

# From Architectures to Applications: A Review of Neural Quantum States

Hannah Lange<sup>1,2,3</sup>, Anka Van de Walle<sup>2,3</sup>, Atiye Abedinnia<sup>4</sup> and Annabelle Bohrdt<sup>2,4</sup>

<sup>1</sup>Max-Planck-Institute for Quantum Optics, Hans-Kopfermann-Str.1, Garching D-85748, Germany

<sup>2</sup>Munich Center for Quantum Science and Technology, Schellingstr. 4, Munich D-80799, Germany

<sup>3</sup>Ludwig-Maximilians-University Munich, Theresienstr. 37, Munich D-80333, Germany

<sup>4</sup>University of Regensburg, Universitätsstr. 31, Regensburg D-93053, Germany

## Abstract.

Due to the exponential growth of the Hilbert space dimension with system size, the simulation of quantum many-body systems has remained a persistent challenge until today. Here, we review a relatively new class of variational states for the simulation of such systems, namely neural quantum states (NQS), which overcome the exponential scaling by compressing the state in terms of the network parameters rather than storing all exponentially many coefficients needed for an exact parameterization of the state. We introduce the commonly used NQS architectures and their various applications for the simulation of ground and excited states, finite temperature and open system states as well as NQS approaches to simulate the dynamics of quantum states. Furthermore, we discuss NQS in the context of quantum state tomography.

Quantum many-body systems are of great interest for many research areas, including physics, biology and chemistry. However, their simulation has remained challenging until today, due to the exponential growth of the Hilbert space with the system size, making it exceedingly difficult to parameterize the wave functions of large systems using exact methods. One common approach to overcome this problem are variational methods, where a certain functional form of the quantum state is assumed, with free parameters to be optimized to obtain the best possible representation of the state under investigation. A well established method based on variational wave functions are tensor networks (TN) [1, 2, 3, 4, 5, 6, 7, 8], among them variants that can be contracted efficiently, like matrix product states (MPS) [9, 10, 11], and which hence allow an efficient evaluation of observables. MPS are restricted to states that obey the area law of entanglement [12, 13], and are hence particularly well suited for one-dimensional gapped quantum systems, although extensions to higher dimensional systems are possible [1, 7, 14]. Another class of methods for the numerical simulation of quantum systems, quantum Monte Carlo (MC) algorithms [15, 16, 17], suffers from the sign problem [18, 19] and slow convergence for large system sizes close to critical points or other challenging statistical physics problems [20].

The ability of sufficiently large neural networks to represent any continuous function [21, 22, 23, 24] motivated their use for the simulation of quantum states, and was pioneered by Carleo and Troyer [25] in 2017. To date, these so-called neural quantum states (NQS) have been shown to overcome many problems that are inherent to some conventional methods such as MPS: *(i)* Some works have demonstrated that NQS are capable of representing volume-law entangled states [26, 27, 28, 29, 30] and can hence in principle be used for a broad range of quantum systems [26, 28, 31, 30, 32, 27, 33]. In particular, it has been shown that in some cases mappings between NQS and efficiently contractable TNs can be established, e.g. in Ref. [34] the authors find that TNs are a subset of the considered NQS [26]. *(ii)* They can be designed to be particularly well suited for two-dimensional problems. Most prominently, some architectures like convolutional neural networks were specifically designed for two-dimensional data; *(iii)* In many cases, they allow for an efficient evaluation of operators, in some cases even global operators like the momentum [35].

NQS are typically used for two distinct tasks: First, they have appeared in the field of quantum state tomography, where they are used for quantum state reconstruction of states prepared in experiments, allowing the estimation of observables that can not be accessed in experiments [36, 37]. Second, they can be used as simulation tools for quantum systems, with a Hamiltonian driven optimization similar to TNs. In this setting no training data is needed [25]. Furthermore, NQS simulations can not

only be applied to represent ground states, but also excited states, finite temperature states, the time evolution of quantum states or open systems.

The goal of this article is to give an overview of the current state of NQS, i.e. existing NQS architectures, their training and their performance in quantum state simulation and tomography in comparison to conventional methods. Previously published overviews on neural quantum states can be found in Refs. [20, 38, 39, 40, 40, 41] in the more general context of neural network applications in quantum physics, or more specifically on NQS and their optimization in Refs. [42, 43, 44, 45].

The outline of this review is as follows: We start with an overview of existing NQS architectures and their application to physical systems as well as commonly used design choices. The second part considers applications of NQS, namely the simulation of ground and excited states, finite temperature states, time evolution and open quantum systems. Furthermore, we review quantum state tomography and hybrid simulation schemes with NQS.

## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>A Short Introduction to Neural Quantum States (NQS)</b> | <b>3</b>  |
| <b>2</b> | <b>NQS Architectures</b>                                   | <b>4</b>  |
| 2.1      | Feed Forward Neural Networks (FFNNs)                       | 4         |
| 2.2      | Restricted Boltzmann Machines (RBMs)                       | 6         |
| 2.3      | Convolutional Neural Networks (CNNs) and Group CNNs        | 7         |
| 2.4      | Graph Neural Networks                                      | 9         |
| 2.5      | Latent Space Representations: Autoencoders                 | 10        |
| 2.6      | Autoregressive Networks                                    | 10        |
| 2.6.1    | Recurrent Neural Networks (RNNs)                           | 11        |
| 2.6.2    | Transformers   | 12        |
| 2.7      | NQS for Fermionic Systems                                  | 13        |
| 2.7.1    | Fermionic Architectures                                    | 13        |
| 2.7.2    | Bosonic Architectures with Jordan Wigner Strings           | 14        |
| 2.8      | Other Design Choices                                       | 15        |
| 2.9      | Open-Source Toolboxes                                      | 16        |
| <b>3</b> | <b>Applications of NQS</b>                                 | <b>16</b> |
| 3.1      | Ground States  | 16        |
| 3.2      | Excited States   | 19        |
| 3.3      | Dynamics   | 20        |
| 3.3.1    | Spectral Functions   | 22        |
| 3.4      | Finite Temperature States                                  | 23        |
| 3.5      | Open Systems   | 24        |
| 3.6      | Learning from Data   | 26        |
| 3.6.1    | Quantum State Tomography                                   | 26        |
| 3.6.2    | Hybrid Training  | 28        |
| <b>4</b> | <b>Summary and Outlook</b>                                 | <b>29</b> |

## 1. A Short Introduction to Neural Quantum States (NQS)

In most cases, neural quantum states (NQS) are used to represent a quantum state

$$|\psi\rangle = \sum_{\sigma} \psi(\sigma) |\sigma\rangle, \quad (1)$$

for a certain basis choice  $|\sigma\rangle$  that is given by the  $d_l$  different local configurations, e.g. the spin configurations or Fock space configurations  $\sigma$ . For a system with  $N$  sites,  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$  consists of e.g.  $\sigma_i = 0, 1$  for spin systems ( $d_l = 2$ ) or  $\sigma_i = 0, 1, 2, \dots, n_{\max}$  for bosonic systems ( $d_l = n_{\max} + 1$ ).

The underlying idea of neural quantum states is to use neural networks in order to represent the wave function coefficients  $\psi(\sigma)$  of the state under investigation. Hereby, the neural network is used as a variational wave function, mapping configurations  $\sigma$  to the respective wave function coefficient  $\psi_{\theta}(\sigma)$ , parameterized by the neural network parameters  $\theta$ . More precisely, the input of the neural network used for the NQS representation are configurations  $\sigma$ , and the output is

$$\psi_{\theta}(\sigma) = \sqrt{p_{\theta}(\sigma)} e^{i\phi_{\theta}(\sigma)}, \quad (2)$$

which is often split into its amplitude  $p_{\theta}(\sigma) = |\psi_{\theta}(\sigma)|^2$  and its phase  $\phi_{\theta}(\sigma) = \text{Im}(\log \psi_{\theta}(\sigma))$ . To feed an input  $\sigma$  into the network,  $\sigma$  is typically one-hot encoded, i.e. the  $d_l$  different local configurations are encoded binary, resulting in a matrix  $\sigma \in \mathbb{R}^{N \times d_l}$  for every configuration  $\sigma$  of length  $N$ . Furthermore, the input is often embedded into a space of dimension  $d_h$ , i.e. using a trainable or physically inspired projection the input is projected onto the  $d_h$  dimensional space that the network is operating on.

The main difficulty of variational approaches is to come up with a good representation  $\psi_{\theta}(\sigma)$  of the true wave function coefficients. Here, the great strength of neural networks, namely their expressive power, comes into play: Neural networks with at least one hidden layer, a sufficient number of parameters and an arbitrary non-linear activation function have the ability to represent continuous functions of any – potentially very complicated – form [21, 22, 23, 24]. This makes them promising candidates for a successful representation  $\psi_{\theta}$  that is close to the exact wave function  $\psi$ . In order to obtain this representation  $\psi_{\theta}$ , the network parameters  $\theta$  are adjusted during the training of the NQS, i.e. starting from some initialization of the neural network parameters  $\theta_0$ , the network parameters are adjusted such that  $\psi_{\theta}$  approximates the true wave function  $\psi$  at the end of the training.

In contrast to most machine learning applications, the training can be done in a self-contained way without the use of external data. In general, the specific design choices for the NQS can have a significant impact on its performance, which will be the focus of Sec. 2: Besides the choice of architecture, e.g. the way how the real and imaginary parts of the wave function coefficients  $\psi_{\theta}$  are modeled. This can be done by splitting  $\psi_{\theta}$  into amplitude  $p_{\theta}$  and phase  $\phi_{\theta}$  parts and using separate networks or separate output nodes / final layers for each part. Another possibility is to use complex network parameters to model the full  $\psi_{\theta}$  with a single network. Furthermore, NQS can be designed to obey certain symmetries like  $U(1)$  symmetries. The performance of the wave function does moreover depend on the optimization and the specific task under consideration, which is discussed in Sec. 3.

Similar to Monte Carlo methods, observables of NQS are evaluated by generating samples  $\{\sigma\}$  from the NQS amplitudes, which are used for the estimation of the respective expectation values. Specifically, for an operator  $\hat{O}$ , the expectation value can be written as

$$\langle \hat{O} \rangle = \frac{\langle \psi_{\theta} | \hat{O} | \psi_{\theta} \rangle}{\langle \psi_{\theta} | \psi_{\theta} \rangle} = \sum_{\sigma, \sigma'} \frac{\langle \psi_{\theta} | \sigma \rangle \langle \sigma | \hat{O} | \sigma' \rangle \langle \sigma' | \psi_{\theta} \rangle}{\sum_{\sigma''} \langle \psi_{\theta} | \sigma'' \rangle \langle \sigma'' | \psi_{\theta} \rangle} = \sum_{\sigma} P_{\theta}(\sigma) \sum_{\sigma'} \frac{\psi_{\theta}(\sigma')}{\psi_{\theta}(\sigma)} \langle \sigma | \hat{O} | \sigma' \rangle \approx \langle O_{\theta}^{\text{loc}}(\sigma) \rangle_{\sigma} \quad (3)$$

with the probability for each configuration  $P_{\theta}(\sigma)$  and the local estimator  $O_{\theta}^{\text{loc}}(\sigma)$ , defined as

$$P_{\theta}(\sigma) = \frac{|\psi_{\theta}(\sigma)|^2}{\sum_{\sigma''} |\psi_{\theta}(\sigma'')|^2} \quad \text{and} \quad O_{\theta}^{\text{loc}}(\sigma) = \sum_{\sigma'} \frac{\psi_{\theta}(\sigma')}{\psi_{\theta}(\sigma)} \langle \sigma | \hat{O} | \sigma' \rangle \quad (4)$$

respectively, as well as the Monte Carlo average  $\langle \cdot \rangle_{\sigma}$ . For operators involving only a limited number of matrix elements,  $O_{\theta}^{\text{loc}}$  can be evaluated very efficiently [46], namely local operators or global operators that do not require the calculation of higher order correlations. The computational cost of Eq. (3)

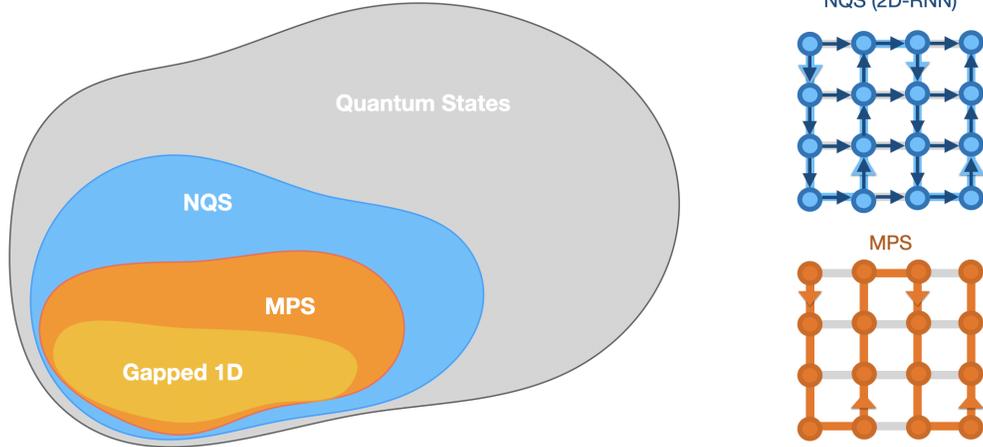


Figure 1: Representative power of Matrix product states (MPS) and neural quantum states (NQS). NQS have been proven to have strictly the same or higher expressive power than practically usable variational tensor networks, see e.g. Ref. [26]. This results from the high connectivity between sites of NQS, here exemplary shown for a two-dimensional recurrent neural network (2D-RNN) as considered in Ref. [49] on the top right, compared to MPS, for which information is passed through the system in a 1D manner (see bottom right). Figures adapted from Refs. [38, 26].

results from the generation of samples  $\sigma$  from  $|\psi_\theta|^2$ , as well as from the evaluation of the wave function amplitudes for  $\sigma$  and its connected samples  $\sigma'$ . The computational cost of the former strongly depends on  $\psi_\theta$  being normalized or not, since in the latter case samples can not directly be generated from the wave function and more elaborate approaches like Metropolis sampling are needed. Normalized NQS, using so-called autoregressive architectures, are the topic of Sec. 2.6.

NQS have been shown to be capable of representing a broad range of quantum states [26, 28, 31, 30, 32, 27, 33]. To compare their expressivity to more conventional variational approaches like TNs and PEPS, the relationship between them has been studied [47, 48, 26, 49]. Some NQS architectures have been proven to have strictly the same or higher expressive power than practically usable variational tensor networks [26], see Fig. 1. In particular, a range of works have shown that some NQS can encode volume law states without exponential cost [26, 27, 28, 29, 30], although there are volume-law entangled states like the ground state of the Sachdev-Ye-Kitaev model that can not be represented efficiently [50]. Furthermore, Ref. [51] develops a combination of TNs and autoregressive NQS, which improves the capabilities compared to both the original TN and NQS. However, in contrast to TNs which are guaranteed to converge after a sufficiently long optimization, the training of NQS involves a non-convex landscape. Hence, it can be challenging to find the actual ground state, even when the NQS ansatz itself is expressive enough to capture it. Advanced training strategies to overcome this issue are discussed in Sec. 3.1.

## 2. NQS Architectures

Neural network quantum states can be implemented using several techniques, including various neural network architectures and different representations of phase and amplitude parts of the wave function. Each architecture comes with its advantages and specialized training strategies, see also Ref. [52]. Additionally, the choice of architecture can also depend on the physical model under investigation. In this section, we discuss commonly used architectures, their application to physical systems in the literature as well as their advantages and downsides compared to other ansätze.

### 2.1. Feed Forward Neural Networks (FFNNs)

A feed forward neural network (FFNN), often represented by the structure of a multi-layer perceptron (MLP), is the fundamental building block of artificial neural networks. It is composed of distinct

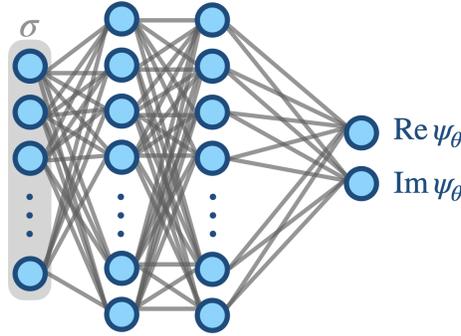


Figure 2: Feed forward neural network (FFNN): The physical state  $\sigma$  (blue circles enclosed by the grey rectangle) is taken as the input. Blue circles denote the nodes of the network, grey lines the connecting weight matrices. The output of the neural network is the wave function coefficient  $\psi(\sigma)$ , here split into its amplitude and phase.

layers of neurons, including an input layer that receives the data, one or more hidden layers where computations are performed, and an output layer that delivers the final result, see Fig. 2. Within each layer  $l$ , each neuron is assigned a bias  $b_j^{(l)}$ , and is linked to neurons in the adjacent layers through connections  $w_{i,j}^{(l)}$ . These weights and biases are crucial as they are iteratively adjusted during the network's training, primarily using backpropagation and optimization techniques like gradient descent. The activation functions applied to each neuron's output introduce non-linearity, enabling the network to model complex relationships. In a fully connected FFNN, every neuron in a layer is connected to all neurons in the next layer. The value of each neuron,  $a_j^{(l)}$ , in layer  $l$ , can be described mathematically as:

$$a_j^{(l)} = f \left( \sum_{i=1}^n w_{ij}^{(l)} \cdot a_i^{(l-1)} + b_j^{(l)} \right), \quad (5)$$

where  $w_{ij}^{(l)}$  represents the weight from the  $i$ -th neuron in layer  $l-1$  to the  $j$ -th neuron in layer  $l$ ,  $a_i^{(l-1)}$  is the activation of the  $i$ -th neuron in layer  $l-1$ , and  $b_j^{(l)}$  is the bias of the  $j$ -th neuron in layer  $l$ . The function  $f$  denotes the activation function. This straightforward, yet powerful structure makes FFNNs a vital component in the field of neural networks and deep learning, and has led to a range of applications in the context of NQS:

Ref. [53] uses FFNNs to describe ground states of different one-dimensional systems, as well as spinless fermions and the frustrated  $J_1 - J_2$  spin-1/2 model in 2D. Ref. [54] explores the possibility to directly target excited states, see Sec. 3.2, and compares the capabilities of FFNNs and restricted Boltzmann machines (Sec. 2.2) to represent excited states of the one-dimensional Heisenberg and Bose-Hubbard models. A Bose-Hubbard model on a ladder with strong magnetic flux is studied using a FFNN in Ref. [55]. In Ref. [56], a FFNN is trained to represent the ground state of the one- and two-dimensional Bose-Hubbard model. By using the particle number as well as the interaction strength  $U$  as additional input parameters to the network, the ground state can be directly obtained without or with little re-training for different Hamiltonian parameters.

Furthermore, FFNNs were applied to simulate quantum systems with continuous degrees of freedom. In Ref. [57], a FFNN is used to simulate the ground state of the Calogero-Sutherland model in one dimension and Efimov bound states in three dimensions, where the particle positions in real space are used as input to the network. Another approach to use FFNNs to simulate continuous quantum systems was taken in Ref. [58] using Radial Basis Function (RBF) networks. RBF networks consist, as FFNNs, of an input layer, one or more hidden layers with RBFs as activation functions, and an output layer. RBFs are of the general form

$$f(x) = \phi(\|\mathbf{x} - \mathbf{c}\|), \quad (6)$$

where  $\mathbf{x}$  is a point in the input space,  $\mathbf{c}$  is the center of the radial basis function,  $\|\cdot\|$  denotes a distance measure, such as the Euclidean distance, and  $\phi$  is a radial function, such as the Gaussian function. The neurons' activation is thus determined by the distance of the input  $\mathbf{x}$  from certain points in the input

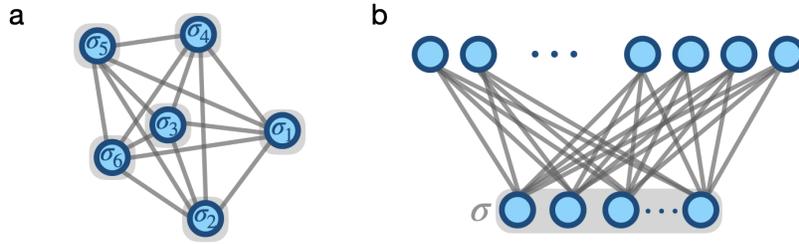


Figure 3: a) Hopfield network: All-to-all connections between the physical sites are learned by the network. b) Restricted Boltzmann machine (RBM): An additional set of hidden, unphysical nodes is introduced. Furthermore, the Boltzmann machine is restricted to only inter-layer connections which connect the physical (visible) layer and the hidden layer.

space  $\mathbf{c}$ , known as centers, which are trainable network parameters. The RBF activation functions strongly depend on the distance to a center point, which allows them to capture variations in data that are radially symmetric. In Ref. [58], the input to the RBF corresponds to the quantum numbers of e.g. an undisturbed quantum harmonic oscillator, and the network parameters are optimized to represent the ground state of a quantum harmonic oscillator with an additional applied field.

## 2.2. Restricted Boltzmann Machines (RBMs)

Restricted Boltzmann machines are energy based models, i.e. they are governed by an energy function  $E_{\theta}(\boldsymbol{\sigma})$  for configurations  $\boldsymbol{\sigma}$ . Using statistical physics, the respective probability distribution of these models is directly related,

$$p_{\theta}(\boldsymbol{\sigma}) = \frac{1}{Z_{\theta}} \exp(-E_{\theta}(\boldsymbol{\sigma})), \quad (7)$$

with the normalization constant  $Z_{\theta}$ . A first example for energy based models were Hopfield networks [59] shown in Fig. 3a, which consist of all-to-all connected nodes with connections  $W_{ij}$  and the biases  $b_i$ , similar to an Ising model with long-range interactions and local magnetic field.

When being used to model physical systems, the number of nodes in a Hopfield model corresponds to the number of physical sites in the system under consideration (*visible nodes*  $\boldsymbol{\sigma}$ ). In contrast, Boltzmann machines (BMs) increase the expressiveness by introducing additional, unphysical nodes (*hidden nodes*  $\mathbf{h}$ ) and the respective connections that increase the expressiveness of the network. With their all-to-all connections between all visible and hidden nodes, BMs are very expressive but can be hard to train. Hence, they are mostly used in their restricted version, see Fig. 3b, where only visible-to-hidden node connections  $W_{ij}$  and no hidden-to-hidden or visible-to-visible node connections are considered. Analogously to statistical physics, the energy of a restricted Boltzmann machine (RBM) is given by

$$E_{\theta}(\boldsymbol{\sigma}, \mathbf{h}) = - \sum_{ij} W_{ij} h_i \sigma_j - \sum_j b_j \sigma_j - \sum_i c_i h_i, \quad (8)$$

with the biases in the visible and hidden layers,  $b_j$  and  $c_i$ , respectively. The corresponding probability for a given input configuration of physical sites  $\boldsymbol{\sigma}$  is given by [60]

$$p_{\theta}(\boldsymbol{\sigma}) = \sum_{\mathbf{h}} p_{\theta}(\boldsymbol{\sigma}, \mathbf{h}) = \sum_{\mathbf{h}} \frac{\exp(-E_{\theta}(\boldsymbol{\sigma}, \mathbf{h}))}{Z_{\theta}}. \quad (9)$$

An overview on the application of RBMs in physics can be found in Ref. [60].

To use RBMs for representing quantum states, apart from the amplitude  $p_{\theta}(\boldsymbol{\sigma}) = |\psi_{\theta}(\boldsymbol{\sigma})|^2$  a phase of the RBM has to be defined. This can be done in several ways, e.g. by making the network parameters  $\boldsymbol{\theta}$  complex [25] or by modeling the phase with a separate RBM [61].

The expressivity of the RBM ansatz is often studied in the framework of tensor networks [48, 47, 62, 63], using the entanglement that can be captured with an ansatz as an indicator for the

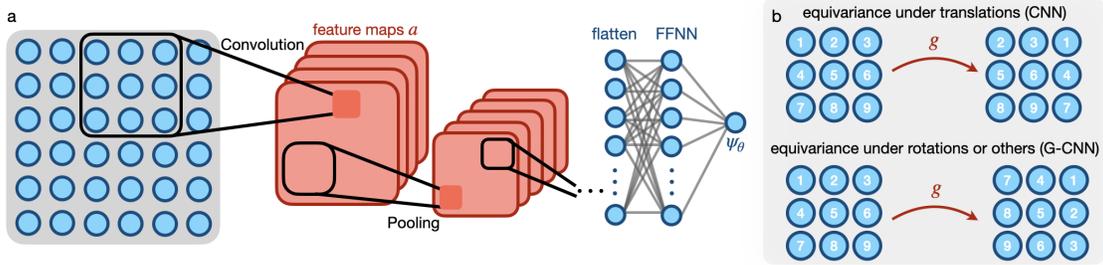


Figure 4: a) Convolutional neural network (CNN) with convolutions as defined in Eq. (10), pooling layers that downsample the spatial dimension and fully connected layers as final layers. b) Comparison of equivariance w.r.t. the translational group present in conventional CNNs (top) and w.r.t. other groups, e.g.  $C_4$  rotations (bottom) that can be implemented in G-CNNs.

representability. For RBMs, the connectivity between visible and hidden layers, that indirectly couples all sites of the physical system, allows for the entanglement entropy to scale with a subregion's volume, in contrast to its area [27]. In particular, this can make RBMs more efficient in capturing volume-law entangled states compared to e.g. MPS or PEPS. However, the efficiency of shallow RBMs to represent general quantum states has limitations, but they can be overcome with deep RBMs [24, 28, 64, 65]. This was further confirmed empirically e.g. in Refs. [66, 67], where random matrix product states were learned with shallow and deep RBMs using supervised approaches. Furthermore, RBMs can, due to their connectivity, straightforwardly be applied to higher dimensions, e.g. 3D systems [68].

RBMs have been used for modeling a large number of physical systems, among them frustrated spin systems [69, 70], spin liquids [71], topologically ordered states [72, 73, 31], the Toric code [72, 74], Bose-Hubbard models [75, 54, 76, 77], strongly interacting fermionic systems [78], boson - fermion coupled systems such as electron - phonon coupled systems [79], and molecules [80]. In these works, often variants of RBMs are used, e.g. the correlator RBM where correlations are introduced into the RBM energy functional based on physical insights [74]. Another modification, the convolutional RBM (CRBM) (see Sec. 2.3), makes use of the fact that physical models are typically translationally invariant and feature local interactions. This is taken into account by introducing an additional convolutional layer between the visible and the hidden layers and is employed e.g. in Refs. [81, 82] for the simulation of Ising and Kiteav models as well as the Hubbard model. Furthermore, the implementation of symmetries was shown to improve the results [83]. In Ref. [84, 85] further symmetries such as non-abelian or anyonic symmetries are considered. RBMs have also been used for the simulation of real-time dynamics [25, 86, 87, 88, 89, 90], see also Sec. 3.3.

### 2.3. Convolutional Neural Networks (CNNs) and Group CNNs

Convolutional neural networks (CNNs) are used in processing data with a grid-like topology, most commonly two-dimensional data like images. The building blocks of CNNs are shown in Fig. 4: First, convolutional layers employ filters or kernels to scan the input, which can detect local patterns and capture spatial relationships. Basically, each filter in the network uses the same weights for different parts of the input, making CNNs translationally invariant. This approach significantly reduces the number of parameters compared to fully connected networks. Second, pooling layers downsample the spatial dimensions, reducing computational complexity while preserving important features. The final layer of a CNN typically consists of one or more fully connected layers. The convolution of the input data/feature map  $a$  with a given filter or kernel  $k$  corresponds to

$$[a * k](x) = \sum_y \sum_{i=1} a_i(y) k_i(x - y). \quad (10)$$

The result of this operation is a new feature map.  $a_i(y)$  refers to the value of the  $i$ -th channel of the input feature at position  $y$ . For an RGB image, for example, there are three channels (red, green, blue), and  $a_i(y)$  is the intensity of one of these colors at pixel  $y$ .  $k_i(x - y)$  is the value of the  $i$ -th channel of the kernel at position  $x - y$  within the kernel. The kernel is slid over the input

image (or feature maps at later layers), and for each position, the kernel values are multiplied with the corresponding input feature values.

In the context of NQS, CNNs are regarded as a viable approach to deal with the properties of square lattice spin systems. The application of CNNs to solve the highly frustrated  $J_1 - J_2$  antiferromagnetic Heisenberg model was first introduced in Ref. [91]. In these systems the sign problem remains a significant challenge for quantum Monte Carlo approaches. Ref. [91] demonstrates how CNNs effectively tackle the challenging problem of finding the ground state of such models. Furthermore, Ref. [30] shows that deep CNNs can encode volume-law entangled states efficiently, requiring only  $O(\sqrt{N})$  parameters to represent a 2D system with  $N$  particles, instead of  $O(N)$  for RBMs or  $O(N^2)$  for fully connected networks. In a typical neural network, adding more layers, i.e. making the network deeper, theoretically allows the network to learn more complex features and improve its performance on tasks. However, in practice, when the network gets too deep, one faces problems such as vanishing gradients, where the gradients (which are used to update the weights in the network during training) become very small and make learning very slow or even stop it entirely. This makes deep CNNs hard to train and often leads to poorer performance. In order to address this gradient issue, Ref. [92, 93] attempted to use deep CNNs for a frustrated  $J_1 - J_2$  model on a square lattice using different techniques for network training, including a variant of stochastic reconfiguration tailored for large parameter numbers [93]. Besides these ground state calculations, CNNs have also been applied for dynamics simulations [87, 94, 95], see also Sec. 3.3.

In Ref. [96], a novel approach is introduced for adapting CNNs to other common lattice structures such as triangular lattices, which are somewhat analogous to sheared square lattices, allowing the application of regular CNN filters. In the same work, the authors consider honeycomb and Kagome lattices, where the key techniques involve augmenting these lattices with strategically placed virtual vertices, effectively transforming them into grid-like structures akin to triangular lattices. This allows for the application of standard CNN convolutional kernels while preserving the unique properties of the original lattices. The method enhances information processing and exchange, expanding the receptive field and enabling the analysis of varied local structures and staggered arrangements unique to these lattices.

CNNs can also be made autoregressive, allowing for sampling directly from the NQS' amplitudes instead of more involved Markov chain sampling, see Sec. 2.6. This property can be enforced by masking the CNN inputs  $\sigma_i, \dots, \sigma_N$  for the  $i$ -th output vector to suppress their contribution to the output, as is done in Ref. [34]. The autoregressive CNN has been used for state reconstruction by [97], see also Sec. 3.6.1.

Although CNNs exhibit translational invariance, they lack the ability to learn additional types of symmetries, such as rotation or mirror symmetries. Typically, data augmentation is employed to train the model for these specific symmetries [92]. In Ref. [98] the wave function was symmetrized in order to incorporate the rotational symmetries, see also Sec. 2.8. A more intrinsic solution is the development of group convolutional neural networks (G-CNNs). These networks extend the capabilities of standard CNNs by using group theory, allowing them to automatically incorporate various symmetries, see Fig. 4b. The key component of G-CNNs is the group convolution operation. An equivariant convolution ensures that if the input is transformed (e.g., rotated), the output feature maps will be transformed in the same way. The group-equivariant convolution of the input/feature map  $a$  with a kernel  $k$  under the group  $G$  evaluated at a group element  $g$  corresponds to

$$[a *_G k](g) = \sum_{h \in G} a(h)k(g^{-1}h). \quad (11)$$

This convolution operation is designed to be equivariant to the transformations in the group  $G$ .  $G$  can be a group of (discrete) rotations, translations, or other transformations. Ref. [99] considers the full wallpaper group, consisting of translation, rotation and mirror symmetry. The first convolution takes the input  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$ , where  $i$  are the positions in the lattice, and transforms it as

$$a^1(g) = f \left( \sum_i k(g^{-1}\mathbf{i})\sigma_i \right), \quad (12)$$

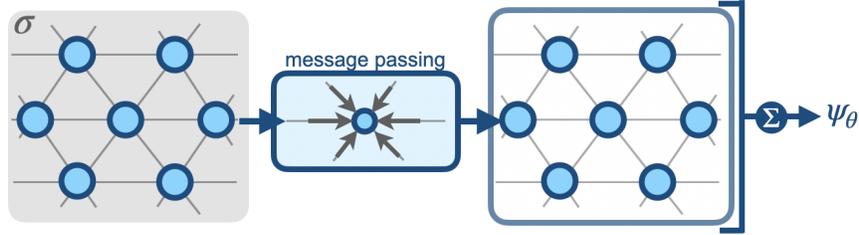


Figure 5: Graph neural network: In each permutation equivariant layer, the weights are updated through typically pairwise message passing, i.e. the value of a given node is updated according to the values of its immediate neighbors. Graph neural networks can easily be applied to any lattice geometry, here we show a hexagonal lattice as an example.

where  $f$  is a point-wise non-linear activation function, to obtain the value of the feature map  $a(g)$  for the group element  $g$ . This first (embedding) layer thus generates equivariant feature maps, which are indexed with group elements, from the input. After repeating the application of group convolution and non-linearity for  $l$  layers, the wave function coefficient  $\psi(\sigma)$  is determined as

$$\psi(\sigma) = \sum_g \chi_{g^{-1}} \exp(a_g^l), \quad (13)$$

where  $\chi_g$  is the character of the symmetry operation  $g$ .

Ref. [99] uses G-CNNs to determine the ground state energy of the  $J_1 - J_2$  Heisenberg model on a square and triangular lattice. Ref. [100] underscores the capability of very deep G-CNNs in achieving high-accuracy results for the same lattices, and furthermore directly calculates low-lying excited states by changing the characters of the symmetry operations  $\chi_g$  accordingly. For a Heisenberg model on a Kagome lattice, one of the most studied models in frustrated magnetism since it is a promising candidate to host exotic spin liquid states, Ref. [101] presents a new ground state: the spinon pair density wave (PDW), which does not break the time-reversal and lattice symmetries. G-CNNs are used to study the ground state of the  $J - J_d$  Heisenberg model on the Maple-Leaf lattice, which results in dimer state paramagnetic and canted magnetic order phases for different values of  $J_d/J$ , [102].

#### 2.4. Graph Neural Networks

Graph neural networks directly take the geometry of the underlying problem into account [103]. In the case of NQS, this means that the lattice geometry of the Hamiltonian under consideration is used as the graph structure. Throughout the graph neural network, this graph structure is kept, see Fig. 5. In Ref. [104], a sublattice encoding, denoting the position of the site within the unit cell, is used as additional input for each site on the lattice. Subsequently, in each permutation equivariant layer of the graph neural network, the values of the nodes are updated through typically pairwise message passing. This means that the value of a given node is updated according to the values of its immediate neighbors, thus directly taking the graph structure into account. The specific details of this updating procedure are design choices, leading e.g. to graph convolutional neural networks [105] or gated graph sequence neural networks [106], where a gated recurrent unit is used.

One advantage of the graph structure and message passing layers is that a transfer to different system sizes is straightforwardly possible, as shown in Ref. [104]. Ref. [107] considers the ground state of the hard-core bosonic  $t - V$  model on different lattice geometries, such as the Kagome and triangular lattices. Since this constitutes a stoquastic Hamiltonian, no sign structure has to be learned. In Ref. [104], the ground state of the  $J_1 - J_2$  Heisenberg model on square, triangular, honeycomb and Kagome lattices is studied, in which case a non-trivial sign structure exists. The performance for using complex network weights as well as separate networks for amplitude and phase are compared.

Permutation equivariant message passing has also been used in the context of neural network backflow transformations to simulate interacting fermions in continuous space in Refs. [108, 109]. Ref. [110] uses a graph neural network to represent a generalized pair amplitude in the context of a BCS type wave function.

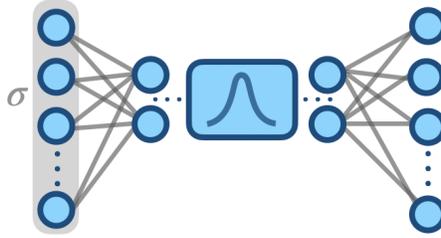


Figure 6: Variational autoencoder consisting of an encoder, the latent space or bottleneck layer, and a decoder. The latent space representation is a multivariate Gaussian, i.e. the encoder generates the values for mean and variances of the Gaussian distribution. By sampling from this Gaussian, the input for the decoder layer is generated.

### 2.5. Latent Space Representations: Autoencoders

An autoencoder consists of an encoder, the latent space or bottleneck layer, and a decoder, see Fig. 6. The encoder, typically consisting of several fully connected or convolutional layers, compresses the input into just a few nodes in the bottleneck layer. The decoder subsequently generates an output based on the information in the bottleneck layer. The network parameters are optimized such that the generated output is as close to the input as possible. In a variational autoencoder [111], the encoder generates the values for mean values and variances in the bottleneck layer, and the values used as input for the decoder are then sampled from a multivariate Gaussian.

Variational autoencoders have been used in Refs. [112, 113, 114] for (quantum) state reconstruction, where the input consists of the measured data, which the autoencoder learns to compress and de-compress using encoder and decoder. After training, the encoder can be dropped, and by sampling random numbers as input to the latent space, new, uncorrelated samples can be generated.

In Ref. [112], this approach is used to reconstruct positive wave functions, i.e. effectively, the probability distribution of the samples in the computational basis is learned. The efficiency of the compression is quantified by the ratio of the number of network parameters to the Hilbert space dimension. In this case, the size of the latent space is always chosen to correspond to the system size. Ref. [113] uses a conditional variational autoencoder to perform state reconstruction of ground states of the 1D transverse field Ising model based on informationally complete positive-operator valued measures. The magnetic field  $h$  is the *condition*, which is used as additional input to the decoder. Furthermore, the autoencoder representation of the quantum state is appealing since its latent space contains information on the state under consideration: In Ref. [114], the low-dimensional latent space representation of finite temperature samples of the 2D Ising model is used to extract physical features. The authors of Ref. [115] determine the minimal size of the latent space needed to reproduce local observations to measure the local complexity of time-evolved states.

### 2.6. Autoregressive Networks

Autoregressive architectures are characterized by their normalized amplitudes  $p_{\theta} = |\psi_{\theta}|^2$ . The use of autoregressive networks for NQS was first proposed by Sharir et al. [34]: At a local configuration  $\sigma_i$ , the authors propose to mask out the local configurations  $j \geq i$ , and only consider sites  $\sigma_{<i} = (\sigma_1, \dots, \sigma_{i-1})$ , such that the network represents  $p_{\theta}(\sigma_i | \sigma_{<i})$ . The total probability is given by

$$p_{\theta}(\sigma) = \prod_i^N p_{\theta}(\sigma_i | \sigma_{<i}). \quad (14)$$

This allows to normalize  $p_{\theta}$  by normalizing each conditional  $p_{\theta}(\sigma_i | \sigma_{<i})$ , and hence to sample directly from the amplitudes instead of more elaborate sampling procedures like Markov chain sampling needed for non-autoregressive architectures. Since the generation of many, uncorrelated samples is crucial for the training, this can yield a speed up and an improvement of the optimization. However, Ref. [116] suggests that the autoregressive sampling can in some cases reduce the expressivity of the neural network wave function. Furthermore, the application of stochastic reconfiguration for optimizing

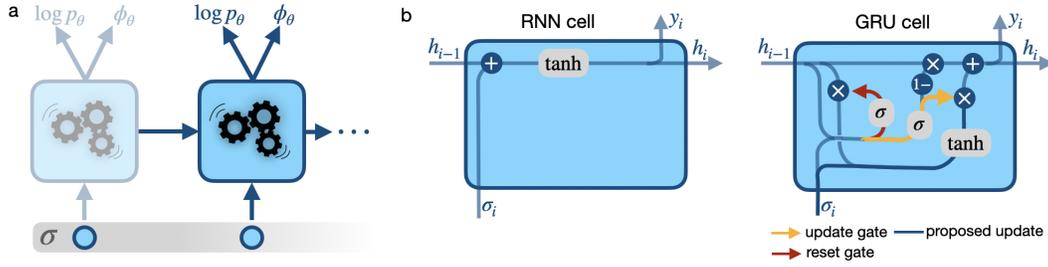


Figure 7: a) Recurrent neural network (RNN). At each site denoted by the blue circles, the local configuration is passed through the RNN cell. Furthermore, information from previous sites is passed to the next cell (horizontal arrows). b) Left: A plain vanilla RNN cell, where the local configuration  $\sigma_i$  and the hidden state  $h_i$  are simply added and passed through a  $\tanh$ -activation function. Right: Gated recurrent unit (GRU), enabling the RNN to capture longer-ranged correlations throughout the system. In a GRU, the proposed update (dark blue), reset and update gates (red and orange) decide how much of the old state is kept.

autoregressive NQS can cause problems, as discussed in Sec. 3.1. For further reading on the potential of autoregressive networks in the context of quantum physics and NQS we refer the reader to Ref. [41].

*2.6.1. Recurrent Neural Networks (RNNs)* Recurrent neural networks (RNNs) consist of several RNN cells, and information is passed from one cell to the next, in a recurrent manner, through the network, as schematically shown in Fig. 7a.

The first applications of RNNs to represent quantum states have considered one-dimensional spin systems [117, 118]. In these cases, the RNN is constructed by  $N$  cells and the information is passed from the first cell corresponding to the first spin of the 1D chain to the last cell in a recurrent fashion. At each lattice site  $i$ , the cell receives a local spin configuration  $\sigma_i$  and the so-called *hidden state*  $h_{i-1}$  that passes information from previous lattice sites through the network. The cell then outputs the updated hidden state  $h_i$  as well as an output  $y_i$  that can be used to calculate the local conditional probability and a local phase of the state representation. Normally, each cell is represented by the same weights (weight sharing), but in some contexts the cells can also be chosen to have different weights [119]. In the former case, the RNN architecture is tailored to model bulk properties and hence becomes particularly effective for large systems [118]. Furthermore, it is possible to iteratively retrain on larger and larger systems, which can improve the performance for large systems [118].

The local amplitude at each RNN cell is given by a conditional probability determined by the previous spin configurations  $\sigma_{<i}$ , i.e.  $p_\theta(\sigma_i|\sigma_{<i})$ . The activation function of the RNN's output layer can be chosen such that the local amplitude is normalized and hence also the total amplitude Eq. (14), making the RNN autoregressive.

To model long-range correlations, it is crucial that the information is passed through the cells in an efficient way. This is usually done by replacing the plain vanilla RNN cells with gated recurrent units (GRU) [120], see Fig. 7b, enabling a long-term memory of the RNN [121]. In one-dimensional settings, this modification yields successful representations of spin systems like Heisenberg and transverse field Ising model [118, 117]. For two-dimensional systems, the hidden states can be passed in a 1D snake through the system, similar to MPS calculations, as e.g. in Ref. [118]. However, it is also possible to pass the information in a 2D fashion through the system, as proposed in Ref. [122] and further improved by introducing a tensorized version of a GRU in Ref. [119]. Furthermore, the authors show that an imposed  $U(1)$  magnetization conservation and spatial symmetries as well as direct implementation of the Marshall sign rule for spin-1/2 systems improve the results, in agreement with other works [123, 117, 118]. This  $U(1)$  symmetry is usually imposed by setting  $p_\theta(\sigma_i|\sigma_{<i}) = 0$  if the system with the new sampled  $\sigma_i$  violates the corresponding conservation law. For example, for the  $U(1)$  particle number conservation of hardcore bosons with local states  $\sigma_i = 0(1)$  corresponding to empty (occupied) sites,  $p_\theta(\sigma_i = 1|\sigma_{<i}) = 0$  with if  $\sigma_{<i}$  is already in the correct particle number sector. With these modifications, RNNs have been applied in many contexts, including the Heisenberg model on square and triangular lattices [124], prototypical states in quantum information [125], states with topological order [126, 127] and fermionic systems using Jordan-Wigner strings [35]. In the context

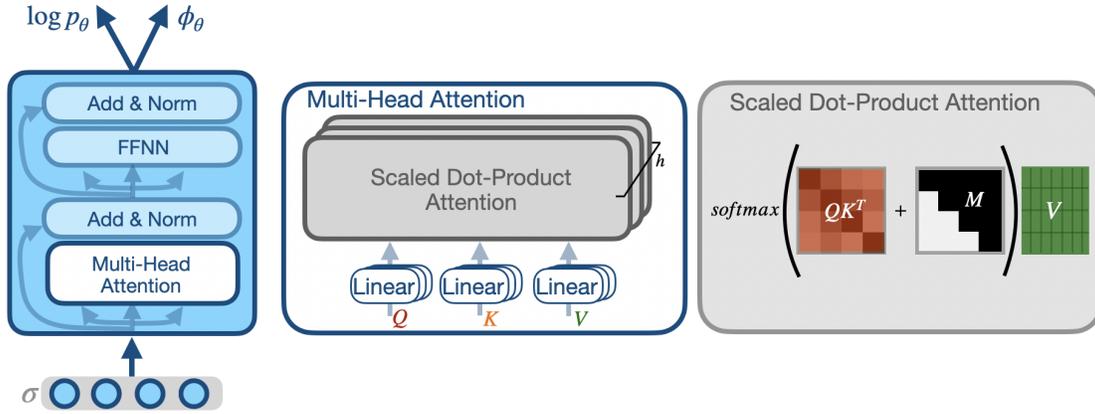


Figure 8: Left: The transformer neural network wave function, adapted from Ref. [132]. The input is passed through the multi-head attention mechanism, addition and normalizing layers (Add and Norm) and a feed forward neural network (FFNN). Middle: The multi-head attention itself consists of  $h$  scaled dot-product attention layers. Right: The scaled dot-product attention according to Eq. (16), based on the attention map  $QK^T$ , which can in principle represent all-to-all correlations throughout the system. Sites of the system that have not been considered yet are masked out by adding the mask  $M$ , before passing it through a softmax activation layer and multiplying with the values  $V$ .

of real-time dynamics simulations, see also Sec. 3.3, RNNs have been used e.g. in Refs. [128, 129, 130].

In order to investigate the expressivity of the RNN ansatz, the authors of Ref. [49] present a mapping from tensor networks to 1D MPS-RNNs and 2D tensorized MPS-RNNs, i.e. RNNs with linear or multilinear update rules for the hidden states and quadratic output layers. For linear update rules and one-dimensional settings, MPSs can be mapped to the 1D MPS-RNN with the same number of variational parameters, but not vice versa, making the latter potentially more expressive than MPS. The 2D version of the MPS-RNN receives hidden states from two directions, inspired from projected entangled pair states (PEPS) [131], and features multilinear updates. This architecture is shown to encode an area law of entanglement entropy, but unlike PEPS, it supports perfect sampling and hence efficient evaluations of the wave function. In particular, receiving hidden states from two directions makes the RNN more efficient in state compression compared to the class of TNs which support wave function evaluation in polynomial time [30].

**2.6.2. Transformers** A transformer model relies entirely on an attention mechanism which draws global dependencies between input and output [132]. This self-attention layer in the transformer setup generates all-to-all interactions between the sites in the system. These trainable connections can potentially represent strong connections or correlations, regardless of their position [133, 134]. The transformer first embeds the different given input elements into a unified feature space. This embedding corresponds to a linear projection, with trainable parameters, of the input elements with a dimension  $d_i$  to elements with an embedded dimension  $d_h$ . The position of the inputs in the sequences are not explicitly modeled in the transformer, but efficiently transformed into abstract representations using positional encoding vectors that are added to the embedded input vectors [134]. Each embedded input element  $\sigma_i$  is projected on a query vector ( $q_i$ ), key vector ( $k_i$ ) and a value vector ( $v_i$ ) of the same dimension  $d_h$  as the embedded input, given by:

$$q_i = \sum_{l=1}^{d_h} W_{i,l}^q \sigma_{i,l}, \quad k_i = \sum_{l=1}^{d_h} W_{i,l}^k \sigma_{i,l}, \quad v_i = \sum_{l=1}^{d_h} W_{i,l}^v \sigma_{i,l}, \quad (15)$$

with the matrices  $W^q$ ,  $W^k$  and  $W^v$  to be the trainable weight matrices of dimension  $d_h \times d_h$ . The query, key and value matrices are then given by  $Q = (q_1, \dots, q_N)$ ,  $K = (k_1, \dots, k_N)$  and  $V = (v_1, \dots, v_N)$ . The model can be made autoregressive by using a masked self-attention layer, which allows connections to all previous elements in the sequence [132]. This autoregressive structure enables efficient exact

sampling from the model [135, 128]. Then, after the masking term  $M$  is added to the signal, a softmax activation function is applied to the masked dot product of the vectors  $Q$  and  $K$ . In multi-headed attention, each query, key and value vector is mapped to  $h$  vectors with trainable weight matrices, with  $h$  the number of attention heads. This is indicated in Fig. 8 (middle). The complete attention formalism can be summarized by

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK}{\sqrt{d_h/h}} + M \right) V, \quad (16)$$

as shown in Fig. 8 (right).

Transformer quantum states can learn ground state properties of various physical systems, such as the 1D transverse field Ising model and the 1D Heisenberg  $J_1 - J_2$  and XYZ model [135, 134]. Comparable results to DMRG calculations have been obtained for a 1D frustrated spin model, with a relatively low number of parameters [134]. In Ref. [128], transformers have shown to be able to simulate the real-time dynamics and steady state in 1D and 2D transverse field Ising and Heisenberg models [128]. In Refs. [136, 137, 138, 139] transformers are used in the context of quantum chemistry calculations.

Small modifications of the transformer model, leading to the so-called vision transformers (ViT), inspired the use of patched transformers. This model splits the system into patches, and can be used to calculate the ground state properties of frustrated spin models [134]. A large patch size enhances the efficiency of the transformer, but on the other hand the network output dimension increases exponentially with the patch size, as it encodes the probability distribution over all possible patch states. To overcome this, large patched transformers are introduced in Ref. [133]. In this ansatz, the output of the patched transformer is passed to a patched RNN as the initial hidden state, which breaks the large inputs into smaller sub-patches, reducing the output dimension. This model has been shown to accurately capture ground state properties and phase transitions of large Rydberg systems, which can compete with quantum Monte Carlo results [133].

In Ref. [135], transformer quantum states have been used to learn ground state properties of a single system, as well as to generalize to different, unseen systems. For the latter, not only the physical degrees of freedom, but also the parameters of the Hamiltonian of the system are used as input. These parameters can be formulated as new elements that have to be passed to the embedding layer. After training the transformer quantum state for a Hamiltonian with different parameters, the transformer is able to generate the ground state for unseen Hamiltonian parameters without any additional training. Although these are with slightly larger error, more accurate results can be achieved with less training than without any a priori training.

In Refs. [140, 141], a transformer with a so-called factored attention is used. In contrast to the conventional attention mechanism (16), where the attention weights  $QK^T$  and the values  $V$  are calculated from both embedded inputs and the positional encoding, factored attention uses  $QK^T$  that depend only on positions, and  $V$  that depend only on the embeddings. Ref. [140] shows that training a model with a single self-attention layer with factored attention can be mapped to solving the inverse Potts problem using the pseudo-likelihood method. This method, in combination with the patched transformer, leads to high quality results for the ground-state energy of the  $J_1 - J_2$  Heisenberg model [140].

## 2.7. NQS for Fermionic Systems

NQS for the representation of fermionic quantum states can be divided into distinctly different ansätze: (i) NQS ansätze that inherently incorporate the fermionic statistics and (ii) bosonic NQS that are antisymmetrized by a Jordan-Wigner transformation, see Fig. 9a and b respectively.

**2.7.1. Fermionic Architectures** The antisymmetry of NQS ansätze with fermionic statistics can be achieved in various ways. The most commonly used ansatz for fermionic variational wave functions is a Slater-Jastrow-inspired ansatz, where the wave function is constructed from an antisymmetric part  $\psi_0$ , typically a Slater determinant, and a Jastrow factor  $\mathcal{J}$  capturing the correlations, i.e.

$$|\psi_{\theta, \nu}\rangle = \sum_{\sigma} \psi_{0, \theta}(\sigma) \mathcal{J}_{\nu}(\sigma) |\sigma\rangle, \quad (17)$$

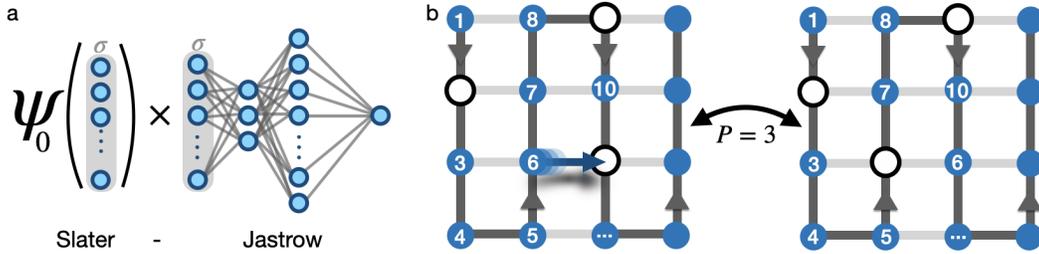


Figure 9: NQS for fermionic systems: a) Fermionic architectures are typically inspired from Slater-Jastrow ansätze with an antisymmetric Slater determinant and a Jastrow factor that captures the correlations throughout the system. b) Bosonic architectures can be made fermionic by applying Jordan-Wigner strings. Effectively, this boils down to counting the number of permutations  $P$  of configurations that are connected by matrix elements when calculating observables and multiplying the respective terms in Eq. (19) by  $(-1)^P$ .

where in principle both  $\psi_0$  and  $\mathcal{J}$  can be parameterized by neural networks with parameters  $\theta$  and  $\nu$ . This is shown schematically in Fig. 9a.

In the setting of first quantization, architectures like FermiNet [142] and PauliNet [143, 144] that use Slater determinants  $\psi_0$  reach high accuracies in *ab initio* molecule simulations. However, the evaluation of Slater determinants is costly in first quantization, which is overcome e.g. in Ref. [145] by an antisymmetric construction of  $\psi_0$  by deep neural networks. However, these approaches often come at the price of a reduced accuracy [146].

For quantum many-body systems, mostly second quantization is used, despite some exceptions e.g. for repulsively interacting, spin-polarized fermions [147], where the authors chose to model  $\psi_0(\sigma)\mathcal{J}(\sigma)$  in Eq. (17) by a single neural network. In the second quantization approaches used in the literature, in many cases machine learning approaches are used to enhance the expressivity of the Slater-Jastrow ansatz. This can be done by employing NNs to parameterize the Jastrow factor  $\mathcal{J}$ , or by using machine learning methods to dress the Slater determinant. For the latter, usually two methods are used [148]: (i) The hidden fermion determinant state, where neural networks are used to replace the standard Slater determinant with a larger determinant which includes single particle orbitals from additional projected hidden fermions [149]. (ii) Neural backflow transformations, which add correlation by making the single particle orbitals of the Slater determinant configuration dependent, with the respective transformation learned by a neural network [150, 151, 143, 142, 109].

One of the first examples for fermionic NQS is the RBM+PP architecture in Ref. [152], with a slightly different ansatz than Eq. (17), i.e.

$$|\psi_\theta\rangle = \sum_{\sigma} \psi_{\text{ref}}(\sigma) \mathcal{F}_\theta(\sigma) |\sigma\rangle, \quad (18)$$

where correlations on top of a reference state  $\psi_{\text{ref}}$  are modeled by a generalized version of an RBM, with additional artificial neurons to mediate entanglement, that is represented by  $\mathcal{F}_\theta$ . In this work, the authors take  $\psi_{\text{ref}}$  to be a pair-product state (PP) that already incorporates some of the entanglement, and test the architecture for the Fermi-Hubbard model.

Furthermore, it is worth mentioning that in simulations of lattice models with a Slater-Jastrow variational wave function, the autocorrelation time increases drastically for large system sizes, motivating the development of a fully autoregressive Slater-Jastrow ansatz by combining a Slater determinant with an autoregressive deep neural network as a Jastrow factor [153].

**2.7.2. Bosonic Architectures with Jordan Wigner Strings** The other way to simulate fermionic systems using NQS are Jordan-Wigner (JW) transformations, which is used to map the bosonic NQS to a fermionic wave function. Hence, per se bosonic architectures can be used and no special fermionic

architecture is needed. The JW transformation is given by

$$\hat{c}_j^{(\dagger)} = \hat{\sigma}^{-(+)} \exp \left( i\pi \sum_{k < j} \hat{\sigma}_k^+ \hat{\sigma}_k^- \right), \quad (19)$$

where indices refer to a one-dimensional labeling of the fermions, the  $\hat{\sigma}^\pm$  are the spin raising and lowering operators and  $\hat{c}$  fulfill the fermionic commutation relations. More precisely, for two fermions at site  $i$  and  $j$ , exchanging these fermions yields a minus sign arising from the argument of the exponential in Eq. (19). This rule does not have to be implemented in the NQS architecture itself, but only on the level of the calculation of expectation values. For an operator  $\hat{O}$  with  $\langle \hat{O} \rangle = \langle O_\theta^{\text{loc}}(\boldsymbol{\sigma}) \rangle_\sigma$  (see Eq. (3)) with

$$O_\theta^{\text{loc}}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} \frac{\langle \boldsymbol{\sigma} | \hat{O} | \boldsymbol{\sigma}' \rangle \langle \boldsymbol{\sigma}' | \psi_\theta \rangle}{\langle \boldsymbol{\sigma} | \psi_\theta \rangle} \eta_{\boldsymbol{\sigma}\boldsymbol{\sigma}'}, \quad (20)$$

each matrix element is multiplied by a factor  $\eta_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} = (-1)^{P_{\boldsymbol{\sigma}\boldsymbol{\sigma}'}}$  if  $\boldsymbol{\sigma}'$  is connected to  $\boldsymbol{\sigma}$  by  $P_{\boldsymbol{\sigma}\boldsymbol{\sigma}'}$  two-particle permutations, see Fig. 9b. This method was applied to simulate molecules [154], for Fermi-Hubbard and  $t - J$  models [155, 35] and solid state systems [156].

Despite its successful application, JW strings come with the disadvantage that the operators in Eq. (19) are highly non-local, which can cause problems for some architectures. Whether antisymmetrizing bosonic networks is as efficient as using inherently fermionic architectures is still under debate [29].

### 2.8. Other Design Choices

Besides the architecture, other design choices can influence the performance of the NQS:

One choice is the way how amplitude and phase of the NQS are calculated. Hereby, amplitude and phase can be learned by two different, real-valued networks, one real-valued network with two separate output nodes or final layers for phase and amplitude or by one network with complex weights. In Ref. [157], a complex-valued RBM and a RBM in which two separate real-valued networks approximate amplitude and phase, are compared for the ground state of the  $J_1 - J_2$  model. In a systematic study on small clusters, they show that the complex RBM outperforms the latter.

A second design choice is how to incorporate symmetries in the NQS training in order to restrict the optimization space to states in the target symmetry sector, hence improving the performance [83, 123, 117, 45]. Firstly, global  $U(1)$  symmetries can be imposed, see e.g. Refs. [123, 117, 158], by restricting to configurations  $\boldsymbol{\sigma}$  that obey the respective symmetry, e.g. magnetization or particle number conservation. For autoregressive architectures, this is done by restricting the conditional probabilities to the targeted symmetry, see e.g. Sec. 2.6.1. For non-autoregressive architectures, the Monte Carlo updates can be chosen such that all generated configurations stay in the same symmetry sector. Second, spatial symmetries can be imposed. There are different symmetrizations used in the literature<sup>‡</sup>:

(i) bare-symmetry:

$$\psi_\theta^S(\boldsymbol{\sigma}) = \exp \left( \frac{1}{|\mathcal{T}|} \sum_{\boldsymbol{\sigma}^S \in \mathcal{T}(\boldsymbol{\sigma})} \log [\psi_\theta(\boldsymbol{\sigma}^S)] \right) \quad (21)$$

(ii) exp-symmetry

$$\psi_\theta^S(\boldsymbol{\sigma}) = \left( \frac{1}{|\mathcal{T}|} \sum_{\boldsymbol{\sigma}^S \in \mathcal{T}(\boldsymbol{\sigma})} \psi_\theta(\boldsymbol{\sigma}^S) \right) \quad (22)$$

<sup>‡</sup> We follow Ref. [45] here.

(iii) sep-symmetry

$$\psi_{\theta}^S(\sigma) = \sqrt{\exp\left(\frac{1}{|\mathcal{T}|} \sum_{\sigma^S \in \mathcal{T}(\sigma)} 2\text{Re}\{\log[\psi_{\theta}(\sigma^S)]\}\right) \cdot \exp\left(i \arg\left(\sum_{\sigma^S \in \mathcal{T}(\sigma)} \exp(i \text{Im}\{\log[\psi_{\theta}(\sigma^S)]\})\right)\right)} \quad (23)$$

for symmetry operations  $\mathcal{T}$  on samples  $\sigma$ . Note that only the last keeps the autoregressive property intact since  $\sum_{\sigma^S \in \mathcal{T}(\sigma)} |\psi_{\theta}^S(\sigma^S)|^2 = \sum_{\sigma^S \in \mathcal{T}(\sigma)} |\psi_{\theta}(\sigma^S)|^2$ . Furthermore, architectures that preserve certain symmetries explicitly can be used, such as group CNNs [99] and gauge equivariant neural Networks [159].

Furthermore, the authors of Ref. [53] show for the exemplary case of a FFNN that the choice of activation function can strongly influence the performance. Lastly, the number of parameters of the NQS can be varied. In Ref. [160] it is argued for the exemplary architecture of an RBM that using overparameterized NQS and subsequently pruning small parameters of the trained model can improve the performance. The compression by pruning is also discussed in Ref. [161]. However, increasing the number of parameters does not always yield an improvement: In Ref. [162] it is shown that the accuracy of an RBM increases for small widths of the hidden layer  $\alpha$ , but saturates at high  $\alpha$ . The authors observe that this behavior coincides with a saturation of the quantum geometric tensor’s rank, see Eq. (31) in Sec. 3.1, i.e. the local dimension of the relevant manifold for the optimized NQS saturates. Lastly, the open-source packages allow to readily optimize code for graphics processing units (GPUs), which can highly accelerate NQS implementations, see also Sec. 2.9. Another direction for fast and energy efficient NQS implementations is spiking neuromorphic hardware [163, 164, 165], as implemented for a RBM in Ref. [164].

For further reading on design choices beyond the discussion provided here, we refer the reader to Reh et al. [45], where the performance of RBMs, CNNs and RNNs with different symmetrization strategies are compared.

### 2.9. Open-Source Toolboxes

There are several toolboxes that provide open-source implementations of NQS: *NetKet* allows for ground state search, dynamics calculations based on TDVP and p-tVMC as well as state tomography using various architectures and comes with many implemented bosonic and fermionic Hamiltonians [166, 167]. *jVMC* [168], designed for computationally efficient variational Monte Carlo, provides several architectures for ground state search and dynamics simulations as well. *FermiNet* [142] provides ground state simulations for atoms and molecules. All of them are based on Google’s JAX library [169]. Lastly, we would like to mention *QuCumber* [170], a RBM based tomography implementation.

## 3. Applications of NQS

### 3.1. Ground States

To represent the ground state of a given system, neural quantum states are normally trained using variational Monte Carlo (VMC) [171, 172]. VMC is based on variational wave functions  $|\psi_{\theta}\rangle$  such as NQS, parameterized by parameters  $\theta$ . To approximate ground states with NQS, the energy

$$E_{\theta} = \frac{\langle \psi_{\theta} | \hat{H} | \psi_{\theta} \rangle}{\langle \psi_{\theta} | \psi_{\theta} \rangle} \geq E_{\text{gs}}, \quad (24)$$

should be as close as possible to the ground state energy  $E_{\text{gs}}$ . For variational wave functions  $\psi_{\theta}$ , this expectation value  $E_{\theta}$  can be evaluated from samples  $\sigma$  drawn from the wave function’s amplitude  $|\psi_{\theta}|^2$  according to Eq. (3). To approximate ground states, NQS are usually trained by minimizing the expectation value of the Hamiltonian,  $E_{\theta} = \langle \hat{H} \rangle \approx \langle H_{\text{loc}}(\sigma) \rangle_{\sigma}$ , i.e. parameters  $\theta$  are updated

according to

$$\begin{aligned}
\partial_{\theta_k} E_{\theta} &= 2\text{Re} \left[ \sum_{\sigma} P_{\theta}(\sigma) \frac{\partial_{\theta_k} \psi_{\theta}^*(\sigma)}{\psi_{\theta}^*(\sigma)} H_{\theta}^{\text{loc}}(\sigma) \right] - 2\text{Re} \left[ \sum_{\sigma'} P_{\theta}(\sigma') \frac{\partial_{\theta_k} \psi_{\theta}^*(\sigma')}{\psi_{\theta}^*(\sigma')} \sum_{\sigma} P_{\theta}(\sigma) H_{\theta}^{\text{loc}}(\sigma) \right] \\
&= \frac{2}{N_s} \text{Re} \left[ \sum_i^{N_s} \partial_{\theta_k} \log \psi_{\theta}^*(\sigma_i) [H_{\text{loc}}(\sigma_i) - \langle H_{\text{loc}}(\sigma') \rangle_{\sigma'}] \right] \\
&= 2\text{Re} \langle \partial_{\theta_k} \log \psi_{\theta}^*(\sigma) [H_{\text{loc}}(\sigma) - \langle H_{\text{loc}}(\sigma') \rangle_{\sigma'}] \rangle_{\sigma}.
\end{aligned} \tag{25}$$

For autoregressive architectures,  $\langle \psi_{\theta} | \psi_{\theta} \rangle = 1$  and hence the second term vanishes when calculating the derivative of Eq. (3). However, often  $H_{\text{loc}}(\sigma)$  is replaced by the covariate  $(H_{\text{loc}}(\sigma) - \langle H_{\text{loc}} \rangle)$  to reduce the variance of the gradients [173, 117], leading to the same expression as Eq. (25). Another approach is to (pre-)train the NQS with experimental or numerical data, see Sec. 3.6.1.

The optimization of NQS can be done with methods commonly used in machine learning, such as stochastic gradient descent, Adam [174] and AdamW [175]. A more elaborate approach is the stochastic reconfiguration (SR) algorithm [176, 177, 15], which incorporates the knowledge of the geometric structure of the parameter space to adjust the gradient direction [178, 179, 180]. The underlying idea§ of SR is to perform an imaginary time evolution of the variational state  $|\psi_{\theta}\rangle$ , i.e.

$$|\psi(\tau + \delta\tau)\rangle = \exp(-\delta\tau \hat{H}) |\psi_{\theta}\rangle \underset{\text{small } \delta\tau}{\approx} |\psi_{\theta}\rangle + |\delta\psi_{\hat{H}}\rangle. \tag{26}$$

For the latter equality, we have assumed small time steps  $\delta\tau$ , when the change of the state from the imaginary time evolution is

$$|\delta\psi_{\hat{H}}\rangle = -\delta\tau \hat{H} |\psi_{\theta}\rangle = -\delta\tau \sum_{\sigma} \psi_{\theta}(\sigma) H^{\text{loc}}(\sigma) |\sigma\rangle. \tag{27}$$

Naturally, the evolved state  $|\psi(\tau + \delta\tau)\rangle$  has less contributions from higher energy states, decreasing the energy with every imaginary time step  $\delta\tau$  [93]. In order to *translate* the evolved state  $|\psi(\tau + \delta\tau)\rangle$  to a parameter update  $\theta \rightarrow \theta'$ , a projection onto the variational manifold of  $\psi_{\theta'}$  is needed, which is done by minimizing the Fubini-Study (FS) distance  $d(\psi(\tau + \delta\tau), \psi_{\theta'})$  [182]. Expanding also for the projected state for small  $\delta\tau$ ,  $|\psi_{\theta'}\rangle = |\psi_{\theta}\rangle + |\delta\psi_{\theta}\rangle$ , with

$$|\delta\psi_{\theta}\rangle = \sum_{k, \sigma} \frac{\partial \psi_{\theta}(\sigma)}{\partial \theta_k} \delta\theta_k |\sigma\rangle = \sum_{\sigma} \psi_{\theta}(\sigma) \sum_k O_k(\sigma) \delta\theta_k |\sigma\rangle \tag{28}$$

and  $O_k(\sigma) = \frac{1}{\psi_{\theta}(\sigma)} \partial_{\theta_k} \psi_{\theta}(\sigma)$  [25], the FS distance can be written as

$$d(\psi(\tau + \delta\tau), \psi_{\theta'}) = \|\bar{O}\delta\theta - \bar{H}^{\text{loc}}\|_2 \tag{29}$$

with  $\|\dots\|_2$  the  $L^2$  norm, the matrix  $\bar{O}_k(\sigma) = \frac{1}{\sqrt{N_s}} (O_k(\sigma) - \langle O_k(\sigma) \rangle_{\sigma})$  and the vector  $\bar{H}^{\text{loc}}$  defined in analogy. This results in the SR equation

$$\bar{O}\delta\theta = \bar{H}^{\text{loc}} \Leftrightarrow \delta\theta = S^{-1} \bar{O}^{\dagger} \bar{H}^{\text{loc}} = S^{-1} F, \tag{30}$$

with the quantum geometric tensor

$$S = \bar{O}^{\dagger} \bar{O}, \tag{31}$$

and the vector of forces

$$F_k = \langle H_{\text{loc}} O_k^* \rangle - \langle H_{\text{loc}} \rangle \langle O_k^* \rangle.$$

Hence, the SR parameter update at the  $p$ -th iteration is given by

$$\theta(p+1) = \theta(p) - \gamma(p) S^{-1} F, \tag{32}$$

§ For the motivation of the SR algorithm from imaginary time evolution, we follow Refs. [93, 181].

with a scaling parameter  $\gamma(p)$ [25].

The SR update Eq. (32) hence involves an inversion of the  $S$  matrix. This inversion comes with the following problems: (i) The matrix  $S$  has to be estimated to a very high precision to avoid instabilities in the optimization, which requires a large number of samples. (ii)  $S$  is not necessarily invertible and hence, often a regularization of  $S$  is needed for a stable optimization, especially if considered close to critical points or in large spin systems [86, 25, 87]. Recent works indicate that the spectra of the  $S$  matrix are distinctly different for non-autoregressive and autoregressive architectures, which can cause problems for regularizing  $S$  for the latter [183, 35]. (iii)  $S$  is a matrix of typically very large dimensions  $N_\theta \times N_\theta$ , making the inversion computationally costly since the complexity of inverting  $S$  scales with  $\mathcal{O}(N_s N_\theta^2 + N_\theta^3)$  when using direct linear solvers [93]. To enlarge the allowed number of parameters by some orders of magnitudes, large-scale supercomputers are needed [184, 185].

To overcome problem (iii), several modifications of SR have been proposed. Among them are iterative solvers such as MINRES [186, 25] which avoid this scaling by iteratively computing the pseudo-inverse of  $S$  [25]. However, for large  $N_s$  and  $N_\theta$  typically also the required number of iterations grows. Another method is the sequential local optimization approach, in which SR only optimizes a portion of all parameters to reduce the time cost [187]. Other recent works have proposed modifications of the SR update rule which involve the inversion of a  $N_s \times N_s$  matrix instead of  $S$ , with  $N_s$  the number of samples to estimate the gradient, which is usually smaller than the number of parameters [141, 93]. Apart from that, the performance of SR can be improved with adaptive learning rate solvers, such as the second order Runge Kutta integrator, allowing for an optimal choice of the learning rate [188].

The optimization of NQS can become very difficult due to the in general very rugged and chaotic optimization landscape with many local minima [188, 182, 189]. To overcome this problem, many techniques have been developed. Among them are *variational neural annealing* that applies an artificial temperature to avoid getting stuck in local minima [119, 124, 190], the application of *symmetries* [83, 123, 117, 45, 99] to enforce a training only in the target symmetry sector, *transfer learning*, i.e. the transfer learned properties of small systems to larger system sizes, [191, 118, 192], and *weight pruning* [160, 161].

Further improvement can be achieved using complementary optimization methods: Firstly, the NQS can be pretrained with external data, see Sec. 3.6. Second, the energy resulting from the VMC optimization can be improved by applying a few *Lanczos* steps after optimizing the network parameters using the techniques described above [193]. Applying the Hamiltonian in the Lanczos algorithm to obtain the next Krylov vector corresponds to minimizing the infidelity to the corresponding state and thus necessitates a separate training, rendering typically only very few Krylov vectors accessible. Similar in spirit to Lanczos algorithms, power methods can be used to find the ground states of (gapped) Hamiltonians [194].

A related approach, also based on applying imaginary time evolution, maximizes the fidelity to a target state at each iteration. In contrast to SR, this is done explicitly, i.e. a small imaginary time step  $\Delta\tau$  has to be applied to the current state using e.g. the Euler [195] or the Heun method [183]. In Ref. [196], the general idea of minimizing the difference between the current parameterization and an explicitly improved wave function is introduced as supervised wave function optimization (SWO). The improved wave function, which constitutes the target in this optimization, can e.g. be obtained through power methods or imaginary time evolution.

Lastly, in Ref. [197] an optimization scheme based on stochastic representations of wave functions is proposed. In this representation, not the configurations, here in terms of particle positions  $\{\mathbf{R}_i\}_{i=1,\dots,N_s}$  of a continuous system, but a set of  $N_s$  samples  $(\mathbf{R}_i, \psi_s(\mathbf{R}_i))$  is used for the optimization. The NQS is given by

$$\psi_s(\mathbf{R}_i) = e^{-\delta\tau\hat{H}}\hat{P}_{s/a}\hat{\psi}_\theta(\mathbf{R})|_{\mathbf{R}=\mathbf{R}_i}, \quad (33)$$

with a variational function  $\psi_\theta$  parameterized by a FFNN and stochastic projection  $\hat{P}_{s/a}$  onto the symmetric or antisymmetric subspace. In order to train the NQS, first a set of samples  $(\mathbf{R}_i, \psi_s(\mathbf{R}_i))$  is generated and projected onto the target subspace. Then, simple regression is applied, with the goal of minimizing the sum of squared residuals between the projected samples and  $\hat{P}_{s/a}\psi_\theta(\mathbf{R}_i)$ . The up-

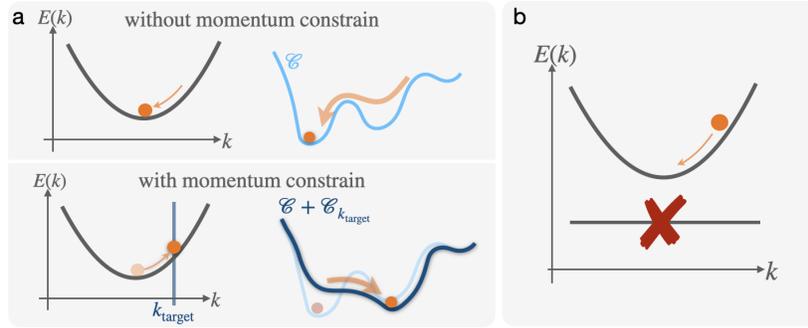


Figure 10: Excited States: a) Optimization for different momentum sectors. Figure adapted from Ref. [35]. b) Low-energy states in the same symmetry by enforcing orthogonality to the lower excited states and the ground state as in Refs. [54, 74].

dated trial function  $\hat{P}_{s/a}\psi_{\theta}$  is then used to generate new sample coordinates  $\mathbf{R}'_i$ , and imaginary time evolution is performed on  $\hat{P}_{s/a}\psi_{\theta}(\mathbf{R})|_{\mathbf{R}=\mathbf{R}'_i}$ . In contrast to VMC, this method does not require that the samples are distributed according to the wave functions' amplitudes. Furthermore, no evaluation of the energy or its gradients is required.

One reason why the optimization is so challenging is the intricate interplay between phase and amplitude parts during the optimization. In some cases, the optimization outcomes are improved by imposing a certain sign structure of the target state, e.g. the Marshall sign rule for the Heisenberg model (restricted to bipartite lattices) [198, 118, 117]. Furthermore, the interplay between phase and amplitude during the optimization can be investigated by considering the partial optimization problem. In Refs. [188, 35], either the exact phases or the exact amplitudes are set and kept constant during the training, such that only amplitude *or* phase, respectively, have to be learned. In both works, considering the  $J_1 - J_2$  model and the bosonic and fermionic  $t - J$  model, the authors find that none of the two optimization strategies can systematically improve the ground state representation, and hence conclude that the interplay between phase and amplitude seems to play a crucial role for the optimization.

### 3.2. Excited States

The methods for calculating excited states using NQS fall into two categories, depending on the type of excited states that are targeted: (i) Lowest energy states in a different symmetry sector than the ground state, e.g. different momentum or magnetization sectors. (ii) Low-energy states in the same symmetry sector as the ground state.

The former usually rely on the same VMC scheme, but with a restriction to the targeted symmetry sector implemented in the wave function or the training loss [54, 35, 79, 157, 199]. This enables e.g. the calculation of quasiparticle dispersions from NQS [54, 35, 79, 74]. For example, in Ref. [35], specific momentum  $\mathbf{k}_{\text{target}}$  sectors are targeted by adding a mean square error between  $\mathbf{k}_{\text{target}}$  and  $\mathbf{k}_{\text{NQS}}$ , see Fig. 10a. Hereby, the momentum of the NQS is given by

$$\mathbf{k}_{\text{NQS}}^{\mu} = \frac{i}{a} \log \langle \psi_{\theta} | \hat{T}_{\mathbf{e}_{\mu}} | \psi_{\theta} \rangle, \quad (34)$$

with the translation operator  $\hat{T}_{\mathbf{e}_{\mu}}$  in direction of the unit vector  $\mathbf{e}_{\mu}$  by a lattice constant  $a$ . In contrast to MPS calculations, this global operator can be evaluated at low computational cost. Another approach to target excited states in different symmetry sectors consists in the use of group convolutional neural networks, as discussed in Section 2.3.

For the second category, Fig. 10b, usually more modifications need to be made. In Refs. [54, 74] excited states are targeted by enforcing orthogonality to the ground state  $\psi_0$  or lower lying excited states  $\psi_i$ . In Ref. [74], the first excited state  $\psi_{1,\theta}$  is calculated by adding the normalized overlap with

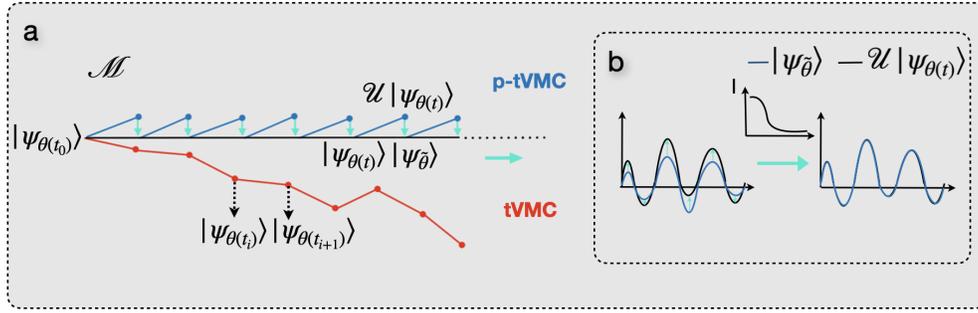


Figure 11: a) Adapted visualisation from Ref. [201] of the p-tVMC method. In the cases when the tVMC evolution of a state  $\psi$  breaks down, p-tVMC can resolve this problem by projecting the exactly evolved state onto the variational manifold  $\mathcal{M}$ . The tVMC evolution of the state (red) starts to break down whereas the p-tVMC method projects the exactly evolved state  $U|\psi_\theta\rangle$  back onto the variational manifold  $\mathcal{M}$  (blue) at each timestep. b) To project the exactly evolved state onto the variational wave function, the distance in the Hilbert space needs to be minimized, resulting in an optimization problem of the infidelity.

the ground state,

$$\frac{\langle \psi_{0,\theta'} | \psi_{1,\theta} \rangle}{\langle \psi_{0,\theta'} | \psi_{0,\theta'} \rangle \langle \psi_{1,\theta} | \psi_{1,\theta} \rangle}, \quad (35)$$

to the cost function. In Ref. [54], the excited state is defined as

$$\psi_1 := \Phi_\theta - \lambda \psi_{0,\theta'}. \quad (36)$$

To enforce orthogonality,  $\langle \psi_1 | \psi_{0,\theta'} \rangle = 0$ , they use  $\lambda = \langle \Phi_\theta | \psi_{0,\theta'} \rangle$ . Both methods require the explicit representation of the ground state  $\psi_{0,\theta'}$ , rendering the calculation of higher excited states computationally demanding.

Another method, requiring no explicit orthogonalization of the different states, transforms the problem of finding the  $K$  lowest excited states of a given system into that of finding the ground state of an expanded system given by all targeted excited states [200]. The ansatz for the expanded ground state is written as

$$\Psi(\mathbf{x}) := \det \begin{pmatrix} \psi_1(\mathbf{x}^1) & \dots & \psi_K(\mathbf{x}^1) \\ \vdots & & \vdots \\ \psi_1(\mathbf{x}^K) & \dots & \psi_K(\mathbf{x}^K) \end{pmatrix}, \quad (37)$$

i.e. an unnormalized Slater determinant of many-particle wave functions  $\psi_i$  instead of single-particle orbitals. Here,  $\mathbf{x}^i$  denotes a set of  $N$  particles  $x_1^i, \dots, x_N^i$ . The Hamiltonian is correspondingly expanded to act on all  $K$  particle sets, and VMC is performed to find the ground state of the expanded system. Subsequently, energies, expectation values, and overlaps in the excited states can be retrieved from  $\Psi(\mathbf{x})$ .

Finally, in Ref. [156] the authors propose a method based on the assumption that one-particle excitations dominate the low-lying spectrum, allowing to construct the excited states by single-particle excitations on top of the ground state.

### 3.3. Dynamics

Neural network quantum states are furthermore capable of describing time-dependent systems. The quantum dynamics can be obtained using time-dependent network weights  $\theta(t)$  [25, 15]. Numerically exact results for timescales comparable to or exceeding the capabilities of TN algorithms have been obtained for the paradigmatic two-dimensional transverse-field Ising model [87]. However, even though NQS are able to capture strongly entangled states, the required number of parameters needed to represent a quantum state after a global quench can grow exponentially in time, according to Ref. [202].

This can potentially render the use of NQS to represent the dynamics of a state inefficient. Dynamically increasing the network size and choosing a network architecture incorporating the symmetry of the state are promising approaches to overcome this problem.

Following the Dirac-Frenkel Variational principle, the time derivative of the neural network weights are to be optimized such that the variational residuals

$$R(\dot{\boldsymbol{\theta}}(t)) = \text{dist}(\partial_t \psi_{\tilde{\boldsymbol{\theta}}}, -i\hat{H}\psi_{\boldsymbol{\theta}}) \quad (38)$$

are minimized [25, 203]. This is achieved within the stochastic approach using time-dependent Variational Monte Carlo method (t-VMC). In most cases t-VMC is used in combination with stochastic reconfiguration, see Sec. 3.1. The iteration scheme to find the ground state energy can be interpreted as an effective imaginary time evolution, such that an iteration scheme to approximate the real-time evolution of the quantum spin system can be derived in an analogous way [86]. At each time step,  $|\psi_{\boldsymbol{\theta}(t)}|^2$  is sampled and the variational residual is evaluated. Minimization of the variational residuals with respect to  $\dot{\boldsymbol{\theta}}$  gives a first order differential equation for the weights  $\boldsymbol{\theta}(t)$  [87]. The final equation to be solved for the variational parameters  $\boldsymbol{\theta}$  is then given by the time-dependent variational principle (TDVP) equation:

$$S(t)\dot{\boldsymbol{\theta}}(t) = -iF(t), \quad (39)$$

with the covariance matrix  $S$  and the vector of forces  $F$  are the same as in section 3.1 .

To solve this equation for the variational parameters, the covariance matrix needs to be inverted. Since this matrix can be non-invertible,  $S^{-1}$  denotes the Moore-Penrose pseudo-inverse. This inversion leads to various problems, see also Sec. 3.1. Moreover, the unstable Moore-Penrose pseudo-inverse of the  $S$  matrix requires choosing a right cut-off tolerance for small singular values. In practice, one finds that the chosen cut-off for the pseudo-inverse of  $S$  can alter t-VMC. Krylov subspace methods (i.e. the conjugate gradient method or MINRES algorithm) avoid this sensitivity problem, but are not always converging [203]. When calculating the dynamics of a system, this error accumulates with each time step. [203, 183]

Regularization schemes for  $S$  have been developed, see Sec. 3.1, but these still require a very large number of samples [203] and impact the accuracy [183]. The choice of regularization method impacts the stability of the dynamics [88]. In Ref. [87], this problem is approached by disregarding the contributions to the TDVP of which there is insufficient information available due to constraints of a finite number of samples  $N_{MC}$ . This still requires the diagonalization of  $S$ , which has a high computational cost ( $O(N_{\boldsymbol{\theta}}^3)$ ) [203].

Successful applications of t-VMC have been obtained for system sizes up to  $N = 256$  spins using an RBM and for time scales up till 10s in Refs. [25, 204, 89]. In Ref. [90], an RBM ansatz is used to capture the dynamics of the 2D Heisenberg model, predicting strong magnon-magnon interactions leading to supermagnonic propagation. In Ref. [95] the dynamics of the 2D transverse-field Ising model around the quantum phase transition has been studied with t-VMC, leading to insights in the quantum Kibble-Zurek mechanism. Numerical simulations obtained with t-VMC have been used to validate experimental data interpretation, as is done in Ref. [130] for a network description of wave function snapshots. However, accessing long times via t-VMC stays challenging. The stability of t-VMC strongly depends on the chosen variational ansatz [183], and is affected by systematic statistical bias or an exponential sample complexity when the wave-function contains zeros. This is also the case for ground state calculations, but is less harmful due to the accumulation of error that affects dynamics [201].

A new method is proposed to circumvent these issues in Refs. [183] and [201]: Projected time-dependent Variational Monte Carlo (p-tVMC). The scheme consists of casting a Runge-Kutta integration scheme into minimizing a variational distance at each time step. Starting again from the Dirac-Frenkel Variational Principle, an  $s$ -order Runge-Kutta approximant is to be used, instead of a first order expansion of the time propagator [183]. In practice, a metric based upon infidelity can now be used, in contrast to the normally used Fubini-Study distance. This infidelity can be estimated through Monte Carlo sampling and should be considered the distance in the Hilbert space between

$U|\psi_{\theta}\rangle$  and  $|\psi_{\hat{\theta}}\rangle$ , which is to be optimized [201]:

$$\min_{\hat{\theta}} I(|\psi_{\hat{\theta}}\rangle, U|\psi_{\theta}\rangle). \quad (40)$$

No expansion of  $|\psi_{\hat{\theta}}\rangle$  with respect to the variational parameters is necessary for p-tVMC, opposed to tVMC where a second order expansion leads to Eq. (39). However, the unitary time evolution  $e^{-i\hat{H}\delta t}$  needs to be decomposed by the Trotter-Suzuki decomposition, leading to a scaling of the number of required samples with the Trotter order. This is costly and breaks translational invariance. Alternatively, an expansion of the time propagator into its Taylor decomposition is needed, typically up to second order in  $\hat{H}$  in order to evaluate long time dynamics. This results in a quadratic increase of the connected matrix elements that need to be computed, and thereby the computational cost. Furthermore, for higher  $s$ -order integration schemes, the computational cost scales with the system size  $N$  as  $N^s$  [183]. In Ref. [201] the use of an RBM avoids this issue by computing the off-diagonal elements in the transverse field Ising model exactly. Because of this, the p-tVMC method only scales linearly with the number of parameters, which makes this a promising method to compute dynamics for large neural network architectures. As opposed to an update rule for the network parameters  $\theta$ , standard gradient descent based techniques to minimize the infidelity can be used. Since p-tVMC is not affected by biases or vanishing SNR, it can simulate dynamics in cases where t-VMC fails or is inefficient [201].

An alternative approach, using the implicit midpoint method, has been explored in Ref. [203]. Here, the network parameters are optimized to minimize the error between the state at the next time step  $t + \Delta t$  and the discrete flow of the implicit midpoint method applied to the Schrödinger equation. This has shown some advantages in preserving the symplectic form of Hamiltonian dynamics while not complicating the network optimization with intermediate quantities [203].

In Ref. [205], a dynamical strong disorder renormalization group approach is used to map the quantum dynamics of a disordered spin chain onto a quantum circuit generated by local unitaries. These local unitaries are applied to the NQS in a supervised scheme, similar to the SWO discussed in section 3.1 and the infidelity minimization in p-tVMC.

Other methods to calculate the dynamics of a system consist of training with time evolved states that have been exactly calculated using ED, such that the evolution of new initial states can be predicted by a neural network without evolving the wave function explicitly with the Hamiltonian [206]. To speed up the simulation of the dynamics of many-body systems, hybrid methods are used such as using neural quantum states with calculations on quantum devices to determine expectation values with high computational cost [204]. Another approach are *variational classical networks* [207, 208, 209], i.e. efficient and perturbatively controlled representations of (time-evolved) wave functions in terms of classical spins, where the latter are used to construct a NQS representation of the state under consideration.

*3.3.1. Spectral Functions* Most of the work discussed so far focus on global quenches. Time-dependent NQS can however also be used to simulate local quenches, such as the response of a system to a local perturbation, relevant for spectral functions. In Ref. [94], the dynamical spin structure factor of different two-dimensional quantum Ising models is calculated by applying t-VMC following a local perturbation (application of  $\hat{S}^z$  operator) on top of the ground state represented by a convolutional neural network. Subsequent Fourier transformation yields the momentum- and frequency-resolved structure factor. A complementary approach is demonstrated in Ref. [210]. Here, the dynamical structure factor of the one- and two-dimensional Heisenberg model is calculated explicitly using a Chebyshev expansion, where the corresponding wave functions are represented as RBMs. In Ref. [211], the Green's function

$$G_{ij}(z) = \langle \psi | \hat{A}_i^\dagger \frac{1}{z - \hat{H}} \hat{A}_j | \psi \rangle \quad (41)$$

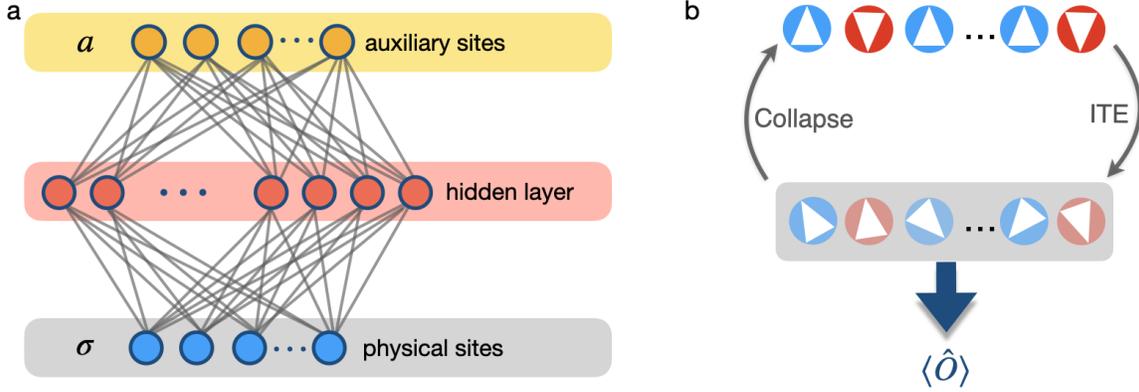


Figure 12: Finite temperature state simulations: a) In the purification method, ancilla sites  $a$  are introduced in addition to the physical sites  $\sigma$ , see Eq. (44). The mixed state of the system is obtained by integrating out the auxiliary sites. b) METTS: Starting from a pure infinite temperature state (top), the state is evolved in imaginary time (ITE) to produce a METT state (bottom). Next, a projective measurement (collapse) to a random basis state is done in order to obtain a new pure state. The expectation value of desired observables is calculated through the ensemble of METT states.

is directly calculated by an extension of the stochastic reconfiguration approach to obtain the *correction vector*

$$|\chi_j(z)\rangle = \frac{1}{z - \hat{H}} \hat{A}_j |\psi\rangle, \quad (42)$$

where the corresponding ground state  $|\psi\rangle$  of the system has been obtained beforehand using SR.

### 3.4. Finite Temperature States

In many experimentally relevant situations, we are dealing with quantum many-body systems at a finite temperature, and in order to compute thermodynamics properties of the system one needs to work with the thermal density matrix

$$\hat{\rho} = \frac{1}{Z} e^{-\beta \hat{H}}. \quad (43)$$

Where  $\hat{H}$  is the Hamiltonian of the system,  $\beta = \frac{1}{k_B T}$  is the inverse temperature, and  $k_B$  is the Boltzmann constant. Hence, the task boils down to evaluating this density matrix efficiently. One approach that has been developed and is commonly applied in the context of MPS is the idea of purification, also known as the thermofield approach [131, 2, 212]. In the purification method, an additional auxiliary site is introduced for each physical site of the system, known as an ancilla. As a result, one deals with a pure instead of a mixed state, where said pure state lives in a higher dimensional Hilbert space. The algorithm then starts from an infinite temperature state, followed by imaginary time evolution to cool down the system to the desired temperature, where imaginary time  $\tau$  here denotes the inverse temperature  $\beta$ . The desired thermal density matrix is then obtained by tracing out the auxiliary degrees of freedom  $a$ , i.e.

$$\hat{\rho}(\sigma, \sigma') = \sum_a \langle \sigma, a | \psi \rangle \langle \psi | \sigma', a \rangle. \quad (44)$$

In the context of neural network quantum states, one approach to purification is through a modified RBM, see Fig. 12a. A similar type of architecture was also used in Ref. [213] to reconstruct mixed states. Refs. [214, 181] employ the purification method to obtain finite temperature expectation values for a Heisenberg chain and a  $6 \times 6$   $J_1 - J_2$  model. On top of the imaginary time evolution, Ref. [215] deals with real-time evolution, leveraging an RNN architecture.

Another promising approach also developed in the context of tensor networks is the idea of minimally entangled typical thermal states (METTS) [216, 217]. METTS is designed to efficiently

sample from the thermal ensemble instead of dealing with the full complexity of the mixed state directly. The idea is to construct an ensemble of pure states, which provides a good approximation of the thermal equilibrium state. Concretely, the trace in the evaluation of finite temperature expectation values can be expanded in terms of an orthonormal basis as

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho} \hat{O}) = \frac{1}{Z} \sum_i \langle \sigma_i | e^{-\beta \hat{H}/2} \hat{O} e^{-\beta \hat{H}/2} | \sigma_i \rangle = \frac{1}{Z} \sum_i P(i) \langle \psi_{\sigma_i}(\beta) | \hat{O} | \psi_{\sigma_i}(\beta) \rangle \quad (45)$$

To this end, one starts from a pure product state  $|\sigma_0\rangle$ . This product state is evolved in imaginary time to generate a state  $|\psi_{\sigma_i}(\beta)\rangle = e^{-\beta \hat{H}/2} |\sigma_i\rangle$ . This procedure gives us a so-called METTS state  $|\psi_{\sigma}(\beta)\rangle$ . After this step, a projective measurement in the computational basis (collapse) is performed in order to produce a new pure state  $|\sigma_1\rangle$  to start over with imaginary time evolution. This procedure of sampling the states  $|\sigma_i\rangle$  ensures that the resulting states represent the thermal ensemble accurately [217]. At the end, we have a set of states  $\{|\psi_{\sigma_0}(\beta)\rangle, |\psi_{\sigma_1}(\beta)\rangle, \dots, |\psi_{\sigma_n}(\beta)\rangle\}$  from which the thermal average of a given operator  $\hat{O}$  can be estimated as:

$$\langle \hat{O} \rangle_{\beta} = \frac{1}{N_s} \sum_{i=1}^n \langle \psi_{\sigma_i}(\beta) | \hat{O} | \psi_{\sigma_i}(\beta) \rangle, \quad (46)$$

where  $N_s$  is the number of METTS state samples. In Refs. [218, 181], the product states  $|\sigma_0\rangle$  are prepared by adjusting the parameters of an RBM correspondingly.

For the imaginary time evolution employed both in the purification and the METTS algorithm, the following equation must be solved:

$$\frac{\partial}{\partial \beta} |\psi(\beta)\rangle = -\frac{1}{2} \hat{H} |\psi(\beta)\rangle, \quad (47)$$

where the new wave function after each imaginary time step  $\delta\tau$  must stay within the variational manifold. This constraint can lead to a modification of  $\beta$ .

Another approach to simulate finite temperature states is based on quantum typicality [181] which utilizes the concept that a single pure state can accurately reproduce the expectation values of an observable in the Gibbs ensemble for large systems. This method approximates an infinite temperature state using a combination of a pair product (PP) wave function  $\Phi_{PP}$  and a neural network component  $\psi_{\theta}$ , i.e.

$$\Psi(\sigma) = \psi_{\theta}(\sigma) \Phi_{PP}(\sigma). \quad (48)$$

Pair product wave functions can model electron interactions within the system, including the prohibition of double occupancy through the use of the Gutzwiller projection. The typical state is then evolved in imaginary time to simulate finite temperatures.

In Ref. [219], a CNN with two input channels  $\sigma$  and  $\sigma'$  is used to represent a mixed state  $\hat{\rho}(\sigma, \sigma')$  of a one-dimensional bosonic system. Starting from an infinite temperature state, imaginary time evolution is performed, such that the output of the network is the corresponding matrix elements of density matrix at the desired temperature. As opposed to e.g. purification, this approach does not guarantee the hermiticity and positive definiteness of the density matrix.

In contrast to the works discussed so far, which all use imaginary time evolution, a recent paper [220] instead minimizes a modified free energy. Here, the von-Neumann entropy is replaced by the second Rényi entropy, which can be evaluated fairly efficiently. The optimization of neural network parameters is guided by the goal of minimizing this approximation to the free energy.

### 3.5. Open Systems

The state of an open quantum system is described by its density operator  $\hat{\rho}$ . This makes the simulation of open systems even more challenging than for closed systems, since for density matrices, the curse of dimensionality is even more pronounced as for wave functions, e.g. for a system of  $N$  spin-1/2 particles the number of coefficients to parameterize  $\hat{\rho}$  scales as  $4^N$  coefficients [129]. The dynamics of open quantum systems is governed by the Lindblad master equation,

$$\dot{\hat{\rho}} = -i \left[ \hat{H}, \hat{\rho} \right] + \sum_i \gamma_i \left( (\hat{L}^i)^\dagger \hat{\rho} \hat{L}^i - \frac{1}{2} \{ (\hat{L}^i)^\dagger \hat{L}^i, \hat{\rho} \} \right) \quad (49)$$

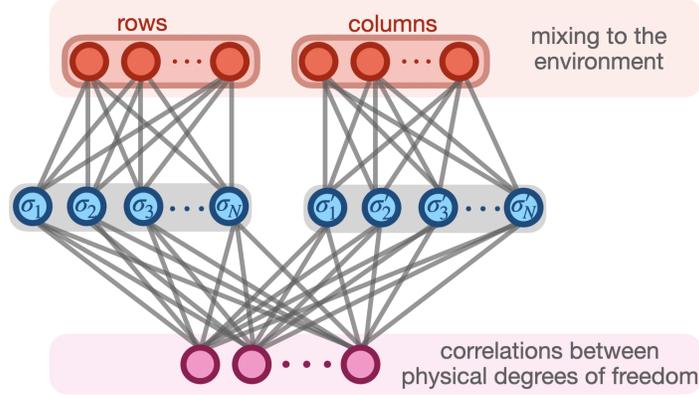


Figure 13: Neural density operator based on a RBM as used in Ref. [221]: The state of the physical system is encoded by the visible (blue) layer, with correlations encoded in the bottom hidden layer (purple) and mixing due to the environment encoded in the top hidden layer (red). The left represents row, the right part column indices of the density matrix  $\hat{\rho}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ .

where  $[\dots](\{\dots\})$  denote (anti-)commutators and  $\hat{\mathcal{L}}$  are so-called jump operators. The first term describes the unitary dynamics of the system given by  $\hat{H}$ , the second the non-unitary dynamics due to the dissipation to the environment with strength  $\gamma$ . Eq. (49) can also be expressed as

$$\frac{d}{dt}\hat{\rho} = \hat{\mathcal{L}}\hat{\rho} \quad (50)$$

with the Liouvillian  $\hat{\mathcal{L}}$ .

In most cases, the second form is used to determine the solution  $\hat{\rho}$  using NQS. In order to do so, a neural representation of the density operator is needed. This is typically realized by (i) using positive operator valued measures (POVMs) [125] or (ii) introducing additional nodes that encode the mixing to the environment [213, 221]. We would like to point out that the works discussed here in the context of (i) and (ii) are inherently different from machine learning approaches to open quantum systems that e.g. aim learn a parameterization of  $\hat{\mathcal{L}}$  and not  $\hat{\rho}$  as e.g. in Refs. [222, 223].

In the POVM approach [125], the density matrix is represented by a probability distribution over measurement outcomes  $a$  of an informationally complete (IC) set of measurement operators  $\hat{M}_a$ , inspired from Born's rule

$$P_{\boldsymbol{\theta}}(a) = \text{Tr}(\hat{M}_a \hat{\rho}_{\text{POVM}}). \quad (51)$$

This leads to the definition

$$\hat{\rho}_{\text{POVM}} = \sum_{a,a'} P_{\boldsymbol{\theta}}(a) T_{a,a'}^{-1} \hat{M}_{a'}, \quad (52)$$

with the overlap matrix  $T_{a,a'} = \text{Tr} \hat{M}_a \hat{M}_{a'}$  and different possible choices of  $\hat{M}$ . The advantage of the IC-POVM representation is that in Eq. (52) only the positive amplitudes  $P_{\boldsymbol{\theta}}(a)$  have to be modeled by a neural network. However,  $\hat{\rho}_{\text{POVM}}$  is, in general, not a positive-definite matrix. This problem does not occur in the purification ansatz [224].

The second approach is e.g. taken in Ref. [221], where an RBM with an additional hidden layer is used, and both visible and hidden state representations are split into two representations for rows and columns of the density matrix  $\hat{\rho}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ , see Fig. 13. Other neural density operators exist, e.g. in terms of CNNs [225, 226] or in form of an autoregressive network, see Ref. [224]. In the latter work, the density matrix is defined as

$$\hat{\rho}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \prod_{i=1}^N \sum_{a=1}^R \psi_{\boldsymbol{\sigma}_{\leq i}, a} (\psi_{\boldsymbol{\sigma}'_{\leq i}, a})^*, \quad (53)$$

with ancillas  $a$  and neural network representations of  $\psi_{\sigma_{\leq i}, a}$ . This is an example of the purification approach discussed in Sec. 3.4. In contrast to the POVM approach, this purification via the ancilla nodes in Eq. (53) makes the density matrix positive semi-definite. Furthermore, each factor  $\sum_{a=1}^R \psi_{\sigma_{\leq i}, a} (\psi_{\sigma'_{\leq i}, a})^*$  in Eq. (53) can be normalized, making the neural network representation of  $\hat{\rho}$  autoregressive.

With these ansätze (i) and (ii), the solution of Eq. (50) can be obtained using different approaches:

*Time dependent solution of the Lindblad equation:* Eq. (50) can be solved directly by minimizing  $\|\frac{d}{dt}\hat{\rho} - \hat{\mathcal{L}}\hat{\rho}\|$  using SR, where  $\|\dots\|$  can e.g. be taken to be the Fubini-Study distance or the trace norm [221, 227]. This is done e.g. in Ref. [221] using an RBM with additional nodes to simulate a 1D anisotropic Heisenberg model or in Ref. [227] using a deep (quantum) FFNN for a dissipative 1D TFIM and 2D  $J_1 - J_2$  spin systems. For systems that lack translational invariance, more elaborate sampling and optimization procedures are necessary [228, 229]. In Refs. [129, 230, 128] the POVM ansatz implemented with autoregressive networks is used for the time dependent solution of 1D and 2D dissipative Heisenberg models and prototypical states from quantum computing. In order to do so, the time evolution has to be represented in the stochastic representation (52), i.e. an operator  $\hat{O}$  is calculated that time evolves  $P_{\theta_t}(a)$ . Then, the parameters  $\theta_{t+1}$  are selected such that the distance between  $\hat{O}P_{\theta_t}(a)$  and  $P_{\theta_{t+1}}(a)$  is minimal. In Refs. [230, 128], this distance is calculated explicitly, e.g. in Ref. [230] the the Kullback-Leibler divergence  $D_{\text{KL}}(\hat{O}P_{\theta_t}(a)|P_{\theta_{t+1}}(a))$ , see Eq. (55), is minimized. In Ref. [128], the network parameters are optimized to minimize the error between the new, time evolved state and the target state given by the discrete flow according to a second-order forward-backward trapezoid method applied to the Lindblad equation.

In Ref. [129], the distance is measured by the Kullback-Leibler or the Hellinger distance, but in this work the distance metrics are expanded around small times, leading to the the time dependent variational principle update, see Eq. (32), for  $P_{\theta_t}(a)$ . This reduces the sampling cost and makes the optimization problem convex.

*Steady states:* Other works use the fact that stationary states  $\hat{\rho}_{ss}$  in open systems fulfill

$$\frac{d}{dt}\hat{\rho}_{ss} = \hat{\mathcal{L}}\hat{\rho}_{ss} = 0. \quad (54)$$

The neural density operator is trained to fulfill this condition by minimizing e.g. the expectation value of  $\hat{\mathcal{L}}$  [231, 232] or the  $L_2$ -norm [233, 226]. Furthermore,  $\hat{\mathcal{L}}^\dagger$  can be applied from the left to Eq. (54), yielding an optimization problem of  $\hat{\mathcal{L}}^\dagger\hat{\mathcal{L}}$  instead of  $\hat{\mathcal{L}}$ , with the advantage that the former operator is hermitean and hence has a real spectrum [234]. In Ref. [232], the dynamics of a 2D dissipative XYZ spin model is simulated using an RBM with additional nodes. Refs. [234, 233] consider 2D transverse-field Ising models and other similar spin systems.

### 3.6. Learning from Data

*3.6.1. Quantum State Tomography* Quantum state tomography (QST), i.e. the reconstruction of a quantum state from measurement data, plays a crucial role for the characterization and verification of quantum devices [235]. For example, it can be applied to compare the experimentally prepared state against the target state to estimate the error of the quantum device under consideration, see Fig. 14. Furthermore, QST enables the evaluation of complex observables that would not be accessible directly from experiments [236].

Full QST relies on two assumptions: (i) Since typically several measurements are needed to infer the quantum state, it is assumed that identical copies of the state can be prepared from which the measurements can be taken. (ii) The set of measurements, described by positive operator valued measures, is informationally complete and hence the probability distribution over measurement outcomes uniquely determine the quantum state via Born's rule. Since these conditions are not fulfilled in most cases, approximate QST schemes are necessary.

Conventional methods for QST, such as linear inversion and maximum likelihood estimation [237, 238], are based on inverting Born's rule and hence suffer from an exponential scaling with the

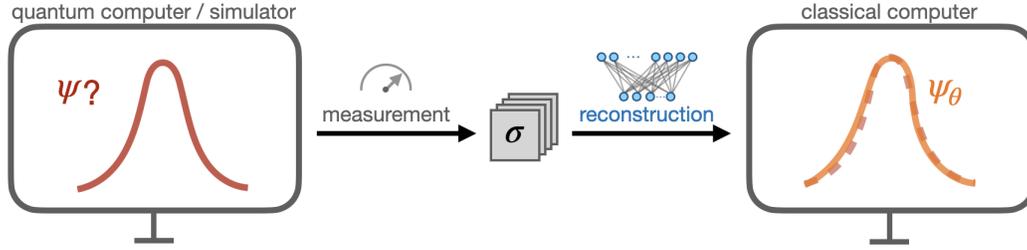


Figure 14: Quantum state tomography: Reconstruction of an unknown quantum state  $\psi$  on a quantum device, by taking measurements  $\sigma$  and reconstruction with a neural network. The network is trained such that the amplitudes of the reconstructed state  $\psi_\theta$  are as close as possible to the real measurement distribution generated from  $\psi$ , see Eq. (55).

system size, resulting from an exponential growth of both the sampling complexity and the number of parameters needed to represent the state. Under these aspects, machine learning techniques have enormous potential for QST: Firstly, machine learning models can learn the structure of a state under consideration, i.e. symmetries or correlations, allowing them to efficiently represent typical physical states with a reduced number of parameters [40]. Furthermore, they have the ability to generalize from an incomplete dataset, tackling the exponential scaling of the sample complexity [239]. In Ref. [240], the authors show that a simple FFNN can outperform conventional methods both in terms of reconstruction time and quality.

The potential of neural quantum states for QST has been explored for various pure and mixed quantum states. One of the first works reconstructs finite temperature states of the 1D and 2D Ising model using real-valued RBMs [241]. Furthermore, highly entangled states with more than a hundred qubits are reconstructed in Ref. [61] using a complex RBM. In these works, the NQS is trained by minimizing the Kullback-Leibler divergence between the measurement distribution  $q$  and the NQS amplitudes  $p_\theta$ ,

$$D_{\text{KL}}(q|p_\theta) = \sum_{\sigma} q(\sigma) \log \left( \frac{q(\sigma)}{p_\theta(\sigma)} \right) \quad (55)$$

on the underlying dataset  $\mathcal{D}$  with measurements  $\sigma \in \mathcal{D}$ . In Ref. [242], instead of  $D_{\text{KL}}$  the classical shadow formalism (see below) is used to approximate the infidelity between target and reconstructed state.

This procedure assumes pure quantum states, which is typically not the case in experimental settings. In some cases, pure representations can nevertheless be used to approximately represent the states under consideration and effect of measurement errors in the training data can be mitigated by modifications of the NQS architecture, as shown in Ref. [243] for data from 1D Rydberg tweezer arrays using an RBM with an additional noise layer. However, in many cases, the reconstruction of the full density matrix is needed. Requirements on the density matrix, such as positivity, can be enforced using a purification scheme with additional latent units [213]. Other works considering mixed state representations involve generative models, neural density operators and POVM representations [125, 244, 245], see Sec. 3.5. In Ref. [246], an iterative scheme to promote any pure state reconstruction to a mixed state reconstruction is proposed. The reconstruction performance can be increased by filtering the experimental data [247] or constraining the density matrix, e.g. to positivity or global symmetries, improve the performance, in particular in the presence of measurement imperfections [248, 123].

To further reduce the amount of measurement data needed for the reconstruction, an efficient evaluation of the typically incomplete set of measurements is needed. In Ref. [249], the authors present an efficient method for constructing approximate classical descriptions of quantum states from very few measurements, so-called classical shadows, with an information-theoretic bound for the precision of estimated expectation values. The procedure relies on randomly selecting unitaries  $\hat{U}$  from an

ensemble of particular unitaries that allow one to calculate

$$\mathcal{M}(\hat{\rho}) = \left\langle \hat{U}^\dagger |\sigma\rangle \langle \sigma| \hat{U} \right\rangle_\sigma \quad (56)$$

from measurements in the computational basis  $|\sigma\rangle$ . The density matrix  $\hat{\rho}$  can be approximated by the so-called classical shadow  $\mathcal{S}$  consisting of  $N_s$  samples, with

$$\mathcal{S} = \left\{ \mathcal{M}^{-1} \left( \left\langle \hat{U}_i^\dagger |\sigma\rangle_i \langle \sigma|_i \hat{U}_i \right\rangle_{\sigma_i} \right) \quad \text{with } i \in 1, \dots, N_s \right\}. \quad (57)$$

Further works consider the effect of local measurements [250] or an efficient choice of measurement configurations [251, 199]. Hereby, adaptive schemes, that incorporate the knowledge gained from previous measurements to propose the next measurement configuration, are of great interest [252, 199]. Moreover, NQS can be pretrained with artificial data before the measurements to enhance the reconstruction [253].

Also in the setting of QST, the choice of NQS depends on the state under consideration, with potential advantages e.g. of autoregressive networks and their perfect sampling [97] or of networks which can represent a high degree of entanglement [136, 254]. Typical architectures are RBMs [241, 61, 255, 37] (see Sec. 2.2), RNNs and transformer networks [136, 254, 123, 37, 256] (see Sec. 2.6), CNNs [97] (see Sec. 2.3), generative adversarial networks [257] and variational autoencoders [112] (see Sec. 2.5).

*3.6.2. Hybrid Training* The ground state search described in section 3.1 typically starts from an NQS with randomly initialized parameters  $\theta_0$ . Assuming convergence of the variational Monte Carlo procedure, the details of this initialization should not matter. However, if the ground state search is challenging, e.g. due to local minima, the question of convergence itself, as well as how many iterations are needed to reach the ground state can crucially depend on the choice of  $\theta_0$ . Using existing data, e.g. from other numerical simulations or an experimental realization, the initial parameters can be chosen such that the ground state search starts from a highly promising region of the parameter space. In this case, the data is used to perform state reconstruction as described in section 3.6.1. Subsequently, the parameters of the same neural quantum state are variationally minimized to find the ground state.

In an experiment, the ground state is typically not perfectly realized. The combination of experimental data from a state close to the ground state with a subsequent numerical ground state search can yield better results than either approach exhibits on its own, as demonstrated for large, interacting two-dimensional Rydberg atom arrays using quantum Monte Carlo data in [258] and using experimental data from large two-dimensional Rydberg atom arrays [259] in [260]. In both cases, a recurrent neural network was used to represent the quantum state. In these examples, the Hamiltonian under consideration is stoquastic, and thus a wave function with real coefficients can represent its ground state. This means in particular that measurements in the computational basis are sufficient for a faithful state reconstruction.

In Ref. [261], molecular Hamiltonians for LiH and H<sub>2</sub>, as well as the one-dimensional lattice Schwinger model are considered. The Hamiltonians under consideration are not stoquastic, and thus measurements in different bases are necessary to faithfully reconstruct their ground states. The computational cost for a single iteration in the reconstruction training is  $2^K$ , where  $K$  is the number of sites on which measurements outside of the computational basis (i.e. in the  $x$ - or  $y$ -basis) are performed. In order to constrain this computational cost, measurements with only a few rotations out of the  $z$ -basis are used for state reconstruction. The ground state is prepared using the variational quantum eigensolver on IBM quantum devices, based on superconducting qubits, as well as using numerical simulations. Again, the results show that this hybrid approach improves the numerical and experimental results to yield lower errors for the ground state energy. Moreover, accurate estimations of more complex observables, such as the entanglement entropy, are directly possible without requiring additional quantum resources.

In Ref. [262] samples from a quantum state produced by the variational quantum eigensolver, a quantum algorithm which generates the ground state on a quantum device, are used. In contrast to the other works discussed in this context, the experimental samples directly replace the samples needed in the VMC, i.e. at the beginning of the training these samples are used to calculate the expectation

values for the VMC optimization instead of samples generated from the trial function. The authors argue that this gives a better estimation of the expectation values at the beginning of the training, speeding up the convergence. After this first stage, the usual VMC algorithm is used. The method is applied for trial wave functions in form of an NQS and the Gutzwiller wave function to find the ground state of the 1D TFIM and Fermi-Hubbard model, respectively.

#### 4. Summary and Outlook

In this review, we discuss neural quantum states (NQS), i.e. variational wave functions that are represented by neural networks. The extensive study of NQS since their proposal in 2017 [25], has revealed that the strengths of neural networks – namely their great expressive power, their ability to compress information very efficiently and their capability to generalize from a given dataset – turn out to be extremely helpful for the simulation of quantum systems:

On the one hand, their expressivity permits the representation of a broad range of quantum states, including a variety of (frustrated) spin systems as well as bosonic and fermionic quantum many-body states in one, two and even three dimensions that we review in this article. The limits of this expressive power are still topic of current research. Second, the efficiency of NQS allows to compress the exponential number of wave function coefficients w.r.t. the system size into a tractable number of network parameters, competitive with state of the art numerical methods. This makes NQS very versatile and applicable in many different contexts in the field of numerical simulation of quantum systems and quantum state tomography. However, finding the targeted state in the huge and complicated optimization landscape with many local minima remains one of the main challenges of the NQS approach, and the optimization depends sensitively on the choice of architecture, hyperparameters and the specific optimization strategies, as we discuss in this review. Advanced training strategies, among them stochastic reconfiguration (SR) which incorporates the geometric structure of the loss landscape, are crucial to overcome this problem. For example, recent results on spin systems obtained with a modified version of SR even start to reach numerical precision in terms of the obtained ground state energies [93]. Furthermore, hybrid approaches, allowing to choose a good starting point of the optimization obtained from training on experimental or numerical data [258, 260] or from the initialization of certain NQS from TNs [49], allow to combine the advantages of existing methods and NQS and are hence promising to overcome this challenge. Finally, the capacity of neural networks to generalize from the training data makes NQS an excellent platform for exploring innovative ideas that go beyond existing methods, such as the creation of toolboxes for simulating entire phase diagrams rather than individual states [135, 56]. Harnessed with these strengths and versatility, neural quantum states offer a new and promising perspective on the challenges posed by simulating quantum many-body systems.

#### *Acknowledgements*

We gratefully acknowledge discussions with Annika Böehler, Fabian Döschl, Fabian Grusdt, Juan Carrasquilla, Lukas Vetter, Markus Schmitt, Moritz Reh, Roeland Wiersema, Roger Melko, Schuyler Moss, Simon Linsel, Stefanie Czischek, and Tizian Blatz. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy—EXC-2111—390814868 and within Project-ID 314695032 – SFB 1277. HL acknowledges support by the International Max Planck Research School.

#### References

- [1] Orús R 2014 *Annals of Physics* **349** 117–158 ISSN 0003-4916 URL <https://www.sciencedirect.com/science/article/pii/S0003491614001596>
- [2] Zvolak M and Vidal G 2004 *Phys. Rev. Lett.* **93**(20) 207205 URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.207205>
- [3] Evenbly G and Vidal G 2014 *Phys. Rev. Lett.* **112**(24) 240502 URL <https://link.aps.org/doi/10.1103/PhysRevLett.112.240502>
- [4] Vidal G 2007 *Phys. Rev. Lett.* **99**(22) 220405 URL <https://link.aps.org/doi/10.1103/PhysRevLett.99.220405>
- [5] Shi Y Y, Duan L M and Vidal G 2006 *Phys. Rev. A* **74**(2) 022320 URL <https://link.aps.org/doi/10.1103/PhysRevA.74.022320>

- [6] Cincio L, Dziarmaga J and Rams M M 2008 *Phys. Rev. Lett.* **100**(24) 240603 URL <https://link.aps.org/doi/10.1103/PhysRevLett.100.240603>
- [7] Verstraete F and Cirac J I 2004 Renormalization algorithms for quantum-many body systems in two and higher dimensions (*Preprint* <https://arxiv.org/abs/cond-mat/0407066>)
- [8] Klümper A, Schadschneider A and Zittartz J 1993 *Europhysics Letters* **24** 293 URL <https://dx.doi.org/10.1209/0295-5075/24/4/010>
- [9] Schollwöck U 2011 *Annals of Physics* **326** 96–192 ISSN 0003-4916 january 2011 Special Issue URL <https://www.sciencedirect.com/science/article/pii/S0003491610001752>
- [10] White S R 1992 *Phys. Rev. Lett.* **69**(19) 2863–2866 URL <https://link.aps.org/doi/10.1103/PhysRevLett.69.2863>
- [11] Schollwöck U 2005 *Rev. Mod. Phys.* **77**(1) 259–315 URL <https://link.aps.org/doi/10.1103/RevModPhys.77.259>
- [12] Hastings M B 2007 *Journal of Statistical Mechanics: Theory and Experiment* **2007** P08024 URL <https://dx.doi.org/10.1088/1742-5468/2007/08/P08024>
- [13] Eisert J, Cramer M and Plenio M B 2010 *Rev. Mod. Phys.* **82**(1) 277–306 URL <https://link.aps.org/doi/10.1103/RevModPhys.82.277>
- [14] Orús R 2019 *Nature Reviews Physics* **1** 538–550 URL <https://doi.org/10.1038/s42254-019-0086-7>
- [15] Becca F and Sorella S 2017 *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press)
- [16] Ceperley D and Alder B 1986 *Science* **231** 555–560 URL <https://www.science.org/doi/abs/10.1126/science.231.4738.555>
- [17] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73**(1) 33–83 URL <https://link.aps.org/doi/10.1103/RevModPhys.73.33>
- [18] Pan G and Meng Z Y 2024 The sign problem in quantum monte carlo simulations *Encyclopedia of Condensed Matter Physics (Second Edition)* ed Chakraborty T (Oxford: Academic Press) pp 879–893 second edition ed ISBN 978-0-323-91408-6 URL <https://www.sciencedirect.com/science/article/pii/B9780323908009000950>
- [19] Troyer M and Wiese U J 2005 *Phys. Rev. Lett.* **94**(17) 170201 URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.170201>
- [20] Carrasquilla J 2020 *Advances in Physics: X* **5** 1797528 (*Preprint* <https://doi.org/10.1080/23746149.2020.1797528>) URL <https://doi.org/10.1080/23746149.2020.1797528>
- [21] Cybenko G 1989 *Mathematics of Control, Signals and Systems* **2** 303–314 URL <https://doi.org/10.1007/BF02551274>
- [22] Hornik K 1991 *Neural Networks* **4** 251–257 ISSN 0893-6080 URL <https://www.sciencedirect.com/science/article/pii/089360809190009T>
- [23] Kim T and Adahl T 2003 *Neural Computation* **15** 1641–1666 ISSN 0899-7667 URL <https://doi.org/10.1162/089976603321891846>
- [24] Le Roux N and Bengio Y 2008 *Neural Comput.* **20** 1631–1649 ISSN 0899-7667 URL <https://doi.org/10.1162/neco.2008.04-07-510>
- [25] Carleo G and Troyer M 2017 *Science* **355** 602–606 URL <https://www.science.org/doi/abs/10.1126/science.aag2302>
- [26] Sharir O, Shashua A and Carleo G 2022 *Phys. Rev. B* **106**(20) 205136 URL <https://link.aps.org/doi/10.1103/PhysRevB.106.205136>
- [27] Deng D L, Li X and Das Sarma S 2017 *Phys. Rev. X* **7**(2) 021021 URL <https://link.aps.org/doi/10.1103/PhysRevX.7.021021>
- [28] Gao X and Duan L M 2017 *Nature Communications* **8** 662 URL <https://doi.org/10.1038/s41467-017-00705-2>
- [29] Denis Z, Sinibaldi A and Carleo G 2023 Comment on “can neural quantum states learn volume-law ground states?” (*Preprint* <https://arxiv.org/pdf/2309.11534.pdf>)
- [30] Levine Y, Sharir O, Cohen N and Shashua A 2019 *Phys. Rev. Lett.* **122**(6) 065301 URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.065301>
- [31] Lu S, Gao X and Duan L M 2019 *Phys. Rev. B* **99**(15) 155136 URL <https://link.aps.org/doi/10.1103/PhysRevB.99.155136>
- [32] Luo D, Chen Z, Hu K, Zhao Z, Hur V M and Clark B K 2023 Gauge-invariant and anyonic-symmetric autoregressive neural network for quantum lattice models URL <https://link.aps.org/doi/10.1103/PhysRevResearch.5.013216>
- [33] Huang Y and Moore J E 2021 *Phys. Rev. Lett.* **127**(17) 170601 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.170601>
- [34] Sharir O, Levine Y, Wies N, Carleo G and Shashua A 2020 *Phys. Rev. Lett.* **124**(2) 020503 URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.020503>
- [35] Lange H, Döschl F, Carrasquilla J and Bohrdt A 2023 Neural network approach to quasiparticle dispersions in doped antiferromagnets (*Preprint* <https://arxiv.org/abs/2310.08578>)
- [36] Torlai G, Mazzola G, Carleo G and Mezzacapo A 2020 *Phys. Rev. Res.* **2**(2) 022060 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.022060>
- [37] Iouchtchenko D, Gonthier J F, Perdomo-Ortiz A and Melko R G 2023 *Machine Learning: Science and Technology* **4** 015016 URL <https://dx.doi.org/10.1088/2632-2153/acb4df>
- [38] Dawid A, Arnold J, Requena B, Gresch A, Płodzień M, Donatella K, Nicoli K A, Stornati P, Koch R, Büttner M, Okuła R, Muñoz-Gil G, Vargas-Hernández R A, Cervera-Lierta A, Carrasquilla J, Dunjko V, Gabrié M, Huembeli P, van Nieuwenburg E, Vicentini F, Wang L, Wetzel S J, Carleo G, Greplová E, Krems R, Marquardt F, Tomza M, Lewenstein M and Dauphin A 2022 Modern applications of machine learning in quantum sciences URL <https://arxiv.org/abs/2204.04198>
- [39] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborova L 2019 *Rev. Mod. Phys.* **91**(4) 045002 URL <https://link.aps.org/doi/10.1103/RevModPhys.91.045002>
- [40] Carrasquilla J and Torlai G 2021 *PRX Quantum* **2**(4) 040201 URL <https://link.aps.org/doi/10.1103/>

- PRXQuantum.2.040201
- [41] Melko R G and Carrasquilla J 2024 *Nature Computational Science* URL <https://doi.org/10.1038/s43588-023-00578-0>
- [42] Jia Z A, Yi B, Zhai R, Wu Y C, Guo G C and Guo G P 2019 *Advanced Quantum Technologies* **2** 1800077 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.201800077>
- [43] Yang Y, Cao H and Zhang Z 2019 *Science China Physics, Mechanics & Astronomy* **63** 210312 URL <https://doi.org/10.1007/s11433-018-9407-5>
- [44] Vivas D R, Madroño J, Bucheli V, Gómez L O and Reina J H 2022 Neural-network quantum states: A systematic review (*Preprint* <https://arxiv.org/abs/2204.12966>)
- [45] Reh M, Schmitt M and Gärttner M 2023 *Phys. Rev. B* **107**(19) 195115 URL <https://link.aps.org/doi/10.1103/PhysRevB.107.195115>
- [46] Van Den Nest M 2011 *Quantum Info. Comput.* **11** 784–812 ISSN 1533-7146 URL <https://dl.acm.org/doi/10.5555/2230936.2230941>
- [47] Chen J, Cheng S, Xie H, Wang L and Xiang T 2018 *Phys. Rev. B* **97**(8) 085104 URL <https://link.aps.org/doi/10.1103/PhysRevB.97.085104>
- [48] Glasser I, Pancotti N, August M, Rodriguez I D and Cirac J I 2018 *Phys. Rev. X* **8**(1) 011006 URL <https://link.aps.org/doi/10.1103/PhysRevX.8.011006>
- [49] Wu D, Rossi R, Vicentini F and Carleo G 2023 *Phys. Rev. Res.* **5**(3) L032001 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.5.L032001>
- [50] Passetti G, Hofmann D, Neitemeier P, Grunwald L, Sentef M A and Kennes D M 2023 *Physical Review Letters* **131** URL <https://doi.org/10.1103/PhysRevLett.131.036502>
- [51] Chen Z, Newhouse L, Chen E, Luo D and Soljačić M 2023 Antn: Bridging autoregressive neural networks and tensor networks for quantum many-body simulation (*Preprint* <https://arxiv.org/abs/2304.01996>)
- [52] Mehta P, Bukov M, Wang C H, Day A G, Richardson C, Fisher C K and Schwab D J 2019 *Physics Reports* **810** 1–124 ISSN 0370-1573 a high-bias, low-variance introduction to Machine Learning for physicists URL <https://www.sciencedirect.com/science/article/pii/S0370157319300766>
- [53] Cai Z and Liu J 2018 *Phys. Rev. B* **97**(3) 035116 URL <https://link.aps.org/doi/10.1103/PhysRevB.97.035116>
- [54] Choo K, Carleo G, Regnault N and Neupert T 2018 *Phys. Rev. Lett.* **121**(16) 167204 URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.167204>
- [55] Çeven K, Oktel M O and Keleş A 2022 *Phys. Rev. A* **106**(6) 063320 URL <https://link.aps.org/doi/10.1103/PhysRevA.106.063320>
- [56] Zhu Z, Mattheakis M, Pan W and Kaxiras E 2023 *Phys. Rev. Res.* **5**(4) 043084 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.5.043084>
- [57] Saito H and Kato M 2018 *Journal of the Physical Society of Japan* **87** 014001 (*Preprint* <https://doi.org/10.7566/JPSJ.87.014001>) URL <https://doi.org/10.7566/JPSJ.87.014001>
- [58] Teng P 2018 *Phys. Rev. E* **98**(3) 033305 URL <https://link.aps.org/doi/10.1103/PhysRevE.98.033305>
- [59] Hopfield J J 1982 *Proc Natl Acad Sci U S A* **79** 2554–2558 ISSN 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking)
- [60] Melko R G, Carleo G, Carrasquilla J and Cirac J I 2019 *Nature Physics* **15** 887–892 URL <https://doi.org/10.1038/s41567-019-0545-1>
- [61] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R and Carleo G 2018 *Nature Phys* **14** 447–450 ISSN 0370-1573 URL <https://doi.org/10.1038/s41567-018-0048-5>
- [62] Clark S R 2018 *Journal of Physics A: Mathematical and Theoretical* **51** 135301 URL <https://dx.doi.org/10.1088/1751-8121/aaaaf2>
- [63] Aoki K I and Kobayashi T 2016 *Modern Physics Letters B* **30** 1650401 (*Preprint* <https://doi.org/10.1142/S0217984916504017>) URL <https://doi.org/10.1142/S0217984916504017>
- [64] Liu D, Ran S J, Wittek P, Peng C, García R B, Su G and Lewenstein M 2019 *New Journal of Physics* **21** 073059 URL <https://doi.org/10.1088/1751-8121/aaaaf2>
- [65] Gan W C and Shu F W 2017 *International Journal of Modern Physics D* **26** 1743020 URL <https://doi.org/10.1142/S0218271817430209>
- [66] Borin A and Abanin D A 2020 *Phys. Rev. B* **101**(19) 195141 URL <https://link.aps.org/doi/10.1103/PhysRevB.101.195141>
- [67] Pastori L, Kaubruegger R and Budich J C 2019 *Phys. Rev. B* **99**(16) 165123 URL <https://link.aps.org/doi/10.1103/PhysRevB.99.165123>
- [68] Sfondrini A, Cerrillo J, Schuch N and Cirac J I 2010 *Phys. Rev. B* **81**(21) 214426 URL <https://link.aps.org/doi/10.1103/PhysRevB.81.214426>
- [69] Westerhout T, Astrakhantsev N, Tikhonov K S, Katsnelson M I and Bagrov A A 2020 *Nature Communications* **11** 1593 URL <https://doi.org/10.1038/s41467-020-15402-w>
- [70] Ferrari F, Becca F and Carrasquilla J 2019 *Phys. Rev. B* **100**(12) 125131 URL <https://link.aps.org/doi/10.1103/PhysRevB.100.125131>
- [71] Li C X, Yang S and Xu J B 2021 *Scientific Reports* **11** 16667 URL <https://doi.org/10.1038/s41598-021-95523-4>
- [72] Deng D L, Li X and Das Sarma S 2017 *Phys. Rev. B* **96**(19) 195145 URL <https://link.aps.org/doi/10.1103/PhysRevB.96.195145>
- [73] Kaubruegger R, Pastori L and Budich J C 2018 *Phys. Rev. B* **97**(19) 195136 URL <https://link.aps.org/doi/10.1103/PhysRevB.97.195136>
- [74] Valenti A, Greplova E, Lindner N H and Huber S D 2022 *Phys. Rev. Res.* **4**(1) L012010 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.4.L012010>
- [75] Saito H 2017 *Journal of the Physical Society of Japan* **86** 093001 (*Preprint* <https://doi.org/10.7566/JPSJ.86.093001>) URL <https://doi.org/10.7566/JPSJ.86.093001>
- [76] McBrien K, Carleo G and Khatami E 2019 *Journal of Physics: Conference Series* **1290** 012005 URL <https://dx.doi.org/10.1088/1742-6596/1290/1/012005>

- [77] Vargas-Calderón V, Vinck-Posada H and González F A 2020 *Journal of the Physical Society of Japan* **89** 094002 URL <https://doi.org/10.7566/JPSJ.89.094002>
- [78] Nomura Y, Darmawan A S, Yamaji Y and Imada M 2017 *Phys. Rev. B* **96**(20) 205152 URL <https://link.aps.org/doi/10.1103/PhysRevB.96.205152>
- [79] Nomura Y 2020 *Journal of the Physical Society of Japan* **89** 054706 (*Preprint* <https://doi.org/10.7566/JPSJ.89.054706>) URL <https://doi.org/10.7566/JPSJ.89.054706>
- [80] Xia R and Kais S 2018 *Nature Communications* **9** 4195 URL <https://doi.org/10.1038/s41467-018-06598-z>
- [81] Alcalde Puente D and Eremin I M 2020 *Phys. Rev. B* **102**(19) 195148 URL <https://link.aps.org/doi/10.1103/PhysRevB.102.195148>
- [82] Karthik V and Medhi A 2024 Convolutional restricted boltzmann machine (crbm) correlated variational wave function for the hubbard model on a square lattice: Mott metal-insulator transition (*Preprint* <https://arxiv.org/abs/2402.02794>)
- [83] Nomura Y 2021 *Journal of Physics: Condensed Matter* **33** 174003 URL <https://dx.doi.org/10.1088/1361-648X/abe268>
- [84] Viejira T, Casert C, Nys J, De Neve W, Haegeman J, Ryckebusch J and Verstraete F 2020 *Phys. Rev. Lett.* **124**(9) 097201 URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.097201>
- [85] Viejira T and Nys J 2021 *Phys. Rev. B* **104**(4) 045123 URL <https://link.aps.org/doi/10.1103/PhysRevB.104.045123>
- [86] Czischek S, Gärttner M and Gasenzer T 2018 *Phys. Rev. B* **98**(2) 024311 URL <https://link.aps.org/doi/10.1103/PhysRevB.98.024311>
- [87] Schmitt M and Heyl M 2020 *Phys. Rev. Lett.* **125**(10) 100503 URL <https://link.aps.org/doi/10.1103/PhysRevLett.125.100503>
- [88] Hofmann D, Fabiani G, Mentink J H, Carleo G and Sentef M A 2022 *SciPost Phys.* **12** 165 URL <https://scipost.org/10.21468/SciPostPhys.12.5.165>
- [89] Fabiani G and Mentink J H 2019 *SciPost Phys.* **7** 004 URL <https://scipost.org/10.21468/SciPostPhys.7.1.004>
- [90] Fabiani G, Bouman M D and Mentink J H 2021 *Phys. Rev. Lett.* **127**(9) 097202 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.097202>
- [91] Liang X, Liu W Y, Lin P Z, Guo G C, Zhang Y S and He L 2018 *Phys. Rev. B* **98**(10) 104426 URL <https://link.aps.org/doi/10.1103/PhysRevB.98.104426>
- [92] Wang J Q, He R Q and Lu Z Y 2023 Variational optimization of the amplitude of neural-network quantum many-body ground states (*Preprint* <https://arxiv.org/abs/2308.09664>)
- [93] Chen A and Heyl M 2023 Efficient optimization of deep neural quantum states toward machine precision (*Preprint* <https://arxiv.org/abs/2302.01941>)
- [94] Mendes-Santos T, Schmitt M and Heyl M 2023 *Phys. Rev. Lett.* **131**(4) 046501 URL <https://link.aps.org/doi/10.1103/PhysRevLett.131.046501>
- [95] Schmitt M, Rams M M, Dziarmaga J, Heyl M and Zurek W H 2022 *Science Advances* **8** eabl6850 (*Preprint* <https://www.science.org/doi/pdf/10.1126/sciadv.abl6850>) URL <https://www.science.org/doi/abs/10.1126/sciadv.abl6850>
- [96] Fu C, Zhang X, Zhang H, Ling H, Xu S and Ji S 2022 Lattice convolutional networks for learning ground states of quantum many-body systems (*Preprint* <https://arxiv.org/pdf/2206.07370.pdf>)
- [97] Schmale T, Reh M and Gärttner M 2022 *npj Quantum Information* **8** 115 URL <https://doi.org/10.1038/s41534-022-00621-4>
- [98] Choo K, Neupert T and Carleo G 2019 *Phys. Rev. B* **100**(12) 125124 URL <https://link.aps.org/doi/10.1103/PhysRevB.100.125124>
- [99] Roth C and MacDonald A H 2021 Group convolutional neural networks improve quantum state accuracy (*Preprint* <https://arxiv.org/abs/2104.05085>)
- [100] Roth C, Szabó A and MacDonald A H 2023 *Phys. Rev. B* **108**(5) 054410 URL <https://link.aps.org/doi/10.1103/PhysRevB.108.054410>
- [101] Duric T, Chung J H, Yang B and Sengupta P 2024 Spin-1/2 kagome heisenberg antiferromagnet: Machine learning discovery of the spinon pair density wave ground state (*Preprint* <https://arxiv.org/abs/2401.02866>)
- [102] Beck J, Bodky J, Motruk J, Müller T, Thomale R and Ghosh P 2024 Phase diagram of the  $j$ - $j_d$  heisenberg model on the maple-leaf lattice: Neural networks and density matrix renormalization group (*Preprint* <https://arxiv.org/abs/2401.04995>)
- [103] Bronstein M M, Bruna J, Cohen T and Velicković P 2021 Geometric deep learning: Grids, groups, graphs, geodesics, and gauges (*Preprint* [2104.13478](https://arxiv.org/abs/2104.13478))
- [104] Kochkov D, Pfaff T, Sanchez-Gonzalez A, Battaglia P and Clark B K 2021 Learning ground states of quantum hamiltonians with graph networks (*Preprint* <https://arxiv.org/abs/2110.06390>)
- [105] Kipf T N and Welling M 2017 Semi-supervised classification with graph convolutional networks (*Preprint* <https://arxiv.org/abs/1609.02907>)
- [106] Li Y, Tarlow D, Brockschmidt M and Zemel R 2017 Gated graph sequence neural networks (*Preprint* <https://arxiv.org/abs/1511.05493>)
- [107] Yang L, Hu W and Li L 2020 Scalable variational monte carlo with graph neural ansatz (*Preprint* <https://arxiv.org/abs/2011.12453>)
- [108] Pescia G, Nys J, Kim J, Lovato A and Carleo G 2023 Message-passing neural quantum states for the homogeneous electron gas (*Preprint* <https://arxiv.org/abs/2305.07240>)
- [109] Kim J, Pescia G, Fore B, Nys J, Carleo G, Gandolfi S, Hjorth-Jensen M and Lovato A 2023 Neural-network quantum states for ultra-cold fermi gases (*Preprint* <https://arxiv.org/abs/2305.08831>)
- [110] Luo D, Dai D D and Fu L 2023 Pairing-based graph neural network for simulating quantum materials (*Preprint* <https://arxiv.org/pdf/2311.02143.pdf>)
- [111] Kingma D P and Welling M 2022 Auto-encoding variational bayes (*Preprint* <https://arxiv.org/abs/1312.6114>)
- [112] Rocchetto A, Grant E, Strelchuk S, Carleo G and Severini S 2018 *npj Quantum Information* **4** 28 URL

- <https://doi.org/10.1038/s41534-018-0077-z>
- [113] Luchnikov I A, Ryzhov A, Stas P J, Filippov S N and Ouerdane H 2019 *Entropy* **21** ISSN 1099-4300 URL <https://www.mdpi.com/1099-4300/21/11/1091>
- [114] Walker N, Tam K M and Jarrell M 2020 *Scientific Reports* **10** 13047 URL <https://doi.org/10.1038/s41598-020-69848-5>
- [115] Schmitt M and Lenarčič Z 2022 *Phys. Rev. B* **106**(4) L041110 URL <https://link.aps.org/doi/10.1103/PhysRevB.106.L041110>
- [116] Bortone M, Rath Y and Booth G H 2023 Impact of conditional modelling for universal autoregressive quantum states (*Preprint* <https://arxiv.org/abs/2306.05917>)
- [117] Hibat-Allah M, Ganahl M, Hayward L E, Melko R G and Carrasquilla J 2020 *Phys. Rev. Res.* **2** 23358 publisher: American Physical Society URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.023358>
- [118] Roth C 2020 Iterative Retraining of Quantum Spin Models Using Recurrent Neural Networks arXiv:2003.06228 [cond-mat, physics:physics] URL <http://arxiv.org/abs/2003.06228>
- [119] Hibat-Allah M, Inack E M, Wiersema R, G J M R and Carrasquilla 2021 *Nature Machine Intelligence* **3** 2522–5839 URL <https://doi.org/10.1038/s42256-021-00401-3>
- [120] Steffen, Maximilian Z H G S A and Udluft 2006 Learning Long Term Dependencies with Recurrent Neural Networks *Artificial Neural Networks – ICANN 2006* ed Andreas, Włodzisław D, D O E K S and Stafylopatis (Springer Berlin Heidelberg) pp 71–80 ISBN 978-3-540-38627-8
- [121] Chung J, Gulcehre C, Cho K and Bengio Y 2014 Empirical evaluation of gated recurrent neural networks on sequence modeling (*Preprint* <https://arxiv.org/abs/1412.3555>)
- [122] Graves A, Fernandez S and Schmidhuber J 2007 Multi-dimensional recurrent neural networks (*Preprint* <https://arxiv.org/abs/0705.2011>)
- [123] Morawetz S, Vlugt I J S D, Carrasquilla J and Melko R G 2021 *Phys. Rev. A* **104** 12401 publisher: American Physical Society URL <https://link.aps.org/doi/10.1103/PhysRevA.104.012401>
- [124] Hibat-Allah M, Melko R G and Carrasquilla J 2022 Supplementing Recurrent Neural Network Wave Functions with Symmetry and Annealing to Improve Accuracy arXiv:2207.14314 [cond-mat, physics:physics] URL <http://arxiv.org/abs/2207.14314>
- [125] Carrasquilla J, Torlai G, Melko R G and Aolita L 2019 *Nature Machine Intelligence* **1** 155–161 ISSN 2522-5839 number: 3 Publisher: Nature Publishing Group URL <https://www.nature.com/articles/s42256-019-0028-1>
- [126] Hibat-Allah M, Melko R G and Carrasquilla J 2023 *Phys. Rev. B* **108**(7) 075152 URL <https://link.aps.org/doi/10.1103/PhysRevB.108.075152>
- [127] Döschl F, Lange H, Palm F, Grusdt F and Bohrdt A 2023 In preparation.
- [128] Luo D, Chen Z, Carrasquilla J and Clark B K 2022 *Phys. Rev. Lett.* **128**(9) 090501 URL <https://link.aps.org/doi/10.1103/PhysRevLett.128.090501>
- [129] Reh M, Schmitt M and Gärtner M 2021 *Phys. Rev. Lett.* **127**(23) 230501 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.230501>
- [130] Mendes-Santos T, Schmitt M, Angelone A, Rodriguez A, Scholl P, Williams H J, Barredo D, Lahaye T, Browaeys A, Heyl M and Dalmonte M 2023 URL <https://arxiv.org/abs/2301.13216>
- [131] Verstraete F, García-Ripoll J J and Cirac J I 2004 *Phys. Rev. Lett.* **93**(20) 207204 URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.207204>
- [132] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 *Advances in neural information processing systems* **30** URL [https://papers.nips.cc/paper\\_files/paper/2017](https://papers.nips.cc/paper_files/paper/2017)
- [133] Sprague K and Czischek S 2023 Variational monte carlo with large patched transformers (*Preprint* <https://arxiv.org/abs/2306.03921>)
- [134] Viteritti L L, Rende R and Becca F 2023 *Physical Review Letters* **130** 236401 ISSN 0031-9007, 1079-7114 arXiv:2211.05504 [cond-mat, physics:quant-ph] URL <http://arxiv.org/abs/2211.05504>
- [135] Zhang Y H and Di Ventra M 2023 *Physical Review B* **107** 075147 ISSN 2469-9950, 2469-9969 arXiv:2208.01758 [cond-mat, physics:physics, physics:quant-ph] URL <http://arxiv.org/abs/2208.01758>
- [136] Cha P, Ginsparg P, Wu F, Carrasquilla J, McMahon P L and Kim E A 2021 *Machine Learning: Science and Technology* **3** 01LT01 URL <https://dx.doi.org/10.1088/2632-2153/ac362b>
- [137] von Glehn I, Spencer J S and Pfau D 2023 A self-attention ansatz for ab-initio quantum chemistry (*Preprint* <https://arxiv.org/abs/2211.13672>)
- [138] Shang H, Guo C, Wu Y, Li Z and Yang J 2023 Solving schrödinger equation with a language model (*Preprint* <https://arxiv.org/abs/2307.09343>)
- [139] Wu Y, Guo C, Fan Y, Zhou P and Shang H 2023 Nnqs-transformer: an efficient and scalable neural network quantum states approach for ab initio quantum chemistry (*Preprint* <https://arxiv.org/abs/2306.16705>)
- [140] Rende R, Gerace F, Laio A and Goldt S 2023 Optimal inference of a generalised Potts model by single-layer transformers with factored attention arXiv:2304.07235 [cond-mat, stat] URL <http://arxiv.org/abs/2304.07235>
- [141] Rende R, Viteritti L L, Bardone L, Becca F and Goldt S 2023 A simple linear algebra identity to optimize Large-Scale Neural Network Quantum States arXiv:2310.05715 [cond-mat] (*Preprint* <http://arxiv.org/abs/2310.05715>)
- [142] Pfau D, Spencer J, de G Matthews A and Foulkes W 2020 *Phys. Rev. Research* **2**(3) 033429 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.033429>
- [143] Hermann J, Schätzle Z and Noé F 2020 *Nature Chemistry* **12**(10) 1755–4349 URL <https://doi.org/10.1038/s41557-020-0544-y>
- [144] Schätzle Z, Hermann J and Noé F 2021 *The Journal of Chemical Physics* **154** 124108 ISSN 0021-9606 URL <https://doi.org/10.1063/5.0032836>
- [145] Han J, Zhang L and E W 2019 *Journal of Computational Physics* **399** 108929 URL <https://doi.org/10.1016/j.jcp.2019.108929>
- [146] Pang T, Yan S and Lin M 2022  $o(n^2)$  universal antisymmetry in fermionic neural networks (*Preprint* <https://arxiv.org/abs/2207.14314>)

- [//arxiv.org/abs/2205.13205](https://arxiv.org/abs/2205.13205))
- [147] Stokes J, Moreno J R, Pnevmatikakis E A and Carleo G 2020 *Phys. Rev. B* **102**(20) 205122 URL <https://link.aps.org/doi/10.1103/PhysRevB.102.205122>
- [148] Liu Z and Clark B K 2023 A unifying view of fermionic neural network quantum states: From neural network backflow to hidden fermion determinant states (*Preprint* <https://arxiv.org/abs/2311.09450>)
- [149] Moreno J R, Carleo G, Georges A and Stokes J 2022 *Proceedings of the National Academy of Sciences* **119** e2122059119 (*Preprint* <https://www.pnas.org/doi/pdf/10.1073/pnas.2122059119>) URL <https://www.pnas.org/doi/abs/10.1073/pnas.2122059119>
- [150] Feynman R P and Cohen M 1956 *Phys. Rev.* **102**(5) 1189–1204 URL <https://link.aps.org/doi/10.1103/PhysRev.102.1189>
- [151] Luo D and Clark B K 2019 *Phys. Rev. Lett.* **122**(22) 226401 URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.226401>
- [152] Nomura Y, Darmawan A S, Yamaji Y and Imada M 2017 *Phys. Rev. B* **96**(20) 205152 URL <https://link.aps.org/doi/10.1103/PhysRevB.96.205152>
- [153] Humeniuk S, Wan Y and Wang L 2022 Autoregressive neural slater-jastrow ansatz for variational monte carlo simulation (*Preprint* <https://arxiv.org/abs/2210.05871>)
- [154] Barrett T D, Malyshev A and Lvovsky A I 2022 *Nature Machine Intelligence* **4** 351–358 URL <https://doi.org/10.1038/s42256-022-00461-z>
- [155] Inui K, Kato Y and Motome Y 2021 *Phys. Rev. Res.* **3**(4) 043126 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.3.043126>
- [156] Yoshioka N, Mizukami W and Nori F 2021 *Communications Physics* **4**(1) 2399–3650 URL <https://doi.org/10.1038/s42005-021-00609-0>
- [157] Viteritti L L, Ferrari F and Becca F 2022 *SciPost Phys.* **12** 166 URL <https://scipost.org/10.21468/SciPostPhys.12.5.166>
- [158] Malyshev A, Arrazola J M and Lvovsky A I 2023 Autoregressive neural quantum states with quantum number symmetries (*Preprint* <https://arxiv.org/abs/2310.04166>)
- [159] Luo D, Carleo G, Clark B K and Stokes J 2021 *Phys. Rev. Lett.* **127**(27) 276402 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.276402>
- [160] Sehayeck D, Golubeva A, Albergo M S, Kulchytskyy B, Torlai G and Melko R G 2019 *Phys. Rev. B* **100**(19) 195125 URL <https://link.aps.org/doi/10.1103/PhysRevB.100.195125>
- [161] Golubeva A and Melko R G 2022 *Phys. Rev. B* **105**(12) 125124 URL <https://link.aps.org/doi/10.1103/PhysRevB.105.125124>
- [162] Dash S, Vicentini F, Ferrero M and Georges A 2024 Efficiency of neural quantum states in light of the quantum geometric tensor (*Preprint* <https://arxiv.org/abs/2402.01565>)
- [163] Klassert R, Baumbach A, Petrovici M A and Gärttner M 2022 *iScience* **25** 104707 ISSN 2589-0042 URL <https://www.sciencedirect.com/science/article/pii/S2589004222009798>
- [164] Czischek S, Baumbach A, Billaudelle S, Cramer B, Kades L, Pawlowski J M, Oberthaler M K, Schemmel J, Petrovici M A, Gasenzer T and Gärttner M 2022 *SciPost Phys.* **12** 039 URL <https://scipost.org/10.21468/SciPostPhys.12.1.039>
- [165] Czischek S, Pawlowski J M, Gasenzer T and Gärttner M 2019 *Phys. Rev. B* **100**(19) 195120 URL <https://link.aps.org/doi/10.1103/PhysRevB.100.195120>
- [166] Vicentini F, Hofmann D, Szabó A, Wu D, Roth C, Giuliani C, Pescia G, Nys J, Vargas-Calderón V, Astrakhantsev N and Carleo G 2022 *SciPost Phys. Codebases* **7** URL <https://scipost.org/10.21468/SciPostPhysCodeb.7>
- [167] Sinibaldi A and Vicentini F 2023 Netket fidelity package URL [https://github.com/netket/netket\\_fidelity](https://github.com/netket/netket_fidelity)
- [168] Schmitt M and Reh M 2022 *SciPost Phys. Codebases* **2** URL <https://scipost.org/10.21468/SciPostPhysCodeb.2>
- [169] Bradbury J, Frostig R, Hawkins P, Johnson M J, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S and Zhang Q 2018 JAX: composable transformations of Python+NumPy programs URL <http://github.com/google/jax>
- [170] Beach M J S, Vlucht I D, Golubeva A, Huembeli P, Kulchytskyy B, Luo X, Melko R G, Merali E and Torlai G 2019 *SciPost Phys.* **7** 009 URL <https://scipost.org/10.21468/SciPostPhys.7.1.009>
- [171] McMillan W L 1965 *Phys. Rev.* **138**(2A) A442–A451 URL <https://link.aps.org/doi/10.1103/PhysRev.138.A442>
- [172] Huang L and Wang L 2017 *Phys. Rev. B* **95**(3) 035105 URL <https://link.aps.org/doi/10.1103/PhysRevB.95.035105>
- [173] Assaraf R and Caffarel M 1999 *Phys. Rev. Lett.* **83**(23) 4682–4685 URL <https://link.aps.org/doi/10.1103/PhysRevLett.83.4682>
- [174] Kingma D P and Ba J 2017 Adam: A method for stochastic optimization (*Preprint* <https://arxiv.org/abs/1412.6980>)
- [175] Loshchilov I and Hutter F 2019 Decoupled weight decay regularization (*Preprint* <https://arxiv.org/abs/1711.05101>)
- [176] Sorella S 1998 *Phys. Rev. Lett.* **80**(20) 4558–4561 URL <https://link.aps.org/doi/10.1103/PhysRevLett.80.4558>
- [177] Sorella S 2001 *Phys. Rev. B* **64**(2) 024512 URL <https://link.aps.org/doi/10.1103/PhysRevB.64.024512>
- [178] Amari S and Douglas S 1998 Why natural gradient? *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98* vol 2 pp 1213–1216 vol.2
- [179] Amari S, Karakida R and Oizumi M 2019 Fisher information and natural gradient learning in random deep networks *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research* vol 89) ed Chaudhuri K and Sugiyama M (PMLR) pp 694–702 URL <https://proceedings.mlr.press/v89/amari19a.html>
- [180] Hackl L, Guaita T, Shi T, Haegeman J, Demler E and Cirac J I 2020 *SciPost Phys.* **9** 048 URL <https://scipost.org/10.21468/SciPostPhys.9.4.048>

- [181] Wagner D, Klümper A and Sirker J 2023 arXiv preprint arXiv:2311.13799
- [182] Park C Y and Kastoryano M J 2020 *Phys. Rev. Res.* **2**(2) 023232 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.023232>
- [183] Donatella K, Denis Z, Le Boité A and Ciuti C 2023 *Phys. Rev. A* **108**(2) 022210 URL <https://link.aps.org/doi/10.1103/PhysRevA.108.022210>
- [184] Li M, Chen J, Xiao Q, Wang F, Jiang Q, Zhao X, Lin R, An H, Liang X and He L 2022 *IEEE Transactions on Parallel and Distributed Systems* **33** 2846–2859
- [185] Zhao X, Li M, Xiao Q, Chen J, Wang F, Shen L, Zhao M, Wu W, An H, He L and Liang X 2022 Ai for quantum mechanics: High performance quantum many-body simulations via deep learning *SC22: International Conference for High Performance Computing, Networking, Storage and Analysis* pp 1–15
- [186] Choi S C T and Saunders M A 2014 *ACM Trans. Math. Softw.* **40** ISSN 0098-3500 URL <https://doi.org/10.1145/2527267>
- [187] Zhang W, Xu X, Wu Z, Balachandran V and Poletti D 2023 *Phys. Rev. B* **107**(16) 165149 URL <https://link.aps.org/doi/10.1103/PhysRevB.107.165149>
- [188] Bukov M, Schmitt M and Dupont M 2021 *SciPost Phys.* **10** 147 URL <https://scipost.org/10.21468/SciPostPhys.10.6.147>
- [189] Inack E M, Morawetz S and Melko R G 2022 *Condensed Matter* **7** ISSN 2410-3896 URL <https://www.mdpi.com/2410-3896/7/2/38>
- [190] Khandoker S A, Abedin J M and Hibat-Allah M 2023 *Machine Learning: Science and Technology* **4** 15026 publisher: IOP Publishing URL <https://dx.doi.org/10.1088/2632-2153/acb895>
- [191] Zen R, My L, Tan R, Hébert F, Gattobigio M, Miniatura C, Poletti D and Bressan S 2020 *Phys. Rev. E* **101**(5) 053301 URL <https://link.aps.org/doi/10.1103/PhysRevE.101.053301>
- [192] Efthymiou S, Beach M J S and Melko R G 2019 *Phys. Rev. B* **99**(7) 075113 URL <https://link.aps.org/doi/10.1103/PhysRevB.99.075113>
- [193] Chen H, Hendry D, Weinberg P and Feiguin A E 2022 Systematic improvement of neural network quantum states using a lanczos recursion (*Preprint* <https://arxiv.org/abs/2206.14307>)
- [194] Giuliani C, Vicentini F, Rossi R and Carleo G 2023 *Quantum* **7** 1096 ISSN 2521-327X URL <https://doi.org/10.22331/q-2023-08-29-1096>
- [195] Ledinauskas E and Anisimovas E 2023 *SciPost Phys.* **15** 229 URL <https://scipost.org/10.21468/SciPostPhys.15.6.229>
- [196] Kochkov D and Clark B K 2018 Variational optimization in the ai era: Computational graph states and supervised wave-function optimization (*Preprint* <https://arxiv.org/abs/1811.12423>)
- [197] Atanasova H, Bernheimer L and Cohen G 2023 *Nature Communications* **14** 3601 URL <https://doi.org/10.1038/s41467-023-39244-4>
- [198] Szabó A and Castelnovo C 2020 *Phys. Rev. Res.* **2**(3) 033075 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.033075>
- [199] Lange H, Kebrić M, Buser M, Schollwöck U, Grusdt F and Bohrdt A 2023 *Quantum* **7** 1129 URL <https://doi.org/10.22331/q-2023-10-09-1129>
- [200] Pfau D, Axelrod S, Sutterud H, von Glehn I and Spencer J S 2023 Natural quantum monte carlo computation of excited states (*Preprint* <https://arxiv.org/abs/2308.16848>)
- [201] Sinibaldi A, Giuliani C, Carleo G and Vicentini F 2023 arXiv preprint arXiv:2305.14294
- [202] Lin S H and Pollmann F 2022 *physica status solidi (b)* **259** 2100172 (*Preprint* <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pssb.202100172>) URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/pssb.202100172>
- [203] Gutiérrez I L and Mendl C B 2022 *Quantum* **6** 627
- [204] Lee C K, Patil P, Zhang S and Hsieh C Y 2021 *Phys. Rev. Res.* **3**(2) 023095 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.3.023095>
- [205] Bureau H and Heyl M 2021 *Phys. Rev. Lett.* **127**(5) 050601 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.050601>
- [206] Zhang Z W, Yang S, Wu Y H, Liu C X, Han Y M, Lee C H, Sun Z, Li G J and Zhang X 2020 *Chinese Physics Letters* **37** 018401
- [207] Schmitt M and Heyl M 2018 *SciPost Phys.* **4** 013 URL <https://scipost.org/10.21468/SciPostPhys.4.2.013>
- [208] Verdel R, Schmitt M, Huang Y P, Karpov P and Heyl M 2021 *Phys. Rev. B* **103**(16) 165103 URL <https://link.aps.org/doi/10.1103/PhysRevB.103.165103>
- [209] Karpov P, Verdel R, Huang Y P, Schmitt M and Heyl M 2021 *Phys. Rev. Lett.* **126**(13) 130401 URL <https://link.aps.org/doi/10.1103/PhysRevLett.126.130401>
- [210] Hendry D, Chen H, Weinberg P and Feiguin A E 2021 *Phys. Rev. B* **104**(20) 205130 URL <https://link.aps.org/doi/10.1103/PhysRevB.104.205130>
- [211] Hendry D and Feiguin A E 2019 *Phys. Rev. B* **100**(24) 245123 URL <https://link.aps.org/doi/10.1103/PhysRevB.100.245123>
- [212] Feiguin A E and White S R 2005 *Phys. Rev. B* **72**(22) 220401 URL <https://link.aps.org/doi/10.1103/PhysRevB.72.220401>
- [213] Torlai G and Melko R G 2018 *Phys. Rev. Lett.* **120**(24) 240503 URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.240503>
- [214] Nomura Y, Yoshioka N and Nori F 2021 *Phys. Rev. Lett.* **127**(6) 060601 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.060601>
- [215] Nys J, Denis Z and Carleo G 2023 arXiv preprint arXiv:2309.07063
- [216] White S R 2009 *Phys. Rev. Lett.* **102**(19) 190601 URL <https://link.aps.org/doi/10.1103/PhysRevLett.102.190601>
- [217] Stoudenmire E M and White S R 2010 *New Journal of Physics* **12** 055026 URL <https://dx.doi.org/10.1088/1367-2630/12/5/055026>

- [218] Hendry D, Chen H and Feiguin A 2022 *Phys. Rev. B* **106**(16) 165111 URL <https://link.aps.org/doi/10.1103/PhysRevB.106.165111>
- [219] Irikura N and Saito H 2020 *Phys. Rev. Res.* **2**(1) 013284 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.013284>
- [220] Lu S, Giudice G and Cirac J I 2024 Variational neural and tensor network approximations of thermal states (*Preprint* <https://arxiv.org/abs/2401.14243>)
- [221] Hartmann M J and Carleo G 2019 *Phys. Rev. Lett.* **122**(25) 250502 URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.250502>
- [222] Mazza P P, Zietlow D, Carollo F, Andergassen S, Martius G and Lesanovsky I 2021 *Physical Review Research* **3** ISSN 2643-1564 URL <http://dx.doi.org/10.1103/PhysRevResearch.3.023084>
- [223] Carnazza F, Carollo F, Zietlow D, Andergassen S, Martius G and Lesanovsky I 2022 *New Journal of Physics* **24** 073033 URL <https://dx.doi.org/10.1088/1367-2630/ac7df6>
- [224] Vicentini F, Rossi R and Carleo G 2022 Positive-definite parametrization of mixed quantum states with deep neural networks (*Preprint* <https://arxiv.org/abs/2206.13488>)
- [225] Herrera Rodríguez L E and Kananenka A A 2021 *The Journal of Physical Chemistry Letters* **12** 2476–2483 URL <https://doi.org/10.1021/acs.jpcllett.1c00079>
- [226] Mellak J, Arrigoni E and von der Linden W 2024 Deep neural networks as variational solutions for correlated open quantum systems (*Preprint* <https://arxiv.org/abs/2401.14179>)
- [227] Liu Z, Duan L M and Deng D L 2022 *Phys. Rev. Res.* **4**(1) 013097 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.4.013097>
- [228] Kaestle O and Carmele A 2021 *Phys. Rev. B* **103**(19) 195420 URL <https://link.aps.org/doi/10.1103/PhysRevB.103.195420>
- [229] Mellak J, Arrigoni E, Pock T and von der Linden W 2023 *Phys. Rev. B* **107**(20) 205102 URL <https://link.aps.org/doi/10.1103/PhysRevB.107.205102>
- [230] Carrasquilla J, Luo D, Pérez F, Milsted A, Clark B K, Volkovs M and Aolita L 2021 *Phys. Rev. A* **104**(3) 032610 URL <https://link.aps.org/doi/10.1103/PhysRevA.104.032610>
- [231] Weimer H 2015 *Phys. Rev. Lett.* **114**(4) 040402 URL <https://link.aps.org/doi/10.1103/PhysRevLett.114.040402>
- [232] Nagy A and Savona V 2019 *Phys. Rev. Lett.* **122**(25) 250501 URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.250501>
- [233] Vicentini F, Biella A, Regnault N and Ciuti C 2019 *Phys. Rev. Lett.* **122**(25) 250503 URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.250503>
- [234] Yoshioka N and Hamazaki R 2019 *Phys. Rev. B* **99**(21) 214306 URL <https://link.aps.org/doi/10.1103/PhysRevB.99.214306>
- [235] Cramer M, Plenio M B, Flammaria S T, Somma R, Gross D, Bartlett S D, Landon-Cardinal O, Poulin D and Liu Y K 2010 *Nature Communications* **1** 149 URL <https://doi.org/10.1038/ncomms1147>
- [236] Torlai G, Mazzola G, Carleo G and Mezzacapo A 2020 *Phys. Rev. Res.* **2**(2) 022060 URL <https://link.aps.org/doi/10.1103/PhysRevResearch.2.022060>
- [237] Häffner H, Hänsel W, Roos C F, Benhelm J, Chek-al kar D, Chwalla M, Körber T, Rapol U D, Riebe M, Schmidt P O, Becher C, Gühne O, Dür W and Blatt R 2005 *Nature* **438** 643–646 URL <https://doi.org/10.1038/nature04279>
- [238] Hradil Z 1997 *Phys. Rev. A* **55**(3) R1561–R1564 URL <https://link.aps.org/doi/10.1103/PhysRevA.55.R1561>
- [239] Lohani S, Kirby B T, Brodsky M, Danaci O and Glasser R T 2020 *Machine Learning: Science and Technology* **1** 035007 URL <https://dx.doi.org/10.1088/2632-2153/ab9a21>
- [240] Koutný D, Motka L, Hradil Z c v, Řeháček J and Sánchez-Soto L L 2022 *Phys. Rev. A* **106**(1) 012409 URL <https://link.aps.org/doi/10.1103/PhysRevA.106.012409>
- [241] Torlai G and Melko R G 2016 *Phys. Rev. B* **94**(16) 165134 URL <https://link.aps.org/doi/10.1103/PhysRevB.94.165134>
- [242] Wei V, Coish W A, Ronagh P and Muschik C A 2023 Neural-shadow quantum state tomography (*Preprint* <https://arxiv.org/abs/2305.01078>)
- [243] Torlai G, Timar B, van Nieuwenburg E P L, Levine H, Omran A, Keesling A, Bernien H, Greiner M, Vuletić V, Lukin M D, Melko R G and Endres M 2019 *Phys. Rev. Lett.* **123**(23) 230504 URL <https://link.aps.org/doi/10.1103/PhysRevLett.123.230504>
- [244] Torlai G and Melko R G 2020 *Annual Review of Condensed Matter Physics* **11** 325–344
- [245] Zhao H, Carleo G and Vicentini F 2023 Empirical sample complexity of neural network mixed state reconstruction (*Preprint* <https://arxiv.org/abs/2307.01840>)
- [246] Melkani A, Gneiting C and Nori F 2020 *Phys. Rev. A* **102**(2) 022412 URL <https://link.aps.org/doi/10.1103/PhysRevA.102.022412>
- [247] Palmieri A M, Kovlakov E, Bianchi F, Yudin D, Straupe S, Biamonte J D and Kulik S 2020 *npj Quantum Information* **6** 20
- [248] Neugebauer M, Fischer L, Jäger A, Czischek S, Jochim S, Weidemüller M and Gärttner M 2020 *Phys. Rev. A* **102**(4) 042604 URL <https://link.aps.org/doi/10.1103/PhysRevA.102.042604>
- [249] Huang H Y, Kueng R and Preskill J 2020 *Nature Physics* **16** 1050–1057 URL <https://doi.org/10.1038/s41567-020-0932-7>
- [250] Xin T, Lu S, Cao N, Anikeeva G, Lu D, Li J, Long G and Zeng B 2019 *npj Quantum Information* **5** 109 URL <https://doi.org/10.1038/s41534-019-0222-3>
- [251] Smith A W R, Gray J and Kim M S 2021 *PRX Quantum* **2**(2) 020348 URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.020348>
- [252] Quek Y, Fort S and Ng H K 2021 Adaptive quantum state tomography with neural networks URL <https://doi.org/10.1038/s41534-021-00436-9>
- [253] Zhu Y, Wu Y D, Bai G, Wang D S, Wang Y and Chiribella G 2022 *Nature Communications* **13** 6222 URL

- <https://doi.org/10.1038/s41467-022-33928-z>
- [254] Ma H, Sun Z, Dong D, Chen C and Rabitz H 2023 Attention-based transformer networks for quantum state tomography (*Preprint* <https://arxiv.org/abs/2305.05433>)
- [255] Tiunov E S, Tiunova V, Ulanov A E, Lvovsky A and Fedorov A K 2020 *Optica* **7** 448–454
- [256] Zhong L, Guo C and Wang X 2022 Quantum state tomography inspired by language modeling (*Preprint* <https://arxiv.org/abs/2212.04940>)
- [257] Ahmed S, Sánchez Muñoz C, Nori F and Kockum A F 2021 *Phys. Rev. Lett.* **127**(14) 140502 URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.140502>
- [258] Czischek S, Moss M S, Radzihovsky M, Merali E and Melko R G 2022 *Phys. Rev. B* **105** 205108 publisher: American Physical Society URL <https://link.aps.org/doi/10.1103/PhysRevB.105.205108>
- [259] Ebadi S, Wang T T, Levine H, Keesling A, Semeghini G, Omran A, Bluvstein D, Samajdar R, Pichler H, Ho W W, Choi S, Sachdev S, Greiner M, Vuletić V and Lukin M D 2021 *Nature* **595** 227–232 URL <https://doi.org/10.1038/s41586-021-03582-4>
- [260] Moss M S, Ebadi S, Wang T T, Semeghini G, Bohrdt A, Lukin M D and Melko R G 2023 Enhancing variational Monte Carlo using a programmable quantum simulator arXiv:2308.02647 [cond-mat, physics:quant-ph] URL <http://arxiv.org/abs/2308.02647>
- [261] Bennewitz E R, Hopfmueller F, Kulchytskyy B, Carrasquilla J and Ronagh P 2022 *Nature Machine Intelligence* **4** 618–624 URL <https://doi.org/10.1038/s42256-022-00509-0>
- [262] Montanaro A and Stanisc S 2023 Accelerating variational quantum monte carlo using the variational quantum eigensolver (*Preprint* <https://arxiv.org/abs/2307.07719>)