Ground state properties of quantum skyrmions described by neural network quantum states

Ashish Joshi [1,*] Robert Peters,[1] and Thore Posske [2,3]

[1]*Department of Physics, Kyoto University, Kyoto 606-8502, Japan*
[2]*I. Institute for Theoretical Physics, Universität Hamburg, Notkestraße 9, 22607 Hamburg, Germany*
[3]*The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany*

We investigate the ground state properties of quantum skyrmions in a ferromagnet using variational Monte Carlo with the neural network quantum state as variational ansatz. We study the ground states of a two-dimensional quantum Heisenberg model in the presence of the Dzyaloshinskii-Moriya interaction (DMI) and show that the ground state accommodates a quantum skyrmion for a large range of parameters, especially at large DMI. The spins in these quantum skyrmions are weakly entangled, and the entanglement increases with decreasing DMI. We also find that the central spin is completely disentangled from the rest of the lattice, establishing a non-destructive way of detecting this type of skyrmion by local magnetization measurements. While neural networks are well suited to detect weakly entangled skyrmions with large DMI, they struggle to describe skyrmions in the small DMI regime due to nearly degenerate ground states and strong entanglement. In this paper, we propose a method to identify this regime and a technique to alleviate the problem. Finally, we analyze the workings of the neural network and explore its limits by pruning. Our work shows that neural network quantum states can be efficiently used to describe the quantum magnetism of large systems exceeding the size manageable in exact diagonalization by far.

I. INTRODUCTION

Classical magnetic skyrmions are magnetic structures with vortex-like configurations characterized by a quantized skyrmion number. After pioneering theoretical work [1,2], skyrmions have been discovered in a variety of materials, including MnSi, FeCoSi, FeGe, and others [3–7], with sizes ranging from micrometers to nanometers. These quasiparticles can be created by competition between the exchange interaction and the anti-symmetric Dzyaloshinskii-Moriya interaction (DMI) [7], or by frustration [8], or dynamically by an electric current [9,10] or by boundary effects [11]. Magnetic skyrmions have potential uses in magnetic storage devices due to their topological protection and ease of motion under electric currents [12–16]. The observation of skyrmions with sizes a few times the atomic lattice spacing raises the question about the importance of quantum effects in these systems, meriting a purely quantum mechanical analysis to study them.

In the past few years, there have been some works addressing the quantum nature of magnetic skyrmions, with the focus on quantum spin systems in the presence of DMI or frustration [8,11,17–21], or very recently in systems with itinerant magnetism [22] and with f-electron systems [23]. Most studies have used exact diagonalization techniques to tackle the problem of quantum skyrmions in spin lattices, which inherently puts a limit on the system size they can consider. Recently, quantum skyrmions on a larger spin lattice were considered by Haller *et al.* [19], using density matrix renormalization group (DMRG) methods. This work discovered a skyrmion lattice phase that would not be tractable using exact diagonalization. Although having many successful applications, DMRG becomes numerically challenging in two dimensions due to increasing entanglement with the system sizes described by the area law [24,25]. On the other hand, due to the presence of DMI, quantum Monte Carlo methods suffer from the sign problem, which slows down the optimization process.

In recent years, artificial neural network-based variational methods have been introduced to approximate the quantum many-body problem, achieving results comparable to state-of-the-art methods [26–34]. In these variational methods, an artificial neural network is used to represent the variational wave function, known as a neural network quantum state (NQS), which then learns the target state using a gradient-based optimization scheme. NQS-based variational methods offer a novel approach to studying a wide range of quantum many-body systems, especially in two and three dimensions where existing methods involve a high level of complexity. NQSs with various structures have been successfully applied to frustrated spin systems in two dimensions [26,30–32,35] and have recently been shown to have the ability to capture long-range quantum entanglement with an expressive capacity greater than conventional methods [36]

In this paper, using NQSs we show that quantum skyrmions (QSs) are the ground states for a wide range of parameters in

*joshi.ashish.42a@st.kyoto-u.ac.jp

the two-dimensional spin-1/2 Heisenberg Hamiltonian with DMI in a ferromagnetic medium. To study quantum entanglement in this system, we calculate Renyi entropy of second order and demonstrate that the entanglement in the QS ground state decreases with increasing DMI. Previous work indicated that the central spin of a quantum skyrmion can have vanishing concurrence with its surrounding spins [19]. Interestingly, we also find that the central spin in the QS ground state is completely disentangled from the rest of the spins within the error bars of our method. This opens up a way of detecting quantum skyrmions experimentally without destroying their quantum nature. While we find stable QSs at large DMI, the variational method is insufficient to learn the ground state wave function at small DMI. An analysis of small systems reveals that the variational method finds a superposition of the ground state and the first excited state due to a tiny excitation gap. Motivated by this, we present a projection-based method to improve the variational ground state in this region. Finally, we analyze the internal structure of our NQS ansatz by inspecting the trained network weights and pruning. While the lowly entangled NQS does not change significantly upon pruning, the performance degrades rapidly with pruning in the highly entangled NQS. Our work shows that an NQS variational ansatz can be used to efficiently approximate spin systems with medium to high DMI at system sizes out of reach for exact methods.

The paper is organized as follows. In Sec. II, we describe the Heisenberg Hamiltonian on a square lattice in the presence of DMI. In Sec. III, we briefly discuss the variational method, along with the neural network structure, with more details in Appendix A. In Sec. IV, we study the different ground states of this system. In Sec. V, we study the entanglement in the ground state by calculating the Renyi entropies. In Sec. VI, we obtain and present insights into the workings of the trained NQS. Finally, we conclude the paper in Sec. VII.

II. MODEL

We study the ground state of the two-dimensional spin-1/2 Heisenberg Hamiltonian on a square lattice in the presence of the Dzyaloshinskii-Moriya interaction (DMI) and a strong external magnetic field at the boundaries that simulates a ferromagnetic background,

$$H = -J \sum_{\langle ij \rangle} \left(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y \right) - A \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z$$
$$- D \sum_{\langle ij \rangle} (u_{ij} \times \hat{z}) \cdot (\sigma_i \times \sigma_j) + B^z \sum_b \sigma_b^z. \quad (1)$$

Here, J is the Heisenberg exchange interaction, A is the Heisenberg anisotropy term, D is the DMI, and B^z is the external magnetic field along the $\hat{z}$ axis acting only on the boundary spins indexed with b. The Pauli operator at the ith lattice site is $\sigma_i = (\sigma_i^x, \sigma_i^y, \sigma_i^z)$ and u_{ij} is the unit vector pointing from site i to site j. We consider $\hbar = 1$. The sum in the first three terms is over the nearest neighbor lattice sites, while the last term only covers the boundary sites.

The Hamiltonian in Eq. (1) can be considered the quantum analog of a classical spin model in which the competition between the noncolinear DMI and the ferromagnetic Heisenberg
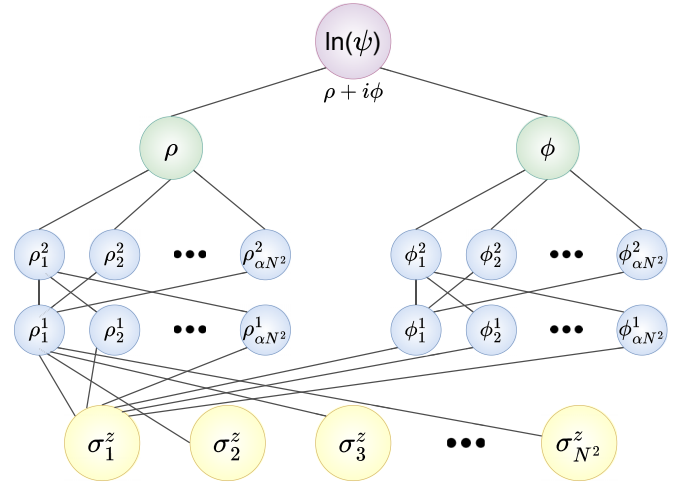

FIG. 1. Neural network structure used as NQS. The inputs are the spin configurations in the σ^z basis, and the output is the logarithm of the wave function. There are two fully connected networks, with two hidden layers in each, to learn the phase and the amplitude part of the wave function separately. Each hidden layer consists of αN^2 neurons.

exchange interaction gives rise to the formation of magnetic skyrmions. In classical systems, these skyrmions are often stabilized by an external magnetic field over the whole lattice. However, we only apply the magnetic field ($B^z = 10J$) to fix the spins at the boundaries. This is the main difference between the Hamiltonian in our work and that in Ref. [19], where the authors study a similar system but with a bulk external magnetic field. Thus our model describes a single quantum skyrmion embedded in a ferromagnetic medium. We leave the study of a quantum skyrmion lattice with NQS and the comparison with DMRG results [19] for future works.

III. METHOD

The idea behind neural network quantum states is to use the output of an artificial neural network to represent the complex-valued coefficients $\psi_\theta(\sigma)$ in the variational wave function,

$$|\psi_\theta\rangle = \sum_\sigma \psi_\theta(\sigma)|\sigma\rangle. \quad (2)$$

Here, $|\sigma\rangle$ are local basis states, which in our case are the eigenstates of the σ^z operators, and θ are the variational parameters of the neural network. In this paper, we use two fully connected feed-forward neural networks to each represent the phase and modulus part of the wave function, see Fig. 1, and take the logarithm of the wave function as the total output [30]

$$\ln(\psi_\theta(\sigma)) = \rho(\sigma) + i\phi(\sigma). \quad (3)$$

The network takes the configuration of the spins on the two-dimensional lattice as the input. Both the phase and the modulus part of the network consist of two fully connected hidden layers with αN^2 neurons in each layer, where N is the length of one side of the lattice, and we use $\alpha = 2$ in this paper. We use the rectified linear unit (reLU) as the nonlinear activation function. The optimization of the variational wave function is achieved by minimizing the loss function L_θ, i.e., the variational energy, with respect to the variational
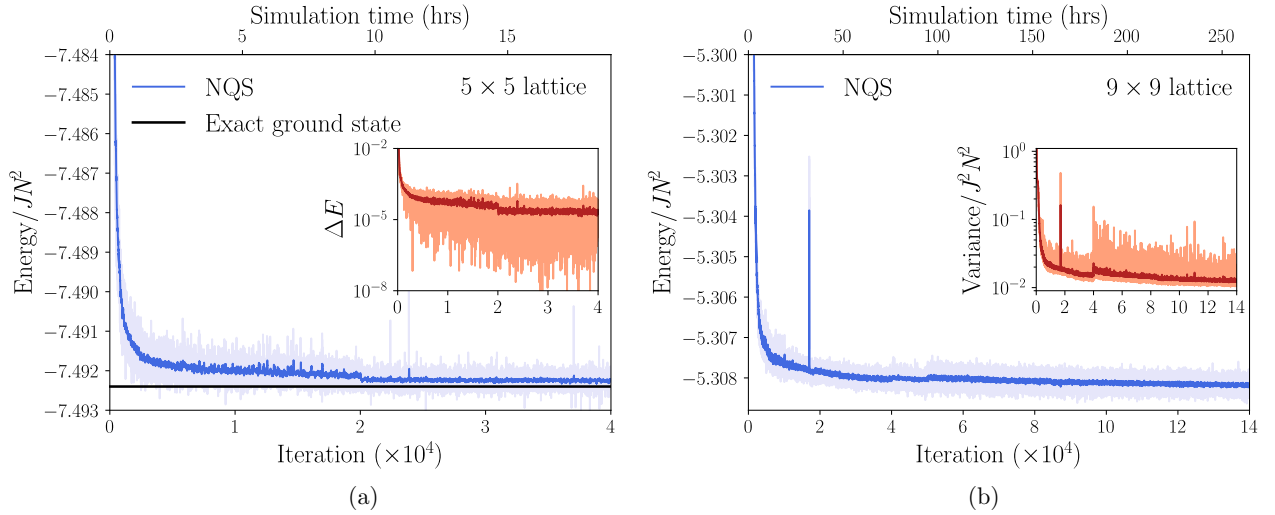
FIG. 2. Convergence of the NQS training procedure: The figure shows the convergence of variational energy per spin to the ground state of a 5×5 lattice (a) and a 9×9 lattice (b) over the number of iterations at the bottom axis and time elapsed at the top axis (see A for hardware specifications). The inset in (a) shows the relative error ΔE in the ground state energy [see Eq. (5)] with respect to the exact ground state energy (black line). The inset in (b) shows the energy variance per spin in dependence on the number of iterations. Light blue and orange lines show the values at each iteration while dark blue and red lines show the moving average over 30 iterations.

parameters

$$L_\theta = \langle \psi_\theta | H | \psi_\theta \rangle. \tag{4}$$

The phase part of the network is trained first while keeping the modulus part constant before optimizing the whole network. This method of optimization results in better learning of the sign structure of the ground state wave function, as demonstrated in Ref. [30] and also found by our testing. We use Adam as the optimizer [37]. The input samples are generated using the Markov chain Monte Carlo. We use NetKet to implement the NQS and Monte Carlo algorithms [38–40]. Details of the optimization procedure and hyperparameters are given in Appendix A.

IV. GROUND STATE

First, we discuss the ability of the NQS ansatz to represent the ground state of the Hamiltonian in Eq. (1). To check that our method works correctly, we compare the NQS ground state energy E_{NQS} for 3×3 and 5×5 spin lattices with the exact ground state energy E_{exact}, obtained using exact diagonalization. We find that the NQS correctly describes all parameter regimes besides the small DMI regime. While the NQS ground state energies are in agreement with the exact energies within the error margin in this regime, the NQS spin expectation values do not match that of the exact ground state. The reason for this problem lies in an almost degeneracy of the ground state with the first excited state resulting in a significant overlap of the NQS ground state with the excited state found by exact diagonalization. Because of the above, we first present our results for the parameter regime where NQS is accurate and discuss the small DMI regime afterward.

The energy convergence plot for the 5×5 lattice at $D = 0.8J$ and $A = 0.3J$ is shown in Fig. 2(a). Here, the NQS correctly describes the quantum skyrmion ground state. The inset shows the relative error ΔE in the ground state energy

over the number of iterations,

$$\Delta E = \frac{|E_{\mathrm{NQS}} - E_{\mathrm{exact}}|}{|E_{\mathrm{exact}}|}. \tag{5}$$

The energy convergence for a 9×9 lattice over the number of iterations for the Hamiltonian parameters $D = 0.5J$ and $A = 0.2J$ is shown in Fig. 2(b).

The ground state diagram for this lattice is shown in Fig. 3(a), depending on the DMI, D, and the anisotropy, A. The quantum skyrmion (QS) is the ground state for a wide range of parameters [triangles in Fig. 3(a)], especially at stronger DMI, which favors a noncolinear alignment of the neighboring spins. The spin expectation value at the ith site, $\langle S_i \rangle = \langle \sigma_i/2 \rangle$, for the ground state at $D = 0.5J$ and $A = 0.2J$ is shown in Fig. 3(c). A fundamental difference from the case of classical magnetic skyrmions is that the expectation value of the length of the spins, $|\langle S_i \rangle|$, is reduced in the QS state. For the ground state in Fig. 3(c), $|\langle S_i \rangle|$ ranges from $0.92\frac{\hbar}{2}$ in the ring around the center to $1.00\frac{\hbar}{2}$ at the boundary and the center of the quantum skyrmion. Among the QS ground states, the minimum of $|\langle S_i \rangle| = 0.90\frac{\hbar}{2}$ is found at $D = 0.6J$ and $A = 0$. The spins are not merely rotated from the boundary to the center, as is the case with classical spins, but are a superposition of the local eigenstates of spin operators in different directions.

For large A and small DMI, the ground state is a ferromagnet (FM) as the spins align in the direction parallel to the boundary fields [squares in Fig. 3(a)]. An example is shown in Fig. 3(d) for $D = 0.1J$ and $A = J$. Now, we discuss the parameter regime where the NQS struggles to find the correct ground state. As both DMI and A decrease, the magnitude of the spin expectation values also decreases. We find that in this regime, marked by circles in Fig. 3(a), the quantum skyrmion only exists as a metastable state for some parameters, observed in the form of a local minimum during the optimization procedure where the NQS is stuck
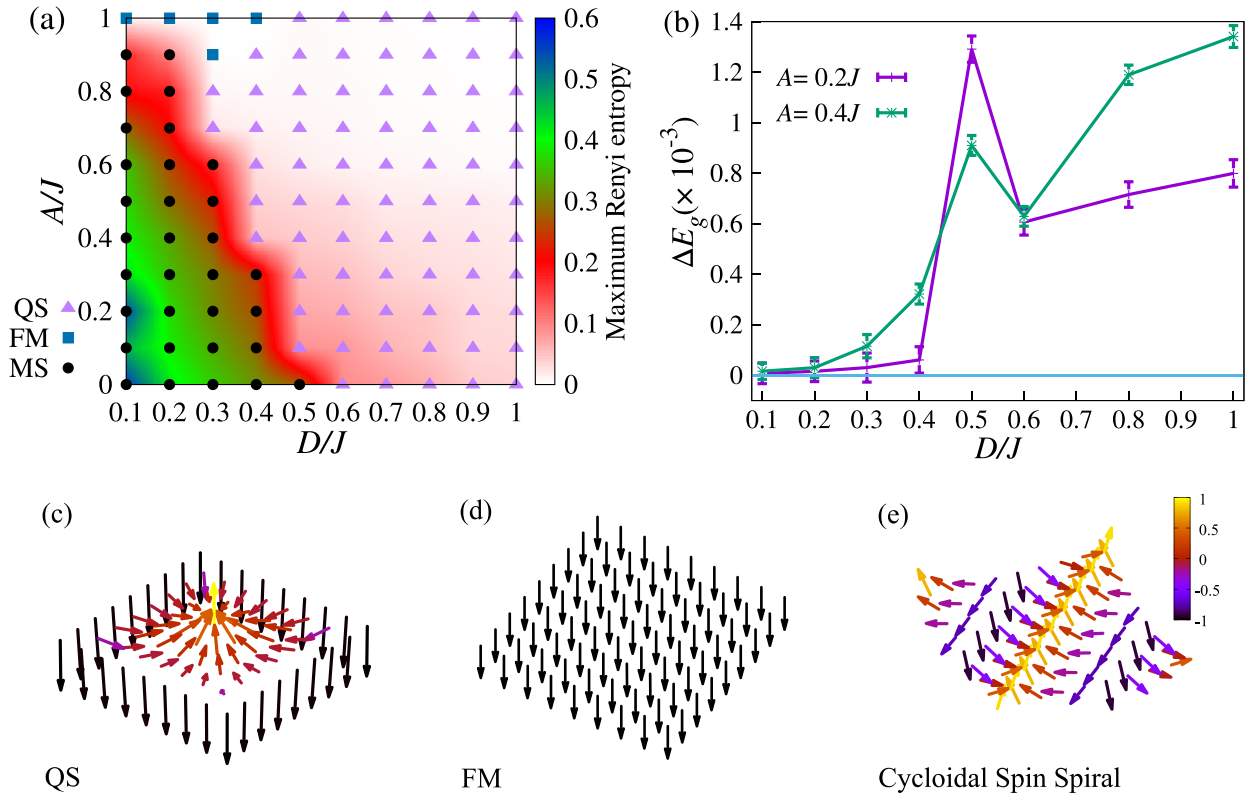
FIG. 3. Ground state diagram and spin expectation values of different ground states of Eq. (1) for a 9×9 square lattice. (a) Ground state diagram. QS denotes the quantum skyrmion state, FM the ferromagnetic state aligned with the boundary fields, and MS the mixed state (see the main text). The color map shows the maximum Renyi entropy. (b) Relative energy gap ΔE_g between the ground state and the first excited state found by the neural network quantum state method over DMI at $A = 0.2J$ and $A = 0.4J$. (c)–(e) Spin expectation values of different ground states. (c) QS at $D = 0.5J$, $A = 0.2J$ and (d) FM at $D = 0.1J$ and $A = J$. (e) For periodic boundary conditions, we obtain a cycloidal spin spiral instead of a quantum skyrmion at $D = J$ and $A = 0.5J$.

for some iterations before converging to the ground state. The ground state is characterized by almost vanishing spin expectation values aligned along the *x* or *y* direction. As mentioned earlier, for small DMI values, the NQS is not able to resolve the nearly degenerate ground state from the first excited state even in smaller lattices. Hence, we label this regime where our method does not find either a QS or an FM ground state as a "mixed state" (MS) [circles in Fig. 3(a)].

A conclusion that must be drawn from this result is that energy convergence cannot be taken as the sole measure of accuracy for the variational ground state. To have an additional metric for quantifying the accuracy of our approach, we calculate the gap between the ground state and the first excited state. This is achieved in the variational Monte Carlo scheme by optimizing a second NQS, $|\psi_\theta^1\rangle$, orthogonal to the ground state NQS, $|\psi_\theta^0\rangle$, by adding an additional term in the loss function

$$L_\theta = \langle \psi_\theta^1 | H | \psi_\theta^1 \rangle + J \left| \langle \psi_\theta^0 | \psi_\theta^1 \rangle \right|^2. \quad (6)$$

We calculate the relative energy gap as $\Delta E_g = (E_0 - E_1)/E_0$, where E_0 and E_1 are the energies corresponding to $|\psi_\theta^0\rangle$ and $|\psi_\theta^1\rangle$ respectively, and plot it over the DMI in Fig. 3(b). For $\Delta E_g < 2 \times 10^{-4}$, we do not obtain a QS or FM ground state. This corresponds to the MS region in the parameter space, where quantum skyrmions with very low spin expectation values might exist for some parameters that our method is

not able to resolve, as found for small systems by exact diagonalization [8,11,20]. This suggests that the NQS-based variational methods generally struggle with almost degenerate states. This scenario observed here for quantum spin systems, is well known from finite size electronic topological systems, which only reach exact degeneracy in the thermodynamic limit.

Projection Monte Carlo techniques exist to improve the variational ground state. However, the presence of complex off-diagonal terms in the Hamiltonian makes it difficult to use them stochastically [41]. We use an alternative projection method to remove the excited state contributions in the variational ground state (see Appendix B). After this improvement of the wave function, we obtain the correct ground state for the 3×3 lattice but not for the 5×5 lattice. Thus, while the NQS is able to represent the correct ground state for all parameters in the case of a 3×3 lattice, it is not able to learn it in the small DMI region in our variational Monte Carlo scheme.

In addition to the spin expectation values, we calculate the skyrmion number C using the normalized spin expectation values, $\mathbf{n}_i = \langle \mathbf{S}_i \rangle / |\langle \mathbf{S}_i \rangle|$, to define quantum skyrmions [11]

$$C = \frac{1}{2\pi} \sum_\Delta \tan^{-1} \left(\frac{\mathbf{n}_i \cdot (\mathbf{n}_j \times \mathbf{n}_k)}{1 + \mathbf{n}_i \cdot \mathbf{n}_j + \mathbf{n}_j \cdot \mathbf{n}_k + \mathbf{n}_k \cdot \mathbf{n}_i} \right), \quad (7)$$

where the sum runs over all elementary triangles Δ of the triangular tessellation of the quadratic lattice, having the sites i, j, and k as corners. C gives the number of times the spins wind around a unit sphere and is an integer for quantum skyrmions. In our model, we find $C = 1$ for the quantum skyrmion ground state and $C = 0$ otherwise. Furthermore, using unnormalized spin expectation values in Eq. (7), $\mathbf{n}_i = \langle 2\boldsymbol{S}_i \rangle$, results in a non-integer number Q that indicates the 'quantum' nature of skyrmions [11], similar to other quantum measures [20]. Q decreases as the entanglement increases and the spin expectation values decrease. For the QS ground states, we find a lower threshold of $Q = 0.9$.

Lastly, we note that using periodic boundary conditions without ferromagnetic boundaries ($B^z = 0$), we do not find a QS ground state. Instead, we obtain a cycloidal spin spiral [Fig. 3(e)], which is consistent with DMRG findings [19] and the fact that unfrustrated classical skyrmions require a magnetic field for stabilization. Here, a QS state minimizes the energy of a finite region of the lattice if the boundary of this region is ferromagnetically ordered. Furthermore, the quantum skyrmion ground state is stable in the presence of an additional bulk magnetic field $B^z_{\text{ext}} \sum_j \sigma^z_j$ with B^z_{ext} up to the order of $2J$ (not shown here), above which the ground state is a ferromagnet aligned along the applied field.

V. QUANTUM ENTANGLEMENT

Entanglement is an important property of quantum systems that is absent in classical systems. In this section, we investigate whether the spins in the ground state are entangled by calculating the Renyi entropy as a measure of entanglement. The Renyi entropy of the order α, where $\alpha \geqslant 0$ and $\alpha \neq 1$, is defined as

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \ln\left(\text{Tr}\left(\rho_A^\alpha\right)\right). \tag{8}$$

Here, ρ_A is the reduced density matrix obtained after splitting the system into two regions A and B and tracing out the degrees of freedom in region B. The Renyi entropy is a non-negative quantity that is zero for a pure state and takes the maximum value $\log(\min(d_1, d_2))$, where d_1 and d_2 are the dimensions of the Hilbert space in region A and B, respectively. We take region A as a single spin and region B as the rest of the lattice to obtain the entanglement of each spin with its environment. We calculate the $\alpha = 2$ Renyi entropy, $S_2(\rho_A)$, using the expectation value of the "Swap" operator [31,42] (see Appendix C).

The maximum Renyi entropy associated with the parameters is shown as a heatmap in the ground state diagram in Fig. 3(a). In the QS state for large values of DMI, we find $S_2(\rho_A) \approx 0$ irrespective of which spin A we consider, which means that these quantum skyrmions can be approximated as product states. However, as we reduce the DMI, we find that the entanglement among the spins increases, with the maximum reaching $\max_A S_2(\rho_A) = 0.09$ at $D = 0.6J$ and $A = 0.0$ for the most entangled spin. Here, the quantum skyrmion cannot be described as a product state. We plot the Renyi entropy $S_2(\rho_A)$ as a heat map over the QS ground state in Fig. 4 for the parameters $D = 0.5J$ and $A = 0.2J$. As the boundary spins are fixed with a large magnetic field, they
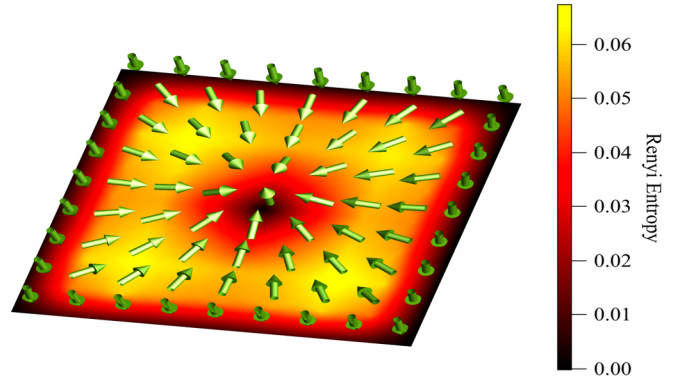

FIG. 4. Renyi entropy of each spin with its environment for the QS at $D = 0.5J$ and $A = 0.2J$.

are not entangled with the rest of the spins. The entropy first increases and then decreases from the boundary to the center, reaching its maximum between the two. One unexpected feature of this QS state is that the central spin is also disentangled from the surrounding spins, even though there is no external magnetic field acting on this site. We find that the Renyi entropy of the central spin is numerically zero for all quantum skyrmions that we obtain in our analysis; there are no accepted spin configurations during the Monte Carlo integration where the central spin points in the opposite direction than the ferromagnetic environment. This means that the QS is a product state of the central spin and a superposition of the rest of the spins. The disentangled central spin can be used to detect quantum skyrmions using the central spin magnetization as an observable in measurements without destroying the quantum nature of the skyrmionic state. We note that our results of the entropy for the QS ground state match with those in [19], in which the authors considered a bulk magnetic field instead of a ferromagnetic boundary. There, the DMRG calculations indicate a vanishing entanglement of the central spin in a quantum skyrmion with the rest of the system for a certain parameter regime. Thus a disentangled central spin might be a general feature of quantum skyrmions.

In the FM parameter region, the entropy is $S_2(\rho_A) = 0$, and these states can be represented as product states of the spins aligned with the boundary fields. Decreasing A for small DMI, we approach the MS, and the entropy reaches its maximum. Thus the difficulties in obtaining a correct solution in this parameter region might also be due to the highly entangled spins that have almost vanishing spin expectation values, along with the small energy gap between the eigenstates.

VI. NETWORK INTERPRETATION

In this final results section, we shift our focus towards interpreting the working and training of the neural network. Understanding how the network learns the target problem is integral to machine learning research and provides insights that cannot be obtained only through the final prediction. However, the interpretation of neural networks is a nontrivial problem, and a large number of neurons in multiple layers, as
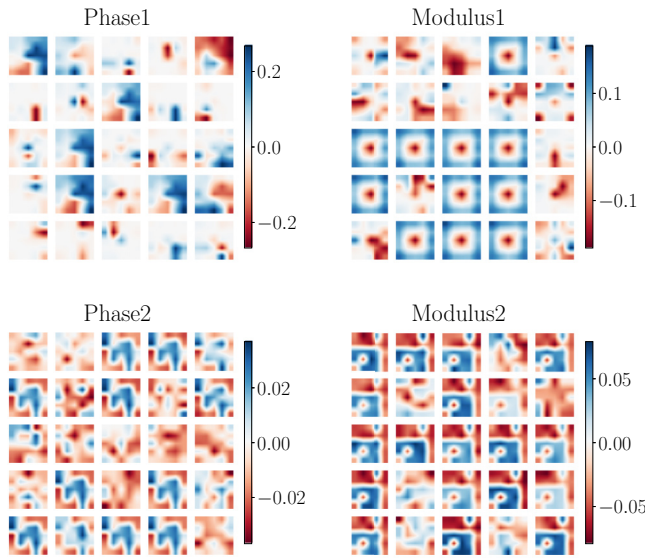
FIG. 5. Weight distribution in the hidden layers of the quantum skyrmion ground state in a 5×5 lattice at $D = J$ and $A = 0.5J$. Phase1 (phase2) and modulus1 (modulus2) denote the phase and modulus parts in the first (second) hidden layer, respectively. Each block shows the weights inside one hidden neuron. While the first hidden layer learns the essential features of the ground state, most of the neurons in the second hidden layer show a similar pattern.

in the present network shown in Fig. 1, makes it even more challenging.

For the case of NQS and many-body physics, inspecting the weights of the neural network may offer clues towards understanding the inner workings of the network [26,30]. To achieve this and to avoid dealing with an unmanageable amount of variational parameters, we study the QS ground state of the 5×5 lattice. We also use a smaller, fully connected feed-forward neural network as our variational ansatz, with two hidden layers and each layer consisting of 25 neurons for the phase and modulus parts, corresponding to $\alpha = 1$ in Fig. 1. We then transfer the results of our analysis to the calculation in the 9×9 lattice.

We plot the weights of all neurons of our NQS after training, in a 5×5 grid for each layer, in Fig. 5. We consider the QS solution at $D = J$ and $A = 0.5J$ for our analysis. Inspecting the weights of the first hidden layer, we see that in the phase part, which is trained first to improve the learning of the sign structure of the wave function, each neuron learns a specific part of the wave function. In the modulus part of the first hidden layer, we find that most of the neurons have a skyrmion-like distribution of the weights. This is because the first hidden layer directly takes the spins as inputs; it learns the most important features of the ground state. However, in the second hidden layer, we find that most of the weights in both the phase and the modulus neurons are distributed in a similar pattern and, visually, do not offer a physical interpretation. This raises two questions: First, whether the second hidden layer is essential in the network, and second, whether the neurons with a similar distribution of weights are redundant and can be removed without loss in the accuracy of the network.

In machine learning, pruning is often used to reduce the number of parameters in a neural network to increase computational efficiency without any loss in the accuracy of the network [43–45]. In most cases, pruning is done post-training by removing the weights with the smallest magnitude and adjusting the remaining weights. After all the pruning steps, only the most important weights are left in the neural network, which can shed some light on the most significant underlying features of the target problem. Pruning could also be important for NQS as a variational ansatz since, with increasing system sizes, the size of the network increases [46].

We analyze the effects of pruning to answer the questions we raised above. Again, we consider the 5×5 lattice with two types of ground states - the low entanglement QS state at $D = J, A = 0.5J$ and the high entanglement MS state at $D = 0.1J, A = 0.1J$ in Figs. 6(a) and 6(b). Starting from the second hidden layer, at each pruning step, 10% of the neurons from both phase and modulus parts are randomly deleted until only one neuron is left in each of them. Then the same procedure is applied to the first hidden layer. After deleting the neurons, the pruned network is trained to adjust the remaining weights (*pr*). Furthermore, a network with the same structure as the pruned network is also trained from scratch (*prsc*) to compare with the *pr* networks.

As metrics for the performance, we use the relative error ΔE_p and the fidelity F between the original network and the *pr* or *prsc* networks [47],

$$\Delta E_p = \frac{|E_{\text{full}} - E_p|}{|E_{\text{full}}|}, \tag{9}$$

$$F = |\langle \psi_{\text{full}} | \psi_p \rangle|^2, \tag{10}$$

where $p = pr, prsc$ (see details in Appendix A). In Figs. 6(a) and 6(b), we plot ΔE_p and F over the pruning for the 5×5 solution, with the maximum Renyi entropies in the insets. For the low entanglement QS solution, the degradation in performance is small even after removing 97% of the weights, and the fidelity stays over 97% for both *pr* and *prsc* networks. However, for the high entanglement MS solution, the *pr* and *prsc* networks show different behavior. The fidelity gradually decreases in the *prsc* network as the weights are removed. The performance of the pruned network in the high entanglement MS state is worse than in the low entanglement solution. This is expected as it becomes considerably more difficult for fewer neurons to describe the highly entangled state correctly. Moreover, the performance degradation in the *pr* network is much more severe than in the *prsc* network. This could be due to the difficulty in leaving the local minimum by the optimizer for the already trained *pr* network, while *prsc* networks have the advantage of starting from random weights and thus more flexibility. The maximum Renyi entropy in both Figs. 6(a) and 6(b) decreases as the weights are removed. Interestingly, it only becomes zero when only one neuron is left in both hidden layers, showing that NQS can represent entanglement even with a minimal number of neurons. Lastly, we note that on reducing the number of neurons, the optimization process becomes unstable and requires much fine-tuning to converge near the ground state.
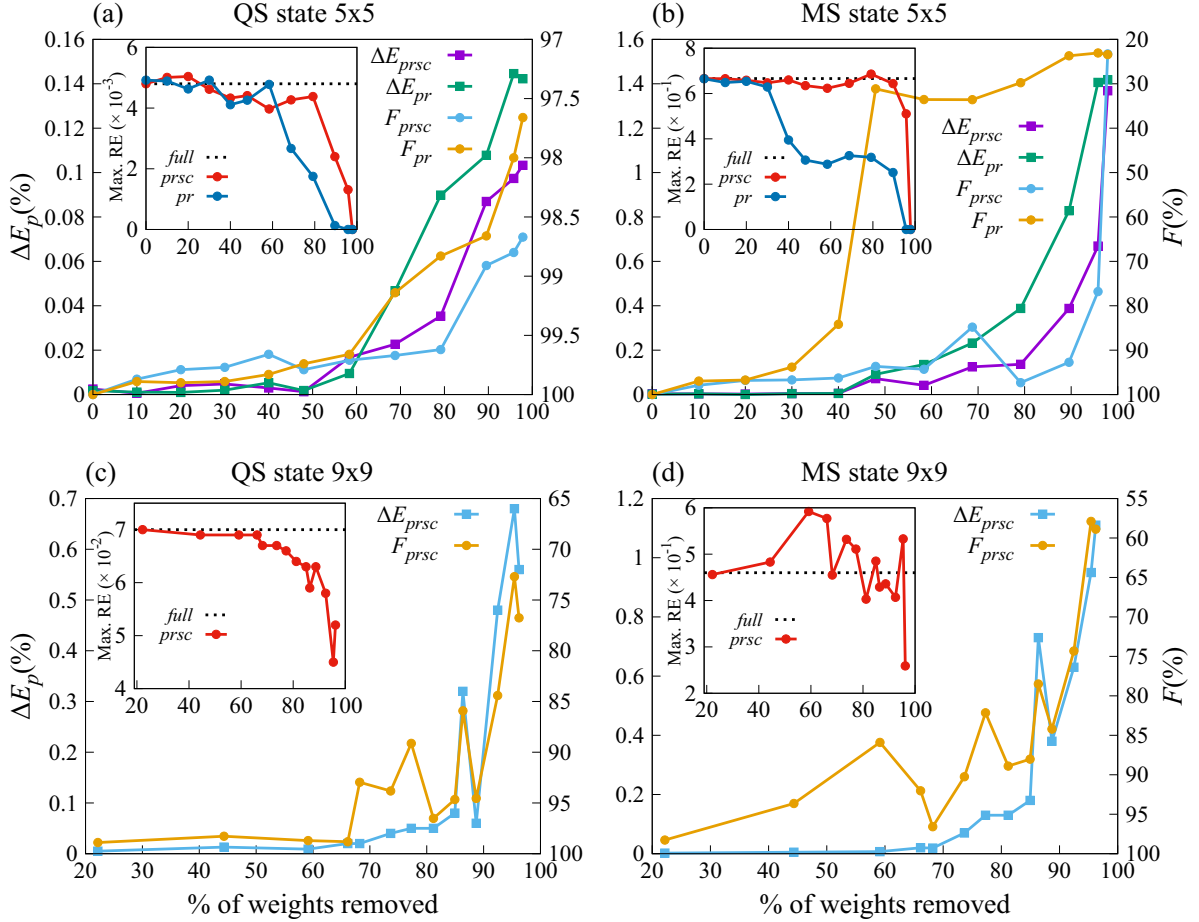
FIG. 6. Performance metrics after pruning the neural network. ΔE_p denotes the relative error in energy and F denotes the fidelity. *pr* denotes a pruned NQS trained after removing the weights from the full NQS and *prsc* denotes an identical network to the *pr* one but trained from scratch. (a) 5×5 lattice QS ground state at $D = J$ and $A = 0.5J$, (b) 5×5 lattice MS ground state at $D = 0.1J$ and $A = 0.1J$, (c) 9×9 lattice QS ground state at $D = 0.5J$ and $A = 0.2J$, and (d) 9×9 lattice MS ground state at $D = 0.1J$ and $A = 0.1J$. The inset shows the maximum values of the Renyi entropies.

In Figs. 6(c) and 6(d), we show the same results for the 9×9 lattice, calculating only *prsc* networks as they have better performance than the *pr*. The degradation in energy and fidelity, while qualitatively similar to the 5×5 case, is more severe. In all four cases, we find that removing neurons from the first hidden layer affects the network's performance more than removing them from the second hidden layer, signifying the importance of the former over the latter. This is seen in the very low error in energy until about half of the total weights are removed, after which the error rises drastically. Does this mean we can remove the second hidden layer entirely without strongly deteriorating the performance? We find that this is not the case because the performance drastically drops, and the optimization, especially in the high entanglement region, becomes unstable with only one hidden layer. We find that (not shown here) having even a single neuron in the second hidden layer results in greater accuracy than having only one hidden layer with as much as four times the number of neurons. Thus increasing the width of the network is not the optimal strategy here. On the other hand, having three or more hidden layers makes the optimization process more challenging, and the network is prone to get stuck in a local minimum. Hence, we conclude that the optimal network for our problem should have

two hidden layers, with a large number of neurons in the first hidden layer and fewer neurons in the second hidden layer.

VII. SUMMARY

In this work, we have studied the ground states of the spin-1/2 Heisenberg model in the presence of Dzyaloshinskii-Moriya interaction and Heisenberg anisotropy on a square lattice with ferromagnetic boundaries using variational Monte Carlo. We use a neural network as the variational wave function, with different parts to learn the phase and amplitude of the wave function. We show that a weakly entangled quantum skyrmion ground state, with the skyrmion number $C = 1$, exists for a wide range of Hamiltonian parameters. The entanglement increases with decreasing DMI. For large DMI values, a product state can describe the QS ground state. Remarkably, the central spin in the QS state is disentangled from the rest of the spins. Furthermore, we analyze the weights of our NQS ansatz and find that while the first hidden layer learns the most important features of the ground state, the second hidden layer is essential to achieve high accuracy. We then test the limits of the NQS by pruning and find that the higher the entanglement, the more deterioration in the performance.

Finally, we emphasize two of our results: First, our finding that the central spin decouples from the rest of the system and points into the opposite direction than the surrounding ferromagnet can be potentially used as a nondestructive detection scheme for quantum skyrmions by local spin measurements, e.g., by a magnetic scanning tunneling microscope. Second, we obtain a region in the parameter space where our method cannot resolve the correct ground state. Instead, we find a superposition between the ground state and the first excited state. This can be traced back to a tiny excitation gap between the ground state and the first excited state and reveals that the NQS ansatz has problems with almost degenerate states, which typically appear in finite size topological systems. While we could devise a scheme to improve the variational state further and separate the ground state from the first excited state in small systems, we could not do this in large spin systems. Thus, while NQS-based variational methods offer an effective tool to study the quantum skyrmion systems at medium to large DMI, they struggle in the small DMI regime. It is an open question whether other methods like DMRG also struggle in this regime. Improvement of the learning algorithm for NQS-based methods and its comparison with established methods will be our focus for future works.

ACKNOWLEDGMENTS

APPENDIX A: OPTIMIZATION PROCEDURE

Given a variational wave function $|\psi_\theta\rangle$, the expectation value of an operator O can be calculated as [26,48],

$$
\begin{aligned}
\langle O \rangle &= \frac{\langle \psi_\theta | O | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \\
&= \frac{\sum_{\sigma,\sigma'} \langle \psi_\theta | \sigma \rangle \langle \sigma | O | \sigma' \rangle \langle \sigma' | \psi_\theta \rangle}{\sum_\sigma |\psi_\theta(\sigma)|^2} \\
&= \frac{\sum_\sigma |\psi_\theta(\sigma)|^2 \sum_{\sigma'} \langle \sigma | O | \sigma' \rangle \frac{\psi_\theta(\sigma')}{\psi_\theta(\sigma)}}{\sum_\sigma |\psi_\theta(\sigma)|^2} \\
&= \sum_\sigma p_\theta(\sigma) O_\theta^{\mathrm{loc}}(\sigma),
\end{aligned}
\tag{A1}
$$

where $\psi_\theta(\sigma) = \langle \sigma | \psi_\theta \rangle$ and

$$
p_\theta(\sigma) = \frac{|\psi_\theta(\sigma)|^2}{\sum_\sigma |\psi_\theta(\sigma)|^2},
\tag{A2}
$$

$$
O_\theta^{\mathrm{loc}}(\sigma) = \sum_{\sigma'} \langle \sigma | O | \sigma' \rangle \frac{\psi_\theta(\sigma')}{\psi_\theta(\sigma)}.
\tag{A3}
$$

Here, $O_\theta^{\mathrm{loc}}(\sigma)$ is the local estimator and $p_\theta(\sigma)$ is the probability distribution of $|\sigma\rangle$. Thus the quantum expectation value of an observable O is the average of a random variable $O_\theta^{\mathrm{loc}}(\sigma)$ over the probability distribution $p_\theta(\sigma)$. Since the sum over all the states $|\sigma\rangle$ scales exponentially with the system size, Markov Chain Monte Carlo, with Metropolis-Hastings algorithm, is used to sample a series of states $|\sigma\rangle_n$ and stochastically estimate the expectation values

$$
\langle O \rangle \approx \frac{1}{N} \sum_{n=1}^{N} O_\theta^{\mathrm{loc}}(\sigma_n),
\tag{A4}
$$

where N is the total number of samples.

The energy of the system can be calculated by taking O to be the Hamiltonian. The NQS is then optimized for the ground state iteratively by minimizing the energy using a gradient descent algorithm. Here, we use the Adam optimizer, with the moments $\beta_1 = 0.9$, and $\beta_2 = 0.999$ [37]. The learning rate η is set to $\eta = 0.001$ for the phase part and increases linearly from 0 to 0.001 over the first 5000 iterations for the modulus part of the NQS. The learning rate is then reduced to $\eta = 0.0001$ after some iterations, evident by a small kink in the energy convergence plots near 20000 iterations in Fig. 2(a) and near 40 000 iterations in Fig. 2(b). We also tried a stochastic gradient descent optimizer with a stochastic reconfiguration method as a preconditioner to the gradient [26] and obtained similar results as with the Adam optimizer but with increased computational cost. A critical step in the optimization procedure is to first optimize the phase part of the network and keep the modulus part constant to facilitate the learning of the phase of the wave function. According to the variational principle, the variational energy is bounded from below by the actual ground state energy, which makes energy a convenient loss function to minimize. We sample by flipping a spin locally N times, each at a random location, where N is the total number of spins in the lattice. This makes one Monte Carlo sweep. We use 10^4 samples for energy calculation and 10^7 for all the other expectation values. The calculations were performed on a Xeon Gold 6338 CPU with multi-threading up to 4 cores.

To calculate the fidelity between two NQSs, $|\psi_1\rangle$ and $|\psi_2\rangle$ (dropping the dependence on θ for clarity), we follow a similar procedure as in Eq. (A1):

$$
\begin{aligned}
F &= \frac{|\langle \psi_1 | \psi_2 \rangle|^2}{\langle \psi_1 | \psi_1 \rangle \langle \psi_2 | \psi_2 \rangle} \\
&= \frac{\sum_{\sigma,\sigma'} \langle \psi_1 | \sigma \rangle \langle \sigma | \psi_2 \rangle \langle \psi_2 | \sigma' \rangle \langle \sigma' | \psi_1 \rangle}{\sum_\sigma |\psi_1(\sigma)|^2 \sum_{\sigma'} |\psi_2(\sigma')|^2} \\
&= \sum_\sigma \frac{|\psi_1(\sigma)|^2}{\sum_\sigma |\psi_1(\sigma)|^2} \frac{\psi_2(\sigma)}{\psi_1(\sigma)} \sum_{\sigma'} \frac{|\psi_2(\sigma')|^2}{\sum_{\sigma'} |\psi_2(\sigma')|^2} \frac{\psi_1(\sigma')}{\psi_2(\sigma')} \\
&= \sum_\sigma p_1(\sigma) \frac{\psi_2(\sigma)}{\psi_1(\sigma)} \sum_{\sigma'} p_2(\sigma') \frac{\psi_1(\sigma')}{\psi_2(\sigma')}.
\end{aligned}
$$

Thus F can be evaluated by first sampling from two different probability distributions corresponding to the two NQSs, and then computing the ratio of the wave function amplitudes.

APPENDIX B: AN ALTERNATIVE PROJECTION METHOD

Variational wave functions can be improved by projection techniques, which require the variational state to have a finite overlap with the exact ground state. Then, the high-energy components can be projected out by applying the "power method." However, this method can be done exactly only for systems manageable by exact diagonalization. In other cases, stochastic methods have to be used, requiring the Hamiltonian's off-diagonal terms to be real and non-negative. When this condition is not fulfilled, as in our case with Eq. (1) with complex off-diagonal terms, there is a fixed-node approximation for Hamiltonians with real and negative off-diagonal terms and its modification fixed-phase approximation for complex off-diagonal terms [41].

Here, we propose another method to filter out high-energy components by projecting the Hamiltonian on a few low-energy states, which can be directly obtained by the variational Monte Carlo scheme introduced in the main text.

Given a Hamiltonian H, its eigenvalue equation is

$$H|\phi\rangle = E|\phi\rangle, \tag{B1}$$

where E and $|\phi\rangle$ are the eigenvalues and eigenvectors, respectively. By expanding this equation in a complete but not necessarily orthonormal basis $|n\rangle$, we obtain the generalized eigenvalue equation

$$\frac{1}{\Omega}\left(\sum_{n_i,n_j}\langle n_j|H|n_i\rangle\langle n_i|\phi\rangle - E\langle n_j|n_i\rangle\langle n_i|\phi\rangle\right) = 0. \tag{B2}$$

Using an incomplete set of states, $|n_i\rangle$, we can define the projection of the Hamiltonian into the space spanned by these states as $H_{\mathbf{proj}} = \langle n_j|H|n_i\rangle$ and the overlap matrix $X = \langle n_j|n_i\rangle$. If $|n_i\rangle$ are approximations of the ground state and the lowest excited states of the Hamiltonian, the ground state of the projected Hamiltonian will be an improved version of the variational ground state of the full Hamiltonian.

In the converged variational NQS ground state $|n_0\rangle$, the main component is the exact ground state with small contributions from the excited states. By optimizing a second NQS, which is nearly orthogonal to the ground state NQS, using the cost function

$$L_\theta = \langle n_1|H|n_1\rangle + J|\langle n_0|n_1\rangle|^2, \tag{B3}$$

as described in the main text in Eq. (6), the first excited state can be approximated as $|n_1\rangle$. This procedure can be repeated to approximate the excited states of H. We then can use these variational low-energy states to calculate the projected Hamiltonian and the overlap matrix in a Markov Chain Monte Carlo scheme. We note that even by using the cost function Eq. (B3), there is no guaranty that the overlap of $|n_0\rangle$ and $|n_1\rangle$ exactly vanishes. We use a similar procedure as in Eq. (A1) to calculate the matrix elements of the projected

Hamiltonian and the overlap matrix. For the projection on two low-energy states, we sample using the product of these two wave functions $|n_0(\sigma)||n_1(\sigma)|$, as it gave us the best results. The projected Hamiltonian and the overlap matrix are then given as

$$\frac{\langle n_i|H|n_j\rangle}{\Omega} = \frac{\sum_{\sigma\sigma'}\frac{|n_0(\sigma)||n_1(\sigma)|}{|n_0(\sigma)||n_1(\sigma)|}n_i^\star(\sigma')n_j(\sigma)\langle\sigma|H|\sigma'\rangle}{\sum_\sigma|n_0(\sigma)||n_1(\sigma)|} \tag{B4}$$

$$\frac{\langle n_i|n_j\rangle}{\Omega} = \frac{\sum_\sigma\frac{|n_0(\sigma)||n_1(\sigma)|}{|n_0(\sigma)||n_1(\sigma)|}n_i^\star(\sigma)n_j(\sigma)}{\sum_\sigma|n_0(\sigma)||n_1(\sigma)|}, \tag{B5}$$

which determines the normalization constant in Eq. (B2) as $\Omega = \sum_\sigma|n_0(\sigma)||n_1(\sigma)|$. The wave functions in Eq. (B2) do not need to be normalized because the overlap matrix X takes care of any factors arising due to the absence of normalization. Hence, this method can be used in the variational Monte Carlo scheme, which usually considers unnormalized wave functions.

Then, by solving the generalized eigenvalue problem, Eq. (B2), we can filter out the high energy components from the variational ground state. The new variational ground state is $|n_0\rangle_{\text{new}} = \sum_i\phi_{0i}|n_i\rangle$, where ϕ_{0i} are the components of the lowest energy eigenvector of Eq. (B2). This procedure is feasible when only a few excited states are mixed in the approximation of the ground state, as the calculation of the excited state itself is variational, and the errors build up with each excited state calculation. This method works well for the 3×3 lattice over the entire parameter range, as the variational ground state has negligible overlap with the second and higher excited states. Then, only the calculation of the first variational excited state is required. However, while it improves the variational energy slightly for larger lattices, we do not obtain the correct ground state in the small DMI and A region of the ground state diagram.

APPENDIX C: RENYI ENTROPY

When a system is divided into two parts, A and B, the variational wave function can be written as

$$|\psi_\theta\rangle = \sum_{\sigma_A\sigma_B}\psi_\theta(\sigma_A\sigma_B)|\sigma_A\rangle|\sigma_B\rangle, \tag{C1}$$

where σ_A and σ_B are the basis states in region A and region B, respectively. The Renyi entropy of order α between A and B is

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha}\ln\left(\text{Tr}\left(\rho_A^\alpha\right)\right), \tag{C2}$$

where ρ_A^α is the reduced density matrix obtained after tracing out the degrees of freedom in region B. To calculate the Renyi entropy of the second order ($\alpha = 2$), we use the replica trick to evaluate the expectation value of the 'Swap' operator on two copies of the variational wave function. The Swap operator swaps the spins in one region with that of another

region between the two wave functions [42]

$$
\text{Swap}_A |\psi_\theta\rangle \otimes |\psi_\theta\rangle = \text{Swap}_A \left(\sum_{\sigma_A \sigma_B} \psi_\theta(\sigma_A \sigma_B) |\sigma_A\rangle |\sigma_B\rangle \right) \otimes \left(\sum_{\sigma'_A \sigma'_B} \psi_\theta(\sigma'_A \sigma'_B) |\sigma'_A\rangle |\sigma'_B\rangle \right)
$$

$$
= \sum_{\sigma_A \sigma_B} \psi_\theta(\sigma_A \sigma_B) \sum_{\sigma'_A \sigma'_B} \psi_\theta(\sigma'_A \sigma'_B) |\sigma'_A\rangle |\sigma_B\rangle \otimes |\sigma_A\rangle |\sigma'_B\rangle, \tag{C3}
$$

where σ and σ' are the basis states for the two copies of the wave function. The expectation value of Swap_A is then given by

$$
\langle \text{Swap}_A \rangle = \frac{\langle \psi_\theta \otimes \psi_\theta | \text{Swap}_A | \psi_\theta \otimes \psi_\theta \rangle}{\langle \psi_\theta \otimes \psi_\theta | \psi_\theta \otimes \psi_\theta \rangle}
$$

$$
= \frac{\sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} \psi_\theta^*(\sigma_A \sigma_B) \psi_\theta^*(\sigma'_A \sigma'_B) \psi_\theta(\sigma'_A \sigma_B) \psi_\theta(\sigma_A \sigma'_B)}{\sum_{\sigma \sigma'} |\langle \psi_\theta \otimes \psi_\theta | \sigma \otimes \sigma' \rangle|^2}
$$

$$
= \text{Tr}(\rho_A^2)
$$

$$
= \exp(-S_2(\rho_A)). \tag{C4}
$$

For the final step we use the definition in Eq. (C2) with $\alpha = 2$. In the Monte Carlo scheme, Eq. (C4) can be evaluated as

$$
\langle \text{Swap}_A \rangle = \sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} \frac{|\psi_\theta(\sigma_A \sigma_B)|^2}{\sum_\sigma |\psi_\theta(\sigma)|^2} \frac{|\psi_\theta(\sigma'_A \sigma'_B)|^2}{\sum_{\sigma'} |\psi_\theta(\sigma')|^2} \cdot \frac{\psi_\theta(\sigma'_A \sigma_B) \psi_\theta(\sigma_A \sigma'_B)}{\psi_\theta(\sigma_A \sigma_B) \psi_\theta(\sigma'_A \sigma'_B)}
$$

$$
= \sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} p_\theta(\sigma) p_\theta(\sigma') \frac{\psi_\theta(\sigma'_A \sigma_B) \psi_\theta(\sigma_A \sigma'_B)}{\psi_\theta(\sigma_A \sigma_B) \psi_\theta(\sigma'_A \sigma'_B)}. \tag{C5}
$$

[1] A. N. Bogdanov and D. A. Yablonskii, Thermodynamically stable 'vortices' in magnetically ordered crystals. The mixed state of magnets, Zh. Eksp. Teor. Fiz. **95**, 178 (1989) [JETP Lett. **68**, 101 (1989)].

[2] A. N. Bogdanov and C. Panagopoulos, Physical foundations and basic properties of magnetic skyrmions, Nat. Rev. Phys. **2**, 492 (2020).

[3] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, Skyrmion lattice in a Chiral magnet, Science **323**, 915 (2009).

[4] X. Z. Yu, Y. Onose, N. Kanazawa, J. H. Park, J. H. Han, Y. Matsui, N. Nagaosa, and Y. Tokura, Real-space observation of a two-dimensional skyrmion crystal, Nature (London) **465**, 901 (2010).

[5] A. Kotani, K. Harada, M. Malac, M. Salomons, M. Hayashida, and S. Mori, Observation of FeGe skyrmions by electron phase microscopy with hole-free phase plate, AIP Adv. **8**, 055216 (2018).

[6] S. Heinze, K. von Bergmann, M. Menzel, J. Brede, A. Kubetzka, R. Wiesendanger, G. Bihlmayer, and S. Blügel, Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions, Nat. Phys. **7**, 713 (2011).

[7] N. Nagaosa and Y. Tokura, Topological properties and dynamics of magnetic skyrmions, Nat. Nanotechnol. **8**, 899 (2013).

[8] V. Lohani, C. Hickey, J. Masell, and A. Rosch, Quantum Skyrmions in Frustrated Ferromagnets, Phys. Rev. X **9**, 041063 (2019).

[9] M. Stier, W. Hausler, T. Posske, G. Gurski, and M. Thorwart, Skyrmion–Anti-Skyrmion Pair Creation by in-Plane Currents, Phys. Rev. Lett. **118**, 267203 (2017).

[10] K. Everschor-Sitte, M. Sitte, T. Valet, A. Abanov, and J. Sinova, Skyrmion production on demand by homogeneous dc currents, New J. Phys. **19**, 092001 (2017).

[11] P. Siegl, E. Y. Vedmedenko, M. Stier, M. Thorwart, and T. Posske, Controlled creation of quantum skyrmions, Phys. Rev. Res. **4**, 023111 (2022).

[12] N. S. Kiselev, A. N. Bogdanov, R. Schäfer, and U. K. Rößler, Chiral skyrmions in thin magnetic films: New objects for magnetic storage technologies? J. Phys. D **44**, 392001 (2011).

[13] N. Romming, C. Hanneken, M. Menzel, J. E. Bickel, B. Wolter, K. von Bergmann, A. Kubetzka, and R. Wiesendanger, Writing and deleting single magnetic skyrmions, Science **341**, 636 (2013).

[14] C. Schütte, J. Iwasaki, A. Rosch, and N. Nagaosa, Inertia, diffusion, and dynamics of a driven skyrmion, Phys. Rev. B **90**, 174434 (2014).

[15] F. Jonietz, S. Mühlbauer, C. Pfleiderer, A. Neubauer, W. Münzer, A. Bauer, T. Adams, R. Georgii, P. Böni, R. A. Duine, K. Everschor, M. Garst, and A. Rosch, Spin transfer torques in MnSi at ultralow current densities, Science **330**, 1648 (2010).

[16] X. Z. Yu, N. Kanazawa, W. Z. Zhang, T. Nagai, T. Hara, K. Kimoto, Y. Matsui, Y. Onose, and Y. Tokura, Skyrmion flow near room temperature in an ultralow current density, Nat. Commun. **3**, 988 (2012).

[17] R. Takashima, H. Ishizuka, and L. Balents, Quantum skyrmions in two-dimensional Chiral magnets, Phys. Rev. B **94**, 134415 (2016).

[18] J. P. Gauyacq and N. Lorente, A model for individual quantal nano-skyrmions, J. Phys.: Condens. Matter **31**, 335001 (2019).

[19] A. Haller, S. Groenendijk, A. Habibi, A. Michels, and T. L. Schmidt, Quantum skyrmion lattices in Heisenberg ferromagnets, Phys. Rev. Res. **4**, 043113 (2022).

[20] O. M. Sotnikov, V. V. Mazurenko, J. Colbois, F. Mila, M. I. Katsnelson, and E. A. Stepanov, Probing the topology of the quantum analog of a classical skyrmion, Phys. Rev. B **103**, L060404 (2021).

[21] K. Mæland and A. Sudbo, Quantum fluctuations in the order parameter of quantum skyrmion crystals, Phys. Rev. B **105**, 224416 (2022).

[22] K. Kobayashi and S. Hayami, Skyrmion and vortex crystals in the Hubbard model, Phys. Rev. B **106**, L140406 (2022).

[23] R. Peters, J. Neuhaus-Steinmetz, and T. Posske, The quantum skyrmion Hall effect in f electron systems, arXiv:2304.08006.

[24] U. Schollwöck, The density-matrix renormalization group, Rev. Mod. Phys. **77**, 259 (2005).

[25] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).

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

[27] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted-boltzmann-machine learning for solving strongly correlated quantum systems, Phys. Rev. B **96**, 205152 (2017).

[28] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and Many-Body Excitations with Neural-Network Quantum States, Phys. Rev. Lett. **121**, 167204 (2018).

[29] H. Saito and M. Kato, Machine learning technique to find quantum many-body ground states of bosons on a lattice, J. Phys. Soc. Jpn. **87**, 014001 (2018).

[30] A. Szabo and C. Castelnovo, Neural network wave functions and the sign problem, Phys. Rev. Res. **2**, 033075 (2020).

[31] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, Phys. Rev. Res. **2**, 023358 (2020).

[32] D. Kochkov, T. Pfaff, A. Sanchez-Gonzalez, P. Battaglia, and B. K. Clark, Learning ground states of quantum Hamiltonians with graph networks, arXiv:2110.06390.

[33] K. Inui, Y. Kato, and Y. Motome, Determinant-free fermionic wave function using feed-forward neural networks, Phys. Rev. Res. **3**, 043126 (2021).

[34] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).

[35] A. Chen, K. Choo, N. Astrakhantsev, and T. Neupert, Neural network evolution strategy for solving quantum sign structures, Phys. Rev. Res. **4**, L022026 (2022).

[36] 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).

[37] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980.

[38] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo, NetKet 3: Machine learning toolbox for many-body quantum systems, SciPost Physics Codebases, 7 (2022).

[39] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. v. Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, NetKet: A machine learning toolkit for many-body quantum systems, SoftwareX **10**, 100311 (2019).

[40] D. Häfner and F. Vicentini, mpi4jax: Zero-copy MPI communication of JAX arrays, J. Open Source Softw. **6**, 3419 (2021).

[41] F. Becca and S. Sorella, Green's function monte carlo, in *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017), pp. 167–188.

[42] M. B. Hastings, I. Gonzalez, A. B. Kallin, and R. G. Melko, Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations, Phys. Rev. Lett. **104**, 157201 (2010).

[43] T.-J. Yang, Y.-H. Chen, and V. Sze, Designing energy-efficient convolutional neural networks using energy-aware pruning, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI* (IEEE, Piscataway, NJ, 2017), pp. 6071–6079.

[44] S. Han, H. Mao, and W. J. Dally, Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding, arXiv:1510.00149.

[45] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag, What is the state of neural network pruning? (2020), arXiv:2003.03033 [cs, stat].

[46] A. Golubeva and R. G. Melko, Pruning a restricted boltzmann machine for quantum state reconstruction, Phys. Rev. B **105**, 125124 (2022).

[47] D. Schwandt, F. Alet, and S. Capponi, Quantum Monte Carlo Simulations of Fidelity at Magnetic Quantum Phase Transitions, Phys. Rev. Lett. **103**, 170501 (2009).

[48] F. Becca and S. Sorella, Variational monte carlo, in *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017), p. 103–130.