenstates at  $\lambda = E_{\text{tar}}$ : any state that is a coherent superposition of different eigenstates incurs a positive variance penalty. We fix  $a + b = 1$  as only their relative weighting is meaningful, while an overall common rescaling would simply rescale the cost function. We therefore use  $a = b = 1/2$  as a simple symmetric benchmark choice rather than as a finely tuned optimum. Unless otherwise stated, we set  $E_{\text{tar}} = 0$ , which matches the embedded scar energy in our testbed models, which we define later in Sec. III.

### B. Gradient-based optimizer: ADAM

There is a choice over what classical optimization scheme is used in hybrid variational quantum algorithms [37,38]. Unless restricted by noise or resources, we use ADAM as it has been shown to be particularly robust and typically converges quickly. ADAM is a stochastic gradient method which utilizes adaptive first- and second-moment estimates of the gradient. For our purposes we estimate the gradient for each parameter using the parameter-shift rule. Let  $\theta_t$  be the parameter vector at iteration  $t$ . Let  $g_t = \nabla_\theta C(\theta_t)$  be the gradient of the cost function, Eq. (2). ADAM maintains two auxiliary vectors: a running exponential average of gradients and a running exponential average of squared gradients (or velocity),

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (4)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t \odot g_t, \quad (5)$$

where  $\odot$  denotes the Hadamard product. We take the hyperparameters  $\beta_1$  and  $\beta_2$  to be 0.9 and 0.999, respectively. Then, we correct for bias as both  $m_0$  and  $v_0$  are 0. Bias-corrected estimates are then

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (6)$$

The update rule for each component is

$$\theta_{t+1,k} = \theta_{t,k} - \alpha \frac{\hat{m}_{t,k}}{\sqrt{\hat{v}_{t,k} + \epsilon}}, \quad (7)$$

where  $\alpha$  is the learning rate and  $\epsilon$  is a numerical stabilization hyperparameter which we take to be  $10^{-8}$ . To update the parameters we need to estimate the gradient  $g_t$ . As the circuit consists of fixed two-qubit entangling gates and single-qubit rotations of the form  $\exp(-i\theta_k G_k/2)$ , where  $G_k$  is a Pauli operator (or tensor product thereof), we make use of the parameter-shift rule (PSR) to calculate the gradient. For an observable  $O$ , the PSR implies that the derivative  $\partial_{\theta_k} \langle O \rangle_\theta$  can be computed as

$$\partial_{\theta_k} \langle O \rangle_\theta = \frac{1}{2} [\langle O \rangle_{\theta + \frac{\pi}{2}} - \langle O \rangle_{\theta - \frac{\pi}{2}}], \quad (8)$$

where only parameter  $\theta_k$  is shifted by  $\pm\pi/2$ . As the cost function for  $\sigma$ -VQE is a linear combination of  $\langle H \rangle_\theta$ ,  $\langle H^2 \rangle_\theta$ , and  $\langle H \rangle_\theta^2$ , we cannot simply apply the PSR directly to the cost function as in the case of a ground-state search. Instead, we have

$$\partial_{\theta_k} C = (a + b) \partial_{\theta_k} \langle H^2 \rangle_\theta - (2aE_{\text{tar}} + 2b\langle H \rangle_\theta) \partial_{\theta_k} \langle H \rangle_\theta. \quad (9)$$

Both  $\partial_{\theta_k} \langle H^2 \rangle_\theta$  and  $\partial_{\theta_k} \langle H \rangle_\theta$  can be evaluated via (8). The full gradient can then be assembled in an unbiased way by using the PSR for each derivative, with an additional evaluation of  $\langle H \rangle_\theta$ .

This method of calculating the gradient is exact, but the calculation of the gradient for each optimization iteration requires  $2P + 1$  estimates of expectation values, where  $P$  denotes the number of parameters and we assume both  $\langle H \rangle$  and  $\langle H^2 \rangle$  can be estimated simultaneously from the same circuit measurement (more on which later).

### C. Gradient-free optimizer: SPSA

Expectation values—and therefore also the cost function—(2) are estimated from a finite number of shots (circuit preparations and executions) on quantum hardware. Using the PSR to calculate the gradient demands  $2P + 1$  expectation values to be estimated. We can avoid explicit PSR gradients by optimizing with a stochastic gradient approximation. This will prove useful in cases where our estimation of the cost function is particularly noisy, or when the number of shots is severely limited. The gradient-free method we will use is simultaneous perturbation stochastic approximation [39]. SPSA replaces the gradient by a simultaneous two-point finite difference along a random direction in parameter space.

We now outline the SPSA method. At iteration  $t$  of the optimization procedure we draw a random perturbation vector  $\Delta_t$ . Each component  $\Delta_{t,k}$  is independently and randomly drawn from  $\{-1, 1\}$ . SPSA then evaluates the cost function at two shifted parameter vectors,

$$\theta_t^{(+)} = \theta_t + c_t \Delta_t, \quad \theta_t^{(-)} = \theta_t - c_t \Delta_t, \quad (10)$$

where  $c_t = c_0(t + 1)^{-\gamma}$  is a decay hyperparameter. The finite-difference approximation of the gradient is given by

$$\hat{g}_{t,k} = \frac{C(\theta_t^{(+)}) - C(\theta_t^{(-)})}{2c_t \Delta_{t,k}}. \quad (11)$$

The advantage of SPSA is that this estimate requires only two cost evaluations per iteration (instead of  $2P + 1$ ), and therefore is independent of the number of parameters. We then update the parameters as

$$\theta_{t+1} = \theta_t - a_t \hat{g}_t, \quad (12)$$

where

$$a_t = \frac{a_0}{(A + t + 1)^\alpha}. \quad (13)$$

The algorithm wraps each updated angle back to the interval  $(-\pi, \pi]$ . SPSA is well suited to NISQ devices as it is known to tolerate device noise and demands a lower shot overhead [35,40]. We have a number of hyperparameters,  $\{a_0, c_0, A, \gamma, \alpha\}$ , and unlike ADAM, SPSA is quite sensitive to their choice. When utilizing SPSA we optimize the hyperparameters using the methods prescribed in Ref. [39].

#### D. Unbiased estimation of the cost function

We now describe how to estimate the expectation values  $\langle H \rangle_\theta$ ,  $\langle H^2 \rangle_\theta$ , and  $\langle H \rangle_\theta^2$  needed for both PSR and SPSA methods from finite shots in a way that is unbiased (with respect to shot noise) and allows for simultaneous use of circuit measurement outcomes for each value.

We work with  $N$  qubits. We denote  $\mathcal{P} = \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes N}$  as the set of Pauli strings. We expand the operators needed for the cost function in the Pauli string basis as

$$H = \alpha_{\mathbb{1}} \mathbb{1} + \sum_i \alpha_i P_i, \quad H^2 = \beta_{\mathbb{1}} \mathbb{1} + \sum_i \beta_i P_i, \quad (14)$$

with real coefficients  $\alpha_i, \beta_i$ . For a given circuit state  $|\psi(\theta)\rangle$ , the expectation values we need are

$$\langle H \rangle_\theta = \alpha_{\mathbb{1}} + \sum_i \alpha_i \langle P_i \rangle_\theta, \quad \langle H^2 \rangle_\theta = \beta_{\mathbb{1}} + \sum_i \beta_i \langle P_i \rangle_\theta. \quad (15)$$

Measuring  $\langle P_i \rangle_\theta$  for all strings separately is inefficient and wastes information that can be shared across shots. Instead, we exploit the QWC structure [41,42] and use a classical shadow-motivated approach to simultaneous estimation of expectation values [29]. A Pauli string  $P_i$  is compatible with a measurement basis  $B \in \{\sigma_x, \sigma_y, \sigma_z\}^{\otimes N}$  if on each qubit  $P_i$  either acts as the identity or  $P_i$  and  $B$  act with the same Pauli axis. We define the coverage indicator for this as

$$\delta_B(i) = \begin{cases} 1, & P_i \text{ covered by } B, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

Let  $\mathcal{B}$  be the set of all bases that cover at least one Pauli string  $P_i$  which has nonzero  $\alpha_i$  or  $\beta_i$ . To measure in a basis  $B$ , local prerotations  $U_B = \bigotimes_j U_{B_j}$  are applied to each qubit in the register after the circuit is executed. This maps measurement in the computational basis to a measurement in basis  $B$ . This mapping is done by  $U_{\sigma_x} = H$ ,  $U_{\sigma_y} = R_x(\pi/2)$ , and  $U_{\sigma_z} = I$ . The bitstring  $q \in \{0, 1\}^N$  is then drawn by measurements on each qubit in the computational basis. The distribution of outcome bitstrings is then

$$p_q \equiv \Pr(q | B, \theta) = |\langle q | U_B | \psi_\theta \rangle|^2. \quad (17)$$

For any  $i$  covered by  $B$ , the eigenvalue of  $P_i$  on the outcome  $q$  is

$$m_i(q) = \prod_{j \in \text{supp}(P_i)} (-1)^{q_j}, \quad (18)$$

where  $\text{supp}(P_i)$  means we consider only those elements of the bitstring that correspond to nonidentity elements of the Pauli string  $P_i$ . By construction,  $\mathbb{E}[m_i(q) | B, \theta] = \langle P_i \rangle_\theta$ . Given a fixed budget of circuit executions, we allocate shots across measurement bases using importance sampling. We define basis weights,

$$w_B^{(H)} = \sum_i |\alpha_i| \delta_B(i), \quad w_B^{(H^2)} = \sum_i |\beta_i| \delta_B(i), \quad (19)$$

and sample a measurement basis  $B$  with

$$p_B = \frac{w_B^{(H)} + w_B^{(H^2)}}{\sum_{B' \in \mathcal{B}} (w_{B'}^{(H)} + w_{B'}^{(H^2)})}. \quad (20)$$

The coverage probability for a string  $i$  is given by

$$\xi(i) = \sum_{B \in \mathcal{B}} p_B \delta_B(i). \quad (21)$$

This is the probability that a random basis draw can measure  $P_i$ . After  $S$  shots, where shot  $s = 1, 2, \dots, S$  has corresponding measurement setting and outcome  $(B^{(s)}, q^{(s)})$ , we form the unbiased estimators

$$\widehat{\langle H \rangle} = \alpha_{\mathbb{1}} + \frac{1}{S} \sum_{s=1}^S \sum_i \frac{\alpha_i \delta_{B^{(s)}}(i)}{\xi(i)} m_i(q^{(s)}), \quad (22)$$

$$\widehat{\langle H^2 \rangle} = \beta_{\mathbb{1}} + \frac{1}{S} \sum_{s=1}^S \sum_i \frac{\beta_i \delta_{B^{(s)}}(i)}{\xi(i)} m_i(q^{(s)}). \quad (23)$$

These estimators are unbiased because each Pauli-string contribution is reweighted by the inverse of its sampling probability  $\xi(i)$ . In addition to reusing each shot across all covered strings, the importance sampling additionally concentrates shots on bases that cover strings with large  $|\alpha_i|$  and  $|\beta_i|$ , which reduces variance for fixed  $S$ .

The cost function, Eq. (2), also relies on estimating  $\langle H \rangle_\theta^2$ . If we naively square our estimator  $\widehat{\langle H \rangle}$  we will introduce bias at finite  $S$ . To avoid this bias, we build a U-statistic for  $\langle H \rangle_\theta^2$ . We isolate the per-shot contribution from (22)

$$Y^{(s)} = \sum_i \frac{\alpha_i \delta_{B^{(s)}}(i)}{\xi(i)} m_i(q^{(s)}). \quad (24)$$

The unbiased estimator is then constructed as

$$\widehat{\langle H \rangle^2} = \frac{(\sum_s Y^{(s)})^2 - \sum_s (Y^{(s)})^2}{S(S-1)}. \quad (25)$$

It satisfies  $\mathbb{E}[\widehat{\langle H \rangle^2}] = \langle H \rangle_\theta^2$  provided that the shots are independent. Our estimator for the cost function is then

$$\widehat{C} = (a + b) \widehat{\langle H^2 \rangle} - 2a E_{\text{tar}} \widehat{\langle H \rangle} - b \widehat{\langle H \rangle^2} + a E_{\text{tar}}^2. \quad (26)$$

Unless otherwise stated, we take  $E_{\text{tar}} = 0$  and  $a = b = 1/2$ , which gives the cost function the following form:

$$\widehat{C} = \widehat{\langle H^2 \rangle} - \frac{1}{2} \widehat{\langle H \rangle^2}. \quad (27)$$

This estimator is what is queried when we need to estimate the full cost function such as for SPSA. We note that ADAM with the PSR only requires estimation of  $\langle H \rangle_\theta$  and  $\langle H^2 \rangle_\theta$ , with the U-statistic only needed for logging an estimate of the cost.

### E. Simulating shot noise

Simulating  $S$  independent circuits per optimization iteration to capture the effect of shot noise is expensive in classical benchmarks, particularly when we simulate a density matrix to capture hardware noise. We avoid this by drawing the shot statistics directly. After choosing an average number of shots per QWC measurement circuit  $\bar{S}_{\text{circ}}$ , we set the total shot budget for one estimator evaluation at fixed circuit parameters  $\theta$  to

$$S = G\bar{S}_{\text{circ}},$$

where  $G = |\mathcal{B}|$  is the number of QWC measurement bases. We then allocate this total budget across bases by sampling a multinomial distribution [43] with probabilities  $p_B$  given by Eq. (20), i.e., the distribution of counts obtained from  $S$  independent draws over the possible measurement bases,

$$\vec{K} \sim \text{Mult}(S, \vec{p}_B), \quad \sum_{B \in \mathcal{B}} K_B = S. \quad (28)$$

This allocation gives, on average, more shots to measurement bases covering Pauli strings with larger coefficients in  $H$  and  $H^2$ . Then, for each basis, we draw the histogram of  $K_B$  total outcome bitstrings in that basis from the multinomial distribution with probabilities  $\vec{p}_q$ ,

$$\vec{M}^{(B)} \sim \text{Mult}(K_B, \vec{p}_q), \quad \sum_{q \in \{0,1\}^N} M_q^{(B)} = K_B, \quad (29)$$

where  $p_q$  is given by Eq. (17). With this we can reconstruct the expectation values using our unbiased estimators as before. This approach samples the same shot and basis-allocation noise as  $S$  independent circuit executions, without explicitly simulating each execution individually.

## III. MODELS WITH EMBEDDED SCAR STATES

We now introduce the Hamiltonians used to benchmark the  $\sigma$ -VQE. The guiding principle is to construct interacting spin models that host at least one eigenstate with anomalously low entanglement and otherwise display Wigner-Dyson spectral statistics. We consider two constructions. The first is a Shiraishi-Mori-type projector embedding of a product scar. The second is a parent Hamiltonian construction for an MPS scar of tunable bond dimension.

### A. Shiraishi-Mori embedding

The projector-embedding approach of Shiraishi and Mori [17] is a general method for embedding many-body scar states into the spectrum of a Hamiltonian. It can be readily extended to arbitrary dimensions and model geometries, though we restrict to 2-local interactions on an open spin chain. Let  $\langle i, j \rangle$  denote a pair of neighboring sites  $i$  and  $j$ . Let  $h_{ij}$  be a two-spin Hermitian operator and  $P_{ij}$  a projector acting on sites  $i$  and  $j$ .

The Shiraishi-Mori Hamiltonian is then

$$H = \sum_{\langle i, j \rangle} P_{ij} h_{ij} P_{ij}, \quad (30)$$

where  $\langle i, j \rangle$  indicates we are summing over all bonds in the model. We independently draw a Haar-random normalized single-qubit state  $|\phi_i\rangle$  for each site  $i$ . The projector for each bond is then

$$P_{ij} = \mathbb{1} \otimes \mathbb{1} - |\phi_i\rangle\langle\phi_i| \otimes |\phi_j\rangle\langle\phi_j|. \quad (31)$$

This embeds the product state  $|\psi_{\text{scar}}\rangle = \otimes_i |\phi_i\rangle$  as a scar state with eigenvalue 0, since it is annihilated by each  $P_{ij}$ . The state  $|\psi_{\text{scar}}\rangle$  is clearly a product state, so its bipartite entanglement entropy is strictly zero. We choose the local Hamiltonian terms to be XXZ type with a symmetry breaking transverse field as in Ref. [44],

$$h = J(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y) + \Delta \sigma_z \otimes \sigma_z + b(\sigma_x \otimes \mathbb{1} + \mathbb{1} \otimes \sigma_x), \quad (32)$$

with  $J = b = 1$  and  $\Delta = 0.7$ . We therefore obtain a model that has a product eigenstate with energy zero, while the rest of the spectrum displays Wigner-Dyson level-spacing statistics, specifically those of the Gaussian orthogonal ensemble (GOE). This setting is ideal for testing whether  $\sigma$ -VQE can selectively prepare the scarred eigenstate over the large number of other nearby eigenstates.

One limitation of the Shiraishi-Mori embedding used here is that the embedded scar is a product state. In numerical and analytical studies of many-body scar phenomenology this is often sufficient, but many physical scar states such as in Rydberg chains [15,45–47] exhibit nontrivial internal entanglement. We therefore introduce a second construction that embeds a target state with tunable entanglement.

### B. Parent Hamiltonian embedding

We next construct Hamiltonians that host an entangled scar eigenstate whose entanglement structure we can tune. We do this by building a parent Hamiltonian for a chosen MPS. We consider a ring of  $N$  qubits. We define the normalized, translationally invariant MPS state with bond dimension  $\chi$  as

$$|\psi_{\text{MPS}}\rangle = C \sum_{s_1, \dots, s_N} \text{Tr}[A^{s_1} A^{s_2} \dots A^{s_N}] |s_1 s_2 \dots s_N\rangle, \quad (33)$$

where  $s_i \in [0, 1]$ ,  $C$  is a normalization constant, and  $A^0, A^1 \in \mathbb{C}^{\chi^2}$ . The bond dimension  $\chi$  controls the amount of entanglement in the MPS state. We construct a Hamiltonian operator that annihilates  $|\psi_{\text{MPS}}\rangle$ . We specify an interaction range  $D$  for the ring of spins. We then tile the ring with  $N$  overlapping length- $D$  blocks. On each block we consider a subspace  $\mathcal{A}$  that supports the reduced density matrix of  $|\psi_{\text{MPS}}\rangle$  on those  $D$  sites. The orthogonal complement  $\mathcal{A}_C$  within the full  $2^D$ -dimensional Hilbert space of the block is the set of local states that we will build the Hamiltonian from. We require

$$2^D > \chi^2 \quad (34)$$

in order for the complement space to be nonempty. Let  $\{\psi_n\}$  be an orthonormal basis for  $\mathcal{A}_C$ . We define a  $D$ -local Hermitian operator akin to an annihilation operator acting on the  $i$ th

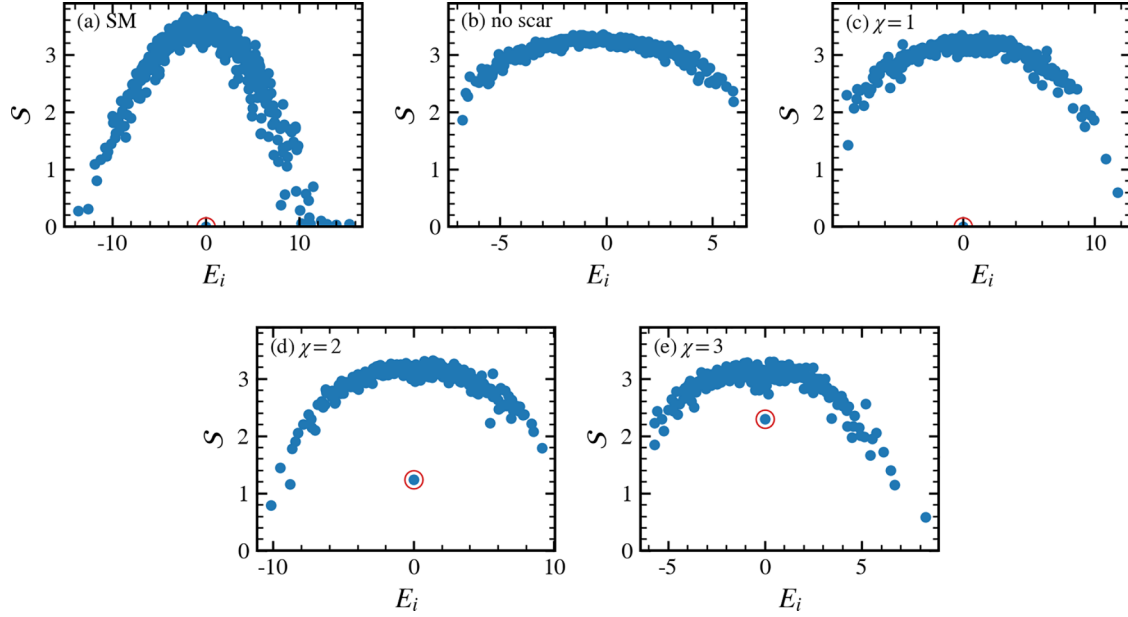


FIG. 2. Bipartite entanglement entropy  $\mathcal{S}$  of eigenstates for the benchmark Hamiltonians. The scar states are circled in red. (a) The  $N = 9$  Shiraishi-Mori (SM) embedding yields a product-state scar at  $E = 0$ . (b) The  $N = 8$  parent Hamiltonian construction without an embedded scar. (c)–(e) The  $N = 8$  parent Hamiltonian construction with an embedded MPS scar of bond dimension  $\chi = 1, 2, 3$ , respectively. Increasing  $\chi$  increases the scar entanglement and moves the scar state closer to the thermal band.

block as

$$h_i = \sum_{m,n \in \mathcal{A}_C} c_{mn} |\psi_m\rangle \langle \psi_n|. \quad (35)$$

As the MPS state has no support in the complement space, a projector operator  $h_i$  constructed from states from  $\mathcal{A}_C$  will annihilate the MPS state. We can choose  $c_{mn}$  to be positive semidefinite. Each  $h_i$  is then positive semidefinite along with satisfying  $h_i |\psi_{\text{MPS}}\rangle = 0$ . The parent Hamiltonian,

$$H_P = \sum_{i=1}^N h_i, \quad (36)$$

will then have  $|\psi_{\text{MPS}}\rangle$  as its ground state. However, we are interested in mid-spectrum scars, not ground states. To embed  $|\psi_{\text{MPS}}\rangle$  into the interior of the spectrum, we instead choose coefficients  $c_{mn}$  that are not all positive. A simple choice is [33]

$$c_{nm} = (-1)^n \delta_{nm}, \quad (37)$$

which alternates signs along the basis of  $\mathcal{A}_C$ . This produces  $h_{ij}$  operators that have both positive and negative eigenvalues. The parent Hamiltonian then has  $|\psi_{\text{MPS}}\rangle$  as an eigenstate with zero eigenvalue but not necessarily as an extremum. We then add small random perturbations supported on  $\mathcal{A}_C$  to ensure that the spectrum has GOE-like level statistics.

The parent Hamiltonian embedding provides a benchmark that is qualitatively different from the Shiraishi-Mori approach. By varying  $\chi$  and the interaction range  $D$  we can tune the entanglement of the scar eigenstate. This allows us to examine the performance of  $\sigma$ -VQE when preparation of the scar requires some entanglement be generated by the shallow circuit.

To validate that our benchmarks contain a single scar low-entanglement eigenstate amid otherwise thermal eigenstates, we plot the bipartite entanglement entropy of the full spectrum in Fig. 2. The bipartite entanglement entropy of each eigenstate  $\rho_i$  (with corresponding eigenvalue  $E_i$ ) is defined using a contiguous real-space bipartition into subsystems  $A$  and  $B$ , where for the open-chain Shiraishi-Mori model  $A$  and  $B$  are the left- and right-hand parts of the chain, while for the parent Hamiltonian ring we cut the ring at a fixed bond and take  $A$  and  $B$  to be two contiguous halves,

$$\mathcal{S} = -\text{Tr}[\rho_i^A \log_2 \rho_i^A], \quad (38)$$

where the reduced density matrices are found by tracing out the support of partition  $B$ , i.e.,  $\rho_i^A = \text{Tr}_B[\rho_i^{AB}]$ .

For the Shiraishi-Mori embedding, the embedded scar is a product state and appears as a clearly isolated eigenstate at  $E = 0$ . For parent Hamiltonian embedding with  $N = 8$ ,  $D = 4$ , tuning the bond dimension  $\chi$  allows us to increase the scar entanglement. As we do this the scar becomes less separated from the thermal band, providing a control parameter for testing how the performance of  $\sigma$ -VQE degrades as the entanglement gap is reduced (and thus the shallow ansatz becomes insufficient to prepare mid-spectrum eigenstates).

#### IV. RESULTS

We benchmark the  $\sigma$ -VQE using both models outlined in Sec. III with the goal of assessing whether a shallow, hardware-efficient ansatz combined with an energy-targeting, variance-penalized objective can reliably converge to a QMBS when such an atypical low-entanglement eigenstate exists near a chosen target energy. We quantify performance using the cost function, Eq. (2), taking  $a = b = 1/2$ . We additionally track the fidelity of the state produced by the circuit  $|\psi_\theta\rangle$

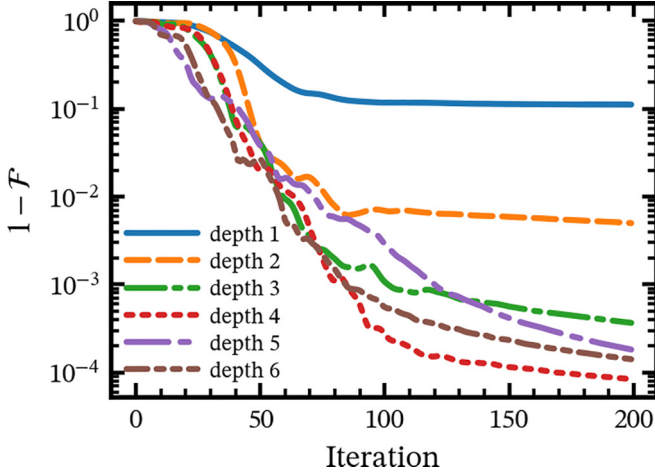


FIG. 3.  $\sigma$ -VQE performance when searching for the SM-embedded nonintegrable  $N = 9$  open XXZ chain. The results shown are from setting the target energy of the cost function to zero, the energy of the scar state. We use ADAM with exact (state-vector) evaluation of all expectation values and parameter-shift gradients. We simulate using 1–6 ansatz layers. We plot the infidelity ( $1 - \mathcal{F}$ ) of the circuit state with the scar state vs iteration of the algorithm. We observe that even a relatively shallow depth of 2 is enough to achieve significant overlap with the scar state in this setting.

with the target scar state  $|\psi_{\text{scar}}\rangle$ ,

$$\mathcal{F}_\theta = |\langle \psi_{\text{scar}} | \psi_\theta \rangle|^2. \quad (39)$$

#### A. State-vector simulations

We first benchmark  $\sigma$ -VQE using exact state-vector simulations for numerical evaluation of the cost objective (and its gradients). We then make these simulations more realistic by using the techniques described in Sec. II E to simulate the effect of shot noise on trainability. In both cases we assume noiseless circuit execution and measurement. Optimization is performed with ADAM using PSR, and the hardware-efficient ansatz is initialized from  $|0\rangle^{\otimes N}$  with small random parameters (of order  $10^{-3}$ ).

Figure 3 shows  $\sigma$ -VQE applied to the  $N = 9$  Shiraishi-Mori-embedded open XXZ chain taking  $E_{\text{tar}} = 0$ , the eigenvalue of the scar state. We run the algorithm with circuit depths ranging from 1 to 6. All other hyperparameters are kept fixed. The main observation is that substantial overlap with the scar is already achievable with low-depth circuits, indicating that the low-entanglement structure of the target state is accessible to the shallow ansatz. The results shown do not suggest a simple monotonic improvement with depth, and we therefore limit ourselves here to the conclusion that low-depth circuits already perform well in this benchmark. To test whether  $\sigma$ -VQE succeeds in lowering the cost function only near anomalously low-entanglement eigenstates, we use the same  $N = 9$  Shiraishi-Mori XXZ Hamiltonian and sweep the target energy  $E_{\text{tar}}$  over a range of values. For each  $E_{\text{tar}}$  value we run a 300-iteration  $\sigma$ -VQE routine, and we additionally repeat the procedure for a control case of a disordered XXZ chain without an embedded scar. As shown in Fig. 4, the cost objective drops substantially only when the target is

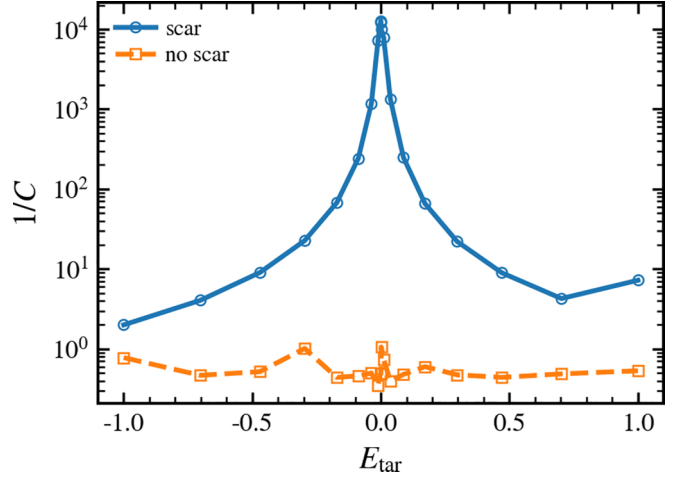


FIG. 4. We run the  $\sigma$ -VQE for the  $N = 9$  Shiraishi-Mori embedded chaotic XXZ model, now varying the target energy that is set in the cost function, Eq. (2). We run the  $\sigma$ -VQE routine for 300 iterations with an ansatz depth of 3. We observe that the algorithm is only successful in reducing the cost function significantly when we specifically target the scar. Eigenstates at other target energies require more entanglement to be generated, and therefore are out of reach of the low-depth circuit. We additionally run the VQE for a control case of the chaotic XXZ model without an embedded scar. The shallow ansatz cannot significantly lower the cost function at all in the control case.

tuned to the scar energy. Away from the scar the optimization largely stalls, consistent with the fact that the shallow ansatz is unable to represent generic midspectrum (volume-law entangled) eigenstates. In the no-scar control case, the same ansatz fails to reduce the cost objective significantly across the scanned energies. This energy selectivity is consistent with the design principle of the  $\sigma$ -VQE. Together, Figs. 3 and 4 illustrate the central mechanism exploited for the algorithm. When entanglement generation is intentionally restricted, successful eigenstate targeting becomes selective for those states that are anomalously low entanglement in an otherwise ETH-obeying system. These noiseless baseline tests establish functionality in the limit of infinite measurement statistics. We next introduce finite-shot sampling of the cost, keeping the underlying circuit evolution noiseless. In this regime, the cost is evaluated via the unbiased estimators defined in Sec. II D. We reiterate that we reuse each measured bitstring across multiple Pauli strings, with simultaneous reconstruction of multiple expectation values within a qubit-wise commuting group to reduce the number of circuits required for accurate measurement of expectation values. Under independent sampling, the variance of the unbiased estimator variance decreases as  $1/S$  due to the central limit theorem, where  $S$  is the shot budget per gradient evaluation. Different grouping schemes for Pauli strings primarily affect the prefactor for this decay. A number of schemes for grouping Pauli strings and creating measurement settings  $\mathcal{B}$  have been proposed in the literature [28,30,48–50] in an effort to reduce this estimation variance and the number of circuits required for the execution of NISQ algorithms and beyond. For our purposes we use qiskit’s default QWC grouping scheme for generating the measurement bases.

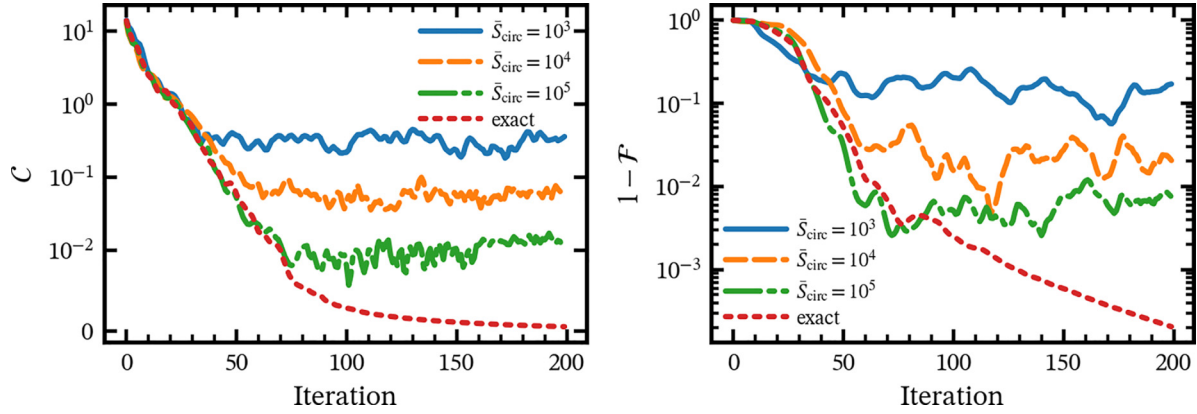


FIG. 5. We simulate the  $\sigma$ -VQE with a depth of 3 for the  $N = 9$  open XXZ chain with a Shiraishi-Mori random-product QMBS state, using average shot budgets  $\bar{S}_{\text{circ}} \in \{10^3, 10^4, 10^5\}$  per physical QWC circuit evaluation. The plotted cost  $\mathcal{C}$  is the exact state-vector cost evaluated along the finite-shot optimization trajectories. The parameter updates themselves are computed using shot-noisy estimates. This probes whether shot-noisy parameter updates still produce progress in the underlying noiseless cost landscape.

Figure 5 shows the same  $\sigma$ -VQE being run for the  $N = 9$  SM embedded chaotic XXZ model using the QWC group estimation scheme. We report the finite-shot data in terms of the average number of shots per physical QWC circuit evaluation,  $\bar{S}_{\text{circ}}$ . This corresponds to a total shot budget  $S_{\text{eval}} = G\bar{S}_{\text{circ}}$ , where  $G$  is the number of QWC measurement groups. In this instance  $G = 193$ , and the budget is divided across groups using the importance-weighted allocation described in Sec. II E.

For the depth-3 ansatz,  $P = 126$ , so each parameter-shift rule ADAM update requires  $2P + 1$  estimator evaluations, giving a total shot count per iteration of  $(2P + 1)G\bar{S}_{\text{circ}}$  not including any additional cost-logging evaluation. We use  $\bar{S}_{\text{circ}} \in \{10^3, 10^4, 10^5\}$ , where  $10^3$  and  $10^4$  are typical reference shot counts per circuit and  $10^5$  is a high-statistics reference.

In the cost panel we plot the exact state-vector cost  $\mathcal{C}$  evaluated at the parameters produced by the finite-shot optimization. The parameter updates are computed using shot-noisy estimates, while the displayed cost is evaluated without shot noise, separating progress in the underlying noiseless cost landscape from fluctuations of the finite-shot cost estimator. We stress that the noiseless curve is a reference optimization trajectory, not a lower bound on the infidelity or cost. Finite-shot gradient estimates generate different stochastic optimization paths, and fidelity is not the optimized quantity, so a finite-shot trajectory can pass through points with lower instantaneous infidelity than the noiseless reference at the same iteration. At smaller  $\bar{S}_{\text{circ}}$ , the larger estimator variance limits the precision with which the cost landscape and gradients are resolved, consistent with the higher apparent optimization floor in the examples shown. We therefore conclude that the unbiased estimator remains able to support  $\sigma$ -VQE optimization under finite-shot sampling, although the total PSR-ADAM circuit overhead remains substantial.

The Shiraishi-Mori benchmark embeds a strict product eigenstate and is therefore “easy” from an entanglement standpoint. As the qubit register is initialized in a product state and we select small initial circuit parameters, optimization is relatively straightforward. To assess performance beyond product scars, we benchmark  $\sigma$ -VQE on the parent Hamiltonian (PH) construction, which embeds a MPS as an exact

eigenstate with bond dimension  $\chi$  (see Sec. III B for details). The bond dimension is a free parameter which allows us to increase the entanglement of the embedded scar, providing a controlled interpolation between product-like and more entangled (yet still subthermal) target states. This is visible directly in the exact-spectrum entanglement diagnostics in Fig. 2. As  $\chi$  increases, the anomalous eigenstate becomes less isolated from the thermal band. Consequently,  $\sigma$ -VQE is forced into a regime where the circuit ansatz must generate enough structure to represent the scar, while still remaining too entanglement limited to approximate generic midspectrum eigenstates at nearby energies.

We benchmark  $\sigma$ -VQE using the parent Hamiltonian construction in Fig. 6. We use  $N = 8$  qubits, interaction range  $D = 4$ , and a depth-4 hardware-efficient ansatz. For each  $\chi \in \{1, 2, 3\}$  and at each target energy  $E_{\text{tar}}$ , we run 300 iterations of the  $\sigma$ -VQE routine using PSR gradient evaluations. The same procedure is repeated for a control Hamiltonian with similar structure in which no scar is embedded. We plot both the inverse cost and the fidelity to the known scar state (when one is present). The inverse cost peaks around  $E_{\text{tar}} = 0$ , the energy of the embedded scar state, reflecting the explicit energy-targeting structure of the  $\sigma$ -VQE objective function. The fidelity exhibits broader plateaus instead of peaks as a function of  $E_{\text{tar}}$  than the inverse of the cost function. The optimization can converge to states with appreciable overlap with the embedded MPS scar, while  $E_{\text{tar}}$  is detuned from the exact eigenvalue of the scar despite the cost function itself being penalized for that detuning. This is consistent with the hypothesis of scar states being particularly attractive low-entanglement solutions for the variational search with shallow circuits. The inverse cost peak is therefore a more suitable diagnostic of the correct energy window. As  $\chi$  increases,  $\sigma$ -VQE performance as a high-fidelity state preparation protocol degrades. Increasing the MPS bond dimension raises the entanglement required to represent the scar and reduces its separation from the thermal band, narrowing the regime where a shallow ansatz can represent the scar while still failing to select for typical eigenstates. The no-scar control provides the corresponding null test. In the absence of an anomalous low-entanglement eigenstate, the same shallow variational

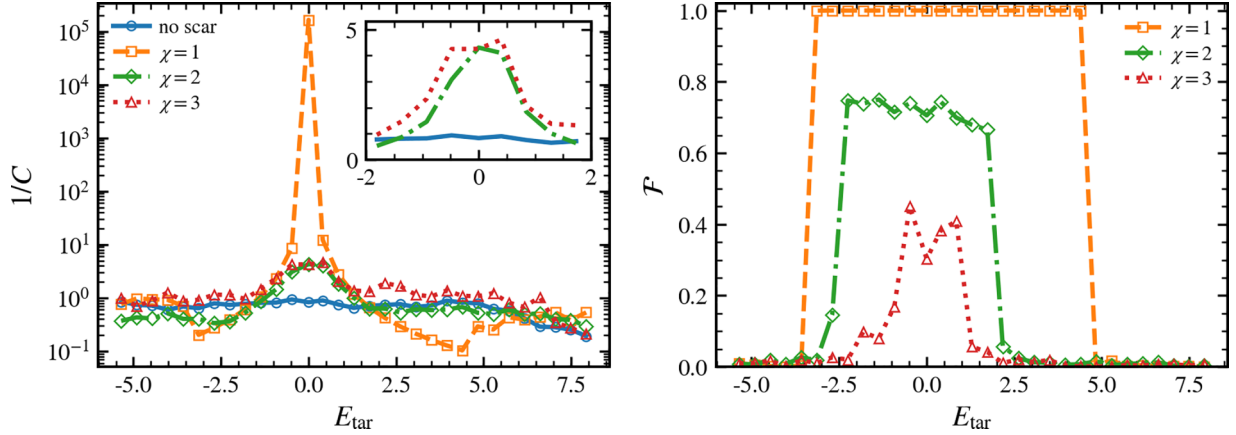


FIG. 6.  $\sigma$ -VQE performance when searching for a random matrix product state QMBS that is embedded via the parent Hamiltonian method taking  $N = 8$ ,  $\chi = 1, 2, 3$ , and  $D = 4$  together with a no-scar control Hamiltonian. For each target energy  $E_{\text{tar}}$ , we run a single depth-4 hardware-efficient ansatz optimization using 300 iterations of PSR gradient evaluations. We plot inverse cost and fidelity to the embedded MPS scar vs  $E_{\text{tar}}$ . Broad fidelity plateaus compared to a narrower inverse-cost peak indicate that the variational search can still find substantial scar overlap even while the energy detuning is penalized by the cost function.

search does not produce a comparable signature with the cost function.

The parent Hamiltonian results support two practical conclusions relevant to scar-hunting in less engineered models. First, the  $\sigma$ -VQE need not be restricted to product-like scars. It can identify and prepare moderately entangled scars so long as they remain atypical relative to the surrounding thermal spectrum. Additionally, the observed fidelity plateaus suggest a useful workflow when the scar energy is not known *a priori*. A coarse sweep in  $E_{\text{tar}}$  can be used to locate candidate windows via an inverse-cost peak, while the broad overlap region indicates that the algorithm can still converge to the relevant low-entanglement structure even before the target energy is perfectly tuned.

We note some further improvements can be made to our strategy if we wish to maximize fidelity with the scar state. While we restrict Fig. 6 to single-run optimization to characterize baseline behavior, standard variational-algorithm strategies may improve robustness in higher  $\chi$  regimes. Examples include restart strategies with different initial circuit parameters to mitigate optimizer sensitivity in rugged landscapes combined with the use of cheaper initial optimizers such as SPSA when one wishes to reduce per-iteration measurement overhead to explore many initial positions in the optimization landscape quickly. More broadly, the VQE literature contains a range of practical techniques that are known to affect convergence. Comparative studies show strong dependence on the classical optimizer, with gradient-based methods often outperforming common derivative-free choices when gradients are accessible [37,38]. Geometry-aware updates such as the quantum natural gradient can further improve performance [51]. Likewise, circuit or parameter initialization strongly affects trainability. A given random initialization can place the optimization on barren plateaus [52], motivating strategies such as identity-block initialization [53]. Additionally, optimizing the circuit structure itself can improve outcomes [54].

Finally, as the estimator variance depends on how shots are distributed across measurement bases, variance-aware shot

allocation can accelerate convergence at a fixed measurement budget. Shot-adaptive approaches allocate shots based on gradient or estimator variance to aid estimation [55–58]. In addition, reinforcement-learning methods for shot assignment have also been proposed [59]. In the present context, such methods are best viewed as ways to improve performance at fixed shot counts, rather than qualitatively changing the mechanism behind  $\sigma$ -VQE.

## B. Hardware demonstration

We performed a proof-of-principle demonstration of  $\sigma$ -VQE on superconducting hardware. Using several IBM Quantum Platform trial allocations, we obtained brief access, totaling approximately 50 min of QPU time on the cloud-based `ibm_fez` on December 10, 2025, using the connected physical-qubit chain 144–143–142. The calibrated coherence times for qubits 144, 143, and 142 were  $(T_1, T_2) = (140.96, 114.86) \mu\text{s}$ ,  $(97.40, 114.54) \mu\text{s}$ , and  $(94.69, 106.03) \mu\text{s}$ , respectively, with readout errors  $3.66 \times 10^{-3}$ ,  $6.35 \times 10^{-3}$ , and  $5.37 \times 10^{-3}$ . The calibrated  $\text{cz}$  errors on 144–143 and 143–142 were  $2.24 \times 10^{-3}$  and  $2.17 \times 10^{-3}$ , while the corresponding  $\text{rzz}$  errors were  $3.40 \times 10^{-3}$  and  $3.48 \times 10^{-3}$ .

We restricted the demonstration to error-mitigation techniques that do not multiply the sampling cost by large constant factors: dynamical decoupling sequences were enabled on idle qubits, and Pauli twirling was applied to two-qubit gates and measurement. In contrast, we did not use methods such as probabilistic error cancellation, error amplification, or zero-noise extrapolation [60,61]. Although powerful, these methods significantly increase circuit volume and/or require repeated evaluations at multiple effective noise levels, making them incompatible with the present shot-limited demonstration and with shallow-budget access regimes.

To implement the estimator described in Sec. II D, we used the `qiskit` sampler primitive rather than the estimator primitive. `sampler` returns raw shot bitstrings for each circuit instance, which was required for our reuse of each bitstring

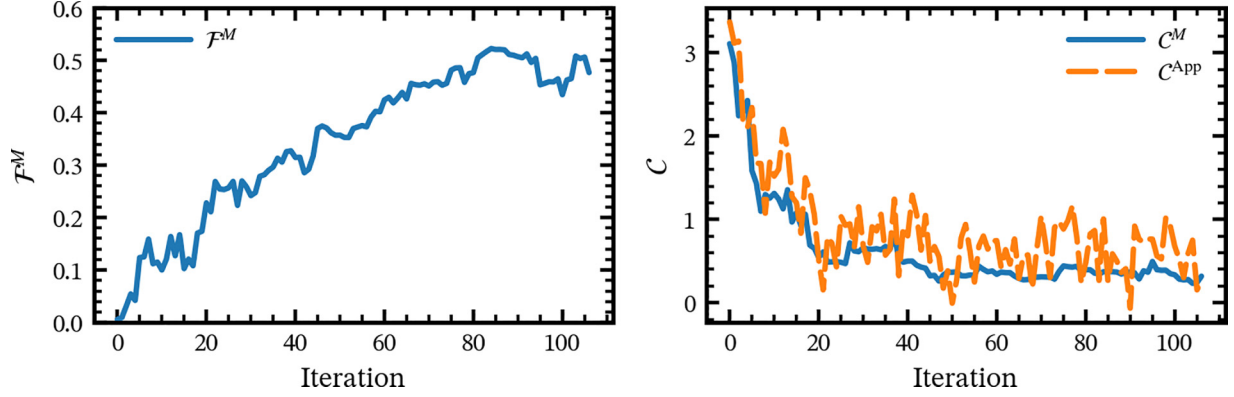


FIG. 7. Hardware implementation of  $\sigma$ -VQE for the  $N = 3$  SM-embedded XXZ chain on `ibm_fez` (a Heron r2 QPU with heavy-hex topology), using a depth-2 hardware-efficient ansatz and SPSA optimization. We plot the on-device proxy cost  $C^{\text{App}}$  computed from shifted evaluations used by SPSA together with an offline reconstructed “model” cost  $C^M$  and model fidelity  $\mathcal{F}^M$  obtained by replaying the hardware parameter trajectory in a noiseless state-vector simulation. We used dynamical decoupling on idle qubits and Pauli twirling on two-qubit gates and readout as error mitigation. Total QPU usage was 50 min, accessed in December 2025.

across multiple Pauli strings within the same QWC measurement group. estimator aggregates observables internally and therefore discards the bitstring-level correlations exploited by our postprocessing. Use of the sampler primitive also limited the error-mitigation techniques available in qiskit.

The limited shot budget prevented us from logging the cost function at each iteration of the optimizer. Since SPSA only requires the shifted evaluations  $C(\theta_i^{(+)})$  and  $C(\theta_i^{(-)})$  to construct its gradient estimate, we devoted the available shots entirely to these evaluations and monitored progress using an on-device proxy that required no additional circuit executions,

$$C_k^{\text{App}} = \frac{C(\theta_k^{(+)}) + C(\theta_k^{(-)})}{2}, \quad (40)$$

where  $C(\theta_k^{(\pm)})$  are the shifted cost function evaluations used for the SPSA optimizer as defined in (11).

To maximize robustness, we intentionally targeted a minimal model, the  $N = 3$  Shiraishi-Mori embedded chaotic XXZ model with a depth-2 hardware-efficient ansatz. The low depth reduced the two-qubit gate count, idle time, and exposure to readout noise. We additionally transpiled the logical chain onto a connected set of physical qubits with the lowest available two-qubit error rates. Deeper circuits and larger  $N$  degrade the optimization signal at fixed shots. Emulation of the noisy hardware suggested that instances beyond  $N = 3$  were not reliably trainable without stronger mitigation (Appendix).

The hardware results are shown in Fig. 7. We used 5 000 shots per SPSA iteration (535 000 total). The on-device proxy  $C_k^{\text{App}}$  shows a clear downward trend, indicating that—even under restricted sampling and lightweight mitigation—the  $\sigma$ -VQE objective retains a usable optimization signal on hardware. To interpret this trajectory beyond a noisy proxy, we also perform an offline “model replay.” We take the parameter sequence  $\{\theta_i\}$  produced by the on-device SPSA loop and evaluate the corresponding “model” cost  $C_k^M$  and “model” fidelity  $\mathcal{F}_M$  in a noiseless state-vector simulation (where the embedded scar is known numerically). The replay confirms that the parameter trajectory learned on the noisy hardware

corresponds to improved model cost and increased overlap with the scar. The significance of the hardware run is not the absolute system size, but that the combination of a shot-frugal optimizer, bitstring-level postprocessing for nonlinear objectives, and lightweight mitigation suffices to obtain an on-device optimization signal.

## V. CONCLUDING REMARKS

In this work we benchmark a hybrid quantum-classical protocol,  $\sigma$ -VQE, for uncovering quantum many-body scar eigenstates. The method combines a nonlinear, energy-targeting objective with a variance penalty, alongside a shallow hardware-efficient circuit ansatz that restricts accessible entanglement to bias optimization toward low-entanglement eigenstates [20]. The resource overhead of the algorithm is reduced by a classical shadow-motivated QWC measurement strategy together with an unbiased finite-shot estimator for the nonlinear cost function, ensuring that finite-sampling effects do not systematically bias the variance-related terms.

Across noiseless simulations,  $\sigma$ -VQE rapidly converges for Shiraishi-Mori projector-embedded product scars, while parent Hamiltonian MPS embeddings provide a controlled view of failure modes as the embedded scar entanglement increases toward typical thermal eigenstates in the same energy window. Finally, a proof-of-principle superconducting-hardware execution (Fig. 7) demonstrates end-to-end feasibility under tight shot constraints using a shot-frugal optimizer, bitstring-level postprocessing, and lightweight mitigation.

These results support interpreting  $\sigma$ -VQE as a diagnostic for scarred structure near a chosen target energy density. If a low-entanglement scar exists near the target energy, the shallow ansatz can represent it and the nonlinear objective supports convergence, whereas the same circuit family fails to realize generic midspectrum eigenstates. The results additionally show that  $\sigma$ -VQE acts as a state-preparation primitive for initializing states with significant overlap with nonthermal states for subsequent dynamical studies, such as probing anomalous relaxation [15,45]. Many applications do not

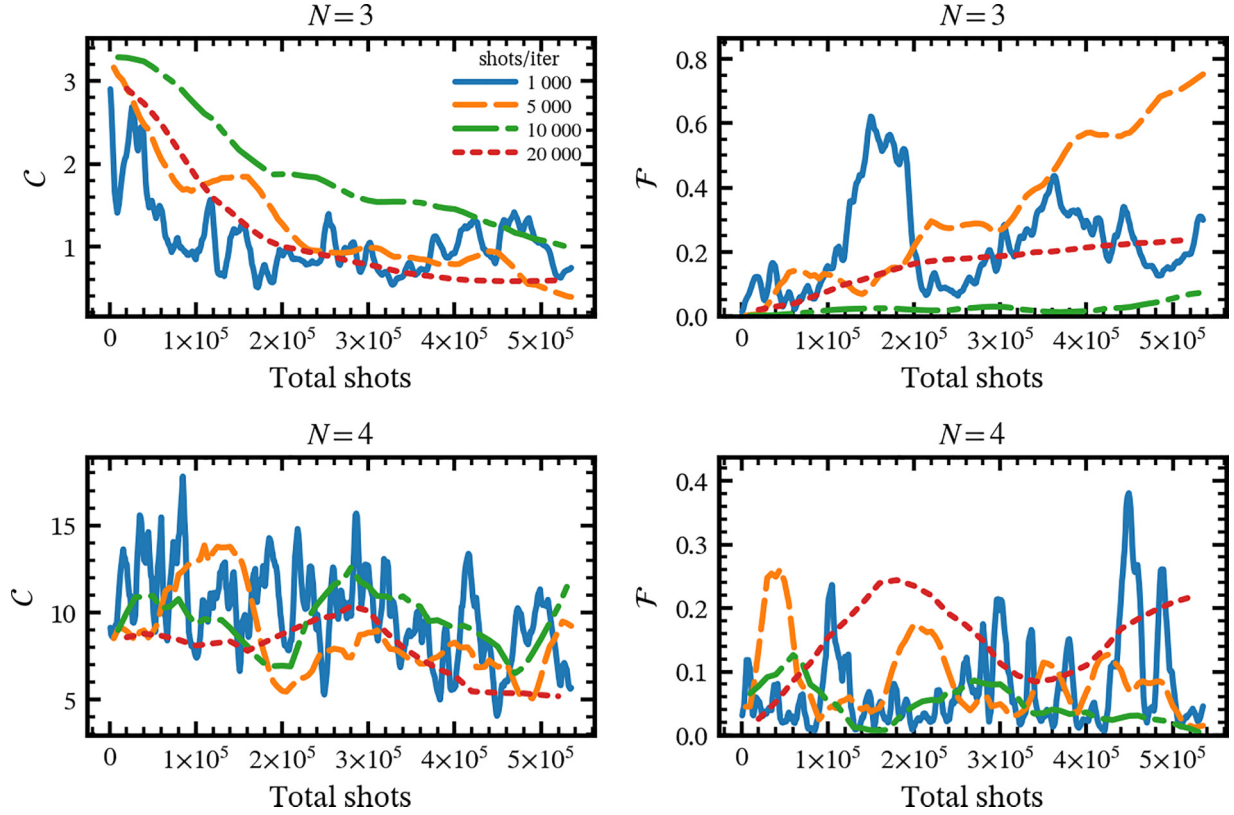


FIG. 8. The results of our noisy emulation using 535 000 total shots (the amount accessed during the hardware run). We use SPSA and a depth-2 hardware-efficient ansatz. Finite-shot outcomes were generated via multinomial sampling from the noisy output distributions. We compare system sizes of  $N = 3$  and  $N = 4$  while using different budgets for shots per SPSA iteration  $S$ . For  $N = 3$ , optimization remains viable, while for  $N = 4$  the noise model prevents reliable convergence within the same shot budget.

require high-fidelity scar eigenstate preparation. Preparing a state with appreciable overlap onto a scar, as seen in the MPS case, is sufficient to exhibit long-lived coherent dynamics, consistent with the wider scar literature [15,45,62,63]. This perspective also aligns with proposed scar-enabled applications such as robust sensing and metrological enhancement [46,64].

$\sigma$ -VQE fits within the broader class of variational excited-state methods [23–26]. As with all VQEs, its practical scaling is constrained by the growth of measurement overhead and the fragility of the optimization signal under finite-shot, noisy execution. In our setting, lightweight mitigation is sufficient for a proof-of-principle hardware demonstration, but scaling to larger systems on current devices will likely require either substantially higher shot budgets or stronger error mitigation to recover a useful optimization signal [60,61,65,66]. Improving measurement efficiency to counteract the growth of Pauli terms (such as through grouping, derandomization, parallelization, and variance-aware shot allocation [28–30,56]) and better device-level noise performance will therefore be central to extending  $\sigma$ -VQE beyond the small-system regime.

#### ACKNOWLEDGMENTS

The authors thank Shane Dooley and Phillip C. Burke for insightful discussions on the models explored in this work. E.C. would particularly like to thank Chloe Smith

and Jack Rivington for conversations on the broader scope of the project near its outset. E.C., G.C., and G.B. acknowledge financial support from the “National Centre for HPC, Big Data and Quantum Computing,” Spoke 10, Project No. CN00000013 QUOVADIS, and from INFN through the project QUANTUM. N.K. and G.C. acknowledge the Spanish State Research Agency through the María de Maeztu Project No. CEX2021-001164-M and the COQUSY Projects No. PID2022-140506NB-C21 and No. PID2022-140506NB-C22, all funded by MCIU/AEI/10.13039/501100011033. N.K. further acknowledges the QuantCom Project No. CNS2024-154720. G.C. further acknowledges the INFOLANET Project No. PID2022-139409NB-I00 and the QuantERA QNet Project No. PCI2024-153410, funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. The data presented in this work are available upon request.

#### DATA AVAILABILITY

The numerical data supporting the figures in this work, including the hardware-experiment data, are available from the corresponding author upon reasonable request.

## APPENDIX: HARDWARE EMULATION

We used qiskit aer to emulate  $\sigma$ -VQE runs under a device-calibrated noise model for IBM Fez. For a chosen physical-qubit chain, we transpiled the variational circuits onto the Fez device layout and constructed a noise model from the backend calibration data, including gate error rates and durations, per-qubit T1 and T2 times, and readout assignment errors. This emulator captures dominant stochastic error channels at the level provided by the calibration snapshot, but it does not include several effects that can be relevant on hardware, such as coherent over- or underrotations and miscalibrations, drift, and crosstalk. The emulation results should therefore be interpreted as an optimistic proxy for trainability rather than a faithful predictor of hardware performance.

We fixed a total budget of 535 000 shots and used SPSA together with a depth-2 hardware-efficient ansatz, reflecting the shot- and depth-frugal regime targeted in the hardware demonstration. We searched over a grid of shots per iteration of the SPSA optimizer ( $S = 10^3, 5 \times 10^3, 10^4, 2 \times 10^4$ )

which trades estimator variance against the number of optimizer iterations available within the fixed total budget. For each choice of  $S$  we generated the requisite number of bitstring outcomes via the multinomial sampling procedure described in Sec. II E. These bitstrings were then used to estimate the shifted cost evaluations for the SPSA gradient approximation. We plot the cost and fidelity with the scar state using access to the full simulated density matrix of the circuit.

The results are shown in Fig. 8. For  $N = 3$  optimization remained viable within the fixed budget. We find that allocating fewer shots per iteration (hence allowing more iterations) generally improved the final outcome until the SPSA updates became dominated by shot noise. In contrast, for  $N = 4$  the same budget and circuit depth did not yield reliable convergence under the Fez-calibrated noise model. This supports the design choice used for the hardware run. Within the available shot budget and without stronger mitigation, increasing  $N$  rapidly degrades the optimization signal, and larger instances are unlikely to be trainable in the same regime.

- 
- [1] W. Ritz, Über eine neue methode zur lösung gewisser variation-sprobleme der mathematischen physik, *J. Reine Angew. Math.* **1909**, 1 (1909).
  - [2] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).
  - [3] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
  - [4] J. Eisert and J. Preskill, Mind the gaps: The fraught road to quantum advantage, *arXiv:2510.19928*.
  - [5] M. da Silva Fonseca, C. Moraes Porto, N. A. Cabrera Carpio, G. de Souza Tavares de Morais, N. H. Morgon, R. A. Nome, and C. J. Villas-Boas, An introduction to variational quantum eigensolver applied to chemistry, *Braz. J. Phys.* **56**, 8 (2026).
  - [6] J. Hu, J. Li, Y. Lin, H. Long, X.-S. Xu, Z. Su, W. Zhang, Y. Zhu, and M.-H. Yung, Benchmarking variational quantum eigensolvers for quantum chemistry, *arXiv:2211.12775*.
  - [7] N. Yoshioka, T. Sato, Y. O. Nakagawa, Y.-y. Ohnishi, and W. Mizukami, Variational quantum simulation for periodic materials, *Phys. Rev. Res.* **4**, 013052 (2022).
  - [8] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruzekiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer, *Nat. Commun.* **13**, 5743 (2022).
  - [9] O. Huijgen, L. Coopmans, P. Najafi, M. Benedetti, and H. J. Kappen, Training quantum Boltzmann machines with the  $\beta$ -variational quantum eigensolver, *Mach. Learn.: Sci. Technol.* **5**, 025017 (2024).
  - [10] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
  - [11] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
  - [12] J. M. Deutsch, Quantum statistical mechanics in a closed system, *Phys. Rev. A* **43**, 2046 (1991).
  - [13] M. Srednicki, Chaos and quantum thermalization, *Phys. Rev. E* **50**, 888 (1994).
  - [14] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics, *Adv. Phys.* **65**, 239 (2016).
  - [15] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, *Nature (London)* **551**, 579 (2017).
  - [16] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Weak ergodicity breaking from quantum many-body scars, *Nat. Phys.* **14**, 745 (2018).
  - [17] N. Shiraishi and T. Mori, Systematic construction of counterexamples to the eigenstate thermalization hypothesis, *Phys. Rev. Lett.* **119**, 030601 (2017).
  - [18] A. Chandran, T. Iadecola, V. Khemani, and R. Moessner, Quantum many-body scars: A quasiparticle perspective, *Annu. Rev. Condens. Matter Phys.* **14**, 443 (2023).
  - [19] S. Moudgalya, B. A. Bernevig, and N. Regnault, Quantum many-body scars and Hilbert space fragmentation: A review of exact results, *Rep. Prog. Phys.* **85**, 086501 (2022).
  - [20] G. Cenedese, M. Bondani, A. Andreanov, M. Carrega, G. Benenti, and D. Rosa, Shallow quantum circuits are robust hunters for quantum many-body scars, *Eur. Phys. J. Plus* **140**, 517 (2025).
  - [21] L. Cadi Tazi and A. J. W. Thom, Folded spectrum VQE: A quantum computing method for the calculation of molecular excited states, *J. Chem. Theory Comput.* **20**, 2491 (2024).
  - [22] F. Zhang, N. Gomes, Y. Yao, P. P. Orth, and T. Iadecola, Adaptive variational quantum eigensolvers for highly excited states, *Phys. Rev. B* **104**, 075159 (2021).
  - [23] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, *Quantum* **3**, 156 (2019).

- [24] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Res.* **1**, 033062 (2019).
- [25] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, *Phys. Rev. A* **95**, 042308 (2017).
- [26] S. Gocho, H. Nakamura, S. Kanno, Q. Gao, T. Kobayashi, T. Inagaki, and M. Hatanaka, Excited state calculations using variational quantum eigensolver with spin-restricted ansätze and automatically-adjusted constraints, *npj Comput. Mater.* **9**, 13 (2023).
- [27] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, *Adv. Quantum Technol.* **2**, 1900070 (2019).
- [28] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* **152**, 124114 (2020).
- [29] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [30] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, *Phys. Rev. Lett.* **127**, 030503 (2021).
- [31] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [32] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [33] P. G. Larsen and A. E. B. Nielsen, Phase transitions in quantum many-body scars, *Phys. Rev. Res.* **6**, L042007 (2024).
- [34] J. Kim, J. Kim, and D. Rosa, Universal effectiveness of high-depth circuits in variational eigenproblems, *Phys. Rev. Res.* **3**, 023203 (2021).
- [35] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [36] S. Liu, S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, Probing many-body localization by excited-state variational quantum eigensolver, *Phys. Rev. B* **107**, 024204 (2023).
- [37] Q. Lu, Comparative studies of optimizer performance in a variational quantum eigensolver, *J. Chem. Phys.* **163**, 234106 (2025).
- [38] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. de Jong, Classical optimizers for noisy intermediate-scale quantum devices, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2020), pp. 267–277.
- [39] J. C. Spall, Implementation of the simultaneous perturbation algorithm for stochastic optimization, *IEEE Trans. Aerosp. Electron. Syst.* **34**, 817 (1998).
- [40] A. Pellow-Jarman, I. Sinayskiy, A. Pillay, and F. Petruccione, A comparison of various classical optimizers for a variational quantum linear solver, *Quantum Inf. Process* **20**, 202 (2021).
- [41] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, Measuring all compatible operators in one series of single-qubit measurements using unitary transformations, *J. Chem. Theory Comput.* **16**, 2400 (2020).
- [42] B. DalFavero, R. Sarkar, J. Rowland, D. Camps, N. P. D. Sawaya, and R. LaRose, Measurement reduction for expectation values via fine-grained commutativity, *Phys. Rev. A* **112**, 052407 (2025).
- [43] S. Ross, *A First Course in Probability*, 10th ed. (Pearson, Hoboken, 2023).
- [44] P. C. Burke and S. Dooley, Taking the temperature of quantum many-body scars, *Phys. Rev. Res.* **7**, 043054 (2025).
- [45] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Quantum scarred eigenstates in a Rydberg atom chain: Entanglement, breakdown of thermalization, and stability to perturbations, *Phys. Rev. B* **98**, 155134 (2018).
- [46] S. Dooley, S. Pappalardi, and J. Goold, Entanglement enhanced metrology with quantum many-body scars, *Phys. Rev. B* **107**, 035123 (2023).
- [47] A. Hudomal, I. Vasić, N. Regnault, and Z. Papić, Quantum scars of bosons with correlated hopping, *Commun. Phys.* **3**, 99 (2020).
- [48] S. Hillmich, C. Hadfield, R. Raymond, A. Mezzacapo, and R. Wille, Decision diagrams for quantum measurements with shallow circuits, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2021), pp. 24–34.
- [49] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).
- [50] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, *J. Chem. Theory Comput.* **16**, 190 (2020).
- [51] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, *Quantum* **4**, 269 (2020).
- [52] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [53] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, *Quantum* **3**, 214 (2019).
- [54] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, *Quantum* **5**, 391 (2021).
- [55] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, An adaptive optimizer for measurement-frugal variational algorithms, *Quantum* **4**, 263 (2020).
- [56] A. Gu, A. Lowe, P. A. Dub, P. J. Coles, and A. Arrasmith, Adaptive shot allocation for fast convergence in variational quantum algorithms, *arXiv:2108.10434*.
- [57] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, Operator sampling for shot-frugal optimization in variational algorithms, *arXiv:2004.06252*.
- [58] L. Zhu, S. Liang, C. Yang, and X. Li, Optimizing shot assignment in variational quantum eigensolver measurement, *J. Chem. Theory Comput.* **20**, 2390 (2024).
- [59] S. Liang, L. Zhu, X. Liu, C. Yang, and X. Li, Artificial-intelligence-driven shot reduction in quantum measurement, *Chem. Phys. Rev.* **5**, 041403 (2024).
- [60] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).

- [61] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, [Nature \(London\) \*\*567\*\*, 491 \(2019\)](#).
- [62] Q. Hummel, K. Richter, and P. Schlagheck, Genuine many-body quantum scars along unstable modes in Bose-Hubbard systems, [Phys. Rev. Lett. \*\*130\*\*, 250402 \(2023\)](#).
- [63] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, *et al.*, Controlling quantum many-body dynamics in driven Rydberg atom arrays, [Science \*\*371\*\*, 1355 \(2021\)](#).
- [64] S. Dooley, Robust quantum sensing in strongly interacting systems with many-body scars, [PRX Quantum \*\*2\*\*, 020330 \(2021\)](#).
- [65] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson, Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver, [Quantum \*\*7\*\*, 1184 \(2023\)](#).
- [66] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, Can error mitigation improve trainability of noisy variational quantum algorithms? [Quantum \*\*8\*\*, 1287 \(2024\)](#).