Automatic order detection and restoration through systematically improvable variational wave functions

Ryan Levy ⬤,[*] Miguel A. Morales ⬤, and Shiwei Zhang ⬤

Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

Variational wave function *Ansätze* are an invaluable tool to study the properties of strongly correlated systems. We propose such a wave function, based on the theory of auxiliary fields and combining aspects of auxiliary-field quantum Monte Carlo and modern variational optimization techniques including automatic differentiation. The resulting *Ansatz*, consisting of several slices of optimized projectors, is highly expressive and systematically improvable. We benchmark this form on the two-dimensional Hubbard model, using both cylindrical and large, fully periodic supercells. The computed ground-state energies are competitive with the best variational results. Moreover, the optimized wave functions predict the correct ground-state order with near full symmetry restoration (i.e., translation invariance) despite initial states with incorrect orders. The *Ansatz* can become a tool for local order prediction, leading to a new paradigm for variational studies of bulk systems. It can also be viewed as an approach to produce accurate and systematically improvable wave functions in a convenient form of nonorthogonal Slater determinants (e.g., for quantum chemistry) at polynomial computational cost.

I. INTRODUCTION

Understanding the ground-state properties of strongly correlated systems has been an important goal in the study of electronic systems. There has been much success in utilizing variational methods, such as variational Monte Carlo (VMC) [1,2] and density matrix renormalization group (DMRG) [3–5], however, each has drawbacks. Tensor network methods, while having incredible variational power, often have challenges of converging large width cylinders or periodic boundary conditions with finite resources [6]. In order to understand these large bulk systems, quantum Monte Carlo (QMC) methods [7,8] are invaluable, however, interesting regimes can be plagued with a sign problem, which renders many variational calculations exponentially difficult. Mitigation of this problem, for example using constrained path auxiliary field QMC (CP-AFQMC) [9], can provide accurate local energies but require more sophisticated methods, such as back propagation, to measure observables that do not commute with the Hamiltonian.

In VMC, a parameterized wave function *Ansatz* $|\psi\rangle$ is optimized against the variational energy expectation $\langle E \rangle = \langle \psi | \hat{H} | \psi \rangle / \langle \psi | \psi \rangle$. For many common *Ansätze*, this evaluation takes place by stochastic sampling via Monte Carlo in an occupation basis. The key to this approach is the interplay between expressivity of the *Ansatz*, and the ability to optimize

it to a global minimum. Often the limiting factor has been that the commonly used forms for fermionic systems lack sufficient expressivity to capture different quantum phases under the same wave function *Ansatz*, thereby making it difficult to detect orders in an unbiased manner.

Recently there has been a number of fermionic wave functions parametrizations introduced that are inspired from machine learning algorithms [10–13]. These are generally systematically improvable and have been shown to obtain very close agreement with the true ground state in small to medium-sized systems. However, in order to optimize vast numbers of parameters, the state may need to be constrained to certain physics, e.g., superconductivity or the wavelength of a charge density wave [12,14]. Furthermore, a wave function *Ansatz* with many parameters can make it more challenging to connect with the underlying physics. In strongly correlated systems with competing orders being separated by small energy differences, the variational energy can cease being an effective signal for detection of different phases. An *Ansatz* with a smaller number of parameters and a simpler, more physically inspired construction can be advantageous. We present a variational wave functions *Ansatz* based on the theory of auxiliary fields, which aims to have $\mathcal{O}(N^2)$ or less optimizable variational parameters and yields a simple form of the wave function that is straightforward to systematically improve. The *Ansatz* builds on a form originally introduced in Ref. [15] by Sorella, as a modified version of auxiliary field quantum Monte Carlo (AFQMC), and given the name variational AFQMC, or VAFQMC. We introduce a further generalized variational form, which defines a framework with which the expressivity can be easily expanded. An extension was recently presented to treat molecules in Ref. [16], with a different form than what we consider. We benchmark our variational wave function in the Hubbard model against existing

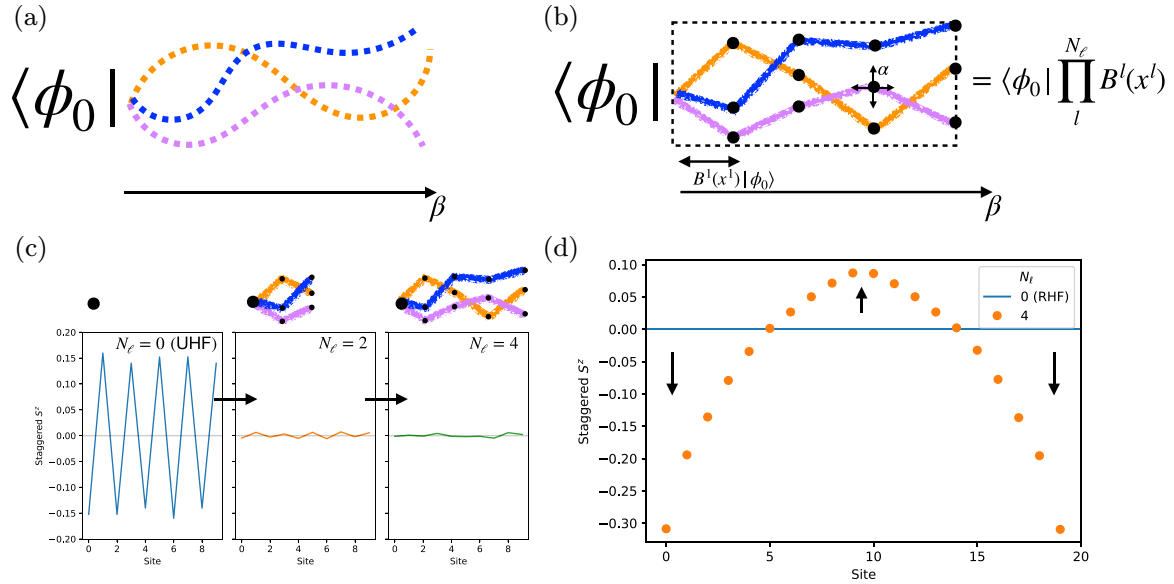---

*[*]rlevy@flatironinstitute.org

FIG. 1. (a) Illustration of standard AFQMC where random walks are taken with many small steps of size τ, following a predefined projector, to reach an imaginary time of β. (b) Illustration of VAFQMC random walk, where a number of potentially large steps are taken, following an effective projector that is variationally optimized, to reach the equivalent imaginary time of β. Each dot represents a projector slice and is a function of optimizable parameters α. (c) Staggered magnetization measurement for the third column of a 10×10 $U = 4$ Hubbard model with $n = 0.8$ density. The first pane shows the initial unrestricted Hartree-Fock (UHF) solution, and subsequent panes show VAFQMC slowly projecting out the incorrect initial order to create a spatially uniform ground state. (d) Staggered magnetization for a 20×4 Hubbard model at $U = 4$ and $n = 0.9$ density, with magnetic pinning fields applied at the edges. Starting from a qualitatively incorrect RHF initial state and optimizing the projection can uncover the correct ground-state local order.

VMC results, as well as DMRG and two projector QMC methods: AFQMC [9,17] and fixed-node diffusion Monte Carlo [14,18].

The rest of the paper is organized as follows. In Sec. II, we introduce our variational formulation and describe the implementation details. Then in Sec. III we benchmark against state-of-the-art VMC, projector QMC, and tensor network methods, showing competitive variational energies and predictive power of local observables. Finally in Sec. IV we conclude and discuss the outlook of VAFQMC in conjunction with other methods.

II. METHODS

We consider a variational form modeled after the projector method [19] such that the ground state $|\psi_0\rangle$ of a Hamiltonian H can be obtained via

$$|\psi_0\rangle = \lim_{N \to \infty} (e^{-\tau H})^N |\psi_T\rangle, \qquad (1)$$

for some trial state $|\psi_T\rangle$, which has nonzero overlap with the ground state. Traditionally, as in AFQMC, a stochastic random walk is formulated to project the trial wave function into the ground state via a Monte Carlo routine. A cartoon representation of this process is depicted in Fig. 1(a). Instead, by fixing the number of projection steps and optimizing the projectors as a wave function *Ansatz*, we can produce an integral representation of the variational state shown in Fig. 1(b).

Our formalism will apply to all Hamiltonians containing general one-body and two-body terms. For concreteness we will use the Hubbard model [20] in this work to both describe our approach and perform benchmark tests. The Hubbard

model is given by

$$\hat{H} = \hat{T} + \hat{V} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.} + U \sum_i n_{\uparrow i} n_{\downarrow i}, \quad (2)$$

where $\langle ij \rangle$ denotes nearest-neighbor hopping, U/t is the strength of the on-site interaction. We restrict the model to two dimensions such that $i = (i_x, i_y)$, and consider lattices of size $N_x \times N_y$. Our focus is on describing ground-state order and approaching the two-dimensional (2D) limit. For this we study both cylindrical geometry, with open boundary condition along x and periodic boundary condition (PBC) along y, and square lattices with fully PBCs along both direction. The cylindrical cells allow us to perform systematic benchmarks on the description of spin and charge density waves, the primary order in the ground state of this model. The fully periodic supercells allow us to test the quality of the wave function, including symmetry restoration, and scaling as we approach the 2D bulk limit.

To apply the projection operator $\exp(-\tau \hat{H})$ to the trial wave function, in AFQMC the propagator is split up into a kinetic and potential part $\exp(-\tau \hat{H}) \approx \exp(-\tau \hat{T}) \exp(-\tau \hat{V})$. This step incurs an error of $\mathcal{O}(\tau)$ and must be corrected, but this is not of concern here due to the variational nature of our wave function as discussed below. The kinetic term can be applied to the trial wave function as a transformation of the Slater determinant, via Thouless theorem [21]. For the potential term, a Hubbard-Stratonovich (HS) decomposition can be applied [22,23], for example the discrete version

$$e^{-\tau U n_{\uparrow i} n_{\downarrow i}} \propto \sum_{x_i = \pm 1} \exp[\lambda_i x_i (n_{\uparrow i} - n_{\downarrow i})], \quad (3)$$

where $\cosh \lambda_i = \exp(U\tau)$.

The form of our variational wave function can be thought of as a sum over auxiliary fields:

$$|\psi\rangle = \sum_{x^l} \prod_l^{N_\ell} e^{\sum_i \lambda_i x_i^l (n_{\uparrow i} - n_{\downarrow i})} e^{-\sum_\sigma \tau_\sigma^l T_\sigma^l} |\phi_0\rangle, \qquad (4)$$

where $T_\sigma^l = t_{ij}^{l\sigma} c_{\sigma i}^\dagger c_{\sigma j}$ and constant terms are ignored.

This can be recast into an effective Hamiltonian form as

$$|\psi\rangle = \prod_l^{N_\ell} e^{-\tau_l \hat{\mathcal{H}}_l} |\phi_0\rangle, \qquad (5)$$

where $\hat{\mathcal{H}}_l$ is determined by optimizable parameters T_σ^l, τ_l, as well as the form of the HS, all of which can be different from that of the actual Hamiltonian H. In other words, we consider variational wave functions of the form

$$|\psi\rangle = \int \prod_l^{N_\ell} B^l(x^l) \, dx^l \, |\phi_0\rangle = \sum_{\{x\}} |x\rangle, \qquad (6)$$

where $B^l(x^l)$ is a one-body propagator of the form $B^l(x^l; \alpha) = \exp[\sum_{ij} \sum_{\sigma\sigma'} h_{i\sigma;j\sigma'}^l (x_{ij}^l; \{\alpha\}) c_{i,\sigma}^\dagger c_{j,\sigma'}]$, with i denoting lattice sites (or basis functions), $|x\rangle$ is the resulting application of $\prod B^l(x)|\phi_0\rangle$, and σ (σ') denoting spins. The effective one-body Hamiltonian in the exponent in B^l contains optimization parameters, denoted by α. For the Hubbard model example we consider here, $B^l(x^l)$ is of the simpler, product form in Eq. (4), with α consisting of $\alpha = \{\lambda_i, \tau_\sigma^l, t_{ij}^{l\sigma}\}$. Here we have separated h^l in the one-body auxiliary-field-independent term into an overall scaling factor τ^l and the hopping terms $T_{i\sigma;j\sigma}^l$, so that no overall scaling factor is involved in the optimization of the latter. These parameters are unconstrained: λ_i is repeated for each projector slice l but can be nonuniform over sites i; τ_σ^l is now distinct from λ_i and is allowed to be positive or negative; and T^l is no longer constrained to be Hermitian. In practice the values of these parameters are optimized directly. Most of these parameters are in T_σ^l, with $\mathcal{O}(N_\ell \cdot N^2)$ parameters. This general variational *Ansatz*, which is restricted to one-body terms, can also be related to shadow wave functions where a kernel of both physical and auxiliary degrees of freedom is parameterized and auxiliary variables are integrated out [24,25]. Shadow wave functions, however, have historically been used in a particle configuration basis unlike Eq. (6).

Unlike Ref. [15], our variational form has no fixed projection time $\beta^{N_\ell} = \sum_l^{N_\ell} (\tau_\uparrow^l + \tau_\downarrow^l)/2$. This allows the *Ansatz* to grow as needed in projection time, to potentially reach the ground state with few effective projector steps N_ℓ. In fact, completely decoupling τ^l from λ_i allows for not only cooling, i.e., $\beta^{l+1} = \beta^l + \tau$, but intermediate heating, i.e., $\beta^{l+1} = \beta^l - \tau$. We show below that the optimized results indeed takes advantage of this structure.

Given a variational form of Eq. (4) or generally Eq. (6), we can evaluate the energy via a double integral

$$\langle E \rangle = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \qquad (7)$$

$$= \frac{\sum_{xx'} \rho(x, x') E_L(x, x') S(x, x')}{\sum_{xx'} \rho(x, x') S(x, x')} \qquad (8)$$

$$= \frac{\langle E_L S \rangle_\rho}{\langle S \rangle_\rho}, \qquad (9)$$

where each x (or x') is a field configuration that produces a Slater determinant $|x\rangle$ ($|x'\rangle$), $E_L(x, x') = \langle x' | \hat{H} | x \rangle / \langle x' | x \rangle$, which is similar to a local energy, $S(x, x') = \langle x' | x \rangle / |\langle x' | x \rangle|$ is the phase or sign of the configuration, and $\rho(x, x') = |\langle x' | x \rangle|$. The probability distribution $|\langle x' | x \rangle|$ can be sampled using Markov chain Monte Carlo (MCMC) through conventional Metropolis algorithm methods [8] where each state is a configuration of fields (x, x'). This procedure is distinctly different from VMC, however, and is instead closer to the evaluation of a N_ℓ-slice path integral in imaginary time. Note that because $\langle x' | x \rangle$ is not necessarily positive semidefinite a sign problem can occur. However, for the problems we consider, the number of slices, N_l, is modest and the average $\langle S \rangle$ remains positive as further discussed below.

Energy optimization is carried out through gradient descent. The derivation of the derivative is provided in Ref. [15], which we repeat here for convenience. The derivative of of parameter α is given by

$$\frac{\partial \langle \hat{H} \rangle_\rho}{\partial \alpha} = \frac{\left\langle S\left(\frac{\partial E_L}{\partial \alpha} + (E_L - \langle \hat{H} \rangle) O\right)\right\rangle_\rho}{\langle S \rangle_\rho}, \qquad (10)$$

where $O = \frac{\partial \ln(S\rho)}{\partial \alpha}$. Both $\frac{\partial E_L}{\partial \alpha}$ and O can be easily evaluated using automatic differentiation. By sampling ρ and accumulating the values in Eq. (10), optimization of the wave function can be performed. This optimization procedure of VAFQMC is therefore dominated by both matrix multiplication (the projection of $|\phi_0\rangle$) and determinant calculations for the local energy, which implies VAFQMC scales as $\approx O(N_\ell \cdot N^3)$. We note that standard practices in AFQMC, for example, compact decomposition of long-range interaction (or Jastrow), the use of force bias [26] or other update schemes, stabilization, low-rank decomposition, fast computations of local energy, etc. [27], can all be applied to this method.

A. Implementation details

We consider the standard Hubbard model with only nearest-neighbor hopping. Two types of calculations are considered, using two different boundary conditions: fully periodic (PBC) and cylindrical boundary conditions (CBC). When using cylindrical boundaries, we apply a weak edge pinning field $H_{\mathrm{pin}} = \sum_{i \in \mathrm{edge}} (-1)^{i_x + i_y} h \hat{S}_i^z$ where $\hat{S}_i^z = (\hat{n}_{\uparrow i} - \hat{n}_{\downarrow i})/2$ to induce a local antiferromagnetic order, where h controls the strength of pinning and is chosen as $h = 1/2$ unless otherwise specified. When pinning is present, energies are reported including the contribution from the additional field, consistent with other publications [28].

Our implementation utilizes Jax [29] to perform automatic differentiation and sampling, alongside the OPTAX package [30] for optimization. The matrix exponential is computed approximately using a $k = 3$ product formula [31]. Optimization generally uses around 180 optimization steps with varying and decaying step size (see Appendix); gradients are accumulated with 189 000 samples and parameters are updated using a random step with the computed sign [11].

This work focuses on the case of $N_\ell = 4$, which gives substantial improvement over, e.g., $N_\ell = 2$ or a single slice.
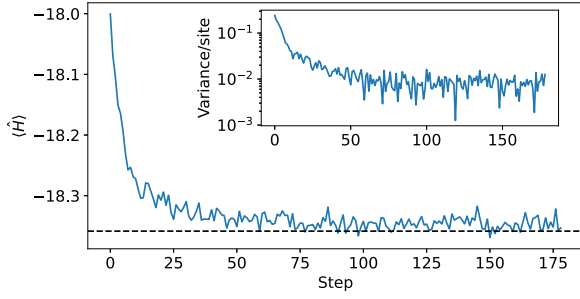
FIG. 2. Variational energy $\langle \hat{H} \rangle$ and variance per site $(\langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2)/N$ (inset) during optimization of a 4×4 Hubbard model at $U = 6$ and $n = 0.625$ using VAFQMC with $N_\ell = 4$. The ground state from exact diagonalization is shown in a black dashed line. Fluctuations are due to the limited sampling during optimization.

While any given calculation can always be reoptimized with additional slices, by fixing N_ℓ we can observe the limits of this form.

We can additionally add a term to optimize the sign similar to Ref. [16], however, in the applications presented here the average sign is almost always very large $\langle S \rangle \geqslant 0.93$. For $U = 8$, the sign can get as low as $\langle S \rangle \approx 0.74$ but this is sufficiently large to resolve observables and derivatives without needing any further constraint in the optimization. We renormalize (i.e., clip) both the sign term derivative and the energy derivative term to be on the same order. An example of the optimization is shown in Fig. 2.

After optimization, parameters are selected based on the lowest observed average energy. Using this parameter set, observables such as the local energy and electron occupation are measured with approximately 10^6 samples with a thermalization of 10^3 sweeps. In order to properly treat the $1/\langle S \rangle$ term, error bars are reported from a jackknife estimate [8].

DMRG results were computed using the ITENSOR library [32,33] with a maximum bond dimension of $m = 15360$. Constrained-path AFQMC results are obtained using

protocols similar to those in Refs. [28,34–36] including self-consistently optimized constraints [37].

III. RESULTS

A. Energetic benchmarks

To test the variational *Ansatz*, we first benchmark against doped large square Hubbard models with PBC as shown in Table I. Some systems at intermediate doping have a reference value from AFQMC [17], which is often effectively exact; we compare to AFQMC the following systems: 6×6 with $n = 2/3$, 8×8 with $n = 0.6875$, and 10×10 with $n = 0.8$ at $U/t = 4$. The VAFQMC *Ansatz* is able to obtain an energy error that is $O(10^{-3})$ with an energy variance per site $(\langle H^2 \rangle - \langle H \rangle^2)/N$ of less than 1, which improves over traditional variational approaches in such systems.

For systems without exact energies, we compare with several state-of-the-art methods, projector and also variational QMC methods. Note that of the projector methods, AFQMC and fixed-node projector diffusion Monte Carlo (DMC) with a VMC optimized trial wave function, only fixed node is variational. Then we compare with VMC energy and energy variance. Additionally note that the fixed-node DMC energies of Ref. [14] are higher than those of Ref. [18]; we report the lowest energy available.

Overall, the variational energy of the VAFQC *Ansatz* is competitive or improves upon those from state-of-the-art VMC under the standard formalism of Slater determinants (or antisymmetrized germinal powers), backflow, and Jastrow factors. For the 8×8 periodic system, at $U = 4$, VAFQMC is lower in energy across densities and, at $U = 8$, E/N from VAFQMC is approximately $0.004t$ lower at $n = 0.78125$ and $0.01t$ higher at $n = 0.875$. More impressively, at $n = 0.78125$ we obtain nearly the same energies for $U = 4$ and $U = 8$ (with $U = 8$ having worse performance) to a very sophisticated VMC wave function, enforcing symmetries such as SU(2) and $k = 0$ momentum projection, augmented by additional Lanczos steps. Likewise when comparing to fixed-node DMC energies across system sizes and densities, we find

TABLE I. Energy comparison between AFQMC [17,18], fixed-node DMC [18], VMC [14], and VAFQMC (this work), on a PBC square Hubbard model with density n and interaction strength U (in units of t). Energies with a * are effectively exact from release constraint calculations of AFQMC [17], and energies with † include SU(2) and momentum projection augmented with a Lanczos step [14]. Additionally shown is the variance $\sigma^2 = \langle H^2 \rangle - \langle H \rangle^2$ of the energy, which can be computed more conveniently in variational methods and which vanishes ($\sigma^2 = 0$) for a true eigenstate of the Hamiltonian. VAFQMC results are optimized with $N_\ell = 4$ projector slices.

| $N_x \times N_y$	6×6		8×8				10×10	
n	2/3	0.6875	0.78125		0.875		0.8	
U	4	4	4	8	4	8	4	8
Projector Methods								
E_{AFQMC}/N	$-1.18525(4)$*	$-1.1858(2)$*	$-1.13253(3)$	$-0.9252(1)$	$-1.01923(6)$	$-0.7616(1)$	$-1.1135(2)$*	$-0.8960(1)$
E_{FN}/N			$-1.128396(9)$	$-0.913753(9)$	$-1.01153(1)$	$-0.74941(2)$	$-1.10934(1)$	$-0.88291(1)$
Variational Methods								
E_{VMC}/N			$-1.12683(1)/$	$-0.91011(2)/$	$-1.00579(1)$	$-0.74220(3)$		
			$-1.13037(1)^\dagger$	$-0.91482(1)^\dagger$				
E_{VAFQMC}/N	$-1.1834(1)$	$-1.18419(7)$	$-1.13048(8)$	$-0.9148(1)$	$-1.01391(8)$	$-0.7469(2)$	$-1.11097(6)$	$-0.8819(1)$
σ^2_{VMC}			$1.43/0.59(2)^\dagger$	$3.46 / 2.58(1)^\dagger$	2.79	6.62		
$\sigma^2_{\mathrm{VAFQMC}}$	0.2(1)	0.3(3)	0.5(3)	2.4(4)	0.6(3)	3.5(5)	0.8(6)	3.9(8)
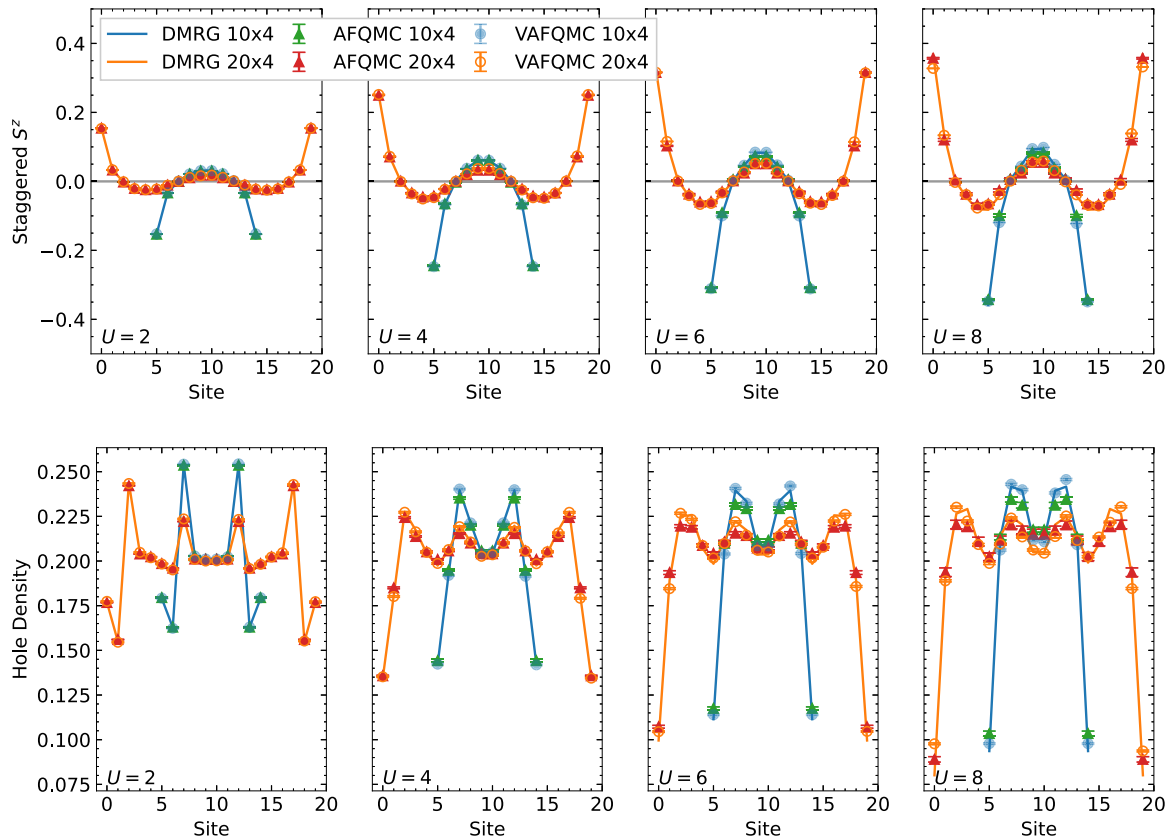
FIG. 3. Comparison of staggered S^z $\langle (-1)^{i_x+i_y} S^z(i_x, i_y) \rangle$ (top) and hole density $\langle 1 - n(i_x, i_y) \rangle$ (bottom) between DMRG, VAFQMC, and AFQMC for a Hubbard model on a cylinder with edge pinning at a series of interaction strengths $U = 2, 4, 6, 8$ and $n = 0.8$ density. Results have been averaged over the width-4 column direction. VAFQMC results are with $N_\ell = 4$ projector slices.

similar or slightly lower energies for $U = 4$ and competitive energies for $U = 8$. This is particularly favorable as VAFQMC can conveniently compute both local and nonlocal observables and correlation functions, which can be challenging or inaccessible within fixed-node DMC and which require back propagation [9] in AFQMC.

We also note for the, e.g., 10×10 $U = 4$ point, the local magnetization $\langle S^z \rangle$ is $< 10^{-2}$ on each site, recovering an unpolarized spin solution despite the initial parameters breaking spin symmetry explicitly. This result comes at a small cost, however, and lower-energy results can be obtained when starting from a UHF trial wave function [e.g., Fig. 1(c)]. Using a UHF initial state $|\phi_0\rangle$ with $U_{\text{UHF}} = 0.5$ (the same as used by AFQMC for its trial wave function), the local magnetization is $\langle S_i^z \rangle \sim O(10^{-2})$ but about $0.05t$ lower in energy. This is not unique to VAFQMC but rather to variational wave functions generally, where local minima of symmetry broken states can be hard to optimize away. In VAFQMC this manifests as requiring more projector slices N_ℓ to modify a strong local magnetization or density in the trial state, compared to a spin-restricted initial *Ansatz*, and thus we suggest utilizing the restricted form to minimize total parameters.

B. Local order prediction and automatic symmetry restoration

As a proxy for phase diagram exploration, we use VAFQMC to explore the local order of cylindrical Hubbard models with antiferromagnetic edge pinning. This model was

used to study the stripe behavior of the ground state [28,38]. In Ref. [28], the stability of the stripe state was determined by comparing the staggered spin and hole densities at different interaction strength U and electron density n as a function of system size. We perform calculations similar to those, but first with small enough system sizes that DMRG can provide quasiexact reference results. We vary the electron density n, or equivalently hole doping of $\delta \equiv 1 - n$, and consider cylinders of width $r/\delta \times N_y$ for some integer r.

With different interaction strengths and system sizes, we find remarkable agreement in the magnetic order between VAFQMC, DMRG, and AFQMC in the top of Fig. 3. The variational *Ansatz* for VAFQMC is completely unconstrained by symmetries, yet can nearly recover both translational symmetry in the y direction and symmetry across the middle of the cylinder. The corresponding hole densities are shown in bottom of Fig. 3. We again see almost complete recovery of the symmetry of the system. There is small disagreement between DMRG and AFQMC as to the hole densities at large U. As DMRG is expected to be nearly exact in these width-4 systems, the discrepancy is likely an indication of small residual bias from the constraint in AFQMC [37]. Remarkably our VAFQMC results are nearly indistinguishable from the DMRG results. This extends to other dopings across similar values of U as well (see Appendix), and illustrates the predictive power of the VAFQMC method. At larger values of U, the variational energy from VAFQMC is 1–2% higher than DMRG and AFQMC. These results are a reminder that
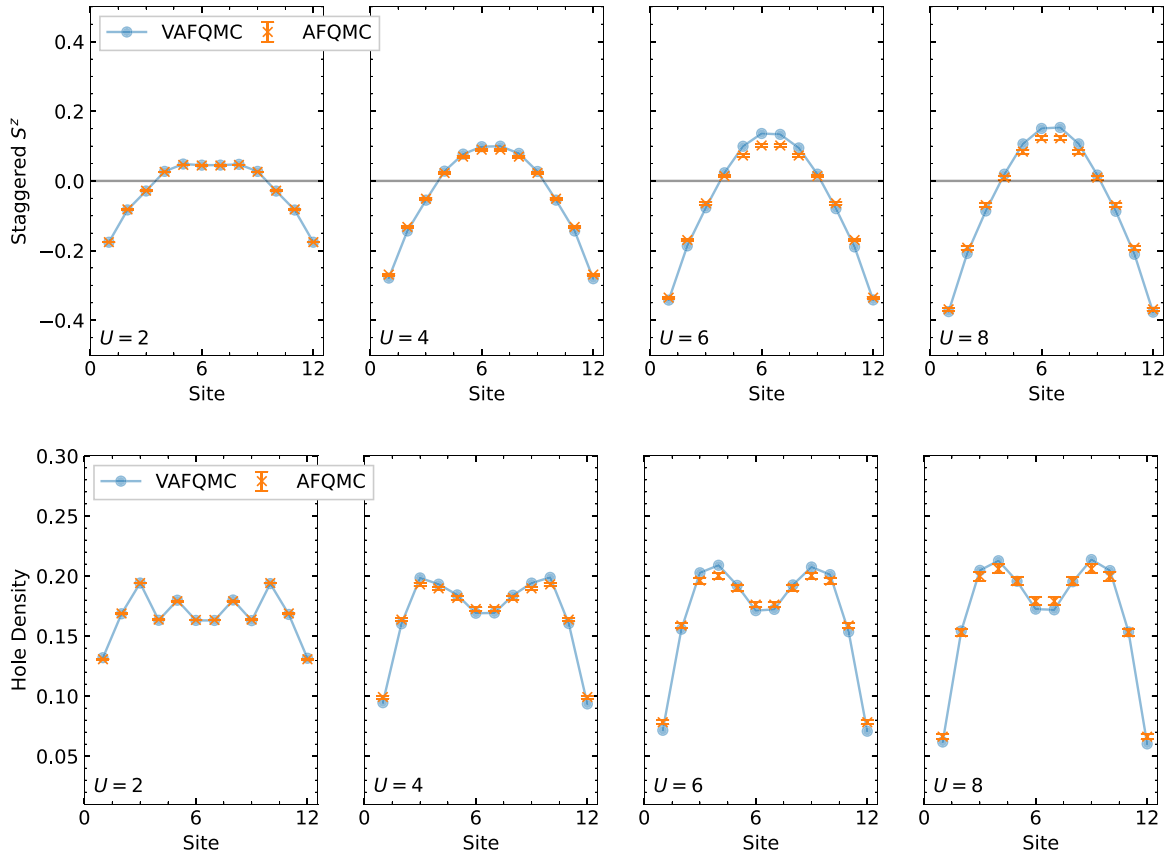
FIG. 4. Comparison of staggered S^z $\langle(-1)^{i_x+i_y}S^z(i_x, i_y)\rangle$ (top) and hole density $\langle 1 - n(i_x, i_y)\rangle$ (bottom) between VAFQMC and AFQMC for a 12×8 Hubbard model on a cylinder with edge pinning at a series of interaction strengths $U = 2, 4, 6, 8$ and $n = 5/6$ density. Results have been averaged over the width-8 column direction. VAFQMC results are with $N_\ell = 4$ projector slices.

the total energy, while very important for a variational *Ansatz*, should not be unduly emphasized, and order parameters and correlation functions often provide a more stringent measure of the predictive power of an *Ansatz*.

We next move beyond width-4 cylinders and consider wider systems, which are important [28,34] to access the thermodynamic limit. In Fig. 4 a comparison of staggered S^z and hole density between AFQMC and VAFQMC is shown for a 12×8 cylinder with $n = 5/6$ density and $U = 2, 4, 6, 8$. Exceptionally, VAFQMC continues to obtain the correct order and stripe wavelength, with reasonable quantitative agreement with AFQMC. Combined the results for larger periodic cells discussed in the previous section, these results show the promise of the VAFQMC *Ansatz* as a predictive approach for extended two-dimensional systems and beyond.

As mentioned, the projection times are unconstrained as full variational parameters in our approach. Unlike the *Ansatz* in Ref. [15], a number of solutions have a negative τ^l value within one of the projector slices, as shown in the left of Fig. 5, similar to that of higher order Trotter expansions [39,40]. As both $T_{ij}^{\sigma l}$ and τ^l can change, we renormalize the τ^l values by the average value of nearest-neighbor hopping terms $\langle T_{\langle ij\rangle}^{\sigma l}\rangle$, which is almost always $O(-t)$. By only considering τ^l values, we can see that the effect of each time slice is not equivalent over a given U. As U increases, the value of τ^l tends to fluctuate between positive and negative. We believe this is to effectively raise the temperature and broaden

the trial wave function in order to (in the next slice) project into the correct manifold. For width-4 systems, τ^l values are around ≈ 0.13 while width-8 systems they are slightly larger at 0.15 on average, but both width cylinders have slices l where $\tau_\sigma^l \approx 0.25$ for example. Width-4 systems also tend to have larger negative τ_σ^l, with $\min_{l,\sigma} \tau_\sigma^l \sim -0.2$ for width-4 cylinders and $\min_{l,\sigma} \tau_\sigma^l \sim -0.07$ for width-8. We emphasize that while some values of τ^l are quite small, the interaction term remains $O(10^{-1})$ and thus the *Ansatz* is still not reduced to fewer slices.

To examine the role that the negative τ values play, in the right of Fig. 5 we plot the $\langle E^l\rangle$ for the optimized four-slice VAFQMC *Ansatz* truncated to l ($l \leqslant 4$) time slices. (This is to be distinguished from a stand-alone l slice or $N_\ell = l$ wave function. In the latter the parameters would be optimized for l slices, while here they are optimized for the entire four-slice wave function.) We subtract the energy of that of the $|\phi_0\rangle$ state (denoted RHF) and rescale by $1/N$. We see that for the first ($l = 1$) and last ($l = 4$) slice the energy always decreases. At intermediate values of l there may be energy values that are significantly worse than the previous time slice, unlike the near monotonic decrease in a QMC projector method. For example consider the 12×4 system at $U = 4$, the energy per site at an intermediate point $l = 2 \to 3$ jumps by $\Delta E/N \sim +0.17$, then lowers in the final slice ($l = 3 \to 4$) by $\Delta E/N \sim -0.2$. The large value of τ^4 at the end of the *Ansatz* appears to be linked to this substantial decrease in
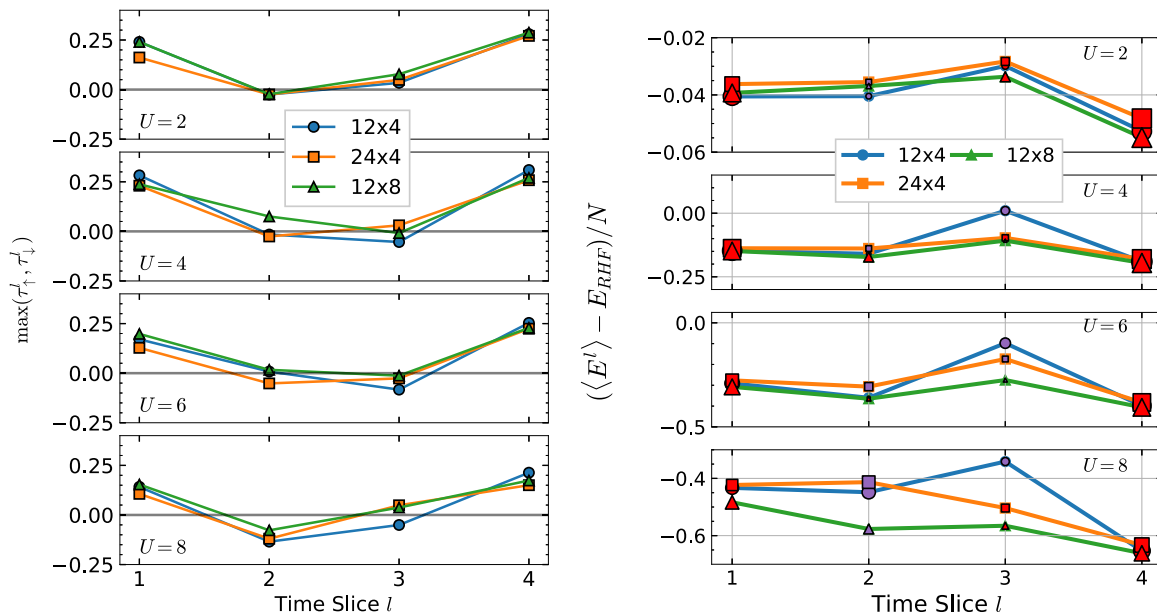
FIG. 5. Left: Optimized values for τ^l maximized over spin $[\tau^l = \max(\tau^l_\uparrow, \tau^l_\downarrow)]$ for the Hubbard model in cylindrical geometry with edge pinning, shown for four values of U and three system sizes. Values have been renormalized by $\langle T^{\sigma l}_{(ij)} \rangle$. The electron density is $n = 5/6$. Right: Energy per site (relative to that from $|\psi_{\rm RHF}\rangle = |\phi_0\rangle$) at a given time slice l (in units of t) for the systems shown on the left. Note that the energy at each l is evaluated with the final parameters fixed at the optimized values for $l = 4$. Marker size is proportional to the renormalized $\tau^l = \max(\tau^l_\uparrow, \tau^l_\downarrow)$ with red denoting positive (equivalent to cooling projection) and purple denoting negative (equivalent to heating projection). Error bars are smaller than the marker size.

energy, to compensate for the heating and cooling of the path.

For lower values of U, often a given slice may have a very small value of τ^l, suggesting that an *Ansatz* with fewer N_ℓ may suffice to capture the dominant physics. This is further suggested by the intermediate values of the energy $\langle E^l \rangle$, which only occupies a range of ~ 0.04/site. We conjecture that the flexibility in $\{\tau^l\}$ can improve the size extensivity of the *Ansatz*. The *Ansatz* can both increase the overall effective β, similar to Ref. [15], to produce a size extensive *Ansatz*, or may be something more complex within the path. The new *Ansatz* based on an effective Hamiltonian given in Eq. (5) is the key for allowing better size consistency with a fixed N_ℓ. More systematic studies on this will be very valuable.

IV. DISCUSSION AND SUMMARY

We presented a variational approach for many-electron systems with an expressive *Ansatz* of $\mathcal{O}(N_\ell \cdot N^2)$ parameters. Although inspired by the idea of an approximate realization of the imaginary-time projection in AFQMC, the *Ansatz* is much more general. The approach shares the idea of variational optimization of a small number of projections with AFQMC-like time slices as in Ref. [15], and we have retained the name VAFQMC from there. However, our approach takes a different philosophy of viewing the *Ansatz* as a variational effective Hamiltonian rather than a variation projection using the physical Hamiltonian. Instead of retaining the form of the Hamiltonian $\hat{H}$ in the projection (e.g., keeping $\hat{T}$ and the basic form of the HS transformation dictated by $\hat{V}$) and creating different orders via $\phi_0\rangle$, we view the *Ansatz* as seeking the most

effective projection (to the true ground state) via variational optimization.

We have emphasized the underlying theory of optimizing a set of effective one-body Hamiltonians or actions defined by auxiliary fields. This wave function, even in our still rudimentary implementation, is already competitive with state of the art VMC and even fixed-node Green's function Monte Carlo. We have demonstrated this on large square Hubbard models, where the approach can systematically obtain the ground-state energy to within or better than 1–2% of the exact results. Furthermore, it nearly restores symmetry automatically via its optimization process and provides accurate physical observables, which allow predictive resolution of magnetic and charge correlations and orders. Broken, e.g., $SU(2)$ symmetry from pinning fields can be additionally restored by hand and used in the original model.

Historically, many VMC wave function *Ansätze* were built around a specific mean-field state (for example in Ref. [15]), or an explicit form that targets a particular type of order. Under this approach, optimization is constrained to the fixed state type, and prediction of the actual order is made by comparing the relative energetic behaviors of two or more such fixed types of states. Despite its considerable success, this approach lacks true predictive power in many situations, either because the order is not known, or because the different types being compared are not well balanced in the *Ansatz* (e.g., one with more variational flexibility than the other). With the onset of neural quantum states (e.g., Refs. [11,12,41]), wave functions are parameterized with partially or fully unconstrained variational freedom to improve expressibility. Likewise, the aim of our VAFQMC approach is to have an *Ansatz*, which is similarly more flexible and more expressive such that local order
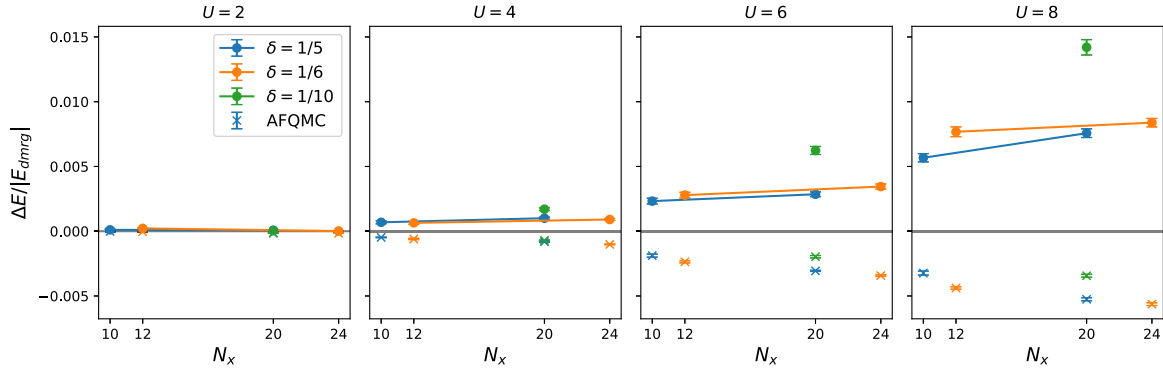
FIG. 6. Relative energy $(\langle E \rangle - E_{\mathrm{DMRG}})/E_{\mathrm{DMRG}}$ comparison for VAFQMC and CP-AFQMC compared to DMRG for various Hubbard models of size $N_x \times 4$ on a cylinder with edge pinning and hole doping δ.

can instead be predicted from optimization. As our examples have shown, with large optimization freedom VAFQMC can function as a predictive *Ansatz* in addition to providing good variational energies in a challenging system operating at realistic parameters and under real-life conditions. It is also notable that this is achieved without the use of neural networks or tensor networks, bringing a new class of unconstrained systemically improvable variational wave functions.

Optimization of an expressive *Ansatz* alone is sometimes still not enough to resolve the hardest problems. For models in which there are many competitive states, e.g., stripes or intertwined orders, the application of pinning fields has become a very useful and essential tool to elucidate the underlying order. The success with which this paradigm has been used [28,34,35,38,42], and this benchmark study of VAFQMC suggests that this is a potential paradigm to better solve problems inaccessible to 1D tensor networks. This should prove to be an invaluable method in the toolbox of computational many-body physics.

Despite the inexpensive *Ansatz* presented here with fixed N_ℓ, we emphasize that VAFQMC can still be systematically improved. One can continue to add more projector slices to potentially improve the energy, but after some point computational costs will increase (and further care must be taken to properly stabilize the long time projection of the trial state). At $N_\ell = 4$ this was functionally unnecessary for the systems we considered. Neural networks can be added within the *Ansatz*, for example as in Ref. [16], to further improve expressibility, particularly in the strongly correlated regime where correlations between auxiliary fields may be useful. Additional considerations may be needed to improve scalability, however. Likewise while the optimization procedure was able to regain translation invariance and minimal local spin symmetry breaking, this should not be relied upon when scaling. Applying symmetry projection [43,44], either within $|\phi_0\rangle$ or for each projected $|x\rangle$ state, is a compelling future direction.

Given the promise the *Ansatz* has already shown at this very early stage of development, we believe the method presents many exciting and interesting new possibilities. In addition to the directions discussed above, there are a number of issues to be addressed or better understood. For example, optimizing to machine precision is currently intractable due to an infinite variance problem. Because the local energies do not have a zero variance principle, sampling can be dominated

by rare events and thus are unable to reduce the statistical noise with more samples. We have checked our results on small systems using the methods of Ref. [45] (see Appendix), however, there may be a better methods to be explored in future work.

APPENDIX A: FURTHER COMPUTATIONAL DETAILS

The step sizes during the first 79 steps of optimization is 0.01 and then is lowered to 0.001 for the remaining 100 steps. The optimal point is selected by the lowest energy during the optimization, normally somewhere after the first 80 steps.

Optimization was performed with 126 independent walkers contributing to the total sample collection, while the additional measurement of the optimized parameters used 120 walkers. Each measurement was collected after $N \cdot N_\ell$ Metropolis updates and values reported in the main text were averaged over walkers and a 1000 sweep thermalization time was used.

Finally the initial values were chosen to correspond to a $\tau_\ell = 0.1$ projection of AFQMC, with a perturbation of $\delta \in [0, 10^{-3})$ for λ and T_σ^ℓ terms. Note that when there is edge pinning, we do not include pinning terms in T_σ^ℓ and $|\phi_0\rangle$ but do include the appropriate cylindrical boundary conditions. Variance measurement was performed with an additional $10\times$ samples.

APPENDIX B: ADDITIONAL HUBBARD CYLINDER COMPARISON

We provide both additional data from the cylindrical Hubbard model systems with edge pinning studied above along with data for $\delta = 1/10$, which is shown in Ref. [28]. This data is again compared to those from DMRG and self-consistent CP-AFQMC in width-4 cylindrical systems.

In Fig. 6, the energy of VAFQMC and AFQMC is compared to that of the obtained DMRG. Note that the DMRG results are not extrapolated in cutoff, and thus may be slightly
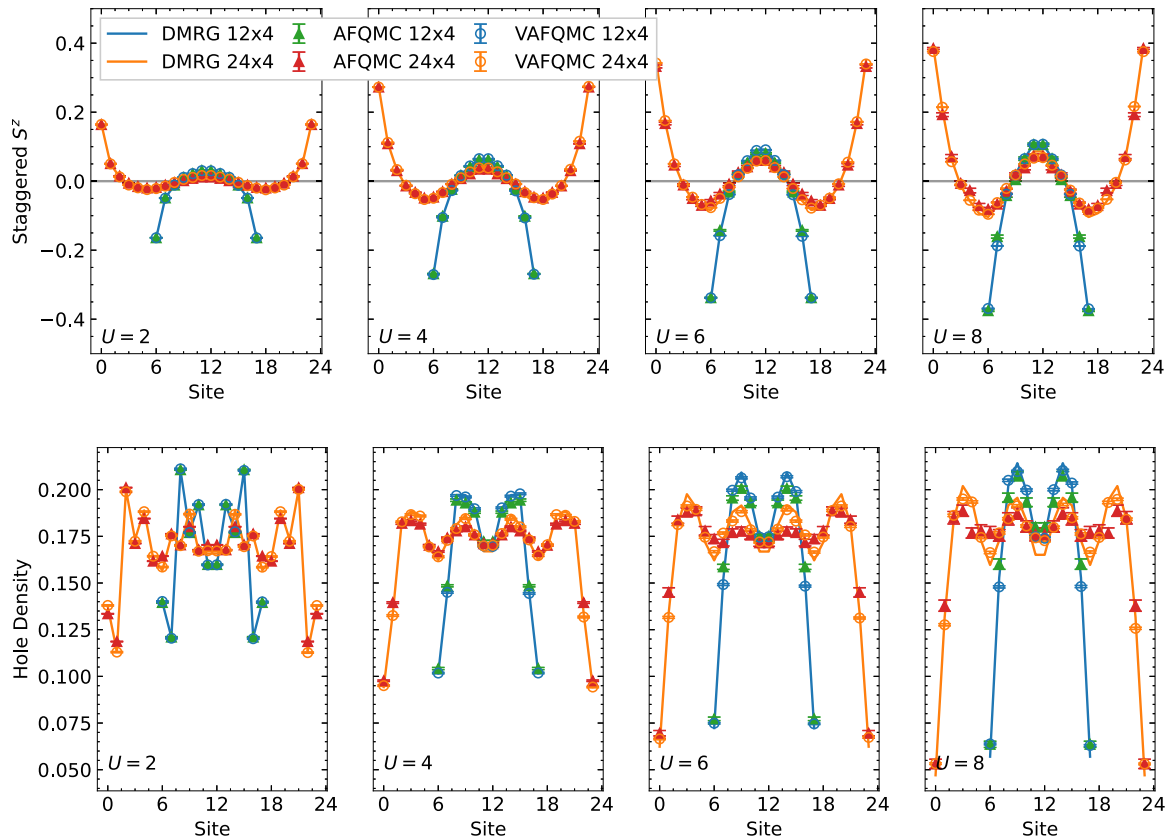
FIG. 7. Comparison of staggered S^z $\langle(-1)^{i_x+i_y}S^z(i_x,i_y)\rangle$ (top) and hole density $\langle 1 - n(i_x,i_y)\rangle$ (bottom) between DMRG, VAFQMC, and AFQMC for a Hubbard model on a cylinder with edge pinning at a series of interaction strengths $U = 2, 4, 6, 8$ and $n = 5/6$ density. Results have been averaged over the width-4 column direction. VAFQMC results are with $N_\ell = 4$ projector slices.

higher in energy to the exact ground-state energy, which is unknown. At $U = 2$, all methods are in agreement for the energy within 10^{-4} relative difference. For larger values of U, AFQMC results trend lower in energy as U gets larger while VAFQMC results trend higher. The largest discrepancies occur at $U = 8$ where VAFQMC is approximately 1–1.5% from DMRG while AFQMC is <1% lower (relatively).

We show further local observables comparisons, of average staggered S^z and hole density, at $n = 5/6$ for 12×4 and 24×4 system sizes in Fig. 7. The VAFQMC results of spin and hole densities mirror those at in Fig. 3 are seen to be in excellent agreement with those from DMRG. While staggered spin densities agree well between the three methods, VAFQMC and DMRG agree on the magnetic order, the charge order has slight discrepancies where VAFQMC and DMRG more closely agree than with AFQMC.

APPENDIX C: INFINITE VARIANCE PROBLEM

The estimator of the variational energy $\langle E \rangle = \langle E_L S \rangle / \langle S \rangle$ can suffer from an infinite variance problem [45], as $\langle x'|x \rangle$ approaches zero. In order to consider the effect on our results, we use the bridge link mitigation of Ref. [45]. A new sampling probability distribution is introduced by inserting an additional (symmetric) time slice in the center of the path, i.e.,

$$\rho'(x, x') = |\langle x'|e^{-\tau H}|x\rangle|. \tag{C1}$$

The projector introduces an additional set of N auxiliary fields to integrate over, and modifies the energy calculation to be

$$\langle E \rangle = \frac{\langle \psi|\hat{H}|\psi\rangle}{\langle \psi|\psi\rangle} \tag{C2}$$

$$= \frac{\langle E_L \, \rho/F \rangle_{\rho'}}{\langle \rho/F \rangle_{\rho'}}, \tag{C3}$$

where $F = F(\tau) = |\langle x'|e^{-\tau\hat{H}}|x\rangle|$ and $\rho = \langle x'|x \rangle$, which has a sign unlike previously. The calculation of F must be done with care, however, so as to not introduce a estimator bias (see Ref. [46]).

In order to consider the effects of the infinite variance, we study a closed-shell configuration ($n = 0.625$) of the 4×4 Hubbard model with PBC at various values of U. The calculation of F is done exactly with a bridge link of $\tau = 0.005$, summing the 2^{16} configurations for each field configuration of x, x'. In Fig. 8 we show the energy for both the standard estimator of Eq. (9) during and after VAFQMC optimization, and with included bridge link Eq. (C3). Optimized energies are often below the exact energy in part due to small sample sizes. These are then raised (potentially above the ground state) with increased measurement, and subsequently the inclusion of the bridge link shifts the energy slightly slower but still within error bars of the original measurement estimate.
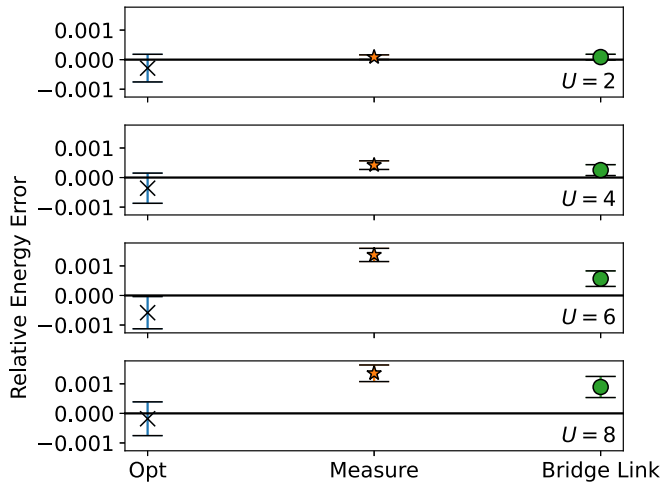
FIG. 8. Relative energy error $(\langle E \rangle - E_0)/E_0$ between VAFQMC and the exact E_0 from exact diagonalization of a 4×4 Hubbard model with PBC at $n = 0.625$. Blue x's ("Opt"), are the energies reported from optimization, orange stars ("Measure") are measured with methods equivalent to Sec. II, and green circles ("Bridge Link") represent bridge link calculations where $F(\tau = 0.005)$ is summed exactly.

APPENDIX D: ADDITIONAL τ^l DATA

In Fig. 9, we present additional τ^l values for square Hubbard models with PBC. The values of τ^l are also
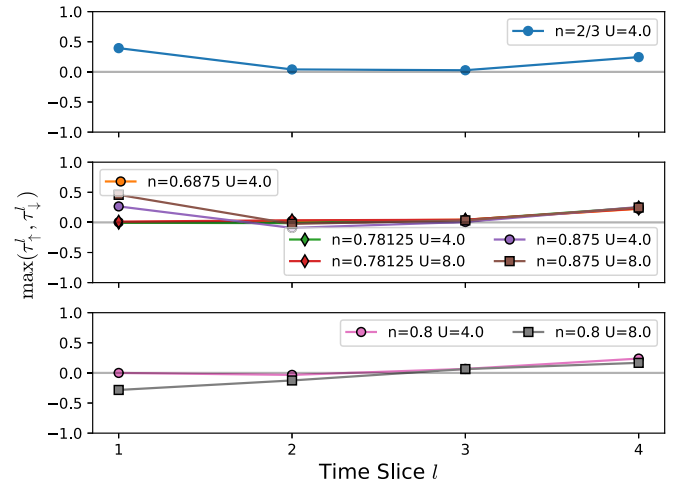

FIG. 9. Optimized values for the maximum τ^l (maximized over spin) for 2D Hubbard models of size (top to bottom) 6×6, 8×8, and 10×10 with PBC at U and n values as shown. Values have been renormalized by $\langle T_{\langle ij \rangle}^{\sigma l} \rangle$.

renormalized by the optimized $\langle T_{\langle ij \rangle}^{\sigma l} \rangle$, which are generally $O(-t)$. For closed shell systems at $U = 4$ (6×6 with $n = 2/3$, 8×8 with $n = 0.6875$), we see similar results to those of cylindrical pinned systems at low U, suggesting that similar results could be obtained with less slices N_ℓ. On average these τ^l values are around ≈ 0.005–0.1 and the corresponding Hubbard-Stratonvich coupling projection time (via λ_i) is of the same order, but can get as large as $\tau^l \approx 0.13$.

[1] W. L. McMillan, Ground state of liquid He^4, Phys. Rev. **138**, A442 (1965).

[2] D. Ceperley, G. V. Chester, and M. H. Kalos, Monte Carlo simulation of a many-fermion study, Phys. Rev. B **16**, 3081 (1977).

[3] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[4] S. R. White, Density-matrix algorithms for quantum renormalization groups, Phys. Rev. B **48**, 10345 (1993).

[5] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. (NY) **326**, 96 (2011).

[6] E. Stoudenmire and S. R. White, Studying two-dimensional systems with the density matrix renormalization group, Annu. Rev. Condens. Matter Phys. **3**, 111 (2012).

[7] D. Ceperley and B. Alder, Quantum Monte Carlo, Science **231**, 555 (1986).

[8] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).

[9] S. Zhang, J. Carlson, and J. E. Gubernatis, Constrained path Monte Carlo method for fermion ground states, Phys. Rev. B **55**, 7464 (1997).

[10] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted boltzmann machine learning for solving strongly correlated quantum systems, Phys. Rev. B **96**, 205152 (2017).

[11] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

[12] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. **119**, e2205789119 (2022).

[13] S. Humeniuk, Y. Wan, and L. Wang, Autoregressive neural Slater-Jastrow ansatz for variational Monte Carlo simulation, SciPost Phys. **14**, 171 (2023).

[14] D. Wu, R. Rossi, F. Vicentini, N. Astrakhantsev, F. Becca, X. Cao, J. Carrasquilla, F. Ferrari, A. Georges, M. Hibat-Allah, M. Imada, A. M. Läuchli, G. Mazzola, A. Mezzacapo, A. Millis, J. R. Moreno, T. Neupert, Y. Nomura, J. Nys, O. Parcollet *et al.*, Variational benchmarks for quantum many-body problems, arXiv:2302.04919 [quant-ph].

[15] S. Sorella, Systematically improvable mean-field variational ansatz for strongly correlated systems: Application to the Hubbard model, Phys. Rev. B **107**, 115133 (2023).

[16] Y. Chen, L. Zhang, Weinan E, and R. Car, Hybrid auxiliary field quantum Monte Carlo for molecular systems, J. Chem. Theory Comput. **19**, 4484 (2023).

[17] H. Shi and S. Zhang, Symmetry in auxiliary-field quantum Monte Carlo calculations, Phys. Rev. B **88**, 125132 (2013).

[18] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos *et al.* (Simons Collaboration on the

Many-Electron Problem), Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms, Phys. Rev. X **5**, 041041 (2015).

[19] N. Trivedi and D. M. Ceperley, Ground-state correlations of quantum antiferromagnets: A green-function Monte Carlo study, Phys. Rev. B **41**, 4552 (1990).

[20] J. Hubbard, Electron correlations in narrow energy bands, Proc. R. Soc. London A **276**, 238 (1963).

[21] D. Thouless, Stability conditions and nuclear rotations in the Hartree-Fock theory, Nucl. Phys. **21**, 225 (1960).

[22] R. L. Stratonovich, On a method of calculating quantum distribution functions, Sov. Phys. Doklady **2**, 416 (1957).

[23] J. E. Hirsch, Discrete Hubbard-Stratonovich transformation for fermion lattice models, Phys. Rev. B **28**, 4059 (1983).

[24] S. Vitiello, K. Runge, and M. H. Kalos, Variational calculations for solid and liquid ^{4}He with a "shadow" wave function, Phys. Rev. Lett. **60**, 1970 (1988).

[25] L. Reatto and G. L. Masserini, Shadow wave function for many-boson systems, Phys. Rev. B **38**, 4516 (1988).

[26] H. Shi, S. Chiesa, and S. Zhang, Ground-state properties of strongly interacting fermi gases in two dimensions, Phys. Rev. A **92**, 033603 (2015).

[27] M. Motta and S. Zhang, Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method, WIREs Comput. Mol. Sci. **8**, e1364 (2018).

[28] H. Xu, H. Shi, E. Vitali, M. Qin, and S. Zhang, Stripes and spin-density waves in the doped two-dimensional Hubbard model: Ground state phase diagram, Phys. Rev. Res. **4**, 013239 (2022).

[29] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, http://github.com/google/jax.

[30] I. Babuschkin, K. Baumli, A. Bell, S. Bhupatiraju, J. Bruce, P. Buchlovsky, D. Budden, T. Cai, A. Clark, I. Danihelka, A. Dedieu, C. Fantacci, J. Godwin, C. Jones, R. Hemsley, T. Hennigan, M. Hessel, S. Hou, S. Kapturowski, T. Keck *et al.*, http://github.com/google-deepmind.

[31] P. Bader, S. Blanes, and F. Casas, Computing the matrix exponential with an optimized Taylor polynomial approximation, Mathematics **7**, 1174 (2019).

[32] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, SciPost Phys. Codebases, 4 (2022).

[33] M. Fishman, S. R. White, and E. M. Stoudenmire, Codebase release 0.3 for ITensor, SciPost Phys. Codebases, 4 (2022).

[34] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S. R. White, and S. Zhang, Coexistence of superconductivity with partially filled stripes in the Hubbard model, arXiv:2303.08376 [cond-mat.supr-con].

[35] M. Qin, C.-M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S. R. White, and S. Zhang, Absence of superconductivity in the pure two-dimensional Hubbard model, Phys. Rev. X **10**, 031016 (2020).

[36] E. Vitali, H. Shi, A. Chiciak, and S. Zhang, Metal-insulator transition in the ground state of the three-band Hubbard model at half filling, Phys. Rev. B **99**, 165116 (2019).

[37] M. Qin, H. Shi, and S. Zhang, Coupling quantum monte carlo and independent-particle calculations: Self-consistent constraint for the sign problem based on the density or the density matrix, Phys. Rev. B **94**, 235119 (2016).

[38] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Stripe order in the underdoped region of the two-dimensional Hubbard model, Science **358**, 1155 (2017).

[39] J. Ostmeyer, Optimised trotter decompositions for classical and quantum computing, J. Phys. A: Math. Theor. **56**, 285303 (2023).

[40] W. Janke and T. Sauer, Properties of higher-order Trotter formulas, Phys. Lett. A **165**, 199 (1992).

[41] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[42] F. F. Assaad and I. F. Herbut, Pinning the order: The nature of quantum criticality in the Hubbard model on honeycomb lattice, Phys. Rev. X **3**, 031010 (2013).

[43] H. Shi, C. A. Jiménez-Hoyos, R. Rodríguez-Guzmán, G. E. Scuseria, and S. Zhang, Symmetry-projected wave functions in quantum Monte Carlo calculations, Phys. Rev. B **89**, 125129 (2014).

[44] J. A. Sheikh, J. Dobaczewski, P. Ring, L. M. Robledo, and C. Yannouleas, Symmetry restoration in mean-field approaches, J. Phys. G: Nucl. Part. Phys. **48**, 123001 (2021).

[45] H. Shi and S. Zhang, Infinite variance in fermion quantum Monte Carlo calculations, Phys. Rev. E **93**, 033303 (2016).

[46] A. Alexandru, P. F. Bedaque, A. Carosso, and H. Oh, Infinite variance problem in fermion models, Phys. Rev. D **107**, 094502 (2023).