Simplicity of mean-field theories in neural quantum states

Fabian Ballar Trigueros ⬤, Tiago Mendes-Santos ⬤, and Markus Heyl ⬤

Theoretical Physics III, Center for Electronic Correlations and Magnetism, Institute of Physics,
University of Augsburg, 86135 Augsburg, Germany

The utilization of artificial neural networks to represent quantum many-body wave functions has garnered significant attention, with enormous recent progress for both ground states and nonequilibrium dynamics. However, quantifying state complexity within this neural quantum states framework remains elusive. In this study, we address this key open question from a complementary point of view: Which states are simple to represent with neural quantum states? Concretely, we show on a general level that ground states of mean-field theories with permutation symmetry require only a limited number of independent neural network parameters. We analytically establish that, in the thermodynamic limit, convergence to the ground state of the fully connected transverse-field Ising model (TFIM), the mean-field Ising model, can be achieved with just one single parameter. Expanding our analysis, we explore the behavior of the one-parameter ansatz under breaking of the permutation symmetry. For that purpose, we consider the TFIM with tunable long-range interactions, characterized by an interaction exponent α. We show analytically that the one-parameter ansatz for the neural quantum state still accurately captures the ground state for a whole range of values for $0 \leqslant \alpha \leqslant 1$, implying a mean-field description of the model in this regime.

I. INTRODUCTION

Representing the wave function of complex quantum matter is exceedingly difficult. Addressing this challenge has prompted the proposal of various techniques and approximations [1–6]. However, each method encounters distinct difficulties. For instance, exact diagonalization becomes computationally intractable for large systems due to the exponentially growing basis. Quantum Monte Carlo methods can encounter the sign problem [7] for many interesting systems. Tensor networks experience an exponentially increasing bond dimension for systems with volume-law scaling of entanglement entropy [8,9]. Recently, neural quantum states (NQSs) were introduced as an alternative method to leverage the expressive power of neural networks to represent the quantum wave function [10].

This approach has delivered remarkable results in discovering ground states and describing dynamics of quantum many-body systems in regimes previously inaccessible via other methods [11–15]. One key feature that distinguishes NQSs from tensor networks is the ability to represent volume-law entangled states [16].

Despite demonstrations of the expressive power of NQSs [17,18], there is a critical need to understand a metric of complexity for it, similar to how bond dimension and circuit depth are utilized to characterize tensor networks and quantum

circuits, respectively. The outstanding question is, What criteria can be employed to quantify the complexity of a neural quantum state? In a first step we approach this question from a complementary side: Which states can be easily described by a neural quantum network? Much like how we know that area-law states can be efficiently described using a tensor network, recognizing the states that NQSs can easily describe might be crucial in advancing our understanding of NQS complexity.

A natural choice is to consider that the complexity is related to the number of independent variational parameters K a trained NQS requires to describe a specific quantum state accurately. In this work, we characterize states that have minimal complexity in this sense under the neural quantum states formalism. In particular, we show that neural quantum states can describe mean-field theories with permutation symmetry with a very small number of parameters. We show that, on a general level, this leads to a reduction in required network parameters from $K \times L$ to just K, where K and L correspond to the number of hidden neurons and the number of physical spins, respectively; see Fig. 1 for an illustration.

We demonstrate this using the fully connected transverse-field Ising model (TFIM). Most importantly, we find that for this model convergence in the thermodynamic limit is achieved even with a single parameter, i.e., $K = 1$. This makes the mean-field solution of the TFIM as simple as possible for NQSs. As a next step, we study the behavior of the network beyond mean-field theory by breaking the permutation symmetry. For that purpose, we consider the long-range interacting TFIM with power-law decaying interactions. This allows us to tune the deviation from mean field through the interaction exponent α in a controlled way. We show analytically that a neural network with a single parameter is sufficient
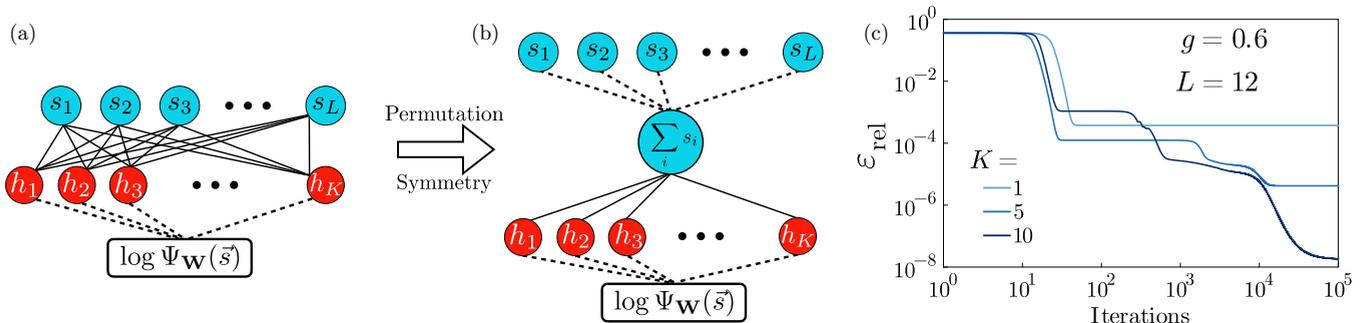
FIG. 1. Permutation-invariant artificial neural networks and training. (a) An unconstrained structure of a general feed-forward network with one hidden layer. Dashed lines carry no variational parameters, whereas solid lines denote the fact that we compute the product between the input spin and a variational parameter. (b) Enforcing permutation symmetry leads effectively to a modified neural network architecture, where the input spin configuration is first transformed into a collective total spin before further processing occurs. (c) Training convergence and scaling of the ground state search process for the fully connected transverse-field Ising model for different numbers of hidden spins K at a fixed system size $L = 12$. Here, $\epsilon_{\rm rel}$ is the relative energy error of the neural quantum state with the exact diagonalization result as a reference.

to obtain the ground state for $\alpha \leqslant 1$ in the thermodynamic limit. As a consequence, we find that the ground state of the long-range interacting TFIM is still described by a mean-field theory for this range of values of α.

II. NEURAL QUANTUM STATES AND PERMUTATION INVARIANCE

Neural quantum states offer a framework that leverages the expressive capabilities and generalization power of neural networks to represent quantum wave functions. This approach relies on the notion of expressing a wave function within a complete basis set denoted as $|s\rangle$, characterized by the expansion

$$|\Psi\rangle = \sum_s \Psi_{\mathbf{W}}(s)|s\rangle. \tag{1}$$

Here, $\Psi_{\mathbf{W}}(s)$ represents the neural network ansatz which depends on a set of variational parameters $\mathbf{W}$ that we learn according to a prescription of choice. In this work we will consider a one-dimensional lattice of size L with periodic boundary conditions and spin-1/2 degrees of freedom. We take as a basis $|s\rangle$, the spin configuration in the computational basis.

We are particularly interested in the structure of $\mathbf{W}$ for a permutation-invariant Hamiltonian as it appears naturally for mean-field theories. For this, we will approximate the wave function by an artificial neural network. So we consider a general feed-forward neural network with L input units, K hidden neurons, no bias, and, last, a single output neuron that returns the value of $\ln[\Psi_{\mathbf{W}}(s)]$. A schematic depiction of the initial architecture can be found in Fig. 1(a).

One important technique to minimize the number of parameters needed to describe the wave function in NQSs is to take advantage of the symmetries at the level of the weight matrix [10], reducing the number of independent parameters. As a consequence, given that the network does not need to find the symmetry on its own, the computational cost of NQSs can be greatly reduced without compromising the accuracy of the model.

The variational ansatz for our architecture can be expressed as

$$\Psi_{\mathbf{W}}(s) = \langle s|\Psi\rangle = \exp\left[\sum_i f(y_i(s))\right], \tag{2}$$

where $y_i(s) \equiv \sum_j W_{ij}s_j$, f denotes the activation function and W_{ij} represents the weight matrix elements. However, to ensure that the output remains invariant under any permutation of the input configuration, we impose a constraint between the elements of $W_{i,j}$. For this, we consider the set of all possible permutations of the form $\pi : s_j \to s_{\pi(j)}$, i.e., $\pi(s) \equiv (s_{\pi(1)}, \ldots, s_{\pi(L)})$ and we require that

$$y_i(s) - y_i(\pi(s)) = \sum_j W_{ij}[s_j - \pi(s)_j] \stackrel{!}{=} 0. \tag{3}$$

Given that due to the nature of permutations π is a bijective mapping and hence there is a unique π^{-1}, we then rewrite our constraint in a more convenient way as

$$\sum_j (W_{ij} - W_{i\pi^{-1}(j)})s_j \stackrel{!}{=} 0. \tag{4}$$

Since this is supposed to hold for any permutation and to be independent of the values of s_j, it follows that the quantity in parentheses has to vanish, e.g., $W_{ij} = W_{i\pi^{-1}(j)}$. Given that we include all possible permutations of our input configurations, this immediately implies that all elements along the rows are constrained to be exactly identical ($W_{ij} = W_i$). Therefore, our weight matrix now has the form

$$\mathbf{W} = \begin{bmatrix} \vec{w}_1 \\ \vdots \\ \vec{w}_K \end{bmatrix}, \quad \vec{w}_i \equiv W_i \underbrace{[1, 1, \cdots, 1]}_{L}. \tag{5}$$

Enforcing the permutation symmetry effectively reduced the number of independent parameters from $L \times K$ in the general weight matrix to K in the constrained case. This is reminiscent of the implementations of symmetry for translational invariance in [10,19,20], where the weight matrix is reduced to a set of circulant matrices [21] and each matrix can be treated as a convolutional kernel that is applied on all

translated versions of the spin configuration s. It is worth mentioning that permutation symmetry imposes a much stronger constraint than translation invariance.

As a key consequence of the above considerations, we can map the symmetry-imposed weight matrix to a new feedforward network that takes as input the total magnetization $M_s = \sum_i s_i$ and returns the value of $\ln \Psi(s)$ [see Fig. 1(b)]. Using this structure, the symmetry-imposed output wave function can be written as

$$\ln(\langle s|\Psi\rangle) = \sum_{k=1}^{K} f(W_k M_s). \tag{6}$$

It is important to note that this neural network wave function is not generally equivalent to a product state ansatz. Therefore, it opens the possibility to capture finite-size effects or correlations that could not be accounted for otherwise. Product states can still be captured by our ansatz, as we will show in Sec. III, where in the thermodynamic limit this ansatz becomes a product state asymptotically.

Therefore, when one aims to describe the ground state of a permutation-invariant Hamiltonian, Eq. (6) provides an ansatz that accurately approximates the exact value for sufficiently many parameters W_k. This can be guaranteed due to the fact that multilayer feed-forward neural networks have been shown to be universal function approximators [22]. Thus, for any targeted error ϵ, there exists a K such that

$$\left| \sum_{k=1}^{K} f(W_k M_s) - \ln \Psi(s) \right| < \epsilon. \tag{7}$$

That is, we may bound our approximation error by an arbitrary amount ϵ by tuning the value of K. The natural questions are, How many parameters are necessary, and how does this number depend on system size? We discuss them in detail in the remainder of this paper.

III. LEARNING THE FULLY CONNECTED TFIM GROUND STATE

We proceed by benchmarking the ansatz proposed in the previous section for a particular permutation-invariant Hamiltonian. This class of systems which give rise to mean-field models with permutation symmetry can be cast as long-range spin or boson models [23]. Concretely, we consider the fully connected TFIM, whose Hamiltonian in terms of the Pauli matrices S^ρ ($\rho = x, y, z$) takes the form

$$H = -\frac{J}{L} \sum_{i \neq j} S_i^z S_j^z - g \sum_i S_i^x. \tag{8}$$

This model exhibits a quantum phase transition at $g_c = 2J$ separating a ferromagnetic phase from a paramagnetic phase [24]. This critical point, as well as ground states and excited states, was studied previously in [23,25,26]. In the remainder of this work, we fix the value of $J = 1$.

Although we expect that the choice of activation function is not crucial, we will take as the ansatz the one defined in Eq. (6), with $f(x) = \ln[\cosh(x)]$ being the activation function, which is the choice that maps our model to a restricted Boltzmann machine, which is a common reference for NQSs
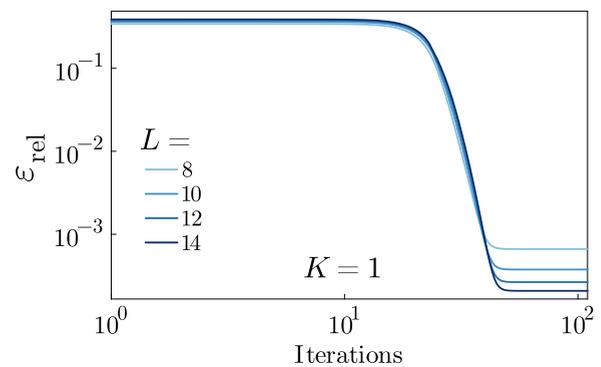
FIG. 2. Relative energy for one variational parameter $K = 1$ under training iterations for several system sizes L. We used exact sampling when averaging over spin configurations.

[10,11,16]. Therefore, the wave function ansatz we will use in the following is given by

$$\ln \Psi_{\mathbf{W}}(s) = \sum_{\ell=1}^{K} \ln \cosh(W_\ell M_s). \tag{9}$$

In order to numerically obtain the ground state of this model we will use stochastic reconfiguration [27,28] to minimize the variational energy $E(W)$ [see Eq. (13)]. It is of particular interest to study how the convergence of the model is affected by the choice of K. That is, for a given system size L how many parameters do we need to converge to the exact ground state?

To evaluate the accuracy of this ansatz for finding the ground state of the fully connected TFIM, we train the model and track the relative energy,

$$\epsilon_{\mathrm{rel}} \equiv \frac{|\langle H\rangle - E_{\mathrm{ED}}|}{E_{\mathrm{ED}}}, \tag{10}$$

where the expectation values are defined as $\langle \cdots \rangle \equiv \langle \Psi_{\mathbf{W}}| \cdots |\Psi_{\mathbf{W}}\rangle$, with appropriate normalization of the wave function, and E_{ED} denotes the ground state energy obtained by exact diagonalization. As a further measure for the accuracy of our obtained wave function, we also study the energy fluctuation density, defined as

$$\sigma^2(H) = \frac{1}{L}[\langle H^2\rangle - \langle H\rangle^2]. \tag{11}$$

If our variational wave function is the targeted ground state, we will have $\sigma^2(H) = 0$ on fundamental grounds.

In Fig. 1(c) we show the results of the ground state search training for several choices of K at a fixed small system size of $L = 12$. The averages were computed by summing over the full Hilbert space basis configurations, which we refer to as exact sampling. We find that as one increases the number of independent variational parameters K, the model is able to decrease the error further. This coincides with our claim emerging from the universal function approximator theorem in Eq. (7), according to which we can reduce our error arbitrarily by taking more parameters K.

At this point, one further question arises: How do the parameter requirements scale with systems size? To address this question, in Fig. 2 we show the relative energy for several

L using only one parameter, $K = 1$. Here, we use exact diagonalization to compute the ground state energy and compare the results to our variational ansatz. It appears that as L grows, the $K = 1$ result improves, signaling that convergence in fact requires fewer parameters as we increase system size. To further study this apparent behavior, in the following section, we will attempt to understand analytically how far we can get with a single-parameter ansatz.

A. Thermodynamic limit

In this section, we demonstrate that the ground state of the fully connected TFIM can be exactly described with only one variational parameter in the thermodynamic limit, building upon the observed trend in finite-size systems in Fig. 2. We argue that this result signals the low complexity of permutation-symmetric models for neural quantum states. By taking the one-parameter case ($K = 1$) we can take the following ansatz:

$$\ln \Psi_W(s) = \ln[\cosh(WM_s)]. \quad (12)$$

Due to the fact that this ansatz has only one variational parameter we can compute quantities such as the magnetization, energy, and energy fluctuations analytically. This will allow us to show the asymptotic convergence to the ground state in the thermodynamic limit. To proceed with this computation we will restrict our system, without loss of generality, to the ferromagnetic phase and assume that L is sufficiently large. Under these assumptions, the typical configurations s are such that $|M_s| \sim L$, and hence, we can approximate the $\ln \cosh$ function as $\ln[\cosh(WM_s)] \approx WM_s$ for $WM_s \sim L \gg 1$. We emphasize that the analytics conducted are applicable to any activation function exhibiting asymptotically linear behavior. This generality in our findings underscores the broader applicability of the results.

Details on all the following computations can be found in the Appendix. We first compute the energy

$$E = \sum_s E_{\text{loc}}(s) \frac{|\langle s|\Psi\rangle|^2}{\langle\Psi|\Psi\rangle}, \quad E_{\text{loc}}(s) \equiv \frac{\langle s|H|\Psi\rangle}{\langle s|\Psi\rangle}, \quad (13)$$

as a function of W [10]. For the energy we obtain

$$E(W) = -J(L-1)\tanh^2(2W)$$
$$+ gL\tanh(2W)\sinh(2W) - gL\cosh(2W). \quad (14)$$

For the ground state we minimize Eq. (14), yielding the value of W in the ground state, which is given by

$$W = \frac{1}{2}\cosh^{-1}\left[\frac{2J}{g}\left(1 - \frac{1}{L}\right)\right]. \quad (15)$$

We can go one step further and relate this single parameter directly to the mean-field order parameter. If we compute the magnetization $M \equiv L^{-1}\sum_i \langle S_i^z\rangle$, we find that the parameter W is related to the magnetization by

$$M = \tanh(2W) = \pm\sqrt{1 - \frac{g^2}{4J^2}}, \quad (16)$$

where the second equality was obtained by substituting the value of W in the ground state equation (15). It is worth mentioning that this equation is identical to the self-consistency

equation for the mean-field solution of the nearest-neighbor Ising model at zero temperature. Then we can identify that $W \sim M$. Hence, it serves as an example of how the network parameters can be directly related to physical quantities, and studying their structure may be beneficial to understand NQS complexity. As a next step, we show that this state is an eigenstate of our Hamiltonian asymptotically in L. To do this, we make use of the energy fluctuations. In the Appendix we obtain as an analytical expression

$$\sigma^2(H) = \frac{g^4 L^2}{8J^2(L-1)^3}. \quad (17)$$

This result indicates that the scale of the energy fluctuations asymptotically vanishes as $1/\sqrt{L}$, implying that, in the thermodynamic limit, our ansatz becomes an exact eigenstate of the Hamiltonian. The one-parameter ansatz also provides plenty of information about the physics of the model. For instance, from Eq. (15) one may also extract the critical g separating the ferromagnetic and paramagnetic regions. This can be detected by noticing that the $\cosh^{-1}$ becomes undefined for $g > 2J$.

IV. BREAKING PERMUTATION SYMMETRY

We now study the effect of deviating from mean field by breaking the permutation symmetry. For this, we will study the convergence to the ground state of the long-range interacting TFIM with periodic boundary conditions (PBCs). The Hamiltonian for this model is given by

$$H = -\frac{J}{\mathcal{N}(L,\alpha)}\sum_{i\neq j}\frac{1}{|i-j|^\alpha}S_i^z S_j^z - g\sum_i S_i^x, \quad (18)$$

where $\mathcal{N}(L,\alpha) \equiv \frac{1}{L-1}\sum_{i\neq j}\frac{1}{|i-j|^\alpha}$ is the so-called Kac normalization factor used to ensure that the energy is extensive. We take $|i-j|$ to be the minimum distance between two lattice sites under PBCs. When $\alpha = 0$, the model is equivalent to the fully connected TFIM studied before, while as $\alpha \to \infty$, the model reduces to the nearest-neighbor transverse-field Ising model. This model can be realized experimentally in trapped-ion simulators for $0 \leqslant \alpha \lesssim 3.5$ [29–35] and is known to go through several phase transitions as a function of α [36–57]. This setup is ideal for our purposes because we can adjust the deviation from mean field upon tuning the value of α, observing the eventual breakdown of our ansatz.

To verify the convergence to the ground state of this model as we vary the interaction range α, we compute the energy fluctuation density $\sigma^2(H)$ of the one-parameter ansatz. We deduce an analytical expression which allows us to establish its system-size dependence:

$$\sigma^2(H) = g^2\tanh^2(2W) - 4Jg\left(1 - \frac{1}{L}\right)\frac{\tanh^2(2W)}{\cosh(2W)}$$
$$+ \frac{4J^2[\tanh^2(2W) - \tanh^4(2W)]}{\mathcal{N}(L,\alpha)^2 L}$$
$$\times \sum_{i\neq j\neq k}\frac{1}{|i-j|^\alpha|j-k|^\alpha}$$
$$+ \frac{2J^2[1 - \tanh^4(2W)]}{\mathcal{N}(L,\alpha)^2 L}\sum_{i\neq j}\frac{1}{|i-j|^{2\alpha}}. \quad (19)$$
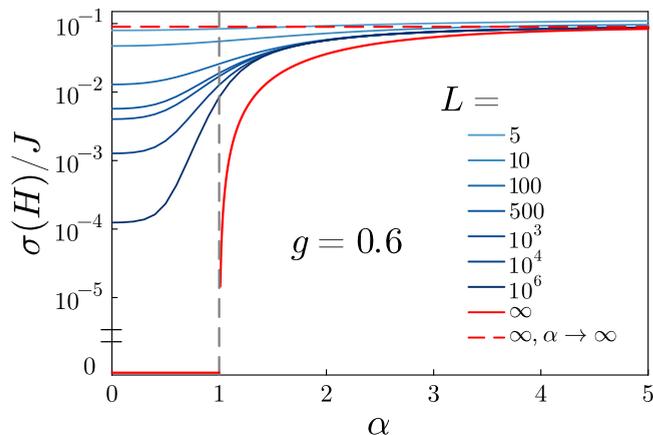
FIG. 3. Normalized energy fluctuations of the long-range interacting transverse-field Ising model are shown as a function of interaction range α for various system sizes. The fluctuations are normalized by the coupling J due to the inaccessibility of the gap in the thermodynamic limit for nonzero values of α. In red we show the limiting curve in the thermodynamic limit.

The details of this computation can be found in the Appendix. Given that we are interested in the thermodynamic limit behavior, we go ahead and rewrite the two remaining sums in terms of the generalized harmonic numbers $H_{n,r} \equiv \sum_{k=1}^{n} k^{-r}$. For $r > 1$, we make use of the fact that, in the limit when $n \to \infty$, $H_{n,r}$ converges to the Riemann zeta function defined as $\zeta(r) \equiv \sum_{k=1}^{\infty} k^{-r}$. This allows us to obtain the energy fluctuations in the thermodynamic limit directly as

$$\sigma^2(H) = \frac{g^4}{16J^2} \begin{cases} \frac{H_{\infty,2\alpha}}{H_{\infty,\alpha}^2}, & \text{if } \alpha \leqslant 1, \\ \frac{\zeta(2\alpha)}{\zeta(\alpha)^2}, & \text{if } \alpha > 1. \end{cases} \tag{20}$$

By analyzing the generalized harmonic series ratio convergence as a function of α in Eq. (20), we find that, in general, for $\alpha \leqslant 1$ the energy fluctuations will vanish. This result implies that even though we have broken the permutation symmetry, our ansatz is still able to exactly approach the ground state for large enough systems, which is remarkable.

For the case of $\alpha > 1$, we no longer have diverging sums and can write the entire expression as a function of the ratio of the Riemann zeta functions. With this, we have the precise scaling with which the energy fluctuations decay as a function of α in the thermodynamic limit. In Fig. 3, we plot the expression σ^2 for different system sizes. Remarkably, we observe that the fluctuations vanish asymptotically as $1/\sqrt{L}$ for $\alpha \leqslant 1/2$ precisely, like for the $\alpha = 0$ mean-field case. In contrast, for $1/2 < \alpha \leqslant 1$ the fluctuations still vanish asymptotically, but with a slower decay. For this last interval the decays at the edges ($\alpha = 1/2$, 1) follow the scalings $\sqrt{\ln(L)/L}$ and $\sqrt{\zeta(2)}/\ln(L)$, respectively. For values of $\alpha > 1$, we also plot the limiting curve, where we see that the energy fluctuations steadily increase until they saturate to an upper bound value of $g^2/4J$. This value comes from evaluating the energy fluctuation density at the ground state with the W given by Eq. (15).

The one-parameter ansatz consequently exactly captures the ground state for a finite range of α values. Interestingly, we find that the one-parameter ansatz can exactly describe quantum states even without permutation symmetry in the thermodynamic limit. Furthermore, it indicates that the long-range interacting TFIM has a simplified ground state within the range $0 < \alpha \leqslant 1$.

This type of long-range Ising model is known to exhibit quantum phase transitions as a function of α and a Kosterlitz-Thouless transition at $\alpha = 1$, and for $\alpha \geqslant 2$ one recovers short-range behavior [29,36–38]. However, the details of the ground state properties in particular in the regime of low α have not been fully settled. For instance, Ref. [37] found that for even larger $\alpha < 5/3$ the ground state might be described by a long-range mean-field theory. It is unclear, however, to what extent the utilized ϵ expansion in this high-α regime remains accurate.

For a comprehensive study of the prior findings related to this model, we would like to further direct the reader's attention to the recent review in Ref. [58]. It is essential to emphasize that while our conclusion of mean-field type behavior may appear intuitive, there has been a notable absence of formal proof establishing this model's ground state as a one-parameter simplified ansatz. While previous computations, employing linear spin wave theory, for instance, indeed yielded results consistent with our own observations, it is worth emphasizing that a crucial difference has also remained. Within linear spin wave theory it has been shown that no inconsistencies are generated [58], which likely makes the approach reliable, but it is still based on some unproven assumptions. Here, with our approach no such assumption is required, providing us with a conclusive and rigorous proof.

V. CONCLUSION

In this work, we investigated a class of models that demonstrate minimal complexity within the framework of NQSs in the sense of the number of parameters required to describe the ground state. By imposing permutation symmetry on our networks, we found that NQSs are particularly well suited to describing permutation-invariant mean-field states. To validate this claim, we investigated the ground state of the fully connected TFIM. In particular, we showed that, even for finite systems, the ground state can be accurately captured using a small number of parameters and this approximation improves as the system size increases.

An important advantage of our approach is the ability to capture finite-size effects, entanglement and quantum correlations, as it does not necessarily assume a product state. This is particularly relevant; as discussed in detail, for instance, in Ref. [23], permutation symmetry does not always imply a product state structure, especially for finite-size systems or for dynamics. For the latter case it is also very important to note that such permutation-invariant Hamiltonians are key for the creation of spin squeezed states through one-axis twisting [59]. Let us note, however, that we furthermore showed that product states, such as those observed in the thermodynamic limit of our model, can still be achieved. Specifically, by opting for an activation function which behaves linearly at least asymptotically, we can witness the emergence of a product state structure for the ground state in the thermodynamic limit.

Motivated by these observations, we proposed a one-parameter ansatz and analytically showed that the ground state can be described by a single variational parameter in the thermodynamic limit. Furthermore, we examined the impact of breaking permutation symmetry by studying the ground state of the long-range interacting TFIM and tuning the interaction range. Surprisingly, we found that the one-parameter ansatz remains effective for values of α up to 1, suggesting that the model is still described by the mean field in this regime and robustness in the presence of weak symmetry breaking.

A potential direction for future research lies in exploring the possibility of learning potential simplifications, such as the one discussed in this work for permutation symmetry, within the NQS parametrization directly from the network parameters. Such an approach may have the ability to shed light on the overall complexity of the network for more complicated scenarios. Additionally, an intriguing angle to consider is whether we can uncover underlying physics in the network parameters, such as the relation found here between the one-parameter model and magnetization. In other words, can we demystify the black-box nature of the neural network and gain insights into the specific physical principles it exploits to represent the quantum state effectively?

While our work has shed some light on which quantum states are easy to capture with NQSs, it remains an open question to identify a general and physical characterization of states which are difficult for NQSs. Recent developments have highlighted that even ground states of frustrated quantum magnets appear to be manageable within NQSs [12], although very deep networks seem to be necessary in order to achieve quantitative accuracy. This suggests that the intricate sign structure in such frustrated quantum magnets might be what is difficult to represent within NQSs. This would apply equivalently to fermionic quantum matter. However, this question will have to be explored in further detail in order to eventually arrive at a physical understanding of NQS complexity in the end.

The data shown in the figures are available from Zenodo [60].

APPENDIX: ANALYTICAL COMPUTATIONS

1. Ground state energy

In this Appendix, we provide further details on the analytical computations discussed throughout the text. We start by recalling our model Hamiltonian and normalization factor:

$$H = \frac{-J}{\mathcal{N}(L, \alpha)} \sum_{i \neq j} \frac{1}{|i - j|^\alpha} S_i^z S_j^z - g \sum_i S_i^x, \quad \text{(A1)}$$

$$\mathcal{N}(L, \alpha) = \frac{1}{L - 1} \sum_{i \neq j} \frac{1}{|i - j|^\alpha}. \quad \text{(A2)}$$

We are interested in computing the local energy, which is defined as $E_{\text{loc}}(s) = \frac{\langle s|H|\Psi \rangle}{\langle s|\Psi \rangle}$. For our model, the local energy takes the form

$$E_{\text{loc}}(s) = \frac{-J}{\mathcal{N}(L, \alpha)} \sum_{i \neq j} \frac{s_i s_j}{|i - j|^\alpha} - g \sum_i \frac{\langle s'(s_i)|\Psi \rangle}{\langle s|\Psi \rangle}, \quad \text{(A3)}$$

where $s'(s_i)$ means we have flipped the spin in the ith site. Using the fact that $\ln\langle s|\Psi \rangle = \ln \cosh(W \sum_i s_i)$ from our ansatz, we can rewrite the second term as

$$E_{\text{loc}}(s) = \frac{-J}{\mathcal{N}(L, \alpha)} \sum_{i \neq j} \frac{s_i s_j}{|i - j|^\alpha}$$
$$- g \sum_i \exp \left\{ \ln \cosh \left(W \sum_i s_i' \right) \right.$$
$$\left. - \ln \cosh \left(W \sum_i s_i \right) \right\}. \quad \text{(A4)}$$

In the ferromagnetic phase, we can use the approximation $\ln \cosh x \approx x$ for $x \sim L \gg 1$. Making this approximation, the second term reduces to

$$E_{\text{loc}}(s) = \frac{-J}{\mathcal{N}(L, \alpha)} \sum_{i \neq j} \frac{s_i s_j}{|i - j|^\alpha} - g \sum_i e^{-2W s_i}. \quad \text{(A5)}$$

To eliminate the exponential term and have all dependence on s_i be polynomial, we can make use of the following identity:

$$e^{-2W s_i} = \cosh(2W) - \sinh(2W)s_i, \quad \text{for } s_i = \pm 1. \quad \text{(A6)}$$

This simplification may not immediately be obvious, but it will be helpful when we sum over all spin configurations. Using this identity, we can simplify our local energy expression to the form

$$E_{\text{loc}}(s) = \frac{-J}{\mathcal{N}(L, \alpha)} \sum_{i \neq j} \frac{s_i s_j}{|i - j|^\alpha} - gL \cosh(2W)$$
$$+ g \sinh(2W) \sum_i s_i. \quad \text{(A7)}$$

With this local energy, we can compute the total energy of the system as follows:

$$E = \sum_{\mathbf{S}} E_{\text{loc}}(s) \frac{|\langle s|\Psi\rangle|^2}{N}; \quad N \equiv \sum_{\mathbf{S}} |\langle s|\Psi\rangle|^2. \tag{A8}$$

The normalization can readily be computed from our ansatz and the exponential term simplification,

$$N = |\langle s|\Psi\rangle|^2 = \sum_{\mathbf{S}} e^{2W \sum_i s_i} = \sum_{\mathbf{S}} \prod_i e^{2W s_i} = \sum_{\mathbf{S}} \prod_i \cosh(2W) - g \sinh(2W) s_i = \cosh^L(2W) 2^L. \tag{A9}$$

Analogously, by substituting into the equation for the total energy, we derive the subsequent expression

$$E = \frac{1}{\cosh^L(2W)2^L} \left[\frac{-J}{\mathcal{N}(L,\alpha)} \sum_{i \neq j} \frac{1}{|i-j|^\alpha} \sum_{\mathbf{S}} s_i s_j \prod_k [\cosh(2W) + \sinh(2W) s_k] \right.$$
$$\left. + g \sinh(2W) \sum_{\mathbf{S}} \sum_i s_i \prod_k [\cosh(2W) + \sinh(2W) s_k] \right] - gL \cosh(2W). \tag{A10}$$

The importance of the polynomial dependence in s_i becomes evident in this expression. Specifically, in the first term, two terms in the product will interact with the external s_i and s_j, respectively, namely, $[\cosh(2W)s_i + \sinh(2W)]$ and $[\cosh(2W)s_j + \sinh(2W)]$. Thus, the first term contains two terms that give $\sinh(2W)$ and $L - 2$ terms that give $\cosh(2W)$, so that the overall expression simplifies to

$$E = \frac{1}{\cosh^L(2W)2^L} \left[\frac{-J}{\mathcal{N}(L,\alpha)} \sum_{i \neq j} \frac{1}{|i-j|^\alpha} 2^L \sinh^2(2W) \cosh^{L-2}(2W) \right.$$
$$\left. + g \sinh(2W) \sum_i 2^L \sinh(2W) \cosh^{L-1}(2W) \right] - gL \cosh(2W). \tag{A11}$$

Having successfully summed over all spin configurations, we arrive at the final expression for the energy:

$$E = -J(L-1)\tanh^2(2W) + gL \tanh(2W)\sinh(2W) - gL \cosh(2W). \tag{A12}$$

The minimization procedure is straightforward from here:

$$\frac{\partial E}{\partial W} = 0 = [1 - \tanh^2(2W)] \underbrace{[-2J(L-1)\tanh(2W) + gL\sinh(2W)]}_{\cosh(2W) = \frac{2J}{gL}(L-1)}. \tag{A13}$$

2. Energy fluctuations

The other quantity we need to compute analytically is the energy fluctuation density,

$$\sigma(H)^2 = \frac{1}{L}[\langle H^2 \rangle - \langle H \rangle^2]. \tag{A14}$$

For the fluctuations, all that remains to be computed is the first term $\langle H^2 \rangle$ since the second term is precisely the energy computed above $\langle H \rangle = E$. In particular, a good simplification we can perform to rewrite the H^2 term in terms of the local energy is

$$\langle H^2 \rangle = \frac{1}{N} \sum_{\mathbf{S}} E_{\text{loc}}(s)^2 |\langle s|\Psi\rangle|^2. \tag{A15}$$

The form of $E_{\text{loc}}(s)^2$ is given by

$$E_{\text{loc}}(s)^2 = \frac{J^2}{\mathcal{N}(L,\alpha)^2} \sum_{i \neq j, n \neq k} \frac{s_i s_j s_n s_k}{|i-j|^\alpha |n-k|^\alpha} + \frac{2JgL}{\mathcal{N}(L,\alpha)} \cosh(2W) \sum_{i \neq j} \frac{s_i s_j}{|i-j|^\alpha} + g^2 L^2 \cosh^2(2W)$$
$$- \frac{2Jg}{\mathcal{N}(L,\alpha)} \sinh(2W) \sum_{i \neq j, k} \frac{s_i s_j s_k}{|i-j|^\alpha} - 2g^2 L \cosh(2W)\sinh(2W) \sum_i s_i + g^2 \sinh^2(2W) \sum_{i \neq j} s_i s_j. \tag{A16}$$

Now, to sum over spin configurations it is convenient to split the sums into sums in which all the indices are different. To give an example of this we consider the expansion of the first term in Eq. (A16),

$$\sum_{i \neq j, n \neq k} \frac{s_i s_j s_n s_k}{|i-j|^\alpha |n-k|^\alpha} = \sum_{i \neq j \neq n \neq k} \frac{s_i s_j s_n s_k}{|i-j|^\alpha |n-k|^\alpha} + 4 \sum_{i \neq j \neq k} \frac{s_i s_k}{|i-j|^\alpha |j-k|^\alpha} + 2 \sum_{i \neq j} \frac{1}{|i-j|^{2\alpha}}. \tag{A17}$$

Analogously, we may rewrite all sums in this way to apply the same method to sum over spin configurations as the one used in the energy calculation. If we follow this process, we find that the value of the expectation value of H^2 is

$$\langle H^2 \rangle = \frac{J^2}{\mathcal{N}(L,\alpha)^2} \left[\tanh^4(2W) \sum_{i \neq j \neq n \neq k} \frac{1}{|i-j|^\alpha |n-k|^\alpha} + 4\tanh^2(2W) \sum_{i \neq j \neq k} \frac{1}{|i-j|^\alpha |j-k|^\alpha} + 2 \sum_{i \neq j} \frac{1}{|i-j|^{2\alpha}} \right]$$

$$+ 2JgL(L-1)\cosh(2W)\tanh^2(2W) - 2Jg(L-1)\sinh(2W)[\tanh^3(2W)(L-2) + 2\tanh(2W)] + g^2 L^2 \cosh^2(2W)$$

$$- 2g^2 L^2 \sinh^2(2W) + g^2 \sinh^2(2W)\tanh^2(2W)L^2 - g^2 \sinh^2(2W)\tanh^2(2W)L + Lg^2 \sinh^2(2W). \tag{A18}$$

With Eqs. (A12) and (A18) we can now compute the fluctuations directly as a function of the variational parameter W. To reach the specific expression in Eq. (19) we reuse the identity in (A17), but now to get rid of the 4 index sum. It is also convenient to write all the sums in terms of the generalized harmonic numbers. To do that we use the following two identities for odd L (the even case can be obtained by subtracting a correction term due to overcounting):

$$\sum_{i \neq j} \frac{1}{|i-j|^\alpha} = 2L H_{\lfloor \frac{L}{2} \rfloor, \alpha}, \tag{A19}$$

$$\sum_{i \neq j \neq k} \frac{1}{|i-j|^\alpha |j-k|^\alpha} = L\left(4H^2_{\lfloor \frac{L}{2} \rfloor, \alpha} - 2H_{\lfloor \frac{L}{2} \rfloor, 2\alpha} \right). \tag{A20}$$

3. Magnetization

To compute the magnetization as expressed in Eq. (16), we calculate the expected value of the S_i^z operator with the one-parameter ansatz. The calculation is performed as follows:

$$\langle S_i^z \rangle = \frac{1}{N} \sum_{\mathbf{s}} s_i \, |\langle s|\Psi \rangle|^2$$

$$= \frac{1}{N} \sum_{\mathbf{s}} s_i \prod_k [\cosh(2W) + \sinh(2W)s_k]$$

$$= \tanh(2W). \tag{A21}$$

[1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[2] S. R. White and A. E. Feiguin, Real-time evolution using the density matrix renormalization group, Phys. Rev. Lett. **93**, 076401 (2004).

[3] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, Phys. Rev. Lett. **93**, 040502 (2004).

[4] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[5] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Quantum Monte Carlo methods for nuclear physics, Rev. Mod. Phys. **87**, 1067 (2015).

[6] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[7] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, Sign problem in the numerical simulation of many-electron systems, Phys. Rev. B **41**, 9301 (1990).

[8] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Inf. Comput. **7**, 401 (2007).

[9] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Phys. Rev. B **73**, 094423 (2006).

[10] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[11] M. Schmitt and M. Heyl, Quantum many-body dynamics in two dimensions with artificial neural networks, Phys. Rev. Lett. **125**, 100503 (2020).

[12] A. Chen and M. Heyl, Efficient optimization of deep neural quantum states toward machine precision, arXiv:2302.01941.

[13] T. Mendes-Santos, M. Schmitt, and M. Heyl, Highly resolved spectral functions of two-dimensional systems with neural quantum states, Phys. Rev. Lett. **131**, 046501 (2023).

[14] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[15] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J_1-J_2 model studied with neural network quantum states, Phys. Rev. B **100**, 125124 (2019).

[16] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, Phys. Rev. X **7**, 021021 (2017).

[17] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, Nat. Commun. **9**, 5322 (2018).

[18] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, Phys. Rev. B **106**, 205136 (2022).

[19] M. Norouzi, M. Ranjbar, and G. Mori, Stacks of convolutional restricted Boltzmann machines for shift-invariant feature learning, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2009), pp. 2735–2742.

[20] K. Sohn and H. Lee, Learning invariant representations with local transformations, arXiv:1206.6418.

[21] R. M. Gray, Toeplitz and circulant matrices: A review, Found. Commun. Inf. Theory **2**, 155 (2005).

[22] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[23] B. Sciolla and G. Biroli, Quantum quenches, dynamical transitions, and off-equilibrium quantum criticality, Phys. Rev. B **88**, 201110(R) (2013).

[24] Due to our definition of the Hamiltonian the critical parameter ratio looks like this. In other conventions it is reported to be $J = g$, which is equivalent.

[25] A. Das, K. Sengupta, D. Sen, and B. K. Chakrabarti, Infinite-range Ising ferromagnet in a time-dependent transverse magnetic field: Quench and ac dynamics near the quantum critical point, Phys. Rev. B **74**, 144423 (2006).

[26] A. Sehrawat, C. Srivastava, and U. Sen, Dynamical phase transitions in the fully connected quantum Ising model: Time period and critical time, Phys. Rev. B **104**, 085105 (2021).

[27] S. Sorella, Generalized Lanczos algorithm for variational quantum Monte Carlo, Phys. Rev. B **64**, 024512 (2001).

[28] S. Sorella, Green function Monte Carlo with stochastic reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).

[29] P. Hauke and L. Tagliacozzo, Spread of correlations in long-range interacting quantum systems, Phys. Rev. Lett. **111**, 207202 (2013).

[30] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, Quasiparticle engineering and entanglement propagation in a quantum many-body system, Nature (London) **511**, 202 (2014).

[31] A. Friedenauer, H. Schmitz, J. T. Glueckert, D. Porras, and T. Schaetz, Simulating a quantum magnet with trapped ions, Nat. Phys. **4**, 757 (2008).

[32] R. Islam, E. Edwards, K. Kim, S. Korenblit, C. Noh, H. Carmichael, G.-D. Lin, L.-M. Duan, C.-C. J. Wang, J. Freericks, and C. Monroe, Onset of a quantum phase transition with a trapped ion quantum simulator, Nat. Commun. **2**, 377 (2011).

[33] C. Schneider, D. Porras, and T. Schaetz, Experimental quantum simulations of many-body physics with trapped ions, Rep. Prog. Phys. **75**, 024401 (2012).

[34] B.-W. Li, Y.-K. Wu, Q.-X. Mei, R. Yao, W.-Q. Lian, M.-L. Cai, Y. Wang, B.-X. Qi, L. Yao, L. He, Z.-C. Zhou, and L.-M. Duan, Probing critical behavior of long-range transverse-field Ising model through quantum Kibble-Zurek mechanism, PRX Quantum **4**, 010302 (2023).

[35] D. Porras and J. I. Cirac, Effective quantum spin systems with trapped ions, Phys. Rev. Lett. **92**, 207901 (2004).

[36] T. Koffel, M. Lewenstein, and L. Tagliacozzo, Entanglement entropy for the long-range Ising chain in a transverse field, Phys. Rev. Lett. **109**, 267203 (2012).

[37] A. Dutta and J. K. Bhattacharjee, Phase transitions in the quantum Ising and rotor models with a long-range interaction, Phys. Rev. B **64**, 184106 (2001).

[38] F. J. Dyson, Existence of a phase transition in a one-dimensional Ising ferromagnet, Commun. Math. Phys. **12**, 91 (1969).

[39] V. Popkov, M. Salerno, and G. Schütz, Entangling power of permutation-invariant quantum states, Phys. Rev. A **72**, 032327 (2005).

[40] M. Knap, D. A. Abanin, and E. Demler, Dissipative dynamics of a driven quantum spin coupled to a bath of ultracold fermions, Phys. Rev. Lett. **111**, 265302 (2013).

[41] F. Pan and J. Draayer, Analytical solutions for the LMG model, Phys. Lett. B **451**, 1 (1999).

[42] L. Affonso, R. Bissacot, E. O. Endo, and S. Handa, Long-range Ising models: Contours, phase transitions and decaying fields, arXiv:2105.06103.

[43] M. Kac and C. J. Thompson, Critical behavior of several lattice models with long range interaction, J. Math. Phys. **10**, 1373 (1969).

[44] Y. Nomura, Investigating network parameters in neural-network quantum states, J. Phys. Soc. Jpn. **91**, 054709 (2022).

[45] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, Phys. Rev. Lett. **122**, 065301 (2019).

[46] G. Jin, X. Yi, L. Zhang, L. Zhang, S. Schewe, and X. Huang, How does weight correlation affect the generalisation ability of deep neural networks, arXiv:2010.05983.

[47] D. Wu *et al.*, Variational benchmarks for quantum many-body problems, arXiv:2302.04919.

[48] E. Granet, Exact mean-field solution of a spin chain with short-range and long-range interactions, SciPost Phys. **14**, 133 (2023).

[49] H. P. Casagrande, B. Xing, M. Dalmonte, A. Rodriguez, V. Balachandran, and D. Poletti, Complexity of spin configurations dynamics due to unitary evolution and periodic projective measurements, Phys. Rev. E **108**, 044128 (2023).

[50] S. Fey and K. P. Schmidt, Critical behavior of quantum magnets with long-range interactions in the thermodynamic limit, Phys. Rev. B **94**, 075156 (2016).

[51] M. E. Fisher, S.-K. Ma, and B. G. Nickel, Critical exponents for long-range interactions, Phys. Rev. Lett. **29**, 917 (1972).

[52] D. Vodola, L. Lepori, E. Ercolessi, and G. Pupillo, Long-range Ising and Kitaev models: Phases, correlations and edge modes, New J. Phys. **18**, 015001 (2015).

[53] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[54] M. C. Tran, A. Y. Guo, C. L. Baldwin, A. Ehrenberg, A. V. Gorshkov, and A. Lucas, Lieb-robinson light cone for power-law interactions, Phys. Rev. Lett. **127**, 160401 (2021).

[55] Z. Eldredge, Z.-X. Gong, J. T. Young, A. H. Moosavian, M. Foss-Feig, and A. V. Gorshkov, Fast quantum state transfer and entanglement renormalization using long-range interactions, Phys. Rev. Lett. **119**, 170503 (2017).

[56] C.-Y. Park and M. J. Kastoryano, Geometry of learning neural quantum states, Phys. Rev. Res. **2**, 023232 (2020).

[57] A. Russomanno, M. Fava, and M. Heyl, Quantum chaos and ensemble inequivalence of quantum long-range Ising chains, Phys. Rev. B **104**, 094309 (2021).

[58] N. Defenu, A. Lerose, and S. Pappalardi, Out-of-equilibrium dynamics of quantum many-body systems with long-range interactions, arXiv:2307.04802.

[59] M. Kitagawa and M. Ueda, Squeezed spin states, Phys. Rev. A **47**, 5138 (1993).

[60] F. B. Trigueros, T. M. Santos, and M. Heyl, Mean-field theories are simple for neural quantum states, Zenodo (2023), doi:10.5281/zenodo.8252947.

[61] P. Weinberg and M. Bukov, QuSpin: A python package for dynamics and exact diagonalisation of quantum many body systems part I: Spin chains, SciPost Phys. **2**, 003 (2017).

[62] M. Schmitt and M. Reh, jVMC: Versatile and performant variational Monte Carlo leveraging automated differentiation and GPU acceleration, SciPost Phys. Codebases 2-r0.1 (2022).

[63] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of PYTHON+NUMPY programs, 2018.