Enhancing Neural Network Backflow

Kieran Loehr* and Bryan K. Clark
The Anthony J. Leggett Institute for Condensed Matter Theory and IQUIST and NCSA center for Artificial Intelligence Innovation and Department of Physics, University of Illinois at Urbana-Champaign, IL 61801, USA
(Dated: November 3, 2025)

Accurately describing the ground state of strongly correlated systems is essential for understanding their emergent properties. Neural Network Backflow (NNBF) is a powerful variational ansatz that enhances mean-field wave functions by introducing configuration-dependent modifications to single-particle orbitals. Although NNBF is theoretically universal in the limit of large networks, we find that practical gains saturate with increasing network size. Instead, significant improvements can be achieved by using a multi-determinant ansatz. We explore efficient ways to generate these multi-determinant expansions without increasing the number of variational parameters. In particular, we study single-step Lanczos and symmetry projection techniques, benchmarking their performance against diffusion Monte Carlo and NNBF applied to alternative mean fields. Benchmarking on a doped periodic square Hubbard model near optimal doping, we find that a Lanczos step, diffusion Monte Carlo, and projection onto a symmetry sector all give similar improvements achieving state-of-the-art energies at minimal cost. By further optimizing the projected symmetrized states directly, we gain significantly in energy. Using this technique we report the lowest variational energies for this Hamiltonian on 4×16 and 4×8 lattices as well as accurate variance extrapolated energies. We also show the evolution of spin, charge, and pair correlation functions as the quality of the variational ansatz improves.

I. INTRODUCTION

Accurate ground state wave-functions of strongly correlated systems are critical to extracting their properties. This is especially true as energy differences between qualitatively different phases of matter can be small. Representing the state exactly is hindered by the exponentially large Hilbert space motivating the use of variational wave-functions.

Recently, the class of neural network quantum states (NQS) [1] which uses machine learning architectures as part of the variational wave-function has shown potential for representing quantum many-body states with significant progress on using NQS for spin systems [2–8]. Fermion NQS are less well developed with some approaches directly mapping second-quantized configuration to amplitudes [9] or modifying the Jastrow factor [10]. Among the most accurate fermion variational wave-functions are the Neural Network Backflow Wavefunctions (NNBF) and low-rank variants such as the hidden fermion approaches which use ML to directly modify the mean-field determinant [11–17]. Despite promise, there is still significant room for improvement in these approaches. Various attempts at building on the standard NNBF form have included using ML architectures other then multi-layer perceptrons (MLP) [18–20] and mean-fields beyond the Slater determinant including BDG [11] as well as pfaffians [21].

In this work, we fix the neural network architecture to an MLP and consider various ways to systematically improve MLP-NNBF. We focus our tests on the 4×16 Hubbard model in periodic boundary conditions at $U = 8$ and hole doped at $\delta = 1/8$ as the paradigmatic example of a physically interesting albeit challenging strongly correlated fermionic system which has been historically used as a benchmark for such methods. We focus on both the accuracy as well as computational efficiency of various improvements. While MLP-NNBF is universal in the limit of enough neurons [11] in practice we find that the benefits of increasing neurons begins to saturate when the cost to evaluate the network becomes approximately equal to the cost of evaluating the determinant(s).

Instead, we find that one can gain significantly by using a sum of multi-determinant MLP-NNBF's [15, 16]. While formally, one could simply independently optimize all these determinants, in practice this is often difficult and time-consuming. Instead, we primarily focus on ways of generating these determinants without increasing the number of parameters. We do this either by using Lanczos steps [22–24] or symmetry projection [25–28] both of which generate a multi-determinant expansion from a single MLP-NNBF. First optimizing the single MLP-NNBF and then applying either Lanczos or symmetry projection gives similar but significant benefits resulting in the lowest energies available for this Hamiltonian; this state-of-the-art calculation took roughly 200 GPU hours. By further optimizing the projected states directly, we gain an additional approximately 45% towards the variance extrapolated ground state at a cost of almost 3000 GPU hours.

We additionally investigate the promotion of the determinant mean field to that of a pfaffian finding that the energy improvement is comparable to increasing neurons, but with a larger computational cost. We also apply diffusion Monte Carlo (DMC) [29] to the single MLP-NNBF, and find that it can frequently give ener-

* Contact author: kloehr@illinois.edu

gies comparable to Lanczos and unoptimized symmetry projection. Finally, we compute spin, charge, and pairing correlations and illustrate how they evolve under increasingly accurate variational wave-functions.

II. METHODS

A. NNBF Ansatz

The NNBF ansatz extends a mean-field wavefunction by introducing configuration-dependant backflow corrections that can alter both the amplitude and sign structure of the state [11]. The amplitude for a configuration $\mathbf{n}$ is

$$\psi_{NNBF}(\mathbf{n}) = \langle \mathbf{n}|\psi_{NNBF}\rangle = \det(\mathbf{u}(\mathbf{n})[\mathbf{n}]) \quad (1)$$

where $\mathbf{u}(\mathbf{n})$ is the output, an $N \times L$ matrix, of an MLP with input $\mathbf{n}$. $[\mathbf{n}]$ notates selecting the columns of $\mathbf{u}(\mathbf{n})$ corresponding to the occupied sites resulting in a $N \times N$ matrix. Here both spin sectors are combined into a single matrix (e.g. N is the sum of the spin up and down electrons and L is twice the number of physical sites), which offers greater representational ability than the product of separate spin matrices [30].

Our MLP is fully connected with two hidden layers each followed by a layer normalization and relu activation. The input layer is set by $|\mathbf{n}\rangle = |n_0, n_1, ..., n_L\rangle$ where $n_i \in \{-1, 1\}$ denotes an empty or occupied spin orbital. The last output layer contains a bias but no non-linear function; this bias can be interpreted as a fixed mean-field solution onto which the neural network provides additive configuration-dependent backflow corrections. The accuracy of this MLP-NNBF can be systematically improved by increasing the number of hidden neurons, n_h.

B. Optimization

To enhance training efficiency and mitigate convergence to local minima, we update our network as

$$\theta_{i+1} = \theta_i - \mathbf{s}_i \delta_L \times \text{sign}[\nabla_\theta E_\theta] \quad (2)$$

where θ_i denotes the parameter vector at iteration i, δ_L is the learning rate, and $\mathbf{s}_i$ is a vector of scalars drawn uniformly from the interval $[0, 1)$ at each iteration [31–34]. The deviation of this update step from the gradient helps prevent the optimizer from becoming trapped in shallow energy basins.

To accelerate convergence, we first optimize the mean-field Slater determinant before introducing the backflow corrections. During the final stages of training, we apply an exponential learning-rate decay to allow the optimizer to refine the variational minimum by flowing to the minima of narrow funnels. Further optimization details are available in Appendix B.

All optimizations reported in this paper were performed using H100 GPUs on GH200 superchips. Most optimizations were performed using a single H100 GPU, but longer optimizations were parallelized across 4 GPUs. GPU times are reported as cumulative GPU usage (e.g. parallelized times are 4 times the wall clock time). All optimization times include the cumulative cost of optimization for a given point which includes (where specified) optimizing smaller ansatz first.

III. RESULTS

To assess the improvements to the NNBF ansatz, we benchmark on the two-dimensional Fermi-Hubbard model on a square lattice

$$\hat{\mathcal{H}} = -t \sum_{\langle ij\rangle,\sigma} (\hat{c}^\dagger_{i,\sigma}\hat{c}_{j,\sigma} + h.c.) + U \sum_i \hat{n}_{i,\uparrow}\hat{n}_{j,\downarrow}. \quad (3)$$

We focus on the computationally challenging regime of strong repulsive interactions $U = 8$, $t = 1$ and small hole doping $\delta = 1/8$ in periodic boundary conditions. We primarily report results on a 4×16 lattice outside of section III B where we report on the 4×8 system. To compare between different lattice sizes more easily, we always report the energy per site and its variance throughout this work.

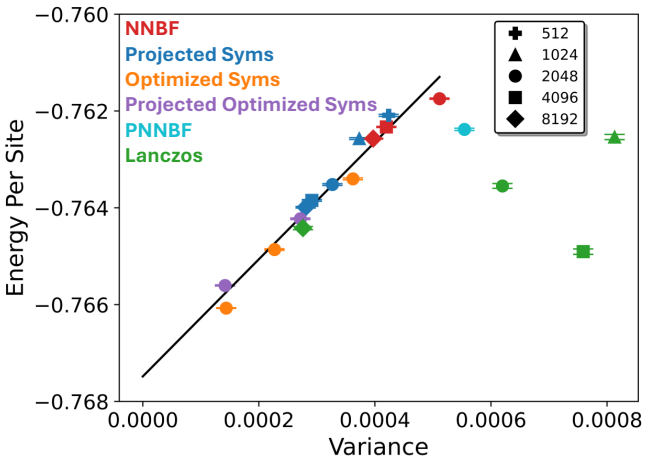

FIG. 1. Energy vs. variance of different variational wave-functions with a linear extrapolation to zero variance taken through the base NNBF and all symmetrized versions (e.g. red, blue, orange and purple points) whose energy was less then -0.7615. Symbols denote n_h.

Variance Extrapolation: Throughout this manuscript, we consider a number of NNBF enhancements measuring both the energy and their variance (see Fig. 1). Close to the true ground state, the energy and variance should be linearly related and, in practice, this can be true even outside this regime if one chooses a uniformly related set of wave-functions. Here we do a linear extrapolation through all sufficiently low energy points of our standard NNBF models and all forms of symmetry projections (discussed in detail

in later sections). We expect the extrapolated energy, −0.76748, which is not necessarily variational, to be a good proxy for the true ground state energy of our Hamiltonian and use this value as the "ground truth" to report fractional improvements of our different methods discussed in this paper. The y-axes of Fig. 2, Fig. 3, and Fig. A1 use this value as its minima.

Increasing neurons: With 64 hidden neurons and a single-layer MLP, NNBF originally reached energies of −0.746 per site [11]. Both significantly increasing neuron width, the number of layers, and using full determinants instead of products of spin determinants yields much more accurate results [13, 30]. In our implementation, a two-layer MLP with 8192 neurons reaches an energy of −0.76258(1) within 400 GPU hours. This energy is lower than recent results using a tensor backflow ansatz with a Lanczos step [24].

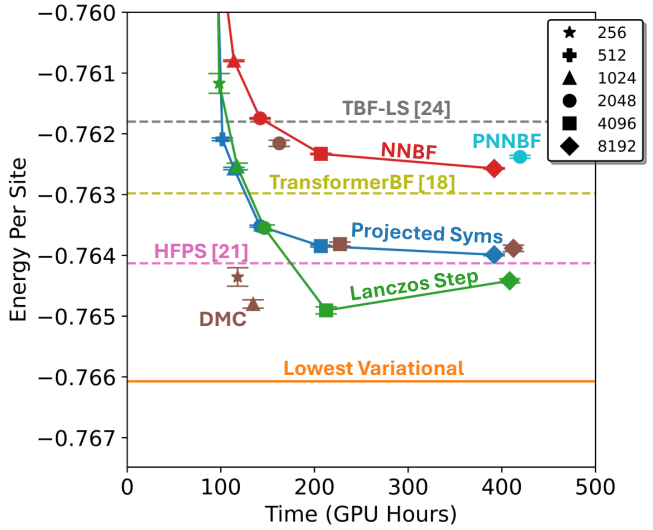

FIG. 2. Variational energy of NNBF approaches as a function of GPU training hours. Symbols denotes n_h. Comparison with other works shown as dotted lines. The lowest variational energy achieved in this work is shown as a solid line. The Lanczos step optimization and DMC for $n_h = 512$ failed and are excluded.

Fig. A1 and Fig. 2 show the dependence of the variational energy on the number of hidden neurons, n_h for standard NNBF. The energy decreases with neurons at small n_h but improves much more slowly as n_h increases; for example increasing from $n_h = 4096$ to $n_h = 8192$ yields only a 4.7% relative improvement with respect to the extrapolated ground state.

At large n_h, increasing n_h is not only less effective, but also increasingly expensive. A fixed depth multi-layer MLP-NNBF scales as $\mathcal{O}(n_o n_h + n_h^2 + \alpha n_h n_o^2 + (\alpha n_o)^3)$ per step where n_o is the number of orbitals and α is the filling fraction. For our system, $n_o = L = 128$ and $\alpha = 0.4375$. For small n_h, the cost of evaluating the determinant, $(\alpha n_o)^3$, dominates and the dependence on n_h is primarily linear through the $n_h n_o$ term. This linear scaling is reflected in the GPU runtime in Fig. 2. Once $n_h > \alpha n_o^2$, the n_h^2 term becomes dominant, and the computational cost grows quadratically with n_h. For our parameters, this crossover happens at $n_h = 7168$ which interestingly coincides with the point where the energy gains also begin to saturate. Thus, beyond a certain network width, the benefit of increasing n_h becomes marginal due to both diminishing variational improvement and rapidly increasing computational cost. This motivates the development of alternative strategies to enhance the NNBF ansatz.

Pffafian NNBF: One natural change is to modify the underlying mean-field away from the Slater Determinant to BDG-NNBF [11] or pfaffians [10, 21, 35–37]. Here we consider a direct pfaffian NNBF (PNNBF) which we define as

$$\psi_{PNNBF}(\mathbf{n}) = \mathrm{pf}\Big(\left(\mathbf{U}(\mathbf{n})\mathbf{J}(\mathbf{n})\mathbf{U}(\mathbf{n})^T\right)[\mathbf{n},\mathbf{n}]\Big) \tag{4}$$

where $\mathbf{U}(\mathbf{n})$ and $\mathbf{J}(\mathbf{n})$ are both $L \times L$ matrices output by the MLP (e.g. the last layer of the network has $2L^2$ nodes) and $[\mathbf{n},\mathbf{n}]$ denotes selecting both the rows and columns corresponding to the occupied sites.

This formulation strictly generalizes the standard NNBF ansatz: if $\mathbf{J}(\mathbf{n})$ is initialized such that $\mathrm{pf}(\mathbf{J}) = 1$ (via output biases) and $\mathbf{U}(\mathbf{n})$ is taken to be the $N \times L$ NNBF matrix output padded with zeros to size $L \times L$, the original determinant based NNBF is recovered.

We study PNNBF with $n_h = 2048$, initializing the parameters by transfer learning from a trained NNBF with the same width and $J(n)$ initialized with 1's and -1's on the super-diagonal and sub-diagonal respectively. As shown in Fig. 2, the resulting energy improves only 11% relative to the NNBF state with the same number of neurons. In contrast, a determinant based NNBF with four times the neurons achieves slightly lower energy than the PNNBF at comparable cost. Training PNNBF from scratch yields an energy consistent with the transfer-learned result but requires approximately 20% more GPU time.

While the PNNBF has the same formal scaling as NNBF, we find that its optimization is significantly slower in practice. The minimal improvement in energy is consistent with previous results on the BDG-NNBF [11] which found the improved mean field reference gives an advantage at fixed n_h but this benefit diminishes as n_h increases.

Optimization-free multi-determinants: An alternative approach to improving NNBF is to introduce multiple determinants [15, 16]. We consider two methods of generating multi-determinant expansions without introducing significant additional variational parameters: symmetry projection and single-step Lanczos. We start by considering how to use these approaches with minimal additional optimization. Later in the manuscript we will see that optimizing with a symmetrized multi-determinant expansion gains additional significant energy at some cost in time.

While symmetries can be spontaneously broken in the thermodynamic limit, on a finite lattice the ground state

must respect the symmetry group G of the Hamiltonian. There are several ways to enforce symmetry in a wavefunction. For Abelian symmetries, one can choose a representative configuration for each equivalence class, $\{g\mathbf{n} : \forall g \in G\}$, and fix the magnitude of the amplitude of all configurations in that class to that of the representative, with phases set to project onto the desired symmetry sector [38]. Such constructions will often have higher energy because forcing different configurations to have the same amplitudes is an additional constraint on the variational freedom. We see this empirically (see Appendix C). Alternatively, equivariant convolution neural networks can encode symmetries into the mean-field orbitals producing a symmetrized wave-function with only a single network evaluation (although sometimes multiple determinant evaluations) [19–21]. In either case, symmetry is enforced by constraining the wave-function which reduces expressivity. In contrast, our goal is to maximize representation ability while keeping the cost constrained.

Here, we enforce symmetrization as

$$\Psi_{sym}(\mathbf{n}) = \sum_g \Pi(\mathbf{n}, g)\Psi(g\mathbf{n}), \tag{5}$$

where $g \in G$ and $\Pi(\mathbf{n}, g) = \pm 1$ is an additional sign coming from fermion permutations. This generates a large multi-determinant expansion in the trivial irrep, though other irreps are straightforward to implement. This requires a number of evaluations of both the network and the determinant equal to the number of symmetries, which scales linearly with system size due to translations. The 4×16 lattice has 4 point group symmetries, 64 translations, and an additional 2 symmetries from spin resulting in a multi-determinant expansion of 512 determinants. As shown in Fig. 2, this projection gives a consistent decrease in energy of approximately -0.002 (equivalent to an improvement between 27% and 34%) for larger neuron number ($n_h > 128$). Despite the large multi-determinant expansion, the optimization cost is unchanged, as the parameters are optimized only on the single-determinant ansatz.

A second approach to generating a nearly optimization-free multi-determinant expansion is to apply a single Lanczos step. Given a trained NNBF state $|\psi\rangle$ we consider $(1 + \alpha H)|\psi\rangle$ and optimize only the scalar coefficient α. Rather then computing $\langle H^3 \rangle$ directly from sampling from $|\psi|^2$, we follow the iterative sampling method of Ref. [22], which samples from $(1 + \alpha H)|\psi\rangle$ with the current optimal α (see Appendix D). Using this method, it takes around 10 GPU hours to determine α, which is negligible on top of the cost of the MLP optimization. The Lanczos step generates determinants connected to $|\psi\rangle$ via the Hamiltonian. Since each spin can hop to up to 4 neighbors, the resulting multi-determinant expansion can have up to $4N$ additional determinants. For our $\delta = 1/8$ model this results in 225 determinants, fewer than in the symmetry projection. As shown in Fig. 2, the Lanczos steps performs similarly to the symmetry projection at all neurons, but gets slightly lower energies, particularly at larger n_h where it provides an additional between 26% and 50% improvement in the energy.

These simple enhancements already achieve state-of-the-art-energies. Symmetry projection alone reaches energies comparable to the best previous results [21] while the multi-determinant expansion generated by a single Lanczos step on top of NNBF with either $n_h = 4096$ or $n_h = 8192$ gives lower energies than all previous reported values. The NNBF training time is significantly faster ($\approx$ 200 GPU hours) compared to previously reported HFPS training times ($\approx$ 1500 GPU hours [21]). This demonstrates that symmetry projection and Lanczos corrections provide an efficient and highly effective methodology for improving NNBF.

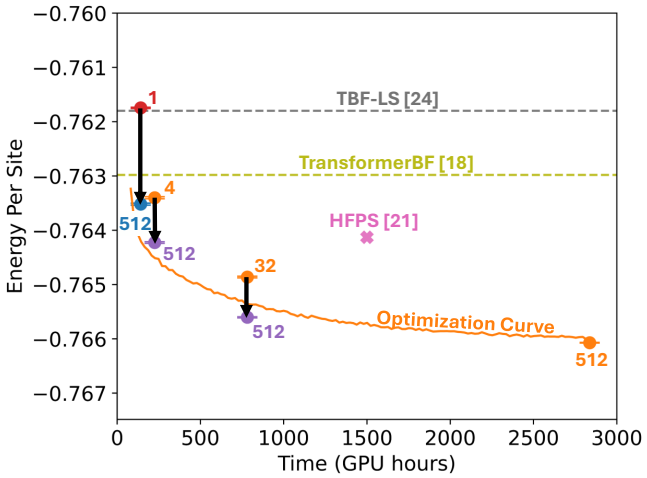

FIG. 3. Variational energy versus GPU training hours for different symmetric operations at $n_h = 2048$. We consider both optimization with a fixed number of symmetries (orange), and projection from these points to the full symmetry sector (purple). The numbers next to each point correspond to the number of symmetries. The projection from the non-symmetrized (red point) to full symmetries (blue point) is also seen in Fig. 2. Comparison with other works shown as dotted lines and a pink cross are all higher in energy then the projected optimized symmetries. The optimization curve is the optimization for the optimized fully symmetric state.

Optimizing multi-determinants: In the previous section, we optimize the MLP-NNBF ansatz first and then project it into a multi-determinant expansion. Instead, one can optimize within the multi-determinant expansion after projection [25]. Here we optimize the energy of Eq. 5 using the parameters of a single MLP using different sets of symmetries including the 4 point group symmetries; 32 symmetries by adding in spin inversion and half translation symmetries; and 512 symmetries by including all translations. After each optimization, we also evaluate the energy after projecting onto the full 512 symmetries.

This procedure systematically improves the ansatz and gets energies substantially lower than other enhancements. As shown in Fig. 3, for the $n_h = 2048$ model,

optimization with four symmetries improves the energy by 29% (43% after full projection), optimization with the 32 symmetries improves the energy by 54% (67% after full projection), and optimization with all 512 symmetries improves the energy by 75% relative to the single MLP ansatz with same n_h. These additional optimizations all use the single-determinant NNBF as a starting point and, in total, take approximately 200, 800, and 3000 GPU hours respectively. Interestingly, as can be seen in Fig. 3, the optimization curve of the fully symmetrized state is almost always lower in energy than any other state for a given number of GPU hours. This suggests that for a fixed GPU budget an effective strategy is to simply optimize the fully symmetrized state until the target number of hours.

We now directly compare energies for ansatzes that are evaluating an equal number of determinants. We find that our ansatz with four determinants generated by optimizing four symmetries has an energy of -0.76340(2) which is slightly lower than Transformer-BF [18] energy of -0.76298 which also uses four determinants. This suggest that either our $n_h = 2048$ MLP has more representation ability than the Transformer-BF or that the transformer optimization is stuck in a local minima. No explicit time is reported in Ref. [18].

The HFPS [21] uses a CNN to augment a pfaffian ansatz using the hidden fermion methodology. The translation equivariance of the CNN architecture and a choice of a sub-lattice symmetric pfaffian ansatz enforces symmetrization without multiple network or pfaffian evaluations. However, to restore the full symmetric invariance, there is a sum over 32 pfaffians to restore all translational, point group and spin symmetries. The HFPS reaches an energy of -0.76413 which is somewhat higher then then $-0.76486(2)$ of our symmetry optimization with 32 determinants. Additionally, our result can be further projected at essentially no cost to the fully symmetric state to reach an energy of $-0.76560(1)$, while the HFPS can not be projected further. We suspect that this improvement in energy is largely due to the fact that the hidden fermion methodology is essentially a low rank NNBF update and therefore has lower representation ability compared to NNBF [13]. The optimization of our 32 determinant MLP-NNBF is faster by approximately a factor of 2 compared with the HFPS.

One could further relax constraints and optimize all the determinants separately. We find that the energy landscape with multi-determinant expansions is challenging, as when we initialize a model with 4 independent determinants starting from the symmetric point, we find no improvement in energy (and actually an increase in energy at fixed learning rate suggesting a need for more adaptive optimization). This suggests that by constraining the ansatz we are helping simplify the energy landscape.

Fixed Node Diffusion Monte Carlo: So far we have primarily considered variational wave-functions which can be compactly represented. Alternatively, we can use quantum Monte Carlo which represents the wavefunction with an exponential number of determinants stochastically. Because of the sign problem, we use the fixed-node diffusion Monte Carlo (FNDMC) starting with the single determinant ansatz. FNDMC takes approximately 20 additional GPU hours. As seen in Fig. 2, the FNDMC decrease in energy has a much larger spread than either the symmetry projection or the Lanczos step spanning from as little as 7% improvement to as high as a 72% improvement from the NNBF at the same n_h. Interestingly, while the standard lore is that FNDMC will generically do better when applied to anzats with variationally smaller energies, we do not find that to be the case here. While the energy of the NNBF decreases monotonically with n_h, the energy of FNDMC vacillates with n_h. This suggests that at these small energy scales it would be better to directly optimize the FNDMC energy.

A. Observables

In this regime, the Hubbard model is believed to have stripes with anti-ferromagnetic (AFM) order separated by one-dimensional hole-rich domain walls over which the polarization of the AFM order flips [18, 21, 39–42]. We compute observables for our different variational wavefunctions including the spin-spin correlations

$$C_s(\mathbf{r}) = \frac{1}{\ell}\sum_{\mathbf{i}} \langle \hat{\mathbf{S}}_{\mathbf{i}+\mathbf{r}} \cdot \hat{\mathbf{S}}_{\mathbf{i}} \rangle - \left(\frac{1}{\ell} \sum_{\mathbf{i}} \langle \hat{\mathbf{S}}_{\mathbf{i}} \rangle \right)^2 \tag{6}$$

and the charge-charge correlations

$$C_c(\mathbf{r}) = \frac{1}{\ell}\sum_{\mathbf{i}} \langle \hat{n}_{\mathbf{i}+\mathbf{r}} \cdot \hat{n}_{\mathbf{i}} \rangle - \left(\frac{1}{\ell} \sum_{\mathbf{i}} \langle \hat{n}_{\mathbf{i}} \rangle \right)^2 \tag{7}$$

where the sum is over the physical site locations and $\ell \equiv L/2$ is the total number of physical sites. In Fig. 4(a,b), we show these observables for our lowest energy variational wave-function. As anticipated, there are two domain walls across which the AFM sub-lattice polarization flips. One can see from the charge-charge correlation, these domain walls are hole-doped. We look at the spin ($\alpha = s$) and charge ($\alpha = c$) Fourier transforms

$$\widetilde{C}_\alpha(\mathbf{k}) = \sum_{\mathbf{r}} C_\alpha(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{8}$$

in Fig. 4(d,e) for a variety of different variational wavefunctions. Looking at slices in the k-space structure factors at $\mathbf{k} = (k_x, \pi)$ for the spin and $\mathbf{k} = (k_x, 0)$ for the charge, we find the shown variational wave-functions look similar and have respective peaks at $\mathbf{k} = (7\pi/8, \pi)$ and $\mathbf{k} = (\pi/4, 0)$. This demonstrates that even higher energy NNBF variational wave-functions capture the spin density wave with period 16 and charge density wave with period 8 successfully. Interestingly, if we go to the $n_h = 128$ state which is much higher in variational energy, there is

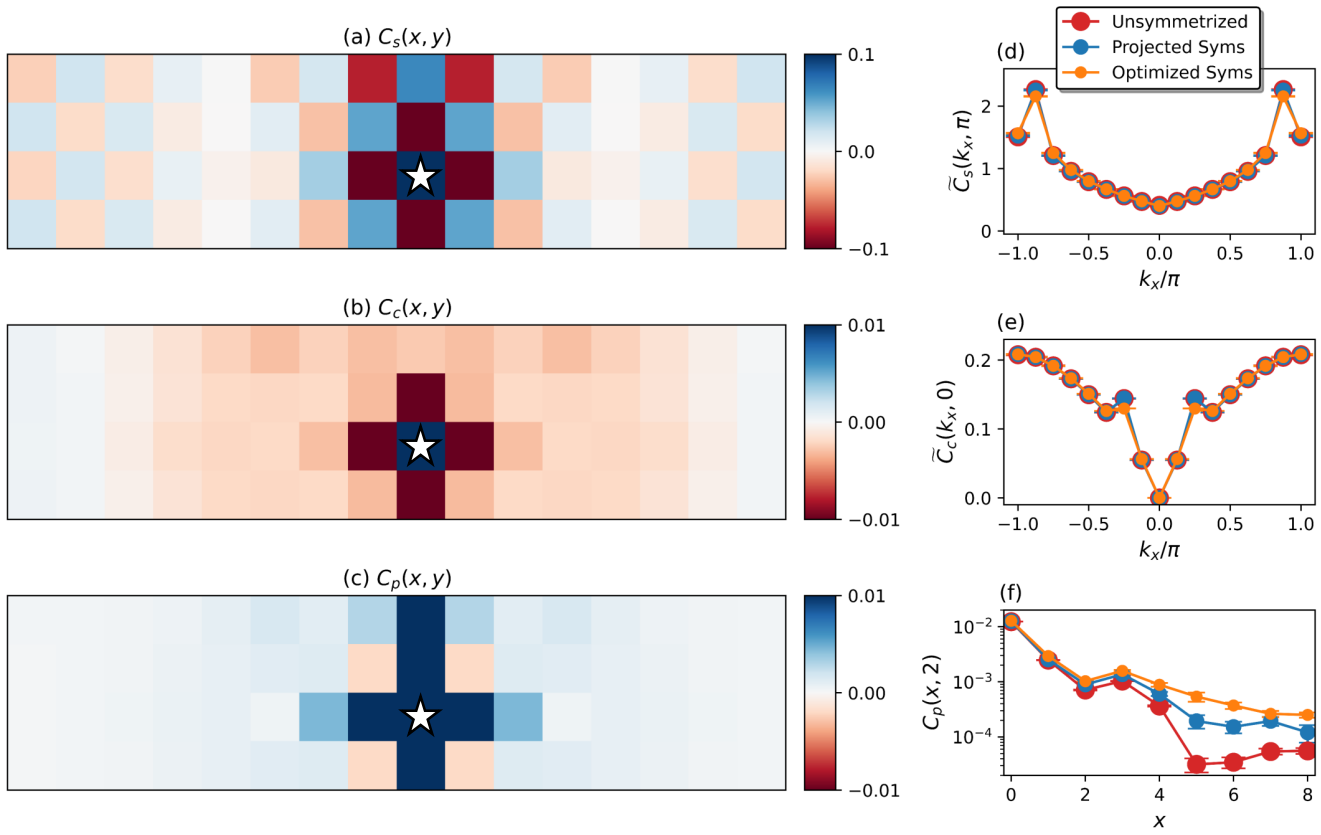

FIG. 4. Observables on the 4×16 lattice. (a-c) The correlation functions C_s, C_c, and C_p in real space for the best variational model, $n_h = 2048$ with optimized symmetries. To show weaker correlation, the values are truncated such that $|C_s| \leq 0.1$ and $|C_c|, |C_p| \leq 0.01$. The central site is marked with a star. (d,e) The spin and charge structure factors of $n_h = 2048$ models with no symmetries, projected symmetries, and optimized symmetries. Peaks appear at $(7\pi/8, \pi)$ and $(\pi/4, 0)$ respectively. (f) The pair-pair correlation $C_p(x, 2)$ on the same models in (d, e). Similar benchmarks for HFPS are shown in Ref. [21].

a non-trivial change in the width of the stripes and therefore respective structure factors (see Fig. A3).

The doped square lattice Hubbard model can also have a tendency towards d-wave pairing. We compute the pair-pair correlation function

$$C_p(\mathbf{r}) = \frac{1}{\ell} \sum_{\mathbf{i}} \langle \hat{\Delta}^\dagger(\mathbf{i} + \mathbf{r}) \hat{\Delta}(\mathbf{i}) \rangle, \tag{9}$$

for d-wave symmetrized pairs

$$\hat{\Delta}(\mathbf{r}) = \frac{1}{4} \sum_{\delta} \frac{1}{\sqrt{2}} \text{sign}(\delta) (\hat{c}_{\mathbf{r},\uparrow} \hat{c}_{\mathbf{r}+\delta,\downarrow} - \hat{c}_{\mathbf{r},\downarrow} \hat{c}_{\mathbf{r}+\delta,\uparrow}). \tag{10}$$

As shown in Fig. 4(f), we find a non-trivial short-range d-wave pairing. A strong short-range d-wave pairing was also seen in Ref. 21 but required a pfaffian and optimization using initialized (but eventually removed) pinning fields to find. Interestingly, the large x pairing function increases as the wave-function quality improves and there are hints that our pairing function is starting to saturate at the largest x.

B. 4×8 Lattice

In addition to the 4×16 lattice, we benchmark these enhancements on a 4×8 lattice and find generically similar results as shown in Fig. 5. Most strikingly, the lowest variational energy, produced by optimizing the fully symmetrized state over 128 determinants, reaches an energy of $-0.768337(2)$ which has a relative error of 1.0×10^{-4} compared to the variance extrapolated "ground truth" energy of -0.768416. This is an order-of-magnitude improvement compared to the previous best variational energies on this model[21].

IV. DISCUSSION

In this work, we have investigated several strategies for improving the NNBF ansatz. Using these strategies, we find the ansatz with the lowest reported variational energies for the $U = 8$ Hubbard model with $\delta = 1/8$ hole doping in periodic boundary conditions for both the 4×16 and 4×8 lattices. While increasing the width of an MLP will decrease the energy, in practice we find that

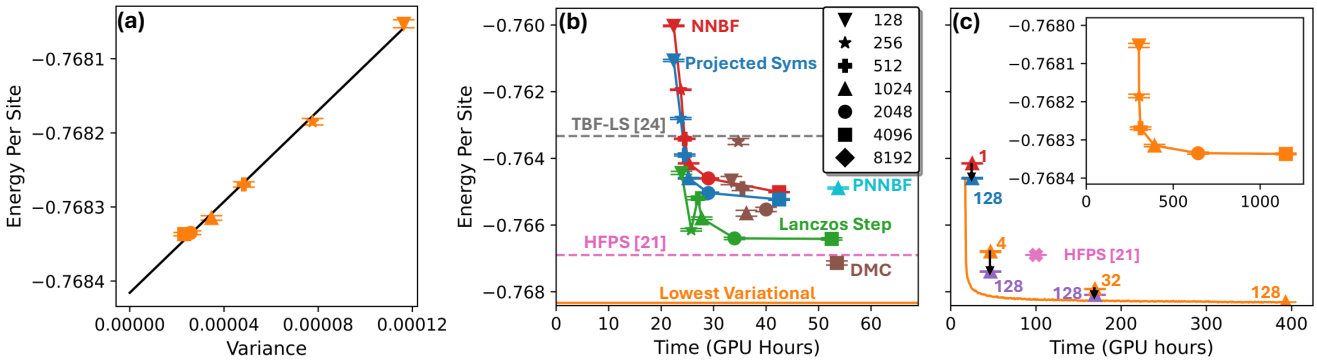

FIG. 5. Energies on the 4×8 lattice. Symbols denote n_h. (a) Energy vs variance of the optimized symmetries at various n_h with variance extrapolated best-fit line which is used as the bottom for the y-axis in (b) and (c). (b) Energy vs GPU time for different methods which optimize only the single mean-field as well as variational energies from other works (dotted lines) and our lowest variational energy (solid orange). (c) Energy vs GPU time for optimized symmetries (orange) and projected optimized symmetries (purple) at $n_h = 1024$ with numbers indicating the number of symmetries. The orange optimization curve is for the projected optimized symmetry with 128 symmetries. Inset contains projected optimized symmetries with 128 symmetries at various n_h.

at large enough neurons there is diminishing returns in this process and a multi-determinant approach becomes preferable. Enhancing an optimized NNBF ansatz with a single Lanczos step, diffusion Monte Carlo, or symmetry projection yields substantial energy improvements at almost no cost in time yielding results for the 16×4 lattice that are both lower in energy and faster (where reported) than previous techniques. By further optimizing after projection, one gets significant additional gains in energies. This is the case even when optimizing with only a small number of such determinants (e.g. 32) and then projecting afterwards. That said, interestingly, at any fixed amount of GPU time, it is nearly optimal to just optimize the fully symmetrized state for that number of GPU hours. The uniformity of symmetrized NNBF wave-functions permitted accurate variance extrapolated energies. We also considered various observables for these systems finding spin and charge density waves that are consistent with previous results and finding that the d-wave pair correlation functions have strong local correlations and show hints of saturation at large x. A natural next step is to apply the symmetrized and post-processed NNBF framework to additional challenging and physically interesting systems.

ACKNOWLEDGMENTS

We acknowledge the helpful conversations with Z. Liu, A. Liu, and J. Hahm. The NQS simulations are implemented using internally developed code with JAX [43], Flax [44], and Optax [45]. This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers. This research used both the DeltaAI advanced computing and data resource, which is supported by the National Science Foundation (award OAC 2320345) and the State of Illinois, and the Delta advanced computing and data resource which is supported by the National Science Foundation (award OAC 2005572) and the State of Illinois. Delta and DeltaAI are joint efforts of the University of Illinois Urbana-Champaign and its National Center for Supercomputing Applications.

[1] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017), https://www.science.org/doi/pdf/10.1126/science.aag2302.
[2] 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, R. Pohle, I. Romero, M. Schmid, J. M. Silvester, S. Sorella, L. F. Tocchio, L. Wang, S. R. White, A. Wietek, Q. Yang, Y. Yang, S. Zhang, and G. Carleo, Variational benchmarks for quantum many-body problems, Science **386**, 296 (2024), https://www.science.org/doi/pdf/10.1126/science.adg9774.
[3] A. Chen and M. Heyl, Empowering deep neural quantum states through efficient optimization, Nature Physics **20**, 1476 (2024).
[4] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, A simple linear algebra identity to optimize large-scale neural network quantum states, Communications Physics **7**, 260 (2024).
[5] T. Đurić, J. H. Chung, B. Yang, and P. Sengupta, Spin-1/2 kagome heisenberg antiferromagnet: Machine learning discovery of the spinon pair-density-wave ground

state, Phys. Rev. X **15**, 011047 (2025).
[6] Y. Nomura and M. Imada, Dirac-type nodal spin liquid revealed by refined quantum many-body solver using neural-network wave function, correlation ratio, and level spectroscopy, Phys. Rev. X **11**, 031034 (2021).
[7] C. Roth, A. Szabó, and A. H. MacDonald, High-accuracy variational monte carlo for frustrated magnets with deep neural networks, Phys. Rev. B **108**, 054410 (2023).
[8] L. L. Viteritti, R. Rende, A. Parola, S. Goldt, and F. Becca, Transformer wave function for two dimensional frustrated magnets: Emergence of a spin-liquid phase in the shastry-sutherland model, Phys. Rev. B **111**, 134411 (2025).
[9] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, Nature Communications **11**, 2368 (2020).
[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, Proceedings of the National Academy of Sciences **119**, e2122059119 (2022), https://www.pnas.org/doi/pdf/10.1073/pnas.2122059119.
[13] Z. Liu and B. K. Clark, Unifying view of fermionic neural network quantum states: From neural network backflow to hidden fermion determinant states, Phys. Rev. B **110**, 115124 (2024).
[14] D. Luo, D. D. Dai, and L. Fu, Simulating moiré quantum matter with neural network (2024), arXiv:2406.17645 [cond-mat.str-el].
[15] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Ab initio solution of the many-electron schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).
[16] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic schrödinger equation, Nature Chemistry **12**, 891 (2020).
[17] A.-J. Liu and B. K. Clark, Neural network backflow for ab initio quantum chemistry, Phys. Rev. B **110**, 115137 (2024).
[18] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, L. Wang, and D. Lv, Solving the hubbard model with neural quantum states (2025), arXiv:2507.02644 [cond-mat.str-el].
[19] I. Romero, J. Nys, and G. Carleo, Spectroscopy of two-dimensional interacting lattice electrons using symmetry-aware neural backflow transformations, Communications Physics **8**, 46 (2025).
[20] D. Luo, G. Carleo, B. K. Clark, and J. Stokes, Gauge equivariant neural networks for quantum lattice gauge theories, Phys. Rev. Lett. **127**, 276402 (2021).
[21] A. Chen, Z.-Q. Wan, A. Sengupta, A. Georges, and C. Roth, Neural network-augmented pfaffian wavefunctions for scalable simulations of interacting fermions (2025), arXiv:2507.10705 [cond-mat.str-el].
[22] S. Sorella, Generalized lanczos algorithm for variational quantum monte carlo, Phys. Rev. B **64**, 024512 (2001).
[23] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, Direct evidence for a gapless Z_2 spin liquid by frustrating néel antiferromagnetism, Phys. Rev. B **88**, 060402 (2013).
[24] Y.-T. Zhou, Z.-W. Zhou, and X. Liang, Solving fermi-hubbard-type models by tensor representations of backflow corrections, Phys. Rev. B **109**, 245107 (2024).
[25] R. Rodríguez-Guzmán, C. A. Jiménez-Hoyos, R. Schutski, and G. E. Scuseria, Multireference symmetry-projected variational approaches for ground and excited states of the one-dimensional hubbard model, Phys. Rev. B **87**, 235129 (2013).
[26] D. Tahara and M. Imada, Variational monte carlo method combined with quantum-number projection and multi-variable optimization, Journal of the Physical Society of Japan **77**, 114701 (2008).
[27] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, Journal of Physics: Condensed Matter **33**, 174003 (2021).
[28] M. Reh, M. Schmitt, and M. Gärttner, Optimizing design choices for neural quantum states, Phys. Rev. B **107**, 195115 (2023).
[29] A. Annarelli, D. Alfè, and A. Zen, A brief introduction to the diffusion monte carlo method and the fixed-node approximation, The Journal of Chemical Physics **161**, 241501 (2024).
[30] J. Lin, G. Goldshlager, and L. Lin, Explicitly antisymmetrized neural network layers for variational monte carlo simulation, Journal of Computational Physics **474**, 111765 (2023).
[31] B. K. Clark, D. A. Abanin, and S. L. Sondhi, Nature of the spin liquid state of the hubbard model on a honeycomb lattice, Phys. Rev. Lett. **107**, 087204 (2011).
[32] J. Lou and A. W. Sandvik, Variational ground states of two-dimensional antiferromagnets in the valence bond basis, Phys. Rev. B **76**, 104432 (2007).
[33] L. Otis and E. Neuscamman, Complementary first and second derivative methods for ansatz optimization in variational monte carlo, Physical Chemistry Chemical Physics **21**, 14491–14510 (2019).
[34] J. Bernstein, Y.-X. Wang, K. Azizzadenesheli, and A. Anandkumar, signSGD: Compressed optimisation for non-convex problems, in *Proceedings of the 35th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 80, edited by J. Dy and A. Krause (PMLR, 2018) pp. 560–569.
[35] B. Fore, J. M. Kim, G. Carleo, M. Hjorth-Jensen, A. Lovato, and M. Piarulli, Dilute neutron star matter from neural-network quantum states, Phys. Rev. Res. **5**, 033062 (2023).
[36] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, Neural-network quantum states for ultra-cold fermi gases, Communications Physics **7**, 148 (2024).
[37] N. Gao and S. Günnemann, Neural pfaffians: Solving many many-electron schrödinger equations, in *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (2024).
[38] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and many-body excitations with neural-network quantum states, Phys. Rev. Lett. **121**, 167204 (2018).
[39] S. Sorella, Systematically improvable mean-field variational ansatz for strongly correlated systems: Application to the hubbard model, Phys. Rev. B **107**, 115133 (2023).
[40] 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), https://www.science.org/doi/pdf/10.1126/science.aam7127.
[41] M. Qin, C.-M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S. R. White, and S. Zhang (Simons Collaboration on the Many-Electron Problem), Absence of superconductivity in the pure two-dimensional hubbard model, Phys. Rev. X **10**, 031016 (2020).
[42] A. Wietek, Fragmented cooper pair condensation in striped superconductors, Phys. Rev. Lett. **129**, 177001 (2022).
[43] 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, JAX: composable transformations of Python+NumPy programs (2018).
[44] J. Heek, A. Levskaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner, and M. van Zee, Flax: A neural network library and ecosystem for JAX (2024).
[45] DeepMind, 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, I. Kemaev, M. King, M. Kunesch, L. Martens, H. Merzic, V. Mikulik, T. Norman, G. Papamakarios, J. Quan, R. Ring, F. Ruiz, A. Sanchez, L. Sartran, R. Schneider, E. Sezener, S. Spencer, S. Srinivasan, M. Stanojević, W. Stokowiec, L. Wang, G. Zhou, and F. Viola, The DeepMind JAX Ecosystem (2020).

Appendix A: Neuron Optimization

In Fig. A1 we present the time taken to perform the optimization of the base NNBF ansatz on the 4×16 lattice with varying neurons zoomed out compared to Fig. 2. As mentioned in section III, we see that the energy decreases rapidly with neurons at small n_h but improves much more slowly as n_h increases. At large n_h, the energy effectively plateaus significantly above the extrapolated ground state.

Appendix B: Optimization Details

Unless otherwise specified, the total number of markov chains is fixed at $N_s = 1024$ and the learning rate is fixed at 3×10^{-4}. We first train the single particle orbitals without backflow for 10^5 VMC iterations. We then multiply all single particle orbitals by a factor of 10^3, set them to the bias of the last layer of the MLP, and randomly initialize the weights to form the initial MLP-NNBF ansatz. The training for the data in Fig. A1 took 3×10^6 VMC iterations, followed by 5×10^5 VMC iterations with an exponential decay learning rate schedule tuning the learning rate to 10^{-7}.

The training for the model with 4 symmetries took approximately 8×10^5 VMC iterations, while the training for the model with 32 symmetries took approximately 5×10^5 VMC iterations. The training for the model with all symmetries took 1.5×10^5 VMC iterations. The training for these models all ended with an exponential decay learning rate schedule tuning the learning rate to 10^{-7}. The total times for training these models is shown in Fig. 3.

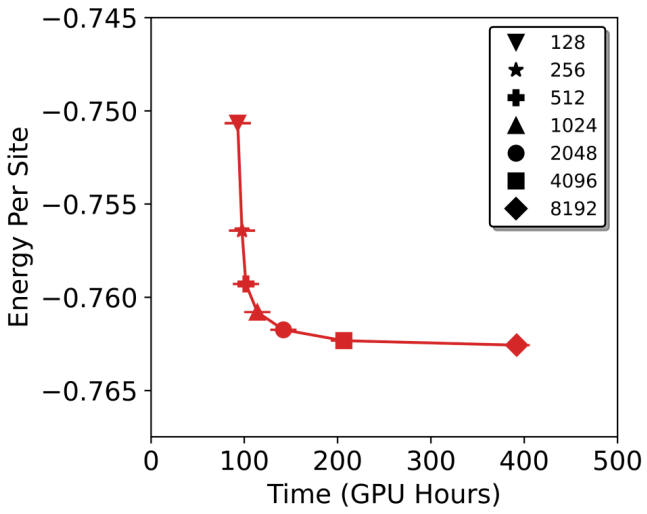

FIG. A1. Variational energy as a function of GPU training time at various n_h demonstrating the trade-off between model expressiveness and computational cost. Symbols denote n_h.

Appendix C: Representative Symmetry

As mentioned in Section II, there are many different ways to ensure a wave-function preserves symmetry. Here we consider symmetrization by choosing a representative configuration for each equivalence class to fix the magnitude of the amplitude for all configurations in that group and with phases then fixed by the symmetry sector. We see empirically in Fig. A2 that choosing a representative and projecting to the trivial irrep initially increases the energy, and while the energy eventually comes down to the base energy in the 4×8 lattice, the energy in the 4×16 lattice ends up higher then the base model.

Appendix D: Lanczos Step

Lanczos steps improve the representation ability of a variational ansatz by projecting the wavefunction into the Krylov subspace of the Hamiltonian. The first Lanczos step can be found by choosing the optimal alpha to construct $|\psi_\alpha\rangle = (1 + \alpha H)\,|\psi_i\rangle$ where $|\psi_i\rangle$ is the previously optimized wavefunction [22–24]. Calculating the optimal α^* requires minimizing the energy

$$E(\alpha) = \frac{\langle\psi_i|\,(1+\alpha H)H(1+\alpha H)\,|\psi_i\rangle}{\langle\psi_i|\,(1+\alpha H)^2\,|\psi_i\rangle} \tag{D1}$$

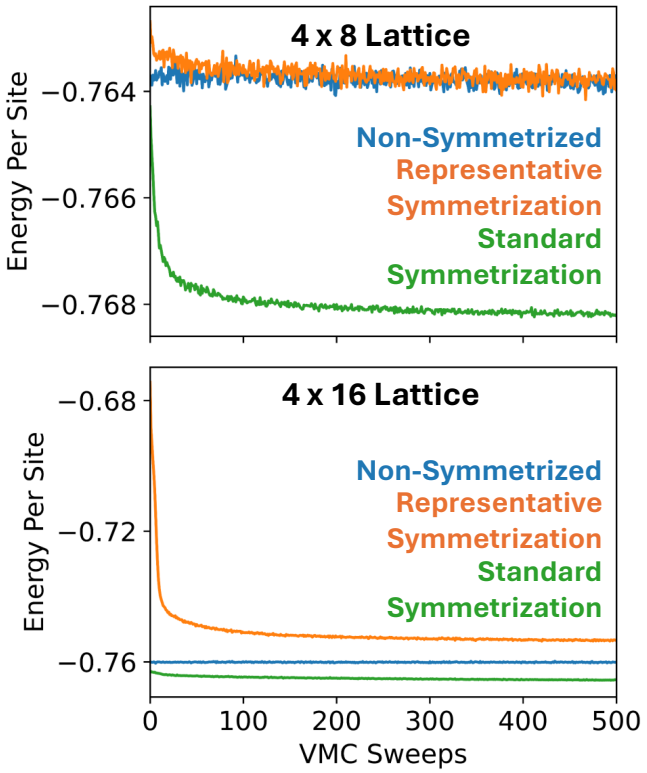

FIG. A2. Energy vs VMC sweep. The representative symmetrization and standard symmetrization are started from the non-symmetrized wavefunction. We see in both the 4×8 ($n_h = 1024$) and 4×16 ($n_h = 2048$) lattices, the representative symmetrization increases the energy, while the standard symmetrization decreases it. Each VMC sweep reports the average energy from 100 VMC steps of each of the 1024 VMC chains.

of the new ansatz. The standard method for calculating a single Lanczos step requires computing the powers of the Hamiltonian

$$h_n = \frac{\langle\psi| H^n |\psi\rangle}{\langle\psi|\psi\rangle} \tag{D2}$$

up to $n = 3$, from which one can analytically solve for α^*.

In practice we implement the algorithm in Appendix C of Ref. [22]. We start by guessing a value of α and computing $E(\alpha)$ from Eq. D1 and

$$\begin{aligned}\chi &= \frac{\langle\psi_\alpha| (1+\alpha H)^{-1} |\psi_\alpha\rangle}{\langle\psi_\alpha |\psi_\alpha\rangle} \\ &= \frac{\langle\psi_\alpha| |\psi_i\rangle}{\langle\psi_\alpha(1+\alpha H) |\psi_i\rangle} \\ &= \langle[1+\alpha E_i(\mathbf{n})]^{-1}\rangle_{|\psi_\alpha|^2}\end{aligned} \tag{D3}$$

where $E_i(\mathbf{n}) = \langle\psi_i| H |\mathbf{n}\rangle / \langle\psi_i |\mathbf{n}\rangle$ is the local energy of the initial wavefunction. From this we can obtain

$$h_2 = [(\chi^{-1} - 2)(1 + \alpha h_1) + 1]/\alpha^2, \tag{D4}$$

and

$$h_3 = [E(\alpha)(1 + 2\alpha h_1 + \alpha^2 h_2) - h_1 - 2\alpha h_2]/\alpha^2, \tag{D5}$$

which can be used to analytically minimize $E(\alpha)$ for the optimal α^*. Note that h_3 is found by sampling an energy expectation value, which is statistically more accurate compared to the direct determination of h_3 from equation D2.

The analytical minimization of $E(\alpha)$ is statistically most accurate as α gets closer to α^*, so we typically iterate quickly by using a small number of samples to estimate $E(\alpha)$ and χ until we are close to α^*. In practice we begin with α near zero, and estimate h_n with 2^{16} samples. We usually find good convergence within 4 iterations, after which we iterate two more times with 2^{18} samples to further lower statistical error. The α^* for the four best wavefunctions in Fig. 2 are listed in Table I.

n_h	α^*
1024	0.03233
2048	0.03875
4096	0.02641
8192	0.02898

TABLE I. Optimized α^* values for different n_h values on the 4×16 Hubbard model. The energies of the resulting wavefunctions $|\psi_{\alpha^*}\rangle = (1 + \alpha^* H) |\psi_i\rangle$ are shown in Fig. 2.

Appendix E: High Energy Observables

As mentioned in Sec. III A, most of the wavefunctions, even prior to enhancement are able to correctly capture the spin and charge density waves. However, the highest energy wavefunction we study, $n_h = 128$ with energy per site of $-0.75066(2)$, does not capture the density waves correctly. As shown in Fig. A3 (c,d), we see a spin density wave with a peak at approximately $\mathbf{k} = (7\pi/16, \pi)$ and a charge density wave with peak at $\mathbf{k} = (3\pi/8, \pi)$. This suggests that a large enough number of neurons is required to capture these structures accurately.

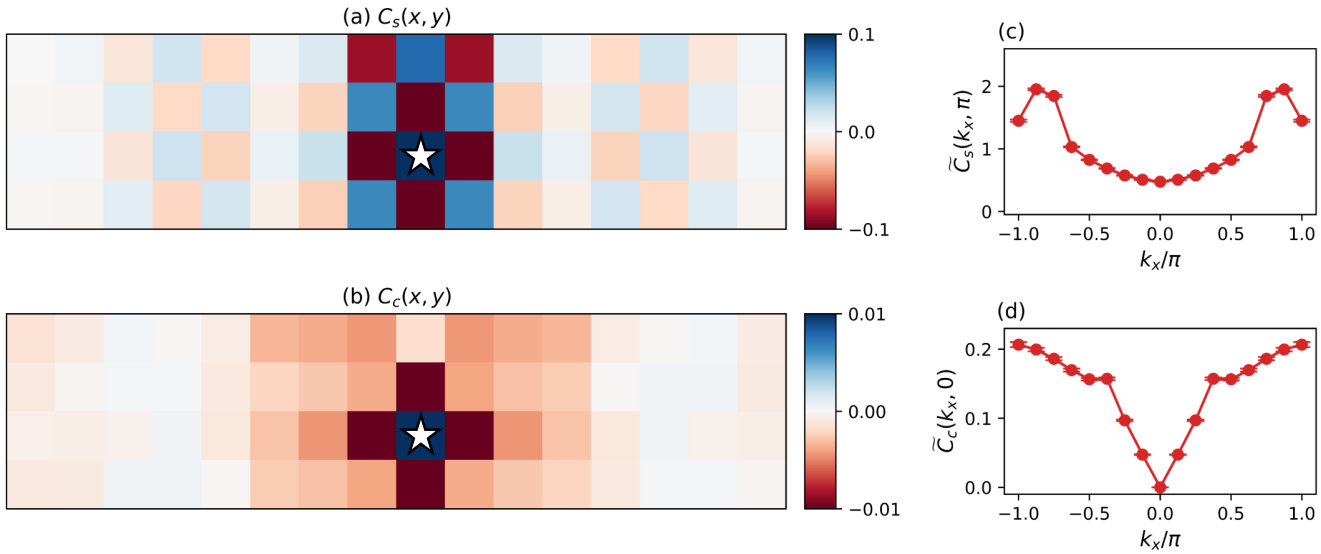

FIG. A3. Spin and charge observables for base NNBF with $n_h = 128$. This ansatz has a high energy per site of $-0.75066(2)$. (a-b) The correlation functions C_s and C_c in real space. The values are truncated such that $|C_s| \leq 0.1$ and $|C_c| \leq 0.01$ to show weaker correlations. The central site is marked with a star. (d-e) The spin and charge structure factors showing cuts at $\widetilde{C}_s(k_x, \pi)$ and $\widetilde{C}_c(k_x, 0)$. Peaks seem to be $(7\pi/16, \pi)$ and $(3\pi/8, 0)$ respectively instead of at the correct values of $(7\pi/8, \pi)$ and $(\pi/4, 0)$ found for more accurate states in Fig. III A.