Thesis

Neural network quantum states and quantum skyrmions

Ashish Joshi

Department of Physics, Kyoto University

June, 2024

Abstract

Classical magnetic skyrmions are topologically protected magnetic structures with vortex-like configurations. 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. However, exact solutions for quantum systems exist only when the number of particles is small and one has to resort to approximation methods to study them. These methods aim to approximate this 'quantum many-body problem' efficiently and accurately. In recent years, machine learning techniques have been increasingly used with the existing numerical methods to tackle the quantum many-body problem. In this thesis we investigate the ground state properties and dynamics of quantum skyrmions 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) with different boundary conditions. We 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 for the case of ferromagnetic boundaries, 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. Then, using the time-dependent variational principle, we study the real-time evolution of quantum skyrmions after a Hamiltonian quench with an nonuniform external magnetic field. We show that field gradients are an effective way of manipulating and moving quantum skyrmions. 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.

Contents

List of Publications

Papers related to this thesis

1. Ashish Joshi, Robert Peters and Thore Posske
   "Ground state properties of quantum skyrmions described by neural network quantum states"
   Physical Review B 108, 094410 (2023).

2. Ashish Joshi, Robert Peters and Thore Posske,
   "Quantum skyrmion dynamics studied by neural network quantum states"
   arXiv:2403.08184.

Paper not related to this thesis

1. Ashish Joshi and Robert Peters
   "Mott transition and magnetism in a fragile topological insulator"
   Physical Review B 103 (16), 165130, 2021.

Chapter 1

Introduction

To make better devices, we need to understand the physical properties of real materials. This understanding can only be obtained at the quantum scale, where interactions of many particles can create states of matter that cannot be explained by the independent electron picture. Although the physical laws necessary to describe interacting quantum systems are well known, the exact application of these laws results in equations that cannot be solved [1]. While modern computers can be used to study interacting quantum systems with a small number of particles, with increasing number of particles, the Hilbert space increases exponentially, which results in an exponential increase in computer memory required to represent the wave function. This is known as the quantum many-body problem. Therefore, approximation methods are needed to study physically relevant properties of quantum many-body systems in a finite amount of time. For more than 50 years, researchers have developed various numerical techniques that can approximate this quantum many-body problem. Among these, variational Monte Carlo [2, 3], density matrix renormalization group [4, 5, 6], dynamical mean field theory [7, 8] and quantum Monte Carlo [9, 10, 11] are some of the most successful ones, shedding light on the physics of interacting quantum systems, which would not have been possible to do with only analytical methods.

One example of such an interacting quantum system is composed of quantum spins in two dimensions in presence of ferromagnetic exchange and anti-symmetric exchange interaction. Due to these competing interactions, the ground state of this system can result in an exotic vortex like state known as a quantum skyrmion [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. Classically, skyrmions are magnetic structures that are topologically protected [24, 25, 26, 27, 28, 29]. Due to their nontrivial topology and ease-of-motion under electric currents, they have potential

applications in memory storage devices [30, 31, 32, 33, 34, 35]. Recently, small sized skyrmions were discovered, with sizes as small as atomic lattice spacing [24, 29]. However, in almost all of the theoretical and numerical studies, the skyrmions are composed of classical spins. The results of these studies might not apply to the nanoscale skyrmions, where quantum effects can become important. Thus, new studies are needed that consider a fully quantum treatment of the composing spins of the skyrmion. Such studies would require using quantum spins, which limits the size of the system that can be considered in exact diagonalization. This is a typical quantum many-body problem, and again, approximation methods are needed to study quantum skyrmions.

The problem in representing quantum many-body wave functions stems from the exponential scaling of the Hilbert space with the system size. However, in most cases we are interested only in the physically relevant states, such as the low energy states, which are usually confined to a corner of the Hilbert space. A class of numerical techniques, grouped under the name "variational methods", attempt to approximate the quantum many-body problem by finding a computationally efficient representation of the physically relevant quantum states within the Hilbert space of interest [36]. In these methods, the wave function is parameterized using a function with adjustable parameters, also known as a variational wavee function or an ansatz. Usually, the form of the ansatz depends of the type of physical problem. Widely used variational methods are variational Monte Carlo (VMC) and tensor networks. In particular, VMC methods are highly flexible and can be used independently of the type of the Hamiltonian. The main limitation of VMC is the ansatz used to approximate the high dimensional many-body wave function as it can have some biases by design or limited representation power.

The problem of approximating high-dimensional functions is not unique in quantum many-body physics. A central problem in computer science and machine learning has been to construct algorithms that can efficiently approximate highly complex tasks like image recognition and generation, speech and language translation and automated driving. For example, in natural language recognition tasks, each word corresponds to a vector in a high dimensional space of all the possible words. While classical machine learning methods like support vector machines, n-gram modeling and dimensionality reduction were somewhat successful [37, 38], the real breakthrough arrived a decade ago with the rapid improvement in artificial neural networks. In 2012, AlexNet [39], a convolutional neural network, achieved 15.3% error on the ImageNet challenge, which was around 11 percentage points lower than the runner up. ImageNet is a large databse of images with more than 14 million hand-labelled images used to train and test machine learning models in image recoginition

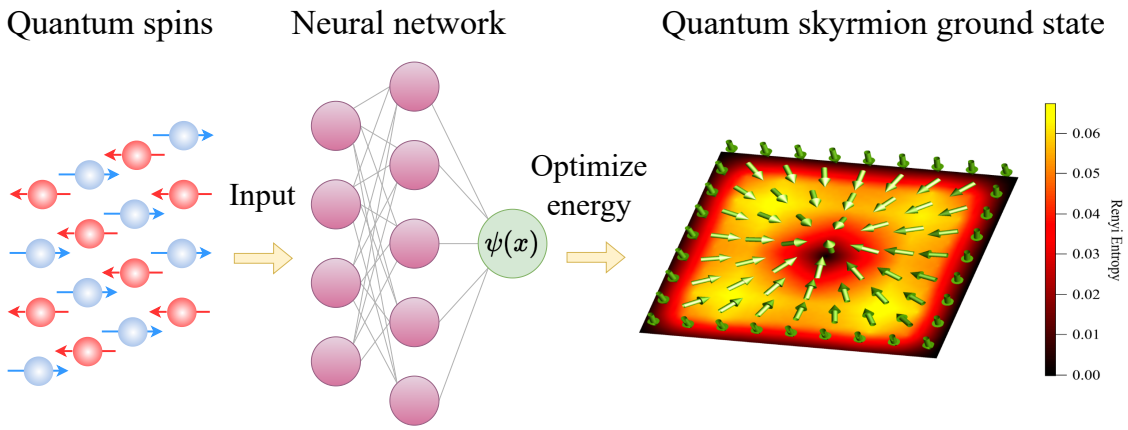


Figure 1.1: Approximating the wave function $\psi(x)$ using an artificial neural network. The input to the neural network x is the spin configuration. The network is optimized by minimzing its energy to obtain the quantum skyrmion ground state. Renyi entropy is shown in the background of the spin expectation values.

tasks [40]. Though researchers had been studying artificial neural networks for more than 30 years, it was only a decade ago that they have gained great popularity and outperformed all the other machine learning models. This success of artificial neural networks is fueled by the advancements in the computing architectures, especially graphical processing units (GPUs), which can perform numerous calculations in parallel and reduce the calculation time by orders of magnitude over a cluster of CPUs. Over the last 10 years, the single chip inference performance of nVidia GPUs has increased by 1000 times [41].

Motivated by the success of artificial neural networks in machine learning tasks, they were proposed as a variational ansatz to approximate the quantum many-body wave function in the seminal work by Carleo and Troyer in 2017 [42]. Known as neural-network quantum states (NQS), the authors showed that they can outperform state-of-the-art tensor networks methods at representing ground states in two dimensions. This opened up a new research avenue to develop NQS-based variational methods for studying quantum many-body systems. The NQS ansatz may bypass the main limitation of variational methods while enjoying all of their benefits. Theoretically, artificial neural networks are universal approximators that can approximate any smooth function given enough parameters [43, 44, 45, 46, 47]. The main difficulty lies in doing this efficiently, i.e. number of parameters scaling at most polynomially with the system size, and with high accuracy. Thus, in the past 6 years, a lot of research has been done in probing the efficacy and limitation of neural networks in quantum systems, as well as applying them to study quantum systems that were out of reach for traditional methods [48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59].

In this thesis, we have used NQS-based variational methods to study the ground state and dynamics of quantum skyrmions. We studied the Heisenberg Hamiltonian with Dzyaloshinskii-Moriya interaction on a two-dimensional quantum spin lattice. We showed that this system hosts quantum skyrmions as ground states for a large range of Hamiltonian parameters and boundary conditions. We also presented a way to move quantum skyrmions using a nonuniform external magnetic field. Figure 1.1 summarizes this thesis in one image. This thesis is organized as follows. In chapter 2, we review classical skyrmions and summarize the developments in quantum skyrmion research. In chapter 3, we present the variational methods used in this thesis and introduce artificial neural networks as a variational ansatz. In chapter 4, we present our work in which we studied the ground state properties of quantum skyrmion embedded in a ferromagnetic medium [21]. In chapter 5, we present our work in which we studied the motion of quantum skyrmions in presence of an external magnetic field [22]. Finally, we present the conclusions of this thesis.

Chapter 2

Magnetic Skyrmions

In this chapter, we give a review of magnetic skyrmions. We start with classical skyrmions in section 2.1 as they have been well studied theoretically and experimentally over the years. Then in section 2.2, we describe the nanoscale skyrmions, also known as quantum skyrmions. This area of research begun only five years ago and deals with simulation of quantum spin lattices.

2.1 Classical Skyrmions

Classical magnetic skyrmions are vortex-like magnetic structures that are topologically protected. First proposed by Skyrme in the 1960s [60] to explain the stability of hadrons in particle physics, skyrmions are now important in various condensed matter systems as well [24, 61, 62, 63, 64]. Skyrmions were experimentally discovered in single crystals of magnetic compounds with a noncentrosymmetric lattice 15 years ago [26, 27]. Due to the absence of inversion symmetry, these skyrmions were stabilized by the Dzyaloshinskii-Moriya interaction (DMI) induced by the spin-orbit coupling. Afterwards, magnetic skyrmions were found in other materials with different mechanisms - frustrated exchange interactions [65] and large interfacial DMIs induced by strong spin-orbit coupling materials interlayered with ferromagnets [29]. Among these mechanisms, skyrmions stabilized by DMIs are the most prominent and will be the focus of this work.

Magnetic skyrmions can be of various types, with a common feature being that the magnetization changes from the center of the skyrmion to its boundary, with the center spin pointing in the opposite direction of the boundary spins. Two skyrmions with different orientation of the spins are shown in Figure 2.1.

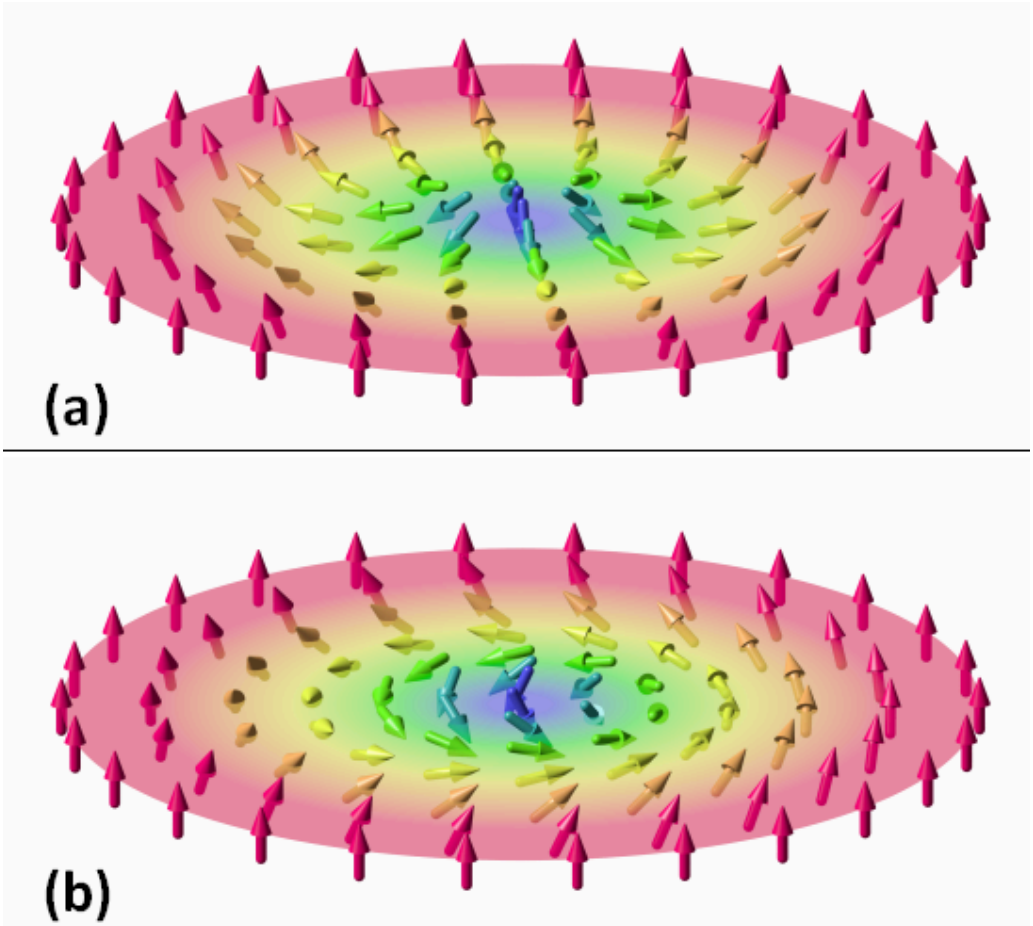


Figure 2.1: Two different types of skyrmions. (a) a Neel type skyrmion and (b) a Bloch type skyrmion [66].

In systems without inversion symmetry, the spin-orbit coupling (SOC) can induce an antisymmetric exchange interaction known as the Dzyaloshinkii-Moriya interaction (DMI) $\mathcal{H}_{i,j}^{\mathrm{DMI}}$ between two neighboring magnetic spins $\mathbf{s}_i$ and $\mathbf{s}_j$,

$$\mathcal{H}_{i,j}^{\mathrm{DMI}} = \mathbf{d}_{i,j} \cdot (\mathbf{s}_i \times \mathbf{s}_j). \tag{2.1}$$

The DMI between two magnetic atoms is mediated by a third atom with large SOC and $\mathbf{d}_{i,j}$ perpendicular to the plane of the triangle formed by the three atoms. DMI can also be induced at the interface between magnetic films and metals with large SOC. DMI is a chiral interaction that favors perpendicular alignment of spins. A skyrmion state may exist because of the stabilization provided by the competition between DMI and ferromagnetic exchange interaction, that prefers colinear spins. Skyrmions can exist as ground states in the form of a skyrmion lattice or a single skyrmion embedded in a ferromagnetic medium or as excited metastable states when an external field is applied [29, 31, 67].

The difference between skyrmions and other vortex-like structures that can also be stabilized by the DMI, like magnetic bubbles, is the well defined chirality that impacts their properties. Skyrmions are topologically protected, and a topological index S, also known as the skyrmion number, can be defined as [24]

$$S = \frac{1}{4\pi} \int \boldsymbol{m} \cdot (\partial_x \boldsymbol{m} \times \partial_y \boldsymbol{m}) \mathrm{d}x\mathrm{d}y. \tag{2.2}$$

Here, S characterizes the winding of $\boldsymbol{m}$, the normalized local magnetization, which can be mapped on a unit sphere. For a skyrmion, the local magnetization covers the whole sphere once (4π) and $S = 1$. As the topological index is nontrivial, it is not possible to continuously deform a skyrmion with $S = 1$ to a configuration with different S, for example, a ferromagnet with $S = 0$. Thus, skyrmions are topologically protected, i.e. they are robust against symmetry preserving perturbations. The quantity in Eq. (2.2) is defined for continuous space. For spins on a lattice, as is used in numerical methods, the discrete version of Eq. (2.2) was given by Berg and Lüscher in [68]. The topological index Q on a lattice is defined as a function of four spins $\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k}, \boldsymbol{s_l}$ of a plaquette u

$$Q = \sum_u \frac{1}{4\pi} \left[(\sigma A)(\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k}) + (\sigma A)(\boldsymbol{s_i}, \boldsymbol{s_k}, \boldsymbol{s_l}) \right]. \tag{2.3}$$

Here, $(\sigma A)(\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k})$ is the signed area of a triangle with corners $\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k}$. Thus, the total lattice skyrmion number is given by adding the signed area of all the elementary triangles. The signed area $(\sigma A)(\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k})$ is given as

$$\exp(\frac{1}{2} i \sigma A) = \frac{1 + \boldsymbol{s_i} \cdot \boldsymbol{s_j} + \boldsymbol{s_j} \cdot \boldsymbol{s_k} + \boldsymbol{s_k} \cdot \boldsymbol{s_i} + i \boldsymbol{s_i} \cdot (\boldsymbol{s_j} \times \boldsymbol{s_k})}{\{2(1 + \boldsymbol{s_i} \cdot \boldsymbol{s_j})(1 + \boldsymbol{s_j} \cdot \boldsymbol{s_k})(1 + \boldsymbol{s_k} \cdot \boldsymbol{s_i})\}^{1/2}}. \tag{2.4}$$

A more convenient expression for Q is given as [13, 17]

$$Q = \frac{1}{2\pi} \sum_{\Delta} \tan^{-1} \frac{8 \boldsymbol{s_i} \cdot (\boldsymbol{s_j} \times \boldsymbol{s_k})}{1 + 4(\boldsymbol{s_i} \cdot \boldsymbol{s_j} + \boldsymbol{s_j} \cdot \boldsymbol{s_k} + \boldsymbol{s_k} \cdot \boldsymbol{s_i})}, \tag{2.5}$$

where the sum runs over all the elementary triangles Δ with corners $\boldsymbol{s_i}, \boldsymbol{s_j}, \boldsymbol{s_k}$. Q is quantized to an integer and any local deformation in the spin field that does not pass through an exceptional configuration will not change it [68].

The interest in magnetic skyrmions is due to their topological stability and chirality and ease of motion under electric current. Many researchers have proposed them to be used as the basis of next generation memory devices such as racetrack memories [35]. Another advantage is the size of skyrmions, they are smaller than the magnetic domains currently being employed in memory devices. Furthermore,

it has been shown in experiments that moving skyrmions requires much lower current than moving magnetic domains [33]. This field of research based on electron spin rather than charge is known as spintronics. Very recently, another application of magnetic skyrmions were proposed, where researchers theoretically showed how skyrmions can be used as qubits [69].

A common method to study classical skyrmions is via many-spin simulations using micromagnetic equations [70]. A simplified equation for the Hamiltonian to study skyrmions on the lattice is given as

$$\mathcal{H} = \mathcal{H}^{\mathrm{H}} + \mathcal{H}^{\mathrm{DMI}} = -\sum_{i,j} J_{i,j}\boldsymbol{s}_i \cdot \boldsymbol{s}_j - \sum_{i,j} \boldsymbol{D}_{i,j} \cdot (\boldsymbol{s}_i \times \boldsymbol{s}_j). \tag{2.6}$$

Here, $\mathcal{H}^{\mathrm{H}}$ is the Heisenberg Hamiltonian which favors a parallel alignment of neighboring spins, and $\mathcal{H}^{\mathrm{DMI}}$ is the Dzyaloshinskii-Moriya interaction that favors perpendicular alignment of neighboring spins. For many-spin simulations, a continuous magnetization field $\boldsymbol{M}(\boldsymbol{r})$, or the unit magnetization field $\boldsymbol{m} = \boldsymbol{M}/M_s$, where M_s is the saturation magnetization, is used. Then, the total energy E can generally be written with various contributions [71]

$$\begin{aligned} E = \int \Big\{ & A(\nabla \boldsymbol{m})^2 + K\boldsymbol{f}(\boldsymbol{m}) + D\boldsymbol{m} \cdot (\boldsymbol{\nabla} \times \boldsymbol{m}) \\ & - \frac{M_s^2}{2\mu_0}\boldsymbol{m} \cdot \boldsymbol{h}_d + M_s H_{ext}\boldsymbol{m} \cdot \boldsymbol{h}_{ext} \Big\} \mathrm{d}V, \end{aligned} \tag{2.7}$$

where A is the exchange stiffness constant, K is the anisotropy energy constant, μ_0 is the permeability of vacuum, D is the strength of the DMI, $\boldsymbol{h}_d$ and $\boldsymbol{h}_{ext}$ are the unit vectors for the demagnetization and external magnetic field (H_{ext}) and $\boldsymbol{f}(\boldsymbol{m})$ is anisotropy function dependent on the unit magnetization $\boldsymbol{m}$. The energy can be minimized to obtain the ground state using numerical techniques like Monte Carlo methods.

The dynamics of spin structures is described by the Landau-Lifshitz-Gilbert (LLG) equation

$$\frac{\partial \boldsymbol{m}}{\partial t} = -\gamma \boldsymbol{m} \times \boldsymbol{H}_{\mathrm{eff}} + \alpha \boldsymbol{m} \times \frac{\partial \boldsymbol{m}}{\partial t} \tag{2.8}$$

where $\boldsymbol{H}_{\mathrm{eff}}$ is the effective magnetic field describing the total effect of all the interactions present in the spin Hamiltonian (see Eq. (2.7))

$$\boldsymbol{H}_{\mathrm{eff}} = -\frac{\partial E(\boldsymbol{m})}{\partial \boldsymbol{m}}. \tag{2.9}$$

The LLG equation in Eq. (2.8) can be used to study the spin dynamics when in

presence of an external magnetic field. Another method to excite the spins is using an electric current, which creates a spin transfer torque (STT) that can be added to the LLG equation [71]

$$\boldsymbol{T}_{\parallel} = -(\boldsymbol{v}_s \cdot \nabla)\boldsymbol{m}, \tag{2.10}$$

$$\boldsymbol{T}_{\perp} = -g\frac{\hbar I_e}{2e}\boldsymbol{m_2} \times (\boldsymbol{m_2} \times \boldsymbol{m_1}). \tag{2.11}$$

Here, $\boldsymbol{T}_{\parallel}$ corresponds to the in-plane electric current with velocity $\boldsymbol{v}_s = \boldsymbol{j}_e P g \mu_B / 2eM_s$ with $\boldsymbol{j}_e$ denoting the current density, P the spin polarization, g the g-factor of an electron, μ_B the Bohr magneton and e the electronic charge. The second equation with $T_{\perp}$ corresponds to the perpendicular current through the ferromagnetic layers $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$, with I_e denoting the total electric current. For the case of skyrmions, a modified version of the LLG equation can be used [24]

$$\frac{\partial \boldsymbol{m}}{\partial t} + \boldsymbol{v}_s \cdot \nabla \boldsymbol{m} = -\gamma \boldsymbol{m} \times \boldsymbol{H}_{\text{eff}} + \boldsymbol{m} \times \left(\alpha \frac{\partial \boldsymbol{m}}{\partial t} + \beta \boldsymbol{v}_s \cdot \nabla \boldsymbol{m}\right) \tag{2.12}$$

where all the terms are as defined before and β denotes the ratio of nonadiabicity of spin transfer. This equation can be used to derive the Thiele's equation of motion for the skyrmion's center of effective mass $\boldsymbol{a}$

$$M_{sk}\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} + \boldsymbol{G} \times (\boldsymbol{j} - \boldsymbol{v}) + \kappa(\alpha \boldsymbol{v} - \beta \boldsymbol{j}) = -\nabla U, \tag{2.13}$$

where $\boldsymbol{v} = \mathrm{d}\boldsymbol{a}/\mathrm{d}t$ is the velocity of skyrmion, M_{sk} is the mass of skyrmion, $\boldsymbol{j}$ is the electron current density, U is the potential caused by the boundary effects and magnetic fields, and $\boldsymbol{G} = 2\pi S \hat{\boldsymbol{z}}$ is the gyrovector with S the skyrmion number defined in Eq. (2.2). Thiele's equation can be used to study the motion of skyrmions and predicts the skyrmion Hall effect, i.e. the diagonal motion of a skyrmion due to a velocity component transverse to the applied current [72, 73, 74].

2.2 Quantum Skyrmions

While most skyrmions were of the sizes ranging from micrometer to nanometer, recent experiments have found skyrmions with sizes a few times the atomic lattice spacing [24, 29]. At this scale, quantum effects can become important. However, as described in the previous section, most theoretical works are done with skyrmions composed of classical spins, which does not take into account any quantum effect. To clarify the importance of quantum effects in these nanoscale skyrmions, also termed as "quantum skyrmions", few works have tried a semiclassical approach [12, 75]. However, a proper understanding of quantum skyrmions can only be obtained with

a full quantum treatment, i.e. by considering quantum spins [16]. Over the last 5 years, many works tackled this problem, most of them for small clusters of quantum spins solvable by exact diagonalization [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. The starting point to study quantum skyrmions is usually the quantum Heisenberg model with Dzyaloshinskii-Moriya interaction (DMI) on a two-dimensional lattice

$$\mathcal{H} = -J \sum_{\langle ij \rangle} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y) - D \sum_{\langle ij \rangle} (\boldsymbol{u}_{ij} \times \hat{\boldsymbol{z}}) \cdot (\boldsymbol{\sigma}_i \times \boldsymbol{\sigma}_j), \tag{2.14}$$

where J is the Heisenberg exchange constant and D is the DMI strength. The sums run over the nearest neighbors and $\boldsymbol{\sigma}_i = (\sigma_i^x, \sigma_i^y, \sigma_i^z)$ are the Pauli matrices at i-th site (taking $\hbar = 1$). While this system may reveal quantum skyrmions as ground states, it is often supplanted with an external magnetic field for stability of quantum skyrmions. In the following, a brief review of quantum skyrmion research is presented.

Initial works studied ground state and excited states of quantum Heisenberg models on a two-dimensional lattice using exact diagonalization. One of the first works considered a quantum skyrmion embedded in a ferromagnet [15]. They modeled seven quantum spins, with various spin momentum $S = 3\hbar/2, \hbar$ and $2\hbar$, surrounded by polarized classical spins. In addition to the exchange and DM interactions (Eq. (2.14)), the Hamiltonian consisted of anisotropy, external magnetic field and four-spin interactions. The authors showed that the ground state is a ferromagnet or a quantum skyrmion state, depending on DMI and four-spin interaction. A phase transition between these states can also be induced by an external magnetic field. As the spins are quantum, the expectation values of their lengths is not constant, a fundamental difference between quantum and classical skyrmions. Finally, the authors discuss the stability of the quantum skyrmion state when interacting with an electron. In [17], the authors address the question of topology in quantum skyrmions. Classical skyrmions are topologically protected, but this does not stay valid for the quantum case, which can become topologically trivial via quantum tunneling. The authors studied a 19 quantum spin cluster with periodic boundaries. Their Hamiltonian consisted of the Hamiltonian in Eq. (2.14) with a uniform external magnetic field $\sum_i B^z \sigma_i^z$ applied along the $\hat{z}$ direction. They show that, unlike classical skyrmions, quantum skyrmions cannot be observed in experiments by probing the local magnetization. They define a new quantity, a three spin correlation function, which can be used to characterize quantum skyrmions. This quantity, termed as the scalar chirality, is given in terms of the expectation value of

the quantum spin operators $\boldsymbol{\sigma}$

$$Q_{sc} = \frac{N_\Delta}{\pi} \langle \boldsymbol{\sigma}_i \cdot (\boldsymbol{\sigma}_j \times \boldsymbol{\sigma}_k) \rangle \,. \tag{2.15}$$

Here, the subscripts i, j, k denote the three spins making an elementary triangle plaquette and N_Δ is the total number of these non-overlapping triangles over the whole lattice. The authors show that the scalar chirality converges to the classical topological invariant defined in Eq. (2.4) for large systems. Scalar chirality stays nearly constant for different Hamiltonian parameters representing quantum skyrmion states and thus can be used to identify them. In addition to the scalar chirality, two additional quantities are defined in [14, 15] to characterize quantum skyrmions

$$Q \quad \text{or} \quad C = \frac{1}{2\pi} \sum_\Delta \tan^{-1} \left(\frac{\mathbf{n}_i \cdot (\mathbf{n}_j \times \mathbf{n}_k)}{1 + \mathbf{n}_i \cdot \mathbf{n}_j + \mathbf{n}_j \cdot \mathbf{n}_k + \mathbf{n}_k \cdot \mathbf{n}_i} \right), \tag{2.16}$$

where the sum runs over all the elementary triangles formed by nearest neighboring sites i, j, k. Using the normalized spin expectation values $\mathbf{n} = \langle \boldsymbol{\sigma} \rangle / |\langle \boldsymbol{\sigma} \rangle|$, we obtain the quantum skyrmion number C, which is equivalent to the skyrmion number for classical spins and is quantized. This takes into account only the angular winding properties of the quantum spins. To also take into account the magnitudes of the spin expectation values, unnormalized spin expectation values $\mathbf{n} = 2\langle \boldsymbol{\sigma} \rangle$ is used, which gives us Q. The authors in [14] find that quantum skyrmions (with $C = \pm 1$) but with small Q have vanishing spin expectation values and are unstable against local perturbations. Again, this property is not seen in classical skyrmions. Q is qualitatively similar to the scalar chirality Q_{sk} and both of them can be used to characterize quantum skyrmions and study their stability. Furthermore, the authors here also described a way of creating quantum skyrmions from a trivial ferromagnetic state by adiabatically rotating the magnetic field at boundaries.

To realize the applications of skyrmions in devices, it is important to understand their motion and stability with time. A few recent works elucidate this in quantum skyrmions [20, 76, 13, 77]. In [77], the authors studied the stability of quantum skyrmions in time with a similar setup as in [17]. They performed local projective measurements and showed that the properties of quantum skyrmions, like the scalar chirality Q_{sk}, does not change much with perturbations. In [76], the authors studied dynamical quantum phase transitions in quantum skyrmions. Unlike the classical phase transitions where the phase transitions are facilitated by thermal fluctuations, quantum phase transitions are purely due to quantum fluctuations. Among these, the phase transitions in out of equilibrium systems are known as dynamical quantum

phase transitions, and it is characterized by a non-analytic behavior of the wave function. In [76], the authors showed that the transition from a ferromagnet to quantum skyrmion is a dynamical quantum phase transition while from a helical phase to quantum skyrmion is not. This property is, of course, a purely quantum feature as well.

The aforementioned works clarified the differences between quantum and classical skyrmions and established methods to characterize them. However, all of them were done on small lattices which can be studied using exact diagonalization. Larger quantum systems cannot be studied exactly and approximation methods are needed. In [16], the authors studied the ground state properties of quantum skyrmions in a large lattice using matrix product states and density matrix renormalization group (DMRG). They discovered a quantum skyrmion lattice phase, which is not tractable using exact diagonalization. Furthermore, they found that the spins in quantum skyrmions are entangled with each other, the entanglement decreases with distance between two spins. However, DMRG becomes very challenging in two dimensions with increasing entanglement. Quantum Monte Carlo, another widely used approximation method for large spin systems, suffers with the negative sign problem due to the presence of complex terms in the Hamiltonian emerging from the DMI [78].

Studying non-equilibrium properties numerically is a much more challenging task than studying the low energy states of quantum many-body systems. Only one work [20] has studied the non-equilibrium properties of quantum skyrmions in large systems. There, the authors considered a localized f-electron system coupled to itinerant conduction electrons in presence of spin-orbit coupling and external magnetic field. They studied this strongly correlated system using dynamical mean field theory and numerical renormalization group. They showed the existence of a quantum skyrmion and a quantum antiskyrmion in f-electrons and c-electrons respectively. Finally, in the linear response limit, they studied the real time evolution of quantum skyrmions in presence of an electric current and showed the onset of quantum skyrmion Hall effect, an analogue of the classical skyrmion Hall effect [72, 73, 74]. However, they could only look at short time intervals because of the linear approximation, and a full nonequilibrium calculation is needed.

It is clear that there is a need of numerical methods that are not plagued by the challenges mentioned above and can be used to study low energy properties and dynamics of quantum skyrmions. Variational methods like variational Monte Carlo do not have these limitations and offer a viable method to numerically study large quantum spin lattices. But here, the main limitation is the choice of variational ansatz, which has to be tailored to the physical problem. Towards this end, very

recently in [23] a variational ansatz was proposed to represent quantum skyrmions as bosonic operators. The authors verified their construction for large spin systems by comparing their results with matrix product state simulations. A more general method, and the focus of this thesis, is to use artificial neural networks to represent the quantum skyrmion wave function [21, 22]. Artificial neural networks are powerful mathematical functions that can approximate complex high-dimensional functions. In the next chapter, we discuss them and their application to quantum many-body problems.

Chapter 3

Variational Methods

In this chapter, we review the theory of variational methods and neural-network quantum states. First, we describe the quantum many-body problem and the exponential complexity associated with it in section 3.1. Then, in section 3.2 we describe the variational principle to obtain ground states and some widely used variational wave functions to approximate the many-body wave function. In section 3.3, we describe the variation Monte Carlo method, which provides a way to estimate the quantum expectation value of an operator using sampling techniques like Markov chain Monte Carlo. In section 3.4, we provide an overview of the time-dependent variational principle to approximate the real time evolution of quantum systems. Finally, in section 3.5, we give a review of the dominant machine learning technique, artificial neural networks, and their application as a variational ansatz. Research in machine learning applications to quantum physics is advancing rapidly and we refer the reader to [79] for a detailed review on the subject.

3.1 The quantum many-body problem

An isolated quantum system is described by the quantum wave function. In interacting quantum systems, e.g. electrons interacting via the repulsive Coulomb force, each particle is affected by all the other particles and the wave function is a very complex object which scales exponentially with the number of particles. Due to this complexity, solving the Hamiltonians of these systems analytically is not possible. Modern high performance computing offers some respite and many algorithms have been developed to solve the Hamiltonians of interacting quantum systems exactly as fast as possible. However, these methods also hit a wall very soon as the number

of particles increase. For example, in the basis of the eigenstates of the spin angular momentum σ^z operator (also known as the computational basis), the wave function for a single spin can be written as $|\Psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$. Here, the coefficients α and β are the probability amplitudes of the spin being aligned along $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively. For N spins, the wave function can be written as

$$|\Psi\rangle = \alpha|\uparrow\uparrow\ldots\uparrow\rangle + \beta|\uparrow\uparrow\ldots\downarrow\rangle + \ldots + \gamma|\downarrow\downarrow\ldots\downarrow\rangle\,. \tag{3.1}$$

The number of coefficients $\alpha, \beta, \gamma...$ scale as 2^N with the number of spins N. Practically, this means that the computer memory required to store the wave function increases exponentially with number of particles and for more than 50 spins, it becomes impossible to calculate the wave function exactly. This exponential scaling of the computational complexity with number of particles is known as the quantum many-body problem.

In real materials, the number of particles run to the order of the Avogadro's number. Since exact calculation of the wave function is not possible except for small number of particles, we need approximation methods to study large systems, which might be closer to approximate real materials. Towards this end, physicists have worked over the last decades to devise clever numerical approximation methods. Some notable ones among these are dynamical mean field theory for strongly correlated electron systems [7, 8], numerical renormalization group methods for quantum impurity models [80, 81], quantum Monte Carlo methods [36, 82], tensor network methods and variational methods for ground states and dynamics of interacting quantum systems [36, 82, 83, 84]. In the next sections, we will describe variational methods, the focus of this thesis, and how it can be used to study quantum many-body systems.

3.2 Variational wave functions and the variational principle

As described in the previous section, the Hilbert space of the quantum many-body systems increases exponentially with the system size. However, we do not often need information about the whole set of states but a small portion of physically relevant states, like the ground state and low-energy states. These physically relevant states are usually confined to a corner of the Hilbert state with limited dimensions. Variational methods aim to represent the physically relevant states by approximating the wave function coefficients using a parameterized function, known as the variational

wave function. The variational wave function $|\Psi_{\boldsymbol{\theta}}\rangle$ depends on the variational parameters $\boldsymbol{\theta}$, which can be optimized to describe the quantum state of interest. For a system of N spins, the variational wave function in the computational basis can be written as

$$|\Psi_{\boldsymbol{\theta}}\rangle = \sum_x \psi_{\boldsymbol{\theta}}(x)\,|x\rangle\,, \tag{3.2}$$

where, $|x\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_N\rangle$ are the local basis states and $\psi_{\boldsymbol{\theta}}(x) = \langle x|\Psi_{\boldsymbol{\theta}}\rangle$. The form of the variational wave function $\psi_{\boldsymbol{\theta}}(x)$ is usually inspired from the physics of the problem, for example, Hartree-Fock wave functions for weakly interacting systems, Jastrow wave function for fermionic systems, Laughlin wave functions for fractional quantum states and BCS wave functions for superconductivity [85]. The variational wave functions can be broadly divided into two categories, depending on whether they calculate the expectation values of operators exactly or approximately. Mean-field states and tensor network states fall under the first category, whereas Jastrow wave functions and neural-network quantum states fall under the second category [79].

One of the simplest forms of variational wave functions is the mean-field ansatz. Using the mean-field approximation, the many-body wave function is taken to be the tensor product of single-particle wave functions. In case of spin systems, the mean-field wave function $\left|\Psi^{\mathrm{MF}}_{\boldsymbol{\theta}}\right\rangle$ can be written as

$$\begin{aligned}\left|\Psi^{\mathrm{MF}}_{\boldsymbol{\theta}}\right\rangle &= \prod_{i=1}^{N} \left|\phi_i(\boldsymbol{\theta}^i_\uparrow, \boldsymbol{\theta}^i_\downarrow)\right\rangle \\ &= \left|\phi_1(\boldsymbol{\theta}^1_\uparrow, \boldsymbol{\theta}^1_\downarrow)\right\rangle \otimes \left|\phi_2(\boldsymbol{\theta}^2_\uparrow, \boldsymbol{\theta}^2_\downarrow)\right\rangle \otimes \cdots \otimes \left|\phi_N(\boldsymbol{\theta}^N_\uparrow, \boldsymbol{\theta}^N_\downarrow)\right\rangle.\end{aligned} \tag{3.3}$$

Here, $\left|\phi_i(\boldsymbol{\theta}^i_\uparrow, \boldsymbol{\theta}^i_\downarrow)\right\rangle$ is the single-body wave function for the spin at the i-th site and N is the total number of spins. Each wave function has two variational parameters $\boldsymbol{\theta}_\uparrow$ and $\boldsymbol{\theta}_\downarrow$ for a total of $2N$ variational parameters. Since this is a product state, the expectation value of any operator can be decomposed to the expectation value of local wave functions and can be calculated exactly. The mean-field wave function can be optimized to find the ground state by minimizing the variational energy using a gradient descent algorithm. Mean-field ansatz is easy to optimize and can give reasonable results in the weak-coupling regime, but it is not suitable to represent to represent moderate to strongly interacting systems.

Tensor networks are an important class of variational wave functions that can capture correlations between the local degrees of freedoms and can calculate the expectation values exactly. Among these, matrix product states have been very successful for one-dimensional systems [86, 83]. The wave function coefficients in

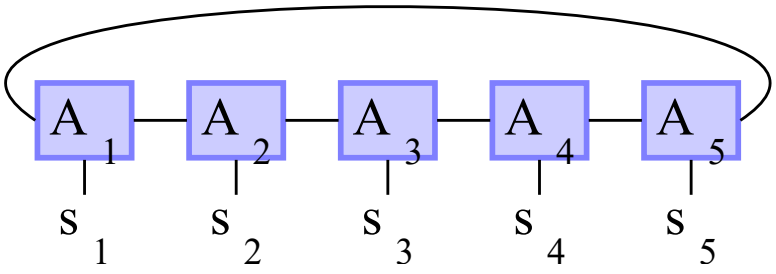


Figure 3.1: Tensor product notation of matrix product states for 5 particles [87].

Eq. (3.1) can be considered a tensor of rank N and can be decomposed into a product of lower dimensional matrices as

$$\left|\Psi_{\boldsymbol{\theta}}^{\mathrm{MPS}}\right\rangle = \boldsymbol{A}_1^{(s_1)} \boldsymbol{A}_2^{(s_2)} \dots \boldsymbol{A}_N^{(s_N)} \left|s_1 s_2 \dots s_N\right\rangle , \tag{3.4}$$

where $\boldsymbol{A}_i$ are the matrices of dimensions χ, also known as the bond dimension, and s_i are the basis states. An exact representation of the wave function requires exponential scaling of ξ with N. The MPS ansatz truncates the dimension of the indices of the tensors $\boldsymbol{A}$ using methods like singular value decomposition. This results in the scaling of variational parameters as $N\chi^2$. For some specific states, such as the gapped Hamiltonians of one-dimensional systems and states under area law entanglement, MPS are the state of the art. However, the bond dimension increase exponentially for two-dimensional systems described by the area law and MPS becomes numerically challenging.

The second type of variational wave functions draws samples from a probability distribution to approximate the expectation values instead of calculating them exactly as it scales exponentially with the system size. The accuracy of the expectation value calculation can be systematically increased by increasing the number of samples. Any function can be used to approximate the exact wave function. However, to be computationally tractable, i.e. the expectation values can be estimated with at most polynomial scaling in number of particles, the variational wave function must satisfy two conditions [79]:

1. The probability amplitude $\psi_{\boldsymbol{\theta}}(x) = \langle x|\Psi_{\boldsymbol{\theta}}\rangle$ can be calculated efficiently for any basis state $|x\rangle$.
2. Samples from the distribution $P(x) = \frac{|\langle x|\Psi_{\boldsymbol{\theta}}\rangle|^2}{\langle\Psi_{\boldsymbol{\theta}}|\Psi_{\boldsymbol{\theta}}\rangle}$ can be generated efficiently.

Here, efficiently means that the computational time scales in polynomial terms with

system size for 1. and with accuracy for 2. An example of this type of variational wave function is the Jastrow wave function

$$\psi_{\boldsymbol{\theta}}^{\mathrm{JW}}(s) = \exp\left[\frac{-1}{2}\sum_{i\neq j}\boldsymbol{\theta}_{ij}s_i s_j\right]. \tag{3.5}$$

Here, the sum runs over all the possible pairs of spins. The pairwise spin correlation is encoded in $\boldsymbol{\theta}_{ij}$. The Jastrow wave function assumes that the two body interactions are the most relevant and allows for local density fluctuations. Advantage of this type of variational wave functions is that a suitable parameterization allows us to consider a wide range of quantum phases like metals, superconductors and insulators.

If the above two conditions are satisfied, the variational wave function can be efficiently optimized to reach the lowest energy state. Thus, variational wave functions offer a flexible approach to study the low-energy properties of various quantum phases and are not bound by the type of Hamiltonian. However, the variational wave function itself is the limiting factor of these methods and may contain a relevant bias that cannot be removed within the chosen parameterization [3]. Recently, artificial neural networks have been used as variational wave functions and promise to alleviate some of the limitations of variational methods. They do not have a bias baked into the form of the variational wave function, and can theoretically represent any continuous function. We discuss artificial neural networks based variational methods in section 3.5.

There are certain properties of variational wave functions that determine its power and applicability to a quantum many-body system:

1. Capacity - it determines the representation power of the variational ansatz. For example, while mean-field states can represent noninteracting and give reasonable results for weakly interacting systems, they cannot represent strongly correlated states. Capacity plays an important role in selecting the type of variational ansatz for a given physical system. States like spin liquids usually result from a highly frustrated spin model and has long range entanglement which is notoriously difficult to be accurately represented by variational wave functions.

2. Scaling - this means the scaling of variational parameters with the system size. A polynomial scaling of parameters represents an efficient representation of the wave function. For example, the bond parameter for matrix product states scales exponentially for two or three dimensional systems with area law entanglement and thus, it cannot be used for more than a few particles.

3. Systematically improvable - another important property of variational wave functions is that it should be possible to improve their accuracy systematically, for example, by increasing the number of parameters.

4. Optimization - existence of efficient optimization algorithms is also crucial to obtain the correct representation of a wave function. The DMRG algorithm is the most efficient for one dimensional states.

As we will discuss in 3.5.3, neural-network quantum states satisfy all of these requirements.

The Variational Principle

Approximating the low-energy states of interacting quantum systems correctly requires the use of the variational principle. Here, we follow the derivation in [85]. Given a Hamiltonian $\mathcal{H}$ and a variational wave function $|\Psi_{\boldsymbol{\theta}}\rangle$, with the variational parameters $\boldsymbol{\theta}$, to approximate the exact ground state $|\Phi_0\rangle$ of $\mathcal{H}$, the variational energy $E_{\boldsymbol{\theta}}$ of $|\Psi_{\boldsymbol{\theta}}\rangle$ can be defined as

$$E_{\boldsymbol{\theta}} = \frac{\langle\Psi_{\boldsymbol{\theta}}|\,\mathcal{H}\,|\Psi_{\boldsymbol{\theta}}\rangle}{\langle\Psi_{\boldsymbol{\theta}}|\Psi_{\boldsymbol{\theta}}\rangle}. \tag{3.6}$$

The variational state can be expanded in terms of the eigenstates $|\Phi_i\rangle$ of the Hamiltonian $\mathcal{H}$ with energies E_i as

$$|\Psi_{\boldsymbol{\theta}}\rangle = \sum_i a_i\,|\Phi_i\rangle\,, \tag{3.7}$$

where $a_i = \langle\Phi_i|\Psi_{\boldsymbol{\theta}}\rangle$ with normalization $\langle\Psi_{\boldsymbol{\theta}}|\Psi_{\boldsymbol{\theta}}\rangle = \sum_i |a_i|^2 = 1$. Using Eq. (3.7), we can obtain

$$E_{\boldsymbol{\theta}} = |a_0|^2 E_0 + \sum_{i>0} |a_i|^2 E_i, \tag{3.8}$$

and using $|a_0|^2 = 1 - \sum_{i>0} |a_i|^2$, we get

$$\epsilon = E_{\boldsymbol{\theta}} - E_0 = \sum_{i\neq 0} |a_i|^2 (E_i - E_0) \geq 0. \tag{3.9}$$

Thus, the energy of variational states is bounded from below by the exact energy. To find the ground state, we have to optimize the variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ to minimize the variational energy $E_{\boldsymbol{\theta}}$. To be considered a good approximation of the exact ground state, the difference between the exact energy and variational energy $\epsilon = E_0 - E_{\boldsymbol{\theta}}$

should be much smaller than the energy gap between the ground state and the first excited state $\Delta = E_1 - E_0$.

3.3 Variational Monte Carlo

Once we have a variational wave function $|\Psi_{\boldsymbol{\theta}}\rangle$, we can optimize it to approximate desired quantum of state of interest, such as the ground state. This requires calculation of the expectation values of observables and their variances, as well as algorithms to find the optimal variational parameters $\boldsymbol{\theta}$. In this section, we present the general framework to achieve this using the variational Monte Carlo (VMC) techniques, which rely on the variational principle described in the previous section. These techniques can be applied to a wide variety of variational wave functions and Hamiltonians, irrespective of the range of interactions and dimensions of the local Hilbert space. This flexibility of VMC methods is due the stochastic approximation of the expectation values, which has a downside of introducing stochastic errors in the observables.

3.3.1 Expectation values

Given a variational wave function $|\Psi_{\boldsymbol{\theta}}\rangle$, and a complete and orthonormal basis set $|x\rangle$ in the Hilbert space such that

$$|\Psi_{\boldsymbol{\theta}}\rangle = \sum_x |x\rangle \langle x|\Psi_{\boldsymbol{\theta}}\rangle = \sum_x \psi_{\boldsymbol{\theta}}(x) |x\rangle . \tag{3.10}$$

The quantum expectation value of any operator $\mathcal{O}$ can be calculated as

$$\begin{aligned} \langle \mathcal{O} \rangle &= \frac{\langle \Psi_{\boldsymbol{\theta}}| \mathcal{O} |\Psi_{\boldsymbol{\theta}}\rangle}{\langle \Psi_{\boldsymbol{\theta}}|\Psi_{\boldsymbol{\theta}}\rangle} \qquad (3.11) \\ &= \frac{\sum_{x,x'} \langle \Psi_{\boldsymbol{\theta}}|x\rangle \langle x|\mathcal{O}|x'\rangle \langle x'|\Psi_{\boldsymbol{\theta}}\rangle}{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2} \\ &= \frac{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2 \sum_{x'} \langle x|\mathcal{O}|x'\rangle \frac{\psi_{\boldsymbol{\theta}}(x')}{\psi_{\boldsymbol{\theta}}(x)}}{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2} \\ &= \sum_x p_{\boldsymbol{\theta}}(x) \mathcal{O}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x), \qquad (3.12) \end{aligned}$$

where

$$p_{\boldsymbol{\theta}}(x) = \frac{|\psi_{\boldsymbol{\theta}}(x)|^2}{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2}, \tag{3.13}$$

$$\mathcal{O}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x) = \sum_{x'} \langle x|\mathcal{O}|x'\rangle \, \frac{\psi_{\boldsymbol{\theta}}(x')}{\psi_{\boldsymbol{\theta}}(x)}. \tag{3.14}$$

Here, $p_{\boldsymbol{\theta}}(x)$ is the Born probability distribution of the state $\psi_{\boldsymbol{\theta}}(x)$ and $\mathcal{O}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ is the local estimator. If $\mathcal{O}$ is the Hamiltonian $\mathcal{H}$, then $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ is called the local energy. Thus, the quantum expectation value of an observable $\mathcal{O}$ is recast into the classical expectation value a random variable $\mathcal{O}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ over the probability distribution $p_{\boldsymbol{\theta}}(x)$. Since the sum over all the states $|x\rangle$ scales exponentially with the system size, various Monte Carlo techniques can be used to approximate Eq. (3.12) by drawing samples $|x_n\rangle$ from the probability distribution $p_{\boldsymbol{\theta}}(x)$ and stochastically estimating the expectation values as

$$\langle \mathcal{O} \rangle \approx \frac{1}{N} \sum_{n=1}^{N} \mathcal{O}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x_n), \tag{3.15}$$

where N is the total number of samples. The error in the estimated expectation value is given by

$$\epsilon_{\mathrm{MC}} = \sqrt{\frac{\sigma^2_{\mathcal{O}}}{N}}, \tag{3.16}$$

where $\sigma^2_{\mathcal{O}}$ is the variance of the local estimator. Thus, the expectation value estimated using Monte Carlo scales with the number of samples as $1/\sqrt{N}$ and converges as $N \to \infty$.

To find the ground state, we minimize the energy $\langle \mathcal{H} \rangle$ with respect to the variational parameters $\boldsymbol{\theta}$. The variance of the random variable $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ plays an important role here because of its zero-variance property. If the variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ coincides with an exact eigenstate of $\mathcal{H}$, so that $\mathcal{H}\,|\Psi_{\boldsymbol{\theta}}\rangle = E\,|\Psi_{\boldsymbol{\theta}}\rangle$, then the local energy $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ is constant

$$\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x) = \frac{\langle x|\mathcal{H}|\Psi_{\boldsymbol{\theta}}\rangle}{\langle x|\Psi_{\boldsymbol{\theta}}\rangle} = E\frac{\langle x|\Psi_{\boldsymbol{\theta}}\rangle}{\langle x|\Psi_{\boldsymbol{\theta}}\rangle} = E. \tag{3.17}$$

Thus, the random variable $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ does not depend on $|x\rangle$ and its variance is zero. While this does not happen in real VMC calculations for general correlated systems, the zero-variance property provides us with a metric to judge the accuracy of our variational approximation. In general, the variance of the local energy $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ should decrease as it approaches an eigenstate of the Hamiltonian $\mathcal{H}$. To calculate the variance of the random variable $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$, we first note that the expectation value of

$\mathcal{H}^2$ can be written as

$$\begin{aligned}\left\langle \mathcal{H}^2 \right\rangle &= \frac{\langle \Psi_{\boldsymbol{\theta}} | \, \mathcal{H}^2 \, | \Psi_{\boldsymbol{\theta}} \rangle}{\langle \Psi_{\boldsymbol{\theta}} | \Psi_{\boldsymbol{\theta}} \rangle} \\ &= \frac{\sum_x \langle \Psi_{\boldsymbol{\theta}} | \mathcal{H} | x \rangle \, \langle x | \mathcal{H} | \Psi_{\boldsymbol{\theta}} \rangle}{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2} \\ &= \frac{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2 |\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)|^2}{\sum_x |\psi_{\boldsymbol{\theta}}(x)|^2}. \end{aligned} \tag{3.18}$$

Using the above equation, we can write the variance of $\mathcal{H}^{\mathrm{loc}}_{\boldsymbol{\theta}}(x)$ as the quantum variance of the Hamiltonian over the variational wave function $\Psi_{\boldsymbol{\theta}}$

$$\sigma^2_{\mathcal{H}} = \frac{\langle \Psi_{\boldsymbol{\theta}} | (\mathcal{H} - E)^2 | \Psi_{\boldsymbol{\theta}} \rangle}{\langle \Psi_{\boldsymbol{\theta}} | \Psi_{\boldsymbol{\theta}} \rangle} \tag{3.19}$$

3.3.2 Markov chain Monte Carlo

To estimate the quantum expectation value in Eq. (3.12), we need to draw samples from the probability distribution $p_{\boldsymbol{\theta}}(x)$. Doing this exactly is still an exponentially hard problem. However, this problem can be approximated using the Markov chain Monte Carlo (MCMC) method to draw samples that are correctly distributed according to $p_{\boldsymbol{\theta}}(x)$. In MCMC, a Markov chain represents a series of samples $x_{(n)}$ obtained by a transition probability $\mathcal{T}(x_{(n)} \to x_{(n+1)})$ which describes the probability of the state $x_{(n)}$ to transition to the state $x_{(n+1)}$. The chain is said to be stationary with some probability distribution $p(x)$ as its equilibrium distribution if $\mathcal{T}$ preserves $p(x)$, i.e.,

$$p(x_{(n+1)}) = \sum_x \mathcal{T}(x_{(n)} \to x_{(n+1)}) p(x_{(n)}). \tag{3.20}$$

To calculate the expectation value of a random variable, as in Eq. (3.15), a stationary Markov chain with N samples is generated from $p_{\boldsymbol{\theta}}(x)$. In practice, multiple Markov chains are used to reduce the correlations between samples. Furthermore, initial samples are discarded as the chain starts from a random distribution and must first converge to the equilibrium distribution before we can use it to estimate observables. This phase is called the *burn-in* phase. Finally, there is still no assurance that the Markov chain has converged and additional metrics might be used to assess its convergence [88, 89].

The Metropolis-Hastings algorithm can be used to construct a Markovian stochastic process that preserves any distribution $p(x)$. Since the process is Markovian, the transition probability depends only on the current state. According to the

Metropolis-Hastings algorithm, the transition probability $\mathcal{T}$ is first written as a product of a local transition rule T and an acceptance probability of the local update A as

$$\mathcal{T}(x_{(n)} \to x_{(n+1)}) = T(x_{(n)} \to x_{(n+1)})A(x_{(n)} \to x_{(n+1)}). \tag{3.21}$$

This separation of $\mathcal{T}$ gives us freedom to easily define a local update rule T instead of a global transitional probability $\mathcal{T}$. The acceptance of the update move to go from $x_{(n)}$ to $x_{(n+1)}$ is given by the probability

$$A(x_{(n)} \to x_{(n+1)}) = \min\left(1, \frac{p(x_{(n+1)})T(x_{(n+1)} \to x_{(n)})}{p(x_{(n)})T(x_{(n)} \to x_{(n+1)})}\right). \tag{3.22}$$

If the transition rule T is symmetric, i.e. $T(x_{(n)} \to x_{(n+1)}) = T(x_{(n+1)} \to x_{(n)})$, then

$$\frac{p(x_{(n+1)})}{p(x_{(n)})} = \left|\frac{\langle x_{(n+1)|}|\Psi_{\boldsymbol{\theta}}\rangle}{\langle x_{(n)}|\Psi_{\boldsymbol{\theta}}\rangle}\right|^2. \tag{3.23}$$

Thus, if the new state has a higher probability amplitude than the old state, it is accepted. If it has lower probability amplitude, it is accepted randomly according to the probability ratio. This also allows us to use unnormalized variational ansatz, which is crucial for using neural-network quantum states. With this procedure, the probability distribution $p(x)$ generated by the Metropolis-Hastings algorithm satisfies the *detailed balance* condition

$$p(x_{(n)})\mathcal{T}(x_{(n)} \to x_{(n+1)}) = p(x_{(n+1)})\mathcal{T}(x_{(n+1)} \to x_{(n)}). \tag{3.24}$$

The detailed balance is a sufficient condition to ensure that the distribution $p(x)$ is stationary. That is, it ensures that regardless of the initial configuration $x_{(0)}$, the Markov chain eventually converges to the correct distribution $p(x)$.

Selecting a transition rule T to update the state $x_{(n)}$ can be quite tricky as we need to make sure that the rule respects the symmetries of the system and is ergodic, i.e., any state with a finite probability can be reached by the Markov chain in a finite number of steps. Even if these conditions are met, it can be the case that the samples generated are not independent but correlated to some degree. Thus, an appropriate choice of the transition rule T is important, and usually depends on the type of system that is considered. For spin systems, flipping a single spin randomly or exchanging two spins randomly are some of the common choices.

3.4 Time-dependent variational principle

In this section, we describe a method to approximate the real time evolution of quantum many-body systems using the time-dependent variational principle (TDVP). The time evolution of a wave function $|\Psi(t)\rangle$ at time t is given by the Schrodinger equation (taking $\hbar = 1$):

$$\frac{d}{dt}|\Psi(t)\rangle = \mathcal{H}|\Psi(t)\rangle . \tag{3.25}$$

For a time-independent Hamiltonian $\mathcal{H}$, the solution of this equation is give by

$$|\Psi(t)\rangle = e^{-i\mathcal{H}t}|\Psi(0)\rangle . \tag{3.26}$$

As the Hilbert space grown exponentially with the system size, exact evolution of equation (3.26) is possible only for small systems and approximation methods are needed. Time dependent mean field approximation is not accurate for interacting many-body systems as it cannot represent the correlation effects. Quantum Monte Carlo methods struggle with the sign problem when the Hamiltonian contains complex off-diagonal elements, as is the case for quantum skyrmions. Tensor network methods can provide accurate results for one dimensional systems, but they struggle in two or three dimensions and with increasing time the entanglement increases, which requires an increasing bond dimension with time. The variational Monte Carlo method described in the previous section to obtain the ground state of many-body Hamiltonians can be extended to deal with the time evolution of variational wave functions.

The main idea behind time-dependent variational principle (TDVP) is that the time dependence of the wave function $\left|\Psi_{\boldsymbol{\theta}(t)}\right\rangle$ is encoded into the variational parameters $\boldsymbol{\theta}(t)$ [36]. The exact infinitesimal real-time evolution of $\psi_{\boldsymbol{\theta}(t)}(x)$ is given as

$$\phi(t+\delta t, x) = \psi_{\boldsymbol{\theta}(t)}(x) - i\delta t \left\langle x\right| H \left|\Psi_{\boldsymbol{\theta}(t)}\right\rangle + \mathcal{O}(\delta t^2). \tag{3.27}$$

Given the variational parameters $\boldsymbol{\theta}(t)$ at time t, our task is to find the new variational parameters $\tilde{\boldsymbol{\theta}}(t+\delta t)$ at time $t+\delta t$ that can approximate the exact infinitesimal time evolution in Eq. (3.27). Expanding $\tilde{\boldsymbol{\theta}}$ in the first order as $\tilde{\boldsymbol{\theta}}(t+\delta t) = \boldsymbol{\theta}(t) + \delta t\dot{\boldsymbol{\theta}}(t) + \mathcal{O}(\delta t^2)$, we can write the new variational wave function amplitudes $\psi_{\tilde{\boldsymbol{\theta}}(t+\delta t)})$ in the first order as

$$\psi_{\tilde{\boldsymbol{\theta}}(t+\delta t)}(x) = \psi_{\boldsymbol{\theta}(t)+\delta t\dot{\boldsymbol{\theta}}(t)}(\sigma) = \psi_{\boldsymbol{\theta}(t)}(\sigma) - \delta t\dot{\boldsymbol{\theta}}(t)\partial_{\boldsymbol{\theta}(t)}\psi_{\boldsymbol{\theta}(t)}(x) + \mathcal{O}(\delta t^2). \tag{3.28}$$

Thus, we can define a cost function between the two wave functions $|\Phi(t+\delta t)\rangle =$

$\sum_x \phi(t+\delta t, x)\,|x\rangle$ (Eq. (3.27)) and $\left|\Psi_{\tilde{\boldsymbol{\theta}}(t+\delta t)}\right\rangle = \sum_x \psi_{\tilde{\boldsymbol{\theta}}(t+\delta t)}(x)\,|x\rangle$ (Eq. (3.28)) to minimize:

$$\mathcal{L}(\tilde{\boldsymbol{\theta}}) = \mathrm{dist}\left(\left|\Psi_{\tilde{\boldsymbol{\theta}}(t+\delta t)}\right\rangle, |\Phi(t+\delta t)\rangle\right). \tag{3.29}$$

The distance $\mathrm{dist}\left(|\psi\rangle, |\phi\rangle\right)$ between two wave functions $|\psi\rangle$ and $|\phi\rangle$ in Eq. (3.29) is given by the Fubini-Study metric

$$\mathrm{dist}(|\psi\rangle, |\phi\rangle) = \cos^{-1}\left(\sqrt{\frac{\langle\psi|\phi\rangle\,\langle\phi|\psi\rangle}{\langle\psi|\psi\rangle\,\langle\phi|\phi\rangle}}\right)^2. \tag{3.30}$$

Minimizing this equation in the limit of $\delta t \to 0$ results in an equation for the time derivative of the variational parameters $\dot{\boldsymbol{\theta}}$ (see [59] for a full derivation)

$$\boldsymbol{S}\dot{\boldsymbol{\theta}} = -i\boldsymbol{F}. \tag{3.31}$$

Here, $\boldsymbol{S}$ is the quantum geometric tensor and $\boldsymbol{F}$ is the force vector defined as (dropping the t dependence of $\boldsymbol{\theta}(t)$ for readability),

$$\begin{aligned} S_{ij} &= \frac{\langle\partial_{\boldsymbol{\theta}_i}\psi_{\boldsymbol{\theta}}|\partial_{\boldsymbol{\theta}_j}\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle} - \frac{\langle\partial_{\boldsymbol{\theta}_i}\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}\frac{\langle\psi_{\boldsymbol{\theta}}|\partial_{\boldsymbol{\theta}_j}\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}, \\ F_i &= \frac{\langle\partial_{\boldsymbol{\theta}_i}\psi_{\boldsymbol{\theta}}|\,H\,|\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle} - \frac{\langle\partial_{\boldsymbol{\theta}_i}\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}\frac{\langle\psi_{\boldsymbol{\theta}}|\,H\,|\psi_{\boldsymbol{\theta}}\rangle}{\langle\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle}. \end{aligned} \tag{3.32}$$

Both $\boldsymbol{S}$ and $\boldsymbol{F}$ can be computed by estimating the expectation values in the Monte Carlo scheme, as described in section 3.3. Finally, Eq. 3.31 can be integrated using any of the established numerical integration methods. A common choice is to use the Euler method or the fourth order Runge-Kutta method.

3.5 Neural network quantum states

In this section we discuss the variational wave functions that use an artificial neural network to represent the wave function, known as neural-network quantum states (NQS). In 3.5.1 we review artificial neural networks. In 3.5.2, we describe how they are used as variational wave functions for quantum many-body systems and in 3.5.3 we describe how to optimize them for quantum systems.

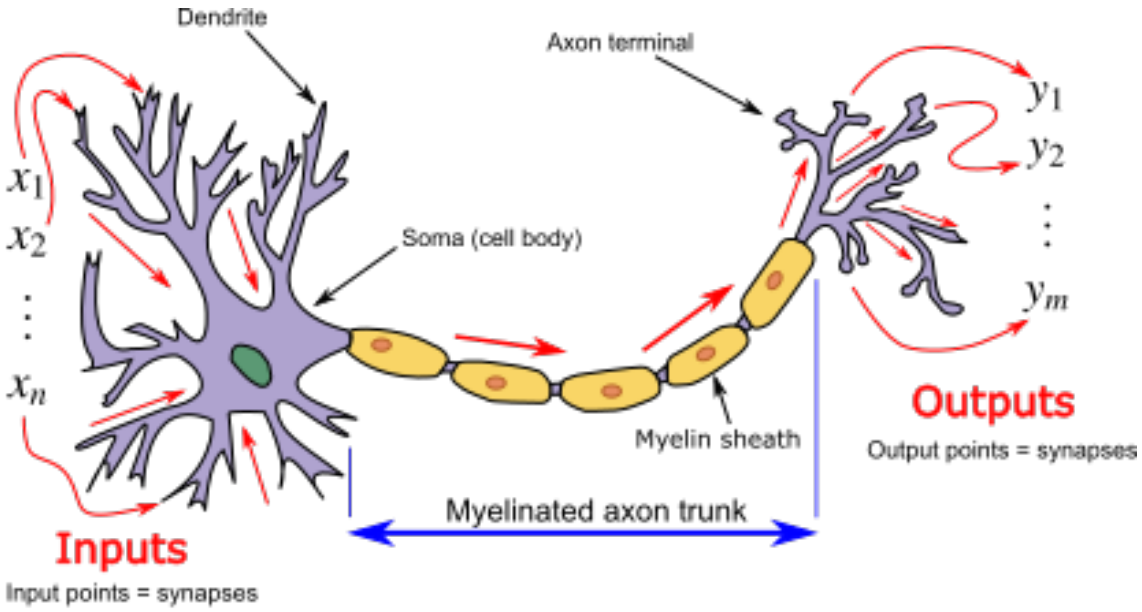


Figure 3.2: A schematic image of a biological neuron. The input signals from other neurons are received at dendrites, denoted by $x_1, x_2, ..., x_n$. Processing of the signal takes place in the cell body and the outputs, $y_1, y_2, ..., y_m$, are transmitted through the axon, which then acts as inputs for other neurons [93]. An artificial neuron is modeled in a similar way.

3.5.1 Machine learning and artificial neural networks

Machine learning is a branch of computer science that deals with statistical algorithms that can learn from data and make predictions about unseen data. They are different from other statistical optimization methods in that there is no explicit instruction given for learning and a general algorithm can work for a wide variety of tasks. Machine learning is now used ubiquitously in every technology and even in research in sciences. Artificial neural networks have outperformed all the machine learning algorithms. They are the basis of highly successful models like alphafold for predicting protein structures and drug discovery [90], natural language models for language translation [91], generation and conversation and text-to-image generators [92].

Artificial neural networks (ANNs) are mathematical functions that can approximate complex nonlinear functions. Inspired from biological neural networks, artificial neural networks form the backbone of modern machine learning and artificial intelligence. Figure 3.2 shows a schematic diagram of a single biological neuron. A biological neural network is made of millions of these neurons connected in a complex manner. On the other hand, an artificial neural network has fewer neurons, called parameters or weights, and is connected in layers in a much simpler manner.

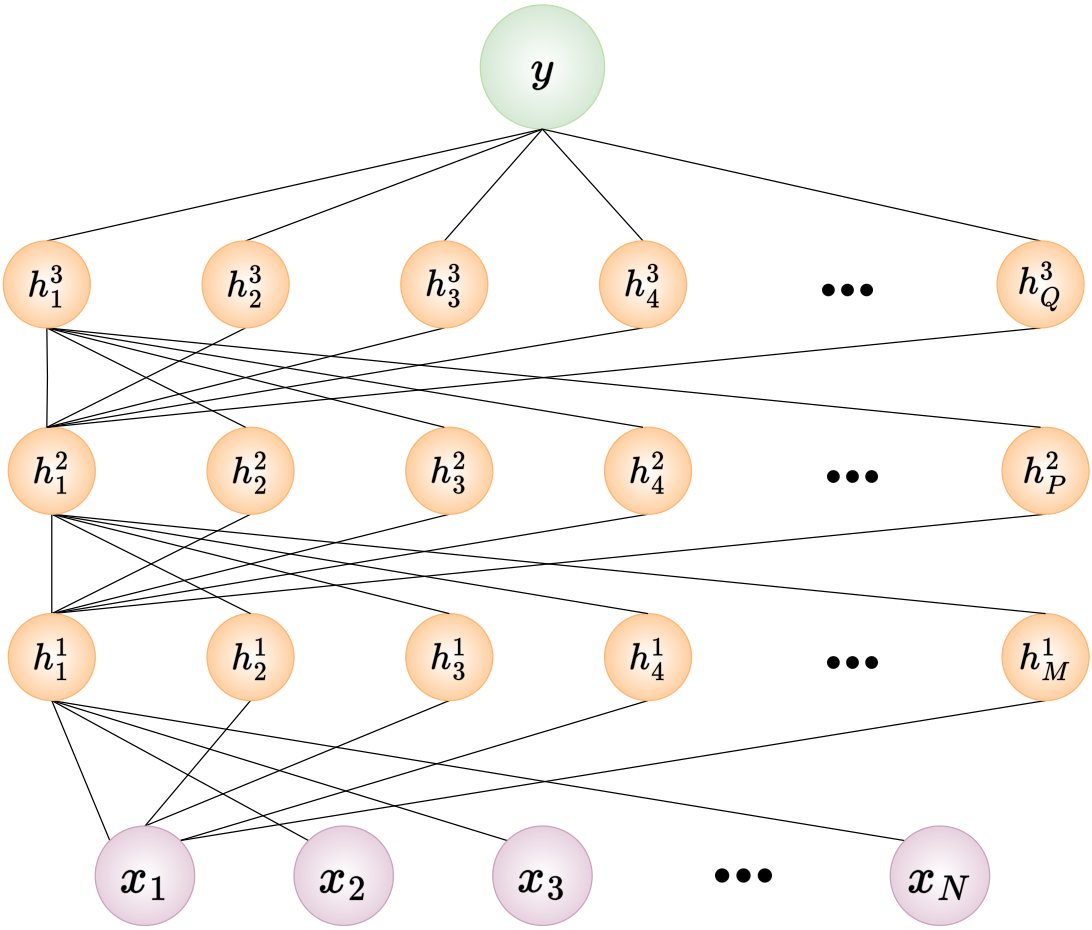


Figure 3.3: A multilayer feed forward neural network with inputs $x_1, x_2, ..., x_N$ and a single output y. There are 3 hidden layers in this network. At each neuron of the hidden layer Eq. (3.33) is carried out. The simplest form of this neural network will have only a single hidden layer.

The goal of an ANN is to approximate a function $\tilde{f}$, given the input x and output y, such that $y = \tilde{f}(x)$. The ANN defines a mapping $y = f(x, \boldsymbol{\theta})$ and learns the parameters $\boldsymbol{\theta}$ that results in the best approximation of $\tilde{f}$. The parameters $\boldsymbol{\theta}$, analogous to the biological nueron, are often connected in multiple layers so that the output of one layer is the input of another. The theoretical power of ANN has been established in the numerous universal approximation theorems that show that an ANN can approximate any continuous function given enough parameters in a single layer, i.e. with increasing the width of the network or with enough layers, i.e. increasing the depth of the network [43, 44, 45, 46, 47].

The simplest form of ANN is a single layer feedforward neural network, in which the information flows in one direction, i.e., there are no feedback connections. Representation power of the neural networks generally increase with increasing hidden layers. A more powerful, 3 hidden layer feedforward architecture is shown in 3.3. The fundamental unit in a feedforward neural network (and most other neural net-

works) is a neuron. An i-th neuron h_i^j in layer j takes inputs v_k^{j-1} from the previous layer $j-1$, weighted by the weights w_{ik}^j and offset by a bias b_i^j. This weighted linear combination then passes through a nonlinear activation function g, which is the most important part of the neural network and differentiates it from a linear approximator. The output of a single neuron is

$$\begin{aligned} h_i^j &= g\left(w_{i,1}^j \cdot v_1^{j-1} + w_{i,2}^j \cdot v_2^{j-1} + \cdots + w_{i,N}^j \cdot v_N^{j-1} + b_i^j\right) \\ &= g\left(\sum_k w_{i,k}^j v_k^{j-1} + b_i^j\right) \end{aligned} \tag{3.33}$$

This equation can be written simply in a vector form for the whole layer j as

$$\boldsymbol{h}^j = g \cdot \left(\boldsymbol{W}^j \boldsymbol{v}^{j-1} + \boldsymbol{b}^j\right), \tag{3.34}$$

where, $\boldsymbol{W}^j$ is the weight matrix for the layer j, $\boldsymbol{v}^{j-1}$ is the input vector from layer $j-1$, $\boldsymbol{b}^j$ is the bias vector for layer j and $g\cdot()$ denotes an element-wise application of the nonlinear activation function g. The choice of g is important and depends on the inputs and outputs of the ANN. For example, a sigmoid function $g(x) = (1+e^{-x})^{-1}$ is a common choice when probabilities are the output of the network. Another popular activation function is the rectified linear unit (reLU), which is defined as $g(x) = \max(0, x)$, which more closely represents a biological neuron firing and has been used in all kinds of neural networks. For stable iterative optimization and calculation of gradients, it is desirable that the activation function is continuously differentiable on almost every part of its domain. For more details refer [38, 37].

The dominance of ANN in machine learning tasks has two main factors behind it. First, the training of neural networks require huge amount of data, which has been possible only in the last decade. Second, the advancements in computing power, especially the graphical processing units that can compute millions of floating point products in parallel, much faster than central processing units. This combination of availability of data and GPUs allows us to train big neural networks with billions of parameters, creating machine learning models that was not possible before [91, 90, 92]. This also opens up the avenue for using advancements in machine learning for solving problems in other fields such as quantum physics.

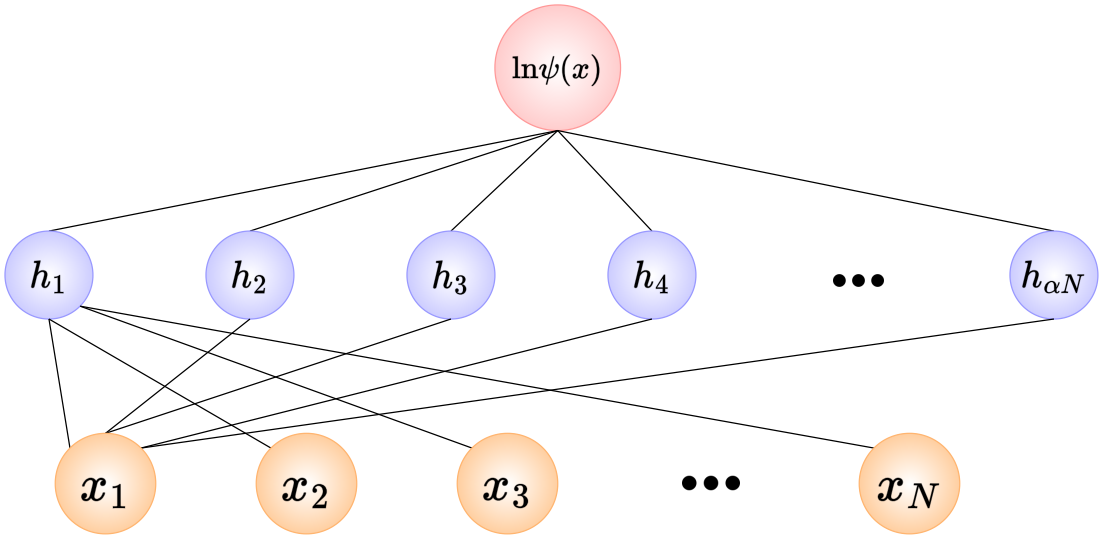


Figure 3.4: A restricted Boltzmann machine (RBM) used to represent the many-body wave function $\psi(x)$. The input layer consists of spin configurations. The output is the log of the wave function Eq. (3.35).

3.5.2 Artificial neural networks as variational wave functions

Due to their high representation power and successes in representing high dimensional data in machine learning community [90, 91, 92], artificial neural networks were proposed as variational functions. In the seminal work by Carleo and Troyer [42], neural-network quantum states (NQS) were used as the variational wave function to study the quantum many-body problem. They showed that NQS can be used to study ground state and unitary time evolution of quantum spin systems, achieving accuracy better than the tensor network methods in two dimensions. Since then, this field has generated a lot of interest and many works have explored different kinds of NQS architectures, capacity and limitations of NQS and application of NQS to a variety of quantum many-body systems [48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 94, 95, 96, 97].

The most commonly used NQS architecture is the restricted Boltzmann machine (RBM), shown in Figure 3.4. The RBM consists of an input layer that takes the spin configurations x as input and a hidden layer with variational parameters $\boldsymbol{\theta} = (a, W, b)$. Here, a are input biases, and W and b are hidden weights and biases, respectively. To capture the sign structure of the ground state wave function, the weights and biases of the RBM are taken to be complex numbers. The total number of spins is given by N, and α is the hidden neuron density, i.e. αN total neurons. To calculate the expectation values in variational Monte Carlo, it is not required for the

wave function to be normalized. The output of the RBM, hence, is the logarithm of the unnormalized wave function

$$\psi_{\boldsymbol{\theta}}(x) = \sum_{\{h_i\}} \exp\left(\sum_i a_i x_i + \sum_i b_i h_i + \sum_{i,j} W_{i,j} h_j x_i\right), \tag{3.35}$$

where $h_i = \{-1, 1\}$ is a set of αN hidden variables. Since there are no intra-layer connections in the hidden layer, the sum over h_i can be traced out and we get

$$\psi_{\boldsymbol{\theta}}(x) = \sum_{\{h_i\}} \exp\left(\sum_i a_i x_i\right) \prod_{j=1}^{\alpha N} 2\cosh\left\{\sum_i W_{i,j} x_i + b_j\right\} \tag{3.36}$$

Since the probability amplitude $\psi_{\boldsymbol{\theta}}$ can span several orders of magnitude, using the logarithm $\ln(\psi_{\boldsymbol{\theta}}(x))$ is better for the stability of the training procedure. The RBM is a special kind of single layer feedforward neural network described in the previous section, with visible biases a and $\ln\cosh(x)$ as the activation function. RBMs have been used to achieve state-of-the-art results for the ground state of quantum spin systems [98, 99], to study real time evolution [42, 100, 101], high temperature properties and fermionic systems [48, 49, 102]. It has been shown that RBMs can efficiently represent quantum states that have volume-law scaling of the entanglement. These states are out of reach for tensor-network approaches, implying the great potential of RBMs, and NQSs in general, for quantum many-body systems [56, 55].

It is possible to implement lattice symmetries like translational symmetry or rotational symmetry in an NQS. The Hamiltonian $\mathcal{H}$ is invariant under the symmetry operation, i.e. it commutes with a set of symmetry operators $\mathcal{T} = \{\hat{T}_k\}_{k=1}^{K}$. This reduces the number of variational parameters and restricts the Hilbert space to only symmetric states which usually results in better accuracy. If the Hamiltonian commutes with an operator, the ground state it is also invariant under the operation of that operator. This means that the probability amplitude for two states $|x\rangle$ and $\hat{T}_k|x\rangle$ must be the same

$$\psi_{\boldsymbol{\theta}}(x) = \psi_{\boldsymbol{\theta}}(\hat{T}_k x). \tag{3.37}$$

This is implemented in an NQS by making the first layer of the network respect the

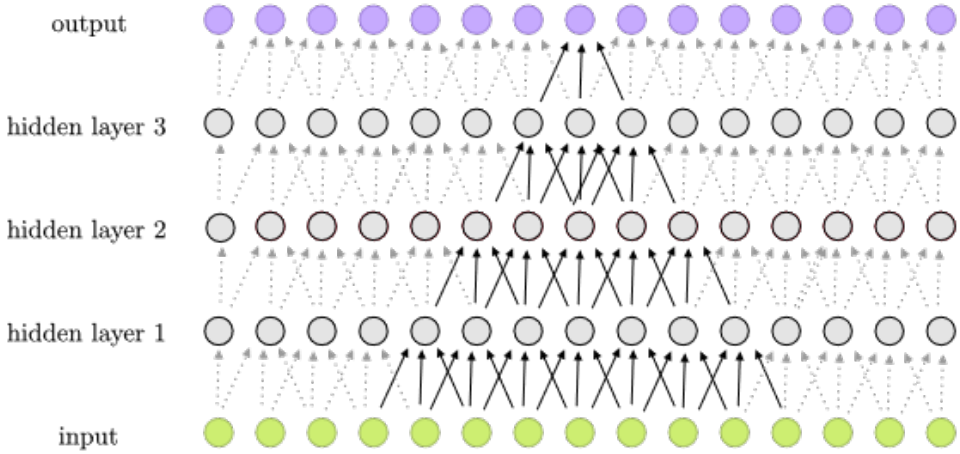


Figure 3.5: A one-dimensional, multilayered convolutional neural network (CNN). The kernel size is 3, meaning each hidden neuron takes the input from the previous 3 neurons. Unlike a fully connected neural network (Figure 3.3), the connections in a CNN are sparse and parameters are shared. This reduces the total number of parameters and the architecture is translation equivariant [103].

symmetries of the Hamiltonian. For an RBM (Eq. (3.35), this can means [42]

$$\psi_{\boldsymbol{\theta}}(x) = \sum_{\{h_{i,k}\}} \exp\Bigg(\sum_{f}^{\alpha} a_f \sum_{k}^{k} \sum_{i}^{N} x_i(k) + \sum_{f}^{\alpha} b_f \sum_{k}^{K} h_{f,k} + \sum_{f}^{\alpha} \sum_{k}^{K} h_{f,k} \sum_{i}^{N} W_i^f x_i(k) \Bigg). \tag{3.38}$$

The visible and hidden biases are now vectors in the feature space a_f and b_f where $f = 1, ..., \alpha$ and the weight matrix has the dimensions $\alpha \times N$. For the case of translational symmetry, we have N translational operators $K = N$. In machine learning language, this means a convolutional neural network with stride as N (see below). The above expression can again be simplified by tracing out the hidden variables

$$\psi_{\boldsymbol{\theta}}(x) = \exp\left(\sum_{f,k,i} a_f x_i(k) \right) \times \prod_{f} \prod_{k} 2\cosh\left(b_f + \sum_{i}^{N} W_i^f x_i(k) \right). \tag{3.39}$$

While RBMs have been quite successful, they struggle in representing frustrated spin systems and scale poorly (αN^2, where N are the input spins) with the system size. In machine learning, the representation power of artificial neural networks increase with increasing depth of the network, i.e. the number of hidden layers. The same procedure for RBMs results in multilayer RBMs, called deep Boltzmann machines and in general, multilayer feedforward neural networks (see Figure 3.3) [55, 49].

Specialized neural network architectures can be used according to the type of input data and symmetries of the Hamiltonian. Convolutional neural networks (CNNs) are appropriate for systems with translational equivariance, such as images. The convolutional operation for an input function $x(t)$ and kernel w is defined as

$$s(t) = (x * w)(t) = \int x(a)w(t-a)\mathrm{d}a. \tag{3.40}$$

The output $s(t)$ is the weighted average of $x(t)$. The output is also known as the feature map. CNNs and their variants are used for quantum systems with translational and other lattice symmetries, as they scale better than RBMs and feedforwards neural networks, use fewer parameters and may result in better accuracy [104]. This is because while an RBM is fully connected, a CNN is sparsely connected while sharing weights as they should be translational equivariant. A simplified architecture for a CNN is shown in Figure 3.5.

Another example of a specialized ANN is a recurrent neural network (RNN) used for sequential data such as text, speech and time series data. In contrast to feedforward architectures like RBMs, DBMs and CNNs, an RNN has bi-directional flow of information, i.e. the output from a nueron affects the input to the same neuron. The hidden neurons of an RNN encodes the history of the previous inputs due to which it can predict the next step in a sequence, see Figure 3.6. In the context of quantum systems, RNNs were used to study the ground state of spin models. A big advantage of an RNN is that is allows for sampling directly from the wave function and thus avoiding the need for Markov chain Monte Carlo sampling, resulting in completely independent samples. A downside is that the computation in RNNs take place sequentially, which is slower than in an RBM which can utilize the parallel processing of GPUs. However, very recently transformer wave functions are taking the spotlight as NQS for frustrated spin systems. Transformers are the state-of-the-art models used in the field of natural language processing and sequential data. In quantum systems, they do not struggle with sequential processing of an RNN and achieve better accuracy than an RBM.

Now, we discuss the characteristics of a variational wave functions in section 3.2 for NQSs:

1. Capacity - It has been shown that NQS can represent all tensor network states in polynomial number of parameters, while the vice versa is not true [56]. This means that NQS have greater capacity than the tensor network states and they can even represent states with volume law of entanglement [106].

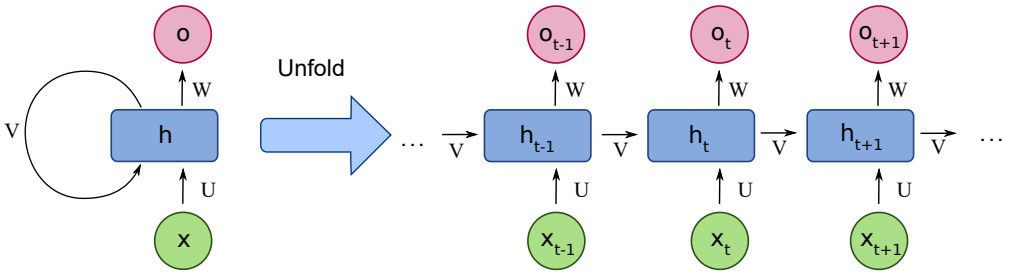


Figure 3.6: A simple recurrent neural network. On left is the compressed architecture. On right is the unfolded architecture in time. The inputs and outputs are sequential and the information flows back to the hidden neuron, which makes RNNs different from the uni-directional feedforward architectures like RBMs and CNNs [105].

2. Scaling - The scaling of variational parameters of NQS, also known as weights, highly depends on the type of neural network architecture. For example, an RBM has poor scaling αN^2 with number of particles N. For translationally symmetric systems, a CNN scales much better. However, more research is needed to shed light on scaling of NQS in general.

3. Systematically improvable - NQSs are systematically improvable as, theoretically, it is possible to increase the accuracy by increasing the number of neurons. Practically, this results in slower training and optimization difficulties with increasing depth.

4. Optimization - One benefit of using NQSs is the advancements in machine learning carry over and one can use the optimization tools developed there for quantum systems. Furthermore, as we describe in the next section, optimization methods for NQSs are well established.

3.5.3 Optimization of NQS

The optimization of neural networks is carried out in an iterative manner by updating the parameters $\boldsymbol{\theta}$ of the neural network such that some loss function $\mathcal{L}(\boldsymbol{\theta})$ is minimized. To find the ground state of a many-body Hamiltonian $\mathcal{H}$ using an NQS $|\psi_{\boldsymbol{\theta}}\rangle$, the loss function is the energy E

$$\mathcal{L}(\boldsymbol{\theta}) = E = \frac{\langle \psi_{\boldsymbol{\theta}} | \mathcal{H} | \psi_{\boldsymbol{\theta}} \rangle}{\langle \psi_{\boldsymbol{\theta}} | \psi_{\boldsymbol{\theta}} \rangle}. \tag{3.41}$$

A common optimization procedure is the gradient descent algorithm, a first order algorithm in which each parameter θ_j at i-th iteration is updated in the opposite direction of the gradient of the loss function with respect to that parameter

$$\theta_j^{i+1} = \theta_j^i - \eta \partial E / \partial \theta_j^i. \tag{3.42}$$

In practice, more sophisticated gradient descent algorithms like stochastic gradient descent and momentum based algorithms like ADAM [107] provide faster convergence with better accuracy.

For the case of variational methods, the first order gradient descent algorithms mentioned above can be further improved to achieve more accurate ground state energies and faster convergence. This is done by combining them with the second order methods such as stochastic reconfiguration [36], which leverages the imaginary time evolution of the quantum states. If the exact ground state of Hamiltonian $\mathcal{H}$ is $|\phi\rangle$, the imaginary time evolution

$$|\psi_{\boldsymbol{\theta}}(\tau)\rangle = e^{-\tau \mathcal{H}} |\psi_{\boldsymbol{\theta}}(0)\rangle \tag{3.43}$$

converges to $|\phi\rangle$ in the limit of $\tau \to \infty$. The procedure to obtain the imaginary time evolution is similar to the one for the real time evolution described in 3.4 and one ends up with an equation of motion similar to Eq. (3.31) but without the complex factor i

$$\boldsymbol{S}\dot{\boldsymbol{\theta}} = \boldsymbol{F}. \tag{3.44}$$

To obtain $\dot{\boldsymbol{\theta}}$, the pseudo-inverse of $\boldsymbol{S}$ has to be calculated. Since this matrix can be singular, a regularization term $\epsilon \to 0$ is added to its diagonal elements $(\boldsymbol{S} + \epsilon \boldsymbol{I})$ for stability of the procedure.

Chapter 4

Ground state properties of quantum skyrmions

With the foundations of the variational Monte Carlo (VMC) and neural-network quantum state (NQS) established in the previous chapter, here we present our first study on the ground state properties of quantum skyrmions using VMC with the NQS as the variational wave function. We study a quantum spin-1/2 Heisenberg model in two dimensions with Dzyaloshinskii-Moriya interaction and Heisenberg anisotropy. We show that quantum skyrmions are the ground states for a wide range of Hamiltonian parameters. To study quantum entanglement in this system, we calculate Renyi entropy of second order and demonstrate that the entanglement in the quantum skyrmion ground state decreases with increasing DMI. Previous work with density matrix renormalization group (DMRG) indicated that the central spin of a quantum skyrmion can have vanishing concurrence with its surrounding spins [16]. Interestingly, we also find that the central spin in the quantum skyrmion 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 quantum skyrmion ground states 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.

We present the model in section 4.1. In section 4.2, we present the details and hyperparameters of the NQS based VMC method. The ground state diagram, energy convergence and spin expectation values are described in section 4.3. Analysis of quantum entanglement in the ground state is given in section 4.4. We provide an interpretation of the parameters of the ground state neural-network quantum state in section 4.5 and summarize this chapter in section 4.6.

4.1 Model

We study a two-dimensional spin-1/2 Heisenberg Hamiltonian on a square lattice in the presence of the Dzyaloshinskii-Moriya interaction (DMI)

$$\begin{aligned} H = &- J \sum_{\langle ij \rangle} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y) - A \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z \\ &- D \sum_{\langle ij \rangle} (\boldsymbol{u}_{ij} \times \hat{\boldsymbol{z}}) \cdot (\boldsymbol{\sigma}_i \times \boldsymbol{\sigma}_j) + B^z \sum_b \sigma_b^z. \end{aligned} \tag{4.1}$$

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 $\hat{\boldsymbol{z}}$-axis acting only on the boundary spins indexed with b. The Pauli operator at the i-th lattice site is $\boldsymbol{\sigma}_i = (\sigma_i^x, \sigma_i^y, \sigma_i^z)$ and $\boldsymbol{u}_{ij}$ 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. (4.1) can be considered the quantum analog of a classical spin model in which the competition between the noncolinear DMI and the ferromagnetic Heisenberg exchange interaction gives rise to the formation of magnetic skyrmions. Our Hamiltonian is similar to the one used in [14], where they considered a quantum spin lattice coupled to a classical boundary and studied this system using exact diagonalization. In classical systems, 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 [16], 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 [16] for future works.

4.2 Method

We use the output of an artificial neural network to represent the complex-valued coefficients $\psi_{\boldsymbol{\theta}}(\sigma)$ in the variational wave function,

$$|\Psi_{\boldsymbol{\theta}}\rangle = \sum_x \psi_{\boldsymbol{\theta}}(x)\,|x\rangle\,. \tag{4.2}$$

Here, $|x\rangle$ are the local basis states, which in our case are the eigenstates of the σ^z operators, and $\boldsymbol{\theta}$ are the variational parameters of the neural network. In this chapter to study the ground states of the Hamiltonian in, we use two fully connected feedforward neural networks (see section 3.5.1) to each represent the phase and modulus part of the wave function, as depicted in Figure 4.1, and take the logarithm of the wave function as the total output [52]

$$\ln\left(\psi_{\boldsymbol{\theta}}(x)\right) = \rho(x) + i\phi(x). \tag{4.3}$$

The NQS takes the configuration of the spins on the two-dimensional lattice as the input. Both the phase and the modulus part of the neural 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 study. We use the rectified linear unit (reLU) as the nonlinear activation function. According to the variational principle (3.2), the variational energy is bounded from below by the exact ground state energy. The optimization of the variational wave function is achieved by minimizing the loss function $L_{\boldsymbol{\theta}}$, i.e., the variational energy, with respect to the variational parameters

$$L_{\boldsymbol{\theta}} = \langle\psi_{\boldsymbol{\theta}}|H|\psi_{\boldsymbol{\theta}}\rangle\,. \tag{4.4}$$

The variational energy is calculated using variational Monte Carlo, as described in section 3.3. The phase part of the neural 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. [52] and also found by our testing. We use Adam as the optimizer [107], with the moments $\beta_1 = 0.9$ and $\beta_2 = 0.999$. 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

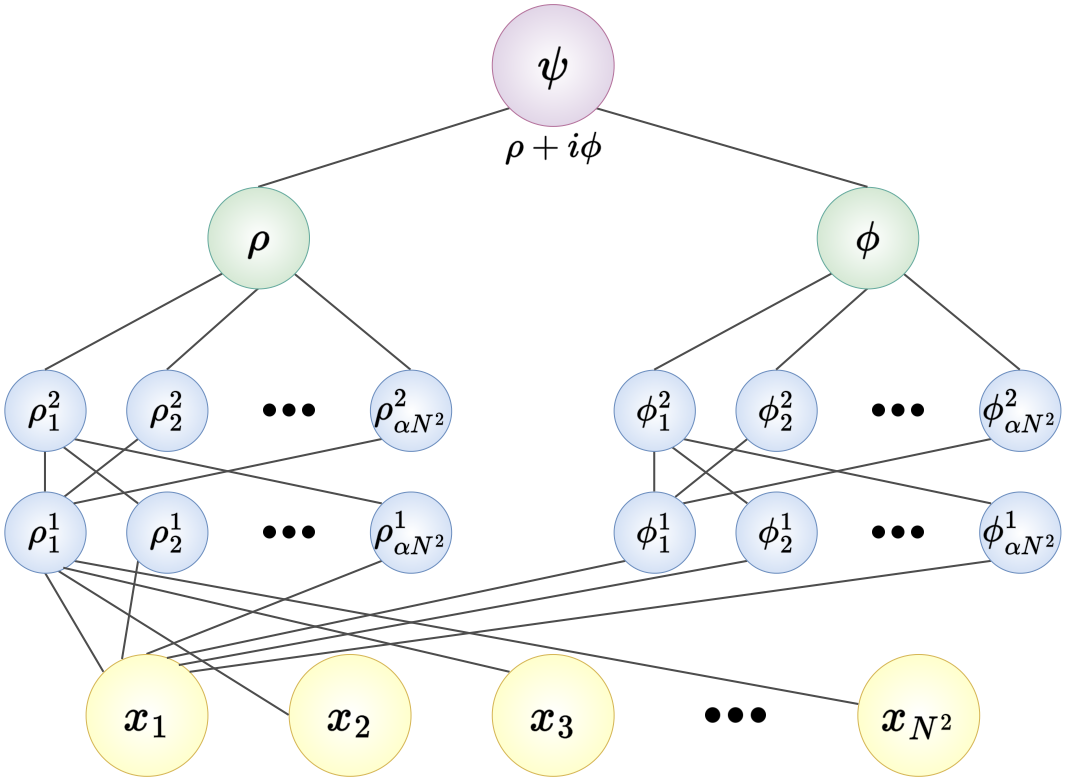


Figure 4.1: Neural network structure used as the NQS [21]. The inputs x_i are the spin configurations in the σ^z basis, and the output is the logarithm of the wave function. There are two fully connected networks, with two hidden layers in each, to learn the phase and the amplitude part of the wave function separately. Each hidden layer consists of αN^2 neurons.

rate is then reduced to $\eta = 0.0001$ after some iterations, depending on the system size. We also tried a stochastic gradient descent optimizer with a stochastic reconfiguration method as a preconditioner to the gradient [42] (see section 3.5.3) and obtained similar results as with the Adam optimizer but with increased computational cost. The input samples are generated using the Markov chain Monte Carlo with Metropolis-Hastings algorithm. The transition rule $T(x_{(n)} \to x_{(n+1)})$ to update the configuration from $x_{(n)}$ to $x_{(n+1)}$ is to flip 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 and generates a new sample. We use 10^4 samples for the energy calculation and 10^7 for all other expectation values. The calculations were performed on a Xeon Gold 6338 CPU with multi-threading up to 4 cores. To implement the NQS and Monte Carlo algorithms, we use NetKet and JAX in python [108, 109, 110].

The iterative procedure to obtain the ground state of Eq. (4.1) is given as:

Algorithm 1 Finding the ground state using NQS

Define an NQS $\psi_{\boldsymbol{\theta}}(\boldsymbol{x})$
Initialize $\boldsymbol{\theta}$ randomly
for $n = 1$ to n_{iter} **do**
    Using MCMC generate N samples $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$
    Calculate the energy gradient w.r.t. $\boldsymbol{\theta}$, $\partial E/\partial \boldsymbol{\theta}$
    Minimize energy by optimizing $\boldsymbol{\theta}$ ▷ e.g. $\boldsymbol{\theta}_j \leftarrow \boldsymbol{\theta}_j - \eta \partial E/\partial \boldsymbol{\theta}_j$
end for

4.3 Ground state analysis

First, we discuss the ability of the NQS ansatz to represent the ground state of the Hamiltonian in Eq. (4.1). 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.

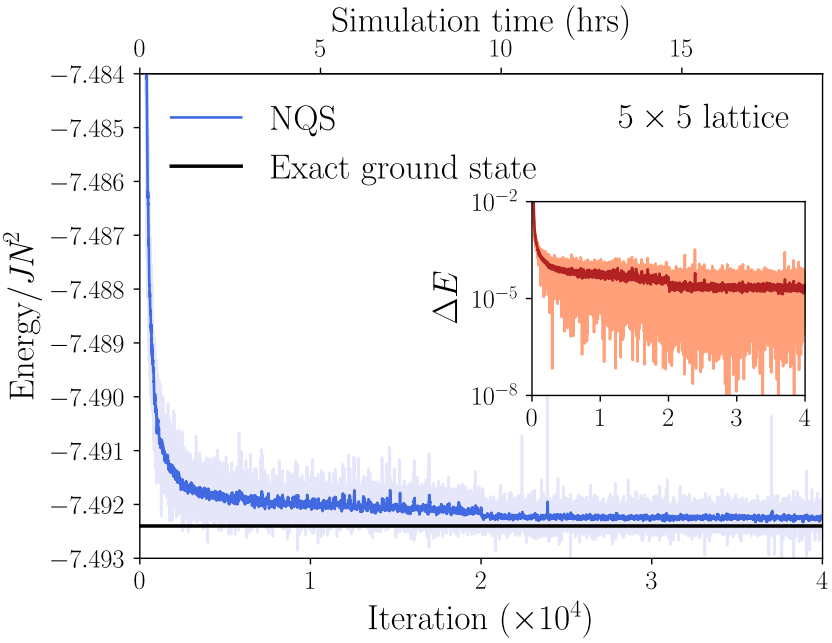
4.3.1 Strong DMI regime

The energy convergence plot for the 5×5 lattice at $D = 0.8J$ and $A = 0.3J$ is shown in Figure 4.2(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,

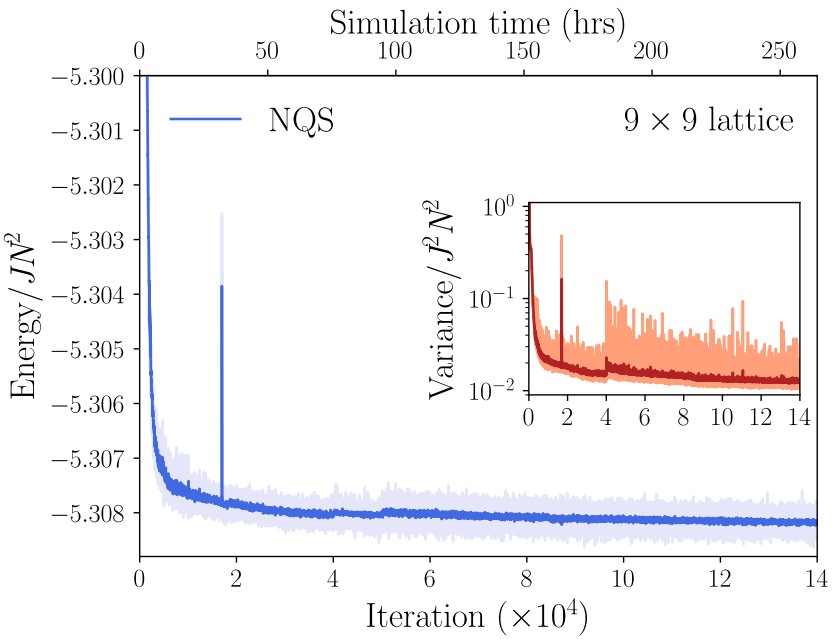
$$\Delta E = \frac{|E_{\text{NQS}} - E_{\text{exact}}|}{|E_{\text{exact}}|}. \tag{4.5}$$

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 Figure 4.2(b).

The ground state diagram for the 9×9 lattice is shown in Figure 4.3, depending on the DMI, D, and the anisotropy, A. The quantum skyrmion is the ground state for a wide range of parameters (triangles in Figure 4.3, especially at stronger DMI, which favors a noncolinear alignment of the neighboring spins. The spin expectation value at the i-th site, $\langle \boldsymbol{S}_i \rangle = \langle \sigma_i/2 \rangle$, for the ground state at $D = 0.5J$ and $A = 0.2J$ is



(a)



(b)

Figure 4.2: 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 4.2 for hardware specifications). The inset in (a) shows the relative error ΔE in the ground state energy (see Eq. (4.5)) 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 [21].

shown in Figure 4.4(a). A fundamental difference from the case of classical magnetic skyrmions is that the expectation value of the length of the spins, $|\langle \boldsymbol{S}_i \rangle|$, is reduced in the quantum skyrmion state. For the ground state in Figure 4.4(a), $|\langle \boldsymbol{S}_i \rangle|$ 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 quantum skyrmion ground states, the minimum of $|\langle \boldsymbol{S}_i \rangle| = 0.90\frac{\hbar}{2}$ 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. Finally, 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 Figure 4.3). An example is shown in Figure 4.4(b) for $D = 0.1J$ and $A = J$.

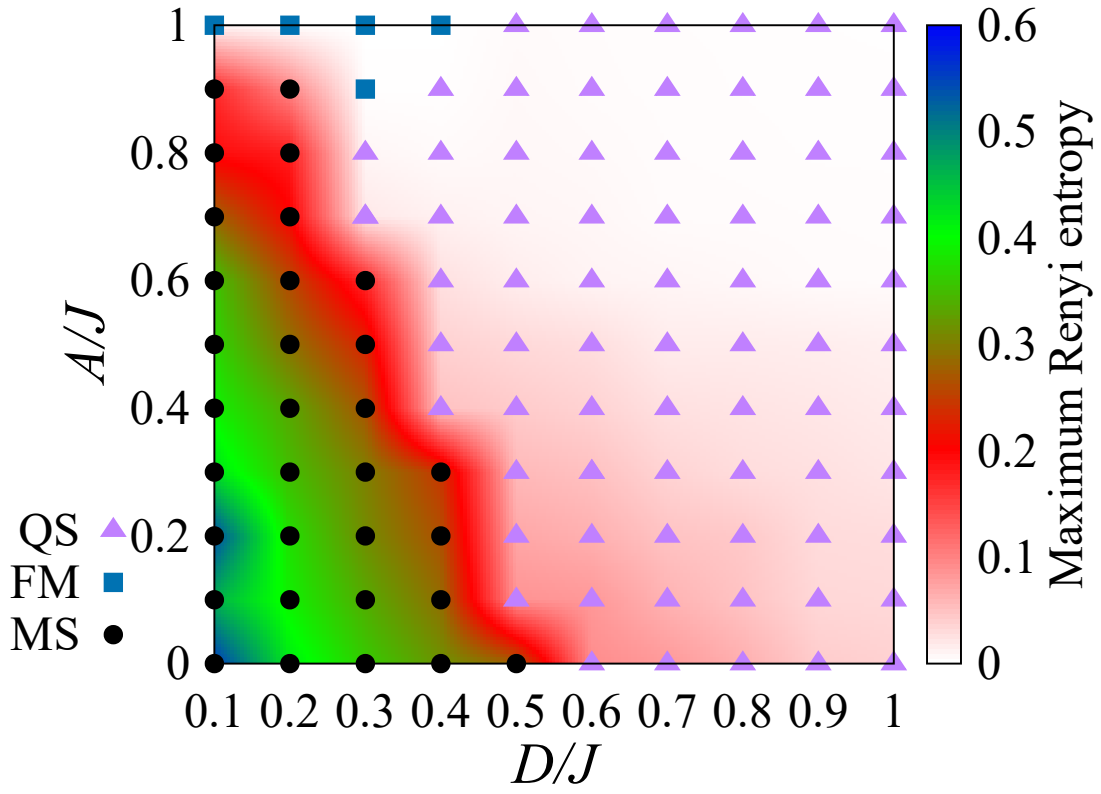


Figure 4.3: Ground state diagram for a 9×9 square lattice. QS denotes the quantum skyrmion state, FM the ferromagnetic state aligned with the boundary fields, and MS the mixed state (see the main text). The color map shows the maximum Renyi entropy between two spins [21].

4.3.2 Weak DMI regime

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 Figure 4.3, 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 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 quantum skyrmion or an FM ground state as a "mixed state" (MS) (circles in Figure 4.3).

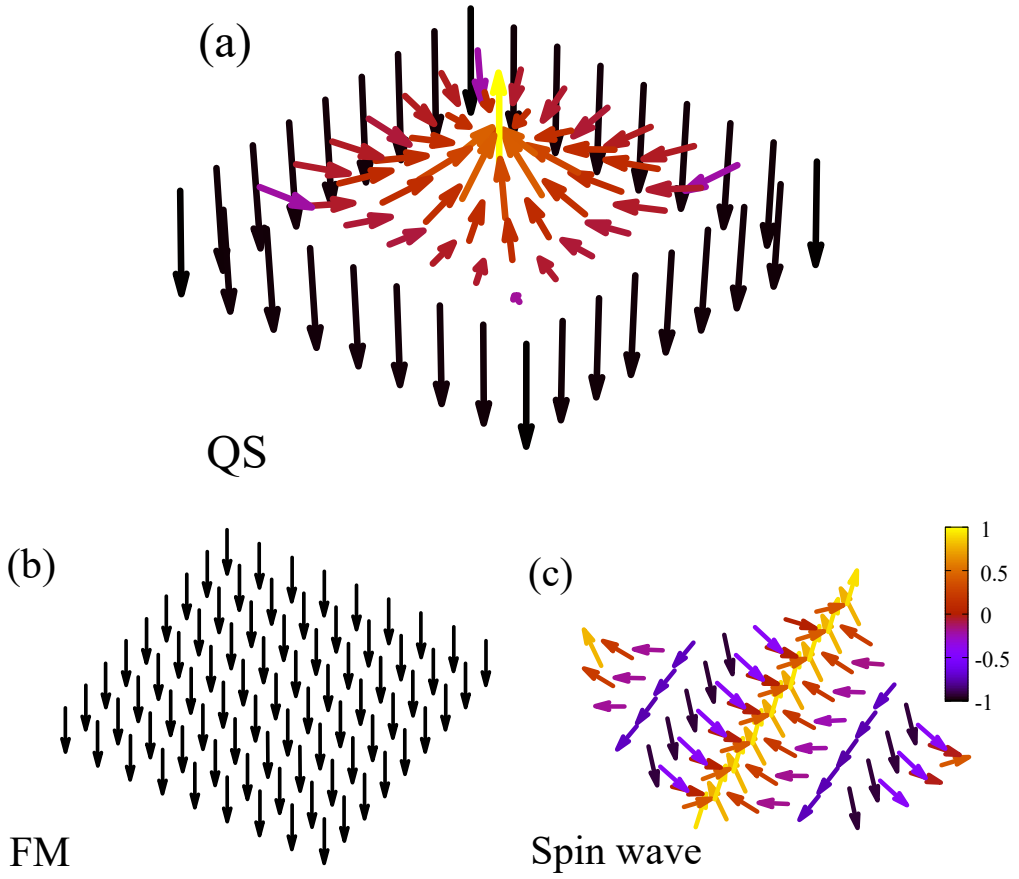


Figure 4.4: Spin expectation values of different ground states of Eq. (4.1) for a 9×9 square lattice. (a) Quantum Skyrmion state (QS) at $D = 0.5J, A = 0.2J$ and (b) Ferromagnetic state (FM) at $D = 0.1J, A = J$. (c) For periodic boundary conditions, we obtain a cycloidal spin spiral instead of a quantum skyrmion at $D = J, A = 0.5J$ [21].

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, $|\psi^1_{\boldsymbol{\theta}}\rangle$, orthogonal to the ground state NQS, $|\psi^0_{\boldsymbol{\theta}}\rangle$, by adding an additional term in the loss function

$$L_{\boldsymbol{\theta}} = \langle \psi^1_{\boldsymbol{\theta}} | H | \psi^1_{\boldsymbol{\theta}} \rangle + J | \langle \psi^0_{\boldsymbol{\theta}} | \psi^1_{\boldsymbol{\theta}} \rangle |^2. \tag{4.6}$$

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 $|\psi^0_{\boldsymbol{\theta}}\rangle$ and $|\psi^1_{\boldsymbol{\theta}}\rangle$ respectively, and plot it over the DMI in Figure 4.5. For $\Delta E_g < 2 \times 10^{-4}$, we do not obtain a quantum skyrmion 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 [14, 17, 13]. 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.

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. (4.1) 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 [9].

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 section 3.3. Given a Hamiltonian H, its eigenvalue equation is

$$H |\phi\rangle = E |\phi\rangle , \tag{4.7}$$

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

$$\frac{1}{\Omega}\left(\sum_{n_i,n_j}\langle n_j|H|n_i\rangle\langle n_i|\phi\rangle - E\langle n_j|n_i\rangle\langle n_i|\phi\rangle\right) = 0. \tag{4.8}$$

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.

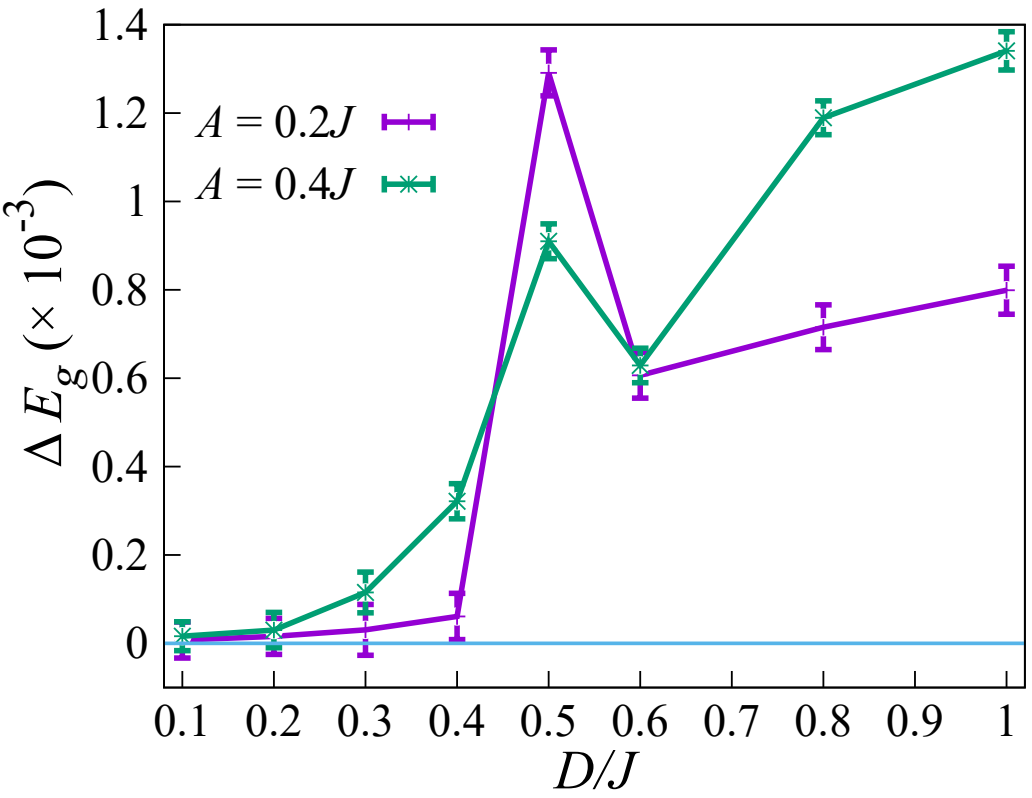


Figure 4.5: Relative energy gap ΔE_g between the ground state and the first excited state found by the NQS method over DMI at $A = 0.2J$ and $A = 0.4J$ [21].

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

$$L_{\boldsymbol{\theta}} = \langle n_1|H|n_1\rangle + J|\langle n_0|n_1\rangle|^2, \tag{4.9}$$

as described in the main text in Eq. (4.6), 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. (4.9), 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. (3.12) 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 $|n_0(x)||n_1(x)|$, as it gave us the best results. The projected Hamiltonian and the overlap matrix are then given as

$$\frac{\langle n_i|H|n_j\rangle}{\Omega} = \frac{\sum_{xx'} \frac{|n_0(x)||n_1(x)|}{|n_0(x)||n_1(x)|} n_i^\star(x') n_j(x) \langle x|H|x'\rangle}{\sum_x |n_0(x)||n_1(x)|} \tag{4.10}$$

$$\frac{\langle n_i|n_j\rangle}{\Omega} = \frac{\sum_x \frac{|n_0(x)||n_1(x)|}{|n_0(x)||n_1(x)|} n_i^\star(x) n_j(x)}{\sum_x |n_0(x)||n_1(x)|} \tag{4.11}$$

which determines the normalization constant in Eq. (4.8) as $\Omega = \sum_x |n_0(x)||n_1(x)|$. The wave functions in Eq. (4.8) 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. (4.8), 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. (4.8). 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.

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. Our model is similar to the one in [14], where the authors found that quantum skyrmions can appear as a ground state for even infinitesimal DMI, which

is not found for classical skyrmions (see Fig. 1 in [14]). These quantum skyrmions in the small DMI regime have vanishing magnitude of spin expectation values. We expect this to be the case in our model as well, where some regions of small DMI host a quantum skyrmion ground state.

4.3.3 Skyrmion number

In addition to the spin expectation values, we calculate the skyrmion number C using the normalized spin expectation values, $\mathbf{n}_i = \langle \boldsymbol{S}_i \rangle / |\langle \boldsymbol{S}_i \rangle|$, to define quantum skyrmions (see section 2.2)

$$C = \frac{1}{2\pi} \sum_{\Delta} \tan^{-1} \left(\frac{\mathbf{n}_i \cdot (\mathbf{n}_j \times \mathbf{n}_k)}{1 + \mathbf{n}_i \cdot \mathbf{n}_j + \mathbf{n}_j \cdot \mathbf{n}_k + \mathbf{n}_k \cdot \mathbf{n}_i} \right), \tag{4.12}$$

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. (4.12), $\mathbf{n}_i = \langle 2\boldsymbol{S}_i \rangle$, results in a non-integer number Q that indicates the 'quantum' nature of skyrmions [14], similar to other quantum measures [17]. Q decreases as the entanglement increases and the spin expectation values decrease. For the quantum skyrmion 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 quantum skyrmion ground state. Instead, we obtain a cycloidal spin spiral (Figure 4.4(c)), which is consistent with DMRG findings [16] and the fact that unfrustrated classical skyrmions require a magnetic field for stabilization. Here, a quantum skyrmion 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.

4.4 Entanglement in quantum skyrmions

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. Renyi entropy generalizes different kinds of entropy measures like Hartley entropy and Shannon's entropy. The Renyi entropy of the order α, where $\alpha \geq 0$ and $\alpha \neq 1$, is defined as,

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha}\log(\mathrm{Tr}(\rho_A^\alpha)). \tag{4.13}$$

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 [53, 111] in the VMC scheme, as described below.

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

$$|\psi_{\boldsymbol{\theta}}\rangle = \sum_{\sigma_A\sigma_B} \psi_{\boldsymbol{\theta}}(\sigma_A\sigma_B)\,|\sigma_A\rangle\,|\sigma_B\rangle\,, \tag{4.14}$$

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

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha}\log(\mathrm{Tr}(\rho_A^\alpha)), \tag{4.15}$$

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 [111]

$$\begin{aligned}
&\mathrm{Swap}_A\,|\psi_{\boldsymbol{\theta}}\rangle \otimes |\psi_{\boldsymbol{\theta}}\rangle = \mathrm{Swap}_A\left(\sum_{\sigma_A\sigma_B} \psi_{\boldsymbol{\theta}}(\sigma_A\sigma_B)\,|\sigma_A\rangle\,|\sigma_B\rangle\right)\\
&\otimes\left(\sum_{\sigma'_A\sigma'_B} \psi_{\boldsymbol{\theta}}(\sigma'_A\sigma'_B)\,|\sigma'_A\rangle\,|\sigma'_B\rangle\right)\\
&= \sum_{\sigma_A\sigma_B} \psi_{\boldsymbol{\theta}}(\sigma_A\sigma_B) \sum_{\sigma'_A\sigma'_B} \psi_{\boldsymbol{\theta}}(\sigma'_A\sigma'_B)\,|\sigma'_A\rangle\,|\sigma_B\rangle \otimes |\sigma_A\rangle\,|\sigma'_B\rangle\,,
\end{aligned} \tag{4.16}$$

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,

$$\begin{aligned}
\langle \text{Swap}_A \rangle &= \frac{\langle \psi_{\boldsymbol{\theta}} \otimes \psi_{\boldsymbol{\theta}} | \, \text{Swap}_A \, | \psi_{\boldsymbol{\theta}} \otimes \psi_{\boldsymbol{\theta}} \rangle}{\langle \psi_{\boldsymbol{\theta}} \otimes \psi_{\boldsymbol{\theta}} | \psi_{\boldsymbol{\theta}} \otimes \psi_{\boldsymbol{\theta}} \rangle} \\
&= \frac{\sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} \psi^*_{\boldsymbol{\theta}}(\sigma_A \sigma_B) \psi^*_{\boldsymbol{\theta}}(\sigma'_A \sigma'_B) \psi_{\boldsymbol{\theta}}(\sigma'_A \sigma_B) \psi_{\boldsymbol{\theta}}(\sigma_A \sigma'_B)}{\sum_{\sigma\sigma'} |\langle \psi_{\boldsymbol{\theta}} \otimes \psi_{\boldsymbol{\theta}} | \sigma \otimes \sigma' \rangle|^2} \\
&= \text{Tr}(\rho_A^2) \\
&= \exp(-S_2(\rho_A)).
\end{aligned} \tag{4.17}$$

For the final step we use the definition in Eq. (4.15) with $\alpha = 2$. In the Monte Carlo scheme, Eq. (4.17) can be evaluated as

$$\begin{aligned}
&\langle \text{Swap}_A \rangle \\
&= \sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} \frac{|\psi_{\boldsymbol{\theta}}(\sigma_A \sigma_B)|^2}{\sum_\sigma |\psi_{\boldsymbol{\theta}}(\sigma)|^2} \frac{|\psi_{\boldsymbol{\theta}}(\sigma'_A \sigma'_B)|^2}{\sum_{\sigma'} |\psi_{\boldsymbol{\theta}}(\sigma')|^2} \cdot \\
&\qquad\qquad \frac{\psi_{\boldsymbol{\theta}}(\sigma'_A \sigma_B) \psi_{\boldsymbol{\theta}}(\sigma_A \sigma'_B)}{\psi_{\boldsymbol{\theta}}(\sigma_A \sigma_B) \psi_{\boldsymbol{\theta}}(\sigma'_A \sigma'_B)} \\
&= \sum_{\sigma_A \sigma_B \sigma'_A \sigma'_B} p_{\boldsymbol{\theta}}(\sigma) p_{\boldsymbol{\theta}}(\sigma') \frac{\psi_{\boldsymbol{\theta}}(\sigma'_A \sigma_B) \psi_{\boldsymbol{\theta}}(\sigma_A \sigma'_B)}{\psi_{\boldsymbol{\theta}}(\sigma_A \sigma_B) \psi_{\boldsymbol{\theta}}(\sigma'_A \sigma'_B)}.
\end{aligned} \tag{4.18}$$

The maximum Renyi entropy associated with the parameters is shown as a heatmap in the ground state diagram in Figure 4.3. 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 quantum skytmion ground state in Figure 4.6 for the parameters $D = 0.5J$ and $A = 0.2J$. As the boundary spins are fixed with a large magnetic field, they 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 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 quantum skyrmion ground state 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.

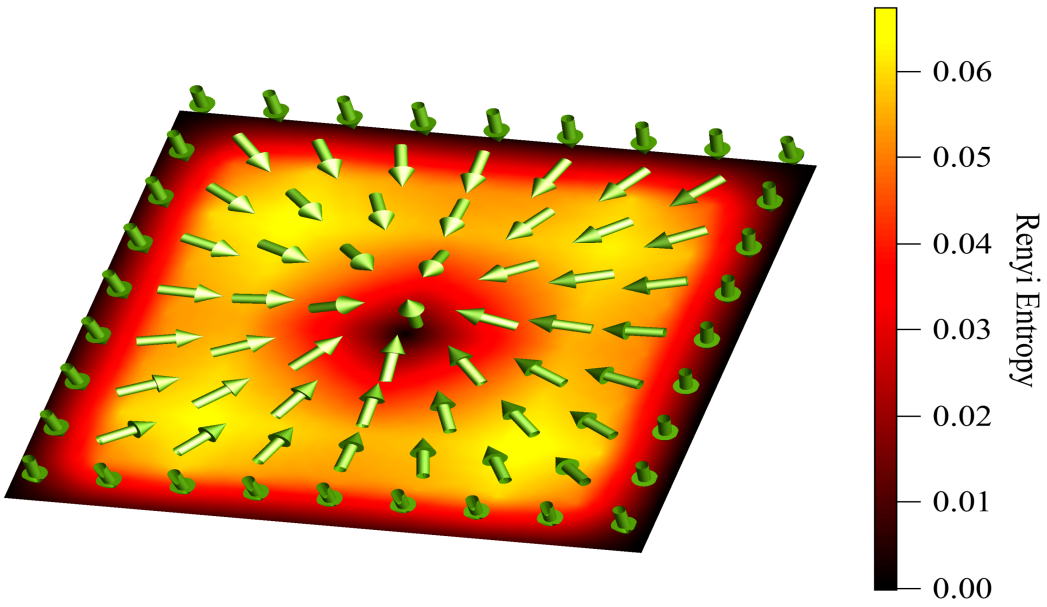


Figure 4.6: Renyi entropy of each spin with its environment for the QS at $D = 0.5J$ and $A = 0.2J$ [21].

We note that our results of the entropy for the QS ground state match with those in [16], 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.

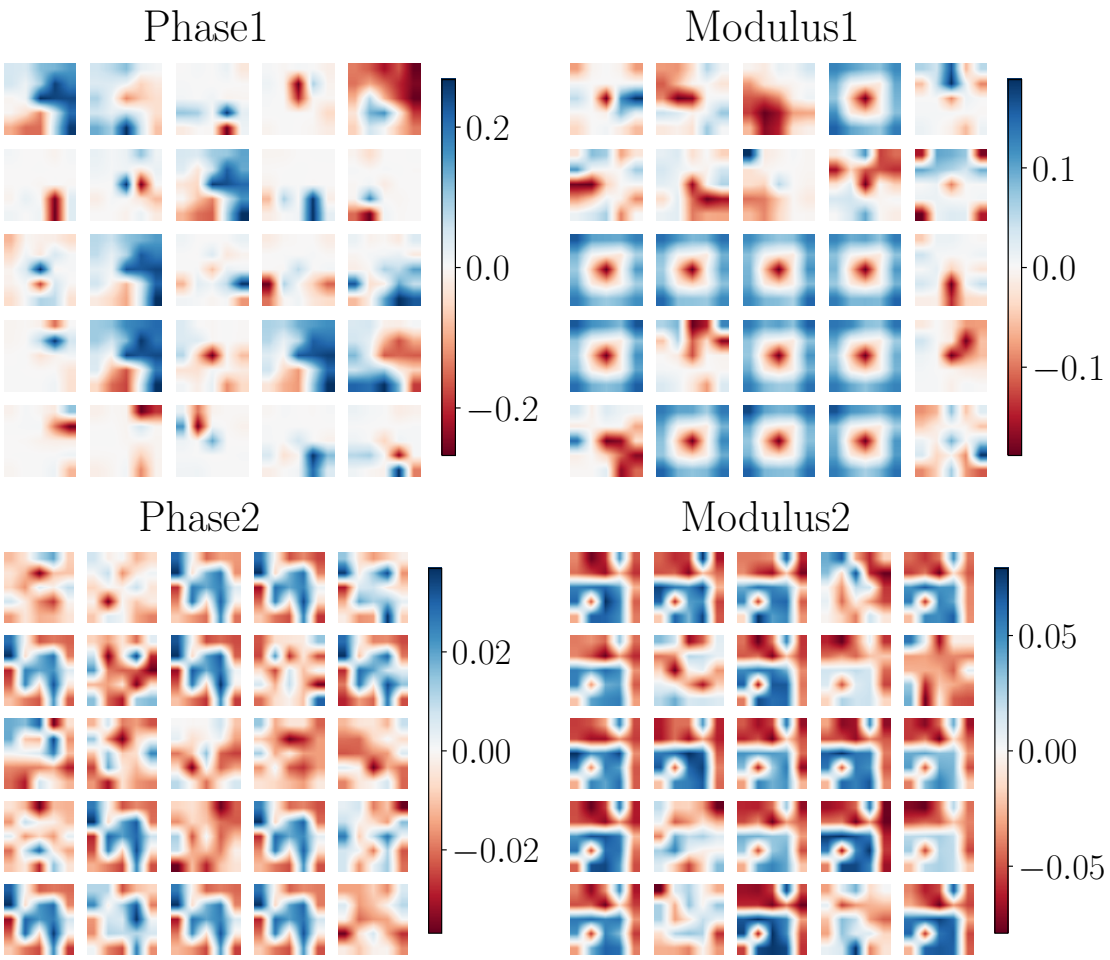


Figure 4.7: Weight distribution in the hidden layers of the quantum skyrmion ground state in a 5×5 lattice at $D = J$ and $A = 0.5J$. Phase1 (Phase2) and Modulus1 (Modulus2) denote the phase and modulus parts in the first (second) hidden layer, respectively. Each block shows the weights inside one hidden 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 [21].

4.5 Neural 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 in the present network shown in Figure 4.1, 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 [42, 52]. 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 Figure 4.1. 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 Figure 4.7. 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 [112, 113, 114]. 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 [115].

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 Figure 4.8(a)-(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 [116],

$$\Delta E_p = \frac{|E_{\text{full}} - E_p|}{|E_{\text{full}}|}, \tag{4.19}$$

$$F = |\langle \psi_{\text{full}} | \psi_p \rangle|^2, \tag{4.20}$$

where $p = pr, prsc$. To calculate the fidelity between two NQSs, $|\psi_1\rangle$ and $|\psi_2\rangle$ (dropping the dependence on $\boldsymbol{\theta}$ for clarity), we follow a similar procedure as in Eq. (3.12),

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

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.

In Figure 4.8(a)-(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 Figure 4.8(a) and (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.

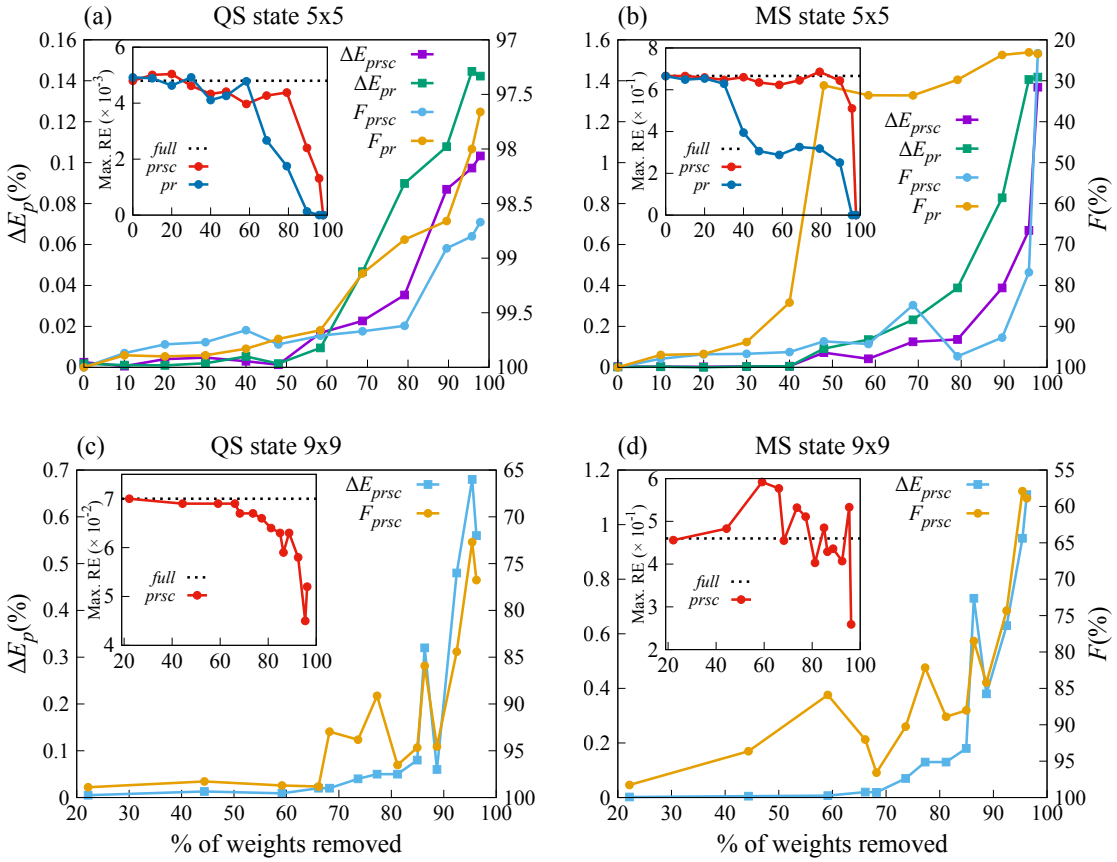


Figure 4.8: Performance metrics after pruning the neural network. ΔE_p denotes the relative error in energy and F denotes the fidelity. *pr* denotes a pruned NQS trained after removing the weights from the full NQS and *prsc* denotes an identical network to the *pr* one but trained from scratch. (a) 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 [21].

In Figure 4.8(c) and (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.

4.6 Summary

In this chapter, 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 is an important task for development of these methods.

Chapter 5

Dynamics of quantum skyrmions

In this chapter we present our second paper in which we study the real-time evolution of a quantum skyrmion lattice in the presence of an external magnetic field gradient using neural-network quantum states (NQS) and time dependent variational Monte Carlo (t-VMC). First, we obtain a quantum skyrmion lattice as the ground state of a two-dimensional spin-1/2 Heisenberg Hamiltonian with Dzyaloshinskii-Moriya interaction (DMI). The spins in this quantum skyrmion lattice have nonzero quantum entanglement, similar to the ground states in previous chapter. Then, we quench the Hamiltonian with a nonuniform external magnetic field and evolve the system according to the time-dependent Schrödinger equation using t-VMC. We show that quantum skyrmions move diagonally to the field gradient, resembling a skyrmion Hall effect, with a velocity that is larger in the direction perpendicular to the magnetic field gradient. The quantum skyrmions interact with each other, leading to the formation of an exceptional configuration with the topological charge of a meron, which causes the decay of a quantum skyrmion. Merons and antimerons are vortex-like spin textures that are quantized to half the skyrmion number N, a topological invariant used to characterize skyrmions (see section 2.2). Our work shows that NQS can be used as a variational ansatz to study the ground state and nonequilibrium properties of quantum skyrmions with system sizes that are not feasible using exact methods.

The quantum spin model we study is described in section 5.1. The details of the method and hyperparameters used are described in section 5.2. We present our ground state results in section 5.3 and time evolution of quantum skyrmions in section 5.4. We summarize this chapter in section 5.5.

5.1 Model

We study the same Hamiltonian as in (4.1) but with different boundary conditions - a spin-1/2 Heisenberg Hamiltonian with DMI and anisotropy on a two-dimensional lattice with periodic boundaries,

$$\begin{aligned} H_0 = &- J \sum_{\langle ij \rangle} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y) - A \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z \\ &- D \sum_{\langle ij \rangle} (\boldsymbol{u}_{ij} \times \hat{\boldsymbol{z}}) \cdot (\boldsymbol{\sigma}_i \times \boldsymbol{\sigma}_j) + B^z \sum_i \sigma_i^z . \end{aligned} \tag{5.1}$$

Here, J is the Heisenberg exchange term, A is the Heisenberg anisotropy term, D is the strength of the DMI, and B^z is the strength of the homogeneous external magnetic field. We take $\hbar = 1$. The Pauli matrices on the i-th site are denoted by $\boldsymbol{\sigma}_i = \{\sigma_i^x, \sigma_i^y, \sigma_i^z\}$ and $\boldsymbol{u}_{ij}$ is the unit vector from site i to site j. The first three terms are summed over the nearest neighbors denoted by $\langle ij \rangle$. A quantum skyrmion state can emerge due to the competition between the ferromagnetic exchange term and the noncolinear DMI term, stabilized by the anisotropy and the external magnetic field.

5.2 Method

To obtain the ground state of the Hamiltonian in Eq. (5.1), we use variational Monte Carlo with a neural-network quantum state (NQS) as the variational ansatz. The many-body wave function is approximated using an artificial neural network that encodes the complex-valued coefficients $\psi_{\boldsymbol{\theta}}(x)$,

$$|\psi_{\boldsymbol{\theta}}\rangle = \sum_x \psi_{\boldsymbol{\theta}}(x) \, |x\rangle \, . \tag{5.2}$$

Here, $\boldsymbol{\theta}$ are the variational parameters, and $|x\rangle$ are the local basis states, which in our case are the eigenvalues of the σ_j^z operators. We use a restricted Boltzmann machine (RBM) with complex weights and biases as the variational wave function. The RBM consists of an input layer that takes the spin configurations $|x\rangle$ as input and a hidden layer with variational parameters $\boldsymbol{\theta} = (a, W, b)$, see Figure 5.1. Here, a are input biases, and W and b are hidden weights and biases, respectively. The length of one side of the lattice is given by L, and α is the hidden unit density. In this study, $\alpha = 2$ is used for both ground state and time evolution calculations. Increasing α increases the expressiveness of the network, resulting in slightly better

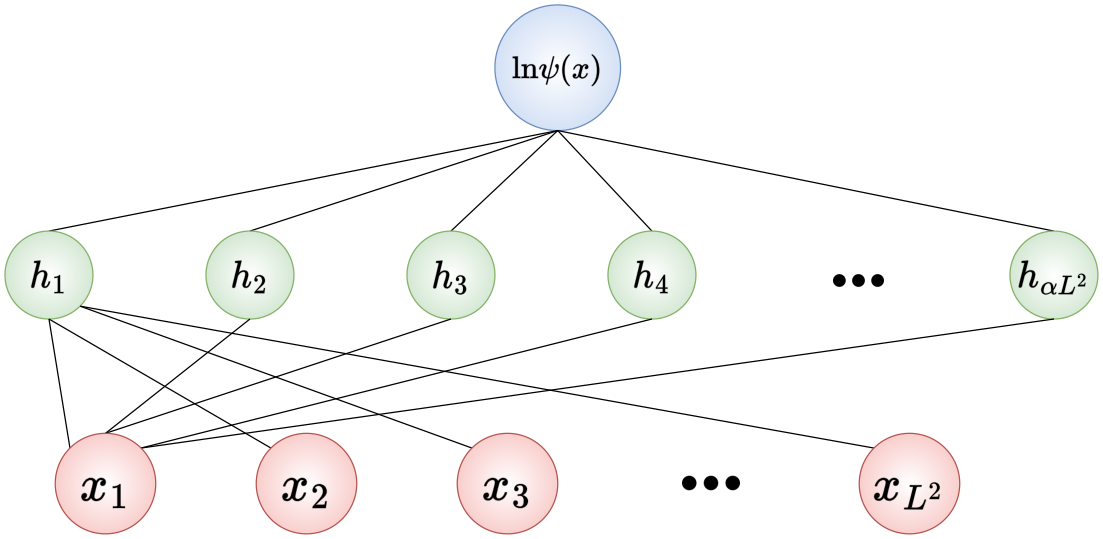


Figure 5.1: Restricted Boltzmann machine used as the neural network quantum state. The inputs are the spin configurations in σ^z basis, and the output is the logarithm of the wave function (Eq. (5.3)). The hidden layer contains αL^2 neurons, where $\alpha = 2$ in our case [22].

energies but with higher computational cost. The output is the logarithm of the unnormalized wave function

$$\ln(\psi_{\boldsymbol{\theta}}(x)) = \sum_{i}^{\alpha L^2} a_i x_i + \ln\cosh[Wx + b]_i. \tag{5.3}$$

It is important to choose differentiable activation functions for deriving the real-time evolution, which relies on the wave function being differentiable at every point of the variational manifold [59]. Thus, instead of the reLU(x) activation function (nondifferentiable at $x = 0$), that was used previously to study the ground states of quantum skyrmions [21], in this work, we use an RBM with ln cosh(x) activation function. The loss function $\mathcal{L}_0(\boldsymbol{\theta})$ for ground state calculations is the energy of the Hamiltonian H_0 which is minimized with respect to the variational parameters $\boldsymbol{\theta}$

$$\mathcal{L}_0(\boldsymbol{\theta}) = \langle\psi_{\boldsymbol{\theta}}|\, H_0\, |\psi_{\boldsymbol{\theta}}\rangle\,. \tag{5.4}$$

The weights and biases are initialized randomly with a normal distribution having a standard deviation of 0.01. To optimize the RBM using gradient descent, we use the Adam optimizer with the moments $\beta_1 = 0.9$ and $\beta_2 = 0.999$ [107]. The learning rate η is varied from $\eta = 10^{-3}$ to $\eta = 10^{-5}$ in the steps of 10^{-1} after every 4×10^4 iterations. Using a stochastic gradient descent optimizer with stochastic reconfiguration [42] gives similar results but with increased computational costs. As the Hilbert space is very large, we use Markov chain Monte Carlo to generate

samples that are used in the computation of expectation values. The samples are generated by flipping one spin randomly, and the process is repeated L^2 times to complete one Monte Carlo sweep. We use 2^{14} samples for energy calculation and 2^{17} samples for all the other expectation values.

For the real-time evolution, we use the time-dependent variational principle, which corresponds to the time dependence of the variational parameters, $\boldsymbol{\theta}(t)$, as described in the section 3.4 . The RBM representing the ground state is used as the initial state at $tJ = 0$. We use a time step of $\delta t = 10^{-4}$. At each time step, the quantum geometric tensor S and the forces vector F are computed with 2^{14} samples. The equation of motion, Eq. (3.31), can be very unstable due to the presence of noise in the calculation of the matrix S [59, 94, 117]. To improve stability, we add a small shift of $\epsilon = 0.01$ to the diagonal elements of the S matrix to regularize the equation of motion. We experimented with different values of ϵ and found that while the quantum skyrmion motion was similar for all $1.0 < \epsilon < 10^{-5}$ qualitatively, a smaller ϵ resulted in unstable energy. Finally, to integrate Eq. (3.31), we use a fourth-order Runge-Kutta integration scheme. Both the ground state optimization and real-time evolution calculations were performed on an NVIDIA A100 GPU.

The procedure to obtain the real time evolution after a quench using NQS is given as:

Algorithm 2 Real time evolution with NQS using TDVP

$t \leftarrow 0$
Quench the Hamiltonian
while $t \leq t_{end}$ **do**
  $t \leftarrow t + \delta t$
  Calculate $\boldsymbol{S}$ and $\boldsymbol{F}$ (Eq. (3.32)) for $\psi_{\boldsymbol{\theta}(t)}$
  Calculate $\boldsymbol{\theta}(t + \delta t)$ (Eq. (3.31)) using fourth order Runge-Kutta scheme
end while

5.3 Ground state analysis

A quantum skyrmion lattice (QSL) is the ground state of the Hamiltonian in Eq. (5.1) for large DMI and finite anisotropy and magnetic field if the lattice size is large enough to accommodate the QSL, consistent with previous findings [16, 21]. The ground state energy E_0 minimization plot for the RBM used to describe the QSL in a 9×9 lattice is shown in Figure 5.2, with the energy variance $\langle \psi_{\boldsymbol{\theta}} | (H_0 - E_0)^2 | \psi_{\boldsymbol{\theta}} \rangle$ in the inset. Here, the Hamiltonian parameters are $D = J$, $A = 0.5J$, and $B^z = J$.

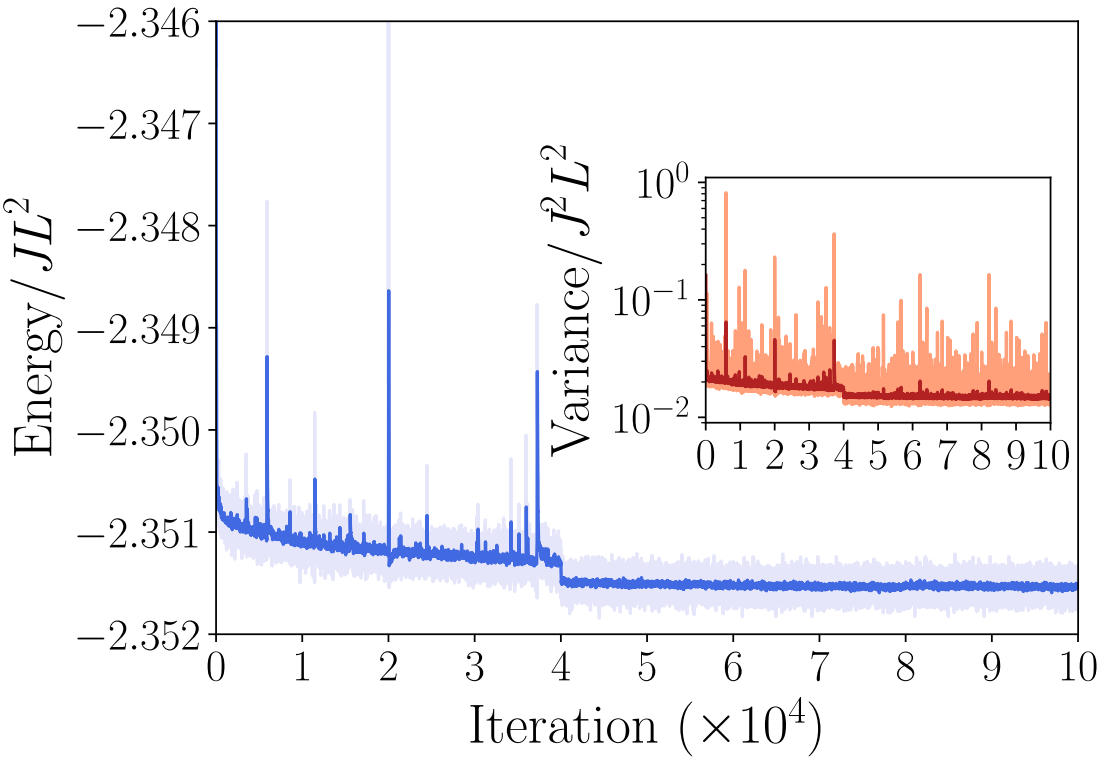


Figure 5.2: Convergence of variational energy per spin over the number of iterations for a 9×9 lattice, with the energy variance per spin in the inset. The lighter color shows the values at each iteration, while the darker color shows the moving average over 30 iterations. The Hamiltonian parameters are $D = J$, $A = 0.5J$ and $B^z = J$ [22].

The RBM converges to a QSL as the variance vanishes. The spin expectation values $\langle \boldsymbol{S} \rangle = \langle \boldsymbol{\sigma} \rangle /2$ in Figure 5.3 show two quantum skyrmions in the ground state, encircled by dashed lines. As this is a quantum spin model, the lengths of the spins are not normalized due to quantum fluctuations, and thus $|\langle \boldsymbol{S} \rangle| < 1/2$.

To characterize quantum skyrmions, we calculate the local skyrmion density for the nearest neighbor spins i, j, and k forming an elemental triangle Δ as [68, 14, 20]

$$\Omega_\Delta = \frac{1}{2\pi} \text{atan2}(\mathbf{n}_i \cdot (\mathbf{n}_j \times \mathbf{n}_k), 1 + \mathbf{n}_i \cdot \mathbf{n}_j + \mathbf{n}_j \cdot \mathbf{n}_k + \mathbf{n}_k \cdot \mathbf{n}_i). \tag{5.5}$$

Here, we use the normalized spin expectation values $\mathbf{n}_i = \langle \boldsymbol{S}_i \rangle / |\langle \boldsymbol{S}_i \rangle|$. The skyrmion number N is given by the sum over all triangles

$$N = \sum_\Delta \Omega_\Delta. \tag{5.6}$$

Using twice the unnormalized spin expectation values in Eq. (5.5) instead of $\mathbf{n}$ results in a non-quantized number Q, which depends on the length $|\langle \boldsymbol{S} \rangle|$ of the

spins and is an indicator of the stability of quantum skyrmions, with $Q \to N$ if and only if the spin expectation values have maximal amplitude [14] and quantum fluctuations completely vanish. For the ground state solution in Figure 5.3, we find $N = 2$ corresponding to two quantum skyrmions in the ground state and $Q = 1.93$, implying that these skyrmions have spin expectation values with magnitude close to $\hbar/2$.

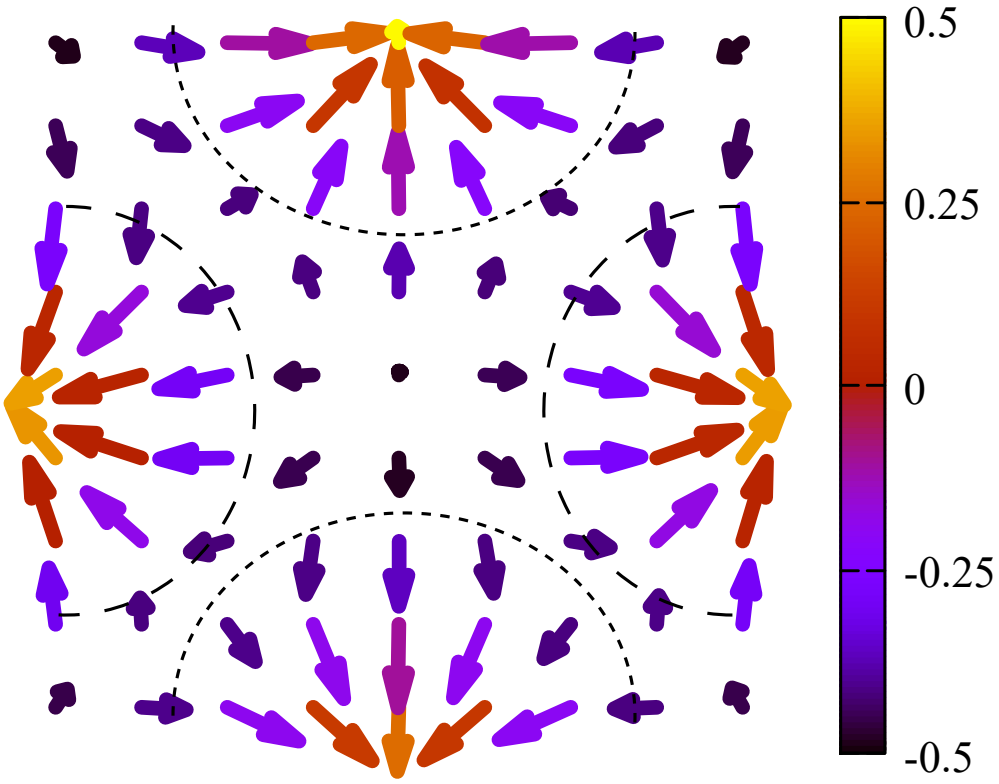


Figure 5.3: Quantum skyrmion lattice ground state spin expectations of Eq. (5.1) at parameters $D = J$, $A = 0.5J$ and $B^z = J$. The skyrmion number $N = 2$ as there are two quantum skyrmions in the lattice. The color map indicates the z component of the spin expectation value [22].

We note that the existence of a QSL depends not only on the DMI D, anisotropy A, and external magnetic field B^z but also on the size of the lattice. For square lattices smaller than 9×9 spins, we do not find any ground state hosting a quantum skyrmion in the parameter range $0 < D/J < 2$, $0 < A/J < 2$ and $0 < Bz < 2$. While it is possible to obtain a QS ground state in smaller lattices when embedded

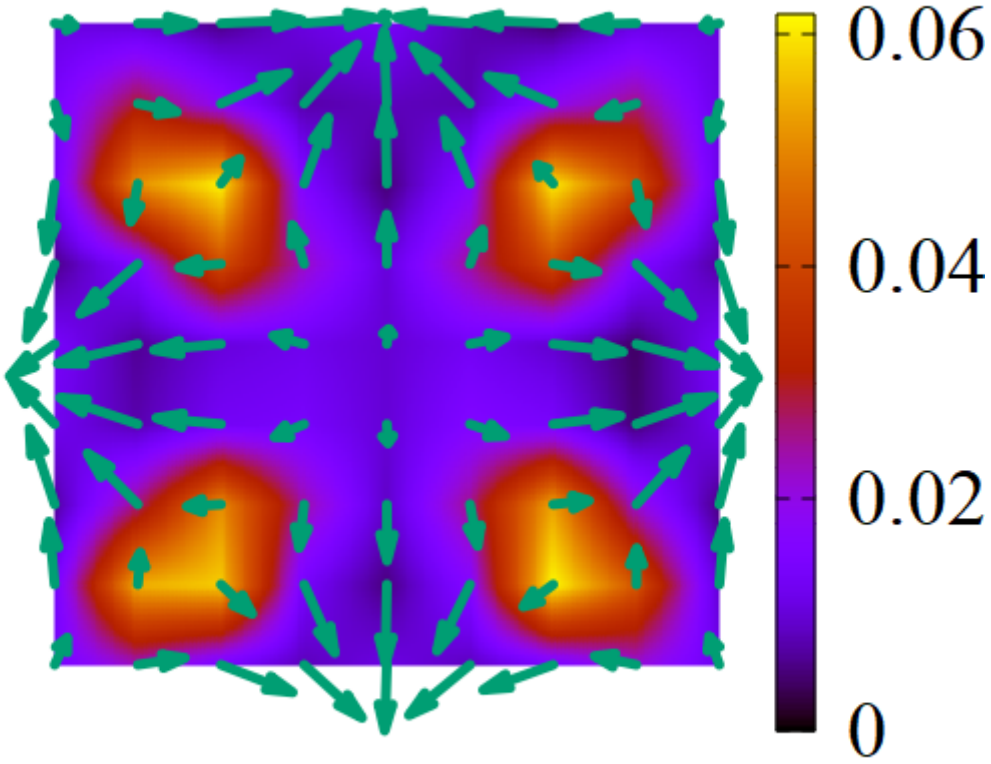


Figure 5.4: Renyi entropy in the ground state shown in Figure 5.3. The heatmap shows the entropy of a single spin with all the other spins. The entropy is largest in the space between two skyrmions [22].

in a ferromagnetic medium [14, 21], in the presence of periodic boundaries, we only obtain a spin spiral or a ferromagnet as the ground state. For larger lattice sizes up to 13×13, we also obtain a QSL as the ground state with $N = 2$ for large DMI.

Next, we study entanglement in the QSL ground state as previously done for single quantum skyrmions in section 4.4 [16, 21]. Using the expectation value of the "Swap" operator, we calculate the second order Renyi entropy $S_2(\rho_A)$ as a measure of entanglement in quantum skyrmions [111, 53, 21],

$$S_2(\rho_A) = -\frac{1}{2}\ln(\mathrm{Tr}(\rho_A^2)). \tag{5.7}$$

Here, ρ_A is the reduced density matrix obtained by dividing the system into subsystems A and B and tracing out the degrees of freedom in subsystem B. In all our Renyi entropy calculations, we take subsystem A to be a single spin and partition B

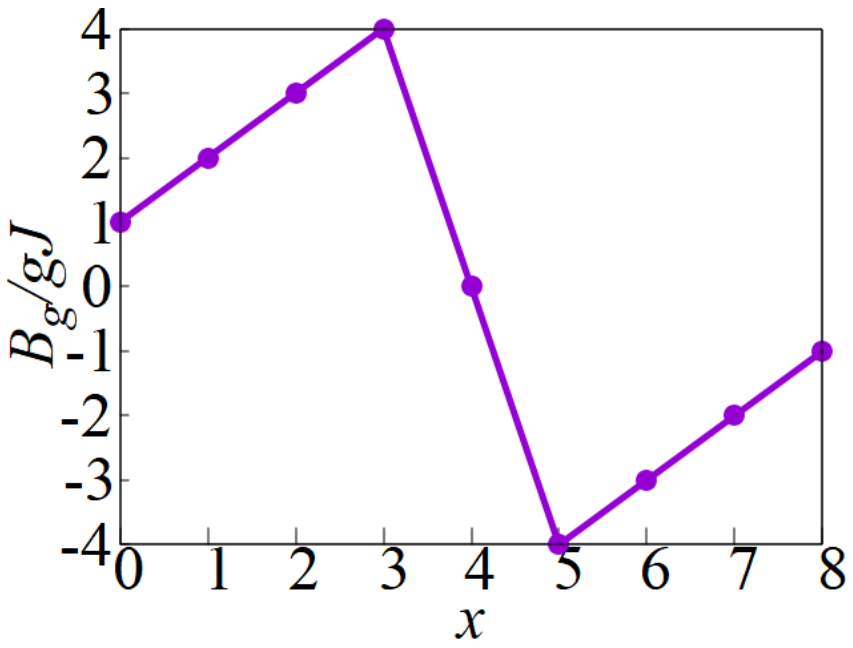
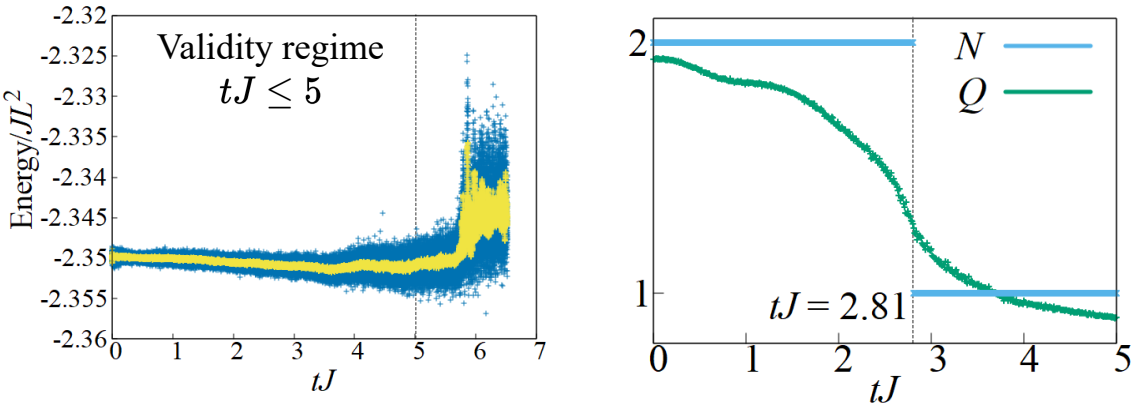


Figure 5.5: Magnetic field gradient used to move quantum skyrmions in Eq. (5.8). The magnetic field points along the z-axis and depends on the x-coordinate [22].

to be the remaining spins to see how the spins are entangled with their environment. The heat map in Figure 5.4 shows the Renyi entropy in the QSL ground state. The entanglement is largest ($S_2(\rho_A) = 0.061$) in the region between two skyrmions and smallest ($S_2(\rho_A) = 0.004$) around the center of the skyrmions. The entropy for the central spin is nonzero ($S_2(\rho_A) = 0.013$), different from the case of a quantum skyrmion embedded in a ferromagnetic medium where the central spin was disentangled from the rest of the lattice. This might be due to different parameter regimes, boundary conditions, and system sizes [16, 21]. As $0 \leq S_2(\rho_A) \leq \ln(2)$, the Renyi entropies are still small in the QSL ground state.

5.4 Real time evolution

In this section, we study the real-time evolution of the QSL ground state after quenching the Hamiltonian with a magnetic field gradient. Magnetic field gradients have been shown to be an effective way of manipulating classical skyrmions and can induce a motion perpendicular to the gradient [118, 119, 120]. We quench the



(a) (b)

Figure 5.6: Energy and skyrmion number with time. (a) Time evolution of energy per spin of the quenched Hamiltonian in (Eq. (5.8)) over time as a quality check for the unitarity of the method. Blue shows the energy at each iteration, and yellow shows the moving average over 30 iterations. (b) Evolution of the normalized skyrmion number N and the unnormalized skyrmion number Q with time. While Q continuously decreases, N is quantized and a transition from $N = 2$ to $N = 1$ takes place at $tJ = 2.81$ [22].

Hamiltonian in Eq. (5.1) with a static, nonuniform magnetic field

$$H_q = H_0 + \sum_i B_i^g \sigma_i^z,$$

$$\text{where,} \quad B_i^g = \begin{cases} g(x_i + 1) & \text{if} \quad 0 \leq x_i < 4 \\ 0 & \text{if} \quad x_i = 4 \\ g(x_i - L) & \text{if} \quad 4 < x_i < L \end{cases} \tag{5.8}$$

Here, g is the strength of the gradient, x_i is the x-coordinate of i-th spin, and $x_i = 4$ is the x-coordinate of the center of one of the skyrmions at $t = 0$. The gradient is along the x-axis. With this B^g, the magnetic field gradient is largest at the center of the middle skyrmion $x_i = 4$ and decreases away from it, see Figure 5.5. The speed at which the quantum skyrmions move depends on the gradient, similar to the classical case [119]. With this choice of B^g, the interaction of quantum skyrmions can be observed in the time scales accessible by our method while maintaining the stability of the nontrivial spin structure. The ground state of the Hamiltonian H_q with magnetic field gradient is a spin spiral phase. Thus, the quench is made from a nontrivial quantum skyrmion phase to a trivial spiral phase. We therefore expect a tendency for the quantum skyrmions to eventually transition to a spiral with $N = 0$. With $g = 0.2J$, Figure 5.6(a) shows the evolution of the energy $E_q = \langle \psi_{\boldsymbol{\theta}(t)} | H_q | \psi_{\boldsymbol{\theta}(t)} \rangle$ with time t. After the quench, the Hamiltonian is time-independent, the time evolution is unitary, and the energy is supposed to be

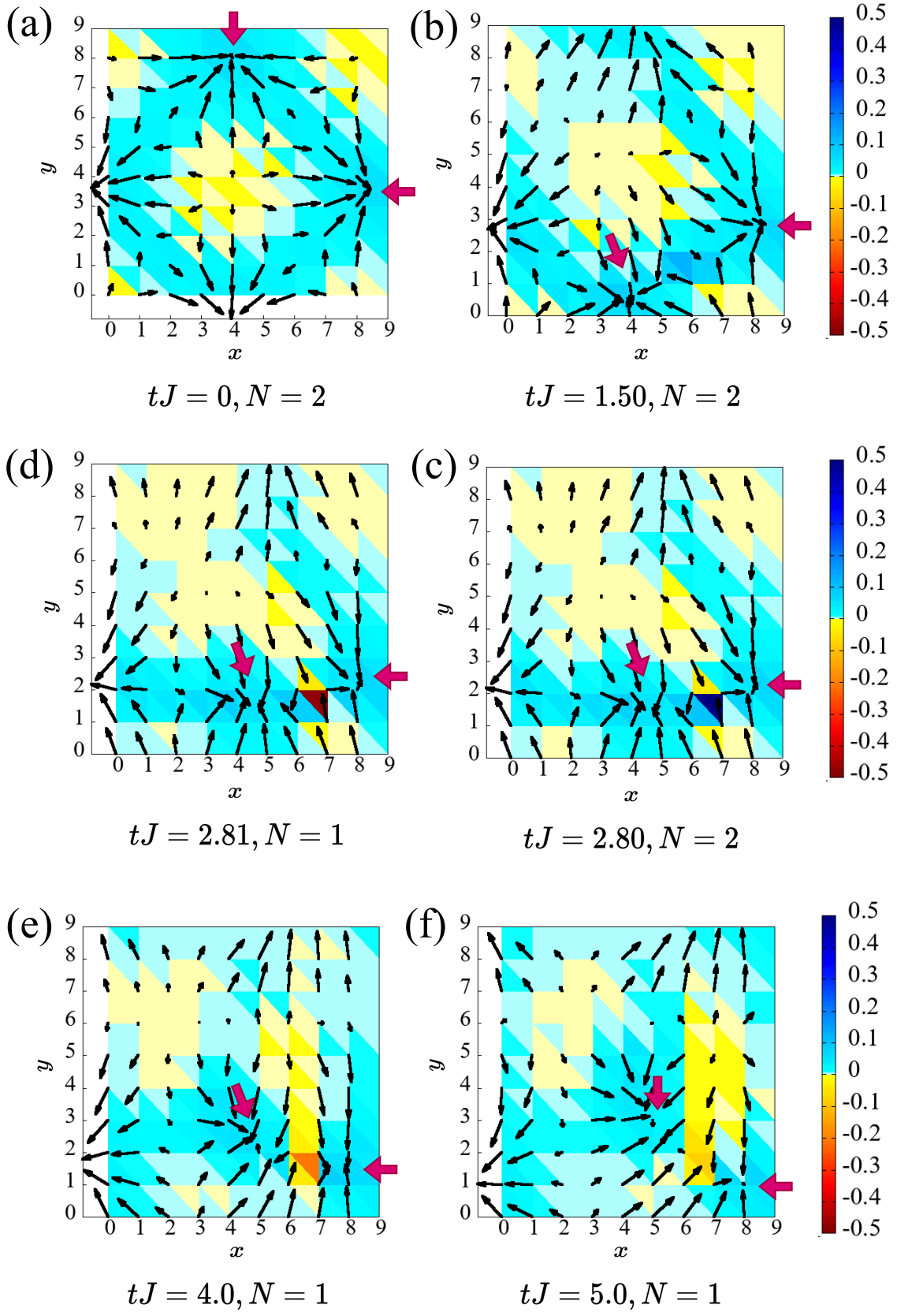


Figure 5.7: Real-time evolution of spin expectation values: snapshots of spin expectation values at different times with the skyrmion density Ω_Δ in the background. The quantum skyrmions (marked by arrows) move towards each other (a)-(b), interact and an exceptional configuration is formed between $tJ = 2.80$ and $tJ = 2.81$ (c)-(d), after which one quantum skyrmion decays and an elongated quantum skyrmion remains (e)-(f) [22].

conserved. While the energy E_q is indeed nearly constant in our simulations, we see that it changes at longer times due to the accumulation of errors [121, 117] and we constrain ourselves to the interval $tJ \leq 5$.

The time evolution of spin expectation values is shown in Figure 5.7. The color plot in the background shows the local skyrmion density Ω_Δ (Eq. (5.5)). The speed of the quantum skyrmions depends on the magnetic field gradient, and as the quantum skyrmion at $x_i = x_c = 4$ experiences a larger gradient than the quantum skyrmion at $x_i = 8$, it moves faster. The speed of the quantum skyrmions is also proportional to the magnitude of g (not shown here). However, a larger g increases the errors in t-VMC and can even destroy the QSL state. The quantum skyrmions move in a Hall-like motion [118], with the velocity perpendicular to the field gradient larger than the velocity parallel to it. The two quantum skyrmions experience opposite magnetic field gradients and move towards each other (Figure 5.7(b)). The skyrmion density Ω_Δ builds up especially for one triangle of spins at $(7, 1)$ as the two quantum skyrmions interact. The skyrmion density reaches a maximum of $\Omega_\Delta = 0.5$ for this triangle at $tJ = 2.80$ Figure 5.7(c). Then, it passes through an exceptional configuration, where the denominator in Eq. (5.5) changes sign [68], and Ω_Δ changes from $\Omega_\Delta \approx 0.5$ to $\Omega_\Delta \approx -0.5$ which results in the change of skyrmion number N from $N = 2$ to $N = 1$ in Figure 5.7(d). Thus, the quantum skyrmion decay is mediated by exceptional configurations carrying the topological charge of a meron. By this, the two quantum skyrmions merge to an elongated quantum skyrmion (Figure 5.7(f)). Although Ω_Δ changes discontinuously at $tJ = 2.81$, the spin expectation values and the wave function do not change discontinuously and the real-time evolution remains valid at this singular point. We also note that we did not observe a dynamical quantum phase transition here, which is accompanied by the nonanalytic behavior of the wave function [76].

To obtain this decay of quantum skyrmions, it is necessary that the two quantum skyrmions interact. By changing the gradient profile, it becomes possible for the two skyrmions to move in the same direction without interaction. Alternatively, starting with a single skyrmion state (achieved by optimizing the ground state RBM in the presence of large pinning fields such that only one skyrmion remains), the time evolution of a single skyrmion can be obtained. In both cases, we do not observe a quantum skyrmion decay. However, when two quantum skyrmions are driven towards each other, they collide, and this interaction leads to the formation of an exceptional configuration and deletion of a quantum skyrmion.

Finally, let us discuss the evolution of the Renyi entropy with time, shown in Figure 5.8. At $tJ = 0$, the entropy is low and concentrated between the two skyrmions,

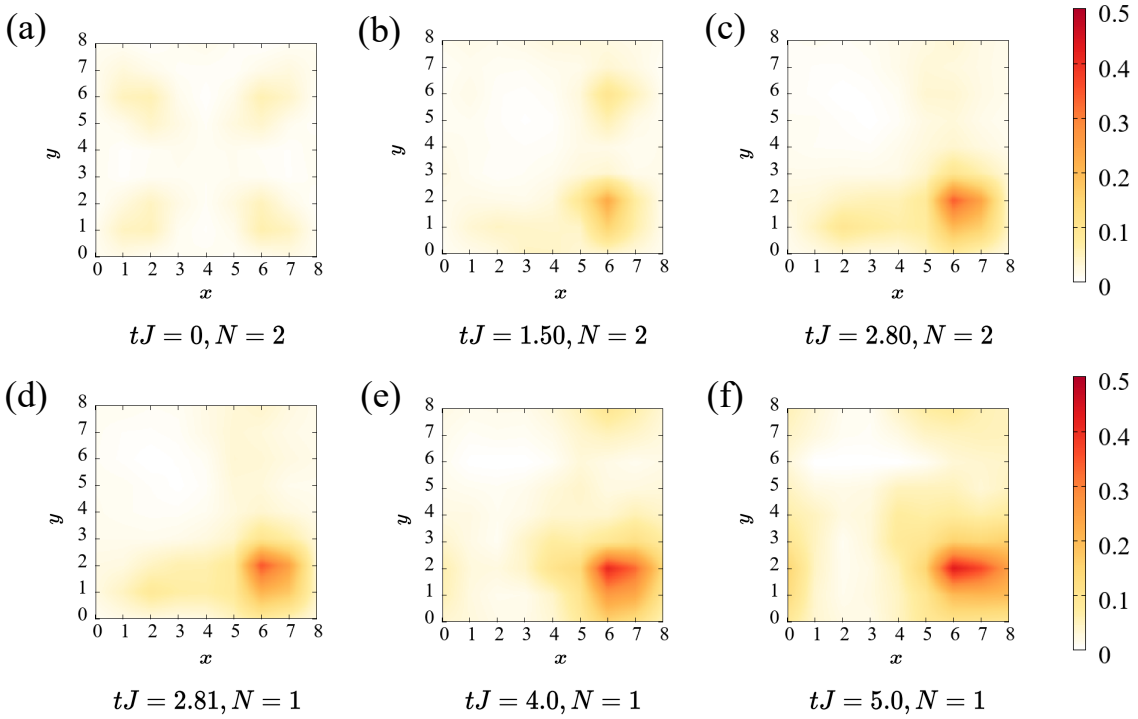


Figure 5.8: Evolution of Renyi entropy over time. The entropy increases when the quantum skyrmions interact [22].

which is shown in Figure 5.8(a). As the quantum skyrmions move toward each other, the entropy between them increases, reaching a maximum of $S_2 = 0.48$ at $tJ = 4.50$. The increase in entropy is due to the interaction between the two quantum skyrmions, and it increases continuously, even after one quantum skyrmion decays. The merging of two quantum skyrmions results in large entropy regions, demonstrating the necessity of quantum calculations to capture the correct behavior of this process.

5.5 Summary

In this chapter, we studied the ground state properties and real-time evolution of quantum skyrmions. Using variational Monte Carlo with a restricted Boltzmann machine as the variational ansatz, we obtained the ground state of a spin-1/2 Heisenberg model in the presence of Dzyaloshinskii-Moriya interaction and Heisenberg anisotropy. The ground state hosts a quantum skyrmion lattice with nonzero Renyi entropy and skyrmion number $N = 2$. The Renyi entropy is largest between the two quantum skyrmions. These quantum skyrmions can be manipulated by applying a magnetic field gradient. The quantum skyrmions move in a direction mostly perpendicular to the gradient, with a small parallel component. The velocity of quantum

skyrmions depends on the magnitude and direction of the gradient. An exceptional configuration with the topological charge of a meron is formed due to the interaction of the time-evolving quantum skyrmions, resulting in a quantum skyrmion decay as the skyrmion number $N = 2$ changes to $N = 1$. Thus, neural network quantum states can effectively approximate the real-time evolution of quantum skyrmions and reveal previously unknown quantum phenomena. Stabilizing longer-time evolution is an interesting aspect for future work.

Conclusion

In this thesis, we have investigated the properties of low-energy states and real time evolution of quantum skyrmion. We used variational Monte Carlo methods and artificial neural networks as the variational wave functions for studying system sizes that are outside the realm of exact diagonalization methods. We showed that a spin-1/2 Heisenberg Hamiltonian on a two-dimensional lattice with Dzyaloshinskii-Moriya interaction can host quantum skyrmions as ground states for a large range of Hamiltonian parameters. The spins in quantum skyrmions are entangled with each other, a property that cannot be observed in classical skyrmions. We showed that depending on the boundary conditions, the central spin of the quantum skyrmion can be disentangled from the rest of the spins. This can be important for establishing ways to measure quantum skyrmions without destroying the entanglement of the whole state. Then, using time-dependent variational principle, we studied the motion of quantum skyrmions and showed that external magnetic field gradients are effective in moving quantum skyrmions. As the quantum skyrmions move towards each other and interact, the entanglement between them increases greatly. This interaction eventually leads to the decay of quantum skyrmions. We also studied how the neural network learns the quantum skyrmion ground states and examined the limits of its representation power by pruning. Our work showed that artificial neural networks based variational Monte Carlo methods are effective in studying physics of large quantum magnetic systems.

Bibliography

[1] Paul Adrien Maurice Dirac and Ralph Howard Fowler. "Quantum mechanics of many-electron systems". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 123.792 (Jan. 1997), pp. 714–733. DOI: 10.1098/rspa.1929.0094.

[2] W. L. McMillan. "Ground State of Liquid ${\mathrm{He}}ˆ{4}$". In: *Physical Review* 138.2A (Apr. 1965), A442–A451. DOI: 10.1103/PhysRev.138.A442.

[3] Federico Becca and Sandro Sorella. "Variational Monte Carlo". In: *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge: Cambridge University Press, 2017, pp. 103–130. DOI: 10.1017/9781316417041.006.

[4] Steven R. White. "Density matrix formulation for quantum renormalization groups". In: *Physical Review Letters* 69.19 (Nov. 1992), pp. 2863–2866. DOI: 10.1103/PhysRevLett.69.2863.

[5] Frank Verstraete et al. "Density matrix renormalization group, 30 years on". In: *Nature Reviews Physics* 5.5 (May 2023), pp. 273–276. ISSN: 2522-5820. DOI: 10.1038/s42254-023-00572-5.

[6] U. Schollwöck. "The density-matrix renormalization group". In: *Reviews of Modern Physics* 77.1 (Apr. 2005), pp. 259–315. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.77.259.

[7] Antoine Georges et al. "Dynamical Mean-Field Theory of Strongly Correlated Fermion Systems and the Limit of Infinite Dimensions". In: *Rev. Mod. Phys.* 68 (Jan. 1996), pp. 13–125. DOI: 10.1103/RevModPhys.68.13.

[8] Antoine Georges. "Strongly Correlated Electron Materials: Dynamical Mean-Field Theory and Electronic Structure". In: *AIP Conference Proceedings* 715 (2004), pp. 3–74. ISSN: 0094243X. DOI: 10.1063/1.1800733.

[9] Federico Becca and Sandro Sorella. "Green's Function Monte Carlo". In: *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge: Cam-

bridge University Press, 2017, pp. 167–188. DOI: 10.1017/9781316417041.009.

[10] David Ceperley and Berni Alder. "Quantum Monte Carlo". In: *Science* 231.4738 (Feb. 1986), pp. 555–560. DOI: 10.1126/science.231.4738.555.

[11] W. M. C. Foulkes et al. "Quantum Monte Carlo simulations of solids". In: *Reviews of Modern Physics* 73.1 (Jan. 2001), pp. 33–83. DOI: 10.1103/RevModPhys.73.33.

[12] Rina Takashima, Hiroaki Ishizuka, and Leon Balents. "Quantum Skyrmions in Two-Dimensional Chiral Magnets". In: *Physical Review B* 94 (Oct. 2016), p. 134415. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.94.134415.

[13] Vivek Lohani et al. "Quantum Skyrmions in Frustrated Ferromagnets". In: *Physical Review X* 9.4 (Dec. 2019), p. 041063. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.9.041063.

[14] Pia Siegl et al. "Controlled Creation of Quantum Skyrmions". In: *Physical Review Research* 4.2 (May 2022), p. 023111. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.4.023111.

[15] J P Gauyacq and N Lorente. "A model for individual quantal nano-skyrmions". In: *Journal of Physics: Condensed Matter* 31.33 (Aug. 2019), p. 335001. ISSN: 0953-8984, 1361-648X. DOI: 10.1088/1361-648X/ab1f3a.

[16] Andreas Haller et al. "Quantum skyrmion lattices in Heisenberg ferromagnets". In: *Physical Review Research* 4.4 (Nov. 2022), p. 043113. DOI: 10.1103/PhysRevResearch.4.043113.

[17] O. M. Sotnikov et al. "Probing the topology of the quantum analog of a classical skyrmion". In: *Physical Review B* 103.6 (Feb. 2021), p. L060404. DOI: 10.1103/PhysRevB.103.L060404.

[18] Vladimir V. Mazurenko et al. "Estimating Patterns of Classical and Quantum Skyrmion States". In: *Journal of the Physical Society of Japan* 92.8 (Aug. 2023), p. 081004. ISSN: 0031-9015. DOI: 10.7566/JPSJ.92.081004.

[19] Kristian Maeland and Asle Sudbo. "Quantum fluctuations in the order parameter of quantum skyrmion crystals". In: *Physical Review B* 105.22 (June 2022), p. 224416. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.105.224416.

[20] Robert Peters, Jannis Neuhaus-Steinmetz, and Thore Posske. "Quantum skyrmion Hall effect in f-electron systems". In: *Physical Review Research* 5.3 (Sept. 2023), p. 033180. DOI: 10.1103/PhysRevResearch.5.033180.

[21] Ashish Joshi, Robert Peters, and Thore Posske. "Ground state properties of quantum skyrmions described by neural network quantum states". In: *Physical Review B* 108.9 (Sept. 2023), p. 094410. DOI: 10.1103/PhysRevB.108.094410.

[22] Ashish Joshi, Robert Peters, and Thore Posske. *Quantum skyrmion dynamics studied by neural network quantum states*. Mar. 2024. DOI: 10.48550/arXiv.2403.08184.

[23] Andreas Haller et al. *Quantum Magnetic Skyrmion Operator*. Apr. 2024. DOI: 10.48550/arXiv.2403.10347.

[24] Naoto Nagaosa and Yoshinori Tokura. "Topological properties and dynamics of magnetic skyrmions". In: *Nature Nanotechnology* 8.12 (Dec. 2013), pp. 899–911. ISSN: 1748-3395. DOI: 10.1038/nnano.2013.243.

[25] Martin Stier et al. "Skyrmion–Anti-Skyrmion Pair Creation by in-Plane Currents". In: *Physical Review Letters* 118.26 (June 2017), p. 267203. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.118.267203.

[26] S. Mühlbauer et al. "Skyrmion Lattice in a Chiral Magnet". In: *Science* 323.5916 (Feb. 2009), pp. 915–919. DOI: 10.1126/science.1166767.

[27] X. Z. Yu et al. "Real-space observation of a two-dimensional skyrmion crystal". In: *Nature* 465.7300 (June 2010), pp. 901–904. ISSN: 1476-4687. DOI: 10.1038/nature09124.

[28] Atsuhiro Kotani et al. "Observation of FeGe skyrmions by electron phase microscopy with hole-free phase plate". In: *AIP Advances* 8.5 (May 2018), p. 055216. DOI: 10.1063/1.5028398.

[29] Stefan Heinze et al. "Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions". In: *Nature Physics* 7.9 (Sept. 2011), pp. 713–718. ISSN: 1745-2481. DOI: 10.1038/nphys2045.

[30] N. S. Kiselev et al. "Chiral skyrmions in thin magnetic films: new objects for magnetic storage technologies?" In: *Journal of Physics D: Applied Physics* 44.39 (Sept. 2011), p. 392001. ISSN: 0022-3727. DOI: 10.1088/0022-3727/44/39/392001.

[31] Niklas Romming et al. "Writing and Deleting Single Magnetic Skyrmions". In: *Science* 341.6146 (Aug. 2013), pp. 636–639. DOI: 10.1126/science.1240573.

[32] Christoph Schutte et al. "Inertia, diffusion, and dynamics of a driven skyrmion". In: *Physical Review B* 90.17 (Nov. 2014), p. 174434. DOI: 10.1103/PhysRevB.90.174434.

[33] F. Jonietz et al. “Spin Transfer Torques in MnSi at Ultralow Current Densities”. In: *Science* 330.6011 (Dec. 2010), pp. 1648–1651. DOI: 10.1126/science.1195709.

[34] X. Z. Yu et al. “Skyrmion flow near room temperature in an ultralow current density”. In: *Nature Communications* 3.1 (Aug. 2012), p. 988. ISSN: 2041-1723. DOI: 10.1038/ncomms1990.

[35] R. Tomasello et al. “A strategy for the design of skyrmion racetrack memories”. In: *Scientific Reports* 4.1 (Oct. 2014), p. 6784. ISSN: 2045-2322. DOI: 10.1038/srep06784.

[36] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017.

[37] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[38] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL: probml.ai.

[39] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 25. 2012.

[40] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. June 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[41] Rick Merritt. *Why GPUs Are Great for AI*. URL: https://blogs.nvidia.com/blog/why-gpus-are-great-for-ai/.

[42] Giuseppe Carleo and Matthias Troyer. “Solving the Quantum Many-Body Problem with Artificial Neural Networks”. In: *Science* 355.6325 (Feb. 2017), pp. 602–606. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag2302.

[43] G. Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of Control, Signals and Systems* 2.4 (Dec. 1989), pp. 303–314. ISSN: 1435-568X. DOI: 10.1007/BF02551274.

[44] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural Networks* 4.2 (Jan. 1991), pp. 251–257. ISSN: 0893-6080. DOI: 10.1016/0893-6080(91)90009-T.

[45] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8.

[46] Moshe Leshno et al. “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function”. In: *Neural Networks* 6 (Jan. 1993), pp. 861–867. ISSN: 0893-6080. DOI: 10.1016/S0893-6080(05)80131-5.

[47] Zhou Lu et al. “The Expressive Power of Neural Networks: A View from the Width”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[48] Yusuke Nomura et al. “Restricted-Boltzmann-Machine Learning for Solving Strongly Correlated Quantum Systems”. In: *Physical Review B* 96.20 (Nov. 2017), p. 205152. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.96.205152.

[49] Yusuke Nomura, Nobuyuki Yoshioka, and Franco Nori. “Purifying Deep Boltzmann Machines for Thermal Quantum States”. In: *Physical Review Letters* 127.6 (Aug. 2021), p. 060601. DOI: 10.1103/PhysRevLett.127.060601.

[50] Kenny Choo et al. “Symmetries and many-body excited states with neural-network quantum states”. In: *Physical Review Letters* 121.16 (Oct. 2018), p. 167204. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.121.167204.

[51] Hiroki Saito and Masaya Kato. “Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice”. In: *Journal of the Physical Society of Japan* 87.1 (Jan. 2018), p. 014001. ISSN: 0031-9015, 1347-4073. DOI: 10.7566/JPSJ.87.014001.

[52] Attila Szabo and Claudio Castelnovo. “Neural network wave functions and the sign problem”. In: *Physical Review Research* 2.3 (July 2020), p. 033075. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.2.033075.

[53] Mohamed Hibat-Allah et al. “Recurrent neural network wave functions”. In: *Physical Review Research* 2.2 (June 2020), p. 023358. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.2.023358.

[54] Dmitrii Kochkov et al. “Learning ground states of quantum Hamiltonians with graph networks”. In: *arXiv:2110.06390 [cond-mat, physics:quant-ph]* (Oct. 2021).

[55] Xun Gao and Lu-Ming Duan. “Efficient representation of quantum many-body states with deep neural networks”. In: *Nature Communications* 8.1 (Sept. 2017), p. 662. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00705-2.

[56] Or Sharir, Amnon Shashua, and Giuseppe Carleo. "Neural tensor contractions and the expressive power of deep neural quantum states". In: *Physical Review B* 106.20 (Nov. 2022), p. 205136. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.106.205136.

[57] Koji Inui, Yasuyuki Kato, and Yukitoshi Motome. "Determinant-free fermionic wave function using feed-forward neural networks". In: *Physical Review Research* 3.4 (Nov. 2021), p. 043126. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.3.043126.

[58] Ao Chen et al. "Neural network evolution strategy for solving quantum sign structures". In: *Physical Review Research* 4.2 (May 2022), p. L022026. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.4.L022026.

[59] Markus Schmitt and Markus Heyl. "Quantum Many-Body Dynamics in Two Dimensions with Artificial Neural Networks". In: *Physical Review Letters* 125.10 (Sept. 2020), p. 100503. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.125.100503.

[60] T. H. R. Skyrme. "A unified field theory of mesons and baryons". In: *Nuclear Physics* 31 (Mar. 1962), pp. 556–569. ISSN: 0029-5582. DOI: 10.1016/0029-5582(62)90775-7.

[61] S. L. Sondhi et al. "Skyrmions and the crossover from the integer to fractional quantum Hall effect at small Zeeman energies". In: *Physical Review B* 47.24 (June 1993), pp. 16419–16426. DOI: 10.1103/PhysRevB.47.16419.

[62] Tin-Lun Ho. "Spinor Bose Condensates in Optical Traps". In: *Physical Review Letters* 81.4 (July 1998), pp. 742–745. DOI: 10.1103/PhysRevLett.81.742.

[63] David C. Wright and N. David Mermin. "Crystalline liquids: the blue phases". In: *Reviews of Modern Physics* 61.2 (Apr. 1989), pp. 385–432. DOI: 10.1103/RevModPhys.61.385.

[64] Suzuki Takao. "A study of magnetization distribution of submicron bubbles in sputtered Ho-Co thin films". In: *Journal of Magnetism and Magnetic Materials* 31-34 (Feb. 1983), pp. 1009–1010. ISSN: 0304-8853. DOI: 10.1016/0304-8853(83)90772-2.

[65] Tsuyoshi Okubo, Sungki Chung, and Hikaru Kawamura. "Multiple- q States and the Skyrmion Lattice of the Triangular-Lattice Heisenberg Antiferromagnet under Magnetic Fields". In: 108.1 (Jan. 2012), p. 017206. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.108.017206.

[66] Karin Everschor-Sitte and Matthias Sitte. *The vector field of two-dimensional magnetic skyrmions: a) a hedgehog skyrmion and b) a spiral skyrmion.* URL: https://en.wikipedia.org/wiki/Magnetic_skyrmion#/media/File:2skyrmions.PNG.

[67] Albert Fert, Nicolas Reyren, and Vincent Cros. "Magnetic skyrmions: advances in physics and potential applications". In: *Nature Reviews Materials* 2.7 (June 2017), pp. 1–15. ISSN: 2058-8437. DOI: 10.1038/natrevmats.2017.31.

[68] B. Berg and M. Lüscher. "Definition and statistical distributions of a topological number in the lattice O(3) σ-model". In: *Nuclear Physics B* 190.2 (Aug. 1981), pp. 412–424. ISSN: 0550-3213. DOI: 10.1016/0550-3213(81)90568-X.

[69] Christina Psaroudaki and Christos Panagopoulos. "Skyrmion Qubits: A New Class of Quantum Logic Elements Based on Nanoscale Magnetization". In: *Physical Review Letters* 127.6 (Aug. 2021), p. 067201. DOI: 10.1103/PhysRevLett.127.067201.

[70] Michael J. Donahue. "OOMMF User's Guide, Version 1.0". In: *NIST* (Sept. 1999).

[71] R. Antos and Y. Otani. "The dynamics of magnetic vortices and skyrmions". In: *Spin Current*. Ed. by Sadamichi Maekawa et al. Oxford University Press, Sept. 2017. ISBN: 978-0-19-878707-5. DOI: 10.1093/oso/9780198787075.003.0022.

[72] Wanjun Jiang et al. "Direct observation of the skyrmion Hall effect". In: *Nature Physics* 13.2 (Feb. 2017), pp. 162–169. ISSN: 1745-2481. DOI: 10.1038/nphys3883.

[73] Roméo Juge et al. "Current-Driven Skyrmion Dynamics and Drive-Dependent Skyrmion Hall Effect in an Ultrathin Film". In: *Physical Review Applied* 12.4 (Oct. 2019), p. 044007. ISSN: 2331-7019. DOI: 10.1103/PhysRevApplied.12.044007.

[74] Seonghoon Woo et al. "Current-driven dynamics and inhibition of the skyrmion Hall effect of ferrimagnetic skyrmions in GdFeCo films". In: *Nature Communications* 9.1 (Mar. 2018), p. 959. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03378-7.

[75] Christina Psaroudaki et al. "Quantum Dynamics of Skyrmions in Chiral Magnets". In: *Physical Review X* 7.4 (Nov. 2017), p. 041045. DOI: 10.1103/PhysRevX.7.041045.

[76] Vipin Vijayan et al. “Topological dynamical quantum phase transition in a quantum skyrmion phase”. In: *Physical Review B* 107.10 (Mar. 2023), p. L100419. DOI: 10.1103/PhysRevB.107.L100419.

[77] Fabio Salvati et al. “Stability of a quantum skyrmion: Projective measurements and the quantum Zeno effect”. In: *Physical Review B* 109.6 (Feb. 2024), p. 064409. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.109.064409.

[78] James Gubernatis, Naoki Kawashima, and Philipp Werner. “Fermion ground state methods”. In: *Quantum Monte Carlo Methods: Algorithms for Lattice Models*. Cambridge University Press, 2016, pp. 338–364.

[79] Anna Dawid et al. *Modern applications of machine learning in quantum sciences*. Nov. 2023. DOI: 10.48550/arXiv.2204.04198.

[80] Kenneth G. Wilson. “The renormalization group: Critical phenomena and the Kondo problem”. In: *Reviews of Modern Physics* 47.4 (Oct. 1975), pp. 773–840. DOI: 10.1103/RevModPhys.47.773.

[81] Ralf Bulla, Theo A. Costi, and Thomas Pruschke. “The numerical renormalization group method for quantum impurity systems”. In: *Reviews of Modern Physics* 80.2 (Apr. 2008), pp. 395–450. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.80.395.

[82] James Gubernatis, Naoki Kawashima, and Philipp Werner. *Quantum Monte Carlo Methods: Algorithms for Lattice Models*. Cambridge University Press, 2016.

[83] Román Orús. “A practical introduction to tensor networks: Matrix product states and projected entangled pair states”. In: *Annals of Physics* 349 (Oct. 2014), pp. 117–158. ISSN: 0003-4916. DOI: 10.1016/j.aop.2014.06.013.

[84] Jacob Biamonte and Ville Bergholm. *Tensor Networks in a Nutshell*. July 2017. DOI: 10.48550/arXiv.1708.00006.

[85] Federico Becca and Sandro Sorella. “Correlated Models and Wave Functions”. In: *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017, pp. 3–36.

[86] F. Verstraete, V. Murg, and J.I. Cirac. “Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems”. In: *Advances in Physics* 57.2 (Mar. 2008), pp. 143–224. ISSN: 0001-8732. DOI: 10.1080/14789940801912366.

[87] Pmigdal. *Penrose graphical notation (tensor diagram notation) of a matrix product state of five particles.* URL: https://en.wikipedia.org/wiki/Matrix_product_state#/media/File:Matrix_product_state_obc_tikz.svg.

[88] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (Nov. 1992), pp. 457–472. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177011136.

[89] Aki Vehtari et al. "Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion)". In: *Bayesian Analysis* 16.2 (June 2021), pp. 667–718. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/20-BA1221.

[90] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[91] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* May 2019. DOI: 10.48550/arXiv.1810.04805.

[92] Aditya Ramesh et al. *Zero-Shot Text-to-Image Generation.* Feb. 2021.

[93] Egm4313.s12. *Neuron and myelinated axon, with signal flow from inputs at dendrites to outputs at axon terminals.* URL: https://en.wikipedia.org/wiki/Biological_neuron_model#/media/File:Neuron3.png.

[94] Alessandro Sinibaldi et al. *Unbiasing time-dependent Variational Monte Carlo by projected quantum evolution.* May 2023.

[95] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. "Transformer Variational Wave Functions for Frustrated Quantum Spin Systems". In: *Physical Review Letters* 130.23 (June 2023), p. 236401. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.130.236401.

[96] Or Sharir, Garnet Kin-Lic Chan, and Anima Anandkumar. *Towards Neural Variational Monte Carlo That Scales Linearly with System Size.* Dec. 2022.

[97] Javier Robledo Moreno et al. "Fermionic Wave Functions from Neural-Network Constrained Hidden States". In: *Proceedings of the National Academy of Sciences* 119.32 (Aug. 2022), e2122059119. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2122059119.

[98] Kenny Choo, Titus Neupert, and Giuseppe Carleo. "Two-dimensional frustrated ${J}_{1}\text{\ensuremath{-}}{J}_{2}$ model studied with neural network quantum states". In: *Physical Review B* 100.12 (Sept. 2019), p. 125124. DOI: 10.1103/PhysRevB.100.125124.

[99] Luciano Loris Viteritti, Francesco Ferrari, and Federico Becca. "Accuracy of restricted Boltzmann machines for the one-dimensional J_1-J_2 Heisenberg model". In: *SciPost Physics* 12.5 (May 2022), p. 166. ISSN: 2542-4653. DOI: 10.21468/SciPostPhys.12.5.166. (Visited on 06/26/2024).

[100] Giammarco Fabiani and Johan H. Mentink. "Investigating ultrafast quantum magnetism with machine learning". In: *SciPost Physics* 7.1 (July 2019), p. 004. ISSN: 2542-4653. DOI: 10.21468/SciPostPhys.7.1.004.

[101] Douglas Hendry and Adrian E. Feiguin. "Machine learning approach to dynamical properties of quantum many-body systems". In: *Physical Review B* 100.24 (Dec. 2019), p. 245123. DOI: 10.1103/PhysRevB.100.245123.

[102] Yusuke Nomura and Masatoshi Imada. "Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy". In: *Physical Review X* 11.3 (Aug. 2021), p. 031034. DOI: 10.1103/PhysRevX.11.031034.

[103] Vicente Oyanedel M. *1D convolutional neural network feed forward example*. URL: https://en.wikipedia.org/wiki/Convolutional_neural_network#/media/File:1D_Convolutional_Neural_Network_feed_forward_example.png.

[104] Christopher Roth, Attila Szabo, and Allan H. MacDonald. "High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks". In: *Physical Review B* 108.5 (Aug. 2023), p. 054410. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.108.054410.

[105] fdeloche. *Compressed (left) and unfolded (right) basic recurrent neural network*. URL: https://en.wikipedia.org/wiki/Recurrent_neural_network#/media/File:Recurrent_neural_network_unfold.svg.

[106] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. "Quantum Entanglement in Neural Network States". In: *Physical Review X* 7.2 (May 2017), p. 021021. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.7.021021.

[107] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 2017.

[108] Filippo Vicentini et al. "NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems". In: *SciPost Physics Codebases* (Aug. 2022), p. 007. ISSN: 2949-804X. DOI: 10.21468/SciPostPhysCodeb.7.

[109] Giuseppe Carleo et al. "NetKet: A machine learning toolkit for many-body quantum systems". In: *SoftwareX* 10 (July 2019). ISSN: 2352-7110. DOI: 10.1016/j.softx.2019.100311.

[110] Dion Häfner and Filippo Vicentini. "mpi4jax: Zero-copy MPI communication of JAX arrays". In: *Journal of Open Source Software* 6.65 (2021), p. 3419. DOI: 10.21105/joss.03419.

[111] Matthew B. Hastings et al. "Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations". In: *Physical Review Letters* 104.15 (Apr. 2010), p. 157201. DOI: 10.1103/PhysRevLett.104.157201.

[112] Tien-Ju Yang, Yu-Hsin Chen, and Vivienne Sze. "Designing Energy-Efficient Convolutional Neural Networks Using Energy-Aware Pruning". In: IEEE Computer Society, July 2017, pp. 6071–6079. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.643.

[113] Song Han, Huizi Mao, and William J. Dally. *Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding*. Feb. 2016.

[114] Davis Blalock et al. *What is the State of Neural Network Pruning?* Mar. 2020.

[115] Anna Golubeva and Roger G. Melko. "Pruning a restricted Boltzmann machine for quantum state reconstruction". In: *Physical Review B* 105.12 (Mar. 2022), p. 125124. DOI: 10.1103/PhysRevB.105.125124.

[116] David Schwandt, Fabien Alet, and Sylvain Capponi. "Quantum Monte Carlo Simulations of Fidelity at Magnetic Quantum Phase Transitions". In: *Physical Review Letters* 103.17 (Oct. 2009), p. 170501. DOI: 10.1103/PhysRevLett.103.170501.

[117] Damian Hofmann et al. "Role of stochastic noise and generalization error in the time propagation of neural-network quantum states". In: *SciPost Physics* 12.5 (May 2022), p. 165. ISSN: 2542-4653. DOI: 10.21468/SciPostPhys.12.5.165.

[118] S. L. Zhang et al. "Manipulation of skyrmion motion by magnetic field gradients". In: *Nature Communications* 9.1 (May 2018), p. 2115. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04563-4.

[119] Stavros Komineas and Nikos Papanicolaou. "Skyrmion dynamics in chiral ferromagnets". In: *Physical Review B* 92.6 (Aug. 2015), p. 064412. ISSN: 1098-0121, 1550-235X. DOI: 10.1103/PhysRevB.92.064412.

[120] Karin Everschor et al. "Rotating skyrmion lattices by spin torques and field or temperature gradients". In: *Physical Review B* 86.5 (Aug. 2012), p. 054432. ISSN: 1098-0121, 1550-235X. DOI: 10.1103/PhysRevB.86.054432.

[121] Kaelan Donatella et al. “Dynamics with autoregressive neural quantum states: Application to critical quench dynamics”. In: *Physical Review A* 108.2 (Aug. 2023), p. 022210. DOI: 10.1103/PhysRevA.108.022210.

Acknowledgments

First of all, I thank my supervisor Dr. Robert Peters, without whom this work would not have been possible. He was always present to answer my questions with patience, and taught me many things about physics, numerical methods and establishing good research practices. During my 5 years of graduate study, I have learned many invaluable lessons from him, both in research and life, and I will always be grateful to him for his support and guidance.

I also thank our collaborator, Dr. Thore Posske, who hosted me in Hamburg in 2022 and provided many fruitful insights and discussions which were instrumental in forming the two papers that are the basis of this thesis.

I thank all the professors in the Condensed Matter Theory group for their guidance over the last 5 years. I have learned many things about physics from them. I also thank my colleagues in the group, for many discussions in physics and outside of it.

I am grateful for the financial support for my graduate studies from the MEXT scholarship by Japanese government. I also thank the Graduate School of Science for the research grant 'Ginfu Fund' in 2022.

Studying in a foreign country can get lonely sometimes and I appreciate all my friends in Kyoto, some of them still here, for making it feel like a second home.

I thank Shrinjini for always being present, supporting and encouraging me from afar, the only constant in a sea of variables.

I am most grateful to my parents, who always supported me and made it possible so that I can fulfill my dream of studying physics. I also thank my sister Arpna for making my life a lot easier.