

# Comparing Symmetrized Determinant Neural Quantum States for the Hubbard Model

Louis Sharma,<sup>1,2,\*</sup> Ahmedeo Shokry,<sup>1,2,3,4,\*</sup> Rajah Nutakki,<sup>1,2,3,4</sup>  
Olivier Simard,<sup>1,2</sup> Michel Ferrero,<sup>1,2</sup> and Filippo Vicentini<sup>1,2,3,4</sup>

<sup>1</sup>*CPHT, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France*

<sup>2</sup>*Collège de France, Université PSL, 11 place Marcelin Berthelot, 75005 Paris, France*

<sup>3</sup>*Inria Paris-Saclay, Bâtiment Alan Turing, 1,  
rue Honoré d'Estienne d'Orves – 91120 Palaiseau*

<sup>4</sup>*LIX, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France*

Accurate simulations of the Hubbard model are crucial to understanding strongly correlated phenomena, where small energy differences between competing orders demand high numerical precision. In this work, Neural Quantum States are used to probe the strongly coupled and underdoped regime of the square-lattice Hubbard model. We systematically compare the Hidden Fermion Determinant State and the Jastrow-Backflow ansatz, parametrized by a Vision Transformer, finding that in practice, their accuracy is similar. We also test different symmetrization strategies, finding that output averaging yields the lowest energies, though it becomes costly for larger system sizes. On cylindrical systems, we consistently observe filled stripes. On the torus, our calculations display features consistent with a doped Mott insulator, including antiferromagnetic correlations and suppressed density fluctuations. Our results demonstrate both the promise and current challenges of neural quantum states for correlated fermions.

## I. INTRODUCTION

The Hubbard model [1, 2] occupies a central role in the understanding of correlated electrons in condensed matter. It is thought to host a broad range of phases of matter from antiferromagnetism, spin- and charge-density waves, to  $d$ -wave superconductivity and even topologically-ordered spin liquids. Much interest stems from its ability to capture essential qualitative properties of experimental correlated quantum systems, such as the high- $T_c$  cuprate [3, 4] and nickelate [5] superconductors or organic charge-transfer crystals  $\kappa$ -(BEDT-TTF)<sub>2</sub> [6–8]. Despite this, even on the square lattice, reaching a consensus on its phase diagram across the entire parameter space (hoppings, interaction strengths and fillings) remains a considerable challenge. Many numerical methods are under active development to tackle this and similar problems, including DMFT [9–11], DMRG [12–14], PEPS [15, 16], quantum Monte Carlo (QMC) [17–21] and determinant expansions [22] with recent notable successes in tackling challenging points of the phase diagram [23–31].

At the same time, the use of neural networks as variational wavefunctions within a variational Monte Carlo framework (neural quantum states) [32, 33], has emerged as a promising method for studying ground state properties of quantum many-body systems and non-equilibrium dynamics [34–39]. In par-

ticular, for ground states of frustrated spin systems, neural quantum states (NQS) have achieved unprecedented accuracy [40–44], building consensus around the nature of ground states in these challenging models.

The method has also been applied to lattice fermionic systems [45–48], with particular focus on the  $t' = 0$  square-lattice Hubbard model in order to benchmark different approaches, such as the neural backflow [49], hidden fermion determinant state [50], hidden fermion Pfaffian state [51] or transformer wavefunction [52]. However, a fair comparison among these architectures does not exist. An advantage of NQS over other methods is the ability to simulate lattices with periodic boundaries, potentially helping to alleviate the difficulties in extrapolating the results of simulations on finite-size systems to the thermodynamic limit.

The aim of this work is to take the lessons learnt from the successful application of NQS to spin systems (neural network architectures, optimization strategies, use of lattice symmetries) and investigate their applicability to fermionic systems. Furthermore, whilst the similarity between the backflow and hidden fermion determinant approaches can be shown explicitly [53], their relative strengths and weaknesses in practice, particularly in combination with lattice symmetries have not been systematically studied. We perform such a study here.

The remainder of the paper is organized as follows. In Sec. II, we present the Hubbard model and in Sec. III the NQS ansätze employed to study its ground state properties. Then, in Sec. IV, we present and discuss results on  $8 \times L$  lattices with

---

\* These two authors contributed equally.

open and periodic boundary conditions.

## II. HUBBARD MODEL

We consider the single-band Fermi-Hubbard model on an  $M$ -site 2d square lattice with nearest neighbor hopping. The Hamiltonian is given by

$$\hat{\mathcal{H}} = -t \sum_{\langle ij \rangle, \sigma} \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (1)$$

where  $t$  is the hopping matrix element between nearest neighbor sites  $\langle ij \rangle$ ,  $\sigma = \{\uparrow, \downarrow\}$  is the spin index,  $\hat{c}_{i\sigma}^\dagger$  and  $\hat{c}_{i\sigma}$  are fermionic operators that add or remove electrons from the single particle states  $|i\sigma\rangle$ . They satisfy the canonical anticommutation relations

$$\{\hat{c}_{i\sigma}, \hat{c}_{j\sigma'}^\dagger\} = \delta_{ij} \delta_{\sigma\sigma'}, \quad \{\hat{c}_{i\sigma}, \hat{c}_{j\sigma'}\} = 0. \quad (2)$$

Finally,  $U$  is the on-site Coulomb repulsion and  $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^\dagger \hat{c}_{i\sigma}$  is the number operator. In this study, we work in sectors of the Hilbert space corresponding to a fixed particle number  $N = N_\uparrow + N_\downarrow$  and fixed magnetization  $2S^z = N_\uparrow - N_\downarrow$ . The corresponding electron (hole) density is  $n = N/M$  ( $n_h = 1 - n$ ).

We focus on the Hubbard model at strong coupling ( $U/t = 8$ ) and hole-doping  $n_h = 1/8$ , a particular challenge due to the severity of the fermionic sign problem [54], non-perturbative parameter regime and presence of competing ground states [2]. At half-filling, the absence of a sign-problem means that QMC can provide numerically exact solutions, showing that the ground state is a Néel antiferromagnet [24, 55]. Upon doping, there is competition between a uniform d-wave superconducting state and stripe states. In a stripe state holes cluster together along one direction, forming a line defect in the antiferromagnetic spin background and a periodic modulation in hole density with wavelength  $\lambda$  [56]. Several results [2, 24, 25, 31, 52, 57–62] find a  $\lambda = 8$  filled stripe (with periodicity  $2\lambda$  in the spin correlations) as the ground state for  $U/t = 8, n_h = 1/8$ .

## III. METHOD

### A. Determinant states for fermionic wavefunctions

A generic second-quantized many-body wavefunction can be written as

$$|\psi\rangle = \sum_{\mathbf{n}} \psi(\mathbf{n}) |\mathbf{n}\rangle \quad (3)$$

where  $|\mathbf{n}\rangle$  denotes states in the Fock basis. A class of fermionic NQS developed in the past few years are those based on determinants, having the advantage of a straightforward connection to a mean-field Slater determinant and indications that this structure is advantageous for obtaining strongly-correlated ground states [63].

Whilst other types of fermionic NQS ansätze exist, such as those based on Pfaffians [51, 64], our focus is on the determinant states which we define in the following.

**Neural Network Backflow:** Originally introduced in Ref. [49], the neural network backflow approach parameterizes an additive transformation of the single-particle orbitals, which we use to construct a linear combination of Slater determinants,

$$\psi_{\text{Bf}}(\mathbf{n}) = \sum_k C_k \prod_{\sigma} \det \left( [\Phi^{\sigma, k}(\mathbf{n})]_{\mathbf{n}} \right). \quad (4)$$

where  $\Phi^{\sigma, k}$  is an  $M \times N_{\sigma}$  matrix and  $[\Phi^{\sigma}(\mathbf{n})]_{\mathbf{n}}$  denotes the matrix obtained by selecting the  $N_{\sigma}$  rows corresponding to the occupied orbitals in the configuration  $\mathbf{n}$ , where  $N_{\sigma}$  is the number of fermions with spin  $\sigma$ . The coefficients  $C_k$  are free parameters. The matrix entries  $\Phi_{ij}^{\sigma}$  are given by the sum of a “mean-field” orbital and an input-dependent correction

$$\Phi_{ij}^{\sigma}(\mathbf{n}) = \Phi_{ij}^{\sigma, (0)} + \tilde{\Phi}_{ij}^{\sigma}(\mathbf{n}). \quad (5)$$

Here,  $\Phi_{ij}^{\sigma, (0)}$  is the  $i$ -th variational mean-field orbital evaluated at the position of the  $j$ -th particle with spin  $\sigma$ , while  $\tilde{\Phi}_{ij}^{\sigma}(\mathbf{n})$  is the input-dependent correction obtained from a neural network.

We also include a Jastrow density-density correlation term.

$$J(\mathbf{n}) = \exp \left( -\frac{1}{2} \sum_{i,j} n_i W_{ij} n_j \right), \quad (6)$$

with  $W$  an upper triangular matrix. The full variational wavefunction is then written as

$$\psi_{\text{JBf}}(\mathbf{n}) = J(\mathbf{n}) \cdot \psi_{\text{Bf}}(\mathbf{n}), \quad (7)$$

which we call the Jastrow-Backflow (JBf). The set of variational parameters includes those of the neural network backflow, the mean-field orbitals, the Jastrow matrix  $W$ , and the Slater coefficients  $C_k$ .

**Hidden Fermion Determinant States:** Instead of adding a many-body correction to non-interacting orbitals, hidden fermion determinant states (HFDS) augment the Fock space of  $N$  fermions with  $M$

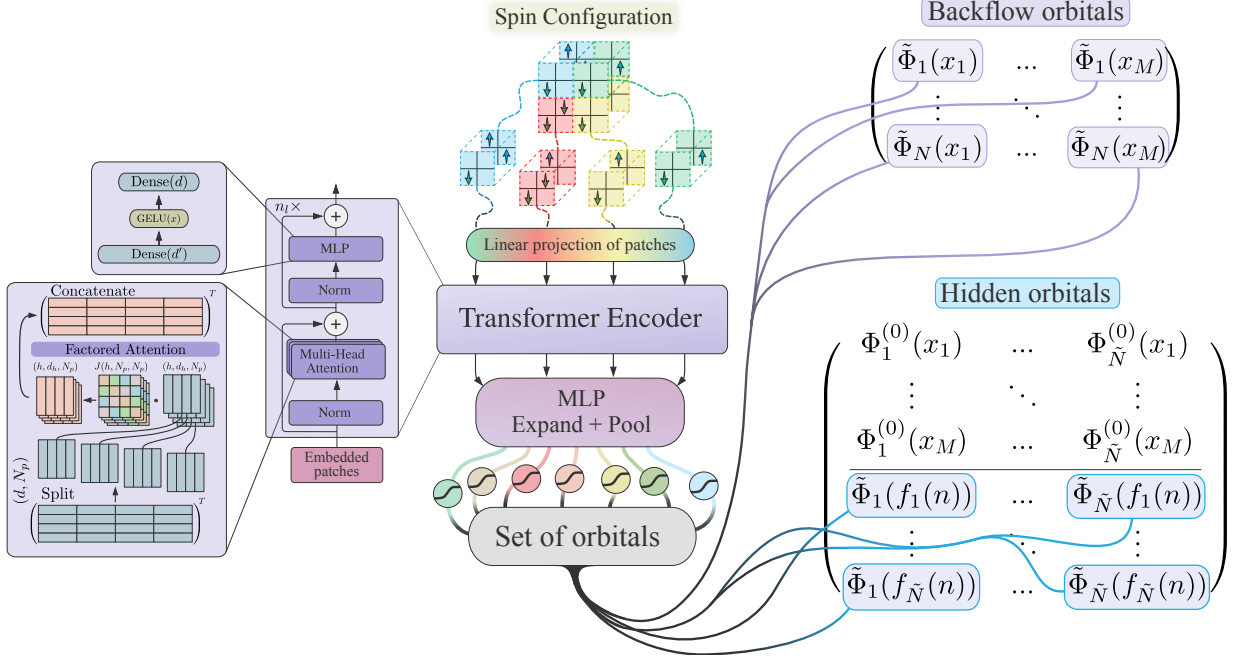


FIG. 1: Schematic of the ViT architecture adapted to the HFDS and JBf fermionic wavefunctions.

modes by adding  $\tilde{N}$  hidden fermions that occupy  $\tilde{M}$  hidden modes. The physical wavefunction amplitudes  $\psi(\mathbf{n})$  are then obtained by projecting the state in the augmented space  $|\Phi\rangle$  assuming that the occupancies of the hidden fermions depend on those of the physical ones. In this work, we consider the case where  $|\Phi\rangle$  is a Slater determinant. Therefore, the physical wavefunction is expressed as

$$\psi(\mathbf{n}) = \prod_{\sigma} \det \left( [\Phi^{\sigma, (0)}]_{\mathbf{n}} + \tilde{\Phi}^{\sigma}(\mathbf{n}) \right), \quad (8)$$

where  $[\Phi^{\sigma, (0)}]_{\mathbf{n}}$  are  $(N_{\sigma} + \tilde{N}) \times (N_{\sigma} + \tilde{N})$  matrices whose last  $\tilde{N}$  rows are zero. They are obtained by slicing the  $N_{\sigma}$  rows of the input-independent  $M \times (N_{\sigma} + \tilde{N})$  *visible* matrices that correspond to the occupancies of  $|\mathbf{n}\rangle$ . The  $\tilde{\Phi}^{\sigma}(\mathbf{n})$  terms are  $(N_{\sigma} + \tilde{N}) \times (N_{\sigma} + \tilde{N})$  matrices whose first  $N_{\sigma}$  rows are zero. They are the output of a neural network evaluated on  $|\mathbf{n}\rangle$ . Additional details regarding the HFDS can be found in appendix A.

## B. Vision Transformer (ViT)

Both ansätze discussed in the previous section rely on a neural network to parameterize the configuration-dependent matrix  $\tilde{\Phi}(\mathbf{n})$ . Originally introduced for sequence modelling [65], transformers, later extended to computer vision (vision trans-

formers) [66], have been successfully applied as NQS for quantum spin systems [43, 67], Rydberg atom arrays [68] and fermionic systems [52, 69]. Here, we apply the vision transformer (ViT) from Ref. [67] as an NQS by using it to parameterize  $\tilde{\Phi}(\mathbf{n})$  for both ansätze. The ViT is made up of three stages: an embedding, encoder and readout.

Configurations from the computational basis of fermionic systems are numerically encoded as  $N_o \times (2S + 1)$  bit strings [70], where  $N_o$  is the number of spatial orbitals and  $S$  is the spin. Each bit of this string specifies the binary occupation of a spin-orbital. Making a connection with computer vision, we interpret the orbitals as the spatial degree of freedom of an image and the different spin values as different channels (features) of the same image. We use this channel-based representation as the starting point for evaluating  $\tilde{\Phi}(\mathbf{n})$ .

First, the input is partitioned into patches [66, 67], which both lowers the computational cost of evaluating the network and acts as a form of inductive bias [71]. As illustrated in Fig. 1, the spin-channel representation of  $\mathbf{n}$  is split into  $N_p$  patches of size  $p = b \times b$ , where  $b$  is the side length of the patch. Together with the spin channels this results in  $N_p$  vectors of size  $2p$ , which are then embedded into a higher-dimensional space of dimension  $d$ .

After the embedding stage, the input is processed through an encoder which employs a factored attention mechanism [67, 72], where the input-dependent

attention matrix is replaced by a fixed, learnable parameter matrix  $\mathbf{J}$ . The result is a  $d$ -dimensional correlated representation that incorporates global contextual information across patches. Attention heads  $h$  are used to introduce parameter-sharing in  $\mathbf{J}$ , and the encoding block is repeated  $n_l$  times.

Finally, the output is pooled over the patch dimension, resulting in a  $d$ -dimensional representation. This is passed through a dense layer with  $d'$  hidden units and a GELU activation. A final dense layer followed by a tanh activation produces the matrices  $\tilde{\Phi}(\mathbf{n})$ .

This fundamental structure is shared by both the HFDS-ViT and JBf-ViT architectures, which can be thought of as functions designed to generate antisymmetric outputs. The primary distinction between these two models lies in their readout shape. The HFDS-ViT readout is a set of hidden matrices with shape  $\tilde{N} \times (N_\sigma + \tilde{N})$  whilst the JBf readout are  $M \times N_\sigma$  matrices. This leads to a crucial difference in computational efficiency. On one hand, when  $\tilde{N}$  is fixed, the HFDS-ViT scales linearly with the system size. On the other, the JBf-ViT scales quadratically.

### C. Symmetries

When the Hamiltonian commutes with a set of operators that obey a group structure, it can be block diagonalized [73]. Each block corresponds to a particular irreducible representation (irrep) of the group. Consequently, the eigenstates of the Hamiltonian can be classified by their irrep [74]. It has been shown that enforcing Hamiltonian symmetries on a variational wavefunction can significantly improve variational energies [50, 75–80]. In general, an unrestricted wavefunction will not exactly belong to a specific irrep after optimization. This symmetry may be restored by using quantum number projection [78, 81]. This is done using

$$\psi_\mu(\mathbf{n}) = \frac{1}{|G|} \sum_{g \in G} \chi_\mu^*(g) \xi_{g^{-1}}(\mathbf{n}) \psi(\mathbf{n} \circ g), \quad (9)$$

where  $\mu$  is the index of the irreducible representation,  $|G|$  is the order of the group,  $\chi_\mu(g)$  is the character of the irreducible representation evaluated on  $g$ ,  $\xi_g(\mathbf{n})$  is the sign of the permutation applied to state  $|\mathbf{n}\rangle$ , and  $\mathbf{n} \circ g$  are the permuted occupation numbers. More details concerning the representation of permutation groups on fermionic Fock space are given in appendix B.

Equation (9) can be implemented using a brute-force  $O(|G|)$  approach, with every amplitude query  $\psi_\mu(\mathbf{n})$  requiring  $|G|$  neural-network evaluations. However, when the order of the group scales with

the lattice size, as is the case for the translation group, this leads to an  $O(M)$  overhead which can be prohibitively expensive for large lattices.

To avoid multiple network evaluations, it is possible to design attention matrices that produce patch-translation equivariant attention vectors [67][82]. To do so, the attention weights need to satisfy  $J_{ij}^\mu = J_{T(i)T(j)}^\mu$ , a derivation of this condition can be found in Section C.

When the output of the last encoder block is summed over the patch dimension, this results in a patch-translation-invariant output. In order to obtain an output that is invariant with respect to the full translation group of the lattice, the output is pooled over the intra-patch translation vectors  $\mathbf{p}$ ,

$$\tilde{\Phi}(\mathbf{n}) = \sum_{\mathbf{p}} f_{\text{ViT}}(\mathbf{n} \circ T_{\mathbf{p}}). \quad (10)$$

Here, we consider a square lattice which is partitioned into  $2 \times 2$  patches, so these vectors are  $\mathbf{p} \in \{\mathbf{0}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_1 + \mathbf{a}_2\}$  where  $\mathbf{a}_i$  are the primitive lattice vectors. This construction leads to  $\tilde{\Phi}(\mathbf{n}) = \tilde{\Phi}(\mathbf{n} \circ T_{\mathbf{R}})$  for all lattice vectors  $\mathbf{R}$ . Consequently, a translationally-symmetrized wavefunction with momentum  $\mathbf{k}$  is given by

$$\psi_{\mathbf{k}}(\mathbf{n}) = \frac{1}{M} \sum_{\mathbf{R}} e^{-i\mathbf{k} \cdot \mathbf{R}} \xi_{-\mathbf{R}}(\mathbf{n}) \times \det \left( [\Phi^{(0)}]_{\mathbf{n} \circ T_{\mathbf{R}}} + \tilde{\Phi}(\mathbf{n}) \right), \quad (11)$$

where  $\mathbf{k}$  is a reciprocal vector in the first Brillouin zone. Equation 11 produces translationally-invariant amplitudes, in which the ViT only needs to be evaluated on the vectors  $\mathbf{p}$  [46]. The number of such vectors remains constant with the size of the system.

We denote the translationally invariant wavefunction produced from equation 11 as **Symm ViT-Ti** and the one produced from equation 9 as **Symm ViT**.

## IV. RESULTS

In this section, we present results obtained using the symmetrization procedures described above, applied to both the HFDS and JBf wavefunctions, each parametrized by a ViT. We focus on the  $8 \times L$  square lattice where  $L \in \{4, 8\}$  using both open and periodic boundary conditions. In the simulations, we fix  $d = 64$ ,  $h = 8$  and  $d' = 128$ . The wavefunctions are optimized using the kernel reformulation [41, 83] of stochastic reconfiguration [84]. The learning rate is gradually annealed using a cosine decay scheduler [85] whilst the diagonal shift of the neural tangent

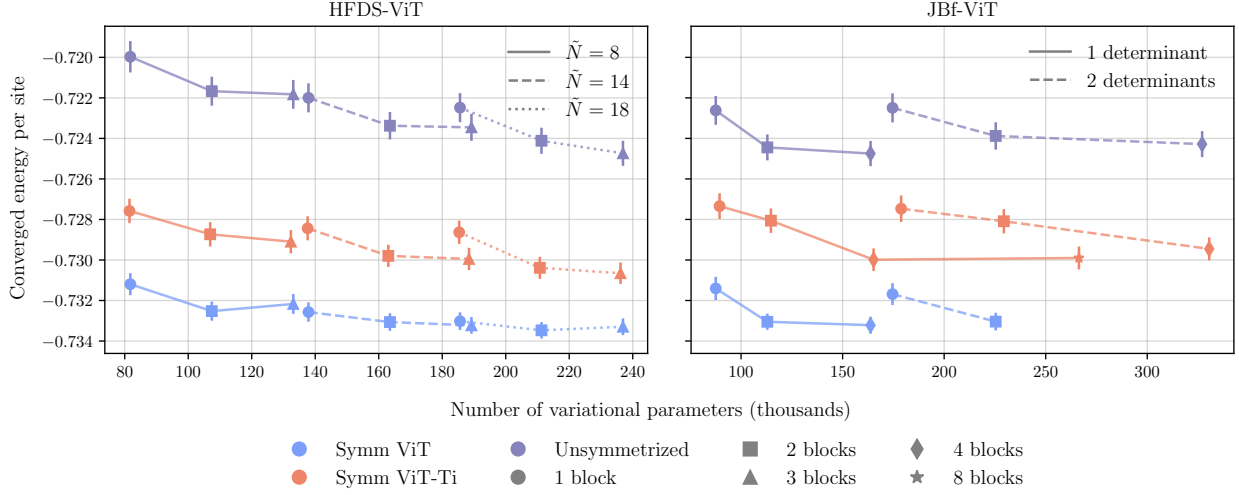


FIG. 2: Comparison of variational energies on a the  $8 \times 4$  cylinder. The plot shows converged energies for both architectures as a function of the number of variational parameters and symmetrization methods. For HFDS-ViT (left), the number of hidden fermions  $\tilde{N}$  is varied. For JbF-ViT, the number of determinants is varied. Symbols indicate different numbers of stacked encoder blocks,  $n_l$ , in the ViT.

kernel is linearly decreased from  $10^{-2}$  to  $10^{-8}$ . We do not use a pinning field and start from a set of random variational parameters.

#### A. $8 \times 4$ cylinder (OBC-PBC)

First, we consider an  $8 \times 4$  lattice with open boundary conditions along the long side and periodic boundary conditions along the short side.

To investigate the effect of the number of parameters in the wavefunction, we systematically increased the number of hidden fermions  $\tilde{N}$  for the HFDS ansätze, the number of determinants for the JbF wavefunction and the number of encoder blocks for both architectures, as shown in Figure 2. Increasing the number of parameters by increasing the number of encoder blocks tends to improve the variational energy, although the improvement saturates at higher values. While increasing the number of hidden fermions in the HFDS also tends to improve variational energies, this is not necessarily the case for using a second determinant in the JbF, with the improvement in variational energy small compared to the cost in terms of number of parameters. In what follows we fix  $n_l = 2$ ,  $\tilde{N} = 18$  and use a single-determinant in the JbF.

To investigate the accuracy gains of the symmetrization procedure, we also compare both symmetrized and unsymmetrized versions of the ansätze in Figure 2. The symmetrized wavefunctions are projected onto irreps of the space group and spin-

parity group. Note that in the case of the Symm ViT-Ti, one can construct a wavefunction that is invariant with respect to the space group whilst only evaluating the network on the point group, whose order does not scale with the system size. For both the JbF and HFDS ansätze, we find that the lowest energy state corresponds to the same non-trivial irrep of the space group with momentum  $k_y = \pi$ .

We find that symmetrized wavefunctions significantly outperform the non symmetrized ones in terms of energy, with the Symm-ViT achieving the lowest energy. This suggests that enforcing symmetry through output averaging, rather than by restricting the model's internal parameters offers a more expressive and accurate approach, consistent with findings in [79].

We also investigated how each approach responds to increasing the number of Monte Carlo samples during optimization. Specifically, we selected the best hyperparameters identified in the previous analysis and re-optimized them using 2048, 4096, 8192 and 16384 samples. The results are presented in Fig. 3. Among the three approaches, the non-symmetrized ViT shows the most significant improvement in energy as the number of samples increases. In contrast, both ViT-Ti and symm-ViT show a more gradual decrease in energy. This suggests that the symmetrized models may already be close to their optimal performance even at lower sample counts, while the unsymmetrized ViT requires more samples to effectively optimize.

To investigate the presence of charge and spin or-

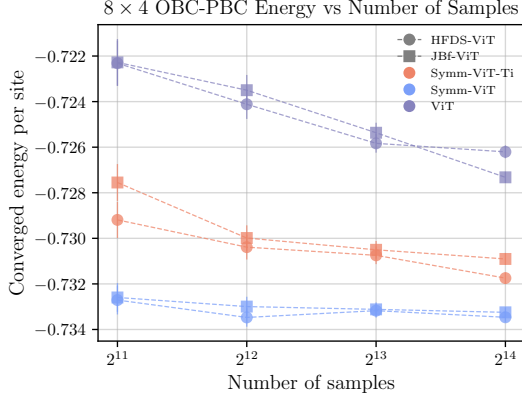


FIG. 3: Comparison of variational energies on the  $8 \times 4$  cylinder. The plot shows converged energies for the presented architectures as a function of the number of Monte Carlo samples and symmetrization methods.

ders, we compute the spin-spin correlation function  $S_{ij} = \langle \hat{S}_i^z \hat{S}_j^z \rangle$  and the hole density  $h_i = 1 - \langle \hat{n}_i \rangle$ , where the expectation value is taken with respect to the ground state and computed stochastically [86]. Plots of these observables for both the JbF and HFDS are presented in Figure 4.

The spin-spin correlation function and hole density exhibit a consistent pattern across all three approaches. For both the HFDS and JbF ansätze, we observe regions of high hole concentration that align with domain walls in the antiferromagnetic order. This is consistent with the  $\lambda = 8$  filled stripe discussed in Section II. While the overall structure of the correlations is similar across the different architectures, we note small differences in the amplitude of the spin-spin signal and the hole density, reflecting subtle variations in how each network captures charge and magnetic ordering.

### B. $8 \times 8$ torus (PBC-PBC)

We now turn to the  $8 \times 8$  square lattice, with periodic boundary conditions in both directions. In this case, we fix the network hyperparameters to  $d = 64$ ,  $h = 8$ ,  $n_l = 2$ , 4096 samples,  $N = 22$  and one determinant in the JbF. For the HFDS, we use  $d' = 128$  whereas for the JbF the latent dimension of the read-out is set  $d' = 64$ , which results in a similar number of parameters. Due to the favorable scaling of the HFDS, we can use a higher  $d'$  for the same number of parameters. First, several identical Symm-ViT wavefunctions are initialized and optimized with respect to the translation group for the  $\mathbf{k}$  points of the irreducible Brillouin zone. After identifying the low-

est energy  $\mathbf{k}$ -point, the optimization is continued in that momentum sector until convergence. For both the JbF-ViT and HFDS-ViT ansätze, we found that  $\mathbf{k} = (\pi/2, \pi/2)$  produces the lowest energy.

To study the magnetic and electronic properties of the converged solution, we compute the spin-spin correlation, the density-density correlation,

$$\rho_{ij} = \langle \hat{n}_i \hat{n}_j \rangle - \langle \hat{n}_i \rangle \langle \hat{n}_j \rangle \quad (12)$$

as well as the associated structure factors. The structure factor  $\tilde{A}_{\mathbf{k}}$  of a translationally-invariant two-point correlation function  $A_{ij}$  is given by

$$\tilde{A}_{\mathbf{k}} = \sum_i e^{i\mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j)} A_{ij}, \quad (13)$$

where  $\mathbf{R}_i$  is position of site  $i$  and  $\mathbf{k}$  a vector in the first Brillouin zone.

Both methods display a similar charge and spin ordering pattern (see Fig 5). The spin-spin signal exhibits long-range antiferromagnetic correlations, characterized by sharp peaks at  $\mathbf{k} = (\pi, \pi)$  in the structure factor. In contrast, the density-density correlations are short ranged and uniform, indicating suppressed density fluctuations. These results are consistent with the physics of a doped Mott insulator. This state is distinct from the  $\lambda = 8$  filled stripe phase, as the required  $2\lambda = 16$  spin density periodicity is incompatible with the  $8 \times 8$  torus.

A similar calculation was performed on this system but with open boundary conditions in one direction. The results on Fig. 4 show that  $\lambda = 8$  stripes are found.

We also explored the performance of the ViT and ViT-Ti architectures on this larger system. Using the same neural network hyperparameters, we trained all models to convergence. For ViT-Ti, we focused on the momentum sector previously identified as energetically favorable, and compared the final energies as well as the physical observables. Surprisingly, the unsymmetrized wavefunction outperforms the ViT-Ti, obtaining an energy lower by about 1.2% see I. A possible explanation for this discrepancy is that enforcing full translational equivariance on the attention weights imposes an overly restrictive constraint on the variational wavefunction. In addition, the unsymmetrized wavefunctions struggle to capture the spin-spin correlation function.

## V. CONCLUSION AND OUTLOOK

In this work, we used the ViT architecture to represent the backflow orbitals of the JbF wavefunction and the hidden orbitals of the HFDS. Our simulations of the  $8 \times L$  Hubbard model, with  $L \in \{4, 8\}$ , in

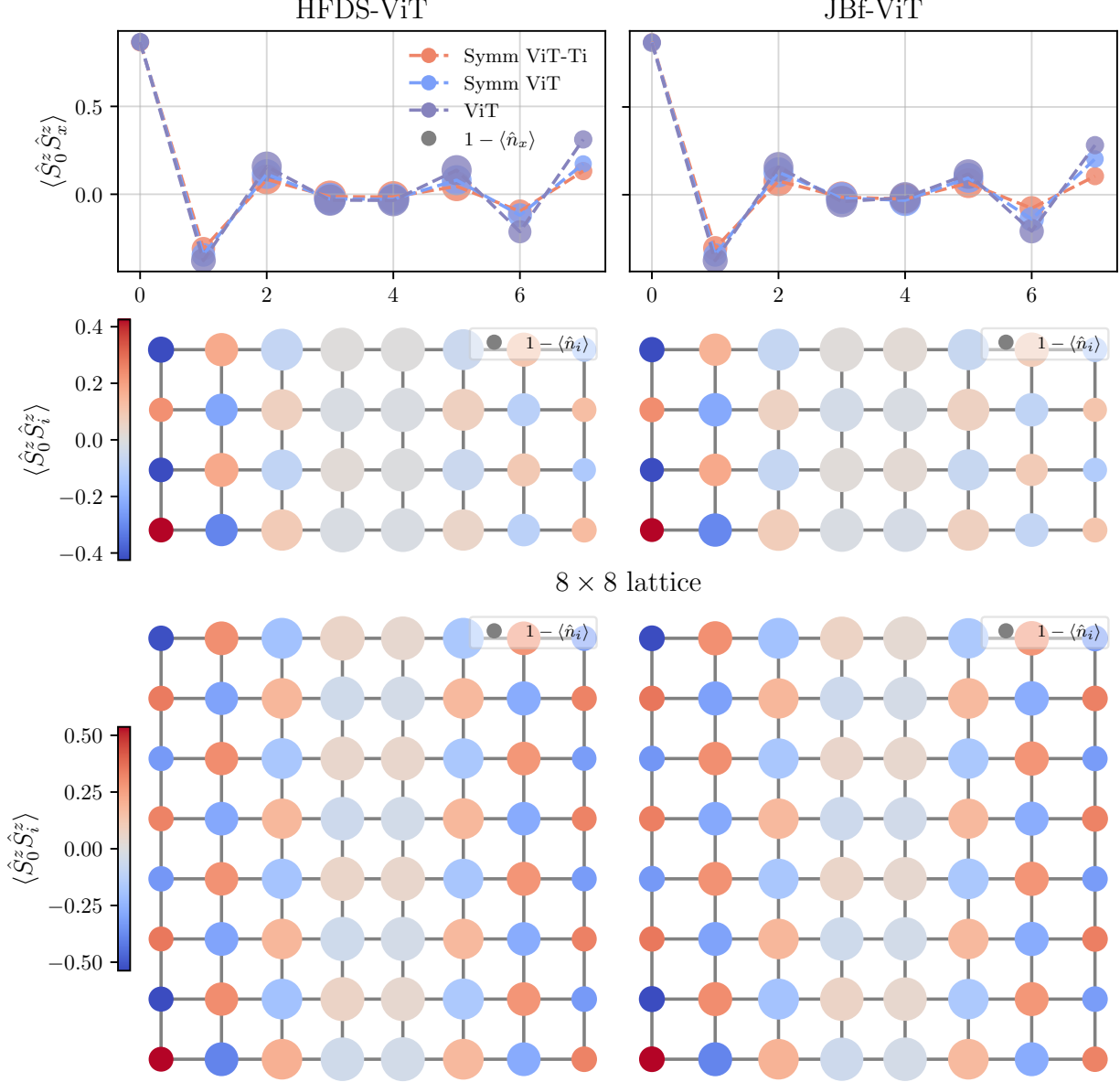


FIG. 4: Comparison of physical observables on the  $8 \times 4$  and  $8 \times 8$  cylinders. The first row shows a horizontal cut of the spin-spin correlation. The  $x$  axis corresponds to the site index and the size of the balls is proportional to the hole density at that site. The central (bottom) row plots show the observables in real space for the lowest energy solution on the  $8 \times 4$  ( $8 \times 8$ ) system.

the strongly correlated, underdoped regime, provide a systematic comparison between these two ansätze.

Our results show that HFDS and JBf wavefunctions achieve comparable variational energies and observables in practice. Although previous studies reported differing performance between the two, we attribute these discrepancies to differences in network architecture, size, and training protocols [49, 50], in line with the findings of Ref. [53]. A key dis-

tinguishing factor arises in scalability: HFDS scales more favorably with system size, as its readout depends on a fixed number of hidden fermions, potentially providing an advantage for simulating larger systems.

We find that enforcing symmetries of the Hamiltonian is crucial in order to obtain lower energies and physically-meaningful observables. Our results show that explicitly imposing symmetries through computationally expensive output averaging outper-



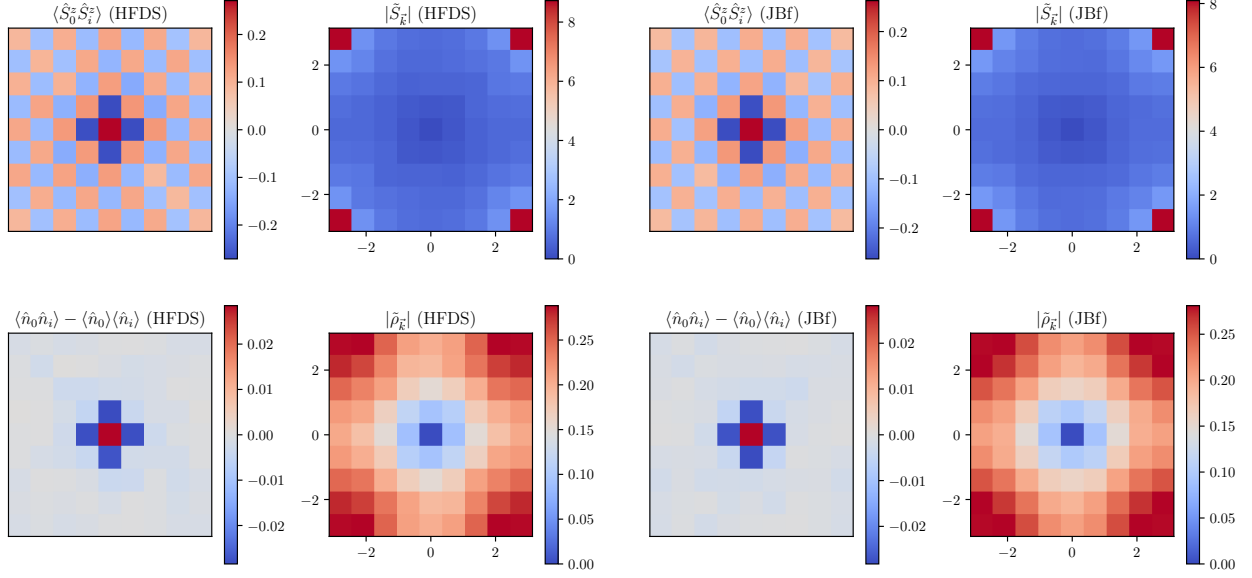


FIG. 5: Comparison of physical observables on the  $8 \times 8$  square lattice (PBC-PBC) with  $U/t = 8$  and  $n_h = 1/8$  hole doping. The results show real space plots of the expectation values of the of spin-spin correlation function, the density-density correlation and the associated structure factors for both ansätze. The expectation values are computed from the converged Symm-ViT variational states, the symmetrization being with respect to the translation group in the  $\mathbf{k} = (\pi/2, \pi/2)$  which was identified as the lowest energy sector.

forms architectures that enforce constraints more efficiently [46].

Finally, both methods converged to the expected  $\lambda = 1/n_h$  filled stripes on cylindrical systems, in line with recent results in the literature [2, 31, 52]. On the  $8 \times 8$  torus, we find physics which is characteristic of a doped Mott-insulator, marked by long range antiferromagnetic correlations and localized, repulsive density fluctuations. We are however not able to access system sizes large enough to resolve the presence of incommensurate spin and charge ordering. Therefore, our results may be impacted by significant finite-size effects

This highlights a crucial challenge for future research, developing methods that incorporate symmetries in an effective and scalable manner. Overcoming this hurdle is essential for probing larger system sizes, in order to obtain better resolution of ground state properties and reliably extrapolating to the thermodynamic limit.

Our results suggest that a path forward for efficiently simulating larger systems could be achieved

by using the HFDS with a reduced number of hidden fermions allowing one to adopt low-rank determinant updates in the spirit of Ref. [51].

## VI. ACKNOWLEDGEMENTS

Simulations were performed with NetKet [87, 88], and at times parallelized with mpi4JAX [89]. This software is built on top of JAX [90] and Flax [91]. We acknowledge insightful discussions with G. Carleo and A. Kahn. F.V. acknowledges support by the French Agence Nationale de la Recherche through the NDQM project ANR-23-CE30-0018. O.S. acknowledges support from the Swiss National Science Foundation through the Postdoc.Mobility fellowship. This project was provided with computing HPC and storage resources by GENCI at IDRIS thanks to the grant 2024-A0170515698 on the supercomputer Jean Zay's A100 partition. We acknowledge EuroHPC Joint Undertaking for awarding us access to Leonardo at CINECA, Italy through grant EHPC-AI-2024A05-006 .



- [2] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, The Hubbard Model: A Computational Perspective, *Annual Review of Condensed Matter Physics* **13**, 275 (2022).
- [3] J. G. Bednorz and K. A. Müller, Possible high  $t_c$  superconductivity in the ba-la-cu-o system, *Zeitschrift für Physik B Condensed Matter* **64**, 189 (1986).
- [4] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen, From quantum matter to high-temperature superconductivity in copper oxides, *Nature* **518**, 179 (2015).
- [5] M. Kitatani, L. Si, O. Janson, R. Arita, Z. Zhong, and K. Held, Nickelate superconductors—a renaissance of the one-band hubbard model, *npj Quantum Materials* **5**, 10.1038/s41535-020-00260-y (2020).
- [6] Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, Spin liquid state in an organic mott insulator with a triangular lattice, *Phys. Rev. Lett.* **91**, 107001 (2003).
- [7] S. Yamashita, Y. Nakazawa, M. Oguni, Y. Oshima, H. Nojiri, Y. Shimizu, K. Miyagawa, and K. Kanoda, Thermodynamic properties of a spin-1/2 spin-liquid state in a  $\kappa$ -type organic salt, *Nature Physics* **4**, 459 (2008).
- [8] M. Yamashita, N. Nakata, Y. Kasahara, T. Sasaki, N. Yoneyama, N. Kobayashi, S. Fujimoto, T. Shibauchi, and Y. Matsuda, Thermal-transport measurements in a quantum spin-liquid state of the frustrated triangular magnet  $\kappa$ -(bedt-ttf) $_2$ cu $_2$ (cn) $_3$ , *Nature Physics* **5**, 44 (2009).
- [9] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [10] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Quantum cluster theories, *Rev. Mod. Phys.* **77**, 1027 (2005).
- [11] G. Rohringer, H. Hafermann, A. Toschi, A. A. Katanin, A. E. Antipov, M. I. Katsnelson, A. I. Lichtenstein, A. N. Rubtsov, and K. Held, Diagrammatic routes to nonlocal correlations beyond dynamical mean field theory, *Rev. Mod. Phys.* **90**, 025003 (2018).
- [12] S. R. White, Density-matrix algorithms for quantum renormalization groups, *Phys. Rev. B* **48**, 10345 (1993).
- [13] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Annals of Physics* **326**, 96 (2011).
- [14] E. M. Stoudenmire and S. R. White, Studying Two Dimensional Systems With the Density Matrix Renormalization Group, *Annual Review of Condensed Matter Physics* **3**, 111 (2012), arXiv:1105.1374 [cond-mat].
- [15] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, Fermionic projected entangled pair states, *Physical Review A* **81**, 052338 (2010).
- [16] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, *Reviews of Modern Physics* **93**, 045003 (2021).
- [17] N. V. Prokof'ev and B. V. Svistunov, Polaron problem by diagrammatic quantum monte carlo, *Phys. Rev. Lett.* **81**, 2514 (1998).
- [18] G. Sugiyama and S. E. Koonin, Auxiliary Field Monte-Carlo for Quantum Many-Body Ground States, *Annals of Physics* **168**, 1 (1986).
- [19] S. Zhang, *Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems*, edited by E. Pavarini, E. Koch, and U. Schollwöck, Schriften Des Forschungszentrums Jülich. Reihe Modeling and Simulation, Vol. Emergent phenomena in correlated matter: lecture notes of the Autumn School Correlated Electrons 2013: at Forschungszentrum Jülich, 23-27 September 2013 (Forschungszentrum, Zentralbibliothek, Jülich, 2013).
- [20] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems*, 1st ed. (Cambridge University Press, Cambridge, 2017).
- [21] E. Kozik, K. Van Houcke, E. Gull, L. Pollet, N. Prokof'ev, B. Svistunov, and M. Troyer, Diagrammatic Monte Carlo for correlated fermions, *EPL (Europhysics Letters)* **90**, 10004 (2010).
- [22] C. Giuliani, J. Nys, R. Martinazzo, G. Carleo, and R. Rossi, Precise quantum chemistry calculations with few slater determinants, arXiv 10.48550/ARXIV.2503.14502 (2025).
- [23] J. P. 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.*, Solutions of the two-dimensional hubbard model: Benchmarks and results from a wide range of numerical algorithms, *Physical Review X* **5**, 041041 (2015).
- [24] 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).
- [25] 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).
- [26] O. Simard, C.-D. Hébert, A. Foley, D. Sénéchal, and A.-M. S. Tremblay, Superfluid stiffness in cuprates: Effect of mott transition and phase competition, *Phys. Rev. B* **100**, 094506 (2019).
- [27] 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).
- [28] B. Xiao, Y.-Y. He, A. Georges, and S. Zhang, Temperature dependence of spin and charge orders in the doped two-dimensional hubbard model, *Phys. Rev. X* **13**, 011007 (2023).
- [29] 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, *Science* **384**, eadh7691 (2024).

- [30] F. Šimkovic, R. Rossi, A. Georges, and M. Ferrero, Origin and fate of the pseudogap in the doped hubbard model, *Science* **385**, eade9194 (2024).
- [31] W.-Y. Liu, H. Zhai, R. Peng, Z.-C. Gu, and G. K. Chan, Accurate simulation of the hubbard model with finite fermionic projected entangled pair states, arXiv preprint arXiv:2502.13454 (2025).
- [32] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [33] F. Vicentini, Machine learning toolbox for quantum many body physics, *Nature Reviews Physics* **3**, 156–156 (2021).
- [34] L. Gravina, V. Savona, and F. Vicentini, Neural projected quantum dynamics: a systematic study, *Quantum* **9**, 1803 (2025).
- [35] M. Schmitt, M. M. Rams, J. Dziarmaga, M. Heyl, and W. H. Zurek, Quantum phase transition dynamics in the two-dimensional transverse-field ising model, *Science Advances* **8**, 10.1126/sciadv.abl6850 (2022).
- [36] A. Sinibaldi, D. Hendry, F. Vicentini, and G. Carleo, Time-dependent neural galerkin method for quantum dynamics (2024), arXiv:2412.11778.
- [37] E. Parnes, N. Barnea, G. Carleo, A. Lovato, N. Rocco, and X. Zhang, Nuclear responses with neural-network quantum states (2025), arXiv:2504.20195 [nucl-th].
- [38] J. Nys, G. Pescia, A. Sinibaldi, and G. Carleo, Ab-initio variational wave functions for the time-dependent many-electron schrödinger equation, *Nature Communications* **15**, 10.1038/s41467-024-53672-w (2024).
- [39] J. Nys, Z. Denis, and G. Carleo, Real-time quantum dynamics of thermal states with neural thermofields, *Physical Review B* **109**, 10.1103/physrevb.109.235120 (2024).
- [40] 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, *Physical Review X* **11**, 031034 (2021).
- [41] A. Chen and M. Heyl, Empowering deep neural quantum states through efficient optimization, *Nature Physics* **20**, 1476 (2024).
- [42] 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).
- [43] 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, *Physical Review B* **111**, 134411 (2025).
- [44] T. Duric, 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 (2024), arXiv:2401.02866 [cond-mat].
- [45] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, *Physical Review B* **96**, 205152 (2017).
- [46] 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).
- [47] H. Lange, A. Böhler, C. Roth, and A. Bohrdt, Simulating the two-dimensional  $tj$  model at finite doping with neural quantum states, arXiv preprint arXiv:2411.10430 (2024).
- [48] L. Zhang and D. Luo, Neural Transformer Backflow for Solving Momentum-Resolved Ground States of Strongly Correlated Materials (2025), arXiv:2509.09275 [cond-mat].
- [49] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, *Physical review letters* **122**, 226401 (2019).
- [50] J. Robledo 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).
- [51] 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].
- [52] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, *et al.*, Solving the hubbard model with neural quantum states, arXiv preprint arXiv:2507.02644 (2025).
- [53] Z. Liu and B. K. Clark, Unifying view of fermionic neural network quantum states: From neural network backflow to hidden fermion determinant states, *Physical Review B* **110**, 115124 (2024).
- [54] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, Sign problem in the numerical simulation of many-electron systems, *Physical Review B* **41**, 9301 (1990).
- [55] C. N. Varney, C.-R. Lee, Z. J. Bai, S. Chiesa, M. Jarrell, and R. T. Scalettar, Quantum Monte Carlo study of the two-dimensional fermion Hubbard model, *Physical Review B* **80**, 075116 (2009).
- [56] M. Vojta, Lattice symmetry breaking in cuprate superconductors: Stripes, nematics, and superconductivity, *Advances in Physics* **58**, 699 (2009).
- [57] G. Ehlers, S. R. White, and R. M. Noack, Hybrid-space density matrix renormalization group study of the doped two-dimensional Hubbard model, *Physical Review B* **95**, 125125 (2017).
- [58] K. Ido, T. Ohgoe, and M. Imada, Competition among various charge-inhomogeneous states and d-wave superconducting state in Hubbard models on square lattices, *Physical Review B* **97**, 045138 (2018).
- [59] A. S. Darmawan, Y. Nomura, Y. Yamaji, and M. Imada, Stripe and superconducting order competing in the Hubbard model on a square lattice studied by a combined variational Monte Carlo and tensor network method, *Physical Review B* **98**, 205132 (2018).

- [60] E. W. Huang, C. B. Mendl, H.-C. Jiang, B. Moritz, and T. P. Devereaux, Stripe order from the perspective of the Hubbard model, *npj Quantum Materials* **3**, 22 (2018).
- [61] H.-C. Jiang and T. P. Devereaux, Superconductivity in the doped Hubbard model and its interplay with next-nearest hopping  $t'$ , *Science* **365**, 1424 (2019).
- [62] L. F. Tocchio, A. Montorsi, and F. Becca, Metallic and insulating stripes and their relation with superconductivity in the doped Hubbard model, *SciPost Physics* **7**, 021 (2019).
- [63] Z. Denis, A. Sinibaldi, and G. Carleo, Comment on “can neural quantum states learn volume-law ground states?”, *Physical Review Letters* **134**, 10.1103/physrevlett.134.079701 (2025).
- [64] 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).
- [65] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* **30** (2017).
- [66] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint arXiv:2010.11929* (2020).
- [67] L. L. Viteritti, R. Rende, and F. Becca, Transformer variational wave functions for frustrated quantum spin systems, *Physical Review Letters* **130**, 236401 (2023).
- [68] K. Sprague and S. Czischek, Variational monte carlo with large patched transformers, *Communications Physics* **7**, 90 (2024).
- [69] M. Geier, K. Nazaryan, T. Zaklama, and L. Fu, Is attention all you need to solve the correlated electron problem?, *arXiv preprint arXiv:2502.05383* (2025).
- [70] S. A. Jafari, Introduction to hubbard model and exact diagonalization, *arXiv preprint arXiv:0807.4878* (2008).
- [71] R. P. Nutakki, A. Shokry, and F. Vicentini, Design principles of deep translationally-symmetric neural quantum states for frustrated magnets, *arXiv preprint arXiv:2505.03466* 10.48550/ARXIV.2505.03466 (2025).
- [72] R. Rende and L. L. Viteritti, Are queries and keys always relevant? a case study on transformer wave functions, *Machine Learning: Science and Technology* **6**, 010501 (2025).
- [73] M. Tinkham, *Group theory and quantum mechanics* (Courier Corporation, 2003).
- [74] C. Roth, A. Szabó, and A. H. MacDonald, High-accuracy variational monte carlo for frustrated magnets with deep neural networks, *Physical Review B* **108**, 054410 (2023).
- [75] 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).
- [76] S. Morita, R. Kaneko, and M. Imada, Quantum Spin Liquid in Spin  $1/2$   $J_1 - J_2$  Heisenberg Model on Square Lattice: Many-Variable Variational Monte Carlo Study Combined with Quantum-Number Projections, *Journal of the Physical Society of Japan* **84**, 024720 (2015).
- [77] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, *Physical Review Research* **2**, 023358 (2020).
- [78] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, *Journal of Physics: Condensed Matter* **33**, 174003 (2021).
- [79] M. Reh, M. Schmitt, and M. Gärttner, Optimizing design choices for neural quantum states, *Physical Review B* **107**, 195115 (2023).
- [80] M. Medvidović and J. R. Moreno, Neural-network quantum states for many-body physics, *The European Physical Journal Plus* **139**, 1 (2024).
- [81] T. Mizusaki and M. Imada, Quantum-number projection in the path-integral renormalization group method, *Physical Review B* **69**, 125110 (2004).
- [82] While it would be desirable to have the neural network output be fully translational-equivariant, after the patching it is only possible to maintain equivariance to translations that send patches into other patches.
- [83] 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).
- [84] S. Sorella, Wave function optimization in the variational monte carlo method, *Phys. Rev. B* **71**, 241103 (2005).
- [85] I. Loshchilov and F. Hutter, Sgdr: Stochastic gradient descent with warm restarts, *arXiv preprint arXiv:1608.03983* (2016).
- [86] K. Binder, D. M. Ceperley, J.-P. Hansen, M. Kalos, D. Landau, D. Levesque, H. Mueller-Krumbhaar, D. Stauffer, and J.-J. Weis, *Monte Carlo methods in statistical physics*, Vol. 7 (Springer Science & Business Media, 2012).
- [87] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Pereira, and F. Vicentini, Netket: A machine learning toolkit for many-body quantum systems, *SoftwareX* **10**, 100311 (2019).
- [88] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, *et al.*, Netket 3: Machine learning toolbox for many-body quantum systems, *SciPost Physics Codebases* , 007 (2022).
- [89] D. Häfner and F. Vicentini, mpi4jax: Zero-copy mpi communication of jax arrays, *Journal of Open Source Software* **6**, 3419 (2021).
- [90] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necoala, A. Paszke, J. VanderPlas, S. Wanderman-Milne,

and Q. Zhang, JAX: Composable transformations of Python+NumPy programs (2018).

- [91] 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).

### Appendix A: Hidden fermion determinant state

The Fock space of a spin  $\frac{1}{2}$  fermionic system on a finite lattice with  $M$  sites is given by all the ways that the  $2M$  single-particle states can be occupied:

$$|\mathbf{n}\rangle = \prod_{i,\sigma} \hat{c}_{i\sigma}^{\dagger n_{i\sigma}} |\emptyset\rangle, \quad (\text{A1})$$

where  $|\emptyset\rangle$  denotes the Fock vacuum and  $n_{i\sigma} \in \{0, 1\}$  due to the Pauli exclusion principle. In the Hidden Fermion formalism, we add  $\tilde{N}$  hidden fermions that can occupy  $\tilde{M}$  hidden modes. The states of this space are now also specified by the occupations of the hidden orbitals:

$$|\mathbf{n}, \tilde{\mathbf{n}}\rangle = \prod_{i,\sigma} \hat{c}_{i\sigma}^{\dagger n_{i\sigma}} \prod_{\nu} \hat{d}_{\nu}^{\dagger \tilde{n}_{\nu}} |\emptyset\rangle. \quad (\text{A2})$$

The occupations of the hidden modes depend on those of the visible ones through a constraint function:  $\tilde{\mathbf{n}} = F(\mathbf{n})$ , such that the amplitudes of the target wavefunction in the Fock basis are obtained by taking the following overlap

$$\psi(\mathbf{n}) = \langle \mathbf{n}, F(\mathbf{n}) | \Phi \rangle, \quad (\text{A3})$$

where  $|\Phi\rangle$  is a state of the augmented space. We consider the case in which  $|\Phi\rangle$  is a Slater determinant. This means that we fix  $\tilde{N}$  and work in sectors of the physical space that have a fixed number of particles and magnetization. The Slater determinant can be written down in second-quantized form as:

$$|\Phi\rangle = \prod_{\alpha=1}^{N_{\text{tot}}} \hat{\varphi}_{\alpha}^{\dagger} |\emptyset\rangle, \quad (\text{A4})$$

where  $N_{\text{tot}} = N_{\uparrow} + \tilde{N} + N_{\downarrow} + \tilde{N}$  and the operators  $\hat{\varphi}_{\alpha}^{\dagger}$  are linear combinations of the visible and hidden operators. All of this information can be stored into an  $M_{\text{tot}} \times N_{\text{tot}}$  Slater matrix  $\Phi$ , which we choose to be block diagonal in the spin basis. It is convenient to represent each Slater matrix in the following block form

$$\Phi = \begin{pmatrix} \Phi^{\downarrow} & 0 \\ 0 & \Phi^{\uparrow} \end{pmatrix}, \quad \Phi^{\sigma} = \Phi^{\sigma,(0)} + \tilde{\Phi}^{\sigma} \quad (\text{A5})$$

where  $\Phi^{\sigma,(0)}$  is the  $(M + \tilde{M}) \times (N_{\sigma} + \tilde{N})$  *visible* matrix whose last  $\tilde{M}$  rows are zero and  $\tilde{\Phi}^{\sigma}$  is the  $(M + \tilde{M}) \times (N_{\sigma} + \tilde{N})$  *hidden* matrix whose first  $M$  rows are set to zero. Given a basis state  $|\mathbf{n}\rangle$ , the wavefunction is given as:

$$\psi(\mathbf{n}) = \det \Phi(\mathbf{n}) = \prod_{\sigma} \det \left( [\Phi^{\sigma,(0)}]_{\mathbf{n}} + \tilde{\Phi}^{\sigma}(\mathbf{n}) \right), \quad (\text{A6})$$

where  $\Phi^{\sigma,(0)}(\mathbf{n})$  (resp.  $\tilde{\Phi}^{\sigma}(\mathbf{n})$ ) is obtained by slicing the rows of the visible (resp. hidden) matrix that correspond to the occupations of  $|\mathbf{n}\rangle$  (resp.  $F(\mathbf{n})$ ). In our implementation, the visible matrices are configuration-independent. The sliced hidden matrix is directly obtained as the output of a neural network.

### Appendix B: Symmetries

Let  $\mathcal{F}$  be a fermionic Fock space with  $m$  single-particle states and  $G$  a subgroup of  $\mathcal{S}_m$ . Elements of  $\mathcal{F}$  are expressed of products of fermionic creation operators acting on the Fock vacuum

$$|\mathbf{n}\rangle = \hat{c}_{\alpha_1}^{\dagger n_{\alpha_1}} \hat{c}_{\alpha_2}^{\dagger n_{\alpha_2}} \dots \hat{c}_{\alpha_m}^{\dagger n_{\alpha_m}} |\emptyset\rangle \quad (\text{B1})$$

where a canonical ordering has been defined, i.e.,  $\alpha_1 < \alpha_2 < \dots < \alpha_m$ . In order to define a unitary representation  $\hat{U}$  of  $G$  on  $\mathcal{F}$  one must specify how the creation operators transform under conjugation and how the representation acts on the Fock vacuum. Consequently, we define

1. For all  $g \in G$ ,  $\hat{U}_g |\emptyset\rangle = |\emptyset\rangle$ .
2. For all  $g \in G$  and all single particle states  $\alpha$ ,  $\hat{U}_g \hat{c}_\alpha^\dagger \hat{U}_g^\dagger = \hat{c}_{g(\alpha)}^\dagger$

where  $g(\alpha)$  denotes the permuted state index. As a result, the action of the representation on an arbitrary Fock state is expressed as:

$$\hat{U}_g |\mathbf{n}\rangle = \xi_g(\mathbf{n}) |\mathbf{n} \circ g^{-1}\rangle. \quad (\text{B2})$$

We then define the symmetrizer projector  $\hat{\mathcal{P}}_\mu$  operator for a particular irreducible representation  $\mu$  as:

$$\hat{\mathcal{P}}_\mu = \frac{d_\mu}{|G|} \sum_{g \in G} \chi_\mu^*(g) \hat{U}_g. \quad (\text{B3})$$

where  $d_\mu$  is the dimension of the representation,  $|G|$  is the group order and  $\chi_\mu(g)$  is the character of the irreducible representation evaluated on element  $g$ . The symmetrizer can thus be used to construct amplitudes that behave like irreps of  $G$ :

$$\psi_\mu(\mathbf{n}) = \langle \mathbf{n} | \hat{\mathcal{P}}_\mu | \psi \rangle = \frac{d_\mu}{|G|} \sum_{g \in G} \chi_\mu^*(g) \langle \mathbf{n} | \hat{U}_g | \psi \rangle = \frac{d_\mu}{|G|} \sum_{g \in G} \chi_\mu^*(g) \xi_{g^{-1}}(\mathbf{n}) \psi(\mathbf{n} \circ g) \quad (\text{B4})$$

This followed from  $\langle \mathbf{n} | \hat{U}_g = \left( \hat{U}_g^\dagger | \mathbf{n} \rangle \right)^\dagger$  and the fact that  $\hat{U}_g^\dagger = \hat{U}_{g^{-1}}$ .

### Appendix C: Equivariant attention vectors

Suppose we have a permutation group  $G$  that permutes the embedded vectors  $\mathbf{x}_j$  to  $\mathbf{x}_{g(j)}$ . The condition for the attention vectors to be equivariant with respect to all  $g$  in  $G$  is:

$$\mathbf{A}_{g(i)}^\mu(\mathbf{x}) = \mathbf{A}_i^\mu(\mathbf{x} \circ g). \quad (\text{C1})$$

This means that:

$$\sum_{j=1}^{N_p} J_{g(i)j}^\mu \mathbf{V}^\mu \mathbf{x}_j = \sum_{j=1}^{N_p} J_{ij}^\mu \mathbf{V}^\mu \mathbf{x}_{g(j)} = \sum_{j=1}^{N_p} J_{ig^{-1}(j)}^\mu \mathbf{V}^\mu \mathbf{x}_j \quad (\text{C2})$$

where we used a change of index  $k = g(j)$ . This leads to  $J_{g(i)j}^\mu = J_{ig^{-1}(j)}^\mu$  or in other words,  $J_{g(i)g(j)}^\mu = J_{ij}^\mu$ .

### Appendix D: Translation invariance of symm(ViT-Ti)

In this section, we show how a translationally invariant state can be used to obtain an invariant wavefunction. Without loss of generality, we present the result for a symmetrized wavefunction in the trivial sector of the translation operator; however, the same reasoning applies to a generic symmetry transformation in any sector.

Starting from equation (9), consider a translated input.

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \frac{1}{|G|} \sum_g \chi^*(g) \xi_{g^{-1}}(\mathbf{n} \circ T_q) \psi(\mathbf{n} \circ T_g T_q).$$

Noting that the fermionic sign function transforms as  $\xi_{g^{-1}}(\mathbf{n} \circ T_q) = \xi_{g^{-1} \cdot q^{-1}}(\mathbf{n})$ , and that the translation operation composes as  $\psi(\mathbf{n} \circ T_g T_q) = \psi(\mathbf{n} \circ T_{g \cdot q})$ , the expression becomes

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \frac{1}{|G|} \sum_g \chi^*(g) \xi_{g^{-1} \cdot q^{-1}}(\mathbf{n}) \psi(\mathbf{n} \circ T_{g \cdot q}).$$

By performing a change of variable  $g' = g \cdot q$  (with the inverse transformation  $g = g' \cdot q^{-1}$ ), the sum becomes:

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \frac{1}{|G|} \sum_{g'} \chi^*(g' \cdot q^{-1}) \xi_{(g^{-1})'}(\mathbf{n}) \psi(\mathbf{n} \circ T_{g'}).$$

Since the characters satisfy the group property  $\chi^*(g' \cdot q^{-1}) = \chi^*(g') \chi^*(q^{-1})$ . Therefore, the expression factors

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \chi^*(q^{-1}) \left[ \frac{1}{|G|} \sum_{g'} \chi^*(g') \xi_{(g^{-1})'}(\mathbf{n}) \psi(\mathbf{n} \circ T_{g'}) \right].$$

Recognizing that the term in brackets is exactly  $\psi_{\text{Sym}}(\mathbf{n})$ , we obtain:

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \chi^*(q^{-1}) \psi_{\text{Sym}}(\mathbf{n}). \quad (\text{D1})$$

This shows that the symmetric state  $\psi_{\text{Sym}}(\mathbf{n})$  transforms according to the character  $\chi^*(q^{-1})$  under any translation  $T_q$ , thereby preserving the invariance and antisymmetry of the wavefunction.

Now, we can reach the same result using a translational invariant network for the backflow transformation. Here, we need to be more explicit with the notation.

The wavefunction is defined as follows:

$$\psi_{\text{Sym}}(\mathbf{n}) = \frac{1}{|G|} \sum_g \chi^*(g) \xi_{g^{-1}}(\mathbf{n}) \cdot J(\mathbf{n}) \prod_{\sigma} \det \left( [\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n})]_{\mathbf{n} \circ T_g} \right),$$

where  $J(\mathbf{n})$  is the Jastrow factor,  $\Phi^{\sigma, (0)}$  denotes the mean-field orbitals matrix, and  $\tilde{\Phi}^{\sigma}(\mathbf{n})$  represents the backflow transformation matrix.

The subscript in  $[\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n})]_{\mathbf{n} \circ T_g}$  indicates that, given the full matrix  $\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n})$ , we select only the occupied orbitals corresponding to the configuration string  $\mathbf{n}$ .

Hence, given a translation of the input by  $T_q$ , the wavefunction becomes:

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \frac{1}{|G|} \sum_g \chi^*(g) \xi_{g^{-1} \cdot q^{-1}}(\mathbf{n}) \cdot J(\mathbf{n} \circ T_q) \prod_{\sigma} \det \left( [\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n} \circ T_q)]_{\mathbf{n} \circ T_{g \cdot q}} \right).$$

But now, assuming that both the backflow transformation and the Jastrow factor yield invariant representations, we have:

$$J(\mathbf{n} \circ T_q) = J(\mathbf{n}) \quad \text{and} \quad \tilde{\Phi}^{\sigma}(\mathbf{n} \circ T_q) = \tilde{\Phi}^{\sigma}(\mathbf{n}).$$

Thus, we can rewrite the expression as:

$$\psi_{\text{Sym}}(\mathbf{n}) = \frac{1}{|G|} \sum_{g'} \chi^*(g' \cdot q^{-1}) \xi_{g'}(\mathbf{n}) \cdot J(\mathbf{n}) \prod_{\sigma} \det \left( [\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n})]_{\mathbf{n} \circ T_{g'}} \right),$$

where we have re-indexed the sum using  $g' = g \cdot q$ .

Since, again, the characters satisfy the group property  $\chi^*(g' \cdot q^{-1}) = \chi^*(g') \chi^*(q^{-1})$ , the expression factors

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \chi^*(q^{-1}) \left\{ \frac{1}{|G|} \sum_{g'} \chi^*(g') \xi_{(g')^{-1}}(\mathbf{n}) \cdot J(\mathbf{n}) \prod_{\sigma} \det \left( [\Phi^{\sigma, (0)} + \tilde{\Phi}^{\sigma}(\mathbf{n})]_{\mathbf{n} \circ T_{g'}} \right) \right\}.$$

Recognizing that the term in brackets is exactly  $\psi_{\text{Sym}}(\mathbf{n})$ , we obtain:

$$\psi_{\text{Sym}}(\mathbf{n} \circ T_q) = \chi^*(q^{-1}) \psi_{\text{Sym}}(\mathbf{n}).$$

Thus, we retrieve exactly the same equation (D1) without needing to evaluate the network  $|G|$  times.



### Appendix E: Variational energies

| Method                        | $E_0/M$     | $\sigma^2/M$ |
|-------------------------------|-------------|--------------|
| HFDS ( $8 \times 4$ )         | -0.73342(8) | 0.0225       |
| HFDS ( $8 \times 8$ Torus)    | -0.7454(9)  | 0.0563       |
| HFDS ( $8 \times 8$ Cylinder) | -0.7259(0)  | 0.0422       |
| JBf ( $8 \times 4$ )          | -0.7332(6)  | 0.0235       |
| JBf ( $8 \times 8$ Torus)     | -0.7458(6)  | 0.0525       |
| JBf ( $8 \times 8$ Cylinder)  | -0.7264(5)  | 0.0334       |

| Method           | $E_0/M$    | $\sigma^2/M$ |
|------------------|------------|--------------|
| HFDS-ViT         | -0.736(1)  | 0.105        |
| HFDS Symm-ViT-Ti | -0.729(4)  | 0.139        |
| JBf-ViT          | -0.7422(3) | 0.0698       |
| JBf Symm-ViT-Ti  | -0.726(9)  | 0.168        |

TABLE I: Converged variational energies per site for different lattices and network architectures. The left table shows results for the  $8 \times 4$  and  $8 \times 8$  cylinders and  $8 \times 8$  torus lattices with  $U/t = 8$  and  $n_h = 1/8$  hole doping. The right table shows results for the  $8 \times 8$  torus with ViT and Symm-ViT-Ti architectures under the same physical parameters. Energies are estimated using 131072 samples.

### Appendix F: Additional results for the $8 \times 8$

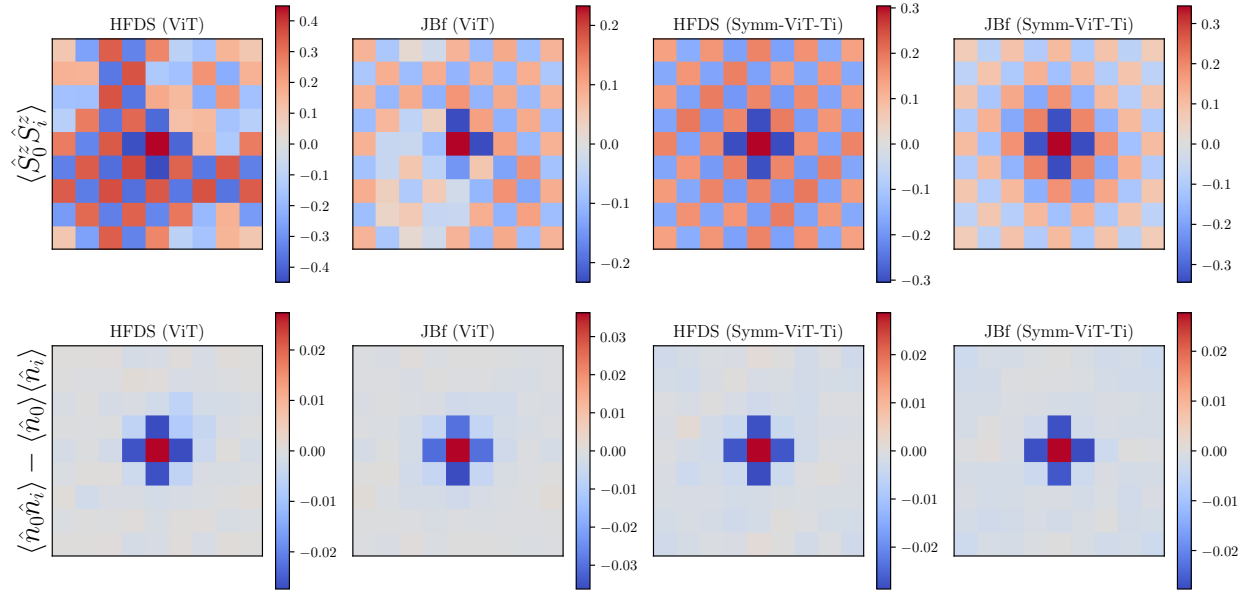


FIG. 6: Results for the physical observables on the  $8 \times 8$  lattice for  $U/t = 8$  and  $n_h = 1/8$ . The top row shows the spin-spin correlation in real space for the HFDS-ViT and JBf wavefunctions parameterized by an unsymmetrized ViT and the Symm-ViT-Ti. The bottom row shows the density-density correlation function.

### Appendix G: Test patching strategies

We conducted a series of experiments to investigate how different patching strategies influence model performance. In particular, we compared a  $2 \times 2$  patching scheme where local groups of orbitals are treated as input units to a baseline where each orbital is treated as an individual patch ( $1 \times 1$ ). This allows us to assess whether incorporating short-range spatial structure improves the model's ability to capture relevant

correlations. We also examined the effect of how spin information is represented, by testing whether concatenating the spin up and spin down subsectors as separate channels of a single input offers advantages over treating them as independent inputs. These experiments provide insight into how local context and spin representation affect the expressiveness and efficiency of the network.

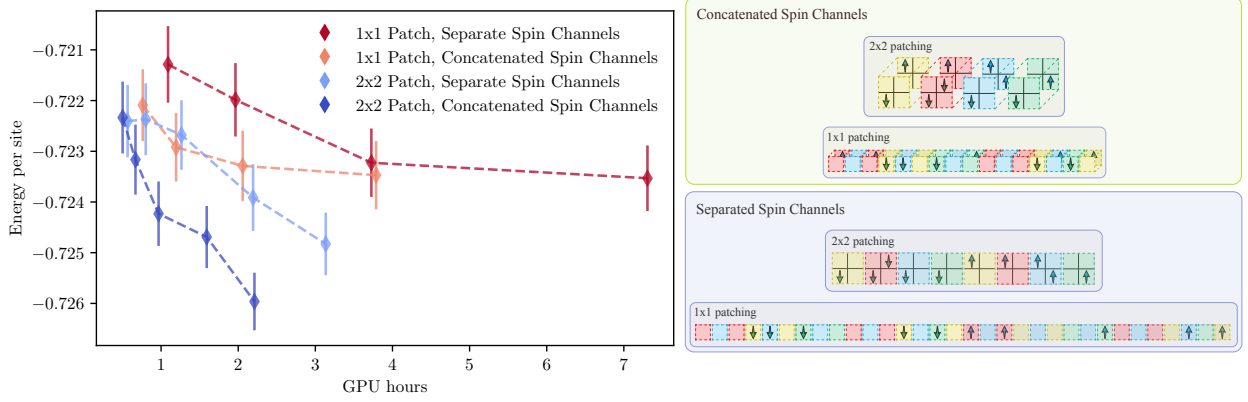


FIG. 7: Left: Effect of different patching strategies on a  $4 \times 8$  PBC-OBC system at  $U = 8$ , with each point corresponding to a different number of encoder blocks. Right: Schematic of the patching schemes—top row:  $2 \times 2$  and  $1 \times 1$  patches with spin concatenated; bottom row: same patches with spin treated separately. The best trade-off between energy and GPU time uses  $2 \times 2$  patches with concatenated spin.