Accurate neural quantum states for interacting lattice bosons

Zakari Denis[1,2] and Giuseppe Carleo[1,2]

[1]Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[2]Center for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

arXiv:2404.07869v2 [quant-ph] 11 Jun 2025

In recent years, neural quantum states have emerged as a powerful variational approach, achieving state-of-the-art accuracy when representing the ground-state wave function of a great variety of quantum many-body systems, including spin lattices, interacting fermions or continuous-variable systems. However, accurate neural representations of the ground state of interacting bosons on a lattice have remained elusive. We introduce a neural backflow Jastrow Ansatz, in which occupation factors are dressed with translationally equivariant many-body features generated by a deep neural network. We show that this neural quantum state is able to faithfully represent the ground state of the 2D Bose-Hubbard Hamiltonian across all values of the interaction strength. We scale our simulations to lattices of dimension up to 20×20 while achieving the best variational energies reported for this model. This enables us to investigate the scaling of the entanglement entropy across the superfluid-to-Mott quantum phase transition, a quantity hard to extract with non-variational approaches.

1 Introduction

Lattice bosonic systems are important in condensed matter because of their unique properties. They are not subject to the exclusion principle, which allows for the appearance of smooth phases of matter in diluted settings. These phases can often be described by classical fields, which simplifies the understanding of many-body phenomena, including the study of $U(1)$ symmetry breaking, superfluidity, and boson condensation. However, simulating these systems beyond mean-field approximations is a challenge. The quantum statistics of bosons result in an exponential growth of the Hilbert space dimension with system size and particle number, complicating simulations. In driven systems, an infinite-dimensional Hilbert space might be needed for a single bosonic mode, limiting the feasibility of exact diagonalization.

Several methods have been proposed to efficiently simulate interacting lattice bosons at zero temperature. Among earliest approaches, path-integral techniques were proposed at finite temperature [1–4]. In particular, quantum Monte Carlo (QMC) simulations using the worm algorithm provided remarkably accurate predictions extrapolated to zero temperature [5, 6]. At zero temperature, Green-function Monte Carlo particularly stands out [7–9], while the reptation quantum Monte Carlo algorithm [10] was extended to lattice systems [11]. However, these techniques do not give direct access to the wave function and restrict the structure of operators which may be efficiently evaluated in the chosen computational basis. Furthermore, they suffer from the sign problem [12], even for bosonic systems in cases of geometric frustration [13–15]. This is more generally the case with non-stoquastic Hamiltonians, such as those of bosonic lattices subjected to artificial magnetic fluxes, which have attracted consistent theoretical interest [16–28] and are of current experimental relevance [29–37].

Variational methods are more flexible in this regard. Variational Monte Carlo (VMC) using a modified Jastrow Ansatz [38–40] reproduced all of the features of the quantum phase transition in all relevant spatial dimensions, though only qualitatively. Tensor networks have been used, in particular, to probe ground-state properties of lattice bosons [41–43]. However, this is re-

Zakari Denis: Email: zakari.denis@epfl.ch
Giuseppe Carleo: Email: giuseppe.carleo@epfl.ch

stricted by entanglement and typically requires to introduce truncations in the local occupation factors. The generic area-law entanglement scaling of lattice bosons [44] and the growth of entanglement upon real-time propagation restrict the efficiency of matrix-product variational representations mainly to the ground state of one-dimensional geometries, such as open chains [45–49] or small-perimeter cylinders [41], and short time evolutions.

In recent years, neural quantum states [50] have emerged as a powerful variational method, consistently demonstrating remarkable accuracy in representing the ground-state wave function of a wide range of nontrivial Hamiltonians. In addition to spin problems, where it stands as the state-of-the-art method for frustrated lattices [51–53], properly tailored networks have demonstrated their effectiveness in addressing problems involving other kinds of degrees of freedom, including bosonic [54], fermionic [55–57] and beyond [58, 59]. Despite these successes, highly accurate neural-network representations of the ground state of lattice bosonic systems have remained elusive. Several works [60–63] have recently contributed to these ongoing efforts, where the most notable advances [64, 65] have been achieved by incorporating some of the physical structure of the system in the variational Ansatz.

In this work, we introduce a new bosonic neural Ansatz based on the concept of backflow [66], originally formulated in continuous space [67–70] and later extended to lattices [71, 72]. Neural-network parametrizations of the backflow transformation [73, 74] have driven the most recent advances in fermionic neural quantum states [55–57]. We provide analytical support to motivate its structure and benchmark it on the celebrated Bose-Hubbard model in two dimensions. We assess its remarkable accuracy and scale our simulations to lattices with up to 20×20 sites without local truncation, allowing us to perform a finite-size scaling analysis around the critical point. We further investigate the entanglement entropy of the system across the phase transition, a matter of theoretical [41, 75, 76] and experimental [77] interest.

The paper is structured as follows: the Bose-Hubbard physics is briefly described in Sec. 2; in Sec. 3, we introduce the neural backflow-Jastrow architecture and formally derive the structure of some of the correlations it can capture; we then apply it to simulate the ground state of the Bose-Hubbard Hamiltonian in Sec. 4, before finally concluding in Sec. 5.

2 Bose Hubbard

In this work, we focus on the paradigmatic Bose-Hubbard model [78], as given by the Hamiltonian

$$\hat{H} = -J \sum_{\langle i,j \rangle} (\hat{a}_i^\dagger \hat{a}_j + \hat{a}_j^\dagger \hat{a}_i) + \frac{U}{2} \sum_i \hat{a}_i^\dagger \hat{a}_i^\dagger \hat{a}_i \hat{a}_i. \quad (1)$$

Here, $\hat{a}_i$ denotes the annihilation operator at site i of the lattice under consideration, satisfying the usual bosonic canonical commutation relations, J is the rate of the hopping of bosons between nearest-neighbor sites, and U quantifies the strength of on-site repulsive interactions. This system can be realized in a number of platforms, notably with ultra-cold atoms loaded in an optical lattice [77, 79–84], or Josephson-junction arrays [85, 86]. In what follows, we shall restrict ourselves to $L \times L$ two-dimensional square lattices at fixed number of particles N and unit density $\bar{n} := N/L^2 = 1$.

This system displays a second-order phase transition resulting from the competition between the kinetic energy and the potential energy. At infinite interaction strength and integer density, the wavefunction is a product of Fock states: the so-called Mott phase. As the ratio J/U increases, the $U(1)$ symmetry under the global transformation $\hat{a}_i \mapsto \hat{a}_i e^{i\alpha}$ spontaneously breaks, and the ground state becomes extensively degenerate. The bosonic population spontaneously condensates into the zero-momentum mode, as revealed by a finite value of the condensate fraction:

$$\rho_0 := \langle \hat{n}(\boldsymbol{k} = 0) \rangle / L^2 \equiv \sum_{ij} \langle \hat{a}_i^\dagger \hat{a}_j \rangle / L^4, \quad (2)$$

which here plays the role of order parameter [87]. In 2D, this phase transition is known to belong to the 3D XY-model universality class [1, 4, 6, 88, 89]. The critical value of the control parameter, $J_c/U = 0.05974(3)$ [$U_c/J = 16.74(1)$], was first precisely determined by a high-order strong coupling expansion [90, 91] and later refined by means of QMC calculations extrapolated to vanishing temperature [6].

3 Variational model

The simplest many-body variational Ansatz capable of qualitatively describing the singular depletion of the condensate while also reproducing the proper behavior of the spatial correlations is the two-body Jastrow wavefunction [92, 93]

$$|\psi_J\rangle = e^{\sum_{ij} \hat{n}_i W_{ij} \hat{n}_j} |\psi_0\rangle \,, \tag{3}$$

where $|\psi_0\rangle = N^{-1/2}(\sum_i \hat{a}_i^\dagger)^N |\mathbf{0}\rangle$ is the wavefunction of the ideal condensate, and $\mathbf{W}$ is a $L^2 \times L^2$ matrix of variational parameters. This can be equivalently written as

$$\ln \psi_J(\boldsymbol{n}) = \sum_{ij} n_i W_{ij} n_j + \ln \psi_0(\boldsymbol{n}), \tag{4}$$

where $\boldsymbol{n} := (n_1, \ldots, n_{L^2})$ and $\psi(\boldsymbol{n}) \equiv \langle \boldsymbol{n}|\psi\rangle$.

In Refs. [38–40], authors showed that the variational accuracy could be greatly enhanced without losing the physical structure and intuition of the two-body Jastrow Ansatz. This is done by adding the following many-body Gutzwiller projector:

$$|\psi_{\mathrm{MBJ}}\rangle = e^{g_{\mathrm{MB}}(\hat{\Pi}_h + \hat{\Pi}_d)} |\psi_{\mathrm{J}}\rangle \,, \tag{5}$$

where g_{MB} is a real variational parameter and the projectors $\hat{\Pi}_h = \sum_i \hat{h}_i \bigotimes_{j\in N(i)} (\hat{\mathbb{1}} - \hat{d})_j$ and $\hat{\Pi}_d = \sum_i \hat{d}_i \bigotimes_{j\in N(i)} (\hat{\mathbb{1}} - \hat{h})_j$, with $\hat{h} = |0\rangle\langle 0|$ and $\hat{d} = |2\rangle\langle 2|$, and $N(i)$ denoting the neighborhood of i on the lattice, measure the amount of isolated holons and doublons, respectively. This many-body term serves as a penalty for localized density inhomogeneities, spatially binding holon-doublon pairs, and thereby favoring the onset of a homogeneous Mott insulating phase.

This variational Ansatz, although simple, is able to qualitatively capture the physics of the transition, including the behavior of the correlations and the structure factor [38–40]. However, the quantitative agreement is not excellent, with an estimated ground-state energy for this Ansatz affected by a relative error of the order of 10% with respect to the exact reference and the transition point shifted by 25% to around $U_c/J \sim 21$.

In what follows, we introduce a flexible generalization of the above construction based on the concept of neural backflow transformation [74].

3.1 Many-body Jastrow

The most straightforward improvement over the Ansatz of Eq. (5) is to lift the Jastrow term to a M-body term. Let us show that the hierarchy of correlations that can be accounted for by a deep convolutional neural network (CNN) encompasses that of such a many-body Jastrow term.

Let us consider a network of depth D built upon the following single-channel convolutional layer with skip connections:

$$h_i^{(n)}(\boldsymbol{n}) = n_i + \tilde{h}_i^{(n)}(\boldsymbol{n}), \tag{6}$$

$$\tilde{h}_i^{(n)}(\boldsymbol{n}) = f\big(\textstyle\sum_j w_{\boldsymbol{r}_{ij}}^{(n)} h_j^{(n-1)}(\boldsymbol{n}) + b^{(n)}\big), \tag{7}$$

with $W_{\boldsymbol{r}} = w_{\boldsymbol{r}} = 0$ for any $\boldsymbol{r}$ such that $|\boldsymbol{r}| > d_F$, where d_F corresponds to the dimension of the convolutional filter. Upon assuming for simplicity that the activation function f belongs to C^2 and possesses a non-vanishing second-order derivative, one has the following recursion relation[1]:

$$\sum_i h_i^{(n)}(\boldsymbol{n}) = \ldots + \frac{1}{2}\sum_{i_0 i_1} n_{i_0} W_{i_0 i_1}^{(n)} h_{i_1}^{(n-1)}(\boldsymbol{n}) = \ldots + \sum_{n=1}^{D} \frac{1}{2^n} \sum_{i_0 \cdots i_n} n_{i_0} W_{i_0 i_1}^{(n)} n_{i_1} W_{i_1 i_2}^{(n-1)} \ldots W_{i_{n-1} i_n}^{(1)} n_{i_n}, \tag{8}$$

that is a linear combination of n-body Jastrow terms with $n \leq D+1$, with symmetric two-body translationally equivariant Jastrow weights given

[1] Here, the Lagrange form of the first-order Taylor expansion of f around the bias b was used: $\tilde{h}_i^{(n)} = f(b) + \sum_j \big[f'(b) w_{\boldsymbol{r}_{ij}}\big] h_j^{(n-1)} + \frac{1}{2}\sum_{jk} \big[f''(b_\star) w_{\boldsymbol{r}_{ij}} w_{\boldsymbol{r}_{ik}}\big] h_j^{(n-1)} h_k^{(n-1)}$. Interestingly, the mean-value theorem, at the core of the Lagrange remainder, was also used to show that the ground state of a system of interacting particles in continuous space is exactly captured by a general backflow transformation [57].

by

$$W_{ij}^{(n)} \equiv W_{ji}^{(n)} \propto \sum_k w_{\boldsymbol{r}_{ki}}^{(n)} w_{\boldsymbol{r}_{kj}}^{(n)}. \tag{9}$$

Trivially, the two-body Jastrow weights can be made independent for each n-body term at the cost of at most M channels in the CNN.

An analogous result is obtained, replacing skip connections by residual connections, as will be later introduced. Therefore, range-d_F M-body Jastrow factors can be generated with a residual neural-network (ResNet) Ansatz with depth $M-1$ and filters of width d_F.

3.2 Generalized many-body Gutzwiller projector

In Eq. (5), the many-body term acts as a partial projection of the wave function onto a manifold with specific values of the local quantum numbers within some finite-size patch. The choice of the specific values of these quantum numbers is set from *a priori* knowledge about the underlying physics. We may loosen this inductive bias at the cost of a larger number of variational parameters.

Let us consider the following many-body Gutzwiller Ansatz:

$$|\psi_{\mathrm{MB}}\rangle = e^{g_{\mathrm{MB}} \sum_i \bigotimes_{\boldsymbol{v}\in\mathcal{V}} |x_{\boldsymbol{v}}\rangle\langle x_{\boldsymbol{v}}|_{T_{\boldsymbol{v}}(i)}} |\psi_0\rangle\,, \tag{10}$$

where g_{MB} is a free variational parameter, $\mathcal{V}$ denotes a set of relative positions with respect to the ith site, thereby defining a *patch*, and $\mathcal{X} = \{x_{\boldsymbol{v}}, \boldsymbol{v} \in \mathcal{V}\}$ some corresponding fixed local configurations given *a priori*. Here, $T_{\boldsymbol{v}}(i)$ denotes the site corresponding to i translated on the lattice by $\boldsymbol{v}$. This Ansatz may be evaluated as

$$\ln \psi_{\mathrm{MB}}(\boldsymbol{n}) = \ln \Pi_{\mathrm{MB}}(\boldsymbol{n}) + \ln \psi_0(\boldsymbol{n}), \tag{11}$$

$$\ln \Pi_{\mathrm{MB}}(\boldsymbol{n}) := g_{\mathrm{MB}} \sum_i \prod_{\boldsymbol{v}\in\mathcal{V}} \mathbb{1}[n_{T_{\boldsymbol{v}}(i)} = x_{\boldsymbol{v}}]; \tag{12}$$

where $\mathbb{1}$ denotes the indicator function. The diagonal many-body projector of Eq. (12) may be represented by a CNN. Indeed, one has that

$$\ln \Pi_{\mathrm{MB}}(\boldsymbol{n}) = \sum_i h_{i,1}^{(2)}(\boldsymbol{n}), \tag{13}$$

$$h_{i,1}^{(2)}(\boldsymbol{n}) = \mathrm{ReLU}\Big(-2g_{\mathrm{MB}}\Delta x \sum_{\mu=1}^{2} h_{i,\mu}^{(1)}(\boldsymbol{n}) + g_{\mathrm{MB}}\Big), \tag{14}$$

$$h_{i,\mu}^{(1)}(\boldsymbol{n}) = \mathrm{ReLU}\Big(\sum_{\boldsymbol{v}\in\mathcal{V}} \frac{(-1)^\mu}{x_{T_{\boldsymbol{v}}(i)}} n_{T_{\boldsymbol{v}}(i)} - (-1)^\mu\Big), \tag{15}$$

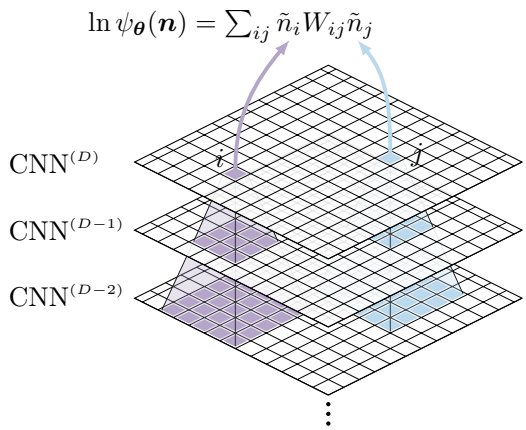

Figure 1: Schematic representation of the neural backflow-Jastrow architecture of Eq. (18). The backflow transformation dresses the bare local quantum numbers entering the Jastrow term with equivariant many-body features. These features bear a dependence on the configuration of all sites within the network's receptive field. The latter is represented by the shaded region for an exemplary deep convolutional neural network with square filters of width $d_F = 3$.

where $h_{i,\mu}^{(n)}$ denotes the μth channel of the ith output of the nth layer of the CNN, and $\Delta x > \max_{x\in\mathcal{X}}(x)$[2]. This corresponds to a two-layer CNN with parameters:

$$\begin{aligned} K_{\boldsymbol{v},\mu,1}^{(1)} &= (-1)^\mu / x_{T_{\boldsymbol{v}}(i)} \in \mathbb{R}^{|\mathcal{V}|\times 2\times 1}, \\ b_\mu^{(1)} &= (-1)^{\mu+1} \in \mathbb{R}^2, \\ K_{\boldsymbol{0},1,\mu'}^{(2)} &= -2g_{\mathrm{MB}}\Delta x \in \mathbb{R}^{1\times 1\times 2}, \\ b_1^{(2)} &= g_{\mathrm{MB}} \in \mathbb{R}. \end{aligned} \tag{16}$$

Therefore, a many-body Gutzwiller Ansatz with K many-body projectors may be represented by a two-layer CNN with a filter geometry matching $\mathcal{V}$ and K channels. Relaxing the conditions in Eq. (16) generalizes the projection to the optimum quantum numbers. In App. A, we show how this architecture can also account for an attractive confining potential for holons and doublons close apart, effectively stabilizing the Mott phase.

[2]The case where the value 0 belongs to $\mathcal{X}$ can be simply worked out by a fixed initial linear transformation of all inputs as $n_i \mapsto n_i + \xi$, $\forall \xi \in \mathbb{R}_+^\star$. In particular, Eq. (5) is recovered using local fluctuations $n_i - \bar{n}$ as an input and identifying holes (doublons) with the value $x = -1$ ($x = +1$)

3.3 Neural backflow Jastrow Ansatz

This motivates the use of a NQS Ansatz of the form [64]

$$\ln \psi_{\boldsymbol{\theta}}(\boldsymbol{n}) = \sum_{i,\mu} h_{i,\mu}^{(D)}(\boldsymbol{n}) + \ln \psi_0(\boldsymbol{n}), \tag{17}$$

where $\boldsymbol{\theta}$ denotes the set of all variational parameters, and $h_{i,\mu}$ is the channel μ of site i of a depth-D residual network with stride equal to one.

Provided the number of channels is large enough, such a network should be able to encompass both the M-body Jastrow term and the many-body Gutzwiller projector. However, the range of the Jastrow terms in such a parametrization is limited by the width of the filters d_F. This is in contrast with the initial Jastrow Ansatz of Eq. (3) which could capture the crucial long-range correlations responsible of the singular behavior of the density structure factor at zero momentum [39].

To circumvent this, we instead introduce the many-body features of the ResNet as an equivariant backflow transformation of a translationally invariant two-body Jastrow with maximum range. This leads to the variational Ansatz that we shall consider throughout this article:

$$\ln \psi_{\boldsymbol{\theta}}(\boldsymbol{n}) = \sum_{i,j=1}^{L^d} \tilde{n}_i W_{d_{ij}} \tilde{n}_j, \tag{18}$$

where the weights only depend on the L^1 minimum-image distance d_{ij} between all pairs (i,j). We also dropped the original mean-field prior ψ_0 which proves useful only for small system sizes. Fed to this Jastrow factor are translation-equivariant many-body features obtained according to the following backflow transformation of the original local occupation factors:

$$\tilde{n}_i = n_i + \sum_{\mu=1}^{\alpha_D} a_\mu h_{i,\mu}^{(D)}. \tag{19}$$

Here, $h_{i,\mu}^{(D)}$ are translation-equivariant features extracted from the original occupations through a residual neural network of depth D with α_D channels, and the mixing weights a_μ both combine all channels and control the magnitude of the backflow transformation. The former are obtained as $h_{i,\mu}^{(2\ell)} = \tilde{h}_{i,\mu}^{(2\ell-1)}$ and $h_{i,\mu}^{(2\ell+1)} = \mathrm{LayerNorm}(h_{i,\mu}^{(2\ell-1)} + \tilde{h}_{i,\mu}^{(2\ell)})$, with

$$\tilde{h}_{i,\mu}^{(\ell)} = \sigma\left(\sum_{\mu'=1}^{\alpha_\ell} \sum_{\boldsymbol{v}\,:\, v_\infty \le d_K} K_{\boldsymbol{v},\mu\mu'}^{(\ell)} h_{T_{\boldsymbol{v}}(i),\mu'}^{(\ell-1)} + b_\mu^{(\ell)} \right), \tag{20}$$

where, σ is an activation function (GELU), α_ℓ the number of channels of the ℓth layer, d_K the kernel size, $T_{\boldsymbol{v}}(i)$ corresponds to the site i translated by $\boldsymbol{v}$, and $(K_{\boldsymbol{v},\mu\mu'}^{(\ell)}, b_\mu^{(\ell)})$ the filters and bias of the ℓth layer. To further improve stability, the input occupations are rescaled as $n_i \mapsto n_i/\bar{n} - 1$, $\forall i$, with $\bar{n} = N_p/L^d$. The architecture of the model is illustrated in Fig. 1.

Let us note that this construction, in presence of a bias term in the last layer of the backflow transformation, is at least as expressive as the bare ResNet. Furthermore it allows one to perform VMC more efficiently by initializing the weights of the Jastrow with those obtained by first optimizing a bare two-body Jastrow Ansatz of the form of Eq. (3), and then optimizing the entire network altogether. Indeed, provided the backflow weights are initialized such that they initially act perturbatively, this allows to start the optimization closer from the ground state and thus with a much lower variance on all Monte Carlo estimates involved in the process.

As detailed in App. B, splitting the layout of the wavefunction into a simple mean-field-like long-range behavior and a complex short-range structure is a convenient design choice performance-wise, especially for gapped Hamiltonians whose ground states exhibit some finite-range quantum correlations.

Very recent variational Ansätze based upon the visual-transformer architecture [53, 94, 95] adopt a reminiscent structure, rooted in representation learning [96], where a strong scission of the model into a deep embedding and a shallow final correlator proves optimal.

4 Numerical results

In what follows, we optimize the parameters $\boldsymbol{\theta}$ of the model defined by Eq. (18) by minimizing the variational expectation value of the energy

$$E_{\boldsymbol{\theta}} = \frac{\langle \psi_{\boldsymbol{\theta}} | \hat{H} | \psi_{\boldsymbol{\theta}} \rangle}{\langle \psi_{\boldsymbol{\theta}} | \psi_{\boldsymbol{\theta}} \rangle}, \tag{21}$$

with $\hat{H}$ as given by Eq. (1). This is achieved with guarantees of exponential convergence via the stochastic-reconfiguration prescription [97], where the parameters are evolved in time according to

$$\dot{\boldsymbol{\theta}} = -\mathbf{S}^{-1} \boldsymbol{F}, \tag{22}$$

Size	U/J	Method	Parameters	$E_{\boldsymbol{\theta}}/JL^2$
8×8	8.5	GFMC [100]	—	−1.544669(8)
		NBFJ	6681	−1.544668(3)
	17.0	GFMC [100]	—	−0.5375(4)
		NBFJ	6681	−0.53804(2)
10×10	16.0	NQS-OH [65]	426	−0.5721(3)
		NBFJ	6683	−0.58972(3)
	20.0	NQS-B [65]	206	−0.4068(9)
		NBFJ	6683	−0.43329(3)

Table 1: Comparison of the variational energy obtained by the neural backflow Jastrow (NBFJ) Ansatz with depth $D = 6$ and those of Green-function Monte Carlo (GFMC) [100] and recent NQS simulations [65].

where $\mathbf{S}$ and $\boldsymbol{F}$ respectively denote the quantum geometric tensor [98] and the vector of forces, as given by

$$S_{kk'} = \mathbb{E}[O_k^\star(\boldsymbol{n})O_{k'}(\boldsymbol{n})] - \mathbb{E}[O_k^\star(\boldsymbol{n})]\mathbb{E}[O_{k'}(\boldsymbol{n})], \tag{23}$$

$$F_k = \mathbb{E}[O_k^\star(\boldsymbol{n})E_{\mathrm{loc}}(\boldsymbol{n})] - \mathbb{E}[O_k^\star(\boldsymbol{n})]\mathbb{E}[E_{\mathrm{loc}}(\boldsymbol{n})], \tag{24}$$

with $O_k(\boldsymbol{n}) := \partial_{\theta_k} \ln \psi_{\boldsymbol{\theta}}(\boldsymbol{n})$ the log-derivative of the variational Ansatz and $E_{\mathrm{loc}}(\boldsymbol{n}) := \langle \boldsymbol{n}| \hat{H} |\psi_{\boldsymbol{\theta}}\rangle / \langle \boldsymbol{n}|\psi_{\boldsymbol{\theta}}\rangle$ the so-called *local energy*. In the above, expectation values are implicitly taken with respect to the Born distribution: $\mathbb{E}[f(\boldsymbol{n})] := \sum_{\boldsymbol{n}} f(\boldsymbol{n})|\psi_{\boldsymbol{\theta}}(\boldsymbol{n})|^2 / \langle \psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle$.

Importantly, all of these quantities can be efficiently estimated by drawing samples from this probability density function thanks to, e.g., Markov chain Monte Carlo and the Metropolis-Hastings algorithm. More details about the variational optimization are given in App. C.

In all following simulations, ResNet-based backflow transformations with square filters of width $d_F = 3$ and $\alpha = 12$ channels per layer were used. A depth of $D = 8$ was used for 20×20 lattices and $D = 6$ for the rest. All models were optimized through NetKet 3.0 [99] on a single GPU.

4.1 Benchmarking

The scaling of the Hilbert-space dimension with system size is rather adverse for bosonic lattices. Therefore, exact-diagonalization reference quantities are restricted to very small systems which are not very representative of the physics at work in the thermodynamic limit.

At moderate system size, ground-state predictions may be efficiently obtained via Green-function Monte Carlo (GFMC) [7–9]. These provide a valuable reference value since they are exact up to a controllable systematic bias. In Table 1, we compare the variational energies achieved by our Ansatz for a 8×8 lattice against a GFMC reference [100], showing perfect agreement. Therein, we further compare our variational energies to the best previously achieved NQS results [65] for a 10×10 system, obtaining values 3 to 6% lower, although with a higher number of parameters.

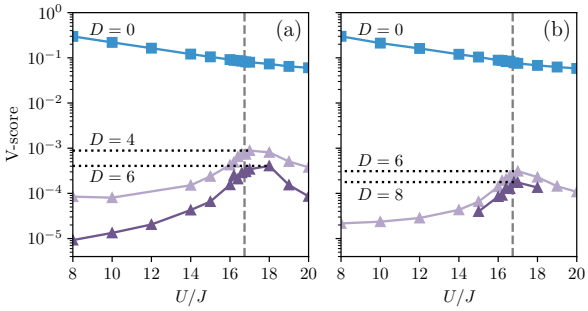

Figure 2: V-score as a function of the depth D of the neural backflow transformation for lattices of size (a) 16×16 and (b) 20×20 at unit filling $\bar{n} = 1$. The gray shaded region indicates the critical value of the control parameter found by QMC [6].

For larger system sizes, instead, we assess the accuracy of our wavefunction by means of the recently introduced variational score (V-score) [101], which allows to systematically compare Ansätze. For bosonic systems, this score can be computed as

$$\text{V-score} = \frac{L^d \mathrm{Var}[E_{\boldsymbol{\theta}}]}{(E_{\boldsymbol{\theta}} - E_{\mathrm{MF}})}, \tag{25}$$

Where L^d is the number of sites of the system, $\mathrm{Var}[E_{\boldsymbol{\theta}}]$ is the variance of the variational energy, and E_{MF} is the mean-field energy. The latter is given by $E_{\mathrm{MF}}/L^d = \bar{n}(U\bar{n}/2 - zJ)$ for Bose Hubbard with coordination number z ($z = 4$ here).

This quantity is intensive. Therefore, it allows one to systematically compare the accuracy of a simulations at varying system sizes. It vanishes when the variational Ansatz exactly matches the ground state by virtue of the zero-variance property. Furthermore, it was empirically verified [101] that the relative error on the variational energy of the ground state positively correlates to the V-score of the Ansatz following a universal trend. This can be used to roughly infer the error, which proves very useful whenever exact diagonalization is not available.

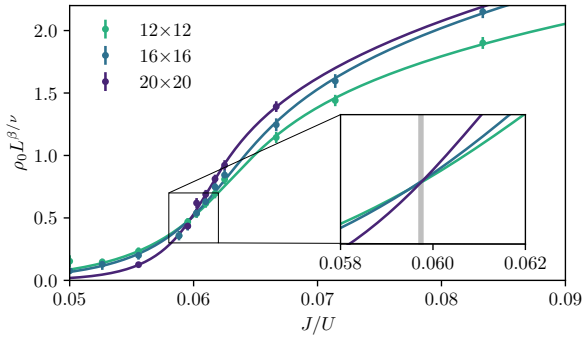

Figure 3: Rescaled condensate fraction as a function of the inverse interaction strength for increasing lattice sizes. Plain lines denote a fit of the scaling function in Eq. (26) to the variational data. The gray shaded region in the inset indicates the confidence interval for the critical value of the control parameter as found by QMC calculations [6]. As expected by scaling theory, all curves intersect for a same abscissa, showing good correspondence with the QMC estimation.

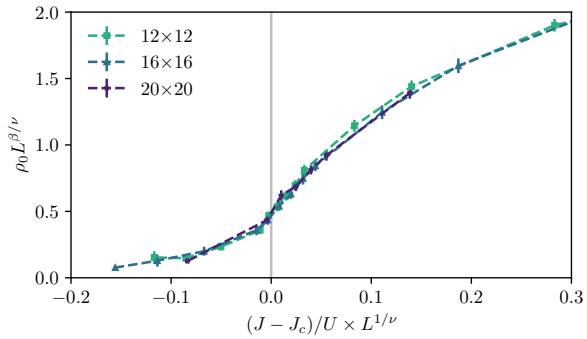

Figure 4: Data collapse of the data of Fig. 3 when plotted against the rescaled centered control parameter, showing the universal behavior of observables close to the critical point.

A V-score in the range 10^{-4}–10^{-6} signals a relative error roughly below 10^{-5}, indicating that the problem has been—for all practical purposes—solved variationally. This is the case, for instance, of spin models in the absence of magnetic frustration.

In Fig. 2, we show the V-score achieved by the backflow-Jastrow Ansatz for Bose Hubbard at unit filling on a periodic 2D lattice with $N = 16{\times}16$ and $N = 20{\times}20$ sites. The problem is clearly harder in the vicinity of the critical point, where the gap closes in the thermodynamic limit. A few ResNet layers in the backflow transformation dramatically improve the accuracy over bare Jastrow, lowering the V-score from 10^{-1} down to about $2{\times}10^{-4}$ for a depth of $D = 8$ layers. We observe that increasing the depth D of the backflow transformation consistently improves the accuracy. Interestingly, the depth required to reach a given accuracy does not seem to depend on the lattice size. Variational energies along with corresponding V-scores for all tested neural-backflow depths at $U/J = 16.8$ ($\sim U_c/J$) are provided in App. D.

4.2 Finite-size scaling

We can exploit the ability of our approach to scale to large system sizes to investigate the universal finite-size scaling of physical quantities of the system. According to the usual finite-size scaling argument, the order parameter ρ_0/N for a lattice of linear size L should depend on the control parameter $(J - J_c)/U$ as $\rho_0/N|_L \sim L^{-\beta/\nu}\tilde{f}(L^{1/\nu}(J - J_c)/U)$ where $\tilde{f}$ is some scaling function. Hence, when plotting $\rho_0/N|_L L^{\beta/\nu}$ against J/U for various values of L, all curves should intersect at the critical value J_c/U, provided the ratio of critical exponents β/ν is correct.

In Fig. 3, we perform the above analysis using the critical exponents of the 3D XY universality class [102]. Upon fitting to our variational data the fitting function

$$f(J/U) = \left[\mathrm{softplus}_a(J/U - b)\right]^c, \tag{26}$$

with $\mathrm{softplus}_\alpha(x) := \ln(1 + e^{\alpha x})/\alpha$, and parameters a, $b \sim J_c/U$ and c, all curves intersect within the confidence intervals of the best available estimation of the critical value of J/U [6]. In Fig. 4, we verify that all of the variational data collapse into a universal curve upon rescaling and centering the abscissa of Fig. 3.

4.3 Entanglement entropy

One of the strengths of variational approaches is the direct access to the wavefunction, in contrast with other techniques such as QMC. This allows us to compute quantities beyond usual linear forms on the wavefunction. In particular, a quantity of interest is the entanglement entropy, which quantizes the degree of entanglement of a subsystem composed of a subset of the lattice A and its complement $\bar{A}$. It is defined as the entropy of the reduced density $\hat{\rho}_A = \mathrm{Tr}_{\bar{A}}[|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|]$. While the von Neumann entropy is typically used, the

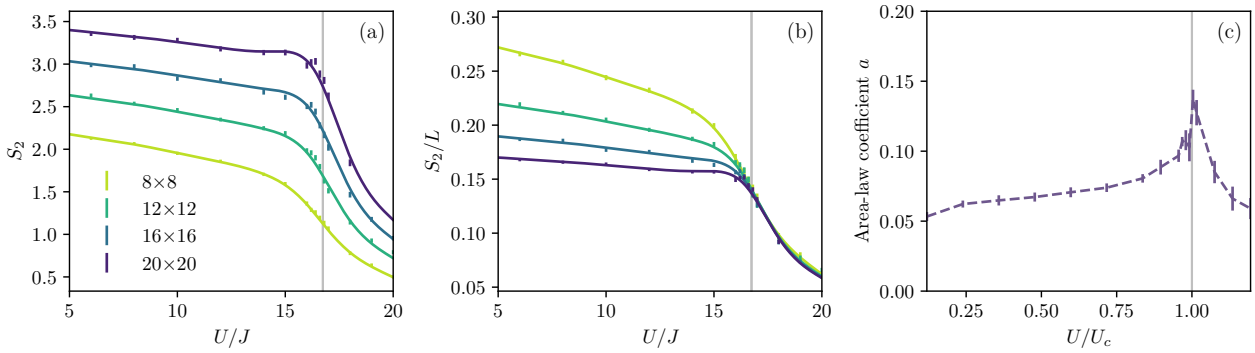

Figure 5: (a) Rényi-2 entropy $S_2(\hat{\rho}_A)$ evaluated on the reduced density matrix of the first half (A) of the lattice. (b) Entanglement entropy rescaled by the typical dimension of the subsystem's boundary length $|\partial A| \sim L$. In the Mott insulating phase, a strict area law is observed while the spontaneous breaking of the continuous symmetry induces a logarithmic correction in the superfluid phase. (c) Area-law coefficient a obtained by fitting Eq. (29) to the variational data for linear sizes $L \in \{8, 10, 12, 14, 16, 20\}$. In panels (a) and (b), errors are estimated by bootstrap; a cubic spline interpolation of the S_2 fit is added as a guide to the eye.

Rényi-2 entanglement entropy, as given by

$$S_2(\hat{\rho}_A) = -\ln(\mathrm{Tr}\,\hat{\rho}_A^2), \tag{27}$$

can be easily estimated via Monte Carlo sampling of two independent replicas through the estimator [103–105]

$$S_2(\hat{\rho}_A) = -\ln \mathbb{E}_{\boldsymbol{n},\boldsymbol{n}'}\left[\frac{\psi_{\boldsymbol{\theta}}(\boldsymbol{n}'_A, \boldsymbol{n}_{\bar{A}})\psi_{\boldsymbol{\theta}}(\boldsymbol{n}_A, \boldsymbol{n}'_{\bar{A}})}{\psi_{\boldsymbol{\theta}}(\boldsymbol{n}_A, \boldsymbol{n}_{\bar{A}})\psi_{\boldsymbol{\theta}}(\boldsymbol{n}'_A, \boldsymbol{n}'_{\bar{A}})}\right]. \tag{28}$$

In this estimator, samples $\boldsymbol{n}$ and $\boldsymbol{n}'$ are first drawn from two independent Markov chains and split into configurations of either complementary subspace $\boldsymbol{n}^{(\prime)} = (\boldsymbol{n}_A^{(\prime)}, \boldsymbol{n}_{\bar{A}}^{(\prime)})$. In the numerator, the wavefunction is evaluated on configurations generated by partially swapping the configurations of the replicas according to $\boldsymbol{n}_A \leftrightarrow \boldsymbol{n}'_A$. Importantly, when evaluating the above estimator, we enforce that $\psi_{\boldsymbol{\theta}}(\boldsymbol{n}) = 0$ whenever $\sum_i n_i \neq N$. In what follows, we shall consider the first half of the lattice, of dimension $L \times L/2$ and boundary length of $|\partial A| = 2L$, as the subsystem A.

Such a system is expected to exhibit an area-law scaling of the entanglement entropy. However, at finite size, the degeneracy of the ground state in the superfluid is slightly lifted. This *tower-of-states* mechanism [75] induces a correction in the superfluid phase scaling as the logarithm of the size of the boundary, resulting into the law [41]

$$S_2(\hat{\rho}_A) = aL + b\mathbb{1}[U > U_c]\ln L + c + O(1/L), \tag{29}$$

where $b = 1/2$ is a universal coefficient related to the number of Goldstone modes (1 for the 2D Bose-Hubbard and 3D XY models).

In Fig. 5(a), we show the Rényi-2 entropy $S_2(\hat{\rho}_A)$ as a function of the strength of the interactions. Upon increasing the lattice size, S_2 approaches a step-like singular behavior. This singular behavior is a signature of bosonic criticality [76]. In Fig. 5(b), we rescale this quantity by the the dimension of the boundary L. As in Ref. [41], we observe a strict area-law dependence of the entanglement entropy in the Mott phase along with a logarithmic departure in the superfluid phase. In Fig. 5(c), we further extract area-law coefficient a of Eq. (29). There, as predicted by Ref. [76], one observes a pronounced cusp at the critical point, a signature of the contribution of the Higgs mode to the entanglement entropy.

5 Conclusion

In this paper, we introduced the neural backflow-Jastrow architecture. We physically motivated its relevance for the study of interacting lattice bosons and analytically proved that it encompasses the structure of various standard variational architectures while providing a natural generalization. We benchmarked our Ansatz on the Bose-Hubbard problem and systematically assessed its variational accuracy, achieving a V-score as low as 2×10^{-4} on a 20×20 lattice. Its great scalability allowed us to scale up VMC

simulations to lattices with up to 20×20 sites on a single GPU with no local Hilbert-space truncation. Thanks to this, we performed a finite-size scaling analysis at zero temperature, showing remarkable agreement with the results obtained in QMC calculations by extrapolating finite-temperature results. Furthermore, we were able to investigate the entanglement properties of the system. We observed the logarithmic correction of the entanglement entropy due to the spontaneous breaking of the $U(1)$ continuous symmetry and were able to extract its universal scaling prefactor, which we found to display a second-order singularity at the critical point.

This work clears the way for the simulation of many bosonic lattice systems which remain out of reach for other techniques. A prime example is the simulation of bosons in the presence of a synthetic magnetic field [43]. Indeed, such systems are intrinsically plagued with the sign problem, thereby ruling out QMC, and set in two spatial dimensions, which challenges matrix produc states, hitherto constrained to narrow-ladder geometries. Another interesting prospect is the extension of this architecture to driven-dissipative scenarios, where accurate neural representations of the density matrix are still unavailable. This would enlarge the scope of the NQS approach to the field of quantum optics.

Acknowledgments

We thank Federico Becca for his valuable early input and for conducting the GFMC simulations presented in this study. Our gratitude extends to Stephen R Clark for sharing data from Ref. [65]. We also appreciate the helpful discussions with Markus Holzmann, Andreas M. Läuchli, David Clément, and Tommaso Roscilde. This work was supported by SEFRI through Grant No. MB22.00051 (NEQS - Neural Quantum Simulation).

All simulations were carried out with NetKet 3 [99]; the code used for this article can be found at Ref. [106].

References

[1] Ghassan George Batrouni, Richard T. Scalettar, and Gergely T. Zimanyi. Quantum critical phenomena in one-dimensional Bose systems. *Physical Review Letters*, 65 (14):1765–1768, 1990. DOI: 10.1103/PhysRevLett.65.1765.

[2] Werner Krauth, Nandini Trivedi, and David Ceperley. Superfluid-insulator transition in disordered boson systems. *Physical Review Letters*, 67(17):2307–2310, 1991. DOI: 10.1103/PhysRevLett.67.2307.

[3] Richard T. Scalettar, Ghassan George Batrouni, and Gergely T. Zimanyi. Localization in interacting, disordered, Bose systems. *Physical Review Letters*, 66(24): 3144–3147, 1991. DOI: 10.1103/PhysRevLett.66.3144.

[4] Ghassan George Batrouni and Richard T. Scalettar. World-line quantum Monte Carlo algorithm for a one-dimensional Bose model. *Physical Review B*, 46(14): 9051–9062, 1992. DOI: 10.1103/PhysRevB.46.9051.

[5] B. Capogrosso-Sansone, N. V. Prokof'ev, and B. V. Svistunov. Phase diagram and thermodynamics of the three-dimensional Bose-Hubbard model. *Physical Review B*, 75(13):134302, 2007. DOI: 10.1103/PhysRevB.75.134302.

[6] Barbara Capogrosso-Sansone, Şebnem Güneş Söyler, Nikolay Prokof'ev, and Boris Svistunov. Monte Carlo study of the two-dimensional Bose-Hubbard model. *Physical Review A*, 77(1):015602, 2008. DOI: 10.1103/PhysRevA.77.015602.

[7] Nandini Trivedi and D. M. Ceperley. Ground-state correlations of quantum antiferromagnets: A Green-function Monte Carlo study. *Physical Review B*, 41(7): 4552–4569, 1990. DOI: 10.1103/PhysRevB.41.4552.

[8] Matteo Calandra Buonaura and Sandro Sorella. Numerical study of the two-dimensional Heisenberg model using a Green function Monte Carlo technique with a fixed number of walkers. *Physical Review B*, 57(18):11446–11456, 1998. DOI: 10.1103/PhysRevB.57.11446.

[9] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 1st edition, 2017. DOI: 10.1017/9781316417041.

[10] Stefano Baroni and Saverio Moroni. Rep-

tation Quantum Monte Carlo: A Method for Unbiased Ground-State Averages and Imaginary-Time Correlations. *Physical Review Letters*, 82(24):4745–4748, 1999. DOI: 10.1103/PhysRevLett.82.4745.

[11] Giuseppe Carleo, Federico Becca, Saverio Moroni, and Stefano Baroni. Reptation quantum Monte Carlo algorithm for lattice Hamiltonians with a directed-update scheme. *Physical Review E*, 82(4):046710, 2010. DOI: 10.1103/PhysRevE.82.046710.

[12] Matthias Troyer and Uwe-Jens Wiese. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Physical Review Letters*, 94(17):170201, 2005. DOI: 10.1103/PhysRevLett.94.170201.

[13] Patrik Henelius and Anders W. Sandvik. Sign problem in Monte Carlo simulations of frustrated quantum spin systems. *Physical Review B*, 62(2):1102–1113, 2000. DOI: 10.1103/PhysRevB.62.1102.

[14] Philipp Hauke, Tommaso Roscilde, Valentin Murg, J. Ignacio Cirac, and Roman Schmied. Modified spin-wave theory with ordering vector optimization: Frustrated bosons on the spatially anisotropic triangular lattice. *New Journal of Physics*, 12(5):053036, 2010. DOI: 10.1088/1367-2630/12/5/053036.

[15] Gaopei Pan and Zi Yang Meng. The sign problem in quantum Monte Carlo simulations. In Tapash Chakraborty, editor, *Encyclopedia of Condensed Matter Physics (Second Edition)*, pages 879–893. Academic Press, Oxford, 2024. DOI: 10.1016/B978-0-323-90800-9.00095-0.

[16] M. Ö. Oktel, M. Niţă, and B. Tanatar. Mean-field theory for Bose-Hubbard model under a magnetic field. *Physical Review B*, 75(4):045133, 2007. DOI: 10.1103/PhysRevB.75.045133.

[17] G. Möller and N. R. Cooper. Condensed ground states of frustrated Bose-Hubbard models. *Physical Review A*, 82(6):063625, 2010. DOI: 10.1103/PhysRevA.82.063625.

[18] Sebastian D. Huber and Netanel H. Lindner. Topological transitions for lattice bosons in a magnetic field. *Proceedings of the National Academy of Sciences*, 108(50):19925–19930, 2011. DOI: 10.1073/pnas.1110813108.

[19] Arya Dhar, Maheswar Maji, Tapan Mishra, R. V. Pai, Subroto Mukerjee, and Arun Paramekanti. Bose-Hubbard model in a strong effective magnetic field: Emergence of a chiral Mott insulator ground state. *Physical Review A*, 85(4):041602, 2012. DOI: 10.1103/PhysRevA.85.041602.

[20] Arya Dhar, Tapan Mishra, Maheswar Maji, R. V. Pai, Subroto Mukerjee, and Arun Paramekanti. Chiral Mott insulator with staggered loop currents in the fully frustrated Bose-Hubbard model. *Physical Review B*, 87(17):174501, 2013. DOI: 10.1103/PhysRevB.87.174501.

[21] Akiyuki Tokuno and Antoine Georges. Ground states of a Bose–Hubbard ladder in an artificial magnetic field: Field-theoretical approach. *New Journal of Physics*, 16(7):073005, 2014. DOI: 10.1088/1367-2630/16/7/073005.

[22] F. Kolley, M. Piraud, I. P. McCulloch, U. Schollwöck, and F. Heidrich-Meisner. Strongly interacting bosons on a three-leg ladder in the presence of a homogeneous flux. *New Journal of Physics*, 17(9):092001, 2015. DOI: 10.1088/1367-2630/17/9/092001.

[23] Yin-Chen He, Subhro Bhattacharjee, R. Moessner, and Frank Pollmann. Bosonic Integer Quantum Hall Effect in an Interacting Lattice Model. *Physical Review Letters*, 115(11):116803, 2015. DOI: 10.1103/PhysRevLett.115.116803.

[24] Christian Romen and Andreas M. Läuchli. Chiral Mott insulators in frustrated Bose-Hubbard models on ladders and two-dimensional lattices: A combined perturbative and density matrix renormalization group study. *Physical Review B*, 98(5):054519, 2018. DOI: 10.1103/PhysRevB.98.054519.

[25] Tian-Sheng Zeng, D. N. Sheng, and W. Zhu. Continuous phase transition between bosonic integer quantum Hall liquid and a trivial insulator: Evidence for deconfined quantum criticality. *Physical Review B*, 101(3):035138, 2020. DOI: 10.1103/PhysRevB.101.035138.

[26] Ya-Fei Song and Shi-Jie Yang. Quantum

phases for bosons in a magnetic lattice with a harmonic trap. *New Journal of Physics*, 22(7):073001, 2020. DOI: 10.1088/1367-2630/ab95e0.

[27] Maximilian Buser, Claudius Hubig, Ulrich Schollwöck, Leticia Tarruell, and Fabian Heidrich-Meisner. Interacting bosonic flux ladders with a synthetic dimension: Ground-state phases and quantum quench dynamics. *Physical Review A*, 102(5):053314, 2020. DOI: 10.1103/PhysRevA.102.053314.

[28] Catalin-Mihai Halati and Thierry Giamarchi. Bose-Hubbard triangular ladder in an artificial gauge field. *Physical Review Research*, 5(1):013126, 2023. DOI: 10.1103/PhysRevResearch.5.013126.

[29] D. Jaksch and P. Zoller. Creation of effective magnetic fields in optical lattices: The Hofstadter butterfly for cold neutral atoms. *New Journal of Physics*, 5(1):56, 2003. DOI: 10.1088/1367-2630/5/1/356.

[30] Fabrice Gerbier and Jean Dalibard. Gauge fields for ultracold atoms in optical superlattices. *New Journal of Physics*, 12 (3):033007, 2010. DOI: 10.1088/1367-2630/12/3/033007.

[31] M. Aidelsburger, M. Atala, M. Lohse, J. T. Barreiro, B. Paredes, and I. Bloch. Realization of the Hofstadter Hamiltonian with Ultracold Atoms in Optical Lattices. *Physical Review Letters*, 111(18):185301, 2013. DOI: 10.1103/PhysRevLett.111.185301.

[32] Hirokazu Miyake, Georgios A. Siviloglou, Colin J. Kennedy, William Cody Burton, and Wolfgang Ketterle. Realizing the Harper Hamiltonian with Laser-Assisted Tunneling in Optical Lattices. *Physical Review Letters*, 111(18):185302, 2013. DOI: 10.1103/PhysRevLett.111.185302.

[33] J. Struck, M. Weinberg, C. Ölschläger, P. Windpassinger, J. Simonet, K. Sengstock, R. Höppner, P. Hauke, A. Eckardt, M. Lewenstein, and L. Mathey. Engineering Ising-XY spin-models in a triangular lattice using tunable artificial gauge fields. *Nature Physics*, 9(11):738–743, 2013. DOI: 10.1038/nphys2750.

[34] Marcos Atala, Monika Aidelsburger, Michael Lohse, Julio T. Barreiro, Belén Paredes, and Immanuel Bloch. Observation of chiral currents with ultracold atoms in bosonic ladders. *Nature Physics*, 10(8): 588–593, 2014. DOI: 10.1038/nphys2998.

[35] N. Goldman, G. Juzeliūnas, P. Öhberg, and I. B. Spielman. Light-induced gauge fields for ultracold atoms. *Reports on Progress in Physics*, 77(12):126401, 2014. DOI: 10.1088/0034-4885/77/12/126401.

[36] A. Sterdyniak, Nigel R. Cooper, and N. Regnault. Bosonic Integer Quantum Hall Effect in Optical Flux Lattices. *Physical Review Letters*, 115(11):116802, 2015. DOI: 10.1103/PhysRevLett.115.116802.

[37] Luca Barbiero, Josep Cabedo, Maciej Lewenstein, Leticia Tarruell, and Alessio Celi. Frustrated magnets without geometrical frustration in bosonic flux ladders. *Physical Review Research*, 5(4): L042008, 2023. DOI: 10.1103/PhysRevResearch.5.L042008.

[38] Manuela Capello, Federico Becca, Michele Fabrizio, Sandro Sorella, and Erio Tosatti. Variational Description of Mott Insulators. *Physical Review Letters*, 94 (2):026406, 2005. DOI: 10.1103/PhysRevLett.94.026406.

[39] Manuela Capello, Federico Becca, Michele Fabrizio, and Sandro Sorella. Superfluid to Mott-Insulator Transition in Bose-Hubbard Models. *Physical Review Letters*, 99(5):056402, 2007. DOI: 10.1103/PhysRevLett.99.056402.

[40] Manuela Capello, Federico Becca, Michele Fabrizio, and Sandro Sorella. Mott transition in bosonic systems: Insights from the variational approach. *Physical Review B*, 77(14):144517, 2008. DOI: 10.1103/PhysRevB.77.144517.

[41] Vincenzo Alba, Masudul Haque, and Andreas M. Läuchli. Entanglement Spectrum of the Two-Dimensional Bose-Hubbard Model. *Physical Review Letters*, 110 (26):260403, 2013. DOI: 10.1103/PhysRevLett.110.260403.

[42] Marek M. Rams, Piotr Czarnik, and Lukasz Cincio. Precise Extrapolation of the Correlation Function Asymptotics in Uniform Tensor Network States with Application to the Bose-Hubbard and XXZ Models. *Physical Review X*, 8(4):041033, 2018. DOI: 10.1103/PhysRevX.8.041033.

[43] Erik Lennart Weerda and Matteo Rizzi. Fractional quantum Hall states with variational projected entangled-pair states: A study of the bosonic Harper-Hofstadter model. *Physical Review B*, 109(24): L241117, 2024. DOI: 10.1103/PhysRevB.109.L241117.

[44] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1):277–306, 2010. DOI: 10.1103/RevModPhys.82.277.

[45] Ramesh V. Pai, Rahul Pandit, H. R. Krishnamurthy, and S. Ramasesha. One-Dimensional Disordered Bosonic Hubbard Model: A Density-Matrix Renormalization Group Study. *Physical Review Letters*, 76 (16):2937–2940, 1996. DOI: 10.1103/PhysRevLett.76.2937.

[46] T. D. Kühner and H. Monien. Phases of the one-dimensional Bose-Hubbard model. *Physical Review B*, 58(22):R14741–R14744, 1998. DOI: 10.1103/PhysRevB.58.R14741.

[47] S. Rapsch, U. Schollwöck, and W. Zwerger. Density matrix renormalization group for disordered bosons in one dimension. *Europhysics Letters*, 46(5):559, 1999. DOI: 10.1209/epl/i1999-00302-7.

[48] Corinna Kollath, Andreas M. Läuchli, and Ehud Altman. Quench Dynamics and Nonequilibrium Phase Diagram of the Bose-Hubbard Model. *Physical Review Letters*, 98(18):180601, 2007. DOI: 10.1103/PhysRevLett.98.180601.

[49] Andreas M. Läuchli and Corinna Kollath. Spreading of correlations and entanglement after a quench in the one-dimensional Bose–Hubbard model. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (05):P05018, 2008. DOI: 10.1088/1742-5468/2008/05/P05018.

[50] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355 (6325):602–606, 2017. DOI: 10.1126/science.aag2302.

[51] Yusuke Nomura and Masatoshi 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(3):031034, 2021. DOI: 10.1103/PhysRevX.11.031034.

[52] Ao Chen and Markus Heyl. Empowering deep neural quantum states through efficient optimization. *Nature Physics*, 20(9): 1476–1481, 2024. DOI: 10.1038/s41567-024-02566-1.

[53] Riccardo Rende, Luciano Loris Viteritti, Lorenzo Bardone, Federico Becca, and Sebastian Goldt. A simple linear algebra identity to optimize large-scale neural network quantum states. *Communications Physics*, 7(1):260, 2024. DOI: 10.1038/s42005-024-01732-4.

[54] Gabriel Pescia, Jiequn Han, Alessandro Lovato, Jianfeng Lu, and Giuseppe Carleo. Neural-network quantum states for periodic systems in continuous space. *Physical Review Research*, 4(2):023138, 2022. DOI: 10.1103/PhysRevResearch.4.023138.

[55] Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic Schrödinger equation. *Nature Chemistry*, 12(10):891–897, 2020. DOI: 10.1038/s41557-020-0544-y.

[56] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Physical Review Research*, 2(3):033429, 2020. DOI: 10.1103/PhysRevResearch.2.033429.

[57] Gabriel Pescia, Jannes Nys, Jane Kim, Alessandro Lovato, and Giuseppe Carleo. Message-passing neural quantum states for the homogeneous electron gas. *Phys. Rev. B*, 110:035108, 2024. DOI: 10.1103/PhysRevB.110.035108.

[58] Di Luo, Giuseppe Carleo, Bryan K. Clark, and James Stokes. Gauge Equivariant Neural Networks for Quantum Lattice Gauge Theories. *Physical Review Letters*, 127 (27):276402, 2021. DOI: 10.1103/PhysRevLett.127.276402.

[59] Matija Medvidović and Dries Sels. Variational Quantum Dynamics of Two-Dimensional Rotor Models. *PRX Quantum*, 4(4):040302, 2023. DOI: 10.1103/PRXQuantum.4.040302.

[60] Hiroki Saito. Solving the Bose–Hubbard Model with Machine Learning. *Journal of*

the Physical Society of Japan, 86(9):093001, 2017. DOI: 10.7566/JPSJ.86.093001.

[61] Kristopher McBrian, Giuseppe Carleo, and Ehsan Khatami. Ground state phase diagram of the one-dimensional Bose-Hubbard model from restricted Boltzmann machines. *Journal of Physics: Conference Series*, 1290(1):012005, 2019. DOI: 10.1088/1742-6596/1290/1/012005.

[62] Vladimir Vargas-Calderón, Herbert Vinck-Posada, and Fabio A. González. Phase Diagram Reconstruction of the Bose–Hubbard Model with a Restricted Boltzmann Machine Wavefunction. *Journal of the Physical Society of Japan*, 89(9):094002, 2020. DOI: 10.7566/JPSJ.89.094002.

[63] Ziyan Zhu, Marios Mattheakis, Weiwei Pan, and Efthimios Kaxiras. HubbardNet: Efficient predictions of the Bose-Hubbard model spectrum with deep neural networks. *Physical Review Research*, 5(4): 043084, 2023. DOI: 10.1103/PhysRevResearch.5.043084.

[64] Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. Symmetries and Many-Body Excitations with Neural-Network Quantum States. *Physical Review Letters*, 121(16):167204, 2018. DOI: 10.1103/PhysRevLett.121.167204.

[65] Michael Y Pei and Stephen R Clark. Specialising neural-network quantum states for the bose hubbard model. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 57(21):215301, 2024. DOI: 10.1088/1361-6455/ad7e87.

[66] R. P. Feynman and Michael Cohen. Energy Spectrum of the Excitations in Liquid Helium. *Physical Review*, 102(5):1189–1204, 1956. DOI: 10.1103/PhysRev.102.1189.

[67] Michael A. Lee, K. E. Schmidt, M. H. Kalos, and G. V. Chester. Green's Function Monte Carlo Method for Liquid ^{3}He. *Physical Review Letters*, 46(11):728–731, 1981. DOI: 10.1103/PhysRevLett.46.728.

[68] Yongkyung Kwon, D. M. Ceperley, and Richard M. Martin. Effects of three-body and backflow correlations in the two-dimensional electron gas. *Physical Review B*, 48(16):12037–12046, 1993. DOI: 10.1103/PhysRevB.48.12037.

[69] Yongkyung Kwon, D. M. Ceperley, and Richard M. Martin. Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study. *Physical Review B*, 58(11):6800–6806, 1998. DOI: 10.1103/PhysRevB.58.6800.

[70] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler. Backflow correlations for the electron gas and metallic hydrogen. *Physical Review E*, 68(4):046707, 2003. DOI: 10.1103/PhysRevE.68.046707.

[71] Luca F. Tocchio, Federico Becca, Alberto Parola, and Sandro Sorella. Role of backflow correlations for the nonmagnetic phase of the t–t' Hubbard model. *Physical Review B*, 78(4):041101, 2008. DOI: 10.1103/PhysRevB.78.041101.

[72] Luca F. Tocchio, Federico Becca, and Claudius Gros. Backflow correlations in the Hubbard model: An efficient tool for the study of the metal-insulator transition and the large-U limit. *Physical Review B*, 83(19):195138, 2011. DOI: 10.1103/PhysRevB.83.195138.

[73] Michele Ruggeri, Saverio Moroni, and Markus Holzmann. Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space. *Physical Review Letters*, 120(20):205302, 2018. DOI: 10.1103/PhysRevLett.120.205302.

[74] Di Luo and Bryan K. Clark. Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions. *Physical Review Letters*, 122 (22):226401, 2019. DOI: 10.1103/PhysRevLett.122.226401.

[75] Max A. Metlitski and Tarun Grover. Entanglement Entropy of Systems with Spontaneously Broken Continuous Symmetry, 2015.

[76] Irénée Frérot and Tommaso Roscilde. Entanglement Entropy across the Superfluid-Insulator Transition: A Signature of Bosonic Criticality. *Physical Review Letters*, 116(19):190401, 2016. DOI: 10.1103/PhysRevLett.116.190401.

[77] Rajibul Islam, Ruichao Ma, Philipp M. Preiss, M. Eric Tai, Alexander Lukin, Matthew Rispoli, and Markus Greiner. Measuring entanglement entropy in a quantum many-body system. *Nature*, 528(7580): 77–83, 2015. DOI: 10.1038/nature15750.

[78] H. A. Gersch and G. C. Knollman. Quantum Cell Model for Bosons. *Physical Review*, 129(2):959–967, 1963. DOI: 10.1103/PhysRev.129.959.

[79] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller. Cold Bosonic Atoms in Optical Lattices. *Physical Review Letters*, 81(15):3108–3111, 1998. DOI: 10.1103/PhysRevLett.81.3108.

[80] Markus Greiner, Olaf Mandel, Tilman Esslinger, Theodor W. Hänsch, and Immanuel Bloch. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature*, 415(6867): 39–44, 2002. DOI: 10.1038/415039a.

[81] I. B. Spielman, W. D. Phillips, and J. V. Porto. Mott-Insulator Transition in a Two-Dimensional Atomic Bose Gas. *Physical Review Letters*, 98(8):080404, 2007. DOI: 10.1103/PhysRevLett.98.080404.

[82] I. B. Spielman, W. D. Phillips, and J. V. Porto. Condensate Fraction in a 2D Bose Gas Measured across the Mott-Insulator Transition. *Physical Review Letters*, 100 (12):120402, 2008. DOI: 10.1103/PhysRevLett.100.120402.

[83] Immanuel Bloch, Jean Dalibard, and Wilhelm Zwerger. Many-body physics with ultracold gases. *Reviews of Modern Physics*, 80(3):885–964, 2008. DOI: 10.1103/RevModPhys.80.885.

[84] Manuel Endres, Takeshi Fukuhara, David Pekker, Marc Cheneau, Peter Schauß, Christian Gross, Eugene Demler, Stefan Kuhr, and Immanuel Bloch. The 'Higgs' amplitude mode at the two-dimensional superfluid/Mott insulator transition. *Nature*, 487(7408):454–458, 2012. DOI: 10.1038/nature11255.

[85] A. van Otterlo, K.-H. Wagenblast, Rosario Fazio, and Gerd Schön. Response of Josephson-junction arrays near the quantum phase transition. *Physical Review B*, 48 (5):3316–3326, 1993. DOI: 10.1103/PhysRevB.48.3316.

[86] C. Bruder, R. Fazio, and G. Schön. The Bose-Hubbard model: from Josephson junction arrays to optical lattices. *Annalen der Physik*, 517(9-10):566–577, 2005. DOI: 10.1002/andp.200551709-1005.

[87] C. N. Yang. Concept of Off-Diagonal Long-Range Order and the Quantum Phases of Liquid He and of Superconductors. *Reviews of Modern Physics*, 34(4):694–704, 1962. DOI: 10.1103/RevModPhys.34.694.

[88] Matthew P. A. Fisher, Peter B. Weichman, G. Grinstein, and Daniel S. Fisher. Boson localization and the superfluid-insulator transition. *Physical Review B*, 40(1):546–570, 1989. DOI: 10.1103/PhysRevB.40.546.

[89] W. Krauth and N. Trivedi. Mott and Superfluid Transitions in a Strongly Interacting Lattice Boson System. *Europhysics Letters*, 14(7):627, 1991. DOI: 10.1209/0295-5075/14/7/003.

[90] J. K. Freericks and H. Monien. Strong-coupling expansions for the pure and disordered Bose-Hubbard model. *Physical Review B*, 53(5):2691–2700, 1996. DOI: 10.1103/PhysRevB.53.2691.

[91] N. Elstner and H. Monien. Dynamics and thermodynamics of the Bose-Hubbard model. *Physical Review B*, 59(19): 12184–12187, 1999. DOI: 10.1103/PhysRevB.59.12184.

[92] A. Bijl. The lowest wave function of the symmetrical many particles system. *Physica*, 7(9):869–886, 1940. DOI: 10.1016/0031-8914(40)90166-5.

[93] Robert Jastrow. Many-Body Problem with Strong Forces. *Physical Review*, 98(5): 1479–1484, 1955. DOI: 10.1103/PhysRev.98.1479.

[94] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. Transformer Variational Wave Functions for Frustrated Quantum Spin Systems. *Physical Review Letters*, 130(23):236401, 2023. DOI: 10.1103/PhysRevLett.130.236401.

[95] Luciano Loris Viteritti, Riccardo Rende, Alberto Parola, Sebastian Goldt, and Federico 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. DOI: 10.1103/PhysRevB.111.134411.

[96] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Ma-*

chine Intelligence, 35(8):1798–1828, 2013. DOI: 10.1109/TPAMI.2013.50.

[97] Sandro Sorella. Wave function optimization in the variational Monte Carlo method. *Physical Review B*, 71(24):241103, 2005. DOI: 10.1103/PhysRevB.71.241103.

[98] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum Natural Gradient. *Quantum*, 4:269, 2020. DOI: 10.22331/q-2020-05-25-269.

[99] Filippo Vicentini, Damian Hofmann, Attila Szabó, Dian Wu, Christopher Roth, Clemens Giuliani, Gabriel Pescia, Jannes Nys, Vladimir Vargas-Calderón, Nikita Astrakhantsev, and Giuseppe Carleo. NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems. *SciPost Physics Codebases*, page 007, 2022. DOI: 10.21468/SciPostPhysCodeb.7.

[100] Federico Becca. Private communication, 2023.

[101] Dian Wu, Riccardo Rossi, Filippo Vicentini, Nikita Astrakhantsev, Federico Becca, Xiaodong Cao, Juan Carrasquilla, Francesco Ferrari, Antoine Georges, Mohamed Hibat-Allah, Masatoshi Imada, Andreas M. Läuchli, Guglielmo Mazzola, Antonio Mezzacapo, Andrew Millis, Javier Robledo Moreno, Titus Neupert, Yusuke Nomura, Jannes Nys, Olivier Parcollet, Rico Pohle, Imelda Romero, Michael Schmid, J. Maxwell Silvester, Sandro Sorella, Luca F. Tocchio, Lei Wang, Steven R. White, Alexander Wietek, Qi Yang, Yiqi Yang, Shiwei Zhang, and Giuseppe Carleo. Variational benchmarks for quantum many-body problems. *Science*, 386(6719):296–301, 2024. DOI: 10.1126/science.adg9774.

[102] Massimo Campostrini, Martin Hasenbusch, Andrea Pelissetto, Paolo Rossi, and Ettore Vicari. Critical behavior of the three-dimensional XY universality class. *Physical Review B*, 63(21):214503, 2001. DOI: 10.1103/PhysRevB.63.214503.

[103] Matthew B. Hastings, Iván González, Ann B. Kallin, and Roger G. Melko. Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations. *Physical Review Letters*, 104(15):157201, 2010. DOI: 10.1103/PhysRevLett.104.157201.

[104] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018. DOI: 10.1038/s41567-018-0048-5.

[105] Jiarui Zhao, Bin-Bin Chen, Yan-Cheng Wang, Zheng Yan, Meng Cheng, and Zi Yang Meng. Measuring Rényi entanglement entropy with high efficiency and precision in quantum Monte Carlo simulations. *npj Quantum Materials*, 7(1):1–9, 2022. DOI: 10.1038/s41535-022-00476-0.

[106] Zakari Denis. cqsl/lattice-bosons: v1.0.0, June 2025. URL `https://doi.org/10.5281/zenodo.15632176`.

A Holon-doublon confinement

Close to the Mott-to-superfluid transition, the state of the system can be expressed as a perturbative expansion around the $U/J \sim +\infty$ limit. To leading order, this yields the following correction to the Mott state $|\psi_\infty\rangle$:

$$|\psi_U\rangle - |\psi_\infty\rangle \propto \frac{J}{U} \sum_{\langle i,j\rangle} (\hat{a}_i^\dagger \hat{a}_j + \hat{a}_j^\dagger \hat{a}_i) |\psi_\infty\rangle + O((J/U)^2). \tag{30}$$

At unit filling this corresponds to the emergence of isolated holon-doublon bound pairs at any neighboring pair of sites. Similarly, higher-order contributions involve additional pairs, or pairs separated by a larger distance. Within the Mott phase, such an expansion converges and holon-doublon pairs must thus be spatially confined exponentially in order to stabilize the phase.

Such a confinement can be encoded in the variational wavefunction in the form of an attractive potential [38]

$$|\psi'\rangle = e^{-\sum_{ij} \hat{h}_i V_{d_{ij}} \hat{d}_j} |\psi\rangle, \tag{31}$$

where $V_d \leq 0$, for $0 < d \leq R$, and $V_d = 0$ otherwise. Upon assuming any holon (doublon) is not surrounded by more than one doublon (holon), the many-body projector of Eq. (5) is a particular case of the above withrange $R = 1$ and $V_1 = -g_{\mathrm{MB}}$. The $R > 1$ case can be easily represented by a convolutional network. Indeed, upon neglecting occupations above 2 bosons per site and assuming that pairs are distant from each other by more than R, one has

$$-\sum_{ij} V_{d_{ij}} \langle \boldsymbol{n}| \hat{h}_i \hat{d}_j |\boldsymbol{n}\rangle = \sum_i h_{i,1}^{(2)}(\boldsymbol{n}), \tag{32}$$

$$h_{i,1}^{(2)}(\boldsymbol{n}) = \mathrm{ReLU}\Big(-V_0 \sum_{\mu=1}^{2} h_{i,\mu}^{(1)}(\boldsymbol{n}) - \sum_{d=1}^{R} V_d h_{i,2+d}^{(1)}(\boldsymbol{n})\Big), \tag{33}$$

$$h_{i,\mu\leq 2}^{(1)}(\boldsymbol{n}) = \mathrm{ReLU}\Big((-1)^\mu (n_i + 1)\Big), \tag{34}$$

$$h_{i,\mu=2+d}^{(1)}(\boldsymbol{n}) = \mathrm{ReLU}\Big(\sum_{\boldsymbol{v}\in\mathcal{V}} \delta_{v,d} n_{T_{\boldsymbol{v}}(i)}\Big), \tag{35}$$

with $V_0 = \sum_{d>0} |V_d|$ and where input occupations are shifted by the unit density such that configurations $n = -1$ and $n = 1$ correspond respectively to a hole and a doublon, such as in the main text. This bears the form of a two-layer convolutional network with parameters

$$\begin{aligned} K_{\boldsymbol{v},\mu\leq 2,1}^{(1)} &= (-1)^\mu \delta_{v,0} \in \mathbb{R}^{|\mathcal{V}|\times 2\times 1}, \quad b_\mu^{(1)} = (-1)^\mu \in \mathbb{R}^2, \\ K_{\boldsymbol{v},\mu=2+d,1}^{(1)} &= \delta_{v,d} \in \mathbb{R}^{|\mathcal{V}|\times R\times 1}, \quad b_\mu^{(1)} = 0 \in \mathbb{R}^R, \\ K_{\boldsymbol{0},1,\mu'}^{(2)} &= -V_\mu \in \mathbb{R}^{1\times 1\times 2}, \quad b_1^{(2)} = 0 \in \mathbb{R}. \end{aligned} \tag{36}$$

B Complexity of the neural backflow Jastrow Ansatz

We here assess the complexity of evaluating our variational wavefunction, as defined in Eq. (18).

The translationally invariant two-body Jastrow operation has a complexity of $O(L^{2d})$ in d spatial dimensions for the simplest algorithm. The neural backflow transformation of Eq. (19) has a complexity scaling as $O(L^d \times D\alpha W^d)$, with D the depth of the network, α the number of convolutional features, W the width of the filters, and L the lattice lateral dimension.

The above estimation concerns the complexity of evaluating amplitudes of the Ansatz at fixed neural-network architecture. However, to keep a comparable relative error on the energy across system sizes, some hyperparameters of the variational wavefunction must carry a dependence in L. In any gapped phase, the system has quantum correlations with fixed correlation length irrespective of system size. In this regime, we expect the long-range correlations to be captured by the Jastrow, while short-range quantum correlations at a scale ξ_c to be accounted for by a convolutional backflow transformation with

a receptive field of size $\sim \xi_c^d$. In this regime, the complexity of the backflow transformation scales as $O(\alpha L^d \xi_c^d)$ for a shallow CNN and as $O(\alpha L^d \xi_c)$ for a deep one, and is thus always proportional to the system size. In a critical region, in contrast, the receptive field must be extensive $\sim L^d$, which incurs a complexity scaling of either $O(\alpha L^{2d})$ or $O(\alpha L^{2d+1})$ for a shallow or deep CNN, respectively.

Scaling the size of the system also has consequences in terms of the number of parameters, which ultimately impacts the complexity scaling of stochastic reconfiguration. In any regime, the Jastrow term has a number of parameters of $\sim dL/2$, whereas for the backflow transformation this scales as $O(\alpha \xi_c^d)$ for a shallow CNN and $O(\alpha \xi_c)$ for a deep one. It follows, that the backflow proposed in this work only involves a constant parameter overhead over Jastrow for gapped systems, and a parameter scaling similar to that of the bare Jastrow in critical regions for deep architectures.

From this analysis, it also appears that it is always better performance-wise to increase the depth of the backflow transformation as the system approaches the critical point rather than to increase the filter size.

B.1 Metropolis-Hastings transition rule

In all simulations, Markov-chain Monte Carlo (MCMC) was used in order to sample from the Born probability distribution $p_{\boldsymbol{\theta}}(\boldsymbol{n}) := |\psi_{\boldsymbol{\theta}}(\boldsymbol{n})|^2/\|\psi_{\boldsymbol{\theta}}\|^2$. At each MCMC step, an update from the current configuration $\boldsymbol{n}$ into a new configuration $\boldsymbol{n}'$ is proposed with probability given by the proposition distribution $g(\boldsymbol{n}'|\boldsymbol{n})$, and accepted with probability

$$p_{\mathrm{acc}}(\boldsymbol{n} \to \boldsymbol{n}') = \min\left(1, \frac{p_{\boldsymbol{\theta}}(\boldsymbol{n}')}{p_{\boldsymbol{\theta}}(\boldsymbol{n})} \frac{g(\boldsymbol{n}|\boldsymbol{n}')}{g(\boldsymbol{n}'|\boldsymbol{n})}\right). \tag{37}$$

Our proposition distribution is induced by the first-quantization local transition rule $g^{(1)}(\mathbf{x}'|\mathbf{x}) = \prod_\mu^N g_{\mathrm{loc}}^{(1)}(\boldsymbol{x}'_\mu|\boldsymbol{x}_\mu)$ with

$$g_{\mathrm{loc}}^{(1)}(\boldsymbol{x}'_\mu|\boldsymbol{x}_\mu) \propto \mathbb{1}\left[\|\boldsymbol{x}_\mu - \boldsymbol{x}'_\mu\| = a\right], \tag{38}$$

where $\boldsymbol{x}_\mu$ denotes the μth particle's position and a is the lattice parameter. This proposal is clearly ergodic and corresponds to the Hamiltonian rule for a non-interacting system. It simply moves a particle picked at random into any of its neighboring sites. In second quantization, this can be achieved through the update kernel $g^{(2)}(\boldsymbol{n}'|\boldsymbol{n}) = \prod_{\langle i,j\rangle} g_{\mathrm{loc}}^{(2)}(n'_i, n'_j|n_i, n_j)$ with

$$g_{\mathrm{loc}}^{(2)}(n'_i, n'_j|n_i, n_j) \propto n_i \delta_{n'_i, n_i-1} \delta_{n'_j, n_j+1}. \tag{39}$$

This corresponds to choosing a site with a probability proportional to its occupation number and then transferring one of the occupations to any neighboring site. This kernel is not symmetric, hence, a correction is needed to enforce detail balance. This is accounted for in Eq. (37) thanks to the following ratio of proposal densities:

$$\frac{g_{\mathrm{loc}}^{(2)}(n_i, n_j|n'_i, n'_j)}{g_{\mathrm{loc}}^{(2)}(n'_i, n'_j|n_i, n_j)} = \frac{n'_j}{n_i}. \tag{40}$$

C Optimization procedure

In all simulations the learning rate was set to $\eta = 1 \times 10^{-3}$, the diagonal-shift regularization of the quantum geometric tensor was set to $\lambda = 5 \times 10^{-4}$ (1×10^{-3}) for the (neural backflow) Jastrow. 8192 samples were used for most of the VMC calculations, and 12 288 for backflow transformations of depth $D = 8$ for the largest lattice, while the number of samples was set to 32 768 for our simulations in Table 1.

D Worst-case variational figures of merit

Variational Monte Carlo proves to be the most demanding in the Mott phase at the vicinity of the critical point U_c^+/J. In Table 2, we compile the variational energies and corresponding variational scores in that challenging region ($U/J = 16.8$) as a function of system size and depth D of the neural backflow transformation.

Lattice size	Depth	$E_{\boldsymbol{\theta}}/JL^2$	V-score
8×8	0	$-0.452(2)$	8.4×10^{-2}
	6	$-0.54783(6)$	5.6×10^{-5}
12×12	0	$-0.444(1)$	7.9×10^{-2}
	6	$-0.54320(7)$	1.6×10^{-4}
16×16	0	$-0.446(1)$	8.1×10^{-2}
	2	$-0.5375(2)$	3.7×10^{-3}
	4	$-0.5411(1)$	7.3×10^{-4}
	6	$-0.54191(7)$	3.0×10^{-4}
20×20	0	$-0.4414(9)$	7.8×10^{-2}
	2	$-0.5373(1)$	3.1×10^{-3}
	4	$-0.54119(5)$	6.8×10^{-4}
	6	$-0.54162(5)$	3.1×10^{-4}
	8	$-0.54182(4)$	1.8×10^{-4}

Table 2: Variational figures of merit at $U/J = 16.8$. 12 288 samples were used in optimizing the networks of depth $D = 6$ and $D = 8$ for the largest system size, 8192 otherwise.