

Neural-network quantum states for periodic systems in continuous space

Gabriel Pescia¹, Jiequn Han², Alessandro Lovato^{3,4,5}, Jianfeng Lu⁶, and Giuseppe Carleo¹¹*Ecole Polytechnique Fédérale de Lausanne (EPFL), Institute of Physics, CH-1015 Lausanne, Switzerland*²*Center for Computational Mathematics, Flatiron Institute, New York, New York 10010, USA*³*Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*⁴*Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*⁵*INFN-TIFPA Trento Institute of Fundamental Physics and Applications, 38123 Trento, Italy*⁶*Departments of Mathematics, Physics, and Chemistry, Duke University, Durham, North Carolina 27708, USA*

(Received 7 January 2022; accepted 12 April 2022; published 20 May 2022)

We introduce a family of neural quantum states for the simulation of strongly interacting systems in the presence of spatial periodicity. Our variational state is parametrized in terms of a permutationally invariant part described by the Deep Sets neural-network architecture. The input coordinates to the Deep Sets are periodically transformed such that they are suitable to directly describe periodic bosonic systems. We show example applications to both one- and two-dimensional interacting quantum gases with Gaussian interactions, as well as to ^4He confined in a one-dimensional geometry. For the one-dimensional systems we find very precise estimations of the ground-state energies and the radial distribution functions of the particles. In two dimensions we obtain good estimations of the ground-state energies, comparable to results obtained from more conventional methods.

DOI: [10.1103/PhysRevResearch.4.023138](https://doi.org/10.1103/PhysRevResearch.4.023138)

I. INTRODUCTION

In recent years, the field of machine learning has seen tremendous progress in various applications of high-dimensional data analysis such as image recognition, language processing, or classification tasks [1,2]. A large portion of this progress is achieved by incorporating *a priori* known structure in the data into the learning algorithm. In this way, it is possible to reduce the set of possible solutions of a learning algorithm to the relevant ones [3–5]. Because of the omnipresence of symmetries in modern physics, the application of these techniques to computational physics has become a very active field of research that has produced promising results, especially in the area of many-body quantum physics [6]. In this field, since the early research work, considerable attention has been devoted to the problem of approximating ground-state wave functions of spin systems through artificial neural networks. These representations, known as neural-network quantum states (NQSs) [7], can encode highly entangled wave functions [8–10] and are routinely used to study correlated quantum systems with discrete degrees of freedom [11–18], often improving upon existing state-of-the-art results.

More recently, NQSs have been extended to study ground-state properties of fermionic systems in continuous space, introducing deep neural-network architectures that by design

satisfy the Pauli exclusion principle. The focus of these approaches has been primarily on relatively small atomic and molecular systems [19–24], as well as nuclear physics [25,26]. These applications have already shown significant improvements over more traditional, physics-inspired wave functions. However, for these tools to also become a compelling, more accurate alternative *ab initio* approach for the prediction of electronic structure properties, and bulk properties in different phases of matter, it is crucial that they are extended to periodic systems. While progress in this direction has been already realized in the context of lattice-based NQSs [27], an open methodological issue is the extension of continuous-space NQSs to efficiently encode periodicity, while preserving other fundamental symmetries, most importantly particle-permutation invariance.

The ability to treat strongly correlated periodic systems in continuous space is of chief importance in a wide range of condensed-matter physics, nuclear physics, and quantum chemistry problems. Emblematic examples of strongly interacting periodic systems in condensed matter are the electron gas [28], the bulk of bosonic and fermionic helium [29–31], supersolids [32,33], and high-pressure hydrogen [34,35]. In nuclear physics, the matter comprising the interior of neutron stars is typically modeled as a periodic system of strongly interacting protons and neutrons [36–38]. It has to be noted that neural networks have been recently employed to reduce the associated finite-size effects, extrapolating quantum Monte Carlo calculations of neutron matter and unitary gas to the thermodynamic limit [39].

In all these problems, physics-driven wave functions combined with quantum Monte Carlo methods have played a crucial role in the understanding of key phenomena, including

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

superfluidity, superconductivity, and crystallization. Notable examples of these wave functions include Jastrow correlators [40] applied to a mean-field Slater determinant, which may include backflow correlations [41–43]. Despite their success at describing key physical phenomena, these ansatz state stems are not systematically improvable and typically require significant adjustments when used on systems different from the one they were originally designed for.

In this paper, we introduce continuous-space NQSs based on a periodic transformation of the single- and two-body coordinates of particles that are suitable for the description of periodic systems of interacting bosons. These NQS variational states are by construction permutation invariant with respect to particle exchanges. We demonstrate their flexibility in solving nonrelativistic Hamiltonians, including those relevant for liquid helium and soft Gaussian cores, for different densities and system sizes.

II. METHODS

Throughout this paper we will consider the many-body Hamiltonian of N interacting particles in d -dimensional continuous space, given by

$$H = T + V = -\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 + V(x), \quad (1)$$

where T and V denote the kinetic and potential energy, respectively. The continuous position variable $x = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ denotes the set of all N single-particle positions $\mathbf{x}_i \in \mathbb{R}^d$. We will confine the single-particle positions to a finite box of length L , equipped with periodic boundaries, such that the methods introduced hereinafter are well suited for the study of the bulk properties of a variety of quantum systems as well as intrinsically periodic structures such as solids.

A. Periodic boundary conditions

In the presence of periodic boundary conditions (PBCs), the bare single-particle coordinates are not well suited as input to our variational ansatz, since they do not reflect the periodicity of the simulation cell. The wave function must be invariant under the translation of a single particle by a primitive lattice vector $\mathbf{L}_j = L\mathbf{e}_j$, i.e., $\psi_\alpha(\mathbf{x}_1, \dots, \mathbf{x}_i + \mathbf{L}_j, \dots, \mathbf{x}_N) = \psi_\alpha(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N)$ ($\mathbf{e}_j$ in our case denotes the Euclidean standard basis vector of $\mathbb{R}^d$) [44].

To respect the boundary conditions, we propose to map the single-particle coordinates to a truncated L -periodic Fourier basis [45]

$$\begin{aligned} \mathbf{x}_i &\mapsto \left(\sin\left(\frac{2\pi k}{L}\mathbf{x}_i\right), \cos\left(\frac{2\pi k}{L}\mathbf{x}_i\right) \right)_{k=1}^K \\ &= \mathbf{x}_i^{(K)} \in \mathbb{R}^{d \cdot 2K}. \end{aligned} \quad (2)$$

Here, $\sin(\mathbf{x}_i)$ and $\cos(\mathbf{x}_i)$ mean applying the trigonometric functions componentwise to the vector $\mathbf{x}_i$. Note that $\mathbf{x}_i^{(K)}$ is invariant under the translation $\mathbf{x}_i \mapsto \mathbf{x}_i + \mathbf{L}_j$, and thus the whole wave-function ansatz is invariant.

In order to respect the periodicity of the simulation box also in the computation of interparticle distances, we will use

$d_{\sin}(\mathbf{x}_i, \mathbf{x}_j) = \|\frac{L}{2} \sin(\frac{\pi}{L}\mathbf{r}_{ij})\|$ with $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ as a surrogate for the ordinary Euclidean distance ($d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|$) in the variational ansatz. Distances in the potential energy will be computed using the minimum image convention [46].

B. Variational ansatz

We construct the variational wave-function ansatz ψ_α as a product of a short- and a long-range part. The former is determined by requiring that the diverging behavior of the potential $V(x)$ at short interparticle distances is compensated by the kinetic energy contribution of the wave function, such that we obtain a finite energy contribution:

$$\lim_{r \rightarrow 0} \left(\frac{\nabla^2 \psi(r)}{\psi(r)} + V(x) \right) < \infty. \quad (3)$$

This is called *Kato's cusp condition* and constitutes a boundary condition for the given many-body system [47].

The long-range part is parametrized by a neural network based on the *Deep Sets* (DSs) architecture, which builds on the fact that a function $f(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is invariant under permutations of its input if and only if it can be decomposed as follows [48,49]:

$$f(\mathbf{x}_1, \dots, \mathbf{x}_N) = \rho(\text{Pool}[\phi(\mathbf{x})]), \quad (4)$$

where $\mathbf{x} \in \mathbb{R}^{N \times d}$ is the matrix of all single-particle coordinates $\mathbf{x}_i$. The two vector functions $\rho(\mathbf{x})$ and $\phi(\mathbf{x})$ depend on the function $f(\mathbf{x}_1, \dots, \mathbf{x}_N)$ [$\phi(\mathbf{x})$ is applied row-wise to $\mathbf{x}$]. The pooling operation $\text{Pool}[\phi(\mathbf{x})]$ denotes the mixing of the feature vectors $\mathbf{y}_i := \phi(\mathbf{x}_i)$ such that the output is permutationally invariant with respect to the particle index (see Fig. 1) [48]. Examples of such pooling operations are the sum pooling $\text{Pool}(\mathbf{y}) = \sum_i \mathbf{y}_i$ and log sum exp pooling ($L\Sigma E$), defined as $L\Sigma E(\mathbf{y}) = \log(\sum_i \exp[\mathbf{y}_i])$, which is a continuous version of the *max* pooling.

The function ϕ can be seen as an encoding of the single-particle coordinates into an appropriate feature space. The pooling collects the relevant features to produce a global feature vector that is then fed to the function ρ , which correlates the combined encodings. To have a flexible variational ansatz applicable to a variety of different systems, the functions ϕ and ρ are parametrized by dense feed-forward neural networks. This ansatz can then be shown to be a universal approximator for the permutation-invariant function class.

The complete neural variational ansatz reads

$$\begin{aligned} \psi_\alpha(\mathbf{x}_1, \dots, \mathbf{x}_N) &= \prod_{i < j} \exp \left[-\frac{1}{2} u_2[d_{\sin}(\mathbf{x}_i, \mathbf{x}_j)] \right] \\ &\quad \times \exp[\rho(\text{Pool}[\phi(\mathbf{x}^{(K)})])], \end{aligned} \quad (5)$$

where $u_2(r)$ is chosen such that Eq. (3) holds, $\mathbf{x}^{(K)}$ is a matrix storing all periodized single-particle coordinates defined in Eq. (2), and we denote the set of all variational parameters used to parametrize the functions u_2 , ρ , and ϕ by α .

As an alternative to the single-particle coordinates $\mathbf{x}_i^{(K)}$ it is also possible to input the periodized two-body distances between the particles $d_{\sin}(\mathbf{x}_i, \mathbf{x}_j) = \|\frac{L}{2} \sin(\frac{\pi}{L}\mathbf{r}_{ij})\|$ and then pool over all indices i, j . This facilitates the learning procedure substantially, though at the price of higher computational complexity of the DSs.

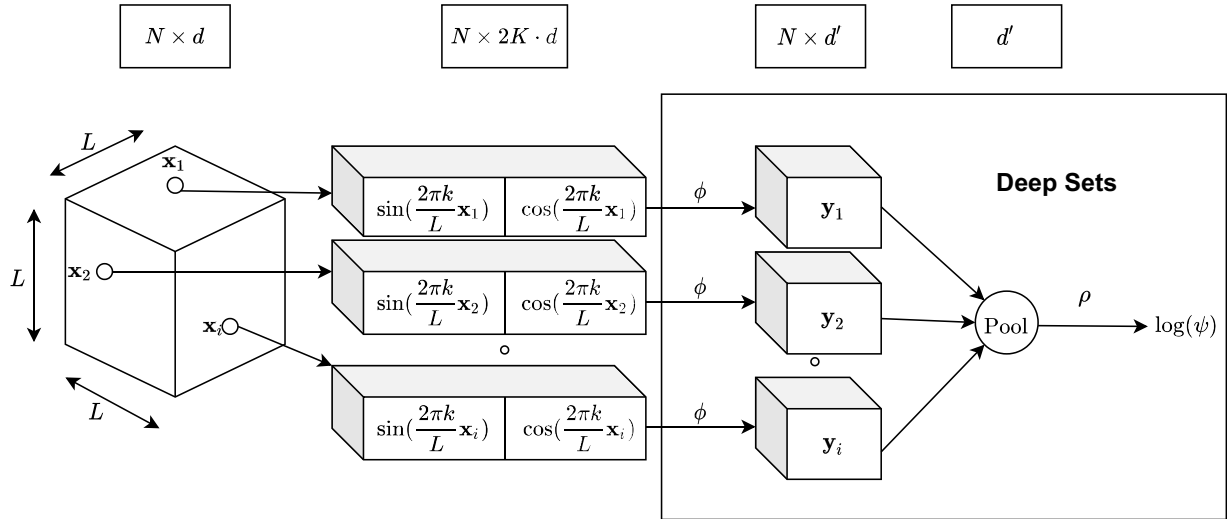


FIG. 1. Illustration of the Deep Sets neural-network architecture used in this paper. Starting from the left, we have a simulation box in d spatial dimensions of extent L . The single-particle coordinates $\mathbf{x}_i \in \mathbb{R}^d$ are transformed to the truncated Fourier basis in Eq. (2), i.e., each single-particle coordinate is mapped to a $(2K \cdot d)$ -dimensional vector. These periodic encodings are then fed to the DSs architecture, defined by the neural networks ϕ and ρ and a pool function. The output of the DSs architecture is the logarithm of the variational wave function. The top row depicts the number of feature vectors and their dimensionality at each stage of the ansatz.

Note that in either case (single- or two-body coordinates) the functional form of this ansatz, e.g., transforming the input coordinates individually followed by a pooling of all the resulting embeddings, allows it to process an arbitrary number of inputs. This feature allows us to train a small system with only a few particles and then use the obtained optimized wave function as an initialization for the variational ansatz for a bigger system, containing more particles. This is only possible if the architecture for the bigger system is exactly the same as for the smaller system, thereby preventing the introduction of additional variational parameters. Consequently, if a big number of parameters is needed to describe the big system, they must already be introduced for the small system size. This has, however, not shown to be a limiting factor in our applications.

The reutilization of an already optimized structure in a context that is similar to, but goes beyond, the one in which it was originally trained, is called *transfer learning* [50] and has been applied to different neural-network architectures and tasks in classical machine learning fields such as image recognition [51], as well as in quantum physics such as identifying phase transitions in spin systems or scaling up *ab initio* computations on spin systems to larger system sizes [52–54]. The most prominent architecture for which transfer learning is easily applicable is the convolutional neural network (CNN). It consists of convolutional layers that are parametrized by filters of fixed size (fixed number of parameters) that move across the whole input to produce a coarse-grained version of data. The size of the input does not matter since the filter explores the whole input, and therefore after being trained on a small system the filters can be applied to a much larger one subsequently.

C. Optimization

The Rayleigh-Ritz principle establishes a lower bound on the expectation value of the Hamiltonian $\langle \psi | H | \psi \rangle / \langle \psi | \psi \rangle \equiv$

$E[\psi] \geq E_0$, allowing us to formulate a variational principle on the ground-state wave function

$$\psi_0(x) = \arg \min_{\psi} E[\psi], \quad (6)$$

where $H|\psi_0\rangle = E_0|\psi_0\rangle$. The exact evaluation of the energy expectation value is not computationally feasible using deterministic integration methods, and we resort to sampling techniques commonly adopted in quantum Monte Carlo. The energy is estimated by accumulating samples of the local energy $E_{\text{loc}}(x) = \langle x | H | \psi_{\alpha} \rangle / \langle x | \psi_{\alpha} \rangle$, where the coordinates x are drawn from the probability distribution $|\psi_{\alpha}(x)|^2$ using the Metropolis-Hastings algorithm.

The variational principle allows us to obtain progressively better approximations to the ground-state wave function by minimizing $E[\psi_{\alpha}]$. The gradient components of the energy with respect to α_i (the i th component of the variational parameter α) are given by

$$G_i = 2 \left(\frac{\langle \partial_i \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} - E[\psi] \frac{\langle \partial_i \psi | \psi \rangle}{\langle \psi | \psi \rangle} \right) \quad (7)$$

and can be efficiently estimated through Monte Carlo sampling. The variational parameters are updated as $\delta \alpha = -\eta S^{-1} G$, where η is the learning rate and

$$S_{ij} = \frac{\langle \partial_i \psi | \partial_j \psi \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \partial_i \psi | \psi \rangle \langle \partial_j \psi | \psi \rangle}{\langle \psi | \psi \rangle^2} \quad (8)$$

is the Fisher-information matrix. This approach, known as the stochastic-reconfiguration (SR) algorithm [55,56], is equivalent to performing imaginary-time evolution in the variational manifold, and it is related to the natural gradient descent method [57] in unsupervised learning.

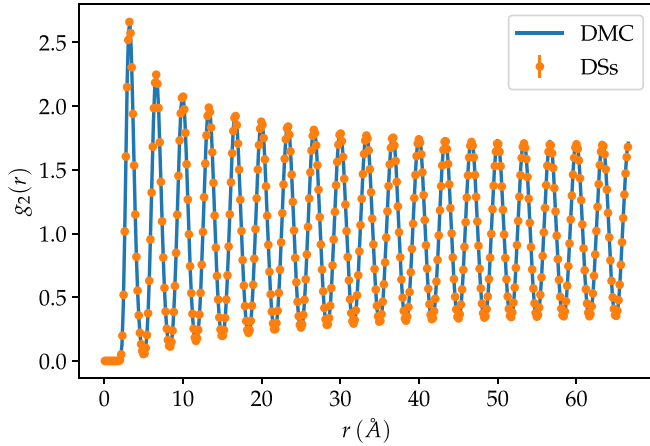


FIG. 2. DMC (solid lines) and DSs (solid circles) two-body density distributions of $N = 40$ particles in one dimension (1D) interacting through the Aziz potential.

III. RESULTS

To test the expressiveness of the proposed variational ansatz in Eq. (5) and, in particular, its compatibility with the periodic transformation in Eq. (2) and the cusp condition in Eq. (3), we examine its performance first on a system of ^4He in $d = 1$ dimension interacting via a Lennard-Jones-like potential. Subsequently, we show the versatility of the ansatz by applying it also to a cusplless system of Gaussian cores in $d = 1$ and $d = 2$ spatial dimensions.

A. Helium

The interparticle interactions of ^4He particles in $d = 1$ spatial dimension are described by an effective two-body potential which qualitatively resembles a Lennard-Jones interaction. Here, we adopt the Aziz-79 potential [58–60].

$$V(\mathbf{x}) = \epsilon \sum_{i < j} V_{\text{Aziz}}[d(\mathbf{x}_i, \mathbf{x}_j)]. \quad (9)$$

For large distances the potential decays rapidly, while for short distances the potential resembles a Lennard-Jones potential exhibiting a $\propto r^{-12}$ divergence. To enforce the cusp condition Eq. (3), this divergence needs to be compensated. It is easy to verify that the McMillan factor $u_2(r) = -(\frac{b}{r^3})$ fulfills Eq. (5), where we treat b as a variational parameter.

For this system we focus on one density only ($D = 0.3 \text{ \AA}^{-1}$) for different system sizes ($N = 10, 20, 40$). We compare the energies obtained from the ordinary DSs model with periodized single-particle coordinates, the DSs with two-body coordinates $d_{\text{sin}}(\mathbf{x}_i, \mathbf{x}_j)$, a traditional Jastrow ansatz, and diffusion Monte Carlo (DMC) [61] computations. Additionally, we plot the two-body density distributions yielded by the optimized DSs and DMC wave functions (Fig. 2). The results are reported in Table I. The energies from the single-particle coordinate DSs are clearly inferior to the ones obtained by any of the other methods for $N > 10$. We suspect that the main limitation of this ansatz is the optimization procedure

TABLE I. Energy per particle E/N for ($d = 1$)-dimensional ^4He . Displayed are three different system sizes ($N = 10, 20, 40$) at $D = 0.3 \text{ \AA}^{-1}$ obtained variationally by a DSs architecture with single- and two-body coordinates, by a Jastrow ansatz, and by means of DMC.

N	Deep Sets	Deep Sets two-body	Jastrow	DMC
10	7.272(7)	7.269(5)	7.273(9)	7.269(3)
20	7.391(9)	7.387(3)	7.424(9)	7.375(2)
40	7.456(3)	7.403(4)	7.406(5)	7.404(3)

rather than its representative power. To support this claim, we display two training curves for $N = 40$ obtained with the single-particle DSs in Fig. 3. In one case the training is started with a random initialization, while in the other case we use the optimized variational state of $N = 10$ particles as initialization to the ansatz for the $N = 40$ system. Not only does the pretrained state reach lower energies with fewer optimization steps, but also the obtained energy estimate for the ground state is considerably lower for the pretrained state.

In contrast to the single-particle case, the two-body DSs do not seem to have any representative or optimization problem. Note that since the variational ansatz contains two-body coordinates in the cusp part, the computational complexity is not increased when using two-body coordinates. The obtained energies are close to the ones obtained from DMC and consistently lower than the Jastrow energies (or within statistical error). Also the two-body density distribution fits the DMC results almost perfectly, in particular, also at large interparticle distance (see Fig. 2). Note the strong oscillations in the two-body distribution, indicating that two particles cannot come closer to each other than the Aziz core. Since in one spatial dimension the particles cannot go around each other, a crystallinelike structure emerges.

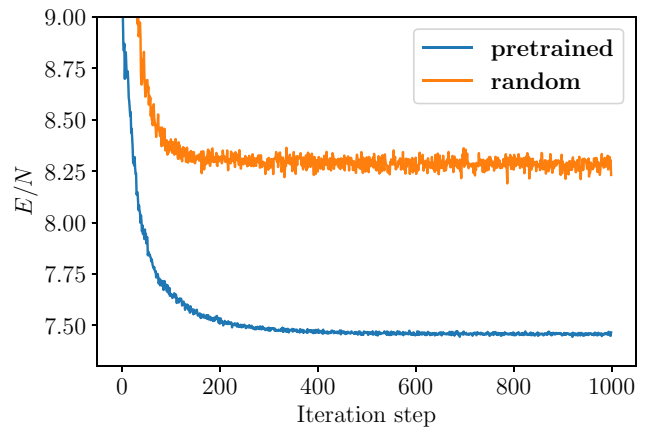


FIG. 3. Convergence pattern of the DSs architecture for $N = 40$ helium particles in 1D. Blue shows the energy for a variational ansatz initialized with the optimized weights for the same system at $N = 10$ particles. Orange shows the energy for randomly initialized weights.

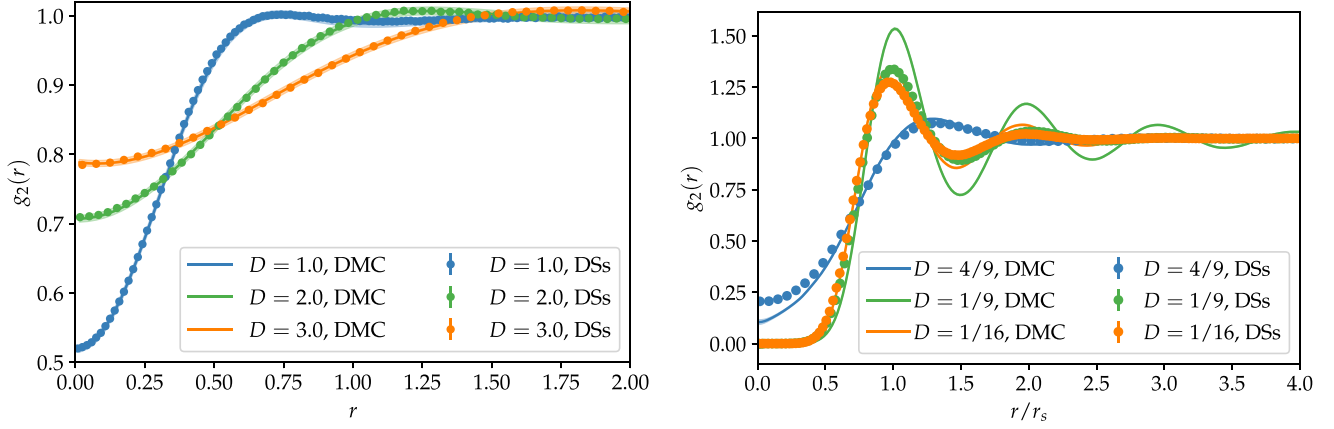


FIG. 4. Left: DMC (solid lines) and DSs (solid circles) two-body density distributions of $N = 50$ particles in 1D at different densities interacting through the Gaussian potential. Right: the two-body density distribution of $N = 64$ particles in 2D at different densities interacting through the Gaussian potential.

B. Gaussian cores

The Hamiltonian of the Gaussian cores is given by Eq. (1) with a Gaussian interaction potential [62]

$$V(\mathbf{x}) = \epsilon \sum_{i < j} \exp \left[-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{2\sigma^2} \right]. \quad (10)$$

Since the interaction does not exhibit divergent behavior in the zero-distance limit, we take $u_2 = 0$ in Eq. (5), and the variational ansatz solely consists of the DSs architecture.

We again benchmark the energies obtained with this NQS approach with variational energies given by DMC computations and a traditional Jastrow ansatz. Additionally, we plot the two-body density distributions yielded by the optimized DSs and DMC wave functions (Fig. 4).

For the one-dimensional system we studied two different system sizes ($N = 20, 50$) at three different densities ($D = \frac{N}{L} = 1, 2, 3$) with $\epsilon = 2$, $\sigma = 2^{-1/2}$. As can be seen in Table II the two variational energies obtained from DSs and Jastrow as well as the DMC energies lie within statistical error of each other. We also obtain almost perfect agreement for the radial distribution functions from the DSs and DMC, indicating that our variational state appropriately describes the ground state of the system. An advantage of the DSs

TABLE II. Energy per particle E/N for $(d = 1)$ -dimensional Gaussian cores. Displayed are two different system sizes ($N = 20, 50$) and three different densities ($D = 1, 2, 3$) obtained variationally by a DSs architecture and a Jastrow ansatz and by means of DMC.

N	D	Deep Sets	Jastrow	DMC
20	1.0	1.2373(4)	1.2373(1)	1.2371(1)
20	2.0	2.9223(5)	2.9224(1)	2.9222(1)
20	3.0	4.6445(7)	4.6447(1)	4.6446(2)
50	1.0	1.2402(5)	1.2408(1)	1.2390(4)
50	2.0	2.9286(3)	2.9310(1)	2.9280(2)
50	3.0	4.6557(1)	4.6597(1)	4.6553(2)

architecture over the Jastrow ansatz is that the computational complexity only scales as $O(N)$ for one evaluation of the ansatz, rather than $O(N^2)$ as is the case for the Jastrow ansatz.

In the case of two spatial dimensions (Table III), we compare the same three models (DSs, Jastrow, and DMC) for two different system sizes ($N = 32, 64$) at three different densities ($D = \frac{N}{L^2} = 4/9, 1/9, 1/16$) with $\hbar^2/(m\epsilon\sigma^2) = 1/30$. Instead of single-particle coordinates, we use the periodized two-body distance between the particles as input to the DSs. We find that the DSs and Jastrow are again very close. However, the DMC computations yield smaller energies. The energy difference is notable in the radial distribution function. In particular, for the density $D = 1/9$, the DSs cannot reproduce the oscillating behavior visible in the DMC result. This might indicate a limitation of the expressiveness of our variational ansatz.

C. Influence of network architecture

For all the computations we used the multilayer perceptron architecture (MLP) with one and two hidden layers, respectively, as parametrization for the functions ϕ and ρ of Eq. (1). We chose the $L\Sigma E$ pooling and sum pooling for single- and two-body input, respectively. The nonlinearity in the neurons of the MLPs was chosen to be the Gaussian error linear unit

TABLE III. Energy per particle E/N for $(d = 2)$ -dimensional Gaussian cores. Displayed are two different system sizes ($N = 32, 64$) and three different densities ($D = 4/9, 1/9, 1/16$) obtained variationally by a DSs architecture and a Jastrow ansatz and by means of DMC.

N	D	Deep Sets two-body	Jastrow	DMC
32	4/9	1.01707(4)	1.02020(3)	1.01355(4)
32	1/9	0.08864(5)	0.08895(2)	0.08546(4)
32	1/16	0.02545(5)	0.02549(1)	0.02434(1)
64	4/9	1.01621(3)	1.02661(4)	1.01429(6)
64	1/9	0.08905(3)	0.08994(2)	0.08519(4)
64	1/16	0.02573(5)	0.02580(2)	0.02443(5)

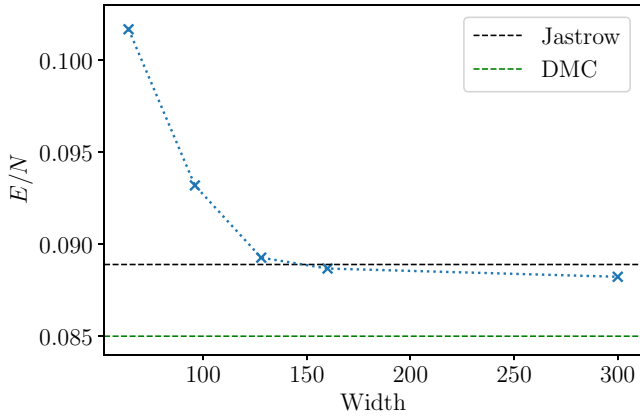


FIG. 5. Final energy of training the DSs with single-particle input for the two-dimensional Gaussian core model with $N = 32$ particles at a density $D = 1/9$ as a function of the width of the neural networks used. The horizontal lines correspond to the Jastrow and DMC results, respectively.

(GELU) activation [63] as a differentiable surrogate for the rectified linear unit (ReLU) activation.

Increasing the number of hidden layers in either of the two MLPs has not increased the accuracy of the results. However, increasing the width of the layers has a significant impact on the quality of the results, in particular, for the ($d = 2$)-dimensional Gaussian cores. In Fig. 5 we show the training curves and final energies for $N = 32$ particles with density $D = 1/9$, interacting via a Gaussian potential in two spatial dimensions. As an ansatz the single-particle DSs are used with $K = 10$ [see Eq. (2)]. Whereas for <128 neurons per layer, the ordinary Jastrow ansatz has an edge over the DSs, our NQS approach starts to outperform the Jastrow ansatz for wider layers. We find the same behavior if we use a two-body neural-network-based Jastrow ansatz, in the form of $\sum_{i<j} \phi[\sin(\frac{\pi}{L} r_{ij})]$, where ϕ is a neural network whose output dimension is 1. This comparison indicates that the Deep Sets approach starts to capture correlations beyond the reach of a two-body Jastrow ansatz. The DMC energies do not seem to be reachable by simply increasing the size of the layers, indicating the necessity of refinements to the current architecture.

IV. DISCUSSION

We have introduced an extension of NQSs to systems in continuous space and with intrinsic spatial periodicity. The approach we discuss in this paper is specialized to bosonic degrees of freedom, and it is based on permutationally equivariant states based on Deep Sets, featuring periodicity under spatial translations. We have successfully applied our periodic DSs neural quantum state to one-dimensional Gaussian cores, confined to a periodic simulation cell, and showed that the obtained system properties typically improve upon traditionally used wave functions based on two-body Jastrow correlators, and are in excellent agreement with exact diffusion Monte Carlo results. An advantage of the NQS-based approach is

that it can be applied to different systems without significant modifications. Specifically, using the exact same ansatz (up to modifying the input) to model interacting ^4He in one spatial dimension, we observe excellent agreement with existing state-of-the-art methods. Modeling the helium system is considerably more difficult than modeling the Gaussian cores because of the rapidly diverging potential at short distances. For the case of two spatial dimensions, we obtain energies that are marginally better than the ordinary Jastrow ansatz. We show that the energy estimate can be improved by increasing the width of the neural networks; however, we are not able to reach the DMC energies even with the largest network used.

An important result for the optimization of our variational state is that the optimized ansatz for small systems can very well be used as an initialization for larger systems (known as transfer learning). However, it remains to be seen whether for even larger systems than were presented in this paper, a higher number of parameters is necessary to capture all the relevant correlations between the particles. If so, these parameters must already be present in the training of the small system when applying the transfer learning method above, potentially making the training of small systems less efficient.

Concluding, we showed that our NQS approach respects the periodicity of the simulation cell as well as the permutation invariance of the wave function with respect to particle exchange, while still yielding high-precision results. It can be envisioned that the current architecture is altered to better account for the locality of the interaction potential as well as further relevant symmetries of the problem. We suspect that a more flexible ansatz than DSs will help systematically improve the accuracy in the two-dimensional case, and we leave this aspect for future studies. Also the architecture might be used to model fermionic systems, which ordinarily are modeled as the product of a symmetric and an antisymmetric part. The DSs can provide the symmetric component of such an ansatz.

ACKNOWLEDGMENTS

The authors acknowledge several insightful discussions with David Ceperley, Markus Holzmann, and Saverio Moroni. The numerical tools used in this paper are based on the open-source software NETKET [64–66] version 3.0. The present research is supported by the Swiss National Science Foundation under Grant No. 200021_200336, by the U.S. Department of Energy (DOE), Office of Science, Office of Nuclear Physics, under Contract No. DE-AC02-06CH11357, the NUCLEI SciDAC program (A.L.), the DOE Early Career Research Program, Argonne LDRD awards, and U.S. National Science Foundation via Reward No. DMS-2012286 (J.L.). Part of the calculations presented in this paper were performed through a CINECA-INFN agreement that provides access to resources on MARCONI at CINECA. This collaboration originates from a SquaREs program hosted by American Institute of Mathematics; we thank AIM for their hospitality.

- [1] *Supervised and Unsupervised Learning for Data Science*, edited by M. W. Berry, A. Mohamed, and B. W. Yap (Springer International, Cham, Switzerland, 2020).
- [2] B. Mahesh, Machine learning algorithms-a review, *Int. J. Sci. Res.* **9**, 381 (2020).
- [3] T. S. Cohen and M. Welling, Group equivariant convolutional networks, in *33rd International Conference on Machine Learning (ICML 2016)*, Journal of Machine Learning Research Workshop and Conference Proceedings (Curran Associates, Red Hook, NY, 2016), Vol. 48, pp. 2990–2999.
- [4] M. Mattheakis, P. Protopapas, D. Sondak, M. Di Giovanni, and E. Kaxiras, Physical symmetries embedded in neural networks, [arXiv:1904.08991](https://arxiv.org/abs/1904.08991) [physics.comp-ph].
- [5] C. Roth and A. H. MacDonald, Group convolutional neural networks improve quantum state accuracy, [arXiv:2104.05085](https://arxiv.org/abs/2104.05085) [quant-ph].
- [6] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [7] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [8] D.-L. Deng, X. Li, and S. Das Sarma, Quantum Entanglement in Neural Network States, *Phys. Rev. X* **7**, 021021 (2017).
- [9] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Equivalence of restricted Boltzmann machines and tensor network states, *Phys. Rev. B* **97**, 085104 (2018).
- [10] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum Entanglement in Deep Learning Architectures, *Phys. Rev. Lett.* **122**, 065301 (2019).
- [11] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and Many-Body Excitations with Neural-Network Quantum States, *Phys. Rev. Lett.* **121**, 167204 (2018).
- [12] F. Ferrari, F. Becca, and J. Carrasquilla, Neural Gutzwiller-projected variational wave functions, *Phys. Rev. B* **100**, 125131 (2019).
- [13] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, *Phys. Rev. Research* **2**, 023358 (2020).
- [14] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, *Nat. Commun.* **11**, 2368 (2020).
- [15] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, *Phys. Rev. Research* **2**, 033075 (2020).
- [16] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, *SciPost Phys.* **10**, 147 (2021).
- [17] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio and Level Spectroscopy, *Phys. Rev. X* **11**, 031034 (2021).
- [18] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, Broken-Symmetry Ground States of the Heisenberg Model on the Pyrochlore Lattice, *Phys. Rev. X* **11**, 041021 (2021).
- [19] J. Han, L. Zhang, and W. E, Solving many-electron Schrödinger equation using deep neural networks, *J. Comput. Phys.* **399**, 108929 (2019).
- [20] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, *Phys. Rev. Research* **2**, 033429 (2020).
- [21] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, *Nat. Chem.* **12**, 891 (2020).
- [22] X. Li, C. Fan, W. Ren, and J. Chen, Fermionic neural network with effective core potential, *Phys. Rev. Research* **4**, 013021 (2022).
- [23] J. Kessler, F. Calcavecchia, and T. D. Kühne, Artificial neural networks as trial wave functions for quantum Monte Carlo, *Adv. Theory Simul.* **4**, 2000269 (2021).
- [24] M. Wilson, N. Gao, F. Wudarski, E. Rieffel, and N. M. Tubman, Simulations of state-of-the-art fermionic neural network wave functions with diffusion Monte Carlo, [arXiv:2103.12570](https://arxiv.org/abs/2103.12570) [physics.chem-ph].
- [25] C. Adams, G. Carleo, A. Lovato, and N. Rocco, Variational Monte Carlo Calculations of $A \leq 4$ Nuclei with an Artificial Neural-Network Correlator Ansatz, *Phys. Rev. Lett.* **127**, 022502 (2021).
- [26] A. Gnech, C. Adams, N. Brawand, G. Carleo, A. Lovato, and N. Rocco, Nuclei with up to $A = 6$ nucleons with artificial neural network wave functions, *Few Body Syst.* **63**, 7 (2022).
- [27] N. Yoshioka, W. Mizukami, and F. Nori, Solving quasiparticle band spectra of real solids using neural-network quantum states, *Commun. Phys.* **4**, 106 (2021).
- [28] T. Dornheim, S. Groth, T. Sjostrom, F. D. Malone, W. M. C. Foulkes, and M. Bonitz, *Ab Initio* Quantum Monte Carlo Simulation of the Warm Dense Electron Gas in the Thermodynamic Limit, *Phys. Rev. Lett.* **117**, 156403 (2016).
- [29] D. Schiff and L. Verlet, Ground State of Liquid Helium-4 and Helium-3, *Phys. Rev.* **160**, 208 (1967).
- [30] W. L. McMillan, Ground State of Liquid He⁴, *Phys. Rev.* **138**, A442 (1965).
- [31] D. M. Ceperley, Path-integrals in the theory of condensed helium, *Rev. Mod. Phys.* **67**, 279 (1995).
- [32] G. V. Chester, Speculations on Bose-Einstein Condensation and Quantum Crystals, *Phys. Rev. A* **2**, 256 (1970).
- [33] L. Tanzi, E. Lucioni, F. Famà, J. Catani, A. Fioretti, C. Gabbanini, R. N. Bisset, L. Santos, and G. Modugno, Observation of a Dipolar Quantum Gas with Metastable Supersolid Properties, *Phys. Rev. Lett.* **122**, 130405 (2019).
- [34] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, Backflow correlations for the electron gas and metallic hydrogen, *Phys. Rev. E* **68**, 046707 (2003).
- [35] S. Azadi, W. M. C. Foulkes, and T. D. Kühne, Quantum Monte Carlo study of high pressure solid molecular hydrogen, *New J. Phys.* **15**, 113005 (2013).
- [36] J. Carlson, J. Morales, Jr., V. R. Pandharipande, and D. G. Ravenhall, Quantum Monte Carlo calculations of neutron matter, *Phys. Rev. C* **68**, 025802 (2003).
- [37] M. Piarulli, I. Bombaci, D. Logoteta, A. Lovato, and R. B. Wiringa, Benchmark calculations of pure neutron matter with realistic nucleon-nucleon interactions, *Phys. Rev. C* **101**, 045801 (2020).
- [38] D. Lonardonì, I. Tews, S. Gandolfi, and J. Carlson, Nuclear and neutron-star matter from local chiral interactions, *Phys. Rev. Research* **2**, 022033(R) (2020).
- [39] N. Ismail and A. Gezerlis, Machine-learning approach to finite-size effects in systems with strongly interacting fermions, *Phys. Rev. C* **104**, 055802 (2021).

- [40] R. Jastrow, Many-Body Problem with Strong Forces, *Phys. Rev.* **98**, 1479 (1955).
- [41] R. P. Feynman and M. Cohen, Energy Spectrum of the Excitations in Liquid Helium, *Phys. Rev.* **102**, 1189 (1956).
- [42] K. E. Schmidt, M. A. Lee, M. H. Kalos, and G. V. Chester, Structure of the Ground State of a Fermion Fluid, *Phys. Rev. Lett.* **47**, 807 (1981).
- [43] M. Ruggeri, S. Moroni, and M. Holzmann, Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space, *Phys. Rev. Lett.* **120**, 205302 (2018).
- [44] P. A. Whitlock and S. A. Vitiello, Quantum Monte Carlo simulations of solid ^4He , in *Large-Scale Scientific Computing*, Lecture Notes in Computer Science (Springer International, Cham, Switzerland, 2006), Vol. 3743, pp. 40–52.
- [45] J. Han, J. Lu, and M. Zhou, Solving high-dimensional eigenvalue problems using deep neural networks: A diffusion Monte Carlo like approach, *J. Comput. Phys.* **423**, 109792 (2020).
- [46] M. P. Allen, Introduction to molecular dynamics simulation, in *Computational Soft Matter: From Synthetic Polymers to Proteins*, Lecture Notes, edited by N. Attig, K. Binder, H. Grubmüller, and K. Kremer, NIC Series (John von Neumann Institute for Computing, Jülich, Germany, 2004), Vol. 23, pp. 1–28.
- [47] T. Kato, On the eigenfunctions of many-particle systems in quantum mechanics, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- [48] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, Deep sets, in *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Curran Associates, Red Hook, NY, 2017), pp. 3392–3402.
- [49] E. Wagstaff, F. B. Fuchs, M. Engelcke, I. Posner, and M. Osborne, On the limitations of representing functions on sets, in *International Conference on Machine Learning (PLMR)*, 2019, pp. 6487–6494.
- [50] R. Caruana, Multitask learning, *Mach. Learn.* **28**, 41 (1997).
- [51] M. Shaha and M. Pawar, Transfer learning for image classification, in *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018* (IEEE, Piscataway, NJ, 2018), pp. 656–660.
- [52] I. Corte, S. Acevedo, M. Arlego, and C. A. Lamas, Exploring neural network training strategies to determine phase transitions in frustrated magnetic models, *Comput. Mater. Sci.* **198**, 110702 (2021).
- [53] R. Zen, L. My, R. Tan, F. Hébert, M. Gattobigio, C. Miniatura, D. Poletti, and S. Bressan, Transfer learning for scalability of neural-network quantum states, *Phys. Rev. E* **101**, 053301 (2020).
- [54] L. Yang, W. Hu, and L. Li, Scalable variational Monte Carlo with graph neural ansatz, in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Advances in Neural Information Processing Systems (Curran Associates, Red Hook, NY, 2020), Vol. 33, pp. XX–XX.
- [55] S. Sorella, Green Function Monte Carlo with Stochastic Reconfiguration, *Phys. Rev. Lett.* **80**, 4558 (1998).
- [56] S. Sorella, Wave function optimization in the variational Monte Carlo method, *Phys. Rev. B* **71**, 241103(R) (2005).
- [57] S.-I. Amari, Natural gradient works efficiently in learning, *Neural Comput.* **10**, 251 (1998).
- [58] G. Bertaina, M. Motta, M. Rossi, E. Vitali, and D. E. Galli, One-Dimensional Liquid ^4He : Dynamical Properties beyond Luttinger-Liquid Theory, *Phys. Rev. Lett.* **116**, 135302 (2016).
- [59] See “Path-integral ground state method” in the Supplemental Material of Ref. [58], pp. 1–5.
- [60] R. A. Aziz and H. H. Chen, An accurate intermolecular potential for argon, *J. Chem. Phys.* **67**, 5719 (1977).
- [61] I. Kosztin, B. Faber, and K. Schulten, Introduction to the diffusion Monte Carlo method, *Am. J. Phys.* **64**, 633 (1996).
- [62] P. Kroiss, M. Boninsegni, and L. Pollet, Ground-state phase diagram of Gaussian-core bosons in two dimensions, *Phys. Rev. B* **93**, 174520 (2016).
- [63] D. Hendrycks and K. Gimpel, Gaussian error linear units (GELUs), [arXiv:1606.08415](https://arxiv.org/abs/1606.08415) [cs.LG].
- [64] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O’Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, NetKet: A machine learning toolkit for many-body quantum systems, *Software X* **10**, 100311 (2019).
- [65] D. Häfner and F. Vicentini, mpi4jax: Zero-copy MPI communication of JAX arrays, *J. Open Source Software* **6**, 3419 (2021).
- [66] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderon, N. Astrakhantsev, and G. Carleo, NetKet 3: Machine learning toolbox for many-body quantum systems, [arXiv:2112.10526](https://arxiv.org/abs/2112.10526) [quant-ph].